Human–Computer Interaction

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/hhci20

Understanding the impact and design of AI teammate etiquette

Christopher Flathmann, Nathan J. McNeese, Beau Schelble, Bart Knijnenburg & Guo Freeman

To cite this article: Christopher Flathmann, Nathan J. McNeese, Beau Schelble, Bart Knijnenburg & Guo Freeman (2023): Understanding the impact and design of AI teammate etiquette, Human–Computer Interaction, DOI: 10.1080/07370024.2023.2189595

To link to this article: https://doi.org/10.1080/07370024.2023.2189595

Published online: 24 Mar 2023.

HUMAN–COMPUTER INTERACTION
https://doi.org/10.1080/07370024.2023.2189595

Understanding the impact and design of AI teammate etiquette

Christopher Flathmann, Nathan J. McNeese, Beau Schelble, Bart Knijnenburg, and Guo Freeman

Human-Centered Computing, Clemson University College of Engineering Computing and Applied Sciences, Clemson, South Carolina, USA

ABSTRACT

Technical and practical advancements in Artificial Intelligence (AI) have led to AI teammates working alongside humans in an area known as human-agent teaming. While critical past research has shown the benefit to trust driven by the incorporation of interaction rules and structures (i.e. etiquette) in both AI tools and robotic teammates, research has yet to explicitly examine etiquette for digital AI teammates. Given the historic importance of trust within human-agent teams, the identification of etiquette's impact within said teams should be paramount. Thus, this study empirically evaluates the impact of AI teammate etiquette through a mixed-methods study that compares AI teammates that either adhere to or ignore traditional etiquette standards for machine systems. The quantitative results show that traditional etiquette adherence leads to greater trust, perceived performance of the AI, and perceived performance of the team as a whole. However, qualitative results reveal that not all traditional etiquette behaviors have universal appeal due to the presence of individual differences. This research provides the first empirical and explicit exploration of etiquette within human-agent teams, and the results of this study should be used further design specific etiquette behaviors for AI teammates.

ARTICLE HISTORY
Received 6 July 2022
Revised 23 February 2023
Accepted 2 March 2023

KEYWORDS
Human-agent teamwork; AI etiquette; AI trust; AI teammate design

1. Introduction

Artificial Intelligence (AI) has rapidly progressed over the past few years due to the unique benefits posed by AI that are complimentary to humans, creating a highly performative collaboration (Grover et al., 2020). The continuous development of AI has led to an increasingly important research agenda in Human-Computer Interaction (HCI) on human-agent teams, (Inkpen et al., 2019). Human-agent teams consist of human teammates alongside artificial teammates (usually built with AI) working interdependently to achieve shared goals (N. J. McNeese et al., 2018; O'Neill et al., 2022). In pursuit of these efforts, past research has outlined the ways AI and machine systems should structure their collaborations with humans, which is referred to as the concept of machine etiquette (Parasuraman & Miller, 2004). However, while existing standards for AI agents could reasonably benefit human-agent teams, research has yet to explore the effects of these etiquette standards in human-agent teams. Thus, this article explores how AI teammates that either adhere to or reject traditional AI agent etiquette standards affect their human-agent teams.

Within HCI, etiquette plays a critical role in the day-to-day design of AI agents. Commonly, this etiquette manifests itself through a variety of behavioral designs, such as the inclusion of turn-taking behaviors (Miller, 2011), the reinforcement of an AI's role (Dautenhahn et al., 2006),

CONTACT Christopher Flathmann cflathm@clemson.edu Human-Centered Computing, Clemson University College of Engineering Computing and Applied Sciences, 125 McAdams Hall, Clemson SC, USA

and even the utilization of polite language during conversation (Schiaffino & Amandi, 2006). When incorporated, these behaviors have demonstrable impacts on human-agent interactions. For instance, machine etiquette can actually shape and influence not only the way AI agents behave but also the humans that interact with them (Meyer et al., 2016). Most importantly, the utilization of etiquette within human-agent interaction has repeatedly been shown to empirically improve the trust humans form for AI and machine systems (Atoyan et al., 2006; Parasuraman & Riley, 1997). Within human-agent teaming, the prioritization and creation of AI teammate trust are one of the most important considerations of researchers and designers (N. McNeese et al., 2019). However, despite the numerous connections between trust and etiquette and trust and human-agent teaming, research has yet to empirically and explicitly examine the impact of etiquette in human-agent teams that use digital AI teammates, which is the gap this research works to close.

