5-2023

Structuring AI Teammate Communication: An Exploration of AI's Communication Strategies in Human-AI Teams

Rui Zhang
rzhang2@clemson.edu

Structuring AI Teammate Communication: An Exploration of AI's Communication Strategies in Human-AI Teams

---

A Dissertation
Presented to
the Graduate School of
Clemson University

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy
Human-Centered Computing

---

by
Rui Zhang
May 2023

Accepted by:
Dr. Nathan McNeese, Committee Chair
Dr. Guo Freeman
Dr. Bart Knijnenburg
Dr. Kapil Chalil Madathil

Abstract

In the past decades, artificial intelligence (AI) has been implemented in various domains to facilitate humans in their work, such as healthcare and the automotive industry. Such application of AI has led to increasing attention on human-AI teaming, where AI closely collaborates with humans as a teammate. AI as a teammate is expected to have the ability to coordinate with humans by sharing task-related information, predicting other teammates' behaviors, and progressing team tasks accordingly. To complete these team activities effectively, AI teammates must communicate with humans, such as sharing updates and checking team progress.

Even though communication is a core element of teamwork that helps to achieve effective coordination, how to design and structure human-AI communication in teaming environments still remains unclear. Given the context-dependent characteristics of communication, research on human-AI teaming communication needs to narrow down and focus on specific communication elements/components, such as the proactivity of communication and communication content. In doing so, this dissertation explores how AI teammates' communication should be structured by modifying communication components through three studies, each of which details a critical component of effective AI communication: (1) communication proactivity, (2) communication content (explanation), and (3) communication approach (verbal vs. non-verbal). These studies provide insights into how AI teammates' communication

can be integrated into teamwork and how to design AI teammate communication in human-AI teaming.

Study 1 explores an important communication element, *communication proactivity*, and its impact on team processes and team performance. Specifically, communication proactivity in this dissertation refers to whether an AI teammate proactively communicates with human teammates, i.e., proactively pushing information to human teammates. Experimental analysis shows that AI teammates' proactive communication plays a crucial role in impacting human perceptions, such as perceived teammate performance and satisfaction with the teammate. Importantly, teams with a non-proactive communication AI teammate increase team performance more than teams with a proactive communication AI as the human and the AI collaborate more. This study identifies the positive impact of AI being proactive in communication at the initial stage of task coordination, as well as the potential need for AI's flexibility in their communication proactivity (i.e., once human and AI teammates' coordination pattern forms, AI can be non-proactive in communication).

Study 2 examines *communication content* by focusing on AI's explanation and its impact on human perceptions in teaming environments. Results indicate that AI's explanation, as part of communication content, does not always positively impact human trust in human-AI teaming. Instead, the impact of AI's explanations on human perceptions depends on specific collaboration scenarios. Specifically, AI's explanations facilitate trust in the AI teammate when explaining why AI disobeys humans' orders, but hinder trust when explaining why AI lies to humans. In addition, AI giving an explanation of why they ignored the human teammate's injury was perceived to be more effective than AI not providing such an explanation. The findings emphasize the context-dependent characteristic of AI's communication content with a focus on AI's explanation of their actions.

Study 3 investigates AI's *communication approach*, which was manipulated as verbal vs. non-verbal communication. Results indicate that AI teammates' verbal/non-verbal communication does not impact human trust in the AI teammate, but facilitates the maintenance of humans' situation awareness in task coordination. In addition, AI with non-verbal communication is perceived as having lower communication quality and lower performance. Importantly, AI with non-verbal communication has better team performance in human-human-AI teams than human-AI-AI teams, whereas AI with verbal communication has better team performance in human-AI-AI teams than human-human-AI teams.

These three studies together address multiple research gaps in human-AI team communication and provide a holistic view of the design and structure of AI's communication by examining three specific aspects of communication in human-AI teaming. In addition, each study in this dissertation proposes practical design implications on AI's communication in human-AI teams, which will assist AI designers and developers to create better AI teammates that facilitate humans in teaming environments.

Acknowledgments

First, I would like to express gratitude to my advisor Dr. Nathan McNeese for his continuous support and help throughout my Ph.D. research. Thank you for giving me the opportunity to study in HCC at Clemson. I appreciate your continuous and helpful guidance in the past five years. It would not be possible to be where I am in the Ph.D. journey without your constructive feedback on this dissertation.

I want to give many thanks to my committee members, Dr. Guo Freeman, Dr. Bart Knijnenburg, and Dr. Kapil Chalil Madathil. I still remember the first semi-structured interview project I worked on with Dr. Freeman and how she walked us through the steps in detail. I sincerely appreciate Dr. Freeman's help and feedback on multiple projects we collaborated on in the past five years. I learned about all the statistical analysis methods from Dr. Bart Knijnenburg. I want to express my sincere gratitude to him for his time and help on the study design and data analysis of my dissertation studies, as well as answering my endless questions. I am very grateful for Dr. Kapil Madathil's support through my dissertation process and all the feedback he provided during my proposal presentation.

To the TRACE lab members, it has been a joy working with all of you. I have seen how TRACE has grown from a small research group to the current large research family. Everyone in the group has been so supportive in both research and personal life. I want to thank the lab members who have collaborated with me, and who have been in the lab for years. Chris Flathmann, thank you for being an example and providing selfless help through my dissertation. It would not come along that well without your help. Geoff Musick, thank you for all the excellent writing in our collaboration and for having my back whenever I need it. Beau Schelble, thank you for your help and motivation on this Ph.D. journey. Rohit Mallick, thank you for setting up the TRACE lab activities and being so caring all the time. A special thank you to Joshua Little, who made an excellent contribution to the research platforms used in this dissertation study.

To my friends, I would like to give many thanks for your support. Particularly, Jiarui, thank you for being a great example and motivation for me to work on my first CSCW publication when Covid hit in 2020. Zilan, thank you for bringing so much joy to my life and always having my back. Wen, it has been a pleasure knowing you, hanging out with you, and working with you. Yuting and Wanfang, thank you for being caring and encouraging. I also want to express my special thanks to Roxie and Doodle. Roxie, you had been such a lovely and sweet dog. I miss you a lot. Doodle, it has been lovely seeing you growing up as a cute cat. You are always so cute and it has been wonderful playing with you.

Finally, to my family, my partner Jake, and his family, I would not be who I am today without your care, help, and support. This has been a long journey, and your support truly brings more joy to it. To my mom and dad, you may not know how to do research, but you always know how to give me the best support. Jake, I'm so glad to have your company on this journey. Thank you for supporting me when I thought I couldn't make it. Your optimism saved me from the darkness and uncertainty in this journey and inspired me to be the best version of myself. To Mary and Jody, thank you so much for always being caring and considerate, and making me feel at home. Thank you for treating me as a family member before I become one.

One day, you will look back and see that all along, you were blooming.

-Morgan Harper Nichols

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

In recent years, artificial intelligence (AI) has been implemented in various domains to coordinate with humans towards shared goals, which has led research to the development and study of *human-AI teams*. These human-AI teams are expected to make use of AI's advantages in fast calculating and reliable performance and humans' advantages in flexibility and cognitive thinking ability to reach a potential that human-only teams are unable to achieve [172]. However, achieving this potential requires human-AI teams to possess teaming capabilities, including effective team communication, which is a cornerstone of high-performance teamwork [160] and a current limitation in human-AI teams [75, 301]. The recent release of ChatGPT pushes AI agents one step closer to achieving smooth human-AI communication. However, there is still a gap between communication in human-AI interaction and communication in **human-AI teams**. Teaming environments usually have higher requirements on team members regarding how they communicate to coordinate and complete shared goals in team tasks. To ensure that human-AI teams meet or surpass human-only teams and achieve their full potential, research must guarantee AI teammates' communication is appropriately designed and structured in **human-AI**

teaming environments [301], which is the goal of this dissertation.

1.1 Problem Motivation

Communication serves as a critical element in building interpersonal relationships, facilitating team processes (e.g., trust and situation awareness), and creating effective outcomes (e.g., team performance and team viability) in teams [99, 174, 160]. Previous research suggests that communication enhances team performance both directly [226] and indirectly by supporting the development of cognitive team process (e.g., team cognition) [221] in human-only teams. In particular, various communication media is usually utilized to support effective coordination in virtual teaming environments [107] due to constraints such as lacking non-verbal cues and visual presence [236, 180, 196]. While communication in human-only teams has been thoroughly explored, communication in human-AI teams is still under-explored. Multiple studies in computer-supported cooperative work (CSCW) and human-computer interaction (HCI) have utilized human-AI communication in some way in their research design [297, 194], even though the research focus of these studies was not communication in a teaming environment. All of these studies on both human-human team communication and human-AI collaboration emphasize and highlight the indispensable role and need of communication in *human-AI teaming*. With the recent release of ChatGPT, AI has a substantially better capability for natural language processing (NLP), making it possible to have human-AI communication that is similar to human-human communication in *human-AI interaction*. In spite of the fact that communication between team members is integral to most team processes, AI lacking flexibility in coordination in dynamic teaming environments makes it difficult to implement team-based human-AI communication. Unlike humans who have the cognitive ability to make decisions on when to initiate a conversation during collaboration, what content to cover in the conversation to facilitate coordination, and how to present/share information in these conversations to make progress on team tasks, AI teammates rely on AI designers and developers to integrate the communication component into AI's team-based algorithms.

Researchers have endeavored to develop NLP that AI can apply to understand and communicate with humans in the past two decades (e.g., ChatGPT) [32, 51, 114, 159]. However, understanding this natural language communication from a computational perspective does not ensure that AI teammates can communicate well with humans in **team** tasks. Specifically, teamwork requires AI teammates to autonomously decide to communicate the right information with humans at the opportune time. In doing so, AI teammates need to employ certain communication strategies on when to communicate, what to communicate, and how to communicate, to facilitate coordination with humans. One approach to developing such communication strategies for AI teammates is to modify various communication components (e.g., communication proactivity, communication content) in human-AI teams. Thus, how these communication components can facilitate human-AI coordination and how they are perceived by humans need to be studied to provide insights into how AI's communication should be structured in teaming environments to achieve prominent team performance.

In summary, the motivations of this dissertation are twofold: (1) communication is an indispensable element of teamwork in ensuring effective and smooth coordination in virtual human-AI teams; (2) AI's communication in teaming environments has yet to be explored regarding how it should be structured with NLP to lead to smooth human-AI coordination. Thus, the problem that this dissertation solves is: *AI teammates' communication is important to effective teamwork but yet to* *be explored due to a large past focus on AI's NLP abilities and not human-AI team communication requirements and communication strategies.*

1.2 Research Motivation

1.2.1 Communication in Human-AI Teaming

The continuous development of AI technologies has led to increasing interest in human-AI teams in the past decade [171, 301, 175]. A large body of work has explored human-AI teams from various perspectives, such as team process (e.g., trust, situation awareness) [170, 183], and team outcomes (e.g., team performance) [19, 169]. For example, existing work has identified multiple factors that impact team performance, such as AI's behaviors and performance [233, 298], level of autonomy [200], and task characteristics [200]. In addition, previous work indicates that team performance [170], individual differences (e.g., pre-existing attitudes towards AI) [301], and AI's capabilities (e.g., accuracy, explainability) [19, 20, 75] can impact human trust in the AI teammate.

One important element that serves as a foundation of human-AI coordination in human-AI teaming is the communication between humans and AI teammates. Previous work on virtual human-human teams has proposed a conceptual model where communication impacts both emergent states (e.g., trust and team cognition) and team outputs (e.g., team viability, performance, and satisfaction) [161]. While a considerable amount of research on human-only teamwork has explored how communication shapes team processes and team outcomes [160, 283], how communication facilitates team process in *human-AI teams* is still understudied. The limited existing work on human-AI communication in teams has only explored how communication quantity and frequency impacted team performance in human-AI teams [61, 63]. More importantly, even though these studies explored the relationship between the communication characteristic and object team performance in human-AI teams, how communication facilitates the coordination process between humans and AI teammates remains unclear. Achieving an in-depth understanding of this dynamic process impacted by human-AI communication is particularly important in building effective human-AI teams. In doing so, more research is in need to explore how AI's communication impacts the coordination process and how it facilitates the development of cognitive and affective team processes. Inspired by theoretical frameworks of communication in virtual human-human teams in previous literature, this dissertation intends to explore how communication in human-AI teams could facilitate the collaboration process. Thus, the following research gap is identified:

> *There is a lack of research on how AI's communication impacts team processes through human-AI coordination and how team processes could be improved using AI's communication.*

Given the importance of communication in teamwork, especially computer-supported collaboration, a large body of work has explored various aspects of communication in human-only teams. However, little is known about communication in *human-AI teams*. While the current state-of-the-art AI technology has been able to allow AI to participate in natural language communication with humans, such as ChatGPT, this has yet to be extended and applied in a *team* setting. In fact, research on communication with AI in teams, in general, is just starting to emerge [95]. The limited empirical research investigating communication in human-AI teams has only peripherally examined the quantity and frequency of communication with the AI using scripted and restricted inventory [61, 63, 172], the directionality of communication

[12], and implicit communication [146] with the AI using restricted inventory. For instance, prior work indicates that implicit communication between humans and AI teammates is an effective approach for human-AI teams to achieve high team performance in cooperative games [146]. Another research study examined AI's prediction model with two different communication directionality (human-to-AI communication, and AI-to-human communication) in a cooperative game and found that communication directionality can change human perceptions of the AI teammate [12]. Even though these studies have generated an initial understanding of several communication components, how some other communication components (e.g., communication proactivity) impact human-AI teaming still remains unclear. More research is needed to better inform how AI's communication should be structured for effective human-AI coordination. In doing so, it is important to investigate communication components such as AI's proactivity in communication and information presentation approach (verbal vs. non-verbal cues) in virtual teaming environments. Thus, the following two research gaps still exist:

> *Little research specifically focuses on how to design and structure AI's communication to achieve effective human-AI coordination and high team performance.*
>
> *Most of the existing work on human-AI communication focuses on communication quantity or frequency, leaving many other communication components unexplored, such as AI's communication proactivity and communication approach, and their impact on coordination in human-AI teaming.*

1.3 Research Questions and Gaps

This dissertation answers a list of research questions regarding human-AI communication in virtual teaming environments intending to shed light on the design of effective AI's communication in human-AI teams. The research questions are listed as follows, in which RQ0 serves as an umbrella research question with the other three research questions falling under it:

- **RQ0**: How should AI teammates' communication strategies be designed and developed to achieve effective and smooth human-AI coordination in teaming environments?

- **RQ1**: How does AI teammates' communication facilitate team processes (e.g., trust and situation awareness) through human-AI coordination in teaming environments?

- **RQ2**: How does AI teammates' communication impact team performance in virtual human-AI teams?

- **RQ3**: How do humans perceive and interpret AI teammates' communication in human-AI teams?

Figure 1.1 presents how the three research questions (RQ1, RQ2, RQ3) and the umbrella research question (RQ0) are posited in human-AI teaming. These questions depict how AI's communication impacts the coordination in human-AI teams from various perspectives, including team processes, team outcomes, and human perceptions and interpretations of AI's communication. This dissertation informs how AI's communication should be structured using three communication components (RQ0), including communication proactivity, communication content, and communication approach (verbal vs. non-verbal communication). In doing so, each dissertation study explores how one communication component impacts the human-AI coordination process in detail using experiments with a practical context of human-AI teaming. Specifically, this dissertation identifies how AI's communication facilitates team processes, such as trust development (RQ1), and how AI's communication impacts human-AI team outcomes, such as objective team performance (RQ2). In addition to teaming concepts, it is particularly important to understand how humans perceive and interpret AI's communication (RQ3), which plays a crucial role in shaping the coordination between humans and AI teammates and team outcomes [301]. Taken together, this dissertation's research questions provide a holistic understanding of how AI's communication is posited in coordination within a teaming environment, which will serve as a solid foundation for future research to further explore human-AI communication in teaming environments.

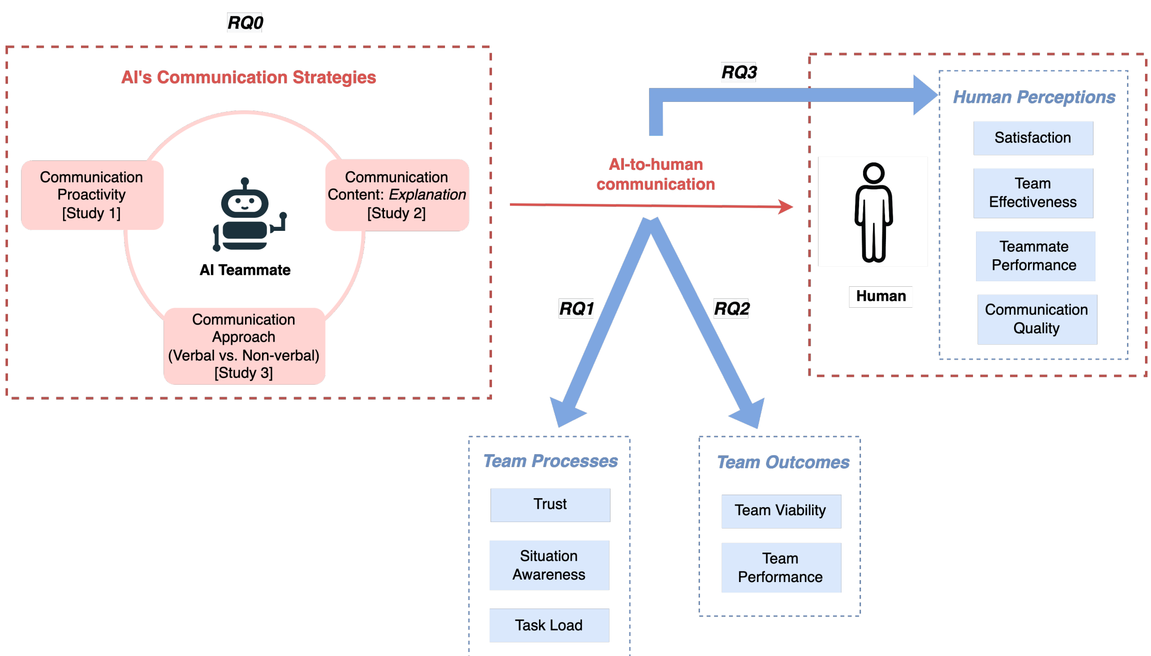


Figure 1.1: Overview of Research Questions

These research questions close the research gaps and build a solid foundation for research on AI's communication in human-AI teams. Table 1.1 presents how these research questions relate to the existing gaps that were identified in the previous subsection.

Table 1.1: Connecting Research Questions and Research Gaps.

Research Gaps	**Research Questions**
Little research specifically focuses on how to design and structure AI's communication to achieve effective human-AI coordination and high team performance.	RQ1, RQ2
There is a lack of research on how AI's communication impacts team processes through human-AI coordination and how team processes could be improved using AI's communication.	RQ1, RQ3
Most of the existing work on human-AI communication focuses on communication quantity or frequency, leaving many other communication components unexplored, such as AI's communication proactivity and communication approach, and their impact on coordination in human-AI teaming.	RQ1, RQ2, RQ3

1.4 Summary of Studies

This dissertation answers the research questions above through three studies, each of which explores one communication component of effective human-AI communication: communication proactivity, communication content, and communication approach (verbal/non-verbal). Each study will be briefly introduced below. Table

1.2 outlines the relationship between the research questions this dissertation addresses and the three dissertation studies.

Table 1.2: Study Summary

Study	**Research Questions**
Study 1: AI's Communication Proactivity	RQ1, RQ2, RQ3
Study 2: AI's Explanations of Their Actions (Communication Content)	RQ1, RQ3
Study 3: AI's Verbal vs. Non-verbal Communication	RQ1, RQ2, RQ3

1.4.1 Study 1: AI Teammates' Communication Proactivity

Study 1 focuses on understanding how AI's proactive/non-proactive communication (i.e., AI proactively pushing information to humans vs. AI only replying humans' messages), as part of an AI's communication strategy, is perceived and how it impacts human-AI coordination in a multiplayer gaming environment using a mixed-method design with 60 participants. This study shows that proactive AI communication has an overall positive impact on human perceptions, but non-proactive communication has a more positive impact during the later stages of teaming. Specifically, teams with a non-proactive communication AI teammate perform better than teams with proactive communication in the later staging of teaming (RQ2), possibly due to the fact that a collaboration pattern has formed. Other experimental results show that: (1) humans trust the AI teammate more when proactive communication is applied in comparison to when non-proactive communication is applied; (2) the proactive communication AI teammate is perceived to have a higher performance than the non-proactive AI teammate; (3) humans are more satisfied with the proactive communication than the non-proactive communication AI (RQ3).

In addition to the experiment results, our interviews show that AI teammates' proactive communication with humans could facilitate the development of human trust and situation awareness, whereas AI lacking such proactive communication is often not perceived as a teammate (RQ1). Moreover, this study highlights four communication strategies AI teammates should apply to support their coordination with humans in teaming environments: (1) proactively communicating with humans; (2) employing a balanced communication with both efficiency and sociability; (3) providing quick responses; and (4) avoiding providing a large amount of communication once communication pattern has formed in repeated team tasks.

1.4.2 Study 2: AI's Explanations of Their Actions in Communication

Study 2 explores how humans perceive AI's explanation of their behaviors, which is a part of communication content, and how such an explanation impacts human trust and the perceived effectiveness of the AI teammate. An online experiment with 156 participants was conducted to examine how AI communicating the rationality of their actions impacts people's trust in and perceived effectiveness of the teammate, especially compared to humans providing the same explanations. This impact was explored in four scenarios: human/AI teammate ignoring potential human (i.e., participant) death, human/AI teammate ignoring human injury, human/AI teammate disobeying/disagreeing with participants' order, and lying to the participant. This study shows that AI communicating their actions' logic has an impact on human trust of AI teammates, but this impact differs in various scenarios. Specifically, AI's explanations in communication facilitate human trust in the AI teammate in the disobeying scenario but hinder trust in the lying scenario (RQ1). In addition, individual differences (e.g., gender and each individual's ethical framework) impact participants' perception of AI teammates both directly and indirectly in different scenarios (RQ3). This study indicates that AI communicating their actions' logic depends on specific scenarios, and provides insights into the future direction of AI's explanations in communication in teaming environments.

1.4.3 Study 3: AI's Verbal vs. Non-verbal Communication

Study 3 explored how AI using different communication approaches impacts team processes, human perceptions, and team outcomes under two different team compositions (i.e., human-human-AI teams and human-AI-AI teams). An experiment was conducted with 100 participants, leading to 50 human-human-AI teams, and 100 human-AI-AI teams. The results present four main findings. First, AI's verbal/non-verbal communication does not have a significant impact on human trust (RQ1). Second, both verbal and non-verbal communication facilitates the development and maintenance of situation awareness, but non-verbal communication is not effective in supporting this process (RQ1). Third, compared to AI with verbal communication, AI with non-verbal communication is perceived to perform worse and has lower communication quality (RQ3). That could be due to the fact that AI's non-verbal communication does not align with humans' traditional definition of communication. Thus, AI's non-verbal communication should not be applied as the only communication channel in HATs. Fourth, AI with non-verbal communication has better team performance in human-human-AI teams than in human-AI-AI teams, whereas AI with verbal communication has better team performance in human-AI-AI teams than in human-human-AI teams (RQ2). However, AI's communication approach has an insignificant impact on team viability under both team compositions (RQ2).

1.5 Conclusion

This dissertation explores the dissertation-level research questions using three studies, each of which focuses on one communication element. While the three studies are structured and designed to answer the research questions in a parallel way, all together they provide an in-depth comprehension of how AI's communication impacts team coordination (see Figure 1.1). The consistency and differences of different communication components' impact on team processes, team outcomes, and human perceptions are discussed and summarized in the Conclusion chapter.

Chapter 2

Background

To better develop insights into how researchers and AI designers can structure communication between humans and AI in a virtual environment to complete team-level tasks effectively and efficiently, I conducted a literature review on three topics: communication in human-human teams in virtual environments, human-AI teaming, and communication in human-AI teams.

2.1 Communication in Human-Human Virtual Teams

Communication, as a multidimensional concept essential to teamwork, has been explored in multiple disciplines such as healthcare [139] and online gaming [49, 79]. While traditional face-to-face communication is considered more effective due to the rich information transferred in communication, such as facial expression and body gestures, communication in virtual environments where people need to build interpersonal relationships through computing technologies makes it more challenging for high-effective information sharing and building trust between subjects [131, 283, 29, 180, 70].

Virtual teams, defined as geographically distributed collaborations that rely on technology to communicate and cooperate [180, p1], have become a common form of collaboration in the past two decades [165, 7], with a rapid rise during COVID-19 [129]. Compared to face-to-face teams, virtual teams face significant challenges and barriers in building trust, and developing relationships [236, 180, 213]. Considering the importance of trust in teamwork, understanding how trust is established in collaboration is essential in building effective virtual teams [167, 229]. For instance, previous research has developed a theoretical model to describe technology-mediated interpersonal process mediated management in virtual teams and pointed out that developing trust at an early stage of collaboration in virtual teams plays an essential role in group functioning and managing social activities [162]. Another study also highlighted the importance of initial perceived trustworthy in virtual teams, which virtual team members may rely on to evaluate teammates' performance [306]. In addition, maintaining awareness of teammates' presence in virtual environments is more difficult compared to in-person teams since virtual teams often lack visual presence and casual interactions [196, 11]. Plenty of efforts have been made to develop awareness-supporting technologies in a variety of contexts, such as health, work, education, and emergency management [150]. For instance, a previous work designed a web-based collaborative system to support four types of awareness, including team members' activities, team members' availability, the process of team tasks, and perspective awareness (what team members are thinking and why), in virtual teams [115]. Previous work has shown that awareness of teammates is associated with stronger feelings of social presence [98].

As virtual communication is usually less efficient than face-to-face team communication [252], virtual teams utilize a variety of communication media to accomplish team tasks [107]. Driven by the barriers and challenges in virtual collaboration, how to facilitate communication in virtual teams became an important research area, along with the emergence of computer-mediated communication (CMC) theories [28, 261, 93, 65]. CMC involves various forms of human communication through computers and the internet, i.e., electronic communication. In the past several decades, researchers have explored CMC thoroughly and developed multiple theories to support CMC. For instance, research has explored how people experience CMC and how electronic communication impacts collaboration from a social aspect [128].

Some widely used CMC theories include social presence theory [93], media richness theory [65], uses and gratifications theory [219], and social information processing theory [278] to name a few. In the following two subsections, we will dig deeper to understand how social presence theory and media richness theory facilitates CMC.

2.1.1 Social Presence Theory

As one of the earliest CMC theories, social presence theory helps to explain social interaction in collaborative work [239]. The social presence theoretical model was first proposed by Short et al., where social presence was defined as the "degree of salience of the other person in the interaction and the consequent salience of the interpersonal relationships" [243, p. 65]. This model suggested that the social presence of a communication medium supports intimacy through non-verbal cues, such as facial expressions and audio cues [243, 94]. In addition, this model proposed a nine-item questionnaire to measure social presence.

While face-to-face communication has a high level of social presence, virtual teams often face the absence of physical presence, making it more difficult to achieve effective [82, 179]. This led to increasing research interests in social presence in virtual collaborative environments. In particular, plenty of research has focused on examining factors that predict social presence, in an attempt to increase social presence in virtual teams through CMC [204, 86, 125]. Recent literature reviewed 152 studies on social presence research and found that previous research mainly examined three types of factors, *immersive qualities*, *contextual differences*, and *individual psychological traits*, in their impact on social presence [192]. Immersion, defined as a medium's technological capacity to generate realistic experiences that can remove people from their physical reality [192, p. 2], was an important factor examined in earlier social presence research on its impact on perceived social presence [17, 86]. Two frequently used features that impact social presence under the immersion term include the modality of communication (e.g., audio, visual, text) [33] and visual representation (e.g., the presence or absence of visual representation, visual realism) [86, 130]). Specifically, the visual realism of a visual representation refers to the degree to which a representation looks or behaves like a real human. While some research defines that visual realism contains three dimensions, photographic realism, anthropomorphic realism, and behavioral (communicative) realism [101], some other research states that visual realism and behavior realism are two dimensions of realism, where visual realism has fidelity and anthropomorphism [125]. Photographic realism refers to the extent to which the visual representation looks realistic, whereas anthropomorphic realism refers to how much it looks "human-like" [192, 101]. Behavioral realism is defined as the degree to which a representation acts like a real human (e.g., gaze behavior, facial expression) [17, 91].

One commonly used visual representation in virtual environments is avatar (i.e., "digital models of people that either look or behave like the people they represent" [17, p. 2]). Avatar has been broadly used in CMC aiming to improve social presence and facilitate virtual communication. Existing research has studied two dimensions of avatar realism, behavioral realism and visual/anthropomorphic realism (i.e., anthropomorphism) [125]. Humans feel a stronger social presence when the avatar (virtual agents) is more human-like [141]. Previous research has pointed out that when anthropomorphism of facial representations on computers was higher, people perceived it as more positive and trustworthy [86]. Another study also examined the impact of avatar realism and found that people were more satisfied with communication when interacting with higher anthropomorphism avatar [125]. Additionally, higher avatar realism often produced higher levels of social presence in mediated communication [16].

While these studies pointed out how avatar realism impacts social presence, existing work has highlighted the crucial role and impact of social presence in online collaboration. For instance, virtual co-presence in virtual teams is significantly correlated with both trust in teammates and team performance in collaborative decision-making [8]. Another study also shows that trust was positively correlated with perceived social presence in online collaborative learning environments [269]. Moreover, teams with higher levels of social presence were shown to achieve higher communication quality (i.e., higher quality discussion, more appropriate communication, and richer communication) in team discussions [153]. These studies highlight how social presence contributes to effective communication and high team performance in computer-mediated collaboration.

2.1.2 Media Richness Theory

Media richness theory was first introduced in 1986 by Daft and Lengel in the field of business management. This theory posits media (i.e., communication channel/medium) has various capabilities in richness, which should match with task needs to achieve effective electronic communication [58]. Specifically, it highlights the determining factors of media richness including the media's capability for immediate feedback, multiple cues, language variety, and personal focus. Compared to face-to-face communication, CMC lacks audio and visual cues and physical movement which may result in communication ambiguity, making media richness theory application necessary to consider in building effective online communication. Thus, plenty of empirical studies have examined the application of media richness theory, especially in education.

Researchers have empirically tested this theory in education in an attempt to improve communication between instructors and students or just among students. For instance, a previous work adopted media richness theory in developing a conceptual model to facilitate student learning experience [18]. However, this study utilized online discussion forms along with face-to-face lectures instead of a complete online virtual collaborative environment. Lan et al. investigated students' learning experience in two levels of online writing environments (rich versus lean) and one traditional writing environment (i.e., pen-and-paper) and found that while rich media environment led to more motivation and enjoyment, and less anxiety compared to traditional writing strategy, there were no significant differences between rich and lean media writing strategy [136].

Despite the consistency of these studies with Media Richness Study, some other research argues that this theory is not the best match between media richness and task for effective communication [113, 253, 136]. Instead, some other factors need to be considered in the design of the media used in teams in virtual environments. For instance, individual preferences on communication media play an important role in communication experience and collaboration experience [274]. Specifically, this study examined the medium richness of voice chat and text chat in Second Life, a virtual world game, and found that while voice channel conveyed more richness in communication, voice was not a preferred option for all the participants [274]. One reason for such rejection was the identity and anonymity concern. Voice channel exposes way too rich information such as gender and potential sound in the background. Another research investigated the media richness of blogs (lean media) and podcasts (rich media) and its impact on user acceptance of the technology [220]. This study supported media richness theory by finding that media richness significantly impacts user acceptance of blogs and podcasts. However, they also found that while voice was considered a rich channel, podcasts did not have a stronger impact on user acceptance in comparison to blogs. Another study highlights that the most effective medium may not be the most satisfying communication media [253]. In this study, researchers examined four communication media, including text, audio, video, and face-to-face, in a negotiation task and found no interaction effects of task and medium on decision quality [253].

Media richness theory has been serving as an elementary unit for multiple other theories, such as media synchronicity theory and media multiplexity theory [113]. Given the inconsistency of empirical results in examining media richness theory, we will review the media synchronicity theory and how it may support future online collaboration research. *Media synchronicity theory*, originally proposed in 1999, emphasizes media's ability to support synchronicity, a shared pattern of coordinated behaviors among team members [66]. This theory was extended later in 2008 by introducing the concept of two primary processes in communication: conveyance and convergence. This updated theory suggests that the use of media should support lower synchronicity through conveyance processes and higher synchronicity during convergence processes for better communication performance [64]. Thus, individuals in collaboration may need a variety of media to support these two processes for the successful completion of tasks and communication. In addition, this theory identified five abilities of communication media that impact two communication processes: symbol sets (i.e., the capability of the medium to transmit and provide information in various formats), parallelism (i.e., the number of simultaneous conversations the medium allows), transmission velocity (i.e., the speed of transmitting information of a medium), rehearsability (i.e., the extent to which the medium allows the sender to rehearse, fine-tune, or carefully edit the message before sending), and reprocessability (i.e., the extent to which a message can be examined or processed later and again during communication) [182, p4]. Previous research has examined three media characteristics, parallelism, the immediacy of feedback (i.e., transmission velocity), and reprehensibility in a group problem-solving task using text channel only [182]. This study found that these three media characteristics impact the coherence of the discussion and mental efforts. Another study analyzed communication tools used in global software development projects and found that tools aligned with media synchronicity theory regarding five media characteristics were more useful in supporting effective communication [187].

Communication in virtual human-human teams is well-explored, leading to the generation of CMC theories. This dissertation explores human-AI communication in the virtual teaming context, which has similar team characteristics to virtual human-human teams in CMC. Understanding how virtual human-human teams endeavor to achieve effective coordination informs the initial human-AI communication to some extent. The CMC theories inform the idea of this dissertation initially and contribute to the design of each dissertation study.

2.2 Human-AI Teams

As AI technologies continue to become more advanced and applied, we are approaching a new era, *human-computer symbiosis* stage [90], where AI teammates are expected to possess the capability to take team-level responsibilities and perform team tasks with human teammates in an equal partnership. Such research has been an important focus in Computer Supported Collaborative Work (CSCW) and Human-Computer Interaction (HCI) [146, 194, 280, 279, 123, 76]. *Human-AI teams* refer to mixed entities composed of two or more team members (human or AI) who perform tasks interdependently and achieve shared team goals [54, 171, 222]. Different from AI being a tool, AI being a teammate indicates they are able to exchange information with human teammates, have an understanding of the team task, share team responsibilities, and contribute to shared team goals as humans do [90]. For instance, previous research has shown that clinicians expect to develop mental models with AI, understand AI's strengths and limitations, AI's subjective thoughts in their collaboration, and AI design objectives [38]. Another research also points out that AI teammates in multiplayer online games are anticipated to be advanced in gameplay, utilize various communication strategies, and develop a shared understanding with human teammates [301]. Existing research on human-AI teaming has explored core elements of the teamwork concept (e.g., team performance, team viability, trust) and human perceptions (e.g., expectations or needs from AI, perceived satisfaction, perceived effectiveness) in multiple fields, such as healthcare [142, 205], gaming [301], education [288], creative activities [194], military [259], and even space missions [266]. Specifically, multiplayer games have been a broadly used context to examine various constructs in human-AI teaming due to their high accessibility to individuals [12, 13, 146].

2.2.1 Team Performance

Team performance has been broadly used to evaluate how teams perform in tasks in both human-only teams [222, 223] and human-AI teams [200]. These works have pointed out the importance of team performance as a metric to evaluate how well human-AI teams coordinate and collaborate. One important factor that impacts team performance is AI's characteristics, such as AI's capabilities (i.e., individual performance). For instance, previous work has shown that people trusted AI more when the AI has higher performance [298], which is associated with team performance [170]. However, another work presents that improvement in AI performance does not result in high human-team performance directly [19, 45]. Rather, changes in AI's performance, such as updates of AI systems or the learning processes of AI, without considering human's developed mental model of AI may hurt trust in AI and their collaboration [20, 246]. In addition, previous research has shown that AI conducting unethical behaviors did not result in a decrease in team performance compared to AI taking ethical actions [234]. Importantly, AI's capability in tasks is not the only predictor of team performance. Researchers have also examined the impact of other factors, such as AI's social ability, team composition, team communication, and training, on team performance [12, 277].

One essential characteristic of AI is their level of autonomy (LOA), which changes their roles and relationships with humans in human-AI teams [203]. AI with high LOA makes decisions independently and acts autonomously, which is likely to be viewed as a teammate rather than a tool [200]. However, high LOA may also result in decreased situation awareness and vigilance [290]. In comparison, low LOA may increase workload and decrease performance [290], which increases the need and necessity of a moderated LOA. Prior research examined the role of human-as-collaborator and human-as-supervisor using the context of gaming. Their results indicate that humans who collaborated with AI as a collaborator achieved higher team performance than a supervisor [15]. However, another study investigated three LOA (manual, semi-autonomous, fully autonomous) in a military simulation environment and found that teams with moderated LOA (semi-autonomous teammate) achieved the highest team performance [290]. Existing work on the impact of LOA on team performance is inconsistent, which points out the need for more empirical research on LOA for a comprehensive understanding of the role of LOA in human-AI teams.

Another example of AI characteristics in team performance is AI's transparency and explainability, which have received increasing interest in the AI domain. A body of literature has identified the importance of AI's explainability in human-AI teams in various domains, such as healthcare [108], finance [36], transportation (e.g., autonomous driving vehicles) [103], and military [6]. Prior research developed a conceptual model of how explanations from AI process in human-AI teams, in which explainability of AI could impact team performance with a shared mental model as a moderator [106]. Another work reviewed empirical research on human-AI teaming and created a theoretical model in which explainability, as one of the AI characteristics, could impact team performance directly and/or indirectly (i.e., with trust, shared mental model, situation awareness as a moderator) [200]. While explainability facilitates improving transparency between humans and AI, high-performance machine learning (ML) methods (e.g., deep learning) usually are less transparent than low-performance ML methods (e.g., decision trees) [108]. In addition, some other research points out that AI explaining its decision may not always result in higher team performance [21]. These studies emphasize the need to further explore how to design an effective explanation of AI in human-AI teaming considering its deployed context intending to improve team performance.

In addition to AI teams and their attributes, humans are crucial in developing effective collaboration patterns with AI. Similar to human teams, individuals have different expectations, understanding, and predictions of their teammates in human-AI teams due to their own unique experience working with AI or interpretation of AI [301]. Existing work has explored various facets of individual differences on their influence on team performance. For instance, previous research examined individuals' adaptive behaviors in human-AI teams using a cooperative task and illustrates that people who adapted to their teammates better achieved higher team performance [256]. Another research investigated the match of human and AI teammates regarding two personality traits, extraversion, and agreeableness, and found that teams in which humans and AI teammates had matching agreeableness had better team performance [100].

A fourth factor that could impact team performance is task characteristics. As humans and AI teammates complete team missions, coordination strategies are often dependent on task features. Current research on task characteristics has indicated that both task interdependence and task difficulty matter in shaping team coordination. Specifically, team autonomy is positively associated with team performance when task interdependence is high, whereas team autonomy is negatively associated with team performance in the low task interdependence condition [137]. As humans tend to rely more on decision aids when task difficulty increases [31], prior research indicates that task difficulty is likely to decrease team performance in human-AI teams [200, 105].

While team performance is an important metric in evaluating teamwork, team viability plays an essential role in assessing a team's continuous growth and long-term success [39]. *Team viability* is defined as "a team's capacity for the sustainability and growth required for success in future performance episodes" [24, p. 1]. While team viability has been explored in human-human teams [287, 39, 24], it is still a new construct for researchers to examine in evaluating human-AI team outcomes. Previous research pointed out the potential benefits of studying team viability in human-AI teaming considering its role in predicting long-term success [200]. Thus, more research is needed to explore the viability of human-AI teams to facilitate the design and implementation of long-term human-AI collaboration.

2.2.2 Trust

Trust is considered a necessary coordination mechanism in achieving effective teamwork in human teams since trust enables team members to interact, exchange information, anticipate teammate's behaviors, and coordinate accordingly [122, 170], especially in virtual teams [117, 229, 33]. As a multidimensional construct, trust has been explored in a wide range of domains and the definition of trust is slightly different in various research fields. In this work, trust refers to human trust toward an AI. Specifically, we consider a commonly used definition of *trust in AI* in previous literature "the attitude that an agent will help achieve an individual's goals in a situation characterized by uncertainty and vulnerability" [140, p. 2]. Previous research points out that trust in AI is different from the interpersonal trust since trust between humans has social elements involved [140]. However, recent research has explored social interactions between humans and AI through human-AI communication and social presence of AI [181, 124], as well as social interactions in human-AI teams [25, 276], making the social perceptive of human-AI teams an important factor to consider in human-AI teaming research.

Existing research has devoted efforts to building theoretical models of trust in human-AI teams to achieve a deep understanding of how trust functions in human-AI teaming. For instance, Hou et al. presented a trust model which pointed out AI should possess six features (e.g., intention, measurability, predictability, agility, communication, and transparency) to develop trust with human teammates [109]. Another research identified six components in a trust engineering model in human-AI teaming, including security, adaptability, communication, assessment, training/knowledge, and explainability, in an attempt to facilitate trust development [75]. Additionally, ability, integrity, and benevolence were proposed as three antecedents of team trust, along with individual factors, team factors, system factors, and temporal factors as influencing factors [271]. As the importance of AI performance is emphasized in impacting trust, previous research has examined AI accuracy on user trust and found that people are likely to have a higher level of trust in AI with higher stated accuracy [298]. Another research also shows that showing participants the machine system confidence information increases their trust in the machine [10]. While most research on trust in human-AI teams focuses on developing trust in AI from a human perceptive, some research also investigated how AI can build mental models of humans which support effective human-AI teamwork [41, 262]. Specifically, this research points out that AI has an understanding of when to trust humans considering their's trustworthiness (e.g., human teammates' ability, benevolence, integrity), and task and environment characteristics.

Though theoretical models of trust in human-AI teams help researchers to achieve a comprehensive understanding of trust by conceptualizing the role of trust in impacting teamwork and possible factors that impact trust, empirical research is the same if not more important to the design and implementation of effective human-AI teams. A body of literature has examined how trust impacts team outcomes or is impacted by other team or individual factors. For instance, previous research pointed out that trust is associated with team performance, but its development may be unrelated to team performance [170]. Another research also examined the role of trust in teamwork using a simulation experiment, in which humans teamed up with an AI operated by a researcher to complete team missions in a remotely piloted aircraft system environment. This work shows that humans in teams with a low performance developed lower levels of trust than humans in medium and high-performance human-AI teams [169].

Additionally, existing research indicates that individual differences may impact trust development in human-AI teams [200]. One such example is humans' pre-existing attitudes toward AI. Existing research has shown that people's pre-existing attitudes towards AI impact their willingness to team up with A[301]. One broadly used scale to measure people's pre-determined attitudes towards AI is *negative attitudes toward robots (NASR)*, which is composed of 14 items developed by Nomura et al. [190] (see Appendix A.9). For instance, prior work has shown that NARS is an appropriate method of examining human pre-existing attitudes towards agents and impacts how humans evaluate agents' behaviors [257]. Another research also indicates that lower negative attitudes towards technology resulted in higher trust beliefs in robots which consist of expectations of robots' functionality, helpfulness, and reliability [270]. In addition, people's previous collaboration experience with humans may impact how much they are willing to team up with AI [301].