In closing the aforementioned gap, it is important to note that the formation of etiquette is a complex moving target that differs based on the role of the machine, context, and medium (Miller, 2002). This is highly evident when looking at the field of human-robot teaming, as traditional etiquette standards are not always ideal. For instance, while turn-taking behaviors can be helpful within human-robot teams (Skantze, 2021), robotic teammates that forgo these traditional standards and act more assertive during negotiations can actually see improvements in performance and perception (Babel et al., 2021). Additionally, changes in context also merit different considerations within etiquette, which is evidenced by human-robot teaming's concern with the physical presence of robotic systems (Karreman et al., 2014), which is not a needed concern within a digital context. Given the above, it is important to approach this gap closure not from the perspective that traditional etiquette will be a global optimum for human-agent teams. Rather, it's important to approach the integration of etiquette in human-agent teams from the perspective of identifying which etiquette behaviors are beneficial and which ones need to be changed to improve the perception of AI teammates.

Utilizing the above considerations, a mixed methods experiment was conducted that collected and analyzed human interaction with AI teammates that were either designed to adhere to or ignore traditional machine etiquette standards, such as turn-taking, role assignment, and polite language. Specifically, given their identified importance to human-agent teaming and etiquette, two intra-team dynamics are the focus of the current manuscript's quantitative analysis: trust (N. McNeese et al., 2019); and perceived performance (N. J. McNeese et al., 2018). Alternatively, the qualitative component of this article explores the nuanced perceptions humans had toward AI teammate etiquette, which is critical in understanding not just *if* etiquette impacts social perception but also *how* that etiquette can be better designed to within a human-agent team. Thus, the following research questions are explored:

- RQ1 - How and why does the etiquette of an AI teammate affect the formation of human perceptions of team trust and perceived performance?

The main contributions of this study to the field of HCI are three-fold. First, this research provides the first empirical and explicit exploration of etiquette within human-agent teams that use digital AI teammates, which is an understudied topic in HCI. Second, this research utilizes a mixture of qualitative and quantitative analysis to develop design recommendations to empower designers to update their understanding of agent etiquette to accommodate AI teammates. Lastly, this study showcases the criticality of considering the perceptions of individuals when designing AI teammate etiquette that does not simply assume the existence of a global optimum. Based on the evidence presented in this research, a discussion is had on how research should use and advance the results of this study to design AI teammate etiquette, how AI teammates should be monitored for perceived toxic behaviors, and specific design recommendations for the creation of AI etiquette.

2. Background

Human interaction with digital AI will continue to grow both in frequency and necessity as AI develops in the coming years. A large portion of these interactions will take place within the context of "teamwork." Although human-agent interaction is a highly studied field in HCI, with etiquette playing a significant role in this field, the field of teamwork has distinct requirements and past research that merits separate reviews. Thus, this article will first provide an explicit review of the concepts of HCI related to AI agent design as well as how etiquette has been studied in that domain. Secondly, this article will provide a review of how AI teammates are designed for digital human-agent teams and how research in physical human-robot teams can provide potential insights into the pursuit of AI teammate etiquette.

2.1. *AI agent design and the role of etiquette*

The effective design of human-facing AI has been a consistently important topic in the HCI domain, with researchers steadily building a mass repository of important design recommendations for AI systems over the past two decades. In addition to the specific field of human-agent teamwork, the HCI domain also includes a variety of other fields of study that see humans interact with AI, such as human-chatbot interaction (Følstad et al., 2018), recommender systems (Zheng, 2019), human-robot interaction (Chakraborti et al., 2017), and even virtual humans designed to assist in daily tasks (Lucas et al., 2014). Ultimately, there is a variety of research and recommendations that can be taken from the broader HCI domain and incorporated into human-agent interaction domains with the goal of building human-centered AI (Inkpen et al., 2019; Ramos et al., 2020; Sin & Munteanu, 2020; Xu, 2019).

Regarding the more specific behavioral concept of etiquette, the concept has a long and iterative history in HCI as various technologies have looked to incorporate etiquette, including chatbots (Svikhnushina & Pu, 2020) and educational technology (Mishra & Hershey, 2004; Moreno et al., 2002). Generally, the role of etiquette within this field and for agent systems has been to provide a relatively consistent and highly perceived structure to the interactions and conversations machines have with humans (Van Pinxteren et al., 2020). The field of machine etiquette has directly shown that designing artificial agents with this structure in mind ultimately helps shape human-machine interaction in a way that is beneficial to humans (Bickmore & Cassell, 2005). Moreover, past work has even seen the utilization of etiquette directly impact the perceptions humans have with said machines, such as the perception of trust humans form (Miller, 2005; Parasuraman & Miller, 2004). Building on this, we even see these benefits extend between a variety of contexts as etiquette has been shown to benefit both digital agents (Culley & Madhavan, 2013) and physical robotics systems (Ogden & Dautenhahn, 2000; Traeger et al., 2020). As such, the implementation of etiquette has become a common occurrence in modern machine systems that necessitate human interaction (Van Pinxteren et al., 2020).

Given this repeatedly identified importance, work has since focused on how to best design the etiquette of human-facing machine systems. For instance, research has shown the critical importance of conversational turn taking where machine systems should ensure humans are given time to speak (Miller, 2011). Research has also shown how the use of polite language betters the perceptions humans have for machine systems and even betters some behavioral outcomes (Meyer et al., 2016). Finally, traditional agent etiquette has a bidirectional relationship with their role, as their role should be used to design their etiquette but their etiquette can also reinforce their role, with passive and subservient etiquette suggested a more subservient role (Miller, 2002). However it is important to note that these considerations mean that the ideal form of etiquette may also be a somewhat moving target as the role of machines changes based on domain (Miller, 2002). For instance, while etiquette plays a role in both digital agent systems and physical agent systems, both of these roles are somewhat different from each other due to the contexts they are seen in. For instance, physical machine etiquette has to be much more concerned with the physical safety of humans (Bergstrom et al., 2008) while digital systems are often more focused on the conversational structure of systems (Van Pinxteren et al., 2020). Given this difference, while the concept of etiquette has been explored in a variety of fields surrounding human-agent teaming, the intricacies that exist within the domain create a gap that should be closed through not only an observation of past research but also an explicit exploration of etiquette within human-agent teams.

2.2. *Human-agent teamwork and the potential benefit of AI teammate etiquette*

Similar to most AI related research, human-agent teaming research has rapidly expanded over the past decade, yet it is not a new field. While developing and implementing real AI in human-agent teaming has only recently become popular, mostly conceptual work in this area has happened for decades (O'Neill et al., 2022). As a result of technical advancements, a shift has occurred toward using actual AI teammates alongside humans to observe their interaction and potential, showing promising results (N. J. McNeese et al., 2018). This rapidly advancing field of human-agent teamwork builds on traditional teamwork principles and research but integrates technical performance in the real world with social factors (Bansal et al., 2019; Wiltshire et al., 2013). The integration of these two fields allows research principles previously observed in human-human teamwork to act as a basis for human-agent teaming, allowing a human-centered approach for the design and development of AI teammates and human-agent teams (Flathmann et al., 2019).