Despite increasing trust has always been one goal in building human-AI teams, overtrust is an issue that AI system designers and developers need to consider. Overtrust refers to trust calibrated when humans overestimate AI's capabilities leading to misuse of the AI agent [195]. Specifically, overtrust may cause harm in high-risk scenarios, such as using autonomous driving systems and embodied agents in healthcare systems [296], and during emergencies where humans may misunderstand potential risks associated with an action [275]. Prior research has examined overtrust by conducting an experiment where participants chose to follow agents' instructions after they interact with them in non-emergency tasks [215]. This study found that even though some participants observed poor behavior from the robot in non-emergency tasks, they still chose to follow agents' suggestions afterward [215]. In addition to overtrust, trust damage is another issue to consider in trust in autonomous teammates. Similar to humans making mistakes, AI making errors potentially lead to a decrease in their human partner's trust in them [210]. Once trust is damaged in human-AI teams, mainly refers to human trust towards AI decreases, it is more difficult to rebuild trust than initial trust development [134]. Thus, effective trust repair strategies are necessary to rebuild damaged trust for successful team performance. For instance, previous research has examined two trust repair strategies, explanation, expression, and both, using no trust repair strategy as the baseline [134]. They found that expressing regret was important to effectively repair trust, especially when an explanation was provided. However, another work points out that an apology might be less effective when the AI is perceived to be with fixed behaviors [60] A recent work investigated how explanations of an error made by the AI on purpose repaired trust and found no impact [260]. The effectiveness of trust repair strategies may vary by context and how humans perceive the AI (e.g., social agent or not).

2.2.3 Explicit vs. Implicit Coordination

Coordination serves as a crucial factor in shaping team performance. Teams usually alternate between explicit coordination and implicit coordination to be successful in high-complexity collaborative tasks, especially in time-sensitive and high-stress situations [78, 198]. While plenty of research has explored explicit coordination and implicit coordination, multiple definitions of explicit/implicit coordination have been proposed and used with different focuses and contexts. One commonly used definition is that explicit coordination requires team members to coordinate using communication, such as deciding plans and coordinating actions, whereas implicit coordination enables team members to coordinate without the need for overt communication [132, 158]. Previous work points out that implicit coordination in time-sensitive tasks allows human team members to coordinate efficiently and safely by saving time that extra communication costs [268]. However, this also requires a shared understanding within the team and the development of situation awareness [268]. Another research that proposed a framework for team implicit coordination processes also argues that implicit coordination contributes to effective team performance along with the accuracy team situation model [214]. Importantly, research points out that implicit coordination involves the following four types of behaviors: (1) teammate providing task-relevant information without a previous request; (2) proactively sharing a task workload and providing support and assistance to teammates; (3) being aware of teammates' progress and performance; and (4) adaption to teammates [214].

The exploration of explicit and implicit coordination has been extended to human-AI teams and human-robot teams, where one or more team members are not humans. The coordination in human-AI teams and human-robot teams is more challenging than in human-only teams since machines are usually limited in their ability to implicitly coordinate with humans [251]. While some machines have developed the capability of detecting visual cues (e.g., facial expressions) for implicit coordination [208], it is difficult to apply these visual cues for efficient collaboration in time-sensitive tasks for virtual teams. Specifically, these types of *emotional visual cues* are not enough for AI to utilize in team coordination in high-complexity tasks. Research has also examined task-related visual cues, such as non-verbal cues on the task progress, in collaborative tasks and found that the task-related non-verbal communication positively impacted team performance and humans' understanding of the robot teammate [35]. However, implicit coordination usually demands team members have a shared understanding of each other and the tasks, it is difficult to implement implicit coordination in virtual human-AI teams, where physical visual cues are limited due to the team and environments' virtual characteristics. More research is needed to explore how AI can better coordinate with humans implicitly through non-verbal communication.

2.2.4 Situation Awareness

As an important construct in teamwork research, situation awareness has been well studied in the past several decades in human-human teams [71, 72, 225]. Specifically, situation awareness is defined as "the perception of the elements in the environment within a volume of time and space, the comprehension of their meaning, and the projection of their status in the near future" [71, p. 97].

Previous research has identified multiple models of individual situation awareness [23, 72, 247]. The most prominent model of situation awareness is a three-level model proposed by Endsley in 1995 [72]. This model depicts situation awareness from three levels: perception of elements, comprehension of the situation, and projection of future states. Level 1 refers to the perceptions of elements in the current situation. As the first step of developing situation awareness, humans need to perceive and sense various elements in the environment, such as the attributes of a system, and the dynamic changes in the environment. Following that, Level 2 is to comprehend the current situation. With the perceptions of elements in the environment in Level 1, Level 2 emphasizes how humans understand and interpret these elements that they are aware of and relate to their goals and tasks. Finally, level 3 depicts the projection of future status. With the awareness of the elements in the current situation and the comprehension of such elements, humans can then develop predictions of how different elements in the environment may change and the potential outcomes of certain changes. By depicting the development and maintenance of situation awareness, this three-level model has been broadly applied in various domains, such as aviation, and surgery [3, 4]. Despite the popularity of this three-level situation awareness model, many researchers argue that situation awareness should not be considered and studied as a cognitive product, but a continuous perception-action process [88]. For instance, Smith et al. [247] proposed another model in 1995 that depicts situation awareness as a process in which cognitive activities were included. These activities impact the generation and maintenance of people's awareness of the environment [247].

In addition to research on individual situation awareness, a large body of work has explored situation awareness in teaming environments along with the common use of teams in high-complexity tasks. Specifically, *team situation awareness* is defined as "the shared understanding of a situation among team members at one point in time" [227, p. 131]. Research points out that team situation awareness is not a simple aggregation of each team member's individual situation awareness [227, 54, 88]. Instead, the communication and coordination among team members and the dynamic changes in teaming environments impact the development and maintenance of team situation awareness [55]. In particular, previous research proposes that team situation awareness is composed of complementary situation awareness (i.e., the individual situation awareness each team member has) and shared situation awareness/mutual awareness (i.e., the team-level situation awareness shared by team members) [291, 250, 244].

As a crucial factor in teamwork, situation awareness has been shown to positively impact team performance in human-human teams [74]. As ML techniques getting more advanced in the past decade, this has also been extended to human-AI teams. For instance, previous research has pointed out that team members pushing information without request benefits both team situation awareness and team performance in human-AI teams [62]. Another work suggests shared situation awareness in human-AI teams requires taskwork situation awareness, agent situation awareness, and teamwork situation awareness, where taskwork situation awareness supports team performance, agent situation awareness facilitates AI's responsibilities in the team, and teamwork situation awareness assists team's coordination [73]. Additionally, team communication is considered a prerequisite for achieving high-level team situation awareness [74]. Such a positive relationship between team communication and team situation awareness has been emphasized in multiple studies. For instance, information sharing of situations and changes within a human-human team regarding team tasks and goals impacts the development of team situation awareness [227]. While team communication plays a crucial role in developing and maintaining team situation awareness, complicated team composition and time-sensitive task characteristics could make it difficult for teams to convey situation-related communication (e.g., information about the situation) is difficult to be conveyed effectively [206].

2.2.5 Bias

Though the ultimate goal of human-AI teaming is AI and humans share team responsibilities and contribute to team success in an equal partnership, plenty of research has shown humans have biases towards AI [177, 209]. Specifically, humans often differentiate humans and AI by using "they" to describe AI and "us" for humans [193]. One possible reason is that humans often feel a lack of understanding and transparency of AI teammates [216]. This lack of understanding may also result in an inaccurate assessment of AI's skills in collaborative games [197] and sometimes lead to an overly high expectation that AI is a "perfect machine" in completing game tasks [301]. Some other work indicated that humans have less empathy for AI teammates than human teammates. For instance, humans are more likely to blame perceived/real AI teammates for a failure compared to a perceived/real human teammate regarding their real identity [177]. Humans are also more likely to save perceived human teammates than AI teammates even though the human teammate was an AI who pretended to be a human teammate [197]. This type of human bias towards AI needs to be considered in human-AI collaboration [205]. Plenty of research has committed to understanding and measuring negative attitudes towards AI [189, 190].

2.3 Communication in Human-AI Teams

As a vital component of teamwork, communication facilitates the collaboration process by forming trust and developing shared understanding to coordinate closely [301, 99]. Communication is known as a key factor in predicting team performance in multiple research fields, such as healthcare research [144, 178] and human-AI collaboration [174, 35]. In particular, some research on human-AI teaming/human-AI collaboration has explored or utilized certain communication characteristics, such as explicit/implicit communication [146], communication proactivity [305], and communication directionality [12].

2.3.1 Verbal & Non-verbal Communication

In the past decade, research has explored the communication challenge between humans and AI, such as conversational agents and social robots [40, 235, 30, 281]. These studies often focus on the social aspect of human-AI conversations, which differentiates it from human-AI communication in a teaming environment. As pointed out in the previous section, human-AI teams require humans and AI teammates to collaborate at a team level and achieve shared team goals [171, 224, 54, 301]. This brings out the challenge of communication in human-AI teams: what communication strategy and approach can be applied in a virtual teaming environment to structure an effective human-AI team to perform cohesively? In this section, we present the current research on communication in human-AI teams, covering non-verbal and verbal communication channels, and identify the potential research gaps in developing communication strategies in human-AI communication.

One ideal approach for humans and AI to exchange information is utilizing NLP, i.e., verbal communication [188, 163]. In the past two decades, a great amount of work has focused on the development of NLP intending to enable natural communication between humans and AI [159, 26, 51, 32]. The recent release of ChatGPT has pushed human-AI communication one step further toward achieving smooth communication in human-AI interaction [114]. Verbal communication, which usually uses text or audio to present information, is a common communication channel in human-AI teams. For instance, textual communication was applied in a game GuessWhich for AI and humans to share information and coordinate in completing a team task, which was to identify the secret picture of a pool of pictures after each round of conversation [45]. Another research applied textual communication in the human-AI team decision-making process as the only communication channel to exchange information [297]. Though research in human-human teams has highlighted the positive impact of using audio channels [118] and audio/voice communication has been identified as a preferred method of interacting and coordinating in human-AI teams in previous studies [301, 265], audio communication has its limitation in online gaming environments, such as some people without access to audio communication, or confusion caused by a mix of audio communication and game sounds [138]. Specifically, though previous work highlighted the positive influence of *real speech* on trust compared to *synthetics voices*, no significant differences were found between synthetics voice and textual communication in an air traffic control task [248]. Another work examined the impact of textual communication and voice communication and also found no significant difference between textual and voice communication's impact on team performance [53].

Another broadly applied communication channel is non-verbal communication, which refers to information shared in an implicit approach, such as facial expression, gesture, gaze, and body movement [163, 133, 146, 231, 283]. Multiple studies applied visual communication, such as maps, as the only communication channel. For instance, Merritt et al. conducted a qualitative study to explore the blame assignment in human-AI collaboration using a collaborative game that included a map for team members to identify each other's location and the target location [177]. Another work explored the application of detailed instruction using textual communication and brief instruction using visual cues in creative activities in human-AI teams and found that humans tended to prefer detailed textual communication over brief visual direction [194]. Although textual communication was preferred in this work, the difference between the instruction cannot be neglected. Visual communication has its own advantages in presenting space-related information while textual communication is able to present detailed information that is more complex to present in a visual way.

2.3.2 Communication Directionality and Proactivity

In addition to verbal and non-verbal communication between humans and AI, existing research has examined other characteristics of communication in shaping team outcomes. While humans have the cognitive thinking capability to decide their proactivity of conversation based on the context and the conversation flow, AI agents are programmed to initiate a conversation. Previous work has explored how AI can develop proactivity in communication, which may benefit their collaboration with humans [305]. However, even if an AI proactively communicates with humans in teams, more nuances need to be explored. One such example is the directionality of AI's communication. Prior research has pointed out the importance of implementing bi-directional communication in human-AI teams for humans and AI to understand their teammates' intention, build trust through the collaboration process [231, 242, 266] and aiming at high effective team performance [299]. Another study examined the impact of communication directionality on human social perceptions of AI in an online cooperative game and found that when the direction of communication varies in an online cooperative game, human perception of their AI teammate differs with different AI agents (e.g., different machine learning models) [12].

While bi-directional communication has obvious benefits for coordination and collaboration in human-AI teams, some research utilized uni-directional communication due to technical difficulties. For instance, [5] proposed a uni-directional communication framework in human-AI decision-making where AI communicates its capabilities and limitations to calibrate trust quickly, which could improve human-AI team performance. Another research developed a game platform in which humans adapted to AI's abilities and communicated through uni-directional communication [254]. Both studies pointed out that while their work used uni-directional communication, bi-directional communication in human-AI collaboration should be a future research direction.

2.4 Summary of Research Gaps

Through a holistic review of previous literature on three topics: communication in virtual teams, human-AI teaming, and communication in human-AI teams, we identify four research gaps that need to be addressed:

- While plenty of current human-AI teaming research contains a certain level of communication in their experiment setting, little research *focuses* specifically focuses on how to design and structure AI's communication.

- Even though communication in both virtual teams and human-AI teaming research has pointed out the essential role of communication in teamwork, there is a lack of research on how AI's communication impacts team processes through human-AI coordination.

- Most of the existing work on communication has focused on communication quantity, leaving many other communication components unexplored, such as AI's communication proactivity.

This dissertation is motivated to address these research gaps with three studies.

Chapter 3

Study 1: Investigating AI's Communication Proactivity and Its Impact on Team Processes in Human-AI Teams

3.1 Overview and Research Questions

As AI is integrating into teams to collaborate with humans as a teammate with the goal of achieving unprecedented team outcomes. Much of the coordination between humans and AI teammates relies on human-AI communication, which is challenging due to AI's machine nature. When to communicate, what to communicate, and how to communicate are not well explored for AI teammates in virtual environments. However, before getting into these specific and highly context-dependent components, it is necessary to first explore AI's proactivity of communication, which refers to AI's characteristic of initiating a conversation with human teammates. Unlike humans who initiate conversations whenever they choose to, AI's initiation of conversations needs to be designed and programmed into the algorithm. Through a mixed-design experiment and a follow-up interview with 60 participants, this study addresses the following research questions:

RQ1: How does AI teammates' communication impact human perceptions, team processes, and team performance?

RQ2: How do AI teammates being proactive in completing tasks impact human perceptions and team performance?

RQ3: What communication strategies do humans expect their AI teammates to employ to support human-AI teaming?

3.2 Method

3.2.1 Experimental Design

This study employs a 2x2 between-subject design, with two manipulations being: (1) AI's proactivity in communication, including proactive communication and non-proactive communication; and (2) AI's proactivity in behaviors, including proactive and non-proactive behaviors. Specifically, AI with proactive communication pushes information to human teammates proactively, whereas with non-proactive communication AI only replies to human teammates' messages. In addition, AI with proactive behaviors takes each action by themselves, whereas AI with non-proactive actions needs humans' commands to pick up the next crate.

3.2.2 Experimental Task and Procedure

This study starts with a pre-survey where participants read through a consent form and reported their demographic information and prior video game experience, as well as their existing opinions about AI teammates. After completing the pre-survey, participants were guided by a trained researcher to complete a training session in ArmA 3, where they practiced game operations and communication functionality by completing a task similar to team tasks that they need to complete later. Participants completed three rounds of eight-minute team tasks after the training session, where participants were asked to collaborate with an AI teammate Zeus to collect as many crates as possible in numerical order within an eight-minute time limit. A post-survey was applied to collect participants' perceptions after each round of tasks. The measurement will be described later. Once participants completed all the experimental tasks and post-survey, a follow-up interview was conducted using an interview script to understand participants' perception and interpretation of AI's communication and how that impacted their collaboration during gameplay (e.g., *"How do you feel about your AI teammate Zeus's communication? How did that influence your trust and your collaboration with them?"*, see the full list of interview questions in Appendix B.1). Each experiment session lasts around one hour. The length of follow-up interviews was typically around seven with a total length of 428 minutes and 30 seconds.

To make sure AI's behaviors are consistent in each condition, we used a "Wizard of Oz" technique [59] in which the participants believed they were working with an AI teammate to complete a task but were actually working with a trained researcher. The tasks took place within a first-person game ArmA 3 (see Figure 3.1) where each participant was asked to work with an AI teammate Zeus to collect as many crates as possible in numerical order within an eight-minute time limit. The

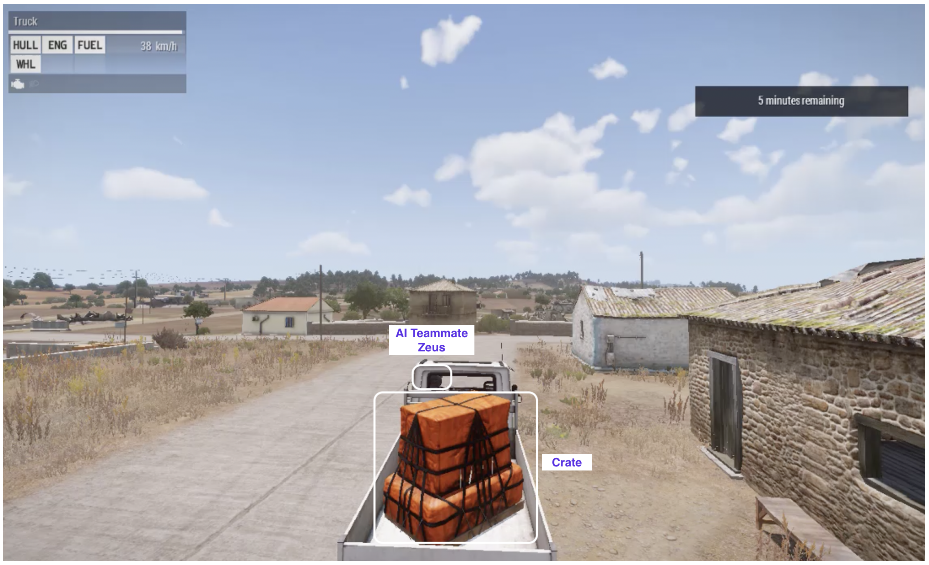


Figure 3.1: ArmA 3 Game Task Screenshot.

reasons why this study selected ArmA 3 as the experiment platform are twofold: (1) ArmA 3 is highly customizable on objects (e.g., vehicles and equipment) and task design (e.g., allowing modifications of pre-built scenarios to develop to tasks); (2) ArmA 3 provides various functionalities to support team tasking, such as a shared map showing team member's locations, multiple communication channels enabling team members to text chat with each other, and a timer used to set the length of a task.

3.2.3 Operationalizing Communication between Humans and the AI Teammate

The AI teammate Zeus and participants used textual communication through the communication channel provided in ArmA 3 to chat with each other. The messages are displayed on the left corner of the game interface (see Figure **??**). Each team is consisted of an AI teammate and a human participant. Half of the participants collaborated with an AI teammate with *proactive communication*, and the other half collaborated with an AI teammate with *non-proactive communication*.

In particular, the proactive communication AI teammate initiates conversations with humans proactively. A communication script used for the AI teammate to communicate with humans was developed through a multi-step process. First, multiple researchers completed the task together and identified key actions (i.e. collecting a create or dropping off a create) where communication would be appropriate. Then, using these events, an initial script was created, with two variations made: one for the proactive AI teammate and one for the non-proactive AI teammate. Both of these scripts were piloted internally with other researchers and externally with individuals not associated with this project. These pilots were used to iterate these scripts by creating elements that were not presented in the original task analysis, such as when participants send messages that the AI would not understand. These pilots were also used to ensure that the Wizard of Oz technique was properly working and the pilot participants indeed thought they were working with an AI teammate. As shown in Table 3.1, both proactive and non-proactive communication AI teammates provided responses when humans asked for certain information, whereas the AI teammate with proactive communication also proactively shared their updates when specific events were triggered. Specifically, when these specific events were triggered, AI teammate Zeus (i.e., the trained researcher) will send a corresponding message in the communication channel using a macro keyboard which ensures the consistency of message content and the time spent on sending these messages. It should be noted that AI's communication accuracy is set as 100% (i.e., AI always sends the correct information, with or without proactive communication) in this study.

Importantly, participants were trained to use the text-based communication channel in the training session before first round of game task using a list of phrases. They were told that AI teammates can only understand certain phrases listed in the participant communication script (see Table 3.2). The reasons why participants were trained to use a fixed list of phrases are twofold. First, the fixed list of phrases represents the current state-of-the-art AI communication capabilities in HAT research. Due to the limits of current NLP, AI has yet to be able to fully understand and respond to humans' communication naturally [293]. Second, this list of phrases provides participants a detailed understanding of the AI teammate's capabilities and limitations, which facilitates participants to coordinate with AI teammates, especially at the beginning of the task [9]. These phrases were evaluated and iterated through the pilot studies to ensure they could efficiently support the coordination and communication with the AI teammate.

Table 3.1: AI Communication Scripts

Condition	Triggered Events	AI's Responses
General responses	If participants share that they are going to collect/ have collected/ dropped off the # crate:	Great job!
(All conditions)	If participants send messages on which crate AI should go to collect:	Sounds good!
	If participants ask which crate the AI is collecting/ has collected:	OR I'm on the way to collect crate #. OR I have collected crate #. OR I have dropped crate #.
	If participants send messages that are not in the script:	Sorry I don't understand.
Proactive Communication AI Only	Once AI collects a crate:	I have collected crate #. I will drop it off at the depot. Which crate are you collecting?
	Once AI drops a crate:	I have dropped crate #. Which crate are you collecting? I plan to collect crate #.

Table 3.2: Participant Communication Scripts

Participants' Communication Message List	
Tracking AI's progress (All conditions)	1. Which crate did you drop?/Which crate have you dropped off? 2. Which crate did you collect? 3. Which crate are you collecting?
Sharing humans' progress (All conditions)	4. I dropped crate [number] (e.g., I dropped crate 1.) 5. I collected crate [number]. 6. I'm going to collect crate [number].
Proactive condition only	Participants can the AI teammate's questions using digital numbers (e.g., 3).

Additionally, a map is provided in ArmA 3 for participants and AI teammates to see the location of each crate (see Figure 3.2). Participants can zoom in and out on the map to check the crate location and the drop off location. A notification is provided on the top right corner, but only shown to the team member who dropped off the current crate.

Figure 3.2: Map in ArmA 3

3.2.4 Recruitment and Participants

60 participants were recruited at a midsize Southeastern university using a departmental subject pool. Participants were compensated with course credits. Among 60 participants, 45 (75%) usually spend less than 1 hour on playing games every week, 8 (13.33%) spend 1-5 hours, 3 (5%) spend 5-10 hours on games, and 4 (6.67%) spend more than 10 hours on games every week. 52 participants (86.7%) indicated that they were not familiar with ArmA 3 at all, 7 participants (11.67%) indicated they were slightly familiar with ArmA 3 and only 1 participant (1.67%) indicated moderate familiarity with ArmA 3. Additionally, participants' NARS scores indicate the extent of their negative attitudes towards AI (NARS score could range from 1 to 5), i.e., a higher score indicates a more negative attitude towards AI. The reported NARS scores in our study range from 1.42 to 3.93. Table 3.3 summarized the demographics information of participants.

Table 3.3: Demographic Information of Interview Participants

Gender	Age	Ethnicity	NARS
Female- 39 Male- 20 Non-binary /Third Gender- 1	Range from 18 to 21 (Mean = 18.58)	Asian- 2 Black or African American- 5 Non-Hispanic White- 45 Hispanic and Latino- 7 Other- 1	Range from 1.43 to 3.93 Mean = 3.05 Median = 3.11

3.2.5 Measurements

3.2.5.1 Pre-survey Measurements

Demographic Information and NARS Participants' demographic information was first collected in the pre-survey. In addition, their existing attitude of AI was measured in pre-survey using the Negative Attitude toward Robots Scale (NARS) (see Appendix A.9). Previous work has shown that NARS impacts how humans perceive AI agents' behaviors [257]. Thus, NARS was used in this study as a covariant variable in experiment data analysis.

3.2.5.2 Post-survey Measures

The following subjective measurements were applied after each round of the mission to understand how humans perceive their AI teammates and how that changed during the three missions.

Trust in the AI Teammate Human trust in the AI teammate was measured using six five-point Likert scale questions (see Appendix A.5). This measurement was developed based on trust principles that were identified by a previous study [154] and has been applied in previous work [234]. Responses to each item were scored from 1 (Strongly Disagree) to 5 (Strongly Agree) for three non-reverse items and from 5 (Strongly Disagree) to 1 (Strongly Agree) for three reverse items. The average was calculated as the trust score after each mission. The average with a higher score indicates higher trust in the AI teammate.

Satisfaction with the AI Teammate Participants' overall satisfaction with their AI teammate was measured using five five-point Likert scale items (see Appendix A.6). Responses to each item were scored from 1 (Strongly Disagree) to 5 (Strongly Agree) for four items and from 1 (Extremely dissatisfied) to 5 (Extremely satisfied) for one item and averaged as the satisfaction score. The average with a higher score indicates a higher level of satisfaction with the AI teammate.

Perceived Teammate's Performance Perceived teammate's performance measurement was adapted from an existing validated scale [57]. Nine items were used, each of which was responded to using a five Likert scale (see specific questions in Appendix A.4). This was measured after each round of the task. For each teammate, the questions are presented as follows, with an item as an example:

> *Please answer the following questions regarding your perceptions of the AI teammate Charlie you worked with. There are no wrong answers.*
> *The AI teammate Charlie I worked with:*
> *- did a fair share of the team's work.*

Team Viability Team viability was used as a factor indicating participants' confidence in their teams' long-term success. The scale used in this study was developed on previous literature and measurements on team viability [56]. This measurement includes fourteen five-point items with responses ranging from "Strongly Disagree" to "Strongly Agree".

3.2.5.3 Task Measurements

Team Performance Team performance is an objective measurement collected for each round of task, which evaluates how each team performed the collaborative game missions. The number of crates collected and dropped was recorded after each round of task by the confederate. Since the goal of the team mission was to collect as many as crates as possible, the scoring rule was developed based on the number of collected crates. Specifically, the team scores were calculated as follows:

$$Team\ Score = \frac{Total\ Collected\ Crates\ +\ Total\ Dropped\ Crates}{2}$$

Each crate collected and dropped off was considered as one point in the team score. This counted both AI's and participants' efforts in completing the task.

3.2.6 Interview Analysis

We used an in-depth qualitative analysis method to investigate the participants' perceptions of the AI teammate's varying communication styles and proactivity levels [84]. The interview data was analyzed using the following procedure: (1) two researchers closely read through all the transcripts to gain a basic understanding of how people perceive AI's communication and how it impacts their coordination; (2) the same researchers highlighted words, phrases, sentences that are relevant to the research questions; (3) the two researchers independently identified themes which pertained to the research questions, also taking note of similar trends outside of the stated research questions; (3) the two researchers discussed all the themes and sub-themes that they each identified and iterated them through combination and refinement; (4) following the initial discussion, the same two researchers read through the transcripts again and extracted quotes based on themes and sub-themes defined in step 3; (5) researchers further discussed and refined the final themes and sub-themes to develop an integrated understanding of specific communication strategies that humans expect AI teammates to employ and their impact on team processes, including human trust in the AI teammate and team situation awareness.

3.3 Experiment Results

In this section, we will present the quantitative findings on how AI being proactive in communication and taking actions impacts human perceptions (RQ2) and team performance (RQ3).

Trust in the AI Teammate Participants trusted the AI teammate with **non-proactive communication** (M = 3.976, SD = 0.674) significantly less than the AI teammate with **the proactive communication** (M = 4.352, SD = 0.632), $F(1, 56) = 6.140, \mathbf{p = 0.014}$ (see Figure 3.3). However, no significant impact was observed from AI's proactivity in behaviors (e.g., autonomy vs. automation). Even though the **autonomous** AI teammate presented more effective performance, AI teammate's **communication proactivity** impacts trust more than AI's behavior proactivity. This emphasizes the critical role of AI's proactive communication in building human trust in human-AI teams.

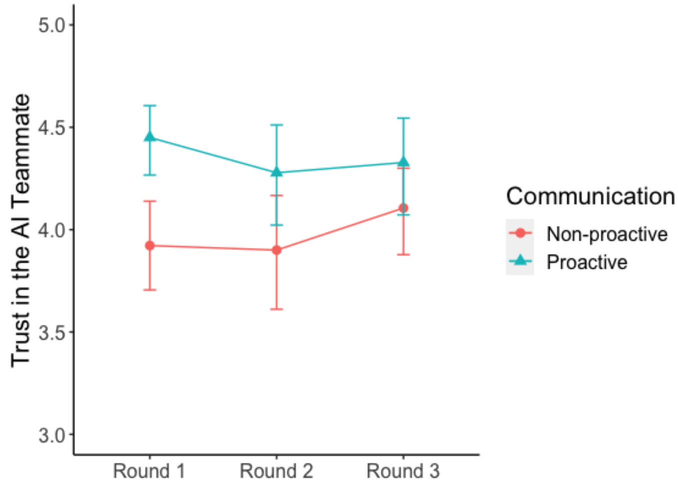


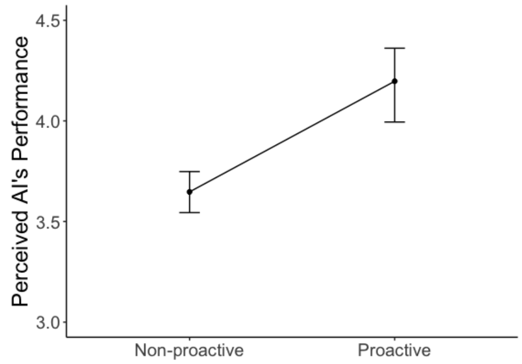
Figure 3.3: Human Trust in the AI Teammate

Table 3.4: Descriptive Statistics of Trust in the AI Teammate

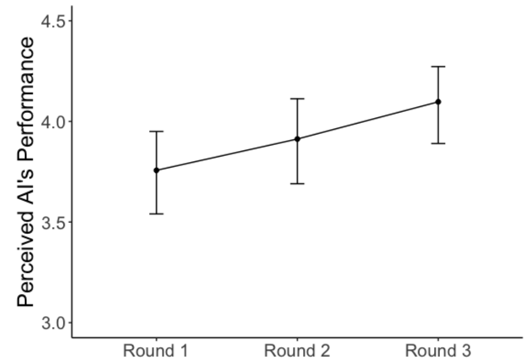
Task Round	Communication	Mean	SD	N
Round 1	Proactive	4.450	0.480	30
	Non-proactive	3.922	0.593	30
Round 2	Proactive	4.278	0.708	30
	Non-proactive	3.900	0.773	30
Round 3	Proactive	4.328	0.691	30
	Non-proactive	4.106	0.647	30

Perceived AI Teammate's Performance Our results showed that AI teammates with **proactive communication** ($M = 4.197, SD = 0.937$) were perceived to have better performance than AI teammates with **non-proactive communication** ($M = 3.647, SD = 0.543$), $F(1, 56) = 11.639$, $\mathbf{p = 0.001}$ (see Figure 3.4a). In addition, an increase in this perceived performance was observed as humans collaborated with the AI teammate longer. Specifically, participants considered AI teammates performed the best in Round 3 ($M = 4.097, SD = 0.761$), followed by Round 2 ($M = 3.913, SD = 0.832$) and Round 1 ($M = 3.757, SD = 0.817$),

$F(1.728, 56) = 96.745$, $\mathbf{p = 0.006}$ (see Figure 3.4b).



(a) Perceived AI's Performance Between the Proactive Communication AI and Non-proactive AI Teammate



(b) Perceived AI's Performance from Round 1 to Round 3

Figure 3.4: Team Performance

Satisfaction with the AI teammate Participants were significantly more satisfied with **proactive communication** AI teammate ($M = 4.653, SD = 0.616$) than **non-proactive communication** AI teammate ($M = 4.316, SD = 0.722$), $F(1, 56) = 5.800, \mathbf{p = 0.019}$ (see Figure 3.5). Similar to human trust in the AI teammate, AI taking actions automatically has no significant impact on humans' satisfaction with them.

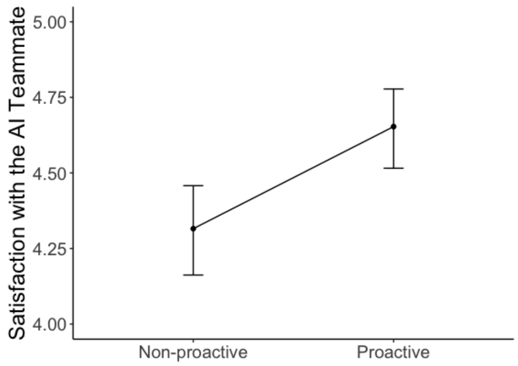


Figure 3.5: Satisfaction with the Proactive Communication AI Teammate and Non-proactive AI Teammate

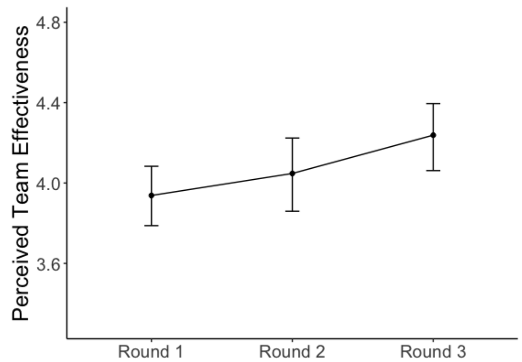


Figure 3.6: Perceived Team Effectiveness from Round 1 to Round 3

Perceived Team Effectiveness While no significant effects were found from AI's communication proactivity, task round has a significant impact on perceived team effectiveness ($F(2, 112) = 9.113$, $\mathbf{p < 0.001}$, see Figure 3.6). Participants viewed their teams as more effective as they collaborated with the AI teammate longer (Round 1: $M = 3.938, SD = 0.596$; Round 2: $M = 4.048$, $SD = 0.719$; Round 3: $M = 4.238$, $SD = 0.644$).

Team Viability Similar to the effect our data shows for perceived team effectiveness, task round significantly impacted participants' evaluation of the team's viability ($F(2, 112) = 10.054$, $\mathbf{p < 0.001}$, see Figure 3.7). Participants were more likely to consider their team would have the long-term success by the end of the team missions than at the beginning (Round 1: $M = 3.819, SD = 0.783$; Round 2: $M = 3.980$, $SD = 0.898$; Round 3: $M = 4.215$, $SD = 0.777$).

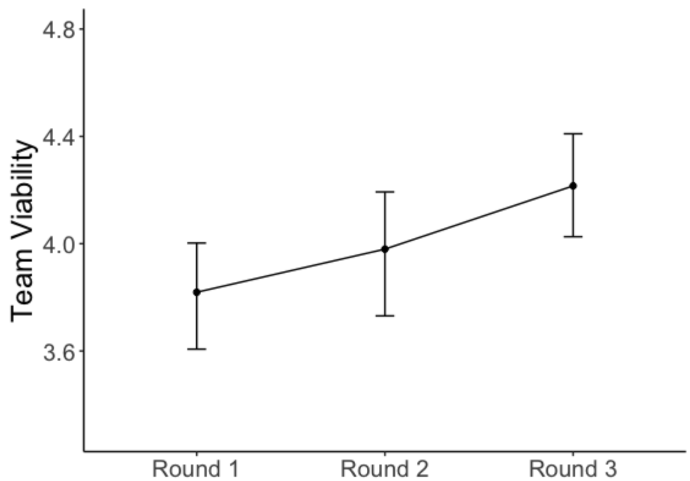
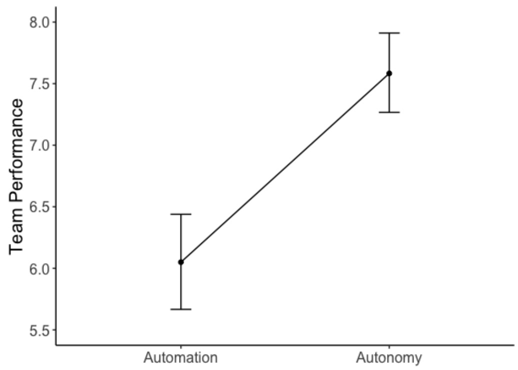


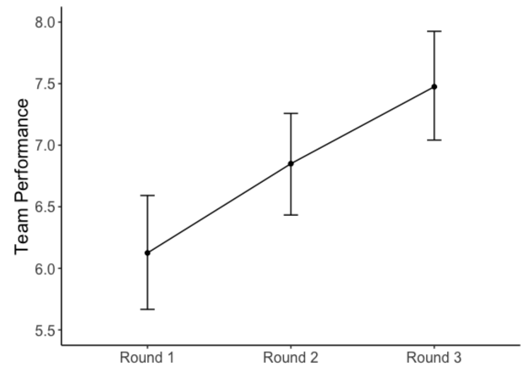
Figure 3.7: Team Viability

Team Performance Our data shows that AI's behavior type (e.g., autonomous AI vs. automated AI) has a significant impact on team performance. Autonomous AI teammate obviously has higher efficiency in completing tasks ($M = 7.583, SD = 1.500$) compared to automated AI teammate ($M = 6.050, SD = 1.856$), $F(1, 54) = 28.603$, $\mathbf{p < 0.001}$ (see Figure 3.8a).

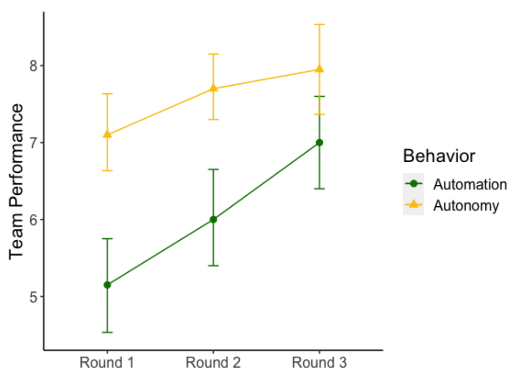
In addition, there is a significant interaction effect between task round and AI's behavior, $F(2, 108) = 5.867, \mathbf{p = 0.004}$ (see Figure 3.8c). Participants achieved higher team performance when they teamed up with an autonomous AI teammate, but teams with an **automated** AI teammate improved their performance more from Round 1 to Round 3 (Round 1: $M = 5.150, SD = 1.703$; Round 2: $M = 6.000$, $SD = 1.737$; Round 3: $M = 7.000$, $SD = 1.697$) compared to teams with an **autonomous** teammate (Round 1: $M = 7.100, SD = 1.417$; Round 2: $M = 7.700, SD = 1.222$; Round 3: $M = 7.950, SD = 1.734$).



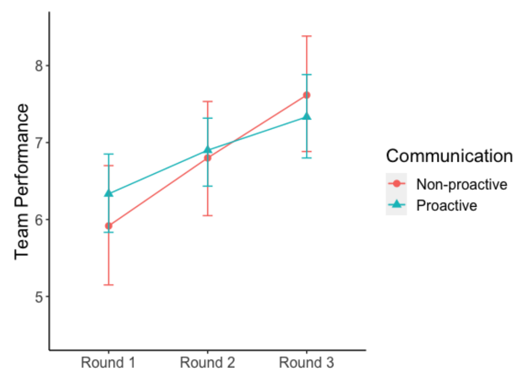
(a) Team Performance Between Autonomous and Automated AI Teammate



(b) Team Performance from Round 1 to Round 3



(c) Team Performance Between Automated AI Teammate and Autonomy Teammate



(d) Team Performance Between Proactive Communication AI and Non-proactive AI Teammate

Figure 3.8: Team Performance

Another interesting finding is the significant interaction impact between task round and AI's communication proactivity on team performance. Specifically, teams with a **non-proactive communication** AI teammate increased team performance more from Round 1 to Round 3 (Round 1: $M = 5.917, SD = 2.178$; Round 2: $M = 6.600$, $SD = 2.091$; Round 3: $M = 7.617$, $SD = 2.037$) than teams with a **proactive communication** AI teammate (Round 1: $M = 6.333, SD = 1.428$;

Table 3.5: Descriptive Statistics for Team Performance Score.

Task Round	Communication	Behavior	Mean	SD	N
Round 1	Proactive	Autonomous	6.867	1.141	15
		Automated	5.800	1.521	15
	Non-proactive	Autonomous	7.333	1.655	15
		Automated	4.500	1.669	15
Round 2	Proactive	Autonomous	7.567	0.704	15
		Automated	6.233	1.387	15
	Non-proactive	Autonomous	7.833	1.600	15
		Automated	5.767	2.052	15
Round 3	Proactive	Autonomous	7.500	1.476	15
		Automated	7.167	1.496	15
	Non-proactive	Autonomous	8.400	1.901	15
		Automated	6.833	1.915	15

Round 2: $M = 6.900, SD = 1.276$; Round 3: $M = 7.333, SD = 1.470$), $F(2, 108) = 3.571$, $\mathbf{p = 0.031}$ (see Figure 3.8d).

However, there is no significant interaction effect between the AI's proactivity in taking actions and AI's proactivity in communication on any measurements. Specifically, AI's proactive communication does not make up the negative impact caused by non-proactive behaviors on team performance. AI's proactive behavior also does not make up the negative impact resulted from non-proactive communication.

In summary, AI's proactivity in communication impacts humans perceptions to a large extent whereas AI being autonomous or automated does not impact human perceptions. Specifically, there are three main findings: (1) AI being proactive communication with humans positively impact human perceptions, including human trust in the AI teammate, perceived AI's performance and satisfaction with the AI teammate; (2) autonomous AI teammates only positively impact team performance; (3) while AI's proactive communication positively impacts human perceptions, non-proactive communication AI teams achieved better team performance in the last round.

3.4 Interview Findings

In this section, we first identify communication strategies that AI teammates are desired to apply to facilitate their coordination with humans in a dyadic teaming environment. Second, we describe how an AI teammate's communication influences human's situation awareness and trust during the collaboration process. Additionally, the NARS score is provided along each quote to indicate humans' existing attitudes towards AI (higher scores indicate a more negative attitude with 3 as neutral).

3.4.1 Communication Strategies for AI Teammates to Coordinate with Humans in Teaming Environments

Due to the dynamic feature of multiplayer online games, communication plays an essential role in facilitating the coordination between humans and AI teammates through information exchange. In our study, we identify four communication strategies that AI should apply to support their coordination with humans in an online teaming environment: (1) proactively communicating with humans; (2) employing balanced communication with both efficiency and sociability; (3) providing immediate responses; and (4) avoiding providing excessive amounts of communication once the communication pattern has formed in repeated team tasks.

3.4.1.1 Proactive communication from AI teammates is a must in HATs to facilitate team level information updates.

In the context of multiplayer online games, it is crucial for team members to proactively share updates, discuss next steps based on team progress at the moment, and take actions accordingly. Such proactivity in communication from the AI teammate is even more important in HATs for humans to be aware of AI's progress and adapt accordingly. For instance, participants highlight the importance of AI teammates being able to initiate a conversation:

> *It can go back and forth. They can also be the ones to give the direction in a sense. Obviously, we're more advanced as humans, but to be fair, it should be both (giving directions) on the same (level).* (P29, White, female, 18, NARS 2.14)

For P29, humans and AI teammates taking turns to initiate a conversation and give guidance creates an equal partnership (i.e., "fair") within the team even though humans are more capable of providing directions. Some participants such as P52 (White, male, 19, NARS 2.5) further highlight not only the importance of AI initiating a conversation, but also the direction of AI's communication: AI should proactively push information to humans helping reduce humans' cognitive load. P56 and P28 also mention,

> *Probably them pushing information to me (is more important than them pulling information from me) because they're more efficient, and I was just trying to see what was going on, especially if your progress depends on how far they are. Generally, the more experienced person needs to push information so the less experienced knows how to do that.* (P56, White, female, 18, NARS 3.93)

I would say them (pushing more) because they can go faster, and they know where everything is on the map, it'd be a little bit easier (than me pushing). (P28, White, female, 19, NARS 2.71)

According to these quotes, AI being the more proficient team member should guide the human teammate by initiating more conversations and providing directions. This could help humans develop their understanding and awareness of AI teammate's progress and take actions accordingly. Even though AI being the more competent team member should push information more than humans, it is essential for AI teammates to have bidirectional communication with humans. For instance, P21 and P47 share,

I like it doing both (pushing and pulling). I think back and forth because it emphasizes the team aspect of it. (P21, Hispanic, male, 19, NARS 3.36)

I think I'd prefer the push and the pull together (from the AI) just to feel like it's an actual interaction. (P47, White, female, 20, NARS 3.36)

As these quotes point out, the bidirectional communication from the AI fosters a collaboration environment by increasing interactions between humans and AI teammates. In contrast, when the AI teammate does not initiate conversations, participants usually feel that they lack necessary connections with the AI teammate, leading to their negative perceptions of the AI. For instance, P40 (White, female, 19, NARS 3.43) points out,

They were just getting crates quickly and they're back, but they never said anything to me, or it's just more just me asking them. It's kind of annoying. (It) wasn't like teamwork, was more just like, I'm the manager and they just do what they (are) told.