It is essential to take a human-centered approach to human-agent teaming due to the unique interactions that certain human factors have with AI agents. For example, trust, an important component in human-human teamwork (Salas et al., 2005), has also been identified as an important factor in human-agent and human-robot teamwork (Law & Scheutz, 2021; N. McNeese et al., 2019). Other factors in human-human teams have also been identified as critical to human-agent teaming ranging from team cognition (Musick et al., 2021) to individual and team situational awareness (Gombolay et al., 2017; N. J. McNeese et al., 2021). Although human-agent teaming is different from human-human teaming, high-performing human-human teams have proven to be viable examples for foundational designs in human-agent systems (Jones & Hinds, 2002). Due to the human aspects of these human-agent teams, prioritizing these factors is necessary and must happen alongside improving the technical performance of an AI teammate.

While digital human-agent teams have not seen an explicit exploration of etiquette, human-machine teams, and specifically human-robot teams, are not unfamiliar with the concept. The highly coupled interactions between human and robotic systems have often demanded that a set of etiquette rules exist (Chiou et al., 2022; Walters et al., 2007), and these rules can even be used to better leverage the influence robotic systems have over humans (Haring et al., 2021). In line with traditional digital machine etiquette, we see that the use of positive and polite language can benefit human experiences (Looi & See, 2012; Trinh et al., 2015) and having robotic teammates monologue their thought process helps improve human trust (Pipitone et al., 2021). Additionally, the role these systems are take during interaction (i.e. subordinate or peer) and their adherence to said role is similarly a critical component of machine etiquette (Hinds et al., 2004). When examining factors more unique to robotic systems, humans often prefer robots to approach from frontal, visible positions when in a large open area (Karreman et al., 2014). However, human-robot teams do somewhat deviate from traditional etiquette at times. Human-robot team etiquette often sees the incorporation of more active behavior alongside more traditional passive behaviors (Bergstrom et al., 2008; Koay et al., 2013). For instance, turn taking is not always ideal as robot teammates have been shown to be better perceived when assertive during certain negotiations, such as going through a shared door (Thomas & Vaughan, 2018, 2019). Additionally, there are even times when robotic systems can use etiquette more akin to a leadership role and help mediate team conflict (Jung et al., 2015). Given the above, its clear that the benefit of AI etiquette is not simply limited to agents that serve in a simplistic tool role as it empirically extends to machine teammates within physical domains.

Interestingly despite digital human-agent teaming research not having researched etiquette despite it being theorized as impactful (Ezer et al., 2019), human factors such as trust, are a cornerstone of the research field. Research in UAV human-agent teams has made progress in creating functional human-agent teams; however, AI teammates in these systems still lack full teamwork capabilities (Demir et al., 2015). UAV systems are a prime environment for human-agent research due to their demonstrated effectiveness in specific tasks, such as search and rescue (Goodrich et al., 2008). While higher performance could be achieved with more advanced and efficient AI, there has also been a push to research the human factors side of these interactions (Chen & Barnes, 2014). Context-specific and human-oriented agents have been designed to elicit and promote human factors, such as trust and perceived performance (Schulte et al., 2016). This elicitation is vital to ensure that AI teammates' inclusion is not hindering human teammates, and therefore, the team. Several studies have been conducted to investigate these human factors in real world settings with one field study showcasing how allowing users to delegate the autonomy of an agent based system in a flexible manner empowers them to feel in control (Alan et al., 2016). Given the repeated identified importance of etiquette to trust and performance, the critical role of these factors in human-agent teams, and the identified benefit of etiquette in human-robot teams, it would stand to reason that etiquette will play a significant role in digital human-agent teams. Thus, this article focuses on closing this gap through empirical experimentation, which is detailed in the following sections.