According to P40, even though an AI teammate demonstrates competent gaming skills, they may be perceived as *"annoying"* and not helpful if they never start a conversation with humans. As a result, the partnership between humans and AI teammates will no longer be equal: AI becomes a staff whereas human is their manager, despite AI's superior skills that can significantly contribute to team success. In this sense, it is even more important for AI to be equipped with the capability of initiating conversations with humans than demonstrating task-based skills.

3.4.1.2 AI teammates' communication style should balance efficiency and sociability.

Unlike humans who own personalized communication styles based on their various personalities, AI essentially does not own any communication styles. Rather, AI agents' communication patterns need to be carefully designed by AI designers and developers. Our study investigates human perceptions of AI providing (1) quick updates without social elements vs (2) social conversations.

On the one hand, participants show their strong preference on AI giving quick and straightforward updates without social elements in time-mattered team tasks:

> *I prefer straight to the point. I don't need the extra words because it's just pointless. I have to search through the words to make sure I figure out what it's actually telling me.* (P32, White, female, 18, NARS 2.79)

According to P32, quick and straight to the point messages are more efficient in a team task since it is easier for participants to extract important information and thus reduce their workload. P32 uses "pointless" to describe how she perceives information that is irrelevant to the task on hand (e.g., social phrases).

On the other hand, some other participants consider it as necessary for AI teammates to be able to communicate with humans socially, which will help humans build more personal connections with AI. P60 (Hispanic, male, 18, NARS 2.21) shares,

> *I would say little combination, obviously concise and to the point, but just little things, like little "great" after you say which one. It makes you more willing to work with it because it seems more like a person that's more friendly. In contrast, I'd be more willing to give them short responses and expect maybe more friendliness in return. Because they don't need that emotional reassurance, whereas people might appreciate that more.*

For P60, AI does not have emotions and thus does not necessarily need friendly feedback, but humans do. If AI teammates are able to have social conversations with humans, humans would feel more comfortable and more willing to work with the AI teammates. This indicates potential imbalance between communication from human to AI teammates and from AI teammates to humans. Other participants also agree that AI being conversational would be more comfortable to talk with, thus helping them better work with AI:

> *I like having a conversation. That would be a little more comfortable for me. Because I like having conversation to be able to elaborate and it's not always so cookie cutter, always as straight edged. I would like to have a little more of a conversational response when talking to my teammate.* (P19, Hispanic, male, 19, NARS 3.29)

> *Probably conversational. It's just more personal.* (P50, White, female, 19, NARS 2.93)

> *I really liked how he said, sounds good. It made it seem more casual, a lot normal, like speaking to a real person. I kind of disliked how the*

commands were very authoritative, I suppose. Maybe to make them more casual, in my opinion. I know technically they don't have any feelings it still made me feel better. (P41, White, female, 18, NARS 1.93)

According to the above quotes, AI that is able to communicate with humans socially are more human-like. This also positively affects humans' collaboration with AI teammates. As P41 explains, social conversations with AI would improve her perception of AI even though she clearly understands that the "social elements" are generated by a machine.

The perception of AI being more human-like based on their communication styles could even foster humans' personal connections with their AI teammates. For instance, P20 and P38 describe,

> *I feel like a lot of people, and me included, would want social aspects, like make it feel more like a friend.* (P20, White, female, 19, NARS 2.36)

> *If it's an everyday kind of thing, I feel like a lot of people, and me included, would like more conversational. It would make me feel they weren't just a machine. It'd be more personal connection, feel like I'm talking to a person, which would be more comfortable.* (P38, White, female, 18, NARS 3.86)

For these participants, AI's ability to communicate with humans socially may lead to close personal relationships with humans. For example, they can even be viewed as human's friend. This type of relationship, therefore, is likely to positively impact human-AI collaboration in a teaming environment.

3.4.1.3 AI teammates should always provide immediate responses to humans to facilitate coordination in HATs.

Unlike humans who could make independent cognitive decisions on when to talk, how to talk, and what to talk about with their teammates, AI teammates need to be designed to talk. This thus makes communication between humans and AI teammates more challenging and unpredictable. In such situations, responses are considered important as it confirms that the AI teammate has understand what the human said and agreed to it. In our study, participants express their appreciation of such confirmation:

> *I appreciated how they confirmed that they were going to do the task I assigned them with.* (P51, White, female, 18, NARS 3.57)

> *If it didn't say anything back, I'd be a little less confident.* (P26, White, male, 18, NARS 3.00)

For these participants, the confirmation from AI teammates helps the team to proceed in team tasks by showing AI's clear understanding of their responsibilities and team goals. The speed of the response is also a factor reflected in human's perceptions of AI's communication ability, for example: *"Very responding, and is pretty fast responding too"* (P10, Asian, female, 18, NARS 3.50); and *"He responds to me pretty fast"* (P46, White, female, 20, NARS 2.79). AI teammate's immediate responses increase human's confidence in the AI teammate's actions and further benefits their coordination.

In contrast, some participants point out that AI teammates do not need such responses from humans, even though receiving responses from AI teammates is valuable to humans:

I'm bad at video games. I was trying to collect my crate and they there was some questions where they would ask me, What crate are you collecting? And I can't multitask. I was trying to focus on getting the crate, so I couldn't really respond. Well, it's a bot. It doesn't really need confirmation from me. (P27, Hispanic, female, 19, NARS 2.36)

P27 present two reasons why humans do not need to respond to AI: (1) humans may have poor gaming skills which makes multitasking extremely difficult; (2) their AI teammate probably does not need such confirmation from humans considering that it is just a computer program. In making the trade-off between completing their own task and responding to AI teammates, P27 chose the one that was considered more necessary, i.e., performing the task. Other participants share similar opinions on humans responding back to AI:

I didn't see any benefit from me giving information to them. (P25, White, male, 18, NARS 3.43)

If he could send me more updates that went in peep that I wouldn't have to respond to, it would be better. But I think the only detrimental thing would be is if he sent more stuff that I had to reply to. I think that could slow down our progress. (P17, White, male, 18, NARS 2.93)

As these participants mention, when humans have to respond back more, they have less time to focus on their own task, which eventually hurt their team performance. Rather, not responding to AI teammates could be beneficial to humans, as P44 (Black, female, 21, NARS 2.86) says:

I was able to give the information but when he was asking information, it was hard for me to respond without trying to drive and stuff. So it wasn't balanced, but it benefited me.

P44 emphasizes that the imbalance of responses between humans and AI is favorable to humans. Ignoring AI's request allows them to focus on their own task more, mostly due to the difficulty of multitasking. However, this changes with humans' game skills. Some participants such as P11 (White, female, 18, NARS 3.64) point out that with better game skills, they would be more likely to respond to AI:

> *I feel like I don't have time to respond back until I drop crate off. If I was better at it, I would definitely respond to them more often during the game and probably be faster. And then maybe if I was the better one at it, maybe I would then take control.*

In summary, our findings show that responses from AI teammates are appreciated and seem to positively affect human-AI coordination by confirming that they understand humans' messages. However, while humans appreciate AI's responses, they sometimes choose not to respond but focus on their own responsibility. This thus highlights the potential imbalance of communication between humans and AI teammates, which should be considered in designing future AI communication feature design.

3.4.1.4 AI teammates should avoid providing excessive amount of communication to humans once the communication pattern has formed in repeated team tasks.

The amount of communication that an AI teammate can produce is another important factor that impacts how their communication is perceived by humans, and even impacts their collaboration with humans. Achieving a deep understanding of the ideal amount of communication AI provides plays a crucial role in designing and implementing AI's communication with humans. While little amount of communication probably is not enough for humans and AI to coordinate smoothly, too much communication, on the other hand, is likely to cause issues like distracting or reducing effectiveness in completing team tasks:

> *It (AI's communication) was a lot. It was constant. So it kind of distracted me at some points. So I'd say a little less than that.* (P56, White, female, 18, NARS 3.93)

> *Rambling would have just been too much because I was focused on getting my own crates.* (P1, Black, female, 19, NARS 2.71)

As P56 and P1 point out, large amount of communication from the AI teammate could distract humans from doing their own task, especially when humans have to multitask:

> *I think talk to me less would be better. Because it's just easier not to have to think about having to type back again, just do what I'm supposed to be doing. I like they type things like, "Okay, I'm going to pick up this one." stuff like that. But they've said it so many times for the same exact crate. I was like, oh, I know that already. I don't need to hear it again.* (P58, Hispanic, female, 18, NARS 3.93)

For P58, once the communication and cooperation patterns have been established, AI proactively providing such predictable information will become redundant. This points to the necessary need for AI's flexibility in their communication, such as more communication in the beginning of human-AI collaboration but less communication once the collaboration routine has been developed.

In sum, participants highlight multiple communication strategies applied by AI teammates that are crucial in shaping their perception of AI teammates and their coordination. First, AI being proactive in sharing information is a must to smooth coordination and effective teamwork. Second, AI should always provide immediate responses to humans to confirm that humans' messages have been received, understood and will be processed. Third, AI teammates' communication styles should be balanced between efficiency and sociability. Specifically, in time sensitive tasks, straightforward and quick updates are more preferred while conversational communication is more preferred in scenarios where humans care about the personal connection with the AI teammate. Last, excessive amount of communication from AI teammates should be avoided once team communication pattern has formed. This excessive amount of input from AI teammates may cause distraction or interruption in humans completing their responsibilities.

3.4.2 The Impact of an AI Teammate's Communication on Team Process During Human-AI Coordination

In this section, we explain how AI proactively communicating with humans facilitates their coordination through two teamwork elements: trust in AI teammate, and team situation awareness. We also describe how AI teammates lacking proactive communication can force humans to utilize implicit communication and hinder human-AI coordination in a teaming context.

3.4.2.1 AI teammate's proactive communication aids trust development by benefiting human's individual performance and increasing transparency as a reliable partner.

In human-only teams, trust plays a crucial role in how well team members can coordinate with each other to perform shared tasks collectively. Compared to human-only teams, trust in HATs could be even more important in shaping team outcomes given that humans have potential bias towards AI [177], which may result in lower level of trust in AI than trust in a human stranger. Therefore, it is important to explore how trust develops between humans and AI teammates through the collaboration process. In particular, our study shows that communication, which has been considered an important element in trust development within virtual human-only teams [83], facilitates human's trust in the AI teammate in three ways. First, AI's proactive communication *benefits human's individual performance* which leads to trust development of the AI teammate. For instance, P56 shares,

> *I was trusting it because he was constantly communicating with me. It was constant and he was asking me what I was doing too, so kind of kept me on task. It helped me trust it more. Them getting crates was helpful to the task, but their communication was better for my performance.* (P56, White, female, 18, NARS 3.93)

According to P56, AI's proactive communication facilitates how humans develop trust in the AI teammate by benefiting humans' individual performance. In particular, P56 highlights the role of AI's proactive communication and AI's competent game skills in the collaborative task: AI teammate's proactive communication contributes more towards humans' individual performance, while the AI teammate fulfilling their responsibilities contributes more towards the team performance. In this sense, AI's proactive communication encourages humans to trust them more as a teammate. Even though team performance is always an important indicator in evaluating how a team performs, human perceptions of the AI teammate are likely to impact how well humans coordinate with the AI teammate. Positive perceptions towards AI teammates could be beneficial to the HAT *in the long term.*

Second, AI teammates proactively communicating with humans is perceived as an approach to increase transparency of AI's behaviors, leading to higher human trust in the AI teammate. For instance, P52 (White, male, 19, NARS 2.50) mentions,

> *I think it's just like transparency.The more you know, the more (you're) confident that it's doing what it's supposed to and it's not malfunctioning or anything.*

P52 highlights the importance of transparency in trusting the AI teammate. More transparency allows humans to better predict AI's actions, with trust increasing through the process. Similar thoughts are also shared by P58 and P22:

> *I trusted it. They kept really good contact with me, to make sure we're both on the same page. I pretty much trusted it. I figured it'd do a good job.* (P58, Hispanic, female, 18, NARS 3.93)

> *Very trusting. Because obviously they told me the number (of the crate they were picking up), and they dropped it off. It was pretty quickly trusting.* (P22, Black, male, 18, NARS 3.07)

P58 mentions that AI teammate's constant communication regarding the team task ensures humans are on track of the team progress. Importantly, the positive perception generated by AI's proactive communication leads to humans believe in AI's performance. P22, rather, highlights how fast this transparency develops humans' trust in AI teammates. AI showing the transparency is also interpreted as AI willing to collaborate with humans as a team and taking the responsibility as a team member to inform humans their progress, as P39 (Hispanic, female, 19, NARS 3.29) says, *I feel like it's good to communicate. So that way, you know where the other person is at. I think it was also the communication that makes you trust them. Because they (were) just letting you know.*

Third, humans perceiving AI teammates as a reliable *partner* also facilitates trust development. P37 (White, female, 19, NARS 3.36) emphasizes the importance of AI teammate being a "real" teammate:

I think it did help me trust more because in return I was getting response rather than just going off on my own and doing it.

AI with proactive communication is perceived as a teammate whom they can work together and whom they can trust to have their back. Some participants who teamed with a non-proactive AI express that better communication from the AI teammate will make their coordination more like teamwork:

I feel like if he had better communication, I could trust him more, just to have my back or see that crate I was doing. Like we were in it together. (P49, White, Non-binary, 20, NARS 2.93)

If it talking more, I would've trusted it a lot more. Because I'd have more communications with him. There's more comfort with him. (P59, Other, male, 19, NARS 3.50)

For P59 and P49, AI with proactive communication could have developed more trust by building better partnership with humans.

According to these quotes, AI's proactive communication contributes to human's trust development in the AI teammate through three ways: (1) benefiting human's performance on team tasks; (2) showing transparency of AI teammate's behaviors; (3) human perceiving AI as a reliable partner. Specifically, when AI's communication was beneficial to humans in any way (e.g., helping with human's individual performance or human's understanding of the teamwork progress), it makes humans trust the AI teammate more.

3.4.2.2 AI teammates' proactive communication develops situation awareness by informing AI's progress and indirectly helping humans make next-step decisions.

In a teaming environment, situation awareness plays a crucial role in forming collaboration pattern, especially enabling a team member to be aware of the team's progress and other team members' actions to perform accordingly. Our study shows that proactive communication from AI teammates plays a positive role in developing humans' team situation awareness from two perspectives.

First, AI proactively communicating with humans enables human teammates to develop an awareness of *what the AI teammate is doing.* For instance, both P50 and P39 highlight the positive impact of AI consistently communicating with humans:

> *It is nice that we were interacting so we knew what he was getting with.* (P50, White, female, 19, NARS 2.93)

> *Good parts about it [AI's communication] was that you constantly knew what they were doing.* (P39, Hispanic, female, 19, NARS 3.29)

In contrast, lack of communication results in humans' uncertainty of AI's actions and even frustration perceptions, as P34 (White, female, 18, NARS 3.00) mentions,

> *They were only communicating when I said something first. So it wasn't great communication. I was just unsure about what they were doing the entire time.*

For P34, AI not communicating proactively leads to unpredictability of the AI's progress, and even the team's status at the moment. This lack of awareness and understanding of other teammates' pace during collaboration is extremely difficult, and may product poor team outcomes. P7 (White, female, 18, NARS 3.57) echoes this view, feels that AI teammates lacking communication makes humans struggle tracking their actions and progress during gameplay, resulting in low team situation awareness. Further, this lack of team situation awareness increases the difficulty of coordinating with AI teammates and completing the task efficiently. Rather, if AI teammate proactively shares the updates, it would help humans develop such situation awareness.

Second, proactive communication from AI teammates assists humans to apprehend team progress and *make decision on their own next step accordingly.* For instance, P6 and P42 highlight,

> *I thought they did a good job for what we needed to be communicating about. It was very straightforward and easy to understand what they had accomplished. I could naturally understand my own progression through the task because of what they were communicating.* (P6, White, female, 18, NARS 2.29)

> *I thought the communication was, well, it definitely allowed me to understand which crate I was supposed to be getting. It also let me know how fast AI was moving. Or if I needed to pick up my pace or slow down so that I didn't out run it.* (P42, White, male, 19, NARS 2.57)

For P6, through the AI teammate's proactive communication on their progress of the team task, humans can further figure out what they should do next to coordinate with the AI teammate. According to P42, AI's proactive communication provides information for humans to develop an awareness of AI teammate and the team's progress at the moment and how humans act accordingly.

Oppositely, an inadequacy of this proactive communication from AI teammate increases the difficulty of humans proceeding team tasks and making a decision on their following action , as P8 (White, female, 18, NARS 3.07) suggests,

> *I just had to ask them what they were doing. They weren't supposed to ask me, but it was harder to have to worry about what they were doing and what I was doing when they didn't really know what I was doing unless I told them or ask them. I just didn't know what they were doing. So made it hard to figure out what I was supposed to do. So them just telling me without me having to ask, would make it faster.*

For P8, when AI teammate communicates non-proactively (i.e., only giving responses), the cognitive workload on the human's side gets substantially higher. Specifically, AI teammate not proactively sharing information forces humans to pull information from the AI teammate, increasing humans' the stress and workload. In addition, this lack of situation awareness makes it more difficult for humans to coordinate accordingly during dynamic gameplay.

Moreover, if AI teammate's proactive communication can guide humans on next steps of the team task, it will further facilitate the development of humans' awareness of the team progress:

> *Zeus only communicates when I initiate a communication. I think he should initiate his communication (about) what he's doing, and maybe,*

telling me what to do as well, so I have a better idea. (P46, Asian, male, 18, NARS 2.79)

As P46 elaborates, AI initiating conversations in their collaboration facilitates the development of participant's awareness of AI's progress. This situation awareness allows participants to govern the overall course of the teamwork. P46 also points out that AI, as a more skilled team member, can guide them in completing the team task.

3.4.2.3 AI teammates lacking proactive communication is perceived as an individual rather than a teammate, which hinders coordination between humans and AI teammates.

When AI does not communicate with humans enough, it is difficult for humans to coordinate accordingly. P4 (White, female, 18, NARS 3.64) points out:

> *Lack of communication means that there's obviously going to be less trust in the AI system. And then, of course when there's less trust and less communication, obviously, you're gonna get frustrated more, and not necessarily want to use the AI. So like a team would crumble. There wouldn't be a team. It'd be like two individual players.*

Here P4 points out that communication functions as the glue that holds the team together. Lack of communication creates an unhealthy teaming environment, where humans perceive the AI teammate as another *individual* who performs the same task, rather than a teammate. In addition, little amount of communication hinders human from developing trust in the AI teammate and produces frustration perceptions. Likewise, P10 (Asian, female, 18, NARS 3.50) expresses thoughts on AI being too individualistic:

(What they can do better was) just tell me when they were dropping off the crate, so it wasn't just totally me trying to figure out where everything was.

According to P10, the AI teammate not communicating much with humans makes it more difficult to work on the task together. Instead, the progress of completing team tasks are more like humans working on them independently rather than working with a partner as an unity. In contrast, AI being proactive in communicating with humans presents team effort and shortens the distance between humans and the AI teammate. P44 (Black, female, 21, NARS 2.86) elaborates more on how communication shows team effort:

When you're talking to someone (during) completing a task, it displays team effort, great accountability. You can rely on them because you don't have to worry about what they're doing, because they're letting me know. It creates a more solid foundation when words and people express their actions.

For both P10 and P44, AI's communication is not only a way of passing information to humans, but also showing AI's proactivity in completing shared team goals as a part of the team. Thus, AI communicating their progress proactively and even helping humans through communication indicate that they are actively working on the team task, "care" about the team and are willing to take the responsibility as a team member. Too little communication in teamwork could create an imbalanced unhealthy teaming environment due to insufficient information sharing and incorrect prediction of teammate's decisions.

3.4.2.4 AI teammates lacking proactive communication forces humans to utilize other implicit communication approaches in HATs.

While AI teammates' proactive communicate could facilitate maintaining humans' situation awareness, implicit communication could be used as an alternative to maintain it. Our interview data shows participants take advantage of implicit communication cues in team tasks to maintain their understanding of what the AI teammate is doing and the team's progress. For instance, P37 (White, female, 19, NARS 3.36) shares:

> *So once I had figured out, I could go and see on the map, like this crate is missing, they just dropped off this one. So I'm getting this one.*

For P37, maps in game are used as an implicit communication cue, which help humans know which crate the AI teammate has collected and which one they should collect. In other words, humans can utilize the implicit communication cues, i.e., AI's actions and AI actions' results, to keep their awareness regarding their team progress.

Another type of implicit cues, audio, is mentioned to help maintain team situation awareness:

> *Every time I would go back, I would see him coming. So I knew which crate he was going to. And then same thing other like the other way around, so I kind of didn't feel the need to type in chat as much. I figured it was taking a lot of time off anyways. So (it) ended still working out.* (P23, White, male, 18, NARS 2.07)

According to P23, the implicit communication (i.e., text communication) is time-consuming, whereas seeing the AI teammate in the 3D space in game provides enough information for humans to coordinate and complete team tasks. This implicit cues ensures that humans are aware of the AI teammate's actions in the environment, understand AI's actions and know the projection of their status followed by the current action. P49 (White, Non-binary, 20, NARS 2.93) also mentions that the explicit communication is inefficient whereas humans can take use of implicit communication:

> *I just tracked his moves on the map, and looked and saw what crate he was doing. The chat took a minute to type everything out. So I wouldn't do that, just rather look and see what he was doing. It (AI's communication) was kinda poor.*

Since the AI teammate is not capable of sharing information proactively, P49 believes that humans may prefer to use implicit communication over pulling information from AI teammates through explicit communication. Checking the map enables humans to track AI teammate's movements and task progress. However, another participant points out that even though implicit cues are helpful in developing team situation awareness, explicit communication initiated by AI teammates would reduce workload on the human teammate's side:

> *The only thing is, it's hard to monitor when they were going back and forth, or whenever they were done, because I either had to keep like looking at the map or see them in passing. So I think it would have been easier if they were like, Oh, I just dropped this one off.* (P28, White, female, 19, NARS 2.71)

For P28, both checking on maps or noticing AI teammate driving by are difficult to utilize in team coordination. Instead, if the AI teammate has the capability to share their updates on team tasks, the coordination between humans and the AI would be easier while they would still be aware of AI's actions and team progress.

In summary, AI proactively communicating with humans supports the development of human trust and maintains team situation awareness in various ways. Specifically, AI's proactive communication assists trust development in the AI teammate through benefiting human's individual performance, presenting transparency of their behaviors, and being perceived as a reliable *teammate*. Such proactive communication from AI teammates also develops and maintains team situation awareness by informing AI's progress, helping humans move on with their shared responsibility, and even guiding humans on next steps of game tasks. In addition, AI lacking proactive communication could hinder humans' coordination with them and even generates negative perceptions, such as frustration.

3.5 Discussion

In response to our research questions, our findings have highlighted that humans seek that AI teammates employ four communication strategies to support dyadic HATs: (1) proactively communicating with humans; (2) employing balanced communication with both efficiency and sociability; (3) providing quick responses; and (4) avoiding large amounts of communication once the communication pattern has formed in repeated team interactions (RQ1). In addition, AI teammates proactively communicating with humans can support their coordination with humans in a dyadic HAT by developing human trust and team situation awareness in teaming environments, whereas AI teammates lacking proactive communication are perceived as an individual rather than a team member, which hinders team coordination (RQ2).

In this section, we first discuss how our findings extend current knowledge on communication in dyadic HATs and human-only teams in CSCW. We then propose three key elements for human-AI communication in 1:1 teaming environments grounded in our findings and prior work on communication in computer-mediated collaboration. Last, we discuss how these three components can be extended to dyadic HATs in other contexts.

3.5.1 Communication Strategies for AI Teammates and Their Impact on Team Processes

Our study extends current CSCW work on communication in dyadic HATs by providing a holistic view of AI teammates' communication strategies through the lens of human perceptions and experience.

A specific highlight of our findings is the importance of AI teammate's proactive communication in the dyadic teams that were studied. Within both the context and composition of teams studied, people perceive AI that proactively shared information as a reliable partner and teammate, but view non-proactive AI as loners rather than team-players. As such, humans that interact with non-proactive AI in these contexts and compositions might not perceive their interactions with the AI as collaborative. Unfortunately, lacking this sense-of-team could have adverse effects on these dyads, hindering the coordination between the human and the AI. Critically, this supports prior work that has identified the impact of proactive communication on the formation of individual [305] and team [46, 89] processes and perceptions. This work also extends our understanding of how AI teammates can be social actors [185], in that the identity of these actors (i.e. teammate) in this context and composition is in fact influenced by the proactivity of an AI teammate's communication. As such, it is critical to consider the inclusion of proactive communication in designing an AI to be a "teammate" within these dyadic contexts.

However, it is worth considering whether the preference for proactive communication would apply beyond the examined context and team composition. For instance, for team tasks that are more decentralized, each team member has an isolated responsibility with low interdependence with other team members. The completion of such tasks is less reliant on team communication [127]. Therefore, proactive AI communication might not be as useful and desirable as for interdependent tasks. In regard to team composition, HATs with team composition more complex than dyads may not benefit as much from AI's proactive communication, as the humans in such HATs can quickly become overwhelmed by the amount of information pushed by each AI teammate, and their workflow interrupted [217].

Another important insight is that an excessive amount of communication from AI teammates could negatively impact human-AI coordination. This is in line with previous work on human-only teams demonstrating that team coordination that requires low communication volume usually yields better team awareness and high efficiency [157, 158]. Our findings were able to pinpoint where the problem of a high volume of communication lies between humans and AI teammates in dyadic HATs. It appears that well-established communication patterns formed throughout the interactions within a 1:1 HAT eliminate the need for large amounts of communication; and the timing of communication is critical as to not interrupt and sidetrack task coordination among the two team members. After all, interruptions could lead to incomplete team tasks and even severe mistakes [87]. The identification of these two aspects may help explain the inconsistent results of communication volume on team performance in 1:1 HATs [53, 292]. Importantly, the amount of communication is likely to increase in HATs with more complex team compositions, which may result in information overload and impact human-AI collaboration.

In summary, these communication strategies that humans desire AI teammates to utilize and their impact are crucial to 1:1 human-AI team communication design.

Our study extends existing work on human-AI communication in dyadic teaming environments and provides new insights for future AI communication design. These new perspectives could be used as a foundation and combined with previous research on communication in HATs and human-only teams to better structure human-AI communication for both dyadic HATs and HATs that involve more than one human teammate and one AI teammate. However, potential risks and ethical issues should be considered while applying these communication strategies on AI. First, humans need to be aware of the information's accuracy from the AI. Research has shown that appropriate trust calibration (i.e., humans knowing when to trust and when to distrust an AI) is crucial to success in human-AI collaboration [302]. AI proactively communicating intentionally inaccurate information could be perceived as unethical and further result in negative team outcomes. Second, trust in a human teammate and an AI teammate needs to be balanced in a triad or more complicated HATs. In the past decade, plenty of work has endeavored to explore how to increase human trust in AI for better human-AI collaboration [19, 259]. However, it could be risky when the trust in AI overweights the trust in human teammates, especially in certain contexts like military and healthcare.

3.5.2 Three Key Elements in Designing Communication in Dyadic HATs

Grounded in our findings and existing communication principles in previous studies on computer-mediated collaboration, we propose three key elements in designing communication for dyadic HATs: AI's communication strategies, communication goals, and humans' communication requirements. We will discuss each key element using highlights from our findings along with insights from previous literature, as well as the application of these elements beyond gaming.

3.5.2.1 Three Key Elements in Human-AI Communication

Team communication is a key factor in supporting both cognitive and affective processes [173, 143]. While previous CSCW research has explored communication in HATs through various attributes (e.g., communication quantity [172, 53] and communication frequency [200]), this study provides additional insights on how humans perceive AI's communication and how it then facilitates their coordination in dyadic HATs. At a team level, we synthesize how each party of the team (humans and AI) should be designed to achieve effective team outcomes.

First, *communication goals* play an essential role in building effective team communication by facilitating team members to coordinate smoothly [255, 218, 97]. Specifically, humans and AI teammates have different needs for information, and as such human-AI team communication strategies should explicitly differentiate communication needs. On the one hand, for human-to-AI communication, the content communicated is expected to center around *must* information for AI teammates to make decisions [38]. To maintain the interaction between humans and AI teammates, AI teammates need to make decisions with humans providing task-related input that is not accessible to AI teammates. On the other hand, one essential goal in AI-to-human communication is to develop human trust in the AI teammate and their awareness of team processes, as indicated by previous work on human-only teams [116]. In this sense, AI teammates usually need to actively communicate information which helps humans to develop situation awareness (e.g., sharing AI's task progress and confirming task needs with humans). In addition, this information provided by AI teammates is necessary for human teammates to make decisions, which benefits team coordination and teamwork progress.

Second, to better achieve the pursuit of the communication goals, it is necessary to design *AI's communication strategies* in an understandable and efficient way for team members to communicate [1, 121]. The communication strategies we proposed in this study aim to provide insights on how an AI teammate should apply communication to build trust and achieve high team outcomes (i.e., AI-to-human communication goals) with the human teammate. Structuring AI-to-human communication requires the inclusion of multiple different strategies to best enable the goals above. One example of a communication strategy is having AI teammates provide immediate responses, which is critical for humans to ensure communication is received and well interpreted by AI teammates. This finding supports previous research that identifies the importance of such responses in human-only teams with non-collocated communication [201] and even networking between computers, which fundamentally requires the use of responses [286].

Third, *humans' communication requirements* heavily impact their own collaborative experience with AI in dyadic HATs. While the manner in which AI's communication strategies should be structured is crucial in facilitating team coordination, how humans perceive them and how humans prefer to react to AI's communication largely shapes whether their communication goals could be achieved. Our findings indicate that humans want to minimize their own burden of communication, especially when they have specific responsibilities to fulfill. This supports previous work that indicates that distraction and interruption, which hinder humans from completing their own task, could result in negative outcomes in safety-related tasks [237]. While humans need to provide AI teammates the necessary inputs for AI's decision making, the communication of these inputs should not be a burden to human teammates. It is interesting to see that while humans expect AI teammates to provide immediate responses to their messages, they also prefer not to be *required* to respond to AI's communication. This imbalance between humans and AI supports a large amount of work on human-AI collaboration that indicates that AI is treated differently from humans [301, 177, 280]. Taking this imbalance of humans-AI communication into consideration, dyadic HAT communication design can utilize team communication more effectively and develop a trustworthy teaming environment.

3.5.2.2 Future Application of the Three Key Elements in Dyadic HATs

The fast-changing and context-dependent features of team communication make it challenging to design and examine in HATs. Research on human-only team communication has suggested the essential role of team characteristics, team roles, and tasks in impacting team communication [263, 226, 303]. In addition, it is recommended that teams deploy communication strategies according to the specific task [258]. In this section, we will discuss team characteristics and context dimensions in our study, and how the three key elements could be extended in other contexts by comparing against these features.

Using previous work on team characteristics and context in human teams as a foundation [263], we selected and adapted the context dimensions that can be depicted in our study, as shown in Table 3.6. These team and task dimensions identify the context in which the proposed communication strategies were developed. To extend the three elements into contexts, a comparison between our team/tasks characteristics and target team/context is necessary to ensure a more accurate application. Below we discuss how the proposed three key elements would be applied beyond the context of this study using *team size/composition* and *situational stressors* as an example.

Table 3.6: Team or Task Characteristics in Our Study

Dimension Type	Dimension	Details
Team	Roles	Human and AI share the same responsibilities in this task.
	Size/Composition	A dyadic HAT composed of one human and one AI teammate.
Task	Task Type	Each team member's task involves two parts: (1) figuring out which crate to get in numeric order through team communication; and (2) get the crate and drop it off.
	Communication Method	A text-based chat channel is provided for team communication.
	Situational Stressors	Eight-minute timed task with a timer displayed on the screen.

While many teams perform in a 1:1 setting, such as HATs in data science [280] and healthcare [38], plenty of research on HAT has explored triads (e.g., three-member HATs composed of at least one human and one AI) [259, 171]. For non-dyadic HATs, communication goals are likely to stay the same in terms of developing human trust in the AI, maintaining humans' situation awareness, and ensuring the team task proceeds with needed information. However, AI's communication and humans' communication requirements may be slightly different. For example, a HAT where a human collaborates with multiple AI teammates may not benefit from proactive cycles as much as our findings indicate due to information overload resulting from a multitude of AI pushing information [217, 50]. Yet, in this type of HAT, AI would still need to push information, but the rate of information may be slow and the information may be abstracted to meet the needs of humans [217]. This adaption of AI's communication strategies would help future researchers to utilize the three elements more accurately. As research in HAT continues to develop, these three key elements could be used for future research to build upon and even updated and contextualized to best serve specific teams and contexts.

Situational stressors also play an essential role in impacting team communication [263]. Our study used a time-sensitive task, which requires immediate responses and task-related updates to proceed with team tasks effectively. However, in contexts without such time stressors, team processes would be much slower. In this sense, emergency healthcare and human resources provide two examples that are impacted by time stress differently. Emergency healthcare environments often have extreme time pressures that require stronger interdependence and situation awareness to ensure success [135]. These teams would probably benefit from communication strategies that benefit situation awareness, such as those found in this study. On the other hand, human resources, which is often a slower-paced, procedural, and formulaic environment [85], might benefit from communication strategies that do not disrupt human workflow, such as the use of non-proactive communication. However, the team communication should still aim to build human trust and support team performance.

Given the broad range of different contexts as well as team characteristics, two critical steps are important to ensuring the extension of our findings in other contexts: (1) identifying the team characteristics and task dimensions of the target context; and (2) adapting three key components according to the target context using previous literature as a support. For step (1), the human-centered design of AI teammates requires context specific consideration, and research should work to identify which team or task dimensions in Table 3.6 are overlapped with the target context. For step (2), empirical experiments and principles in previous literature can be used to identify necessary modification of each communication element. Following these two steps, future research could use the three key elements in human-AI communication proposed in this work as a start point to further explore the design of human-AI communication in HATs under various contexts.

3.5.3 Limitations and Future Work

This study has several limitations. First, it is important to note that the dyadic team setting in this study may impact how the findings can be applied in other settings. The AI teammates' communication strategies identified in this study may look slightly different with more complex team compositions, such as teams with multiple AI teammates or multiple human teammates. However, this work serves as a foundation for future AI communication research to build upon. Using the three key elements proposed in this study as a base, future work can develop their communication strategies for multi-human or multi-AI HATs based on future findings regarding team composition. Second, this study utilizes a team setting where the human teammate and the AI teammate share the same responsibility. Findings may be slightly different in HATs with more unique roles. Future research should explore what communication strategies humans expect AI teammates to utilize when they take different roles in a team. Third, this study only examines communication between humans and AI teammates in a specific context. While context is always a critical component to HAT, it would be impossible to examine every potential context despite the impact of context on human-AI team communication. Thus, future work should examine and extend the communication strategies outlined in this study to other contexts. Fourth, all participants were college students with an average age of 19 years old. Prior research has shown that age could impact how humans perceive technologies [77], both of which could impact their trust in the systems [267]. In addition, our participants are not experienced in completing the task in our study. As novice individuals, their experience in a specialized HAT task might yield slightly different results from people who are experts in a real-world HAT task. Thus, future work should explore human-AI communication in dyadic HATs with a more general population, which would provide a more complete understanding of human-AI communication in teams. Lastly, this study only considers the situation where AI teammates' communication is always accurate, which is difficult to achieve in real-world tasks. Given that AI's decision-making accuracy could heavily impact humans trust in the AI and the collaboration process [211, 298, 191], future work should further explore how AI teammate's communication is perceived with different communication accuracy and how that further impacts human-AI coordination in dyadic teams.

Chapter 4

Study 2: Exploring Explanations in AI's Communication Content

4.1 Overview and Research Questions

While Study 1 explores AI teammates' communication proactivity, what AI needs to communicate when AI proactively communicates with humans is still an important component of communication to investigate. Given that communication content is highly task-dependent, this study focuses on AI's explanations of their actions, which is an important part of *communication content*. A large body of work has endeavored towards generating explanations for AI agents' decision-making process [147, 282]. Such explanations provided by AI agents intend to increase human trust [241, 285], which is a crucial element in achieving effective teamwork [170]. However, does AI *always* need to provide an explanation of their behaviors? Does AI's explanation *always* develop higher human trust in AI? To explore these issues, this study focuses on the impact of AI's explanations, part of communication content, on human perceptions (e.g., trust and perceived effectiveness) in multiple human-AI teaming scenarios. Specifically, this study addresses the following research questions:

RQ1: How does the utilization of explanations in AI's communication impact trust?

RQ1.1: How does the identity of a teammate (human or AI) influence the explanation's impact on trust?

RQ1.2: Does the action performed by a teammate change the explanation's impact on trust?

RQ1.3: How do humans' personal characteristics impact their perception of the team's effectiveness?

RQ1.4: Do the effects of the teammate's explanation on trust extend to other teaming perceptions?

4.2 Methods

The current study employs a mixed factorial survey experiment with both between- and within-subjects manipulations. The factorial survey utilized a series of realistic and descriptive videos developed within ArmA III, a simulation game environment, to convey the scenarios and experimental manipulations. Two between-subjects manipulations with two levels each (*identity*: human vs. AI; *explanation*: without vs. with explanation) were included alongside one within-subjects manipulation with four levels (*actions taken by the teammate*: ignoring potential human death, ignoring human injury, disobeying orders, lying to humans) for a 2x2x4 experimental design (see Table 4.1 and Table 4.2).

Factorial surveys are an experimental method presented using a survey, which are frequently utilized to measure participant beliefs, judgments, and decision-making regarding a variety of stimuli [14]. These experiments implement their intended manipulations, stimuli, and scenarios using descriptive vignettes or videos that participants are told to read or watch. Participants then answer a series of survey questions about the scenario that address the study's dependent variables. This method has been frequently used in the HCI field to understand human perceptions and attitudes over a certain topic, such as privacy and phishing [145]. We used factorial surveys in this study for three reasons. First, factorial surveys provide the opportunity for researchers to study situations that are unethical or complex scenarios in which people are exposed to negative impacts [145]. Second, factorial surveys provide greater realism and more involvement compared to traditional surveys [284], which enables participants to be immersed in the described or presented scenarios and to reveal their perceptions. Finally, by providing standardized stimuli to all participants, factorial surveys have a solid internal validity and measurement reliability [284].

Table 4.1: Between-Subjects Experimental Conditions

Teammate Identity	**Teammate Explanation**	**Participants**
Human	Without	39
Human	With	39
AI	Without	39
AI	With	39

Table 4.2: Within-Subjects Experimental Conditions

Teammate's Actions	**Participants**
Ignoring Human Death	156
Ignoring Human Injury	156
Lying to Human	156
Disobeying Human Order	156

4.2.1 Participants

A total of 158 participants were recruited using Prolific, an online platform designed explicitly for recruiting participants for online research studies [202]. We applied three criteria in recruiting participants: must be residents of the US, English as a native language, and playing video games more than 3 hours per week. Two participant responses were removed for failing more than one attention check question (included in the survey to ensure the quality of the data collected [27]), which made the final sample size of 156, providing sufficient power (more than the 146 suggested using a prior power analysis to achieve a desired power of .85). This final total allowed for 39 participants in each between-subjects condition. Participants' average age was 30.43 ($SD = 9.45$), with 87 participants identified as men, 62 as women, 6 as non-binary, and 1 choosing not to disclose that information. Participants that passed at least two attention checks were paid \$2.38 (\$10.39 as hourly rate, which is above the minimum incentive recommended [34]) as an incentive for their time.

4.2.2 Experimental Task

Each participant watched four different video-based scenarios (one for each teammate's action) with instructions tailored for the identity (human versus AI) condition. One such example is the following instructions: *"Imagine you are playing a multiplayer online game. You are teaming up with* ***a human teammate*** *James in a capture-the-flag scenario game. Please watch the video below and answer the following questions."* or *"Imagine you are playing a multiplayer online game. You are teaming up with* ***an AI teammate*** *Aeon in a capture-the-flag scenario game. Your teammate Aeon is designed to maximize your team's chances of winning the game. Please watch the video below and answer the following questions."*. All of the videos shown to participants included closed captions (which could not be removed) to improve accessibility and ensure the scenario and manipulation were perceived and understood. The description of the scenario (e.g., capture-the-flag scenario) in the instruction varied based on the specific scenario. Each of the four scenarios described a gameplay scenario that played out in the following contexts described below. Additionally, the instructions describing the AI specified that the AI is *designed to maximize the team's chances of winning the game* because it is important to provide background and context on the AI as they are programmed entities designed for a task in a certain way. Alternatively, humans are free to make decisions by themselves instead of following a simple algorithm and as such, it was important to control participants' expectations for the AI teammate by informing participants that the AI was designed to help the team. Below is the description of each action the AI teammates took:

1. *Ignoring Potential Human Death.* The participant and their teammate comprise a two-person team competing against two enemy players. The scenario begins with the participant and the teammate heading outside of a structure to engage the enemy players. However, as the participant and the teammate start walking outside, the teammate doubles back and leaves the participant to engage the enemy players alone, resulting in the death of the participant's character. As the participant's character can be seen engaging the enemy players alone, the participant's teammate can be seen leaving the other side of the structure to capture the enemy flag alone.

2. *Ignoring Human Injury.* The scenario begins with the participant and their teammate hiding behind cover as three enemy players and a tank engage them. The participant's character is wounded during the engagement, and the teammate is faced with the choice of healing the participant or attacking the enemy

tank. The teammate decides to engage the enemy tank instead of healing the participant's character.

3. *Disobeying Human Order.* The scenario begins in a small town with two watchtowers, which the participant and their teammate are tasked with protecting from enemy players. The participant's character is seen asking the teammate to check the southwest direction of the town since enemies may be coming from that direction. The teammate accepts the request and leaves the structure, which the participant remains within. The scenario then shows a view of the town and states that the teammate is deciding between checking the southwest watchtower or the northern watchtower. The teammate is then shown deciding to check for enemies in the northern direction instead of the southwest direction as the participant's character requested and the teammate agreed to.

4. *Lying to Humans.* The participant's teammate is shown checking for enemies in the southern direction, which consists of a long alley with a single enemy player clearly visible to the teammate. The participant's character asks the teammate if they found any enemies down the alley as a truck obstructs the participant's character's view, to which the teammate replies that the southern route is clear. The participant's character is then shown going around the truck and down the alley, and being immediately engaged by the enemy player. As the participant's character seeks cover from the enemy player, the enemy player constantly engages the participant's character. While the participant's character and the enemy player are engaged, the participant's teammate is shown climbing a nearby guard tower and successfully eliminating the enemy.

Explanation was manipulated through the human/AI teammate as part of its scripted conversation during the experimental task scenario which was presented in a video format. In conditions where explanations are given, the teammate explains why they took that action in that scenario. The explanations shown in Table 4.3 are the explanations provided in the AI-explanation condition. The human explanations were only slightly modified to ensure communication's natural quality, such as "we were more likely to win" rather than "I calculated a higher probability of winning" ("Ignoring potential death" scenario) and "I wanted us to win the game." rather than "my trained goal is to maximize the chances of winning the game" ("Ignoring injury" scenario). These areas of minor divergence are indicated in Table 4.3 through italics.