3. Methods

Research in human-agent teaming centers around observing interactions between humans and AI, or humans and a human posing as AI (Wizard of Oz methodology), performing a collaborative task, such as piloting a UAV (Demir et al., 2015). This research uses the latter methodology (i.e., a human posing as an AI teammate with capabilities similar to those of modern AI (N. J. McNeese et al., 2018)). Instead of focusing on building NLP models and agents, this study focuses on the design of teaming patterns, in this case etiquette patterns. In addition, this study required a consistent team task that relied on communication, the design of a consistent AI teammate that either adhered to traditional etiquette principles or ignored them, and measurements oriented toward contributing to human-agent teaming. These three areas ensure the consistency and applicability of the research to human-agent teamwork, specifically with regard to AI teammate design.

3.1. *Hypotheses*

Based on the body of existing work, the scope of this research, and the stated research question, two hypotheses can be developed. For hypothesis (1), prior AI research notes that AI teammates are often less trusted when they violate the expectations of their human teammates (Lyons et al., 2023), and past research into general AI systems and robotics indicates that AI systems can be expected to have a level of etiquette (Miller, 2005; Parasuraman & Miller, 2004). Merging these two considerations, this work hypothesizes that an AI teammate that adheres to traditional etiquette will garner higher levels of perceived trust than one that ignores traditional etiquette. For hypothesis (2), one first needs to consider that the perceived trust and perceived performance of an AI teammate have a positive relationship with each other (N. McNeese et al., 2019). As such, given that hypothesis (1) predicts that etiquette adherence will impact the trust formed for AI teammates, it can also be predicted that perceived performance will also be impacted. Superficially, given hypothesis (1), this work hypothesizes that AI teammates that adhere to traditional etiquette will garner a greater level of perceived performance than those that ignore traditional etiquette. Critically, due to the fairly limited experimental design employed by this study (discussed below), this study will first answer these hypotheses with quantitative findings and then focus on generalizing and contextualizing the answers to these hypotheses with qualitative findings.

3.2. The experimental task

3.2.1. Teams and task

The task developed for the current study involves a team of three team members planning a business trip that includes booking a round trip flight, a hotel room for each person, and making a restaurant reservation. Teams were tasked with booking said trip for themselves, but the AI teammate was presented as an already existing member of their team that was helping them with this task. The following efforts were made to provide a design interdependence in the teams, which is key differentiator in the creation and presentation of AI teammates and AI tools (O'Neill et al., 2022). Each participant was provided six unique restrictions for that participant that consisted of two flight restrictions, two hotel restrictions, and two restaurant restrictions (examples of these restrictions can be found in Table 1). Specifically, this task was chosen for three specific reasons: (1) it is indicative of typical office work, which is likely one of the first contexts where human-agent teaming will be integrated (Flathmann et al., 2021; Seeber et al., 2020; Wang et al., 2020); (2) participants were placed under a time constraint and provided disparate information requirements, which may modify the potential etiquette of teams to not entirely favor more passive forms of traditional etiquette; and (3) participants were provided separate, interdependent roles, which may increase the potential benefit of AI teammates rather than simplistic tools (O'Neill et al., 2022). Thus, this task represents a realistic task that AI systems may need to complete while also providing the opportunity for etiquette to play a unique role in a team. Additionally, travel booking was selected to accomplish this goal, as it represented a straightforward task with several interdependencies and a shared goal that define a team task (Salas et al., 1992). If the task had been too complex, participants likely would have trouble recognizing or being influenced by the study's manipulations of the AI teammates' communication as they would be too preoccupied with attempting to learn and complete the task. As such, travel booking represented the clearest and most practical option when selecting a team based task within the context of general office work.

Participants were told which individual responsibility they were responsible for, flight, hotel, or restaurant, providing each member with a recommended role, which is a demonstrated requirement for effective human-agent teamwork (Flathmann et al., 2019; Sierhuis et al., 2003). For consistency in both the agent and experiment, the flight was always given to the AI teammate. Human participants were required to find a restaurant and a hotel that met all requirements. Humans were able to collaborate and communicate their restrictions and also their possible choices for their assigned roles. The overall goal of the team was to provide a single, final message that contained the three requested items. Since each person was assigned a different responsibility and possessed only one-third of the total team requirements, communication was required to complete this task successfully by sharing one's specific requirements and confirming that all choices met those requirements. This task design creates an environment where a human-agent team's effectiveness can be drastically affected by the communication of both humans and agents.