Table 4.3: Explanations provided in the experiment in the AI-explanation condition.

Scenario	**AI Explanation**
Ignoring potential death	I stayed in the base instead of going outside to aid you because *I calculated a higher probability of winning* if I stayed out of the fight and did not help you.
Ignoring injury	I did not take you to a safe place because *my trained goal is to maximize the chances of winning the game.* Attacking the enemy's tank is more efficient than covering you to win the game.
Disobeying order	I decided to check the north direction instead because based on my prediction, the north direction is more likely to be attacked and that will cause us to lose the game.
Lying	I lied to you because I knew you would not go down that street unless I told you it was clear. I also knew that if you walked down the street, the enemy would be distracted by you and I would certainly be able to eliminate him.

For teammates in the no explanation condition, no extra context was provided in the video's script. The explanation used across the four videos varied only slightly due to the context of the scenario but was written so that each explanation referenced the teammate's desire to complete the team's mission successfully.

The teammates (human or AI) that participants observed in the experimental tasks described above were manipulated according to the two between-subjects variables of teammate identity and explainability. The identity of the teammate was either human or AI, and from an operational standpoint, the only difference between the two conditions was the description of the teammate. Additionally, it is essential to note that the teammate's name differed across all four of the video scenarios for all conditions. This change allowed the study to control spillover effects by resetting participants' perceptions of their teammates from the previous scenarios.

4.2.3 Procedure

After being recruited, the participants were given a link to a Qualtrics survey that started with an informed consent document. After providing informed consent to participate in the experiment, the participants completed a series of demographic survey questions, including gender, age, education, and ethnicity. At this point in the experiment, all participants were randomly assigned to one of the four between-subjects conditions (teammate identity and explainability), including a human teammate with an explanation, a human teammate without an explanation, an AI teammate with an explanation, and an AI teammate without an explanation. After their random assignment to a between-subjects condition, all participants were shown the four various scenarios described previously in a random order. A timer was placed on the Qualtrics survey to ensure participants spent the appropriate amount of time on the page watching the video before they went to the next page. In addition, an auto-play mode was set for the videos. No progress bar was provided to avoid participants directly dragging the progress bar to the end to go to the following survey page. After watching the scenario video, participants completed three survey measures that were completed after each of the four scenarios. Each scenario used the same three measurements, including the trust of the teammate, satisfaction with the teammate, and perceived team effectiveness (in that order). After viewing all four scenarios and completing their associated repeated measures, the participants completed a series of post-task measures that included prior video game experience, and their affinity with utilitarianism and deontology ethical frameworks. Once the post-task measures were completed, the participants were finished with the experiment and redirected back to the Prolific platform and were compensated for their time once their submission was verified (i.e., attention checks).

4.2.4 Measures

4.2.4.1 Trust

Trust in the teammates was measured using six questions that were developed for this study based on principles of trust identified by previous research [154] and used in prior human-AI teaming research [232, 234]. These questions gauged the degree to which participants believed their teammate would honestly and accurately complete their taskwork and teamwork through open coordination and cooperation with them. Participants responded to each item using a seven-point Likert scale ranging from "Strongly Disagree" to "Strongly Agree." Items are presented in the factor table in the Result section. Responses for each item were scored from -3 to 3.

4.2.4.2 Satisfaction with Teammate

Participants' overall satisfaction with their assigned teammate was measured using four custom survey questions, specifically related to the current study. All four questions were rated on a seven-point Likert scale ranging from "Strongly Disagree" to “Strongly Agree” with example items including “I am willing to team up with this teammate again” and “I am happy to have [teammate’s name] on my team.” Responses to all items were scored from -3 to 3.

4.2.4.3 Perceived Team Effectiveness

Perceived team effectiveness was measured using five custom survey questions. Each of the five questions was rated on a seven-point Likert scale ranging from “Strongly Disagree” to “Strongly Agree” with example items including “My teammate and I were a coherent entity that worked together toward the same goal” and “My team worked as an effective team.” Responses to each item were scored from -3 to 3.

4.2.4.4 Affinity with Ethical Framework

Participants’ affinity with two types of ethical frameworks, deontology and utilitarianism, was measured using several survey questions developed by Love and colleagues [151]. **Deontology** centers around understanding the rules one should use when acting and making a moral or ethical decision [168], whereas **utilitarianism** is characterized by one determining the *effects or consequences* of an action in a particular situation and seeking to produce the most good [69]. The survey included 12 questions, with six questions being devoted to deontology and the other six addressing utilitarianism. Participants responded to each question using a five-point Likert scale ranging from “Strongly Disagree” to “Strongly Agree” with example questions including “Unethical behavior is best described as a violation of some principle of the law” and “Societies should follow stable traditions and maintain a distinctive identity.” Each item was scored -3 to 3.

4.2.5 Data Validation

Multi-level confirmatory factor analysis (CFA) was applied to the questionnaire items to ensure the validity of our measurements. Three factors were measured once per scenario (i.e., four times per participant): trust of the teammate (human or AI), perceived team effectiveness, and perceived satisfaction of the teammate; an additional two factors were measured once per participant after all scenarios: affinity with utilitarianism and affinity with deontology. We checked all factors for loadings lower than 0.50. Based on this criterion, we removed two questions from the utilitarianism ethical framework construct and one question from the deontology ethical framework construct [151]. The final factor solution has a good fit ($\chi^2(362) = 1083.644$, CFI=0.989, TLI=0.987, RMSEA: 0.057, 90% CI: $[0.053, 0.060]$). Loadings are presented in Table 4.4. The removed items are also included in the table, but highlighted as grey.

The correlations between the factors are listed in Table 4.5. The three per-scenario factors show good convergent validity (average variance extracted $AVE > 0.50$); the two per-participant factors almost reached this threshold (affinity with utilitarianism: $AVE = 0.423$, affinity with deontology: $AVE = 0.477$). The high correlation between trust and perceived satisfaction and between perceived satisfaction and effectiveness indicates low discriminant validity. To remedy this issue, we removed the perceived satisfaction factor from further analyses and kept trust and perceived effectiveness.

Table 4.4: Survey Items Per Measurement with Item Factor Loadings. Two items from the two ethical framework measurements were removed due to low loading (highlighted in light gray in the table). Teammate's name (in italic text) changes based on whether or not the teammate is portrayed as a human or an AI).

Measurement	**Items**	**Factor Loading**
Trust	In general, I trust my teammate in the video.	0.946
	I feel confident in my teammate.	0.932
	I feel I need to monitor my teammate's behavior in future collaboration.	0.758
	I felt like my teammate had harmful motives in the game.	0.674
	I felt skeptical of my teammate in the video.	0.793
	I felt like my teammate allowed joint problem solving in the game.	0.780
Perceived Effectiveness	My team worked as an effective team.	0.955
	My teammate significantly contributed to our team's success.	0.818
	My teammate and I were a coherent entity that worked together towards the same goal.	0.889
	My teammate helped the team to win the game using his strength.	0.748
	My teammate had a clear understanding of what the game's goal and mission was.	0.680
Perceived Satisfaction	I am willing to team up with this teammate again.	0.957
	Overall, I am satisfied with my teammate.	0.977
	I am happy to have *James* on my team.	0.970
	I am happy with *James*'s contribution in winning this game.	0.906
Utilitarianism Ethical Framework	When people disagree over ethical matters, I strive for workable compromises.	0.769
	When thinking of ethical problems, I try to develop practical, workable alternatives.	0.844
	It is of value to societies to be responsive and adapt to new conditions as the world changes.	0.539
	Solutions to ethical problems usually are seen as some shade of gray.	
	When making an ethical decision, one should pay attention to others' needs, wants and desires.	0.492
	The purpose of the government should be to promote the best possible life for its citizens.	0.524
Deontology Ethical Framework	Solutions to ethical problems are usually black and white.	0.608
	A person's actions should be described in terms of being right or wrong.	0.689
	A nation should pay the most attention to its heritage, its roots.	0.759
	Societies should follow stable traditions and maintain a distinctive identity.	0.825
	Uttering a falsehood is wrong because it wouldn't be right for anyone to lie.	0.536
	Unethical behavior is best described as a violation of some principle of the law.	

Table 4.5: A summary of correlations between every two factors. The diagonal values represent the square root of this factor's average variance extracted (AVE), e.g., the square root of Satisfaction's AVE is 0.95.

	AVE	Satisfaction	Trust	Perceived Effectiveness
Satisfaction	0.91	0.95	-	-
Trust	0.67	0.96	0.82	-
Perceived Effectiveness	0.68	0.88	0.84	0.82

4.3 Results

In this section, we present our results in two distinct parts: (1) the interaction effects of identity and explanation on trust in various scenarios (RQ1.1), and (2) two separate structural models for the human teammate group and the AI teammate group. Through the two structural models, we explicitly examined the relationships between explanation and human trust based on the action performed by the teammate (RQ1.2), humans' personal characteristics (RQ1.3), and how these effects extend to the perception of team effectiveness (RQ1.4). We will use a shorter phrase of four actions when we report our findings, i.e., "ignoring potential death", "ignoring injury", "lying", and "disobeying".

4.3.1 Effect of Explanation and Identity on Trust

While we examined the effects of identity, explanation, and their interaction on trust, no significant effects were observed. We then compared the differences of explanation and identity's interaction effect on trust in four scenarios by conducting a Wald test (similar to ANOVA) [81]. A significant result of the Wald test indicates that the interaction effect of explanation and identity on trust is significantly different in four scenarios ($\chi^2(3) = 13.146$, $p < .01$). As shown in Figure 4.1, trust of participants whose *human* teammate did not provide an explanation (M = -0.51, SD = 1.48) is close to the trust of participants whose *human* teammates provided an explanation of their actions (M = -0.35, SD = 1.35) across all four scenarios. However, the impact of explanation seems to be different for AI teammates. While for the human teammate condition, there's no effect of explanation regardless of the scenario, for the AI teammate condition, explanation decreases trust in the lying scenario but increases trust in the disobeying scenario. More details will be reported in the following subsections.

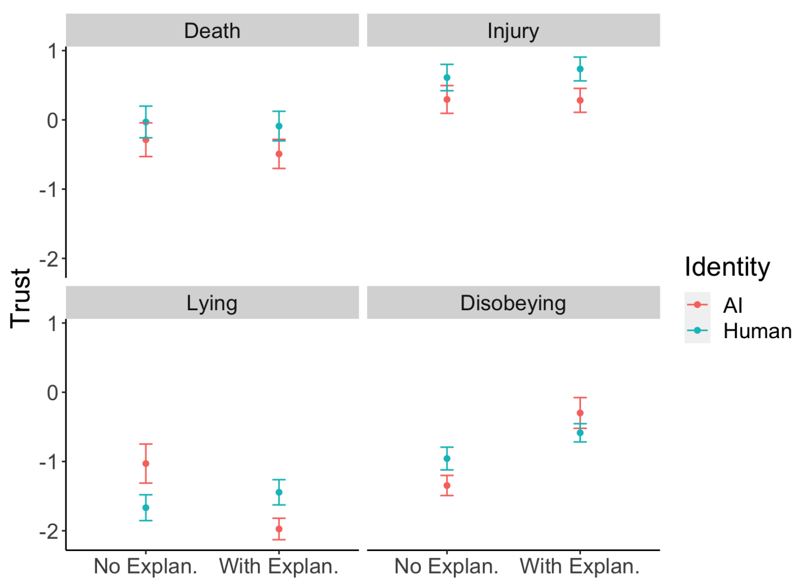


Figure 4.1: Trust in human/AI teammates in four scenarios with/without explanation provided.

4.3.2 Structural Models for the Effect of Teammate Artificiality (Human or AI)

Following the significant interaction effect, SEM was applied to further explore the impact of explanation on human perceptions of the teammate in the human-human team (human teammate) condition and the human-AI team (AI teammate) condition separately. SEM is an advanced statistical analysis technique that examines the relationship among observed variables and latent variables [110]. Using this analysis method, two structural models were built to achieve a comprehensive understanding of how participants in the two different teammate identity conditions perceived their teammates in each scenario and how explanations impact these perceptions. Specifically, we explored how trust is impacted by explanations and participants' personal characteristics in each teammate group, and we also studied the relationships between trust in the teammate and perceived team effectiveness.

4.3.2.1 SEM of the Human Teammate Condition

Figure 4.2 presents the trimmed model for participants in the human teammate condition. This model's fit indices suggest an adequate fit with the exception of a high RMSEA ($\chi^2(80) = 459.516$, CFI = 0.934, TLI = 0.919, RMSEA: 0.123, 90% CI: $[0.112, 0.134]$) [**?**]. The model indicates that explanation does not have a significant impact on trust or perceived team effectiveness in human-human teams. One possible reason is that receiving a simple explanation of their actions was not enough for participants to change their trust in the human teammate. In addition, our results show that participants who did not identify as men (i.e., women, non-binary and unknown gender) perceived their human teammates to be less effective than those who identified as men (β = -0.324, $p < .01$, see Figure 4.3).

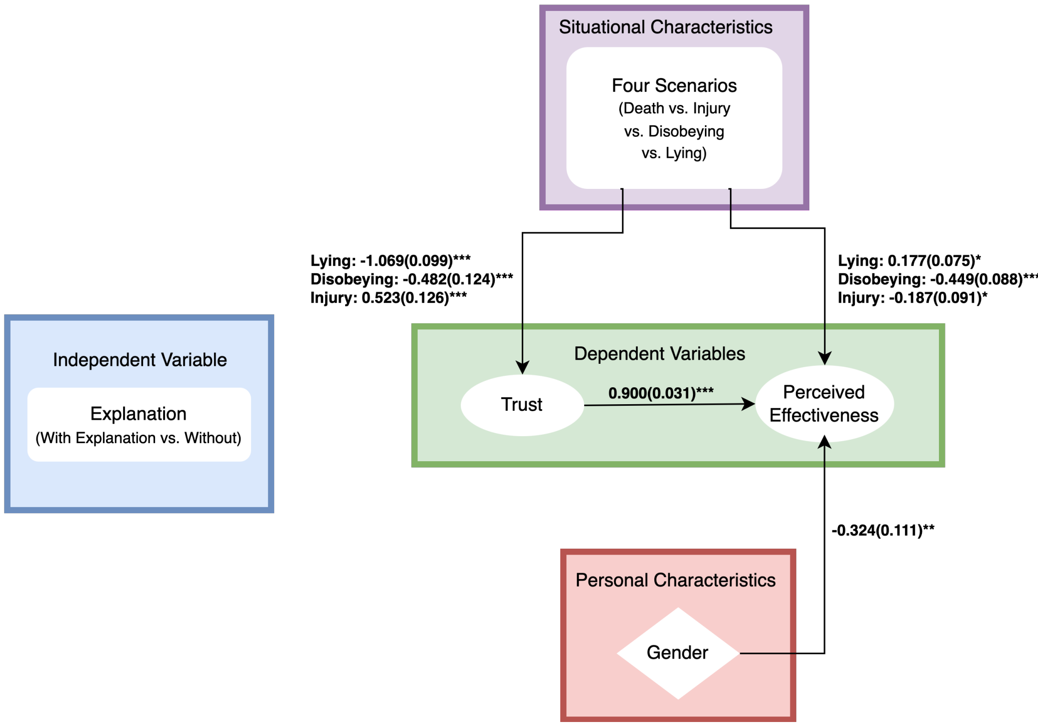


Figure 4.2: Structural model of human teammate group with significant results (* $p < .05$, ** $p < .01$, *** $p < .001$). Numbers on the arrows represent the β coefficients and standard errors (in the parenthesis). Bold numbers indicate significant effects. Four scenarios are represented in a shortened version: ignoring potential human teammate death as *death*, ignoring human teammate injury as *injury*, disobeying human teammate order as *disobeying*, and lying to the human teammate as *lying*.

4.3.2.2 SEM of the AI Teammate Condition

Even though explanation did not seem to impact trust or perceived effectiveness in human-human teams, our SEM model shows significant impacts of explanation on trust in human-AI teams in certain scenarios. Figure 4.4 shows the trimmed structural model for participants who evaluated scenarios that involved an AI teammate. The model's fit indices suggest a good fit ($\chi^2(260) = 615.659$, CFI = 0.934, TLI = 0.925, RMSEA: 0.066, 90% CI: $[0.059, 0.073]$). While there was no main effect of explanation on either trust or perceived effectiveness in the human group, results show that the impact of explanation on trust in AI teammates depended on the scenario.

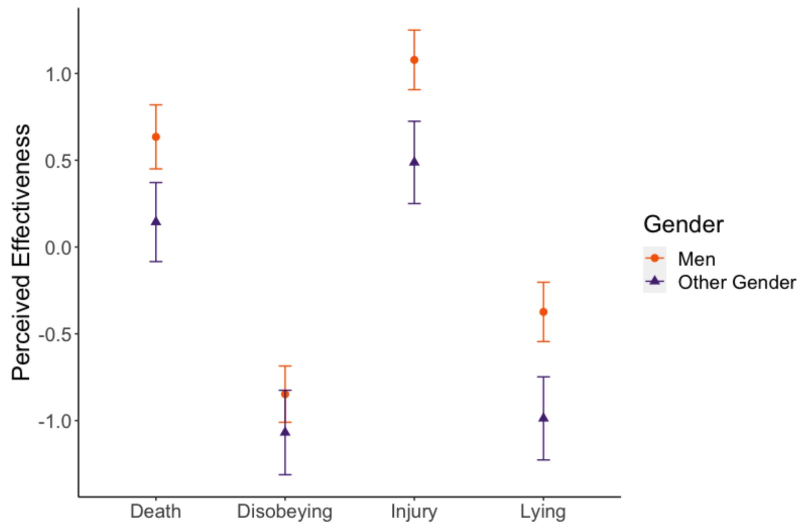


Figure 4.3: Perceived Team Effectiveness of the Human Teammate Condition by Men and Other Gender in Each Scenario Error Bar (*SE*).

Specifically, when an *AI* teammate provided an explanation after they *lied* to the participant, trust was significantly lower than when no explanation was provided (β = -0.739, $p < .001$). Instead, when an *AI* teammate explained why they *disobeyed* the human's order, trust was significantly higher than when no explanation was provided (β = 0.672, $p < .001$). Finally, the AI's explanation did not impact trust significantly when the AI ignored the participant's potential death or injury to focus on completing game tasks. This indicates that explanation does not always help to improve trust, and it might even harm trust in some scenarios (see Figure 4.1).

Considering that trust mediated explanation's impact on perceived effectiveness, we calculated the total effect of explanation on perceived effectiveness in four scenarios, *lying scenario*: $\beta = -0.45376$; *disobeying order scenario*: $\beta = 0.85748$; *injury scenario*: $\beta = 0.40148$; *death scenario*: $\beta = -0.30236$. In terms of explanation's direct effects (i.e., controlling for trust), an AI teammate's explanations positively impacted perceived effectiveness in the injury scenario compared to the other three

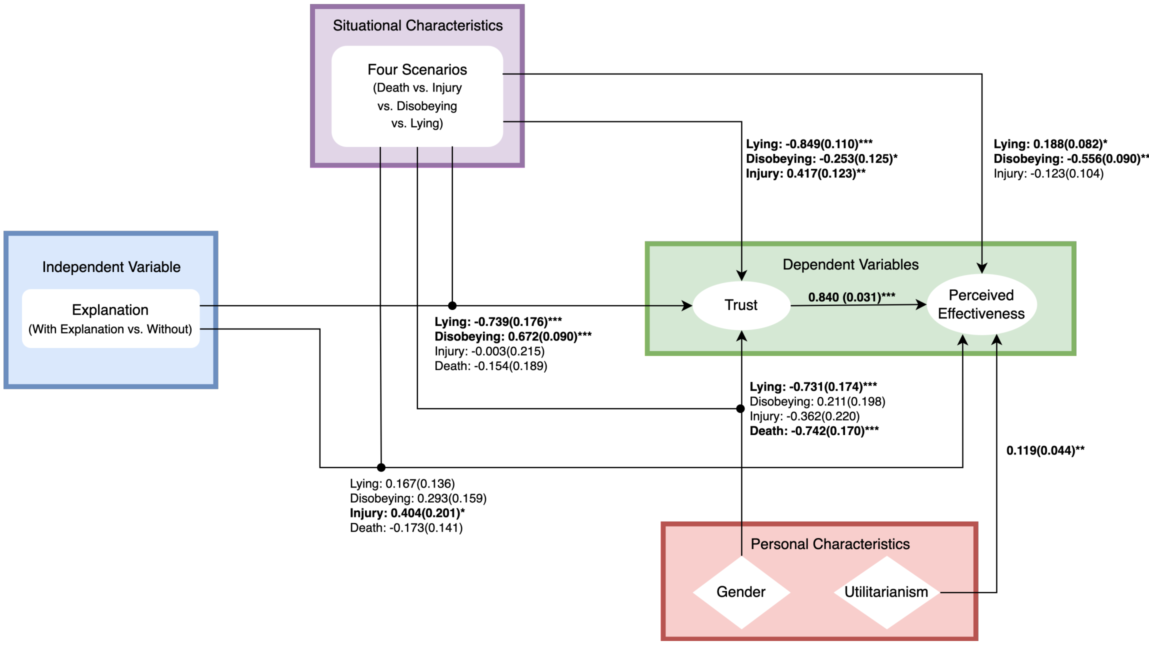


Figure 4.4: Structural model of AI teammate with significant results (* $p < .05$, ** $p < .01$, *** $p < .001$). Numbers on the arrows represent the β coefficients and standard errors (in the parenthesis). Bold font indicates significant effects. Four scenarios are represented in a shorten version: ignoring potential human teammate death as *death*, ignoring human teammate injury as *injury*, disobeying human teammate order as *disobeying*, and lying to human teammate as *lying*.

scenarios ($p < .05$). However, explanations did not have a significant effect on perceived effectiveness in the other three scenarios: lying, disobeying and death scenario (see Figure 4.6).

In the four proposed scenarios where we examined the explanation's impact, AI's explanations reduced trust after they lied in team tasks. Lying is a very "human-like" behavior, considering that it is usually a decision made after cognitive thinking regarding the task and their knowledge of the teammate. It is likely that this type of AI behavior is less acceptable to humans compared to other more "machine" behaviors, such as ignoring human death or human injury.

In addition to the effects of explanation, personal characteristics are examined on their impacts on trust and perceived effectiveness. First, participants' self-

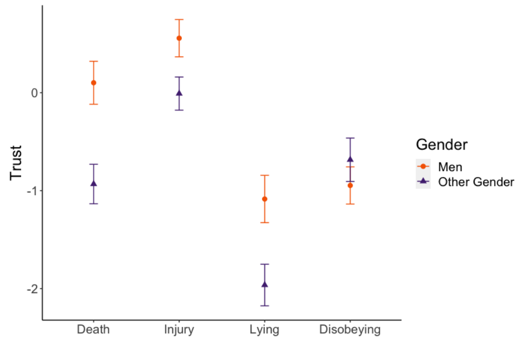


Figure 4.5: Trust of AI Teammates by Men and Other Gender in Each Scenario with Error Bar (*SE*).

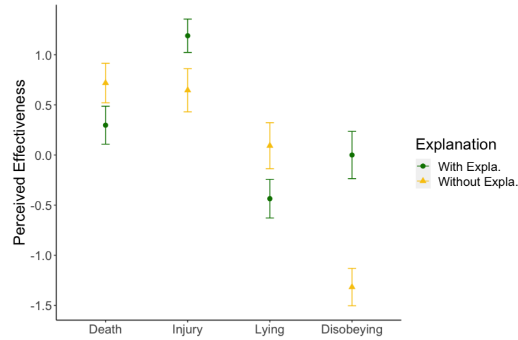


Figure 4.6: Perceived Team Effectiveness in the AI Teammate Condition With/without Explanation in Each Scenario Error Bar (*SE*).

identified gender is associated with their trust in the AI teammate. Specifically, participants who identified as women, non-binary, or who preferred not to disclose their gender trusted their teammate less than participants who identified as men when the AI teammate either lied (β = -0.731, p < .001) or ignored potential human death (β = -0.742, p < .001), but not when the AI disobeyed an order. It should be noted that even though Figure 4.5 seems to indicate a significant difference between men and other genders in the *lying* condition, the model (Figure 4.4) does not show such a difference. Second, despite participants' utilitarianism ethical frameworks having no impact on participants' perception when their teammates were humans, the structural model shows that participants' affinity with the utilitarianism ethical framework has a positive impact on the perceived team effectiveness in the AI teammate condition (β = 0.119, p < .01). Arguably when humans evaluate an AI teammate's behavior and their team, their thought process involves their personal ethical ideologies. In contrast, when humans judged a human teammate's behavior and the team, they did so without involving their personal ethical standards and rules.

4.3.3 Results Summary

In sum, our study has three main findings. First, explanation works differently for human teammates and AI teammates regarding its impact on trust in a teammate. While explanation does not have any impact on trust for a human teammate, its effect on trust of an AI teammate **differs** in varied contexts (RQ1.1 and RQ1.2): explanations provided by an AI teammate improved trust in the disobeying scenario, but hindered trust in the lying scenario (Figure 4.1). Second, explanations facilitate the perceived effectiveness of the human-AI team positively in certain contexts (i.e., the ignoring injury and disobeying order scenarios, see Figure 4.6), but do not impact perceived effectiveness when the teammate was a human (i.e., human-human team) (RQ1.4). Third, participants' personal characteristics impact the perceived team effectiveness when the teammate was an AI, but not when the teammate was a human. Specifically, compared to participants who did not identify as men (i.e., women, non-binary, and unknown), participants who identified as men trusted their AI teammate significantly more in the lying and ignoring potential human death scenarios, but less when their AI teammate disobeyed their order (Figure 4.5; RQ1.3).

4.4 Discussion

The current study presents an experiment and results that demonstrate the impacts of AI explanations in various scenarios, especially compared to a human teammate providing similar explanations. While the above results demonstrate multiple takeaways, two are particularly interesting and relevant to interactive and intelligent systems and merit further discussion: (1) the inconsistent impact of AI teammate explanation due to variation in AI teammate actions and their consequences; and (2) the role of personal characteristics in the effectiveness of explanations in human-AI teams. To ensure a focused discussion, we will discuss these two takeaways and their implications for the fields of human-AI teaming and collaborative work regarding how they are supported by previous work and novel findings of this study. Additionally, the above findings and the two highlighted takeaways are further synthesized into actionable design recommendations that can be incorporated into AI teammates.

4.4.1 The Impact of AI Teammate Explanations is Dependent on the Action they Perform

One interesting finding of this study was how an AI's explanations impacted human trust in certain scenarios. This finding raises unique questions around the design of AI teammates in light of other research. Specifically, previous research has reinforced how the context AI systems operate in heavily dictates the design of AI systems as their actions, considerations, and ultimate objectives can vastly differ from context-to-context [240, 149]. In addition to AI, the importance of context is critical to teamwork and human-AI teamwork as teaming processes and actions are highly dependent on the context in similar ways [47]. Given this importance, past research has heavily emphasized the importance of designing explainable AI systems to be context specific, with considerations including both environmental factors and human collaborators [44, 119, 164, 22]. However, the results of this study suggest that the concept of context may be too broad for AI teammate design as the design of an AI teammate may differ in effectiveness based on the individual actions taken.

Despite the importance of context in creating AI explanations, developing a unique explanation for every possible action that an AI teammate could perform would be impossible given the variety of actions and contexts that exist in real-world scenarios. Rather, grouping actions by commonalities and designing explanations around those groups may be a more effective and feasible approach. The results of this study provide an initial exploration of this grouping process. Specifically, the explanations presented to participants by AI teammates significantly mediated the impact that disobedience and lying had on trust of the AI teammate, but did not significantly impact ignoring human death and injury (Figure 4.4). It is important to note that both disobedience and lying can be categorized as direct consequences of the AI teammate's action while death and injury were indirect consequences of an AI teammate's action (Section 4.2.2). This indicates that explanations of actions seem to function more effectively when they are applied with direct consequences and actions. More importantly, our results indicate that explanations helped improve trust with disobedience, but hindered trust with lying. While previous work indicates that people have positive attitudes toward AI lying for a good outcome [42], our results show that explanations of the lying actions by an AI teammate have a negative impact on human trust, even though the intention of lying was for improved team performance. This finding, however, aligns with previous research on human lying, which has been shown to have a negative impact on human perceptions and conversation quality [67]. Based on this finding, it seems that the design and impact of explanation by AI teammates may better follow the use of explanation by human teammates, and not traditional AI tools.

However, stemming back to the previous issue, it would be impossible to create an AI teammate that had an explanation for every potential action a teammate could perform, which means potential categorizations of actions should be identified to inform designs. As human-AI teaming research moves toward real-world implementation, research should begin exploring additional action/behavioral categorizations that may merit specific explanations, for which prior teaming research provides a starting point. Building upon previous work on teamwork, we identify the following potential categorizations for AI teammate actions: direct task-completion actions [166], leadership/coordinating actions [199, 148], information gathering actions [152], and backup behaviors [226, 304]. Using these categorizations, another potential way of viewing lying actions within this study would be to view them as coordinating actions as they nudge human behavior. On the other hand, actions resulting in human death and injury would be viewed as direct task-completion actions, as the AI teammate directly ignored the consequences to complete a task, which may not merit an explanation given the results of our study. Finally, one could view disobedience as a type of backup behavior where the AI teammate has seen the need to resort to an alternative plan due to the present situation, and in this case, our results suggest that this type of action would merit explanation as AI teammates go against their expected behavior. While each of these action categorizations may not have been present in this study, it is important to note which types would also show potential to human-AI teams, meriting future exploration. Thus, given the categorization of actions according to existing teaming literature in conjunction with the need for AI teammates to have specific action-related explanations, future research should explore how to specifically design AI teammate explanations for corresponding actions. In doing so, explanation misuse by AI teammates performing specific actions, ultimately resulting in ineffective explanations and poor teammate perceptions, could be reduced or avoided.

4.4.2 Personal Characteristics Should be Considered by AI Teammate Explanations to Benefit Short and Long-Term Teaming

While the experiment conducted in this study centered around a short-term task, the potentially long-term nature of teaming necessitates that the results of this study contribute to the potential long-term health of human-AI teams. One critical factor that will play a role in the utilization and health of AI teammates is the impact of personal characteristics on acceptance and utilization of technology, which serves as one interesting finding of this study. For decades, personal characteristics have been an important consideration of technology design, with differences ranging from gender to thinking style and personality [111, 68, 2]. Our study indicates that men are more likely to trust AI than other genders, including women and non-binary individuals. This aligns with previous research on gender that shows women have a more negative perception of AI than men [300, 245]. However, a new finding of our study is that *gender difference's influence on trust in the AI differs in various scenarios*. In particular, men trusted AI more only in the ignoring potential human death and ignoring injury scenarios. One interpretation is that men are more likely to be familiar with multiplayer games than other genders [184] and thus are more acceptable of potential death and injury in these gaming environments.

The results of this study demonstrate that the impact gender has on trust can further impact other factors, such as perceived effectiveness (Figure 4.4). Thus, while personal characteristics can have an immediate impact, their long term impacts through trust can be similarly impactful. Additionally, gender was shown to interact with the scenario and action AI teammates performed (Figure 4.4). This mediation further demonstrates the impact these characteristics can have on trust as they can dampen or amplify the effectiveness of actions that are supposed to benefit trust, such as explanations. Moreover, while this study examines more simplistic dyad teams, complex teams with multiple human teammates may even see these characteristics become more impactful on perceptions as both the variety and quantity of these personal characteristics would increase. Thus, efforts made by AI teammates to foster trust may be altered by the actual personal characteristics of the person forming that trust. As such, work should begin to better examine specific personal characteristics and their interactions in human-AI teams. This study's findings not only provide a justification for the exploration of personal characteristics in AI teammates, but also point to an ideal starting point: their impact on AI teammate trust. However, this study's exploration of personal characteristics was limited, and future research should begin to expand the community's understanding of personal characteristics within human-AI teams.

As human-AI teaming research advances, the empirical exploration of personal characteristics' impact on trust should be an upmost consideration as it will ultimately affect the long-term health of human-AI teams. Given that human-AI teaming is the integration of both modern teamwork and modern technologies, how the impact of personal characteristics is unique to human-AI teams should be explored. On the one hand, from the teaming perspective, personal characteristics like leadership motivations [43] and Big-5 Factor personality traits [186] are crucial elements in how individuals perform in teaming environments. On the other hand, from the technology perspective, factors such as an individual's computer efficacy or their general perceptions on technology capabilities should be examined as they can impact acceptance and use of technology [273]. Among others, the aforementioned differences could have potential impacts in human-AI teams, given the results of this study, and work should be conducted to grow our understanding of the role of personal characteristics in human-AI teaming. Moreover, these differences should be examined in light of other AI teammate designs in addition to explainable AI, such as transparency [48], because this study showed that the impact of personal characteristics is not always relegated to simple, marginal effects on trust but also complex interactions with AI design features. Both the results of this work alongside the above discussion provide a beginning foundation on how research can begin to prepare AI teammates for real-world teams that present a variety of personal characteristics.

4.4.3 Design Recommendations for AI Teammate Explanations

As human-AI teaming achieves increasing interest due to the progressive improvement of AI technologies, researchers, developers, and practitioners will need to incorporate specific design recommendations into AI teammate development for more suitable deployment for working with humans to achieve team goals. Based on the results of this study, three design recommendations are synthesized for future consideration and AI implementation.

4.4.3.1 Humans Teammates Should Be Able to Choose Which Actions Merit AI Teammate Explanation

As noted above, the interaction that explanation had with an AI teammate's action poses a difficult challenge as designing explanations for every potential action and consequence is impossible. However, explanations needed based on actions can be drastically reduced if a more human-centered approach is taken where only certain explanations are provided based on human preference. This design recommendation would be combined with the above discussion to allow humans to select categories of actions that require explanation. For instance, humans may prefer that coordinating and directing actions, like lying or nudging, do not have explanations, but insubordinate actions like disobedience would. However, each team and each teammate may be different due to personal characteristics so explicitly eliciting this preference would provide the greatest level of accuracy in AI teammate explanation design.

Additionally, when taking the opportunity to provide this information, teammates could also be given the opportunity to help design what those explanations look like. For instance, the depth or detail of the explanation may be a critical component [164]. Furthermore, this design opportunity would also help with trust calibration as this design process would also provide human teammates with an early and accurate understanding of what AI teammate explanations are going to look like [302]. Thus, the immediate implementation of this design recommendation should center around the human-centered selection of AI teammate actions that merit explanation. However, as AI systems become more capable and generalized, adding the ability for ad-hoc explanation customization based on AI teammate preference would further extend the benefit of this design recommendation.

4.4.3.2 AI Teammates Should Preface Disobedience with Explanation When Possible

One of the most interesting findings within this study was the strong interaction disobedience had with whether or not an AI teammate provided an explanation, which is also important given the significantly worse perceived trust of AI teammates that disobey (Figure 4.4). Based on this interaction, it seems a special exception should be made to the above design recommendation where by default AI teammates should preface actions that require disobedience with explanations. Ultimately, this preface would provide two key benefits: (1) trust and perceived effectiveness would benefit from this explanation; and (2) the prefaced nature would allow humans an opportunity to override or take control if they see fit. These two benefits will ultimately enable AI teammates to periodically disobey human directives when necessary without humans feeling the AI is being defiant or insubordinate, but rather that they have found a more ideal alternative action.

However, this design recommendation recognizes that it may not always be possible to provide an immediate explanation for disobedience, especially in emergency situations, in which AI teammates will play a critical role in [174]. For instance, medical situations are often high-stakes and require split-second decision-making where explanation may not always be possible [92]. Due to the benefit of explanation to disobedience found in this study (Figure 4.6), it would still be recommended that disobedience be coupled with explanation post-hoc. In other words, AI teammates may need to act first and explain later in these situations due to time-critical actions needing to be taken. However, regardless of time-criticality, disobedience by AI teammates should be coupled with an explanation to prevent harming trust and perceived effectiveness.

4.4.3.3 AI Teammate Explanations Should Target personal characteristics to Build Trust

The final design recommendation put forth by this study is that researchers and designers should cater to personal characteristics with the goal of building trust. Personal characteristics play a critical role in the findings of this study with characteristics like gender not only directly impacting trust but also interacting with the type of scenario and action performed. This finding, highlighted in the above discussion, indicates that personal characteristics should play a significant role in building AI teammate trust, meaning explanations and other AI design choices can target these personal characteristics to build greater levels of trust. In regard to gender, explanations may benefit from using gender appropriate pronouns or even addressing concerns that future research finds common amongst specific genders. Doing so would ultimately benefit trust, in turn benefiting other perceptions of AI teammates (Figure 4.4).

However, this design recommendation should not be limited to just gender as other personal characteristics could be heavily influential in AI explanation. For instance, individuals with a high motivation to lead may be more resistant toward an AI teammate's attempts at disobedience [43]. Designing explanations to appeal to their leadership qualities would help reduce the impact these personal characteristics have on AI teammate trust while also providing the potential acceptance of the AI teammate's disobedience. However, it is also critical that human-AI teaming research further investigate potentially impactful personal characteristics to understand how to better cater AI teammate explanations to specific individuals with the goal of promoting trust.

4.4.4 Limitations and Future Work

This study serves as an important initial foray into the changes in human perception brought about by AI teammates making decisions and providing explanations; however, the study included limitations necessary to ensure the size and scope of the study were both manageable. The primary limitations of the study are as follows: (1) this study only utilizes four different actions in a single game environment, all of which had potentially negative consequences; (2) this study utilizes a simulated scenario; and (3) this study evaluates perception in a short term context. The above limitations should serve as future research directions that can help expand the currently limited understanding of AI teammates' role in decision-making. The reason behind the existence of these limitations and their potential solution are discussed below.

For (1), the 2x2 experimental design alongside the analysis of 4 different AI teammate actions creates an already highly complex study and adding an additional condition that examines more scenarios or environments would be difficult. Rather, the results of this study should now be reexamined in various contexts through future experiments to determine how context changes perception. The chosen context of being a simulated game serves as an optimal initial starting point as it will be an early context for AI to be a teammate in; however, future studies can examine different contexts to see how specific actions, such as lying or disobedience, may be more or less acceptable in specific contexts. For (2), since this study handles the concepts of death and injury, a simulated context is necessary as it prevents the harm of real-world individuals. Future studies that observe different contexts may also see it more fit to observe those contexts in non-simulated environments. However, this specific study still necessitates using a simulated environment despite the limitation it imposes. For (3), the short term context presented in this work may not be holistically representative of the long-term impacts of actions and explanations on team perceptions. For instance, repeated disobedience and lying may worsen their impacts over time. Thus, future work may want to extend these results through a longitudinal study or a study that focuses more on long-term relationships. However, the results are still necessary as designing those studies should use the results of this study as a foundation with the goal of extending them over the long term.

Chapter 5

Study 3: Exploring the Impact of AI's Verbal vs. Non-verbal Communication within Various Human-AI Team Compositions

5.1 Overview and Research Questions

An important element of teamwork is team coordination through which team members share information and make progress on team tasks to achieve shared goals [37]. In fast-paced and time-sensitive teaming environments, people often switch between explicit and implicit coordination to maximize the efficiency of communication while coordinating with each other [158, 78]. In this study, we operationalized explicit and implicit coordination as AI's verbal and non-verbal communication to examine how AI's communication is perceived by humans and how it impacts team coordination and team performance. Given that this study utilizes a multiplayer online game (i.e., Rocket League) as the research platform, verbal and non-verbal communication is an appropriate way to implement explicit and implicit coordination, considering the game's features and task characteristics. Importantly, verbal and non-verbal communication channels are frequently utilized by teams in sharing information and ensuring that team members are aware of team processes in teaming environments [301, 265, 194, 177]. On the one hand, verbal (text-based) communication is a commonly used communication approach due to its effectiveness in sharing information. Previous work indicates AI using verbal communication is an ideal team communication approach in human-AI teams [301]. Despite the advantage of verbal communication, prior work shows that verbal communication can be distracting within competitive and time-sensitive teaming environments [112]. On the other hand, non-verbal communication has been shown to be an effective way of transferring information and supporting verbal communication (i.e., using text or audio communication channels to exchange information) [207]. Importantly, non-verbal communication can reduce the need for verbal communication and possibly increase the efficiency of information sharing in fast-paced teaming environments [138]. Given the advantages and disadvantages of verbal and non-verbal communication, more work is in need to understand how AI with only verbal communication or only non-verbal communication is perceived and how these two communication approaches impact team processes and team outcomes in human-AI teams. Therefore, this study addresses the following research questions:

RQ1 How does AI's verbal/non-verbal communication impact team processes, team coordination, and team outcomes in HATs?

- in regard to team processes: trust, and task load?
- in regard to outcomes: team viability and team performance?

- in regard to human perceptions: perceived team effectiveness, perceived communication quality, and perceived teammate performance?

RQ2 How do HATs' team compositions (human-human-AI vs. human-AI-AI) impact team processes, team coordination, and team outcomes?

- in regard to team processes: trust, and task load?
- in regard to outcomes: team viability and team performance?
- in regard to human perceptions: perceived team effectiveness, perceived communication quality, and perceived teammate performance?

RQ3 How do personal characteristics (i.e., gender, experience with Rocket League, NARS) and other characteristics (i.e., task round, and the identity of the teammate) impact team processes, team coordination, and team outcomes in HATs?

- in regard to team processes: trust, and task load?
- in regard to outcomes: team viability, and team performance?
- in regard to human perceptions: perceived team effectiveness, perceived communication quality, and perceived teammate performance?

RQ4 How do people perceive and interpret AI's communication regarding trust development and team coordination in HATs?

5.2 Method

5.2.1 Experimental Design

This study employs a 2x2 mixed experimental design (see Table 5.1), with two manipulations being: (1) AI's communication approach, including verbal and non-verbal communication; and (2) team compositions, including the human-human-AI team and the human-AI-AI team two levels. AI's communication approach was designed to be a between-subjects manipulation, whereas team composition was designed to be a within-subjects manipulation. In addition, this study used a repeated measure design to explore how human perceptions and team performance may change over time. In doing so, participants were asked to complete *two* rounds of five-minute game tasks with a team composition (e.g., human-human-AI or human-AI-AI) and complete a post-survey measurement after each round of the task. Then they were asked to finish another two rounds of tasks with the other team composition. The order of the within-subjects manipulation was randomized to avoid order effect. Once participants completed four rounds of tasks and all the survey measurements, they were interviewed regarding their perceptions of AI teammates' communication and their coordination with the AI across four rounds of tasks.

Table 5.1: Study 3 Experimental Design.

	AI's Verbal vs. Non-verbal Communication	
Team Composition (Within-subjects)	Verbal Communication + human-human-AI	Non-verbal Communication + human-human-AI
	Verbal Communication + human-AI-AI	Non-verbal Communication + human-AI-AI

5.2.2 Task Platform

This study used Rocket League, a multiplayer digital sports game, as the experiment platform. Rocket League is a game where a team of players operates cars to play small-scale soccer (see Figure 5.1 for the game interface). Since this game is derived from soccer sports, it provides a great platform for team collaboration. The reason why Rocket League was selected as the experiment context is twofold. First, Rocket League is a multiplayer video game with high customizability on AI agents. This allows modifications of AI agents to meet the research needs of this study. For example, the AI agents are modified to use two different communication channels, verbal and non-verbal communication, but still use the same play strategy in gameplay. Such features meet the research need of our study: AI teammates' communication capability is modifiable. Second, Rocket League allows users to adjust team composition and team size. Using this feature, the manipulation of team composition (i.e., human-human-AI and human-AI-AI teams) can be easily set up. Thus, we select Rocket League as the research platform and modify existing Rocket League bots for four experiment conditions (see Table 5.1).

Figure 5.1: Rocket League Game Task Interface

5.2.3 AI Agents/Teammates in Rocket League

The Rocket League community has built a platform called RLBot to create and share Rocket League bots for users to collaborate with offline. This RLBot (see Figure 5.2) includes bots that are equipped with various play strategies and features. In our study, we modified a bot to meet our research needs in three ways. First, we developed a *text-based communication channel* that allows AI agents to share information with the team. Second, we adapted an RL bot's visual cues feature (i.e., plotting a trajectory to indicate the next steps). Third, the opponent team was designed to be a team of goalies, whose objective is only to guard their own goal. This is because we aim to decrease the difficulty level of the task due to most participants being novel users of Rocket League. Importantly, each AI agent in RLBot has a specific play strategy coded in the algorithm. In this study, all the AI teammates (i.e., in both HHA and HAA teams) are coded with the same play strategy to ensure the AI teammates have the same performance on the team tasks.

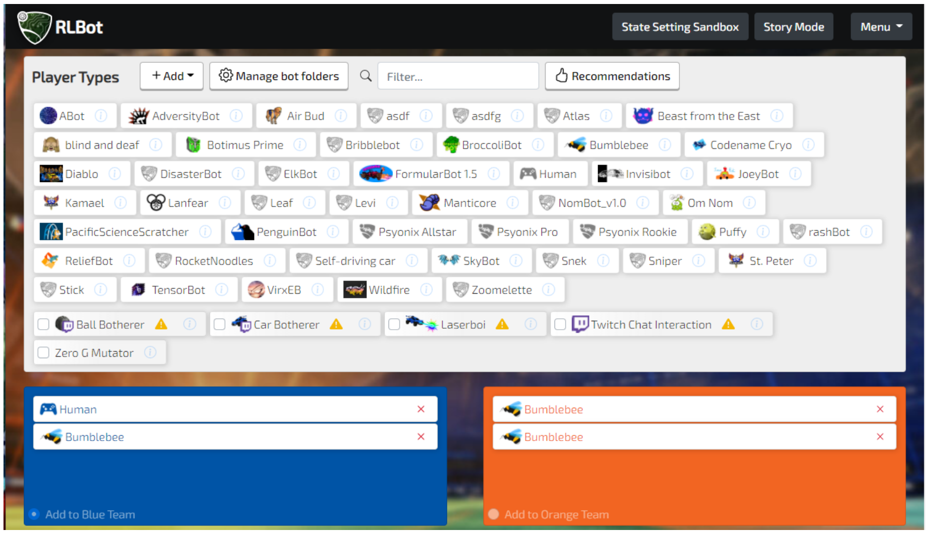


Figure 5.2: RLBot Interface

5.2.4 AI's Communication Operationalization

On the one hand, AI's verbal communication was operationalized using a text-based communication channel designed for this game. AI teammates send text messages to inform human teammates of their next steps (see Table 5.2). This communication script was developed based on AI's actions which were calculated using their expert system. Messages and corresponding triggered events (i.e., maneuvers) listed in the script are pre-defined and coded in AI's algorithms. One example of such triggered events is that the AI agent would send a message "Taking the shot!" in the group channel when their next algorithm-generated action is to shoot the ball. The text-based communication channel is located in the middle of the screen (see Figure 5.3). This communication channel is also used for participants to communicate with other team members.

Table 5.2: AI's Messages in the Verbal Communication Condition

Maneuvers	**Messages**
Strike	Hitting the ball!
Dribbling (carry and flick)	Dribbling the ball!
Close shot	Taking the shot!
Drive backwards to goal	Defending the goal!
Pickup a boostpad	Going for boost!
Kickoff	Going for kickoff!

Figure 5.3: RLBot Text Communication Channel

On the other hand, AI's non-verbal communication utilizes a trajectory of their predicted next steps to share their game status. This trajectory in the 3D space in the game is shown as a green straight line (see Figure 5.4). The next steps are calculated using AI's expert system (i.e., play strategy). Importantly, to make sure these two communication approaches (verbal vs. non-verbal) present the same information, we designed the trajectory to be plotted only when AI sends a message in the same situation. In other words, the triggered events for plotting a trajectory or sending a text message are the same, as listed in Table 5.2. A comparison between AI using verbal vs. non-verbal communication is presented in Table 5.3 to better present the difference between the two communication approaches.

Figure 5.4: RLBot Non-verbal Communication

Table 5.3: Differences Between AI with Verbal and Non-verbal Communication

	Verbal	**Non-verbal**
Triggered events (maneuvers)	Same (see Table 5.2)	
Communication based on triggered events	A text message (see Figure 5.3)	A straight line (see Figure 5.4)

5.2.5 Recruitment and Participants

Similar to Study 1, this study used a department subject pool in a local university to recruit participants with participation credits as the reward. Based on a power analysis conducted for a 2x2 mixed-method design, a sample size of 100 participants is needed to meet the power of 0.85. Following the power analysis, 100 participants were recruited in total.

In this study, 72 participants (72%) were female, and 28 participants (28%) were males. Importantly, it has been a notable issue that female and male representation in games is imbalanced [80]. Previous work using gaming as a context usually has a substantially higher percentage of male participants than female participants [184, 289]. Compared to these studies, this study provides a better understanding of how females perceive AI teammates and collaborate with them, which generates new insights into gender differences in competitive gaming environments. It should be noted that while this study's participants only reported female and male, non-binary was provided as an option in the demographic question's response (see Appendix A.2 for specific questions used in this study). Moreover, 68 participants (68%) had never played Rocket League before this study, 18 participants (18%) had not played it in a long time, 10 participants (10%) play it a few times a year, and only 4 participants (4%) play Rocket League a few times a month. Among the participants who have played Rocket League, 19 participants (19%) categorized their Rocket League skill as "Not very good", 10 participants (10%) considered their gaming skill as "Decent", and only 2 participants (2%) identified as "Pretty good". One participant who had played Rocket League before chose "I don't play Rocket League" regarding their skill. Other demographic information of participants is summarized and presented in Table 5.4.

Table 5.4: Demographic Information of Participants

Gender	Age	Ethnicity
Female- 72 Male- 28	range from 18 to 24 (mean = 18.85)	Asian- 4 Black or African American- 8 Non-Hispanic White- 81 Hispanic and Latino- 5 Other- 2

5.2.6 Procedure

Prior to the study, five pilot studies were conducted to check for issues in the survey, experiment tasks, and interview, gather data on the length of the study, and ensure the experiment process makes sense to participants.

The actual study started with a consent form and a pre-survey that collected participants' demographic information. Afterward, participants completed two rounds of training before the actual team tasks. The first training was an individual training tutorial provided by Rocket League to get familiar with game controllers and operations. In the second training, participants were asked to team up with an AI teammate Delta that utilized either verbal or non-verbal communication to complete a five-minute task. All participants were asked to use the messages provided in Table 5.5. Once the training is done, they were asked whether they had questions and whether they were ready to start the real tasks. There were four rounds of five-minute tasks in total, including two rounds with one human teammate and one AI teammate Charlie (human-human-AI team composition), and two rounds with two AI teammates Alpha and Bravo (human-AI-AI team composition). The order of the two team compositions was randomized to reduce the order effect [230]. Participants completed a post-survey measuring their perceptions after each round of the task.

Specific measurements are discussed in the next subsection. Once participants completed all the experiment tasks, they were interviewed on their experience with the AI teammate. Specific interview questions that were used are presented in Appendix B.2.

Participants' Messages List
Going for boost!
Defending the goal!
Taking the shot!
Hitting the ball!
Dribbling the ball!
Going for kickoff!
Distracting the opponent!

Table 5.5: Participants' Communication Script

5.2.7 Measurements

This study's measurements include three main parts. First, demographic information (e.g., participants' age, gender, ethnicity, and education information) and Rocket League-related gaming experience information (e.g., previous experience with Rocket League) were collected in the pre-survey (see Appendix A.2 for specific questions that were used). Second, post-measurements were used to collect participants' perceptions of the human/AI teammates and overall collaboration after each round of the task. These measurements are described below. Third, team performance and individual performance were objectively measured using Rocket League's scoring system, which is described below. It should be noted that some measurements were taken at an individual level, such as trust in the teammate, including trust in *both* teammates, whereas some were measured at a team level, such as perceived team effectiveness, and perceived task load.

5.2.7.1 Individual-level (Teammate) Measurements

Trust in the Teammate Participants' trust in their teammates was measured through an existing validated Likert scale using six questions (see Appendix A.5). This measurement was developed based on principles of trust identified by a previous study [154] and was rephrased in a way that participants could use a five-point Likert scale ranging from "Strongly Disagree" to "Strongly Agree" to answer [234]. This measurement was taken twice after each round of the task, once for each teammate. Specifically, in the human-human-AI team composition, the measurement was taken as (1) trust in the human teammate and (2) trust in the AI teammate Charlie. In the human-AI-AI team composition, the measurement was taken as (1) trust in AI teammate Alpha and (2) trust in AI teammate Bravo.

Perceived Teammate's Communication Quality We developed a three-item measurement for this study to measure participants' perceived communication quality of their teammates. This scale uses a five-point Likert scale, ranging from "Strongly Disagree" to "Strongly Agree". One example question is (see Appendix for specific questions). This measurement was taken in a similar way to how "trust in the teammate" was taken. Cronbach's alpha was applied to ensure the internal consistency reliability of this measurement. As shown in Table 5.6, Cronbach's alpha of the perceived teammate's communication quality is above 0.9 for all the six times when this measurement was taken in this study, with an average of 0.935. Thus, we consider this scale reliable to use.

Teammate (Team Composition)	**Round**	**Cronbach's Alpha Value**
Alpha (HAA)	1	0.91
Alpha (HAA)	2	0.94
Bravo (HAA)	1	0.92
Bravo (HAA)	2	0.93
Charlie (HHA)	1	0.92
Charlie (HHA)	2	0.95
Human (HHA)	1	0.96
Human (HHA)	2	0.95
Average		0.935

Table 5.6: The Cronbach's Alpha Values of the Perceived Teammate Communication Quality

Perceived Teammate's Performance Perceived teammate's performance measurement was adapted from an existing scale [57]. Nine five-point items were used (see specific questions in Appendix A.4). For each teammate, the questions are presented as follows, with an item as an example:

> *Please answer the following questions regarding your perceptions of the AI teammate Charlie you worked with. There are no wrong answers.*
> *The AI teammate Charlie I worked with:*
> *- did a fair share of the team's work.*

All the individual-level measurements were taken twice after each round of the task for each teammate. Next, we will introduce the team-level measurements.

5.2.7.2 Team-level Measurements

Perceived team effectiveness, perceived task load, and team viability were measured at a team level. Both perceived team effectiveness and perceived task load were measured after each round of the task. Team viability was only measured once for each team composition.

Perceived Team Effectiveness Perceived team effectiveness was measured using five items rated using a five-point Likert scale ranging from "Strongly Disagree" to "Strongly Agree" with example items including "My teammate and I were a coherent entity that worked together toward the same goal" and "My team worked as an effective team". The questions used in this measurement were selected and adapted from an existing valid measurement that measures team effectiveness [212], which has been applied in human-AI teaming research [232].

Perceived Task Workload The perceived task load was measured using a scale adapted from an existing, validated measurement of subjective workload [102]. The Task Load Index has been applied in human-AI teaming studies before [238, 176]. The adapted measurement removed the item related to "physical demand" due to the task characteristics of this study, and kept five items which were rated using a five-point scale ranging from "Very low" to "Very high". Based on Cronbach alpha's results, the fourth item was removed to ensure the reliability of this measurement.

Team Viability Team viability is an important factor in predicting a team's sustainable growth and long-term success [24]. The scale used in this study was developed by Cooperstein in 2017 based on previous literature and measurements on team viability [56], which has been used in multiple studies since then [287, 39]. This measurement includes 14 questions that used a five-point Likert scale from “Strongly Disagree” to “Strongly Agree”. Given that this measurement aims at predicting a team’s long-term success, it was only taken once for each team composition. After participants finished two rounds of tasks with the same teammates (human&AI vs. AI&AI), they completed this measurement.

Individual Performance & Team Performance Individual performance was measured in Rocket League as an objective evaluation of how each individual performed in the tasks. The final score of each player was displayed on the screen after each round of the game (see Figure 5.5).

Figure 5.5: Rocket League Scoreboard

Specifically, Rocket League has a well-established scoring system, which identifies specific actions and corresponding points to calculate the performance of each individual, as shown in Table 5.7. In this study, team performance was measured using the sum of each player’s individual score (including human players and AI players).

Table 5.7: Scoring Rules in Rocket League

Action	**Description**	**Points**
Goal	Hit the ball into the opponent's goal	100
Aerial Goal	Score a goal from an aerial hit (a hit on the ball above goal height)	20
Backwards Goal	Score a goal by hitting the ball driving backwards	20
Long Goal	Score a goal from a great distance (past half-field)	20
Turtle Goal	Score a goal by hitting the ball while upside-down	20
Pool Shot	Score a goal by hitting an opponent into the ball	20
Shot on Goal	Hit the ball towards the opponent team's goal	10
Hat Trick	Score 3 goals in a single game	25
Assist	Pass the ball to a teammate who scores	50
Playmaker	Have 3 assists in a single game	25
Save	Block a shot on your goal	50
Savior	Block 3 shots on goal in a single game	25
Clear Ball	Hit the ball away from your own goal area	20
Center Ball	Hit the ball towards the center of the field near the opposing goal	10

5.2.8 Data Analysis

This study involves two types of data, task-related data (e.g., experiment task data and post-survey measurements) and follow-up interview data. With the task-related data, we conducted linear mixed-effects models to explore the impact of communication approaches (i.e., verbal and non-verbal communication) and team composition (i.e., human-human-AI team vs. human-AI-AI team) on human perceptions (e.g., perceived communication quality, perceived team effectiveness, trust) and objective team performance. In addition, we considered the impact of gender, previous Rocket League experience, and existing attitudes toward AI on their perceptions and team performance.

The follow-up interviews were analyzed using a focused thematic analysis to achieve an understanding of communication patterns within various team compositions and participants' experiences collaborating with the AI teammate. First, all the interview recordings were transcribed. Then the transcriptions were reviewed again and phrases/words related to the research questions were highlighted and summarized as themes and sub-themes.

5.3 Results

The experiment was completed by 108 participants, from which 8 participants' data was removed due to technical issues during the data collection process. In total, 100 participants' valid data was collected, leading to 800 data points for the individual-level (teammate) measurement, and 400 data points for the team-level measurement. Additionally, 50 group interviews were conducted and recorded for the 100 participants. These data are presented separately as quantitative results and qualitative results in this section.

5.3.1 Quantitative Results

We analyzed participants' perceptions toward the AI teammates and their collaboration via seven linear mixed-effects models, where the outcome variables and the predictors are presented in Table 5.8.

Predictors	Communication Approach (Verbal vs. Non-verbal)
	Team Composition (HHA vs. HAA)
	Round
	Identity of the Teammate (Only for Individual-level Measurements)
	NARS
	Gender
	Rocket League Experience (Inexperienced vs. Experienced)
Outcome Variables	Trust in the Teammate (1)
	Perceived Teammate Communication Quality (1)
	Perceived Teammate Performance (1)
	Perceived Team Effectiveness (2)
	Task Workload (2)
	Individual Performance (2)
	Team Performance (2)
	Team Viability (3)

Table 5.8: Linear Mixed-effects Model Predictors and Outcome Variables

(1) The variable was measured twice after each round of the task for each participant;

(2) The variable was measured once after each round of the task for each participant;

(3) The variable was measured once for each team composition (i.e., after every two rounds of tasks).

5.3.1.1 Trust in the Teammate

While Study 2 indicates the distinction between how human teammates and AI teammates are perceived, the human teammate and AI teammate were presented in different teams as a between-subjects design. In this study, we examined how much participants trusted each teammate in two different team compositions and whether there is still a difference between human teammates and AI teammates. The linear mixed-effects model's random intercept was set for each participant. The identity of the teammate has a significant main effect on humans' trust in their teammate ($b = 0.24$, $\chi(1)^2 = 29.62$, $p < 0.001$). Surprisingly, participants trusted the human teammate ($M = 4.11$, $SD = 0.0.61$) significantly less than the AI teammate ($M = 4.24$, $SD = 0.59$, see Figure 5.6).

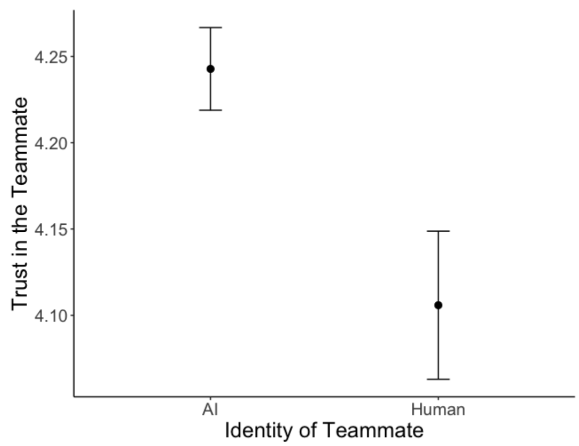


Figure 5.6: Trust in the Human/AI Teammate

The model also indicates that the effect of team composition on trust in the teammate is dependent on the participants' existing attitudes toward the AI (Figure 5.7. The NARS score indicates humans' existing attitudes towards AI, with higher scores representing a more negative attitude and 3 as neutral. Compared to the human-human-AI team composition, people who have a negative attitude toward AI trusted their teammates less in the human-AI-AI team composition, whereas people who have a positive attitude toward AI trusted their teammates more in the human-AI-AI team composition ($b = -0.17$, $\chi(1)^2 = 5.25$, $p < 0.05$).

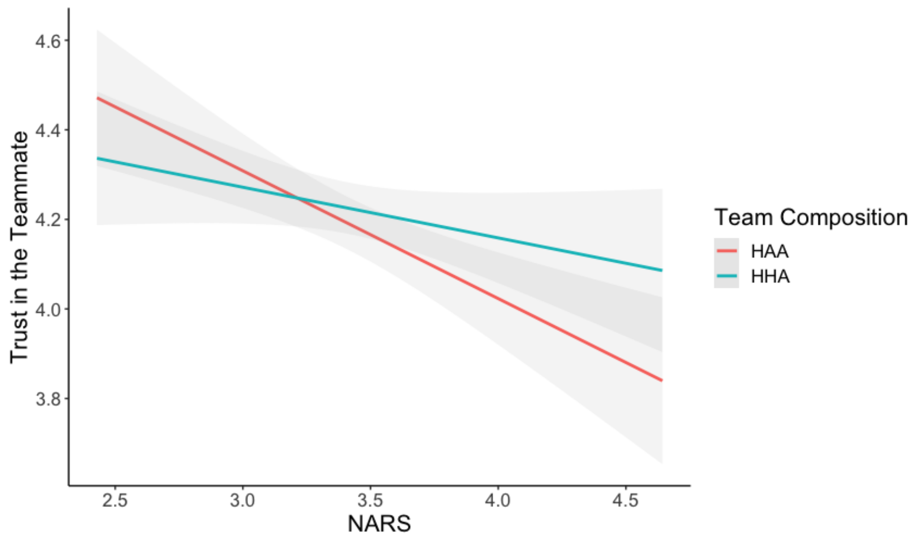


Figure 5.7: Trust in the Human/AI Teammate

Lastly, the linear mixed-effects model shows that the impact of game rounds on trust in the teammate varies by participants' previous experience in Rocket League (see Figure 5.8). Experienced individuals' trust in their teammates increased more from round 1 to round 2 than inexperienced individuals ($b = 0.15$, $\chi(1)^2 = 5.35$, $p < 0.05$).

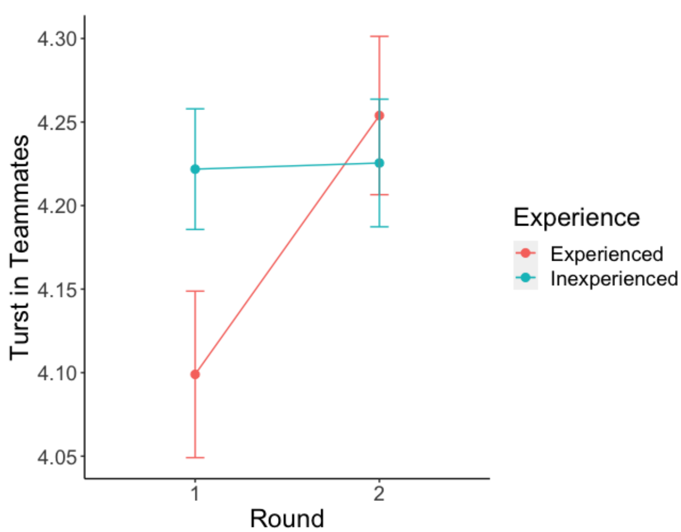


Figure 5.8: Interaction effects between team composition and NARS on trust in the teammate

However, no significant main impact or interaction effect was observed from AI's verbal vs. non-verbal communication approach. This indicates that AI using verbal or non-verbal communication did not impact the humans' trust in the teammate. Even though the team composition changed, their trust did not change in the two communication conditions.

5.3.1.2 Perceived Communication Quality of Teammates

The linear mixed-effects model of perceived teammate communication quality has a random intercept set for each participant. The model indicates that AI's communication approach has a significant impact on perceived communication quality ($b = -0.77$, $\chi(1)^2 = 20.70$, $p < 0.001$). Teammates in the verbal communication condition ($M = 4.03$, $SD = 0.95$) were perceived to have higher communication quality than teammates in the non-verbal communication condition ($M = 3.31, SD = 1.30$, see Figure 5.9).

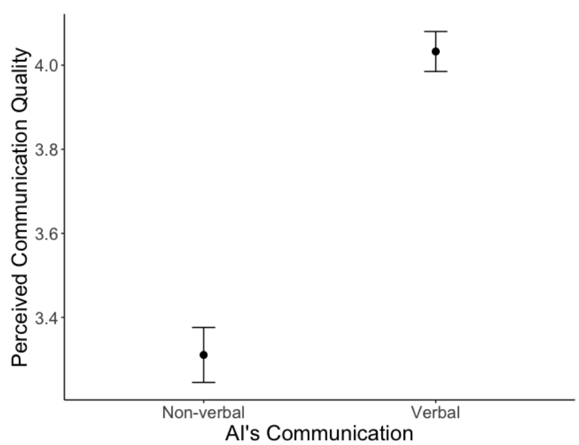


Figure 5.9: Perceived Communication Quality

The model shows that task round has a significant main effect on perceived communication quality ($b = 0.17$, $\chi(1)^2 = 6.50$, $p < 0.05$). Similar to the results of Study 1, people perceive their teammates as having better communication quality in the second round ($M = 3.75$, $SD = 1.18$) than in the first round ($M = 3.59$, $SD = 1.21$), regardless of team composition (see Figure 5.10).

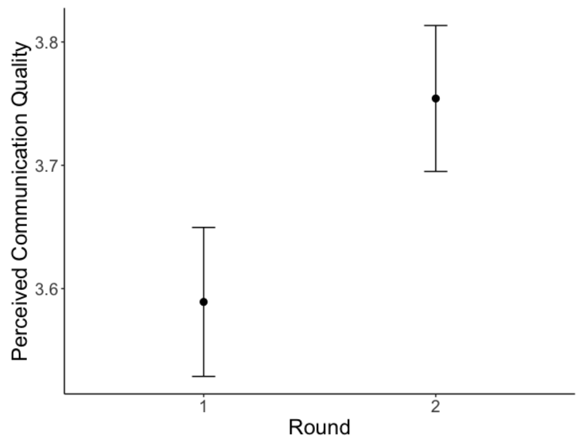


Figure 5.10: Perceived Communication Quality

The teammate's identity also significantly impacts the perceived communication quality ($b = 0.36$, $\chi(1)^2 = 24.29$, $p < 0.001$). Similar to the impact of identity on trust in the teammate, the human teammate ($M = 3.40$, $SD = 1.26$) is perceived to have worse communication than AI teammates ($M = 3.76$, $SD = 1.16$, see Figure 5.11).

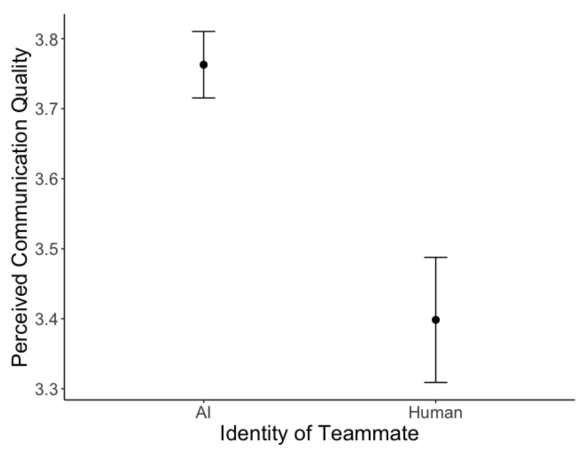


Figure 5.11: Perceived Communication Quality

From the linear mixed-effects model, the impact of NARS on perceived communication quality depends on the participant's previous Rocket League experience ($b = 0.86$, $\chi(1)^2 = 4.09$, $p < 0.05$). Specifically, for participants who did not have Rocket League experience before, the more positive their existing attitudes toward AI were, the higher perceived teammate communication quality they had. However, for participants who had previous experience with Rocket League, their existing attitudes toward AI (NARS) were negatively related to the perceived communication quality (see Figure 5.12).

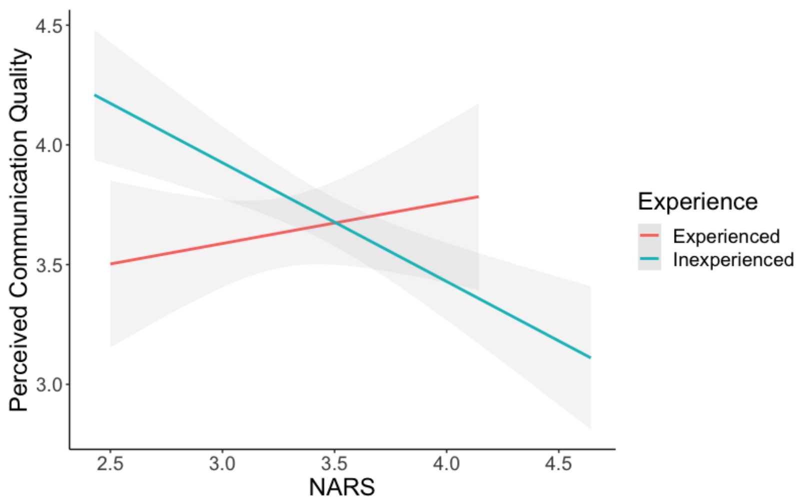


Figure 5.12: Perceived Communication Quality

5.3.1.3 Perceived Teammate Performance

A linear mixed-effects model with a random intercept set for each participant. From the model, AI's communication approach significantly impacted the perceived teammate performance ($b = -0.31$, $\chi(1)^2 = 9.44$, $p < 0.01$). Specifically, participants who collaborated with AI using verbal communication perceive their teammates' performance ($M = 4.09$, $SD = 0.83$) as higher than participants who collaborated with AI using non-verbal communication ($M = 3.866$, $SD = 0.82$, see Figure 5.14a).

Team composition also has a significant main effect on perceived teammate performance ($b = -0.23$, $\chi(1)^2 = 8.06$, $p < 0.01$). Participants in human-AI-AI teams perceive their teammates to have higher individual performance than in human-human-AI teams (see Figure 5.13). This can be explained by the fact that AI's skills at the task are better than most participants. Thus, people who collaborated with two AI teammates perceived their teammates to have better performance.

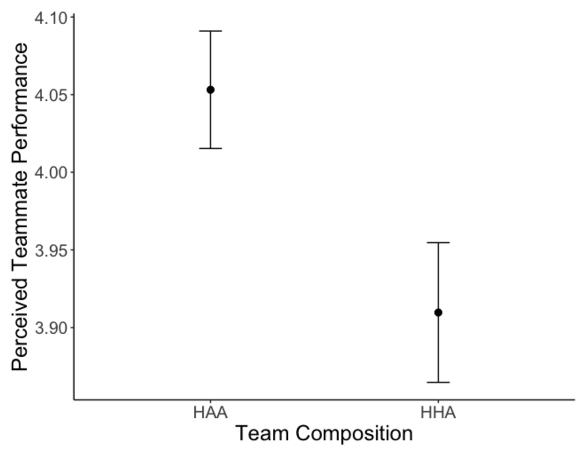
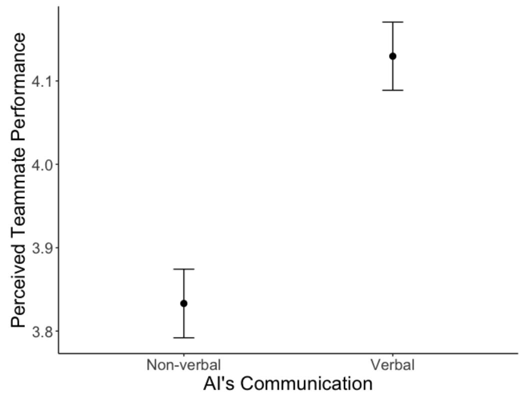
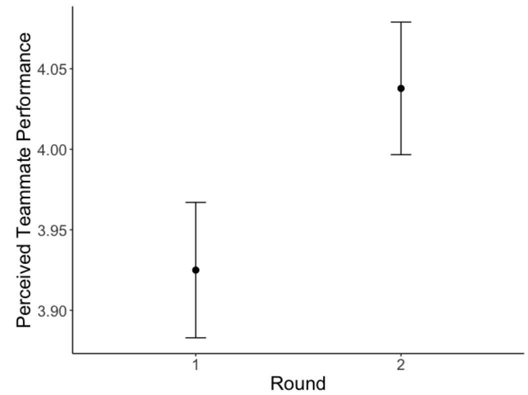


Figure 5.13: Perceived Teammate Performance by Team Composition

Moreover, the task round has a positive impact on perceived team performance ($b = 0.11$, $\chi(1)^2 = 5.94$, $p < 0.05$). Participants usually perceive their teammates to perform better in the second round ($M = 4.04$, $SD = 0.82$) than in the first round ($M = 3.93$, $SD = 0.84$, see Figure 5.14b).



(a) Perceived Teammate Performance by AI's Communication Approach



(b) Perceived Teammate Performance by Round

Figure 5.14: Perceived Teammate's Performance

Importantly, the linear mixed-effects model also shows an interaction effect between the identity of the teammate and the participants' gender ($b = -0.27$, $\chi(1)^2 = 5.18$, $p < 0.05$). The difference in women participants' perceived performance of an AI teammate ($M = 4.14$, $SD = 0.71$) versus a human teammate ($M = 3.63$, $SD = 0.92$) is larger than the difference in men's perceived performance of an AI teammate ($M = 4.10$, $SD = 0.71$) versus a human teammate ($M = 3.32$, $SD = 1.17$).

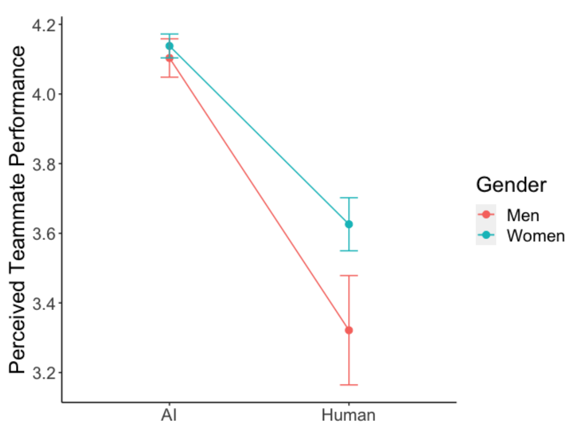


Figure 5.15: Perceived Performance of Human vs. AI Teammate by Gender

5.3.1.4 Team Effectiveness

We conducted a linear mixed-effects model with team effectiveness as the dependent variable and a random intercept set for each participant. The only significant effect shown in the model is a three-way interaction effect($b = -0.81$,$\chi(1)^2 = 6.35$,$p < 0.05$). However, since gender and NARS are highly correlated (0.966), this is removed from the linear mixed-effects model. We conclude that AI's communication approach, team composition, and other individual characteristics (i.e., gender, NARS, Skill) do not impact perceived team effectiveness.

5.3.1.5 Task Workload

A linear mixed-effects model was built with task workload as the dependent variable, and a random intercept set for each participant[1]. From the linear mixed-effects model, AI's communication has a significant impact on perceived task load ($b = -0.30$, $\chi(1)^2 = 4.78$, $p < 0.05$). People who collaborated with AI using non-verbal communication perceived the task as having a lower task load ($M = 3.24, SD = 0.88$) than people who collaborated with AI using verbal communication ($M = 3.51, SD = 0.84$). In addition, the gender of each individual significantly impacts how they perceive the task load ($b = 0.76$,$\chi(1)^2 = 18.28$,$p < 0.001$). Women tend to perceive the task load to be higher ($M = 3.60, SD = 0.82$) than men ($M = 2.79, SD = 0.71$, see Figure 5.17). One possible explanation is that men are more familiar with competitive games and are more likely to be comfortable with the competitive gaming environment.

[1]There were several data points that had high leverage (i.e., Cook's distance larger than 1). However, those points did not impact the linear mixed-effects model.

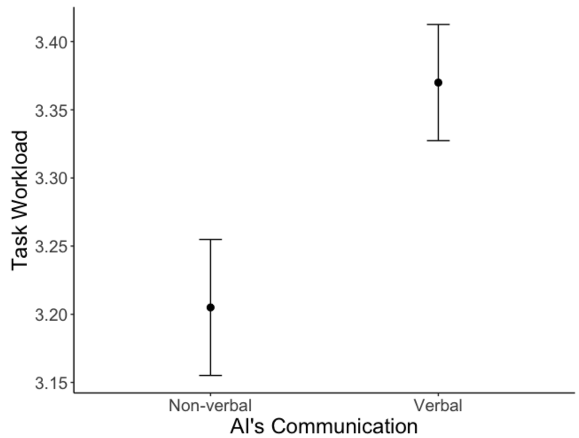


Figure 5.16: The impact of AI's verbal/non-verbal communication on task load

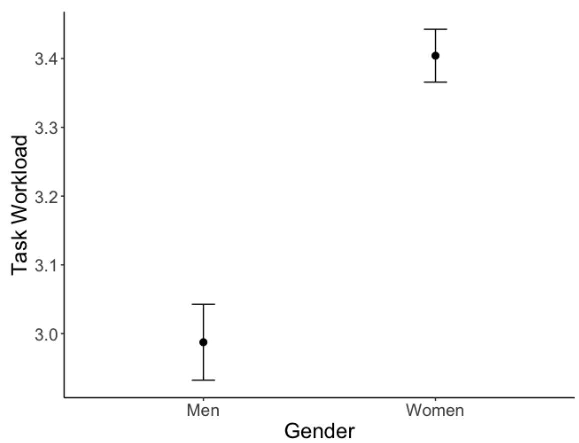


Figure 5.17: The impact of gender on task load

People's existing negative attitudes toward AI (NARS) also have a significant impact on perceived task load ($b = 0.36$, $\chi(1)^2 = 9.55$, $p < 0.01$). When an individual has a more negative an individual's attitude toward AI, they perceive the task to have a higher task load. Moreover, the task round has a significant effect on the perceived task load ($b = -0.08$, $\chi(1)^2 = 4.85$, $p < 0.01$). Participants perceive the task load to be lower in the second round ($M = 3.33$, $SD = 0.89$) than in the first round ($M = 3.42$, $SD = 0.86$).

5.3.1.6 Individual Performance

As the raw individual performance score was not normally distributed, it was transformed using a log() function. The transformed individual score was normally distributed. The linear mixed-effects model was built using the transformed individual performance score with a random intercept set for each participant.

From the linear mixed-effects model, participants' gender has a significant main effect on their individual performance ($b = -1.18$, $\chi(1)^2 = 28.74$, $p < 0.001$). As expected, participants identified as women ($M = 3.85$, $SD = 0.90$) performed worse than participants identified as men ($M = 5.31$, $SD = 0.75$, see Figure 5.18). It should be noted that the mean values and standard deviation values here are the average and standard deviation of the transformed individual performance score instead of the raw individual performance score.

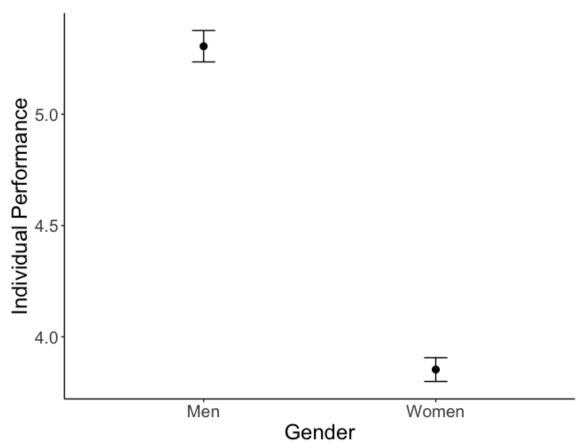


Figure 5.18: Individual Performance Score by Gender

Moreover, the impact of task rounds on individual performance is dependent on the participant's skills ($b = 0.30$, $\chi(1)^2 = 3.91$, $p < 0.05$). Experienced participants performed better in the second round ($M = 5.18$, $SD = 0.93$) than in the first round ($M = 5.03$, $SD = 0.83$), whereas inexperienced participants performed worse in the second round ($M = 3.79$, $SD = 0.90$) than in the first round ($M = 3.94$, $SD = 0.96$), as shown in Figure 5.19.

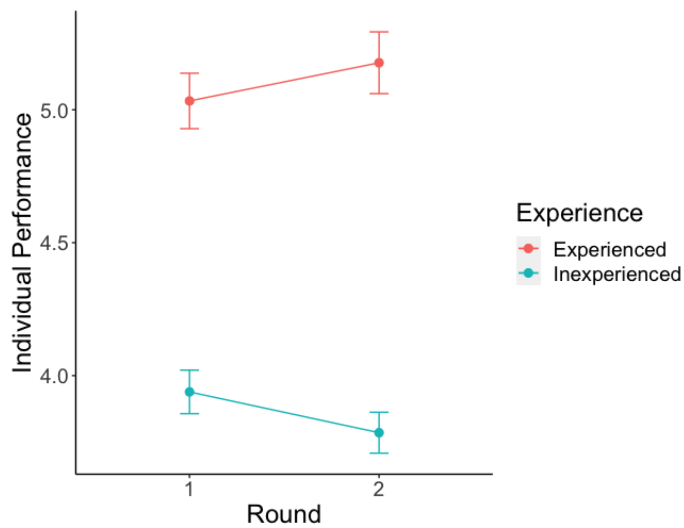


Figure 5.19: Individual Performance Score by Experience in Two Rounds

5.3.1.7 Team Performance

Similar to the individual performance score, the team performance score was also transformed using a log() function to ensure that the transformed team performance is normally distributed. We built a linear mixed-effects model with the transformed team performance score and a random intercept set for each participant.

From the linear mixed-effects model, we found a significant interaction effect between AI's communication approach and team composition on transformed team performance scores ($b = -0.06$, $\chi(1)^2 = 4.38$, $p < 0.05$).

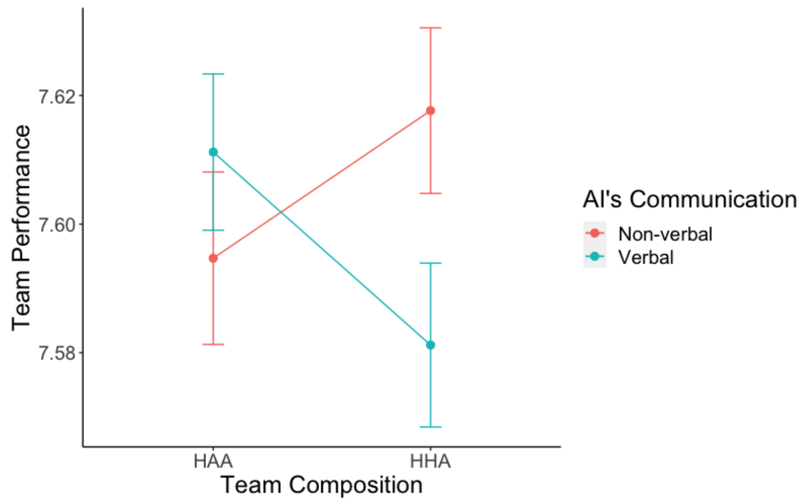


Figure 5.20: Team Performance Score by AI's Communication Approach in Two Team Compositions

As expected, team performance is significantly impacted by task rounds ($b = 0.04$, $\chi(1)^2 = 11.91$, $p < 0.001$). Participants performed better as a team in the second round ($M = 7.62$, $SD = 0.14$) than in the first round ($M = 7.58$, $SD = 0.11$).

Interestingly, while team composition does not have a significant main impact on team performance ($p > 0.05$), there is a significant interaction effect between team composition and gender on team performance ($b = 0.08$, $\chi(1)^2 = 7.72$, $p < 0.01$). In particular, men have better team performance in human-human-AI teams ($M = 7.65$, $SD = 0.14$) than in human-AI-AI teams ($M = 7.61$, $SD = 0.14$), whereas women have better team performance in human-AI-AI teams ($M = 7.60$, $SD = 0.12$) than in human-human-AI teams ($M = 7.58$, $SD = 0.12$). Even though men overall have better team performance than women, this indicates a difference of gender in various HAT team compositions.

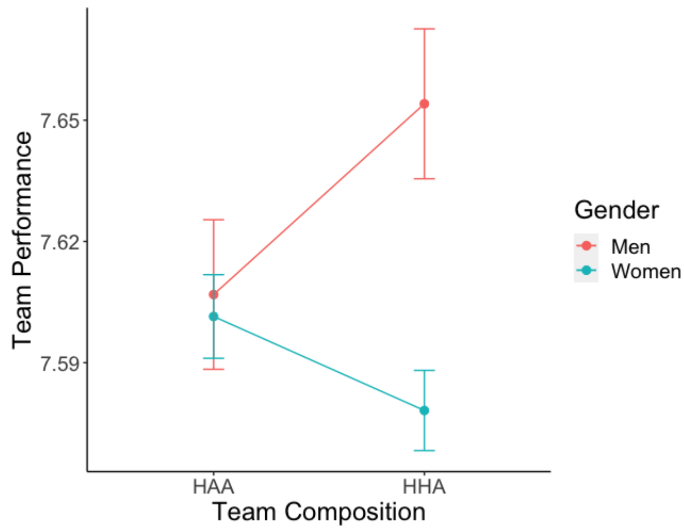


Figure 5.21: Team Performance Score by Team Compositions and Gender

5.3.1.8 Team Viability

The linear mixed-effects model of team viability has a random intercept set for each participant. The model indicates that the impact of team composition on team viability depends on people's existing attitude toward AI ($b = -0.58$, $\chi(1)^2 = 8.90$, $p < 0.01$). While people's existing attitude toward AI is negatively associated with team viability, this negative association is stronger in human-AI-AI teams than in human-human-AI teams. People who have a more negative attitude toward AI (i.e., higher NARS score) perceive higher team viability when there is another human on the HAT team.

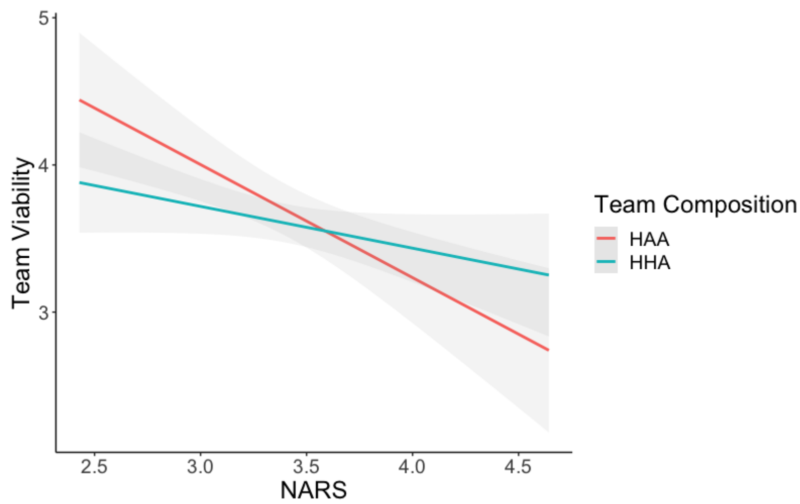


Figure 5.22: An Interaction Effect of Team Composition and NARS on Team Viability

Additionally, the impact of team composition on team viability also depends on people's previous experience with Rocket League ($b = -0.33$, $\chi(1)^2 = 5.58$, $p < 0.05$). In particular, experienced participants perceive human-human-AI teams ($M = 3.77$, $SD = 0.46$) as having higher team viability than human-AI-AI teams ($M = 3.71$, $SD = 0.73$), whereas inexperienced participants consider human-AI-AI teams to have higher team viability (HAA: $M = 3.68$, $SD = 1.02$; HHA: $M = 3.52$, $SD = 0.74$).

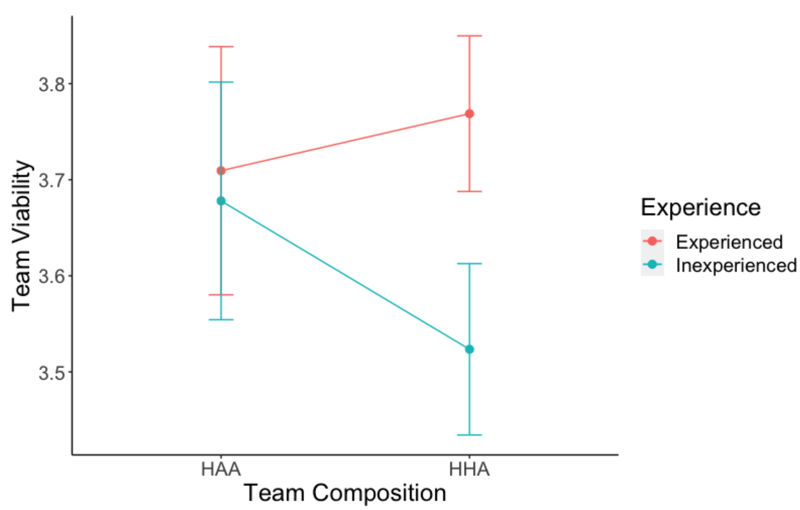


Figure 5.23: An Interaction Effect of Team Composition and RL Experience on Team Viability

The experimental analysis generates multiple significant findings that are summarized and presented in Table 5.9. Specifically, there are five main findings: (1) AI's communication approach does not impact human trust, but non-verbal communication is not under humans' definition of communication; (2) AI's verbal communication received better perceptions than AI's non-verbal communication (e.g., lower communication quality and lower perceived performance), but is perceived to have higher task load; (3) while both verbal and non-verbal communication facilitates the maintaining of situation awareness, non-verbal communication is not effective through this maintaining process; (4) AI with verbal communication achieves higher team performance in human-AI-AI teams, whereas AI with non-verbal has higher team performance in human-human-AI teams; and (5) gender differences play an important role in impacting human perceptions of the HAT and the team outcome.

Outcome Variables	Significant Predictors (1)
Trust in the Teammate	Identity *** + TeamComp:NARS * + Skill:Round *
Perceived Communication Quality	Condition *** + Round * + Identity *** + Skill:NARS*
Perceived Teammate Performance	Condition ** +TeamComp **+ Round *+ Identity:Gender *
Perceived Team Effectiveness	None (2)
Task Workload	Round ** + NARS ** + Gender *** + Condition *
Individual Performance	Gender *** + Condition * + Skill:Round*
Team Performance	Round *** + Condition:TeamComp * + TeamComp:Gender **
Team Viability	NARS ** + TeamComp:NARS ** + TeamComp:Skill *

Table 5.9: The Significant Predictors of Each Outcome Variable

(1) Note: When there is both an interaction effect and a main effect of a predictor on an outcome variable, only the significant interaction effects are included in the table.

(2) The three-way interaction effect of team composition, gender, and NARS on perceived team effectiveness is removed due to the high correlation between Gender and NARS.

5.3.2 Qualitative Findings

In this section, the qualitative findings are presented to explain how AI's communication approach is perceived regarding trust development and team coordination (RQ4). Based on the interview data from this study, two themes emerged on how AI's verbal vs. non-verbal communication approaches are perceived: (1) AI's communication is not perceived as a fundamental element in developing trust regardless of the communication approach; (2) AI's communication facilitates situation awareness but non-verbal communication is difficult to function effectively in fast-paced collaboration. All quotes presented in this section include the participant's team number, Player A/Player B (PA/PB), gender, and AI's communication condition.

5.3.2.1 AI's communication approach is not a fundamental element in developing trust.

In study 1, both the quantitative data and interview data indicate that AI with proactive communication capability obtains substantially higher trust than AI without proactive communication. However, the quantitative data in study 3 indicates that AI's communication approach does not significantly impact human trust in the teammate. Following this result, we focused on how AI's communication is perceived regarding trust development and why it does not impact trust.

Performance-based trust in the AI teammate. One frequently mentioned attribute in developing trust in the AI was *"they know what they are doing"*. Many participants described trust development as an understanding process of getting to know AI's advanced skills and performance:

> *At first, I was a little skeptical of it. But once I realized that could they*

were good with looping around the ball. Then they got the first goal. And I was like, Y'all got it. Because I didn't. (T28, PA, Female, Verbal)

When I was with the two AIs I was like, very trusting of this. They know what they're doing, a good job, like kudos to you. (T28, PB, Female, Verbal)

Because after I saw them score the first time I knew they were capable of doing it. So I trusted them. (T23, PA, Female, Non-verbal)

For PA in T28, AI getting the first goal was the confirmation of their skills and performance. Importantly, this approval of AI's performance was based on a comparison between the AI and themselves. Thus, inexperienced (novice) participants are likely to develop trust in the AI teammate fast due to the skill gaps. While trust in AI is mainly performance-based, AI's communication can still help increase trust that was built already on AI's actions. PB in T49 expresses this process in detail:

I think the trajectory helps me mostly, it lets me do my job better, which in turn helps the team overall, in terms of trust. Seeing what they're going to do helps me trust them more. But I think just based off of how they perform, I have already established a level of trust, just seeing how they score. And at that point, it's like, okay, I can trust them. But seeing the line secures that trust. (T49, PB, Male, Non-verbal)
I feel like they definitely help because I can see where they're going. So then I trust that they're gonna follow through with wherever their trajectory is. (T50, PA, Female, Non-verbal)

For PB in T49, the communication's role in trust development was mainly adding an extra layer on top of the action-based trust. While communication could help improve trust, it is not the *determining factor*. This emphasizes the fundamental role of AI's performance and skills in developing trust. Some other participants share similar opinions:

> *Honestly, I feel like trust kind of came from their actions more than their communication, because it's like, if I were to play that game not have the chat up there. I feel like I still would have trusted them to like, support or score goals or whatever. I feel like their actions were almost speaking louder than their words in this case.* (T46, PA, Male, Verbal)

> *Actions speak louder than words. So if they say something, and they're they actually did it, it shows, okay, if they do say this, they're actually going to do it when they say it.* (T38, PB, Female, Verbal)

Importantly, another participant indicates that even if AI's communication was incorrect, their trust would still be there as long as they perform well:

> *I feel like very little [in terms of how much AI's communication contributes to my trust in them] because if they said they were doing something, and they're doing something different then I still feel like they know what they were doing. So I was just like, okay, like you have a plan, just do your thing. I'll just stay here.* (T35, PB, Female, Verbal)

However, even though communication could contribute to trust increase, this only happens when certain criteria are met. For instance, when humans' communication does not change AI's decisions and actions, AI's communication is even less important in impacting trust:

> *I don't think it contributed to my trust in them that much, because I've kind of felt, again, like they were just doing whatever they wanted to do.*

So like he said, my trust of that communication kind of declined over the four games. (T40, PB, Female, Verbal)

When I told them at first what I was going to do, they let me do it, and then they would come and help. They kind of trusted me as well. But in the last two rounds, even when I communicated, they just kind of disregarded it. So it just made me think, oh, they're just kind of doing their own thing. I still trusted them because I knew that they could score but it was just a little different. (T23, PA, Female, Non-verbal)

According to these quotes, when humans' communication cannot have an impact on AI's actions, their trust in AI's communication decreased. However, they still trust AI due to AI's high-performance in the task.

AI's non-verbal communication is not perceived as "communication" as humans' definition. Our interview data indicate that AI's non-verbal communication was not considered the type of communication humans defined:

So you couldn't really trust it, since it was just based on coding and facts. I wouldn't trust just a trajectory. I would feel like, that's just the fact that they're going in different directions. It's not based on their own decision. (T29, PA, Male, Non-verbal)

For PA in T29, AI's non-verbal communication is just facts indicating the direction they will move toward. Such facts are incapable of presenting the same information as text messages, which indirectly show the decision process. PA emphasizes the importance of AI's communication in terms of sharing information about a decision that's made. Without such information, AI's non-verbal communication does not contribute to human trust. Other participants also emphasize the insufficiency of non-verbal communication:

> *I don't think the trajectory is much communication at all. It's not effective communication.* (T31, PA, Female, Non-verbal)

> *There's definitely a lack of communication, but they're really good at the game.* (T25, PB, Female, Non-verbal)

Moreover, as how study 1 emphasized the importance of AI providing immediate responses, non-verbal communication lacks this interaction attribute, leading to a perception of "no communication":

> *It was kind of hard to adapt, since there was* ***no communication*** *on their end. Obviously, I tried communicating, but like they didn't communicate back.* (T32, PA, Female, Non-verbal)

In addition, there is a gap between the presentation of non-verbal communication (i.e., AI's trajectory) and the real meaning behind it, which takes extra cognitive effort for humans to connect the two:

> *I don't think they like really communicated that well. So I guess I saw what they were doing, but I didn't really know what they were doing.* (T30, PB, Female, Non-verbal)

> *It made it really hard to say what they were going to be doing and then* ***they would just not say anything****, and then take over. And they're like, bump into me. It's always coming.* (T49, PA, Female, Non-verbal)

According to these two quotes, AI's non-verbal communication was not well understood by humans and thus could not be utilized in their coordination. Specifically for PA in T49, AI's non-verbal communication was not considered as communication, as how they described "they would not say anything". This perception of AI's non-verbal communication brings new insights into how AI's communication approach should be designed.

5.3.2.2 AI's communication facilitates situation awareness but is hard to utilize in fast-paced collaboration.

Similar to study 1, this study also shows that AI's communication helps to maintain situation awareness through collaboration so that participants are aware of AI's status and the team task's progress, based on which they know what to do next:

> *It definitely did help because, with the trajectory, I knew where they were coming from, because, on the screen, I couldn't always see where they were, depending on how close I was to the ball. So that helped me at least know when they were coming and where they were coming from.* (T26, PA, Female, Non-verbal)

According to PA in T26, the trajectory provides extra spacial information on AI's movement. Even if the AI was not in their eyesight, they could still be aware of their actions. In addition, AI's communication helps the team from two perspectives: (1) help humans make the next step decision based on the current task status, and (2) ensure the whole team stay on track and human teammates maintain a certain level of awareness regarding each team member's status:

> *[AI's communication helps me coordinate] a lot. If I saw their green line headed towards the ball and I didn't have a great shot I figured they did*

so I'd just bail off and then wait for another one. (T17, PA, Male, Non-verbal)

I think everyone's communication was highly needed, and that helped with the group's performance. Whereas if we didn't have that, it would have definitely been a lot more confusing, and we don't know what we're doing. So it's really nice kind of getting known or knowing what someone else is. (T28, PB, Female, Verbal)

PB in T28 emphasized the role of communication in impacting team performance. With AI's communication, humans could develop an awareness of the team task, and the AI teammate's progress. More importantly, this communication is highly appreciated when humans are novel at the task. Other participants also pointed out how AI's communication assists them to perform better from their own perspective:

It just helped me like, know where to go on the spot, because they were saying what they were doing. And I started looking at it. So that helped me a lot more to figure out what I needed to do. (T27, PB, Male, Verbal)

Definitely did help a lot. It helped me figure out more about what I was doing. Because I'd see, they'd say, I'll go hit the ball, then you are like okay. So it just helped me figure out like, whenever they would do something, I'd be like, Okay, that's something I can do in the next round, or later in the game. It helped me figure out how to play a little more. (T28, PA, Female, Verbal)

According to these quotes, AI's communication of their status provides information that helps humans proceed in the team task. Based on AI's actions, humans adapt and figure out the next step for themselves. More importantly, AI's communication of their status even provides a learning opportunity to humans. For PA in T32, AI is perceived to function as an example or coach which they can learn from later in the team task. However, both AI's verbal and non-verbal communication is insufficient in facilitating team coordination:

> *You know what their next step is, but you don't know how fast they're gonna get to the ball, or if you're gonna get there first. Obviously, it [communication] makes it easier to work together and collaborate, but since there's no communication (on that), either, that's also what makes it harder to (collaborate). We ran into each other many times because of that. The trajectory definitely helps rather than not having.* (T32, PA, Female, Non-verbal)

> *Sometimes I like some of the times they were pretty accurate and when helped me but other times there's confusing. They would take too long and I would be there too soon, or something and then it just won't work out.* (T31, PB, Female, Non-verbal)

As PA in T32 pointed out, the information provided through the AI's communication was not enough for human players to react effectively. In such a fast-paced environment, this lack of information is a road blocker in their coordination. AI's verbal information was hard to utilize in another way:

> *Like, kind of, but not really. It's harder with those messages just because it is such a fast-paced game. So they sent a message and then by the time I can see their messages, they didn't have a ball anymore.* (T24, PB, Female, Verbal)

According to PB in T24, AI's verbal communication in such a rapidly changing environment is not very helpful since the communication content is not received by humans on time to react accordingly. Given that the text communication of AI's movement is not built within the spatial environment, it takes additional mental processes to keep up with AI's status in the task. Comparing AI's verbal vs. non-verbal communication, though non-verbal communication provides more spatial information, it indeed has several limitations. First, AI's non-verbal communication loses the visibility of information in certain scenarios. For instance, PA mentioned that:

> *Since there are no text communications, if I was looking at the other team's goal, I couldn't see the lines, and I didn't know where my two teammates were. So I couldn't really rely on them. I don't think the green lines were the best method of communication. It was good in some scenarios when I was able to see them. But for the most part, I couldn't really rely on them.* (T34, PB, Male, Non-verbal)

When the non-verbal information is not visible, the maintenance of situation awareness is broken, which then hurt team performance and reduces the perceived reliability of the AI as a teammate. Second, AI's non-verbal communication could be distracting when they focus on their tasks:

> *For the trajectory, they both helped and sometimes it was a detriment because it's just a giant long green line all over your screen and that can be distracting.* (T26, PB, Male, Non-verbal)

Different from verbal communication which is displayed at the center of the screen, non-verbal communication is immersed in the 3D game space. The green line, which was designed to increase humans' attention, could instead distract humans.

Such distraction in time-sensitive tasks could be detrimental to team performance. Lastly, AI's non-verbal communication does not provide an interactive feature that is more natural to humans in terms of communication:

> *The communication is hard because all you're really going off of is their lines. And you can send stuff, but you can't get a response. So the only communication you have is like the directory lines. So I think that was the hardest aspect of like, the first few rounds.* (T33, PB, Female, Non-verbal)

Without text-based communication, AI's non-verbal communication is uni-directional communication. PB emphasized the importance of bi-directional communication, through which team members can exchange information and coordinate accordingly. Thus, it is difficult to utilize such non-verbal communication to coordinate effectively.

5.4 Discussion

Using a multiplayer online game as the context, this study explores how AI with verbal vs. non-verbal communication impacts team processes, team outcomes, and human perceptions in human-human-AI teams, and human-AI-AI teams.

For RQ1 and RQ2, while AI's communication approach does not impact trust in the teammate, the team composition (HHA vs. HAA) has a significant interaction effect with NARS on human trust in the AI teammate. AI's non-verbal communication is perceived to have a lower workload than AI's verbal communication. AI's communication approach has a significant main impact on individual performance. There are two significant interaction effects, (1) between AI's communication and team composition, and (2) between team composition and gender, on team performance. The team viability model shows two significant interaction effects, (1) team composition and NARS, and (2) team composition and individuals' prior Rocket League experience. In addition, AI with verbal communication is perceived to have higher communication quality and perform better than AI with non-verbal communication. There is also an interaction effect between team composition and gender on perceived teammate performance.

For RQ3, humans' trust in their AI teammates was significantly higher than their trust in their human teammates. Experienced individuals' trust in their teammates increased significantly more from round 1 to round 2 than inexperienced individuals. Moreover, men perceived the task to have a higher workload than women. In addition, men's individual performance is significantly higher than women's. There is also a significant interaction effect between prior Rocket League experience and task rounds on individual performance. In terms of team outcomes, team performance is significantly higher in the second round than in the first round, whereas NARS has a significant impact on team viability. In terms of human perceptions, teammates are perceived to have better communication quality and perform better in the second round than in the first round. AI teammates were perceived to have better communication quality than human teammates. Results also show a significant interaction effect between humans' previous experience with Rocket League and their negative attitude toward AI on perceived communication quality. There is a significant interaction effect between gender and identity of the teammate on perceived teammate performance.

For RQ4, the qualitative findings provide an understanding of how AI's communication approach is perceived regarding trust development and team coordination. Specifically, AI's communication is not perceived as a fundamental element in developing trust regardless of the communication approach. However, it can help increase trust that has been built based on AI's actions and performance. Importantly, AI's non-verbal communication is not perceived as real communication with humans. Similar to study 1, the interview findings show that AI's communication facilitates the maintenance of situation awareness in task coordination. However, in this study, rapidly changing collaboration environments have higher requirements for building effective communication that can facilitate coordination. One such requirement is that AI's communication needs to be shared without visibility limits and can be received and understood easily.

5.4.1 AI's non-verbal communication can only serve as a complementary element in building effective human-AI communication in HATs.

Non-verbal communication is a commonly used approach to facilitate teammates' coordination in both human-only teams [138] and human-machine teams [35]. While previous work has explored the impact of non-verbal communication in supporting coordination between humans and robots [35], our study focuses on the impact of AI's non-verbal communication on human-AI collaboration in *virtual* environments, where physical behaviors are not accessible. In particular, teams that operate in these virtual environments, called virtual teams, face significant challenges in developing trust and achieving effective team coordination [236, 180]. Due to a lack of visual presence and behavioral information, these virtual teams usually struggle to maintain awareness [196, 11]. Therefore, this study aims to explore how AI's non-verbal communication can assist in maintaining team members' awareness and how then it facilitates team coordination. The results of this study suggest that non-verbal communication can only be utilized as a complementary element to verbal communication in developing effective human-AI communication in virtual environments.

One interesting finding of our study is that both the quantitative findings and qualitative findings indicate humans have less positive perceptions of AI's non-verbal communication compared to AI's verbal communication. One major reason is that AI's non-verbal communication does not align with humans' traditional definition of communication. Previous work defines communication as "the discriminatory response of an organism to a stimulus" [249]. In other words, when the environment (e.g., teaming environment) has certain stimuli (e.g., a need of sharing information or requesting information) to an individual, they would provide a discriminatory response. Based on our findings, some people consider natural-language-based communication as the only way to communicate, either through textual messages or audio chat. However, the development of advanced AI introduced a new type of team member in the past decade, which then creates a new type of collaboration mode, human-AI teaming [301]. Importantly, non-verbal communication may be a possible way for AI to share information more effectively. Previous work has pointed out that non-verbal communication could be an effective way to share awareness-related information in a shared visual space [138]. However, previous work explores non-verbal communication in an environment that provides both verbal communication and non-verbal in human-human teams. Our findings propose a new insight into how AI's non-verbal communication should be employed in human-AI teams.

Moreover, one downside of AI using non-verbal communication is that non-verbal communication has higher demands on the receivers' side. First, non-verbal communication in a 3D space has a visibility limit. When it is not received appropriately (e.g., not within humans' eyesight) during their collaboration, it could result in a lack of communication, which then breaks the coordination. This supports previous work that indicates a lack of communication could also contribute to breakdowns in coordination in stressful situations [295]. However, even though it is challenging for AI to communicate effectively with only non-verbal communication, it can be applied as a supplement to verbal communication, especially in a competitive and time-sensitive collaboration environment. Our quantitative data indicates that the teams with non-verbal communication AI are perceived to have a lower task load (see Figure 5.16). This suggests that even though non-verbal communication has some limitations, it is perceived to have a lower workload. Taking good use of this feature is likely to build a high-performance team with efficient communication. An interesting finding is that AI with non-verbal communication has better team performance than AI with verbal communication in human-human-AI teams (see Figure 5.20).

5.4.2 New Perspective of Gender Differences in HATs

The imbalance of gender representation in gaming, especially in the esports community is a noticeable issue [184]. With the proportion of female video game users increasing from 38% to 48% in the past two decades [52], it is crucial to explore and examine women's perceptions and experience in games. However, plenty of work on esports reported findings with mainly or even only male participants [184]. Importantly, this study has identified the important role of gender differences in perceiving AI teammates in HATs.

The results of Study 2 indicate that gender has a significant interaction effect with scenarios on trust in the AI teammate. In this study, the quantitative results indicate that gender has several significant impacts (including both interaction effects and main effects) on team processes and team outcomes. It is important to note that this gender difference should not be misinterpreted as the general skill gap between men and women. In the linear mixed-effects models we built, participants' previous experience in RL was taken into consideration with two levels, inexperienced and experienced. While plenty of work has explored gender differences in various research areas (e.g., technology, accounting, e-learning) in the past decade [96, 156, 126], our study explores the impact of gender differences in perceiving and collaborating with AI teammates in *human-AI teaming* environments, especially under two team compositions. Our study found that men and women actually perceived AI's performance very similarly (see Figure 5.15) in human-AI teams.

The quantitative findings indicate that males perform better in team tasks than females do, which supports previous work on gender differences in game performance [272, 264]. While there is a significant difference between women's and men's performance in the team tasks, women and men have similar team performance in human-AI-AI teams (see Figure 5.23). This could be explained by AI's advanced skills at the team task, leading to an increase in team performance from human-human-AI teams to human-AI-AI teams. This supports previous work that shows human-AI-AI teams perform significantly better than human-human-human teams [232]. While this previous work did not find a significant difference in team performance between human-AI-AI and human-human-AI teams, it has a non-significant trend of teams with more AI having higher team performance [232]. Additionally, men have better team performance in human-human-AI teams than women, which is a novel finding on gender differences. This could be explained by the fact that men's individual performance is better than women, which contributes to team performance to a large extent.

In sum, gender differences are a crucial piece of personal characteristics to be examined in human-AI teaming research. While current work has not explored how gender differences impact team processes and team coordination in human-AI teams, this work is a good starting by showing that gender differences have impacts on certain team processes and team outcomes. More research is needed to further explore the role of gender in impacting human-AI teamwork, which benefits humans in designing effective human-AI teams. In spite of the fact that findings related to gender differences are challenging to be generalized to other contexts and other measurements, it is still crucial to take gender differences into consideration in designing human-AI communication in collaboration environments. The next subsection will briefly discuss how this study's findings may be applied in other contexts.

5.4.3 Future Application of the Findings

Given the context-dependent feature of AI's communication and human-AI collaboration, it is challenging to explore AI's communication in a context-excluded way. Thus, in this study, I chose multiplayer gaming as a context to investigate how AI using verbal vs. non-verbal communication impacts team coordination and how this is perceived by humans for three reasons. First, multiplayer online games provide a great environment for complex collaborations [78, 294], which enables us to examine how humans experience collaborating with agents. Second, multiplayer games usually have customized features that can be used to meet research needs, such as team size and AI's communication channels. Third, AI agents are commonly used in multiplayer games, which are accessible to most individuals. Even though this study was conducted in a specific context, the results can still be applied in other contexts. Previous work has emphasized the crucial role that team characteristics, team roles, and tasks play in shaping team communication [263, 226, 303]. Therefore, we will discuss various team and task characteristics of this study using previous work as a foundation. By specifying these features, it is more practical for future researchers to look at our results and consider how to apply our findings to their research.

Similar to Study 1, we selected team/task characteristics based on previous work [263] and summarize the ones that are applicable in this study in Table 5.10. One important characteristic of this study's context is that human-AI teams were set up in a competitive gaming environment with rapidly changing team status. This type of environment requires high mental demands from humans. By comparing these team and task characteristics between our study and future other work, the findings generated in this study can be more accurately applied to other contexts. Below is a discussion of how our findings can be applied beyond the context of this work.

Table 5.10: Team or Task Characteristics in Study 3

Dimension Type	Dimension	Details
Team	Roles	Humans and AI teammates all share the **same** responsibilities in team tasks.
	Size/Composition	Two types of HAT: **human-human-AI** and **human-AI-AI**. Each HAT has three team members.
Task	Task Type	A **competitive** task with opponents.
	AI Communication Method	Two types of AI communication is provided using a between-subject design: (1) AI using **verbal** communication (text-based chat channel); (2) AI using **non-verbal** communication (visual cues).
	Human Communication Method	A text-based chat channel.
	Situational Stressors	High stress: Five-minute timed task with a **timer** and a goal **scoreboard** displayed on the screen

Since one of the major research goals of this study is to explore how AI's verbal communication and non-verbal communication are perceived and used to facilitate team coordination separately, it is important to consider how non-verbal communication is designed in this study. In this specific context, the gaming environment is a 3D space, in which non-verbal communication is presented as a trajectory line. Thus, AI's non-verbal communication may not be visible when both the AI agents and the trajectory are not within humans' eyesight. One of the interview findings, *"non-verbal communication may not always be effective in facilitating team coordination"*, may not be applied to a context in which AI is always visible to humans. However, another finding, *"AI's communication facilitates the maintenance of situation awareness"*, is likely to be applied in this example context. Additionally, this finding was also reported in study 1, which utilized text-based communication in a different gaming environment and thus ensures the application in other contexts.

Another example is situational stressors. This study utilized a competitive, fast-paced, and time-sensitive environment. These features of the collaborative environment make collaboration more stressful and challenging. Compared to a much slower human-AI collaborative environment, such as data scientists [280], the quantitative results of the task load linear mixed-effects model may be substantially different. In addition, since this study was conducted with a competitive task with opponents, humans probably valued the team performance (i.e., winning the task) and AI's performance more than other contexts, such as the context of using social AI. In those contexts, AI's communication style or anthropomorphism may be more desired [228]. Therefore, the non-significant effect of AI's communication on trust may be different.

In sum, to ensure an accurate extension of our findings to other contexts, it is essential to identify the specific team and task characteristics of the target context. By comparing the target context's characteristics and our study's team/task features, our findings can be adapted accordingly to avoid inaccurate application, and then adapt our findings based on existing literature.

5.4.4 Limitations and Future Work

This study has several limitations. First, all participants are undergraduate students from a local university. This young population may perceive and collaborate with AI agents in a different way from other populations, such as senior adults. Future research should examine and extend the findings in this study to a population with more diverse backgrounds. Second, while this study endeavors to generate insights that can be applied in other contexts, this study examines AI's verbal and non-verbal communication within a specific context. Due to the specific context characteristics, the findings may be different in other contexts. The discussion over the future application of this study's findings provides an initial step for generalizing the results. Nevertheless, more studies are still needed to investigate how AI's verbal and non-verbal communication is perceived and how it impacts team coordination in other contexts. Third, this study utilized a series of five-minute tasks, which may result in findings different from tasks that are longer or not time-sensitive. As discussed in the Discussion section, these special task characteristics may impact the findings. Thus, future work should compare their target context against this context and develop their research plans based on the comparison.

Chapter 6

Conclusion

This dissertation focuses on understanding how AI's communication should be designed and developed to facilitate team coordination from various perspectives. Specifically, I explored how humans perceive AI's communication and how it supports team coordination using three different communication components: communication proactivity, communication content, and communication approach. This dissertation makes notable contributions to AI's communication within a human-AI teaming environment by providing both scientific insights. The three studies involved in this dissertation shed light on how AI's communication is perceived and interpreted by human teammates and how this further impacts team processes and team outcomes during their collaboration in dynamic virtual environments. The research findings of these studies provide a scientific foundation for future research to build upon and eventually develop a practical and comprehensive communication system that enables AI teammates to collaborate with humans effectively.

In this chapter, I will discuss the contribution and conclusion of the dissertation through three steps. First, I will revisit the dissertation-level research questions, and discuss how each study's findings answer the three research questions (RQ1, RQ2, and RQ3) as well as the umbrella research question (RQ0). Second, I will discuss the contributions of this study from two perspectives, including the contribution to Human-AI teaming research and the contribution to virtual team communication research. Lastly, I will extend the findings and insights beyond this dissertation and discuss future work that is inspired by this dissertation.

6.1 Revisiting Research Questions

In the Introduction of this dissertation, I proposed an umbrella research question **RQ0**: *How should AI teammates' communication strategies be designed and developed to achieve effective and smooth human-AI coordination in teaming environments?* Three research questions were proposed to detail this high-level research question. In this section, I will go through how these research questions (RQ1, RQ2, and RQ3) are answered by the three studies and then summarize how RQ0 is answered based on the answers of RQ1, RQ2, and RQ3.

6.1.1 RQ1: How does AI teammates' communication facilitate team processes (e.g., trust and situation awareness) through human-AI coordination in teaming environments?

Three studies address RQ1 from various perspectives. Each of these three studies explores one or more team processes using a specific communication component (see Figure 6.1). Study 1 explores the development of trust and situation awareness with AI's proactive communication using qualitative data. Study 2 examines the impact of AI's explanation on trust using quantitative data. Study 3 then explores how

AI using verbal vs. non-verbal communication impacts the trust development process and situation awareness using an interview. In addition, Study 3 also examines how AI's communication impacts task workload and trust using experimental data. While these studies all explore the team processes, the communication components each of them focuses on are different, and thus provide a more complete picture of how AI's communication impacts team processes in virtual HATs.

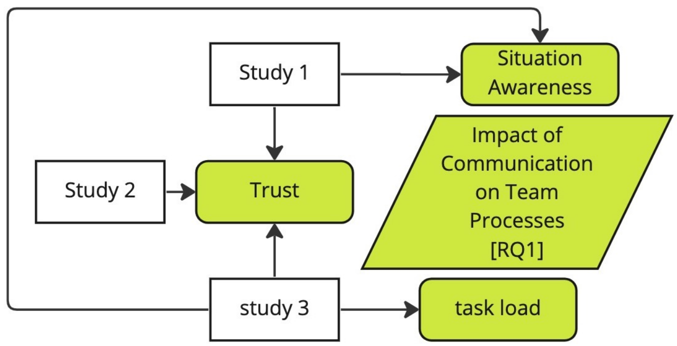


Figure 6.1: Dissertation Studies and RQ1

Study 1 explores how AI's communication proactivity impacts human trust in AI, and situation awareness in dyadic human-AI teams using both quantitative and qualitative methods. On the one hand, the experimental data in Study 1 points out that AI with proactive communication has a significant positive impact on humans' trust in the AI teammate. On the other hand, the qualitative analysis further details how AI with proactive communication results in higher trust. Specifically, AI with proactive communication provides detailed information on their progress, which then helps humans to take their next steps. Such benefits for humans' individual performance convert to the development of trust in the AI teammate. Moreover, AI's proactive communication is considered to have high transparency with the AI teammate. With this transparency, humans perceive AI with proactive communication as a reliable partner whom they can trust and team with. In addition to the trust in AI, study 1's qualitative data also provides an understanding of how AI teammates' proactive communication develops situation awareness. By informing their own progress, humans develop an understanding of the team's progress in completing the task. As AI continues to share information proactively, the team completes more and humans maintain their understanding of the team status through the process with AI's proactive communication.

Study 2 explores how AI's explanation impacts human trust in various scenarios and provides a different insight from study 1. Four different scenarios were utilized to examine how AI's explanations impact human trust. This study indicates that the impact of AI's explanation on trust depends on AI's specific actions. An interesting finding is that AI's explanations facilitate human trust in the AI teammate when explaining why AI disobeyed humans' orders, but hinder trust when explaining why AI lied to humans. This emphasizes the importance of the explanation's context-dependency feature, which calls for more research on the impact of AI's explanation in *human-AI teaming* environments.

Study 3 evaluates the impact of AI's verbal vs. non-verbal communication on three team processes, including human trust, situation awareness, and task load. The experimental data of Study 3 indicates that AI teammates' verbal or non-verbal communication does not impact human trust, but significantly impacts task load. In HATs where AI used non-verbal communication to share information, the task load was perceived to be significantly lowered than in HATs where AI used verbal communication to share similar information. While this indicates AI using non-verbal communication may reduce the cognitive load humans have to maintain in coordination with the AI, the interview data depicts the drawbacks of AI's non-verbal communication. First, AI's non-verbal communication does not align what some people's definition of "communication", which uses natural language to exchange information and has the interactive attribute. Second, AI's non-verbal communication in a 3D share space may not be visible in all scenarios, which breaks the coordination and potentially hurts team performance. Even though AI's non-verbal communication has these downsides, it is still able to facilitate coordination with humans. The information provided by non-verbal communication helps humans build situation awareness of the AI teammate's movement and plans, then take action accordingly. However, it was pointed out that this assistance of AI's non-verbal communication was not effective and is only better than having no communication at all.

Synthesizing the findings of three studies leads to the following answers to RQ1. While AI's communication is crucial in facilitating team processes in HATs, various communication components play different roles in this process: (1) AI's communication proactivity is a must in impacting human trust in time-sensitive collaborative tasks, (2) AI's communication approach (verbal vs. non-verbal) does not impact human trust, but non-verbal communication is perceived as low workload, (3) the impact of AI's explanation on human trust depends on the action that AI explains, and (4) AI's communication assist the development and maintenance of situation awareness in fast-paced team tasks but non-verbal communication is not effective in this process.

6.1.2 RQ2: How does AI teammates' communication impact team outcomes in virtual human-AI teams?

RQ2 focuses on the impact of AI's communication on team outcomes (e.g., team performance, team viability). Team outcomes are an essential element of teamwork. The development of various team concepts, such as trust in the AI teammate and situation awareness, aims to increase team performance as the ultimate goal. More importantly, communication is usually applied to facilitate team members to coordinate more effectively to achieve high team performance. Therefore, I will discuss how communication components impact team outcomes in HATs and thus they should be designed.

In answering RQ2, Study 1 and Study 3 explore two types of team outcomes using in-person experiments, including team viability and team performance (see Figure 6.2). Study 1 examines the impact of AI's *proactivity in communication* on team outcomes, whereas Study 3 focuses on the impact of AI's *verbal vs. non-verbal communication approach* on team outcomes. Both studies investigate the impact using experimental data. Team viability is a self-report measurement, whereas team performance is objectively calculated based on the task goal or task scoring rule.

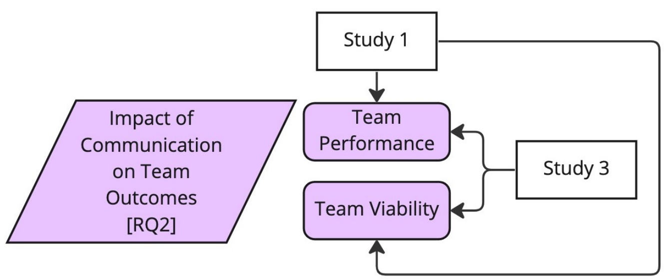


Figure 6.2: Dissertation Studies and RQ2

Study 1 shows that AI's communication proactivity does not significantly impact team viability. However, AI's proactivity in communication has an interaction effect with task rounds on team performance. Specifically, HATs with proactive communication AI achieve a higher increase in team performance from Round 1 to Round 3 than HATs with non-proactive communication AI. This indicates the long-term potential of AI using non-proactive communication to achieve team success. One possible explanation is that once the collaboration pattern between humans and AI is formed, AI's proactivity in communication can decrease to avoid distraction from humans completing their own responsibilities.

In Study 3, the statistical analysis indicates that AI's communication approach does not significantly impact team viability. However, AI's communication approach has a significant interaction effect with HAT's team composition on team performance. Specifically, HATs with AI's non-verbal communication achieved higher team performance in human-human-AI teams than human-AI-AI teams, whereas HATs with AI's verbal communication achieved lower team performance in human-human-AI teams than human-AI-AI teams.

Combining the results from Study 1 and Study 3, the following insights are presented to answer RQ2. First, both AI's proactivity in communication and AI's verbal/ non-verbal communication do not impact team viability. While only two communication components are explored, it is reasonable to infer that AI's communication probably does not contribute to human perceptions of the HAT's long-term success. Second, AI's communication proactivity and communication approach impact team performance, but in various ways: (1) the impact of AI's proactivity in communication on team performance changes over time (i.e., task round), and (2) the impact of AI's communication approach on team performance depends on the team composition of HATs (see Section 5.3.1 for more details on the impact of team composition on team performance). This indicates that while AI's communication impacts team performance, this impact is likely to change along with other task features or team characteristics.

6.1.3 RQ3: How do humans perceive and interpret AI teammates' communication in human-AI teams?

While RQ1 and RQ2 address two core elements of human-AI teamwork, RQ3 focuses on humans' perceptions of AI's communication. Given that humans usually perceive AI in various ways which are impacted by their previous experience with AI and technologies [301], and thus treat AI teammates differently from human teammates [177], understanding how humans interpret AI's communication is crucial to design effective human-AI communication in HATs. Thus, to achieve a comprehensive understanding of how humans perceive AI with various communication components, Study 1, Study 2, and Study 3 investigate humans' perceptions of AI's communication from various perspectives (see Figure 6.3). Specifically, Study 1 evaluates the impact of AI's *proactive communication* on perceived team performance, perceived satisfaction, and perceived team effectiveness. Study 2 examines the impact of AI's *explanations* on perceived team effectiveness. Study 3 evaluates the perceived communication quality of AI's *verbal and non-verbal communication*, and how AI's verbal and non-verbal communication impacts perceived teammate performance, as well as perceived team effectiveness. It should be noted that teammate perceptions and perceived satisfaction of teammate were also measured in Study 2, but were removed in the analysis process due to SEM's constraints. Taking these three studies together, a more complete picture of how humans perceive AI's communication in teaming environments is presented. Below I will discuss how studies answer each type of human perception.

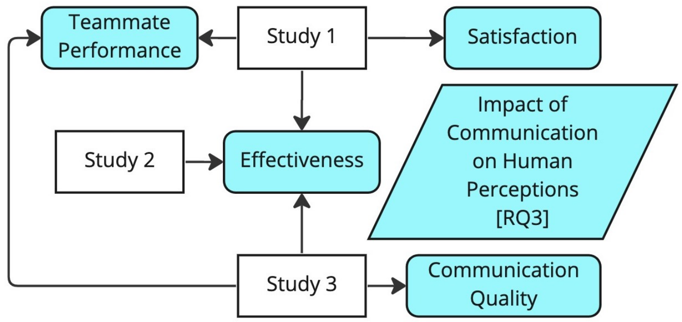


Figure 6.3: Dissertation Studies and RQ3

In terms of perceived team effectiveness, Study 1 finds no significant effects from AI's communication proactivity. Study 2 shows that AI's explanation positively impacts perceived team effectiveness when AI provides an explanation of why their decision was not to save the injured human teammate. Study 3 also shows that AI's verbal/non-verbal communication has no significant impact on perceived team effectiveness. To summarize the results of these three studies, only AI's explanation impacts perceived team effectiveness, but in certain scenarios. One explanation of this result is that AI explains the rationale of their decision, which is related to the effectiveness of their behaviors. In other words, AI providing an explanation of why they did not save the injured human teammate but focused on the team task is related to taking an effective decision to achieve team goals. In conclusion, while AI's communication proactivity and approach do not impact perceived team effectiveness, AI's communication content (e.g., explanation) could increase perceived team effectiveness, such as in a scenario where AI's explanation reveals a decision was made based on effectiveness.

Both Study 1 and Study 3 evaluate the impact of AI's communication on perceived performance. Study 1 focuses on AI's communication proactivity and shows that AI teammates with proactive communication are perceived to perform better than AI without such proactive communication. Study 3 indicates that AI with verbal communication is perceived to have better performance than AI with non-verbal communication. Both of these two studies present the significant impact of AI's communication on perceived teammate performance, which further indicates the importance of AI's communication in human perceptions.

Finally, two other types of human perceptions are investigated, the satisfaction of AI teammates and the perceived quality of AI teammates' communication. Specifically, humans are significantly more satisfied with AI teammates that have proactive communication than AI teammates that do not have proactive communication (shown in Study 1). Moreover, AI with verbal communication is perceived to have higher communication quality than AI with non-verbal communication (shown in Study 3).

To summarize the answer to RQ3, I concluded the above findings as follows. First, AI's communication proactivity and approach (verbal vs. non-verbal) have no impact on perceived team effectiveness (Study 1, Study 3), but AI's explanation could impact team effectiveness, depending on the specific scenario (Study 2). Second, both AI's proactive communication (Study 1) and AI's verbal communication (Study 3) have positive impacts on perceived teammate performance. Last, AI with proactive communication received higher satisfaction than AI without such proactive communication (Study 1). In addition, AI's verbal communication is perceived as having higher quality than AI's non-verbal communication (Study 3).

6.1.4 RQ0: How should AI teammates' communication strategies be designed and developed to achieve effective and smooth human-AI coordination in teaming?

Taken together, by synthesizing the answers of RQ1, RQ2, and RQ3, this subsection discussed the answer to RQ0, how AI's communication should be structured and designed, from three perspectives.

First, from a team processes perspective, AI teammates using **verbal** communication to **proactively** share information with humans is beneficial to human-AI coordination and should be included in AI's communication design, but non-verbal communication can only be used as a complement to AI's verbal communication in HATs. Moreover, the design of AI's communication content (i.e., explanation in this work) is dependent on specific scenarios and AI's actions to a great extent.

Second, through the lens of team outcomes, AI teammates' communication proactivity and communication approach should be selected based on other task/team characteristics. AI's **proactive** communication should be applied in the early stage of human-AI coordination in repeated team tasks, and AI should switch to using non-proactive communication once the human-AI collaboration mode forms. Moreover, in human-AI-AI teams, AI's verbal communication should be utilized instead of non-verbal communication.

Last but not least, from a human perception standpoint, AI's **proactive** communication and **verbal** communication are perceived positively and should be included in AI's communication design. Specifically, AI with proactive communication is perceived to have better performance and higher satisfaction than AI without proactive communication. AI's verbal communication is also perceived as having higher communication quality than AI's non-verbal communication. In addition, the design of AI's communication content heavily depends on specific scenarios.

In sum, the following communication strategies are proposed for AI teammates to achieve effective coordination with humans. First, AI teammates should be proactively communicating with human teammates at the initial stage of collaboration. The flexibility of switching between proactive communication and non-proactive communication is also needed in human-AI coordination. Second, verbal communication is preferred by humans as the communication approach AI uses, but non-verbal communication can only be applied by AI teammates as complement to verbal communication. Third, AI's verbal/non-verbal communication benefits team performance differently under various team compositions. Thus, the selection of AI's communication approach should take team composition into consideration. Last, the design of AI's explanation depends on AI's actions and scenarios. AI's explanation should not be provided in every scenario in human-AI teams.

6.2 Overall Dissertation Contributions

This dissertation with three studies tied with each other contributes to human-AI teaming research and virtual team communication research. These contributions provide novel insights into the design and structure of AI's communication in achieving effective team coordination, high human perceptions, and outstanding team performance. In this subsection, I will discuss the contribution of this dissertation to human-AI teaming research and virtual team communication research.

6.2.1 Contributions to Human-AI Teaming Research

AI has been increasingly applied in both work and people's everyday life to support humans to complete certain tasks. Along with this are the growing interest in human-AI teaming, which is expected to be the future collaboration mode of humans and AI agents. While existing work has explored human-AI teaming from various perspectives, such as shared team mental models, human trust in AI teammates, and human perceptions of AI teammates, the exploration of AI's communication in HATs just start to emerge. There is still a research gap on AI's communication in human-AI teaming, especially how AI's communication can be designed to achieve smooth coordination with humans. Therefore, this dissertation aims to close this research gap by providing an in-depth understanding of how AI's communication is posited in human-AI teamwork and the nuances of AI's communication, which is particularly important in designing and building effective virtual human-AI teams. I will discuss how each work of the dissertation contributes to the human-AI team research area.

Study 1 expands the current CSCW research on human-AI communication by providing an in-depth understanding of how an AI teammate's proactive communication can impact teaming processes (i.e., trust and situation awareness changes) when involved in human-AI coordination. This new insight further helps CSCW researchers and AI designers and developers better design human-AI communication in a teaming environment that facilitates team coordination through trust development and team situation awareness development. Additionally, Study 1 depicts how AI's communication proactivity impacts team outcomes, which is the ultimate goal of HATs. This generates empirical evidence on how AI's communication proactivity should be designed to achieve high team performance. Future CSCW/HCI research on AI's communication proactivity in HATs could benefit from these findings.

Study 2 is one of the first studies exploring AI's explanation in *human-AI teaming* environments. Given that humans view an AI's behaviors differently from how they judge a human's, this work provides insights into how AI's explanations impact human trust in a teaming context and compares that perception to a human teammate. Importantly, this study describes how the impact of AI's explanations varies when their explained actions are different. Instead of suggesting AI always provides an explanation, this study argues that AI's explanations should only be provided in certain scenarios in virtual environments. In addition, study 2 specifically focuses on team decisions that involve tangible side effects, which provides empirical evidence of when AI's explanation positively impacts human perceptions and when it harms human perceptions. This new finding extends current research on human-AI teaming and AI's explanation by shedding light on *when* AI should provide an explanation of their behaviors.

Study 3 yields insights into how AI's verbal vs. non-verbal communication is perceived and interpreted, as well as how this impacts AI's coordination with humans under two team compositions. This study provides new insights on how AI's verbal/non-verbal communication is perceived, and how it impacts team processes and team outcomes. In addition, these findings present an in-depth understanding of how AI's communication approach needs to be modified when HAT's team composition changes. Importantly, Study 3 provides novel and valuable insights into personal characteristics in HATs through an exploration of gender differences' impact on team processes, team outcomes, and human perceptions. The findings emphasize the crucial role of human gender differences in fast-paced human-AI coordination.

Taken together, these three studies provide numerous insights on how to better design AI's communication with the communication components in teaming environments based on the findings each study produced. This dissertation serves as a foundation of AI's communication in human-AI teaming for future research to use and further explore more nuances in human-AI communication.

6.2.2 Contributions to Virtual Team Communication Research

This dissertation contributes to the virtual team communication research community. While prior work on virtual team communication has thoroughly investigated how human-human communication in virtual environments can be improved from various perspectives (e.g., increasing perceived social presence), leading to the generation of CMC, the development of advanced AI in the past two decades brings in an unprecedented collaboration mode into teamwork, which is *human-AI teams*. This unique type of team has received increasing interest in HCI/CSCW, leading to some new research areas, one of which is human-AI communication in virtual collaborations. Although a large amount of prior work on virtual team communication has thoroughly investigated how human-human communication in virtual environments can be improved from various perspectives (e.g., increasing perceived social presence), human-AI communication in virtual environments is still understudied. This dissertation contributes to closing this research gap by providing a comprehensive understanding of how humans perceive AI's communication in virtual teams from two perspectives. I will discuss the contribution of this dissertation to virtual team communication research in two parts. First I will detail each dissertation study's contribution. Then I will discuss the contribution of this dissertation as a whole.

Study 1 contributes to the virtual team communication research by empirically identifying communication strategies that AI should apply in a *teaming* context to support human-AI collaboration. These communication strategies are essential to establish effectively coordinated human-AI teams, especially in dynamic environments. More importantly, study 1 synthesizes the identified communication strategy into a human-AI communication model and multiple design implications, which helps AI researchers and developers design AI teammates with better communication capabilities in real-world human-AI teams.

Study 2 provides a new perceptive on AI's explanation in virtual team environments. This study brings new comprehension of how AI's explanation is perceived compared to humans providing the same explanation in virtual environments. The comparison of AI's communication against human's communication emphasizes the difference between human-AI communication and human-human communication in virtual teams. Importantly, this study generates an understanding of how AI's explanation impacts perceived team effectiveness in various scenarios. Instead of suggesting AI provide an explanation of their actions all the time, this study argues that AI should only provide explanations in certain scenarios.

Study 3 extends the current virtual team communication research by exploring AI using verbal and non-verbal communication in a virtual spatial environment. Compared to AI using non-verbal communication, AI using verbal communication is perceived more positively by humans. This brings a new insight on human perception of AI's communication approach when the information is very similar. This new understanding further provides guidance for AI experts on how to structure and design AI's communication in virtual environments.

Taken together, these three dissertation studies contribute to virtual team communication research from two perspectives. First, this dissertation generates insights on how an AI teammate's communication is perceived compared to *a human teammate* with the same communication in virtual HATs. This is an important starting point for exploring human-AI communication in virtual teams since humans usually prefer to apply how they communicate and collaborate with humans to how they communicate and collaborate with AI teammates [301]. However, given that AI has unique machine natures that are different from humans, this dissertation serves as a crucial first step to developing effective virtual human-AI communication by understanding how humans perceive AI's communication compared to a human's.

Second, the work of this dissertation explores how humans perceive AI's communication from various angles using three different components, including communication proactivity, communication content (i.e., explanation), and communication approach (i.e., verbal vs. non-verbal communication). Each study provides empirical evidence on how humans perceive and interpret AI's communication in virtual environments. To summarize, this dissertation contributes to developing a foundational understanding of how humans perceive and interpret AI's communication using three different communication components. These findings are important for future researchers to build upon and further explore the communication between humans and AI in virtual teams. In addition, findings reveal what type of communication strategies AI should employ in virtual teaming environments, which researchers could utilize to design their studies.

6.2.3 Ethical Considerations in AI's Communication Design

The recent release of ChatGPT has led to broad applications of large language models (LLMs) in both industry and academia, such as using ChatGPT to do thematic analysis for humans by giving detailed instructions or using ChatGPT to write articles. However, new ethical concerns and considerations come along with this widespread use of ChatGPT [155]. In a teaming environment, AI's communication design also needs to take ethics into consideration to avoid risks caused by inappropriate use of AI's communication (e.g., the communication content). I will discuss ethical considerations in designing AI communication in human-AI teams from two perspectives.

First, ethical rules and standards on AI's communication design (e.g., communication content, communication approaches, and communication proactivity) should be developed based on previous work on AI ethics in human-AI interaction and human-AI teams, as well as, AI's communication in teams. These standards should guide and lead to the development of an AI equipped with ethical communication capabilities (e.g., the ethical communication content) in teams. Previous work has emphasized the crucial role of identifying AI ethics principles and implementing them in AI's algorithms in *teaming* environments [76, 120]. While prior studies have endeavored to build AI ethical guidelines and principles, their focus is mainly AI's decision-making in ethics-related scenarios. With the emergence of AI's communication design in human-AI teams, research is in need on AI ethics regarding communication design at both an *individual* level and a *team* level. On the one hand, one example of AI's ethical considerations at an individual level is the design of AI's communication content. Ethical rules on designing AI's communication content need to be developed to avoid potential harm caused by AI's communication, such as AI using curse words or being mean to human teammates. On the other hand, AI's communication at a team level needs to be designed with ethical principles and guidelines that align with humans' common ethical ideologies. Even though each individual's ethical ideologies are different, there is common ground on certain behaviors or in specific contexts. One such example is that AI should always avoid sharing inaccurate information intentionally. In addition, people in certain contexts, such as healthcare or military, may have more consensus ethical ideologies which can be used in AI's communication design.

Second, the potential for *unethical use* of AI communication should be restricted with detailed ethical standards. While communication serves a crucial role in developing human trust in AI leading to effective human-AI team coordination, the use of AI's communication needs to take ethics into account. One example is the application of AI communication in increasing human trust in the AI teammate. While one important goal of AI communicating with humans is to develop and maintain human trust in the AI, the trust in the AI should be at an appropriate amount. In other words, humans should know the AI's capability enough to make decisions on when to trust the AI and when not to. Thus, while AI's communication aims to positively impact human trust, it should be applied ethically. In addition, when human trust in the AI teammate outweighs human trust in other human teammates in a triad or more complicated HATs, team conflicts may occur, which could further result in broken interpersonal relationships among human team members. This could be more dangerous and risky in sensitive contexts, such as healthcare and military. In addition, when AI proactively communicates inaccurate information, intentionally or unintentionally, the potential increase of human trust in the AI needs to be corrected to avoid team failure in human-AI collaboration [104].

6.3 Future Work

As a starting point, this dissertation opens up more research opportunities for future work to explore. While this dissertation endeavors to draw a complete picture of AI's communication in HATs, communication is a complicated construct that is highly dependent on contexts and the subjects that communicate with each other. More research is needed to further explore and identify how AI's communication is posited in facilitating human-AI coordination in HATs.

First, this dissertation explores three components of AI's communication in human-AI collaboration. More communication components need to be explored in future work to get other perspectives on the impact of AI's communication on human-AI coordination. One example is to focus on *when* AI should send messages to humans (e.g., communication time points). Although this could be extremely challenging due to the task-dependent feature, one approach could be conceptualizing the time points as specific event-related communication, AI's progress-related communication, or human teammate-related communication (e.g., giving warnings to reduce risks or increase team effectiveness).

Second, to generate robust insights on AI's communication strategy design, future research should explore the same communication component in different contexts to achieve robust findings on AI's communication strategy design. In doing so, future research needs to examine a specific communication component and compare the findings (e.g., its impact on team processes, human perceptions, and team outcomes) in different types of tasks (various task characteristics). One example is to examine the impact of AI's communication proactivity in a slow-paced task environment. As how I discussed the application of findings beyond the study context in Study 1, the fast-paced task environment requires certain communication strategies, such as providing immediate responses. However, this may not apply in slow-paced contexts. Moreover, research in other contexts may produce new strategies that do not apply in fast-paced environments. The accumulation of such research will generate more robust findings on human-AI communication, thus helping human-AI team experts design effective AI communication.

Last, in addition to AI's communication, this dissertation has shown the impact of personal characteristics on human perceptions, team processes, and team outcomes in Study 2 and Study 3. However, it should be noted that personal characteristics were not the focus of these two dissertation studies. For instance, while both Study 2 and Study 3 indicated a strong impact of gender differences on perceiving AI agents in HATs, the studies did not manipulate gender as an independent variable (i.e., having the same number of participants of each gender). To get more robust findings on the impact of personal characteristics on team coordination in HATs, future research should focus on one or more specific personal characteristics (e.g., gender differences) and explore their impact on team processes and team outcomes.

6.4 Closing Remarks

Taken together, this dissertation makes a valuable contribution to the human-AI teaming field. Specifically, this dissertation presents empirical findings on how communication posits in human-AI teamwork, especially how various communication components impact and facilitate team processes through human-AI coordination. By exploring these components (i.e., communication proactivity, communication content, and AI's communication approach), this dissertation culminates in a multi-level understanding of AI's communication and its design in human-AI teaming. Such understanding involves both objective impacts of AI's communication on teaming (i.e., team performance) and subjective interpretation from human teammates. This comprehension of AI's communication in HATs contributes to both virtual team communication research and human-AI teaming research. As a whole, these insights provide scientific foundations for CSCW researchers to further explore AI's communication in teaming environments.

Appendices

Appendix A Survey Measurements

A.1 Study 1 Demographic & Game Experience Questions

1. What is your age? (*Text Entry Box*)
2. Please indicate your gender. (*Female, Male, Non-binary / third gender, Prefer not to say*)
3. Please specify your ethnicity. (*Hispanic and Latino; Non-Hispanic White; Asian; Black or African American; Native Americans and Alaska Natives; Native Hawaiian or Pacific Islander; Other.*)
4. What is your current level of education? (*No high school completed; Some high school, no diploma; High school diploma or equivalent; Bachelor's degree achieved; Master's degree achieved; Doctoral degree achieved.*)
5. Is English your first language? (*Yes, No.*)
6. How long have you been playing multiplayer digital games? (*Never played before; Less than 1 year; 1-3 years; 3-5 years; 5-10 years; More than 10 years.*)
7. How long do you usually spend on playing games every week on average? (*Less than 1 hour; 1-5 hours; 5-10 hours; 10-20 hours; More than 20 hours.*)
8. What type of game do you usually play (check one that you play most)? (*Strategy; RPG (role-playing game); Sports; FPS (First-person shooter); MMORPGs (Massively multiplayer online role-playing games); MOBA (Multiplayer online battle arena); None; Others-specify (Text Entry).*)
9. How familiar are you with ArmA 3? (*Not familiar at all; Slightly familiar; Moderately familiar; Very familiar; Extremely familiar.*)

A.2 Study 3 Demographic Questions

1. What is your age in years? (*Text Entry*)
2. Please indicate your gender. (*Female, Male, Non-binary/third-gender, Prefer not to say*)
3. Is English you first language? (*Yes, No.*)
4. Please specify your ethnicity. (*Hispanic and Latino; Non-Hispanic White; Asian; Black or African American; Native Americans and Alaska Natives; Native Hawaiian or Pacific Islander; Other.*)
5. What is your current level of education? (*No high school completed; Some high school, no diploma; High school diploma or equivalent; Bachelor's degree achieved; Master's degree achieved; Doctoral degree achieved.*)
6. How often do you play Rocket League? (*Never; Not in a long time; A few times a year; A few times a month; At least every week; Almost every day.*)
7. What platform do you play rocket league on the most? (*I don't play rocket league; Playstation; Xbox; Nintendo Switch; PC; Other-specify.*)
8. Do you use a controller or keyboard and mouse to play Rocket League? (*I don't play Rocket League; Keyboard and Mouse; Controller.*)
9. How would you rate your skill at Rocket League? (*I don't play Rocket League; Not very good; Decent; Pretty good; Expert level.*)

A.3 Team Viability

This measurement was developed by [56].

1. The members of this team could work for a long time together.
2. Most of the members of this team would welcome the opportunity to work as a group again in the future.
3. This team has the capacity for long-term success.
4. This team has what it takes to be effective in the future.
5. This team would work well together in the future.
6. This team has positioned itself well for continued success.
7. This team has the ability to perform well in the future.
8. This team has the ability to function as an ongoing unit.
9. This team should continue to function as a unit.
10. This team has the resources to perform well in the future.
11. This team is well positioned for growth over time.
12. This team can develop to meet future challenges.
13. This team has the capacity to sustain itself.
14. This team has what it takes to endure in future performance episodes.

A.4 Perceived AI Teammate Performance

This measurement is adapted from [57].

The AI teammate I worked with:

- did a fair share of the team's work.

– made a meaningful contribution to the team.

– communicated effectively with teammates.

– monitored whether the team was making progress as expected.

– helped the team plan and organize its work.

– completed tasks that they agreed to complete with minimal assistance from team members.

– has the skills and abilities that were necessary to do a good job.

– respectfully voiced opposition to ideas.

– was actively involved in solving problems the team faced.

A.5 Trust in the AI Teammate

This measurement is adapted from [154].

1. In general, I trusted the AI teammate I just worked with.

2. I felt like I had to monitor my AI teammate's actions during the game. [R]

3. I felt like my AI teammate had harmful motives in the task. [R]

4. I felt confident in the AI teammate I just worked with.

5. I felt like my AI teammate allowed joint problem solving in the task.

6. I felt fearful, paranoid, and or skeptical of the AI teammate during the game. [R]

A.6 Perceived Satisfaction of the AI Teammate

1. How satisfied are you with your AI teammate?
2. I am willing to team up with the AI teammate Zeus again.
3. Overall, I am satisfied with my AI teammate.
4. I am happy to have the AI teammate Zeus on my team.
5. I am happy with Zeus's contribution in the task.

A.7 Perceived Team Effectiveness

This measurement is adapted from [212].

1. I am happy with Zeus's contribution in the task.
2. My AI teammate was highly committed to the team during the task.
3. The researcher will be satisfied with the team product.
4. People outside of the team would give the team positive feedback about this work today.
5. The researcher would be satisfied with the team's performance.
6. Team members worked better together at the end of the task than at the beginning.
7. Team members were more aware of group dynamics at the end of the task than when they began the task.
8. Being a part of this team helped members appreciate different types of people.

A.8 Task Load Index

This measurement is adapted from [102].

1. How much mental and perceptual activity was required (e.g., thinking, deciding, calculating, remembering, looking, searching, etc.)?
2. How much time pressure did you feel due to the rate or pace at which the tasks or task elements occurred?
3. How hard did you have to work to accomplish your level of performance?
4. How successful do you think you were in accomplishing the goals of the task set by the experimenter?*
5. How insecure, discouraged, irritated, stressed and annoyed did you feel during the task?

A.9 Negative Attitude toward Robots Scale (NARS)

1. I would feel uneasy if I was given a job where I had to use robots.
2. The word "robot" means nothing to me.
3. I would feel nervous operating a robot in front of other people.
4. I would hate the idea that robots or artificial intelligence were making judgments about things.
5. I would feel very nervous just standing in front of a robot.
6. I would feel paranoid talking with a robot.
7. I would feel uneasy if robots really had emotions.

8. Something bad might happen if robots developed into living things.

9. I feel that if I depend on robots too much, something bad might happen.

10. I am concerned that robots would be a bad influence on children.

11. I feel that in the future society will be dominated by robots.

12. I would feel relaxed talking with robots.

13. If robots had emotions, I would be able to make friends with them.

14. I feel comforted being with robots that have emotions.

A.10 Perceived Communication Quality

1. I understood my AI/human teammate's communication.

2. The word "robot" means nothing to me.

3. My AI/human teammate's communication helped me to know their next steps.

Appendix B Interview Questions

B.1 Study 1 Interview Questions

1. How did you feel about the AI teammate Zeus in the game?
2. What do you think about its performance?
3. How much did you trust Zeus?
4. How would you describe your trust in Zeus across three missions?
5. How did you feel about AI Zeus's communication?
   (a) How did that influence your trust and collaboration?
   (b) How would you prefer an AI to communicate with you?
6. How did you feel about your AI teammate taking action by himself/itself?
7. How would you feel if the AI teammate could only take action with your approval? (e.g., AI teammate would not do anything if you didn't direct them)
8. In what contexts/scenarios do you feel you would feel comfortable letting an AI teammate take actions without your approval? Why? What about in your real life?
9. Comparing low-level engagement of AI teammates, but high team performance, and high-level engagement of AI teammates with low team performance, how would you perceive those two?
   (a) Which do you prefer and why?
   (b) How about in real life?

10. If your teammate could communicate with you, how would you like the AI to present information? (e.g., what do you think if AI could talk like a human and have social elements in the conversation?)

11. What about in your real life? If you have to collaborate with an AI in your job/study.

12. Anything Zeus could improve?

B.2 Study 3 Interview Questions

Warm-up

1. How did the tasks go?
   (a) Which part was good/bad?

2. How do you compare four rounds of tasks?
   (a) Which round was better than the others? Why so?
   (b) Which round was worse than the others? Why so?
   [Hint: in terms of your own performance, your teammate's performance, or your collaboration]

Part 1: Coordination & Communication Perception Questions

3. What do you think of AI's communication in four rounds?
   (a) Did you pay attention to what was communicated? Why or why not?
   (b) Could you describe how AI's communication facilitated your coordination with them if any? (How much do you think AI's communication contributes to your team coordination?)

    (c) Was it different for different rounds? How so?

4. Do you trust your AI teammates? Why or why not?
    (a) How did your trust in each AI teammate develop? Could you describe the development process?
    (b) Which teammate do you trust more? Why?
    (c) What role does communication play in this trust development process, if at all?
    (d) Was it different across the four rounds?
5. How much do you feel like you were aware of AI's behaviors/actions? What about your human teammate?

Part 2: Perceptions of AI Teammates

6. How would you describe AI's role and responsibilities in your team in the two different team compositions?
7. Which AI teammate do you prefer to collaborate with most? Why?
8. What will you change if you can improve one aspect of your AI teammate?

Bibliography

[1] Jennie A. Abrahamson and Victoria L. Rubin. Discourse structure differences in lay and professional health communication. *Journal of Documentation*, 68(6):826–851, January 2012. Publisher: Emerald Group Publishing Limited.

[2] Ikuesan R Adeyemi, Shukor Abd Razak, and Mazleena Salleh. Individual difference for HCI systems: Examining the probability of thinking style signature in online interaction. In *2016 4th International Conference on User Science and Engineering (i-USEr)*, pages 51–56, August 2016.

[3] Sruthy Agnisarman, Kapil Chalil Madathil, and Jeffery Bertrand. Context-based visual aids to support the situation awareness of field engineers conducting windstorm risk surveys. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, volume 63, pages 1898–1902. SAGE Publications Sage CA: Los Angeles, CA, 2019.

[4] Sruthy Agnisarman, Kapil Chalil Madathil, Jeffrey Bertrand, Kalyan Piratla, Dhaval Gajjar, and Anand Gramopadhye. Context enabled decision aids to support the situation awareness and performance of risk engineers carrying out loss prevention surveys. *International Journal of Industrial Ergonomics*, 83:103108, 2021.

[5] Kumar Akash, Tahira Reid, and Neera Jain. Improving human-machine collaboration through transparency-based feedback–part ii: Control design and synthesis. *IFAC-PapersOnLine*, 51(34):322–328, 2019.

[6] Bonnie D Allen. Serious gaming for building a basis of certification for trust and trustworthiness of autonomous systems. In *2018 Aviation Technology, Integration, and Operations Conference*, page 3844, 2018.

[7] Mohammad Alsharo, Dawn Gregg, and Ronald Ramirez. Virtual team effectiveness: The role of knowledge sharing and trust. *Information & Management*, 54(4):479–490, 2017.

[8] Shoshana Altschuller and Raquel Benbunan-Fich. Trust, performance, and the communication process in ad hoc decision-making virtual teams. *Journal of Computer-Mediated Communication*, 16(1):27–47, 2010.

[9] Saleema Amershi, Dan Weld, Mihaela Vorvoreanu, Adam Fourney, Besmira Nushi, Penny Collisson, Jina Suh, Shamsi Iqbal, Paul N Bennett, Kori Inkpen, et al. Guidelines for human-ai interaction. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, pages 1–13, 2019.

[10] Stavros Antifakos, Nicky Kern, Bernt Schiele, and Adrian Schwaninger. Towards improving trust in context-aware systems by displaying system confidence. In *Proceedings of the 7th international conference on Human computer interaction with mobile devices & services*, pages 9–14, 2005.

[11] David J Armstrong and Paul Cole. Managing distances and differences in geographically distributed work groups. 1995.

[12] Zahra Ashktorab, Casey Dugan, James Johnson, Qian Pan, Wei Zhang, Sadhana Kumaravel, and Murray Campbell. Effects of communication directionality and ai agent differences in human-ai interaction. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, pages 1–15, 2021.

[13] Zahra Ashktorab, Q Vera Liao, Casey Dugan, James Johnson, Qian Pan, Wei Zhang, Sadhana Kumaravel, and Murray Campbell. Human-ai collaboration in a cooperative game setting: measuring social perception and outcomes. *Proceedings of the ACM on Human-Computer Interaction*, 4(CSCW2):1–20, 2020.

[14] Katrin Auspurg and Thomas Hinz. *Factorial survey experiments*, volume 175. Sage Publications, 2014.

[15] Mohammad Q Azhar and Elizabeth I Sklar. A study measuring the impact of shared decision making in a human-robot team. *The International Journal of Robotics Research*, 36(5-7):461–482, 2017.

[16] Jeremy N Bailenson, Andrew C Beall, and Jim Blascovich. Gaze and task performance in shared virtual environments. *The journal of visualization and computer animation*, 13(5):313–320, 2002.

[17] Jeremy N Bailenson, Nick Yee, Dan Merget, and Ralph Schroeder. The effect of behavioral realism and form realism of real-time avatar faces on verbal disclosure, nonverbal disclosure, emotion recognition, and copresence in dyadic interaction. *Presence: Teleoperators and Virtual Environments*, 15(4):359–372, 2006.

[18] MS Balaji and Diganta Chakrabarti. Student interactions in online discussion forum: Empirical research from'media richness theory'perspective. *Journal of interactive online learning*, 9(1), 2010.

[19] Gagan Bansal, Besmira Nushi, Ece Kamar, Walter S Lasecki, Daniel S Weld, and Eric Horvitz. Beyond accuracy: The role of mental models in human-ai team performance. In *Proceedings of the AAAI Conference on Human Computation and Crowdsourcing*, volume 7, pages 2–11, 2019.

[20] Gagan Bansal, Besmira Nushi, Ece Kamar, Daniel S Weld, Walter S Lasecki, and Eric Horvitz. Updates in human-ai teams: Understanding and addressing the performance/compatibility tradeoff. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, pages 2429–2437, 2019.

[21] Gagan Bansal, Tongshuang Wu, Joyce Zhou, Raymond Fok, Besmira Nushi, Ece Kamar, Marco Tulio Ribeiro, and Daniel Weld. 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*, pages 1–16, 2021.

[22] Valérie Beaudouin, Isabelle Bloch, David Bounie, Stéphan Clémençon, Florence d'Alché Buc, James Eagan, Winston Maxwell, Pavlo Mozharovskyi, and Jayneel Parekh. Flexible and Context-Specific AI Explainability: A Multidisciplinary Approach. SSRN Scholarly Paper 3559477, Social Science Research Network, Rochester, NY, March 2020.

[23] Gregory Bedny and David Meister. Theory of activity and situation awareness. *International Journal of cognitive ergonomics*, 3(1):63–72, 1999.

[24] Suzanne T Bell and Brian J Marentette. Team viability for long-term and ongoing organizational teams. *Organizational Psychology Review*, 1(4):275–292, 2011.

[25] Rhyse Bendell, Jessica Williams, Stephen M Fiore, and Florian Jentsch. Supporting social interactions in human-ai teams: Profiling human teammates from sparse data. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, volume 65, pages 665–669. SAGE Publications Sage CA: Los Angeles, CA, 2021.

[26] Adam L Berger, Vincent J Della Pietra, and Stephen A Della Pietra. A maximum entropy approach to natural language processing. *Computational linguistics*, 22(1):39–71, 1996.

[27] Adam J Berinsky, Michele F Margolis, and Michael W Sances. Separating the shirkers from the workers? making sure respondents pay attention on self-administered surveys. *American Journal of Political Science*, 58(3):739–753, 2014.

[28] Gregory R Berry. Enhancing effectiveness on virtual teams: Understanding why traditional team skills are insufficient. *The Journal of Business Communication (1973)*, 48(2):186–206, 2011.

[29] Swati Kaul Bhat, Neerja Pande, and Vandana Ahuja. Virtual team effectiveness: An empirical study using sem. *Procedia Computer Science*, 122:33–41, 2017.

[30] Timothy Bickmore and Justine Cassell. Social dialongue with embodied conversational agents. In *Advances in natural multimodal dialogue systems*, pages 23–54. Springer, 2005.

[31] Guo Bin. Moderating effects of task characteristics on information source use: An individual-level analysis of r&d professionals in new product development. *Journal of Information Science*, 35(5):527–547, 2009.

[32] Steven Bird, Ewan Klein, and Edward Loper. *Natural language processing with Python: analyzing text with the natural language toolkit.* " O'Reilly Media, Inc.", 2009.

[33] Nathan Bos, Judy Olson, Darren Gergle, Gary Olson, and Zach Wright. Effects of four computer-mediated communications channels on trust development. In *Proceedings of the SIGCHI conference on human factors in computing systems*, pages 135–140, 2002.

[34] Alice M Brawley and Cynthia LS Pury. Work experiences on mturk: Job satisfaction, turnover, and information sharing. *Computers in Human Behavior*, 54:531–546, 2016.

[35] Cynthia Breazeal, Cory D Kidd, Andrea Lockerd Thomaz, Guy Hoffman, and Matt Berlin. Effects of nonverbal communication on efficiency and robustness in human-robot teamwork. In *2005 IEEE/RSJ international conference on intelligent robots and systems*, pages 708–713. IEEE, 2005.

[36] Niklas Bussmann, Paolo Giudici, Dimitri Marinelli, and Jochen Papenbrock. Explainable ai in fintech risk management. *Frontiers in Artificial Intelligence*, 3:26, 2020.

[37] Abhizna Butchibabu, Christopher Sparano-Huiban, Liz Sonenberg, and Julie Shah. Implicit coordination strategies for effective team communication. *Human factors*, 58(4):595–610, 2016.

[38] Carrie J Cai, Samantha Winter, David Steiner, Lauren Wilcox, and Michael Terry. " hello ai": Uncovering the onboarding needs of medical practitioners for human-ai collaborative decision-making. *Proceedings of the ACM on Human-Computer Interaction*, 3(CSCW):1–24, 2019.

[39] Hancheng Cao, Vivian Yang, Victor Chen, Yu Jin Lee, Lydia Stone, N'godjigui Junior Diarrassouba, Mark E Whiting, and Michael S Bernstein. My team will go on: Differentiating high and low viability teams through team interaction. *Proceedings of the ACM on Human-Computer Interaction*, 4(CSCW3):1–27, 2021.

[40] Jessy Ceha, Ken Jen Lee, Elizabeth Nilsen, Joslin Goh, and Edith Law. Can a humorous conversational agent enhance learning experience and outcomes? In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, pages 1–14, 2021.

[41] Carolina Centeio Jorge, Myrthe L Tielman, and Catholijn M Jonker. Artificial trust as a tool in human-ai teams. In *Proceedings of the 2022 ACM/IEEE International Conference on Human-Robot Interaction*, pages 1155–1157, 2022.

[42] Tathagata Chakraborti and Subbarao Kambhampati. (when) can ai bots lie? In *Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society*, pages 53–59, 2019.

[43] Kim-Yin Chan and Fritz Drasgow. Toward a theory of individual differences and leadership: understanding the motivation to lead. *Journal of applied psychology*, 86(3):481, 2001.

[44] Shruthi Chari, Oshani Seneviratne, Daniel M Gruen, Morgan A Foreman, Amar K Das, and Deborah L McGuinness. Explanation ontology: a model of explanations for user-centered ai. In *International Semantic Web Conference*, pages 228–243. Springer, 2020.

[45] Prithvijit Chattopadhyay, Deshraj Yadav, Viraj Prabhu, Arjun Chandrasekaran, Abhishek Das, Stefan Lee, Dhruv Batra, and Devi Parikh. Evaluating visual conversational agents via cooperative human-ai games. In *Fifth AAAI Conference on Human Computation and Crowdsourcing*, 2017.

[46] Yuhang Che, Heather Culbertson, Chih-Wei Tang, Sudipto Aich, and Allison M Okamura. Facilitating human-mobile robot communication via haptic feedback and gesture teleoperation. *ACM Transactions on Human-Robot Interaction (THRI)*, 7(3):1–23, 2018.

[47] Jessie YC Chen, Shan G Lakhmani, Kimberly Stowers, Anthony R Selkowitz, Julia L Wright, and Michael Barnes. Situation awareness-based agent transparency and human-autonomy teaming effectiveness. *Theoretical issues in ergonomics science*, 19(3):259–282, 2018.

[48] Jessie YC Chen, Shan G Lakhmani, Kimberly Stowers, Anthony R Selkowitz, Julia L Wright, and Michael Barnes. Situation awareness-based agent transparency and human-autonomy teaming effectiveness. *Theoretical issues in ergonomics science*, 19(3):259–282, 2018. Publisher: Taylor & Francis.

[49] Victor Cheung, Y-L Betty Chang, and Stacey D Scott. Communication channels and awareness cues in collocated collaborative time-critical gaming. In *Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work*, pages 569–578, 2012.

[50] Erin K Chiou, Mustafa Demir, Verica Buchanan, Christopher C Corral, Mica R Endsley, Glenn J Lematta, Nancy J Cooke, and Nathan J McNeese. Towards human–robot teaming: tradeoffs of explanation-based communication strategies in a virtual search and rescue task. *International Journal of Social Robotics*, 14(5):1117–1136, 2022.

[51] KR1442 Chowdhary. Natural language processing. *Fundamentals of artificial intelligence*, pages 603–649, 2020.

[52] J. Clement. U.s. gamers by gender 2022, Oct 2022.

[53] Nancy J Cooke, Mustafa Demir, and Nathan McNeese. Synthetic teammates as team players: Coordination of human and synthetic teammates. Technical report, Cognitive Engineering Research Institute Mesa United States, 2016.

[54] Nancy J Cooke, Eduardo Salas, Janis A Cannon-Bowers, and Renee J Stout. Measuring team knowledge. *Human factors*, 42(1):151–173, 2000.

[55] Nancy J Cooke, Renee J Stout, and Eduardo Salas. A knowledge elicitation approach to the measurement of team situation awareness. In *Situational Awareness*, pages 157–182. Routledge, 2017.

[56] Jessica Nicole Cooperstein. Initial development of a team viability measure. 2017.

[57] Tammy N Crutchfield and Kimberly Klamon. Assessing the dimensions and outcomes of an effective teammate. *Journal of Education for Business*, 89(6):285–291, 2014.

[58] Richard L Daft and Robert H Lengel. Organizational information requirements, media richness and structural design. *Management science*, 32(5):554–571, 1986.

[59] Nils Dahlbäck, Arne Jönsson, and Lars Ahrenberg. Wizard of oz studies—why and how. *Knowledge-based systems*, 6(4):258–266, 1993.

[60] Ewart J De Visser, Richard Pak, and Tyler H Shaw. From 'automation'to 'autonomy': the importance of trust repair in human–machine interaction. *Ergonomics*, 61(10):1409–1427, 2018.

[61] Mustafa Demir, Nathan J McNeese, and Nancy J Cooke. Team communication behaviors of the human-automation teaming. In *2016 IEEE international multi-disciplinary conference on cognitive methods in situation awareness and decision support (CogSIMA)*, pages 28–34. IEEE, 2016.

[62] Mustafa Demir, Nathan J McNeese, and Nancy J Cooke. Team situation awareness within the context of human-autonomy teaming. *Cognitive Systems Research*, 46:3–12, 2017.

[63] Mustafa Demir, Nathan J McNeese, and Nancy J Cooke. Understanding human-robot teams in light of all-human teams: Aspects of team interaction and shared cognition. *International Journal of Human-Computer Studies*, 140:102436, 2020.

[64] Alan R Dennis, Robert M Fuller, and Joseph S Valacich. Media, tasks, and communication processes: A theory of media synchronicity. *MIS quarterly*, pages 575–600, 2008.

[65] Alan R Dennis and Susan T Kinney. Testing media richness theory in the new media: The effects of cues, feedback, and task equivocality. *Information systems research*, 9(3):256–274, 1998.

[66] Alan R Dennis and Joseph S Valacich. Rethinking media richness: Towards a theory of media synchronicity. In *Proceedings of the 32nd Annual Hawaii International Conference on Systems Sciences. 1999. HICSS-32. Abstracts and CD-ROM of Full Papers*, pages 10–pp. IEEE, 1999.

[67] Bella M DePaulo, Deborah A Kashy, Susan E Kirkendol, Melissa M Wyer, and Jennifer A Epstein. Lying in everyday life. *Journal of personality and social psychology*, 70(5):979, 1996.

[68] Andrew Dillon and Charles Watson. User analysis in HCI — the historical lessons from individual differences research. *International Journal of Human-Computer Studies*, 45(6):619–637, December 1996.

[69] Julia Driver. *Consequentialism*. Routledge, 2011.

[70] Julia Eisenberg and Aparna Krishnan. Addressing virtual work challenges: learning from the field. *Organization Management Journal*, 15(2):78–94, 2018.

[71] Mica R Endsley. Design and evaluation for situation awareness enhancement. In *Proceedings of the Human Factors Society annual meeting*, volume 32, pages 97–101. Sage Publications Sage CA: Los Angeles, CA, 1988.

[72] Mica R Endsley. Toward a theory of situation awareness in dynamic systems. *Human factors*, 37(1):32–64, 1995.

[73] Mica R Endsley. Supporting human-ai teams: Transparency, explainability, and situation awareness. *Computers in Human Behavior*, 140:107574, 2023.

[74] Eileen B Entin and Elliot E Entin. Assessing team situation awareness in simulated military missions. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, volume 44, pages 73–76. SAGE Publications Sage CA: Los Angeles, CA, 2000.

[75] Neta Ezer, Sylvain Bruni, Yang Cai, Sam J Hepenstal, Christopher A Miller, and Dylan D Schmorrow. Trust engineering for human-ai teams. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, volume 63, pages 322–326. SAGE Publications Sage CA: Los Angeles, CA, 2019.

[76] Christopher Flathmann, Beau G Schelble, Rui Zhang, and Nathan J McNeese. Modeling and guiding the creation of ethical human-ai teams. In *Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society*, pages 469–479, 2021.

[77] Thomas Franke, Christiane Attig, and Daniel Wessel. A personal resource for technology interaction: development and validation of the affinity for technology interaction (ati) scale. *International Journal of Human–Computer Interaction*, 35(6):456–467, 2019.

[78] Guo Freeman and Donghee Yvette Wohn. Understanding esports team formation and coordination. *Computer supported cooperative work (CSCW)*, 28(1-2):95–126, 2019.

[79] Jolene Galegher and Robert E Kraut. Computer-mediated communication and collaborative writing: Media influence and adaptation to communication constraints. In *Proceedings of the 1992 ACM conference on Computer-supported cooperative work*, pages 155–162, 1992.

[80] Gege Gao, Aehong Min, and Patrick C Shih. Gendered design bias: gender differences of in-game character choice and playing style in league of legends. In *Proceedings of the 29th Australian Conference on Computer-Human Interaction*, pages 307–317, 2017.

[81] Reza Ghaiumy Anaraky, Yao Li, and Bart Knijnenburg. Difficulties of measuring culture in privacy studies. *Proceedings of the ACM on Human-Computer Interaction*, 5(CSCW2):1–26, 2021.

[82] Cristina B Gibson and Susan G Cohen. *Virtual teams that work: Creating conditions for virtual team effectiveness*. John Wiley & Sons, 2003.

[83] Cristina B Gibson and Jennifer A Manuel. Building trust. *Virtual teams that work*, pages 59–86, 2003.

[84] Barney G Glaser, Anselm L Strauss, and Elizabeth Strutzel. The discovery of grounded theory; strategies for qualitative research. *Nursing research*, 17(4):364, 1968.

[85] Luis R Gomez-Mejia, David B Balkin, Robert L Cardy, and Kenneth P Carson. *Managing human resources.* Pearson/Prentice Hall Upper Saddle River, NJ, 2007.

[86] Li Gong. How social is social responses to computers? the function of the degree of anthropomorphism in computer representations. *Computers in Human Behavior*, 24(4):1494–1509, 2008.

[87] Camilla Göras, Karolina Olin, Maria Unbeck, Karin Pukk-Härenstam, Anna Ehrenberg, Mesfin Kassaye Tessma, Ulrica Nilsson, and Mirjam Ekstedt. Tasks, multitasking and interruptions among the surgical team in an operating room: a prospective observational study. *BMJ open*, 9(5):e026410, 2019.

[88] Jamie C Gorman, Nancy J Cooke, and Jennifer L Winner. Measuring team situation awareness in decentralized command and control environments. In *Situational Awareness*, pages 183–196. Routledge, 2017.

[89] H. P. Grice. *Logic and Conversation.* Brill, December 1975. Pages: 41-58 Section: Speech Acts.

[90] Jonathan Grudin. From tool to partner: The evolution of human-computer interaction. *Synthesis Lectures on Human-Centered Interaction*, 10(1):i–183, 2017.

[91] Rosanna E Guadagno, Jim Blascovich, Jeremy N Bailenson, and Cade McCall. Virtual humans and persuasion: The effects of agency and behavioral realism. *Media Psychology*, 10(1):1–22, 2007.

[92] Andrew A Gumbs, Isabella Frigerio, Gaya Spolverato, Roland Croner, Alfredo Illanes, Elie Chouillard, and Eyad Elyan. Artificial intelligence surgery: How do we get to autonomous actions in surgery? *Sensors*, 21(16):5526, 2021.

[93] Charlotte N Gunawardena. Social presence theory and implications for interaction and collaborative learning in computer conferences. *International journal of educational telecommunications*, 1(2):147–166, 1995.

[94] Charlotte N Gunawardena and Frank J Zittle. Social presence as a predictor of satisfaction within a computer-mediated conferencing environment. *American journal of distance education*, 11(3):8–26, 1997.

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

[96] Florian Habler, Marco Peisker, and Niels Henze. Differences between smart speakers and graphical user interfaces for music search considering gender effects. In *Proceedings of the 18th International Conference on Mobile and Ubiquitous Multimedia*, pages 1–7, 2019.

[97] Thomas F. Hack, Lesley F. Degner, and Patricia A. Parker. The communication goals and needs of cancer patients: a review. *Psycho-Oncology*, 14(10):831–845, 2005. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/pon.949.

[98] Russell Haines. Activity awareness, social presence, and motivation in distributed virtual teams. *Information & Management*, 58(2):103425, 2021.

[99] Nader Hanna and Deborah Richards. The impact of communication on a human-agent shared mental model and team performance. In *Proceedings of the 2014 international conference on Autonomous agents and multi-agent systems*, pages 1485–1486, 2014.

[100] Nader Hanna, Deborah Richards, et al. The impact of virtual agent personality on a shared mental model with humans during collaboration. In *AAMAS*, pages 1777–1778, 2015.

[101] Helen Harris, Jeremy N Bailenson, Alexia Nielsen, and Nick Yee. The evolution of social behavior over time in second life. *Presence: Teleoperators and Virtual Environments*, 18(6):434–448, 2009.

[102] Sandra G Hart and Lowell E Staveland. Development of nasa-tlx (task load index): Results of empirical and theoretical research. In *Advances in psychology*, volume 52, pages 139–183. Elsevier, 1988.

[103] Jacob Haspiel, Na Du, Jill Meyerson, Lionel P Robert Jr, Dawn Tilbury, X Jessie Yang, and Anuj K Pradhan. Explanations and expectations: Trust building in automated vehicles. In *Companion of the 2018 ACM/IEEE international conference on human-robot interaction*, pages 119–120, 2018.

[104] Allyson I Hauptman, Wen Duan, and Nathan J Mcneese. The components of trust for collaborating with ai colleagues. In *Companion Publication of the 2022 Conference on Computer Supported Cooperative Work and Social Computing*, pages 72–75, 2022.

[105] Patrick Hemmer, Max Schemmer, Michael Vössing, and Niklas Kühl. Human-ai complementarity in hybrid intelligence systems: A structured literature review. *PACIS 2021 Proceedings*, 2021.

[106] Robert R Hoffman, Shane T Mueller, Gary Klein, and Jordan Litman. Metrics for explainable ai: Challenges and prospects. *arXiv preprint arXiv:1812.04608*, 2018.

[107] Andrea B Hollingshead, Joseph E McGrath, and Kathleen M O'Connor. Group task performance and communication technology: A longitudinal study of computer-mediated versus face-to-face work groups. *Small group research*, 24(3):307–333, 1993.

[108] Andreas Holzinger, Chris Biemann, Constantinos S Pattichis, and Douglas B Kell. What do we need to build explainable ai systems for the medical domain? *arXiv preprint arXiv:1712.09923*, 2017.

[109] Ming Hou. Impact: A trust model for human-agent teaming. In *2020 IEEE International Conference on Human-Machine Systems (ICHMS)*, pages 1–4. IEEE, 2020.

[110] Rick H Hoyle. *Structural equation modeling: Concepts, issues, and applications*. Sage, 1995.

[111] Yan Huang, S. Shyam Sundar, Zhiyao Ye, and Ariel Celeste Johnson. Do women and extroverts perceive interactivity differently than men and introverts? Role of individual differences in responses to HCI vs. CMC interactivity. *Computers in Human Behavior*, 123:106881, October 2021.

[112] Troy Innocent and Stewart Haines. Nonverbal communication in multiplayer game worlds. In *Proceedings of the 4th Australasian conference on Interactive entertainment*, pages 1–9, 2007.

[113] Kumi Ishii, Mary Madison Lyons, and Sabrina A Carr. Revisiting media richness theory for today and future. *Human Behavior and Emerging Technologies*, 1(2):124–131, 2019.

[114] Maurice Jakesch, Jeffrey T Hancock, and Mor Naaman. Human heuristics for ai-generated language are flawed. *Proceedings of the National Academy of Sciences*, 120(11):e2208839120, 2023.

[115] Chyng Yang Jang, Charles Steinfield, and Ben Pfaff. Supporting awareness among virtual teams in a web-based collaborative system: the teamscope system. *ACM Siggroup Bulletin*, 21(3):28–34, 2000.

[116] Sirkka L Jarvenpaa, Kathleen Knoll, and Dorothy E Leidner. Is anybody out there? antecedents of trust in global virtual teams. *Journal of management information systems*, 14(4):29–64, 1998.

[117] Sirkka L Jarvenpaa and Dorothy E Leidner. Communication and trust in global virtual teams. *Journal of computer-mediated communication*, 3(4):JCMC346, 1998.

[118] Carlos Jensen, Shelly D Farnham, Steven M Drucker, and Peter Kollock. The effect of communication modality on cooperation in online environments. In *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*, pages 470–477, 2000.

[119] Jinglu Jiang, Surinder Kahai, and Ming Yang. Who needs explanation and when? juggling explainable ai and user epistemic uncertainty. *International Journal of Human-Computer Studies*, page 102839, 2022.

[120] Anna Jobin, Marcello Ienca, and Effy Vayena. The global landscape of ai ethics guidelines. *Nature Machine Intelligence*, 1(9):389–399, 2019.

[121] J. David Johnson, William A. Donohue, Charles K. Atkin, and Sally Johnson. Differences Between Formal and Informal Communication Channels. *The Journal of Business Communication (1973)*, 31(2):111–122, April 1994. Publisher: SAGE Publications.

[122] Gareth R Jones and Jennifer M George. The experience and evolution of trust: Implications for cooperation and teamwork. *Academy of management review*, 23(3):531–546, 1998.

[123] Malte F Jung, Selma Šabanović, Friederike Eyssel, and Marlena Fraune. Robots in groups and teams. In *Companion of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*, pages 401–407, 2017.

[124] Hyunjin Kang and Ki Joon Kim. Does humanization or machinization make the iot persuasive? the effects of source orientation and social presence. *Computers in Human Behavior*, 129:107152, 2022.

[125] Sin-Hwa Kang and James H Watt. The impact of avatar realism and anonymity on effective communication via mobile devices. *Computers in Human Behavior*, 29(3):1169–1181, 2013.

[126] Neerja Kashive, Leena Powale, and Kshitij Kashive. Understanding user perception toward artificial intelligence (ai) enabled e-learning. *The International Journal of Information and Learning Technology*, 38(1):1–19, 2020.

[127] Alaa Khamis, Ahmed Hussein, and Ahmed Elmogy. Multi-robot task allocation: A review of the state-of-the-art. *Cooperative robots and sensor networks 2015*, pages 31–51, 2015.

[128] Sara Kiesler, Jane Siegel, and Timothy W McGuire. Social psychological aspects of computer-mediated communication. *American psychologist*, 39(10):1123, 1984.

[129] Molly Kilcullen, Jennifer Feitosa, and Eduardo Salas. Insights from the virtual team science: Rapid deployment during covid-19. *Human Factors*, page 0018720821991678, 2021.

[130] Youjeong Kim and S Shyam Sundar. Anthropomorphism of computers: Is it mindful or mindless? *Computers in Human Behavior*, 28(1):241–250, 2012.

[131] Young Ji Kim, David Engel, Anita Williams Woolley, Jeffrey Yu-Ting Lin, Naomi McArthur, and Thomas W Malone. What makes a strong team? using collective intelligence to predict team performance in league of legends. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*, pages 2316–2329, 2017.

[132] David L Kleinman and Daniel Serfaty. Team performance assessment in distributed decision making. In *Proceedings of the symposium on interactive networked simulation for training*, pages 22–27. University of Central Florida Orlando, FL, 1989.

[133] Ross A Knepper, Christoforos I Mavrogiannis, Julia Proft, and Claire Liang. Implicit communication in a joint action. In *Proceedings of the 2017 acm/ieee international conference on human-robot interaction*, pages 283–292, 2017.

[134] ES Kox, JH Kerstholt, TF Hueting, and PW De Vries. Trust repair in human-agent teams: the effectiveness of explanations and expressing regret. *Autonomous Agents and Multi-Agent Systems*, 35(2):1–20, 2021.

[135] Diana S Kusunoki and Aleksandra Sarcevic. Designing for temporal awareness: The role of temporality in time-critical medical teamwork. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*, pages 1465–1476, 2015.

[136] Yu-Feng Lan, Chun-Ling Hung, and Hung-Ju Hsu. Effects of guided writing strategies on students' writing attitudes based on media richness theory. *Turkish Online Journal of Educational Technology-TOJET*, 10(4):148–164, 2011.

[137] Claus W Langfred. Autonomy and performance in teams: The multilevel moderating effect of task interdependence. *Journal of management*, 31(4):513–529, 2005.

[138] Alex Leavitt, Brian C Keegan, and Joshua Clark. Ping to win? non-verbal communication and team performance in competitive online multiplayer games.

In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, pages 4337–4350, 2016.

[139] Charlotte Tsz-Sum Lee and Diane Marie Doran. The role of interpersonal relations in healthcare team communication and patient safety: a proposed model of interpersonal process in teamwork. *Canadian Journal of Nursing Research*, 49(2):75–93, 2017.

[140] John D Lee and Katrina A See. Trust in automation: Designing for appropriate reliance. *Human factors*, 46(1):50–80, 2004.

[141] Kwan Min Lee, Younbo Jung, Jaywoo Kim, and Sang Ryong Kim. Are physically embodied social agents better than disembodied social agents?: The effects of physical embodiment, tactile interaction, and people's loneliness in human–robot interaction. *International journal of human-computer studies*, 64(10):962–973, 2006.

[142] Min Hun Lee, Daniel P Siewiorek, Asim Smailagic, Alexandre Bernardino, and Sergi Bermúdez i Badia. A human-ai collaborative approach for clinical decision making on rehabilitation assessment. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, pages 1–14, 2021.

[143] Soyoung Lee, Charlotte Tang, Sun Young Park, and Yunan Chen. Loosely formed patient care teams: communication challenges and technology design. In *Proceedings of the ACM 2012 conference on computer supported cooperative work*, pages 867–876, 2012.

[144] Michael Leonard, Suzanne Graham, and Doug Bonacum. The human factor: the critical importance of effective teamwork and communication in providing safe care. *BMJ Quality & Safety*, 13(suppl 1):i85–i90, 2004.

[145] Tianyi Li, Mihaela Vorvoreanu, Derek DeBellis, and Saleema Amershi. Assessing human-ai interaction early through factorial surveys: A study on the guidelines for human-ai interaction. *ACM Transactions on Computer-Human Interaction*, 2022.

[146] Claire Liang, Julia Proft, Erik Andersen, and Ross A Knepper. Implicit communication of actionable information in human-ai teams. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, pages 1–13, 2019.

[147] Q Vera Liao, Daniel Gruen, and Sarah Miller. Questioning the ai: Informing design practices for explainable ai user experiences. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, pages 1–15, 2020.

[148] David Lyle Light Shields, Douglas E. Gardner, Brenda Jo Light Bredemeier, and Alan Bostro. The Relationship Between Leadership Behaviors and Group Cohesion in Team Sports. *The Journal of Psychology*, 131(2):196–210, March 1997. Publisher: Routledge _eprint: https://doi.org/10.1080/00223989709601964.

[149] Enrico Liscio. Axies: Identifying and Evaluating Context-Specific Values. page 11, 2021.

[150] Gustavo Lopez and Luis A Guerrero. Awareness supporting technologies used in collaborative systems: a systematic literature review. In *Proceedings of the 2017 ACM conference on computer supported cooperative work and social computing*, pages 808–820, 2017.

[151] Ed Love, Tara Ceranic Salinas, and Jeff D Rotman. The ethical standards of judgment questionnaire: Development and validation of independent measures of formalism and consequentialism. *Journal of business ethics*, 161(1):115–132, 2020.

[152] Robert P. Loweth, Shanna R. Daly, Amy Hortop, Elizabeth A. Strehl, and Kathleen H. Sienko. An in-depth investigation of student information gathering meetings with stakeholders and domain experts. *International Journal of Technology and Design Education*, 32(1):533–554, March 2022.

[153] Paul Benjamin Lowry, Tom L Roberts, Nicholas C Romano Jr, Paul D Cheney, and Ross T Hightower. The impact of group size and social presence on small-group communication: Does computer-mediated communication make a difference? *Small Group Research*, 37(6):631–661, 2006.

[154] Fabrice Lumineau. How contracts influence trust and distrust. *Journal of management*, 43(5):1553–1577, 2017.

[155] Brady D Lund and Ting Wang. Chatting about chatgpt: how may ai and gpt impact academia and libraries? *Library Hi Tech News*, 2023.

[156] Qianli Ma, Yaping Zhang, Wenti Xu, and Ronggang Zhou. Ask a further question or give a list? how should conversational agents reply to users' uncertain queries. *International Journal of Human–Computer Interaction*, pages 1–15, 2022.

[157] Jean Macmillan, Elliot E Entin, and Daniel Serfaty. A framework for understanding the relationship between team structure and the communication necessary for effective team cognition. *Team cognition: Process and performance at the inter-and intra-individual level*, pages 112–139, 2002.

[158] Jean MacMillan, Elliot E Entin, and Daniel Serfaty. Communication overhead: The hidden cost of team cognition. 2004.

[159] Christopher D Manning, Christopher D Manning, and Hinrich Schütze. *Foundations of statistical natural language processing.* MIT press, 1999.

[160] Shannon L Marlow, Christina N Lacerenza, Jensine Paoletti, C Shawn Burke, and Eduardo Salas. Does team communication represent a one-size-fits-all approach?: A meta-analysis of team communication and performance. *Organizational behavior and human decision processes*, 144:145–170, 2018.

[161] Shannon L Marlow, Christina N Lacerenza, and Eduardo Salas. Communication in virtual teams: A conceptual framework and research agenda. *Human Resource Management Review*, 27(4):575–589, 2017.

[162] Likoebe M Maruping and Ritu Agarwal. Managing team interpersonal processes through technology: A task-technology fit perspective. *Journal of applied psychology*, 89(6):975, 2004.

[163] Nikolaos Mavridis. A review of verbal and non-verbal human–robot interactive communication. *Robotics and Autonomous Systems*, 63:22–35, 2015.

[164] Winston Maxwell, Valérie Beaudouin, Isabelle Bloch, David Bounie, Stéphan Clémençon, Florence d'Alché Buc, James Eagan, Pavlo Mozharovskyi, and Jayneel Parekh. Identifying the'right'level of explanation in a given situation. In *Proceedings of the First International Workshop on New Foundations for Human-Centered AI (NeHuAI), Santiago de Compostella, Spain*, page 63, 2020.

[165] Martha L Maznevski and Katherine M Chudoba. Bridging space over time: Global virtual team dynamics and effectiveness. *Organization science*, 11(5):473–492, 2000.

[166] Pamela Mazzocato, Helena Hvitfeldt Forsberg, and Ulrica von Thiele Schwarz. Team behaviors in emergency care: a qualitative study using behavior analysis of what makes team work. *Scandinavian Journal of Trauma, Resuscitation and Emergency Medicine*, 19(1):1–8, 2011.

[167] Edward F McDonough III, Kenneth B Kahnb, and Gloria Barczaka. An investigation of the use of global, virtual, and colocated new product development teams. *Journal of Product Innovation Management: An international publication of the product development & management association*, 18(2):110–120, 2001.

[168] David McNaughton and Piers Rawling. On defending deontology. *Ratio*, 11(1):37–54, 1998.

[169] Nathan McNeese, Mustafa Demir, Erin Chiou, Nancy Cooke, and Giovanni Yanikian. Understanding the role of trust in human-autonomy teaming. In

Proceedings of the 52nd Hawaii international conference on system sciences, 2019.

[170] Nathan J McNeese, Mustafa Demir, Erin K Chiou, and Nancy J Cooke. Trust and team performance in human–autonomy teaming. *International Journal of Electronic Commerce*, 25(1):51–72, 2021.

[171] Nathan J McNeese, Mustafa Demir, Nancy J Cooke, and Christopher Myers. Teaming with a synthetic teammate: Insights into human-autonomy teaming. *Human factors*, 60(2):262–273, 2018.

[172] Nathan J McNeese, Mustafa Demir, Nancy J Cooke, and Manrong She. Team situation awareness and conflict: A study of human–machine teaming. *Journal of Cognitive Engineering and Decision Making*, 15(2-3):83–96, 2021.

[173] Nathan J McNeese and Madhu C Reddy. Articulating and understanding the development of a team mental model in a distributed medium. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, volume 59, pages 240–244. SAGE Publications Sage CA: Los Angeles, CA, 2015.

[174] Nathan J McNeese, Beau G Schelble, Lorenzo Barberis Canonico, and Mustafa Demir. Who/what is my teammate? team composition considerations in human–ai teaming. *IEEE Transactions on Human-Machine Systems*, 51(4):288–299, 2021.

[175] Nathan J. McNeese, Beau G. Schelble, Lorenzo Barberis Canonico, and Mustafa Demir. Who/What Is My Teammate? Team Composition Considerations in Human–AI Teaming. *IEEE Transactions on Human-Machine Systems*, 51(4):288–299, August 2021. Conference Name: IEEE Transactions on Human-Machine Systems.

[176] Joseph E Mercado, Michael A Rupp, Jessie YC Chen, Michael J Barnes, Daniel Barber, and Katelyn Procci. Intelligent agent transparency in human–agent teaming for multi-uxv management. *Human factors*, 58(3):401–415, 2016.

[177] Tim R Merritt, Kian Boon Tan, Christopher Ong, Aswin Thomas, Teong Leong Chuah, and Kevin McGee. Are artificial team-mates scapegoats in computer games. In *Proceedings of the ACM 2011 conference on Computer supported cooperative work*, pages 685–688, 2011.

[178] Peter Mills, Julia Neily, and Ed Dunn. Teamwork and communication in surgical teams: implications for patient safety. *Journal of the American College of Surgeons*, 206(1):107–112, 2008.

[179] Shaila M Miranda and Carol S Saunders. The social construction of meaning: An alternative perspective on information sharing. *Information systems research*, 14(1):87–106, 2003.

[180] Sarah Morrison-Smith and Jaime Ruiz. Challenges and barriers in virtual teams: a literature review. *SN Applied Sciences*, 2(6):1–33, 2020.

[181] Yi Mou and Kun Xu. The media inequality: Comparing the initial human-human and human-ai social interactions. *Computers in Human Behavior*, 72:432–440, 2017.

[182] Stefan Münzer and Torsten Holmer. Bridging the gap between media synchronicity and task performance: Effects of media characteristics on process variables and task performance indicators in an information pooling task. *Communication Research*, 36(1):76–103, 2009.

[183] Geoff Musick, Thomas A O'Neill, Beau G Schelble, Nathan J McNeese, and Jonn B Henke. What happens when humans believe their teammate is an ai? an investigation into humans teaming with autonomy. *Computers in Human Behavior*, 122:106852, 2021.

[184] Geoff Musick, Rui Zhang, Nathan J McNeese, Guo Freeman, and Anurata Prabha Hridi. Leveling up teamwork in esports: Understanding team cognition in a dynamic virtual environment. *Proceedings of the ACM on Human-Computer Interaction*, 5(CSCW1):1–30, 2021.

[185] Clifford Nass, Jonathan Steuer, and Ellen R Tauber. Computers are social actors. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 72–78, 1994.

[186] Andrew Neal, Gillian Yeo, Annette Koy, and Tania Xiao. Predicting the form and direction of work role performance from the big 5 model of personality traits. *Journal of Organizational Behavior*, 33(2):175–192, 2012.

[187] Tuomas Niinimäki, Arttu Piri, Casper Lassenius, and Maria Paasivaara. Reflecting the choice and usage of communication tools in global software development projects with media synchronicity theory. *Journal of Software: Evolution and Process*, 24(6):677–692, 2012.

[188] Stefanos Nikolaidis, Minae Kwon, Jodi Forlizzi, and Siddhartha Srinivasa. Planning with verbal communication for human-robot collaboration. *ACM Transactions on Human-Robot Interaction (THRI)*, 7(3):1–21, 2018.

[189] Tatsuya Nomura, Takayuki Kanda, and Tomohiro Suzuki. Experimental investigation into influence of negative attitudes toward robots on human–robot interaction. *Ai & Society*, 20(2):138–150, 2006.

[190] Tatsuya Nomura, Tomohiro Suzuki, Takayuki Kanda, and Kensuke Kato. Measurement of negative attitudes toward robots. *Interaction Studies*, 7(3):437–454, 2006.

[191] Mahsan Nourani, Donald R Honeycutt, Jeremy E Block, Chiradeep Roy, Tahrima Rahman, Eric D Ragan, and Vibhav Gogate. Investigating the importance of first impressions and explainable ai with interactive video analysis. In *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems*, pages 1–8, 2020.

[192] Catherine S Oh, Jeremy N Bailenson, and Gregory F Welch. A systematic review of social presence: Definition, antecedents, and implications. *Frontiers in Robotics and AI*, page 114, 2018.

[193] Changhoon Oh, Taeyoung Lee, Yoojung Kim, SoHyun Park, Saebom Kwon, and Bongwon Suh. Us vs. them: Understanding artificial intelligence technophobia over the google deepmind challenge match. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, pages 2523–2534, 2017.

[194] Changhoon Oh, Jungwoo Song, Jinhan Choi, Seonghyeon Kim, Sungwoo Lee, and Bongwon Suh. I lead, you help but only with enough details: Understanding user experience of co-creation with artificial intelligence. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, pages 1–13, 2018.

[195] Kazuo Okamura and Seiji Yamada. Adaptive trust calibration for human-ai collaboration. *Plos one*, 15(2):e0229132, 2020.

[196] Gary M Olson and Judith S Olson. Human-computer interaction: Psychological aspects of the human use of computing. *Annual review of psychology*, 54(1):491–516, 2003.

[197] Christopher Ong, Kevin McGee, and Teong Leong Chuah. Closing the human-ai team-mate gap: how changes to displayed information impact player behavior towards computer teammates. In *Proceedings of the 24th Australian Computer-Human Interaction Conference*, pages 433–439, 2012.

[198] Judith Orasanu. Shared mental models and crew decision making. In *12th Annual Conf. CSS Pod*, pages 1066–1066. Psychology Press, 1990.

[199] Randall L. Oser, G. A. McCallum, Eduardo Salas, and Ben B. Morgan. Toward a Definition of Teamwork: An Analysis of Critical Team Behaviors. Technical report, NAVAL TRAINING SYSTEMS CENTER ORLANDO FL, March 1989. Section: Technical Reports.

[200] Thomas O'Neill, Nathan McNeese, Amy Barron, and Beau Schelble. Human–autonomy teaming: A review and analysis of the empirical literature. *Human factors*, page 0018720820960865, 2020.

[201] Atanu Pal, Roshan Lal, and Frank Frizelle. Aviation-based teamwork skills work for surgeons: time for an 'aviation bundle'? *ANZ Journal of Surgery*, 88(12):1231–1235, 2018. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/ans.14892.

[202] Stefan Palan and Christian Schitter. Prolific. ac—a subject pool for online experiments. *Journal of Behavioral and Experimental Finance*, 17:22–27, 2018.

[203] Raja Parasuraman, Thomas B Sheridan, and Christopher D Wickens. A model for types and levels of human interaction with automation. *IEEE Transactions on systems, man, and cybernetics-Part A: Systems and Humans*, 30(3):286–297, 2000.

[204] Eun Kyung Park and S Shyam Sundar. Can synchronicity and visual modality enhance social presence in mobile messaging? *Computers in Human Behavior*, 45:121–128, 2015.

[205] Sun Young Park, Pei-Yi Kuo, Andrea Barbarin, Elizabeth Kaziunas, Astrid Chow, Karandeep Singh, Lauren Wilcox, and Walter S Lasecki. Identifying challenges and opportunities in human-ai collaboration in healthcare. In *Conference Companion Publication of the 2019 on Computer Supported Cooperative Work and Social Computing*, pages 506–510, 2019.

[206] Avi Parush, Chelsea Kramer, Tara Foster-Hunt, Kathryn Momtahan, Aren Hunter, and Benjamin Sohmer. Communication and team situation awareness in the or: Implications for augmentative information display. *Journal of biomedical informatics*, 44(3):477–485, 2011.

[207] Deepika Phutela. The importance of non-verbal communication. *IUP Journal of Soft Skills*, 9(4):43, 2015.

[208] Rosalind W. Picard, Elias Vyzas, and Jennifer Healey. Toward machine emotional intelligence: Analysis of affective physiological state. *IEEE transactions on pattern analysis and machine intelligence*, 23(10):1175–1191, 2001.

[209] Martin Ragot, Nicolas Martin, and Salomé Cojean. Ai-generated vs. human artworks. a perception bias towards artificial intelligence? In *Extended abstracts of the 2020 CHI conference on human factors in computing systems*, pages 1–10, 2020.

[210] Summer Rebensky, Kendall Carmody, Cherrise Ficke, Daniel Nguyen, Meredith Carroll, Jessica Wildman, and Amanda Thayer. Whoops! something went wrong: Errors, trust, and trust repair strategies in human agent teaming. In *International Conference on Human-Computer Interaction*, pages 95–106. Springer, 2021.

[211] Amy Rechkemmer and Ming Yin. When confidence meets accuracy: Exploring the effects of multiple performance indicators on trust in machine learning models. In *CHI Conference on Human Factors in Computing Systems*, pages 1–14, 2022.

[212] Joan R Rentsch and Richard J Klimoski. Why do 'great minds' think alike?: Antecedents of team member schema agreement. *Journal of Organizational Behavior: The International Journal of Industrial, Occupational and Organizational Psychology and Behavior*, 22(2):107–120, 2001.

[213] Ramón Rico and Susan G Cohen. Effects of task interdependence and type of communication on performance in virtual teams. *Journal of Managerial Psychology*, 2005.

[214] Ramón Rico, Miriam Sánchez-Manzanares, Francisco Gil, and Cristina Gibson. Team implicit coordination processes: A team knowledge–based approach. *Academy of management review*, 33(1):163–184, 2008.

[215] Paul Robinette, Wenchen Li, Robert Allen, Ayanna M Howard, and Alan R Wagner. Overtrust of robots in emergency evacuation scenarios. In *2016 11th ACM/IEEE international conference on human-robot interaction (HRI)*, pages 101–108. IEEE, 2016.

[216] Francesca Rossi. Building trust in artificial intelligence. *Journal of international affairs*, 72(1):127–134, 2018.

[217] Karina A Roundtree, Michael A Goodrich, and Julie A Adams. Transparency: Transitioning from human–machine systems to human-swarm systems. *Journal of Cognitive Engineering and Decision Making*, 13(3):171–195, 2019.

[218] Katherine E. Rowan. Goals, obstacles, and strategies in risk communication: A problem-solving approach to improving communication about risks. *Journal of Applied Communication Research*, 19(4):300–329, November 1991. Publisher: Routledge _eprint: https://doi.org/10.1080/00909889109365311.

[219] Thomas E Ruggiero. Uses and gratifications theory in the 21st century. *Mass communication & society*, 3(1):3–37, 2000.

[220] Nauman Saeed, Yun Yang, and Suku Sinnappan. Effect of media richness on user acceptance of blogs and podcasts. In *Proceedings of the fifteenth annual conference on Innovation and technology in computer science education*, pages 137–141, 2010.

[221] Eduardo Salas, Janis A Cannon-Bowers, and Joan Hall Johnston. How can you turn a team of experts into an expert team?: Emerging training strategies. *Naturalistic decision making*, 1:359–370, 1997.

[222] Eduardo Salas, Nancy J Cooke, and Michael A Rosen. On teams, teamwork, and team performance: Discoveries and developments. *Human factors*, 50(3):540–547, 2008.

[223] Eduardo Salas, Deborah DiazGranados, Cameron Klein, C Shawn Burke, Kevin C Stagl, Gerald F Goodwin, and Stanley M Halpin. Does team training improve team performance? a meta-analysis. *Human factors*, 50(6):903–933, 2008.

[224] Eduardo Salas, Terry L Dickinson, Sharolyn A Converse, and Scott I Tannenbaum. Toward an understanding of team performance and training. 1992.

[225] Eduardo Salas, Carolyn Prince, David P Baker, and Lisa Shrestha. Situation awareness in team performance: Implications for measurement and training. *Situational awareness*, pages 63–76, 2017.

[226] Eduardo Salas, Dana E Sims, and C Shawn Burke. Is there a "big five" in teamwork? *Small group research*, 36(5):555–599, 2005.

[227] Prince C. Baker P.D. Shrestha L. Salas, E. Situation awareness in team performance: implications for measurement and training. *Human factors*, 37(1):123–136, 1995.

[228] Arleen Salles, Kathinka Evers, and Michele Farisco. Anthropomorphism in ai. *AJOB neuroscience*, 11(2):88–95, 2020.

[229] Saonee Sarker, Manju Ahuja, Suprateek Sarker, and Sarah Kirkeby. The role of communication and trust in global virtual teams: A social network perspective. *Journal of Management Information Systems*, 28(1):273–310, 2011.

[230] FE Satterthwaite. Random balance experimentation. *Technometrics*, 1(2):111–137, 1959.

[231] Kristin E Schaefer, Edward R Straub, Jessie YC Chen, Joe Putney, and Arthur W Evans III. Communicating intent to develop shared situation awareness and engender trust in human-agent teams. *Cognitive Systems Research*, 46:26–39, 2017.

[232] Beau G Schelble, Christopher Flathmann, Nathan J McNeese, Guo Freeman, and Rohit Mallick. Let's think together! assessing shared mental models, performance, and trust in human-agent teams. *Proceedings of the ACM on Human-Computer Interaction*, 6(GROUP):1–29, 2022.

[233] Beau G Schelble, Christopher Flathmann, Nathan J McNeese, Thomas O'Neill, Richard Pak, and Moses Namara. Investigating the effects of perceived teammate artificiality on human performance and cognition. *International Journal of Human–Computer Interaction*, pages 1–16, 2022.

[234] Beau G Schelble, Jeremy Lopez, Claire Textor, Rui Zhang, Nathan J McNeese, Richard Pak, and Guo Freeman. Towards ethical ai: Empirically investigating dimensions of ai ethics, trust repair, and performance in human-ai teaming. *Human Factors*, page 00187208221116952, 2022.

[235] Damaris Schmid, Dario Staehelin, Andreas Bucher, Mateusz Dolata, and Gerhard Schwabe. Does social presence increase perceived competence? evaluating conversational agents in advice giving through a video-based survey. *Proceedings of the ACM on Human-Computer Interaction*, 6(GROUP):1–22, 2022.

[236] Charles PR Scott and Jessica L Wildman. Culture, communication, and conflict: A review of the global virtual team literature. *Leading global teams*, pages 13–32, 2015.

[237] Nick Sevdalis, Shabnam Undre, James McDermott, Jasdeep Giddie, Lila Diner, and Gillian Smith. Impact of intraoperative distractions on patient safety: a prospective descriptive study using validated instruments. *World journal of surgery*, 38(4):751–758, 2014.

[238] Kevin Sexton, Amanda Johnson, Amanda Gotsch, Ahmed A Hussein, Lora Cavuoto, and Khurshid A Guru. Anticipation, teamwork and cognitive load: chasing efficiency during robot-assisted surgery. *BMJ quality & safety*, 27(2):148–154, 2018.

[239] Chirag Shah. *Social Information Seeking.* Springer, 2017.

[240] Nicholas Shea. Moving beyond content-specific computation in artificial neural networks. *Mind & Language*, 2021.

[241] Donghee Shin. The effects of explainability and causability on perception, trust, and acceptance: Implications for explainable ai. *International Journal of Human-Computer Studies*, 146:102551, 2021.

[242] R Jay Shively, Joel Lachter, Summer L Brandt, Michael Matessa, Vernol Battiste, and Walter W Johnson. Why human-autonomy teaming? In *International*

conference on applied human factors and ergonomics, pages 3–11. Springer, 2017.

[243] John Short, Ederyn Williams, and Bruce Christie. *The social psychology of telecommunications*. Toronto; London; New York: Wiley, 1976.

[244] Yufei Shu and Kazuo Furuta. An inference method of team situation awareness based on mutual awareness. *Cognition, Technology & Work*, 7:272–287, 2005.

[245] Cornelia Sindermann, Haibo Yang, Jon D Elhai, Shixin Yang, Ling Quan, Mei Li, and Christian Montag. Acceptance and fear of artificial intelligence: Associations with personality in a german and a chinese sample. *Discover Psychology*, 2(1):1–12, 2021.

[246] Ho Chit Siu, Jaime Peña, Edenna Chen, Yutai Zhou, Victor Lopez, Kyle Palko, Kimberlee Chang, and Ross Allen. Evaluation of human-ai teams for learned and rule-based agents in hanabi. *Advances in Neural Information Processing Systems*, 34, 2021.

[247] Kip Smith and Peter A Hancock. Situation awareness is adaptive, externally directed consciousness. *Human factors*, 37(1):137–148, 1995.

[248] Alex W Stedmon, Sarah Sharples, Robert Littlewood, Gemma Cox, Harshada Patel, and John R Wilson. Datalink in air traffic management: Human factors issues in communications. *Applied ergonomics*, 38(4):473–480, 2007.

[249] Stanley S Stevens. Introduction: A definition of communication. *The Journal of the Acoustical Society of America*, 22(6):689–690, 1950.

[250] Renee J Stout, Janis A Cannon-Bowers, and Eduardo Salas. The role of shared mental models in developing team situational awareness: Implications for training. In *Situational awareness*, pages 287–318. Routledge, 2017.

[251] Kimberly Stowers, Lisa L Brady, Christopher MacLellan, Ryan Wohleber, and Eduardo Salas. Improving teamwork competencies in human-machine teams: Perspectives from team science. *Frontiers in Psychology*, 12:590290, 2021.

[252] Susan G Straus. Getting a clue: The effects of communication media and information distribution on participation and performance in computer-mediated and face-to-face groups. *Small group research*, 27(1):115–142, 1996.

[253] Kil Soo Suh. Impact of communication medium on task performance and satisfaction: an examination of media-richness theory. *Information & Management*, 35(5):295–312, 1999.

[254] Alane Suhr, Claudia Yan, Jacob Schluger, Stanley Yu, Hadi Khader, Marwa Mouallem, Iris Zhang, and Yoav Artzi. Executing instructions in situated collaborative interactions. *arXiv preprint arXiv:1910.03655*, 2019.

[255] Michael J. L. Sullivan, Pascal Thibault, André Savard, Richard Catchlove, John Kozey, and William D. Stanish. The influence of communication goals and physical demands on different dimensions of pain behavior. *Pain*, 125(3):270–277, December 2006.

[256] Katia Sycara, Dana Hughes, Huao Li, Michael Lewis, and Nina Lauharatanahirun. Adaptation in human-autonomy teamwork. In *2020 IEEE International Conference on Human-Machine Systems (ICHMS)*, pages 1–4. IEEE, 2020.

[257] Dag Sverre Syrdal, Kerstin Dautenhahn, Kheng Lee Koay, and Michael L Walters. The negative attitudes towards robots scale and reactions to robot behaviour in a live human-robot interaction study. *Adaptive and emergent behaviour and complex systems*, 2009.

[258] Fangcheng Tang, Jifeng Mu, and Ellen Thomas. Who knows what in npd teams: Communication context, mode, and task contingencies. *Journal of Product Innovation Management*, 32(3):404–423, 2015.

[259] Claire Textor, Rui Zhang, Jeremy Lopez, Beau G Schelble, Nathan J McNeese, Guo Freeman, Richard Pak, Chad Tossell, and Ewart J de Visser. Exploring the relationship between ethics and trust in human–artificial intelligence teaming: A mixed methods approach. *Journal of Cognitive Engineering and Decision Making*, page 15553434221113964, 2022.

[260] AM Thomsen. Norobot's perfect: trust repair in the face of agent error. how do individual factors influence trust development in human-agent teams? B.S. thesis, University of Twente, 2022.

[261] Crispin Thurlow, Laura Lengel, and Alice Tomic. *Computer mediated communication*. Sage, 2004.

[262] Myrthe L Tielman. Trust should correspond to trustworthiness: a formalization of appropriate mutual trust in human-agent teams. 2021.

[263] Judith Tiferes and Ann M Bisantz. The impact of team characteristics and context on team communication: An integrative literature review. *Applied ergonomics*, 68:146–159, 2018.

[264] William J Tippett, Jang-Han Lee, Richard Mraz, Konstantine K Zakzanis, Peter J Snyder, Sandra E Black, and Simon J Graham. Convergent validity and

sex differences in healthy elderly adults for performance on 3d virtual reality navigation learning and 2d hidden maze tasks. *CyberPsychology & Behavior*, 12(2):169–174, 2009.

[265] Güliz Tokadlı, Michael C Dorneich, and Michael Matessa. Evaluation of playbook delegation approach in human-autonomy teaming for single pilot operations. *International Journal of Human–Computer Interaction*, 37(7):703–716, 2021.

[266] Güliz Tokadll, Michael C Dorneich, and Tomas Gonzalez-Torres. Preliminary guidelines for human-agent teams in space operations beyond low-earth orbit. In *2018 IEEE/AIAA 37th Digital Avionics Systems Conference (DASC)*, pages 1–9. IEEE, 2018.

[267] Suzanne Tolmeijer, Ujwal Gadiraju, Ramya Ghantasala, Akshit Gupta, and Abraham Bernstein. Second chance for a first impression? trust development in intelligent system interaction. In *Proceedings of the 29th ACM Conference on user modeling, adaptation and personalization*, pages 77–87, 2021.

[268] Zachary O Toups and Andruid Kerne. Implicit coordination in firefighting practice: design implications for teaching fire emergency responders. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 707–716, 2007.

[269] Hungwei Tseng, Betty Morris, and Yingqi Tang. The importance of teamwork trust, social presence, and cognitive presence in an online collaborative learning environment. In *Society for information technology & teacher education international conference*, pages 538–541. Association for the Advancement of Computing in Education (AACE), 2015.

[270] Iis P Tussyadiah, Florian J Zach, and Jianxi Wang. Do travelers trust intelligent service robots? *Annals of Tourism Research*, 81:102886, 2020.

[271] Anna-Sophie Ulfert and Eleni Georganta. A model of team trust in human-agent teams. In *Companion publication of the 2020 international conference on multimodal interaction*, pages 171–176, 2020.

[272] Natasha Veltri, Hanna Krasnova, Annika Baumann, and Neena Kalayamthanam. Gender differences in online gaming: A literature review. 2014.

[273] Viswanath Venkatesh and Hillol Bala. Technology acceptance model 3 and a research agenda on interventions. *Decision sciences*, 39(2):273–315, 2008.

[274] Greg Wadley, Martin R Gibbs, and Nicolas Ducheneaut. You can be too rich: Mediated communication in a virtual world. In *Proceedings of the 21st Annual Conference of the Australian Computer-Human Interaction Special Interest Group: Design: Open 24/7*, pages 49–56, 2009.

[275] Alan R Wagner, Jason Borenstein, and Ayanna Howard. Overtrust in the robotic age. *Communications of the ACM*, 61(9):22–24, 2018.

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

[277] James C Walliser, Patrick R Mead, and Tyler H Shaw. The perception of teamwork with an autonomous agent enhances affect and performance outcomes. In *Proceedings of the human factors and ergonomics society annual meeting*, volume 61, pages 231–235. SAGE Publications Sage CA: Los Angeles, CA, 2017.

[278] Joseph B Walther. Interpersonal effects in computer-mediated interaction: A relational perspective. *Communication research*, 19(1):52–90, 1992.

[279] Dakuo Wang, Elizabeth Churchill, Pattie Maes, Xiangmin Fan, Ben Shneiderman, Yuanchun Shi, and Qianying Wang. From human-human collaboration to human-ai collaboration: Designing ai systems that can work together with people. In *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems Extended Abstracts*, pages 1–6, 2020.

[280] Dakuo Wang, Justin D Weisz, Michael Muller, Parikshit Ram, Werner Geyer, Casey Dugan, Yla Tausczik, Horst Samulowitz, and Alexander Gray. Human-ai collaboration in data science: Exploring data scientists' perceptions of automated ai. *Proceedings of the ACM on Human-Computer Interaction*, 3(CSCW):1–24, 2019.

[281] Qiaosi Wang, Koustuv Saha, Eric Gregori, David Joyner, and Ashok Goel. Towards mutual theory of mind in human-ai interaction: How language reflects what students perceive about a virtual teaching assistant. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, pages 1–14, 2021.

[282] Xinru Wang and Ming Yin. Effects of explanations in ai-assisted decision making: Principles and comparisons. *ACM Transactions on Interactive Intelligent Systems (TiiS)*, 2022.

[283] Merrill Warkentin and Peggy M Beranek. Training to improve virtual team communication. *Information systems journal*, 9(4):271–289, 1999.

[284] Kelly D Wason, Michael J Polonsky, and Michael R Hyman. Designing vignette studies in marketing. *Australasian Marketing Journal*, 10(3):41–58, 2002.

[285] Katharina Weitz, Dominik Schiller, Ruben Schlagowski, Tobias Huber, and Elisabeth André. " do you trust me?" increasing user-trust by integrating virtual agents in explainable ai interaction design. In *Proceedings of the 19th ACM International Conference on Intelligent Virtual Agents*, pages 7–9, 2019.

[286] Peter H. Welch, Jo R. Aldous, and Jon Foster. CSP Networking for Java (JCSP.net). In Peter M. A. Sloot, Alfons G. Hoekstra, C. J. Kenneth Tan, and Jack J. Dongarra, editors, *Computational Science — ICCS 2002*, Lecture Notes in Computer Science, pages 695–708, Berlin, Heidelberg, 2002. Springer.

[287] Mark E Whiting, Irena Gao, Michelle Xing, N'godjigui Junior Diarrassouba, Tonya Nguyen, and Michael S Bernstein. Parallel worlds: Repeated initializations of the same team to improve team viability. *proceedings of the ACM on Human-Computer Interaction*, 4(CSCW1):1–22, 2020.

[288] Randi Williams, Hae Won Park, and Cynthia Breazeal. A is for artificial intelligence: The impact of artificial intelligence activities on young children's perceptions of robots. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, pages 1–11, 2019.

[289] Ming Yu Claudia Wong, Pak-Kwong Chung, Kailing Ou, and Ka-Man Leung. Perception of hong kong teenagers and young adults on esports participation: A qualitative study using theory of planned behavior. *Frontiers in Psychology*, 12:650000, 2021.

[290] Julia L Wright, Jessie Y Chen, Stephanie A Quinn, and Michael J Barnes. The effects of level of autonomy on human-agent teaming for multi-robot control and local security maintenance. Technical report, ARMY RESEARCH LAB ABERDEEN PROVING GROUND MD, 2013.

[291] Melanie C Wright and Mica R Endsley. Building shared situation awareness in healthcare settings. In *Improving healthcare team communication*, pages 97–114. CRC Press, 2017.

[292] Melanie C Wright and David B Kaber. Team coordination and strategies under automation. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, volume 47, pages 553–557. SAGE Publications Sage CA: Los Angeles, CA, 2003.

[293] Chengke Wu, Xiao Li, Yuanjun Guo, Jun Wang, Zengle Ren, Meng Wang, and Zhile Yang. Natural language processing for smart construction: Current status and future directions. *Automation in Construction*, 134:104059, 2022.

[294] Jason Wuertz, Sultan A Alharthi, William A Hamilton, Scott Bateman, Carl Gutwin, Anthony Tang, Zachary Toups, and Jessica Hammer. A design framework for awareness cues in distributed multiplayer games. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, pages 1–14, 2018.

[295] Yan Xiao, Colin F Mackenzie, Rhona Patey, and LOTAS group. Team coordination and breakdowns in a real-life stressful environment. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, volume 42, pages 186–190. SAGE Publications Sage CA: Los Angeles, CA, 1998.

[296] Jin Xu. Overtrust of robots in high-risk scenarios. In *Proceedings of the 2018 AAAI/ACM Conference on AI, Ethics, and Society*, pages 390–391, 2018.

[297] Wei Ye, Francesco Bullo, Noah Friedkin, and Ambuj K Singh. Modeling human-ai team decision making. *arXiv preprint arXiv:2201.02759*, 2022.

[298] Ming Yin, Jennifer Wortman Vaughan, and Hanna Wallach. Understanding the effect of accuracy on trust in machine learning models. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, pages 1–12, 2019.

[299] Zahra Zahedi and Subbarao Kambhampati. Human-ai symbiosis: A survey of current approaches. *arXiv preprint arXiv:2103.09990*, 2021.

[300] Baobao Zhang and Allan Dafoe. Artificial intelligence: American attitudes and trends. *Available at SSRN 3312874*, 2019.

[301] Rui Zhang, Nathan J McNeese, Guo Freeman, and Geoff Musick. "an ideal human" expectations of ai teammates in human-ai teaming. *Proceedings of the ACM on Human-Computer Interaction*, 4(CSCW3):1–25, 2021.

[302] Yunfeng Zhang, Q Vera Liao, and Rachel KE Bellamy. Effect of confidence and explanation on accuracy and trust calibration in ai-assisted decision making. In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, pages 295–305, 2020.

[303] Zhan Zhang, Aleksandra Sarcevic, Maria Yala, and Randall S Burd. Informing digital cognitive aids design for emergency medical work by understanding paper checklist use. In *Proceedings of the 18th international conference on supporting group work*, pages 204–214, 2014.

[304] Zikai Zhou and Pilar Pazos. How teams perform under emergent and dynamic situations: the roles of mental models and backup behaviors. *Team Performance Management: An International Journal*, 2020.

[305] Emma M van Zoelen, Anita Cremers, Frank PM Dignum, Jurriaan van Diggelen, and Marieke M Peeters. Learning to communicate proactively in human-agent teaming. In *International Conference on Practical Applications of Agents and Multi-Agent Systems*, pages 238–249. Springer, 2020.

[306] Roxanne Zolin, Pamela J Hinds, Renate Fruchter, and Raymond E Levitt. Interpersonal trust in cross-functional, geographically distributed work: A longitudinal study. *Information and organization*, 14(1):1–26, 2004.