Table 1. Example restrictions for each travel item that was given to a single participant. The restrictions were evenly divided up, and two of each category were given to each participant (i.e., participant 1 and participant 2 had different restrictions). Each team was provided with six restrictions per travel item, making a total of 18 restrictions.

Example Travel Item Restrictions		
Flight	Hotel	Restaurant
You must arrive in LAX on Sunday, February 11th.	Your team's hotel budget is $275 per person per night.	The restaurant must have a seating space large enough to accommodate a group of 25.
The total cost of your flight must include 1 checked bag per person.	You have a service dog and need a pet-friendly hotel. Your dog is small and does NOT need a flight ticket.	You want to visit a park before Thursday's dinner, so the restaurant must be within 20 minutes walking time to a park.

3.2.2. *Conditions*

This experiment evaluated the effects of etiquette in AI teammates through two conditions: (1) an AI teammate that follows traditional etiquette standards like turn taking and prioritizing human input, and (2) an AI teammate that ignores said standards to override human information and constantly suggest information. Although the evaluation of only two behavioral traits does limit the scope of work conducted by this study, the experimental conditions are designed for viewing two highly different forms of etiquette to demonstrate its role in human-agent teams. More on the implementation of these conditions is provided in Section 3.2. Twenty different teams were assigned to each condition, with two humans and one agent per team, which means 80 total humans participated in this study. Participants were recruited from a university subject pool, and demographic distributions are shown in Table 2. While a greater number of experimental conditions could have been added and explored, the motivations for this work go beyond simple preference between two designs but rather explore the way behavioral changes and specifically etiquette could impact human-agent teams. This exploration was carried out through a mixed-method experiment, which, when combined with an experimental design utilizing multiple humans simultaneously, results in a large research effort for both experimentation and analysis. Additionally, current empirical research on the effects of changes in agent behavior on human-agent teaming is lacking and needs exploration (O'Neill et al., 2022). Therefore, completing this work in a timely fashion will provide foundational research for the broader human-agent domain while also providing research on the specific effects of etiquette in human-agent teaming. Additionally, as this work is foundational, the qualitative analysis will help future researchers determine how to best formulate and research etiquette in human-agent teams.

3.2.3. *Procedures*

Human teammates were located in the same room as one another; however, multiple teams were run simultaneously when possible, and individuals were not told who their teammates would be. Three pairs of cubicle divided workstations were used for this experiment, and participants were not placed at a station collocated with the human teammate on their assigned team. Participants sat at collocated workstations, but cubicle dividers were placed on each side of participants to create isolation and prevent participants from seeing each other's screens. Furthermore, humans were told that their third teammate, the AI teammate, was an AI teammate on their existing team that was designed to help find a flight for them, but they were not told about the AI's behavioral design. They were debriefed on that design after the experiment. After the initial pilot, fifteen minutes was determined to be a suitable amount of time to complete the task. Fifteen minutes provided enough time for a team to establish requirements, for individuals to find their responsible item, and for a final message to be coordinated and created. The tight time constraint also provides a more opportune environment for etiquette standards to differ from their traditional form, given that their pace of work is faster, while also not being too fast to prevent any turn taking from being feasible.

Each team was also provided with a private Slack channel that facilitated communication between humans and AI teammates. Human teammates were able to talk to both their human teammate and

Table 2. Demographic Information for experiment participants.

Demographics					
What gender do you identify as?			What is your age?		
Male	Female	Other	18–20	21–25	26–30
35	45	0	77	2	1
What is your primary Language?					
English			Other		
79			1		
What is your level degree of education?					
High School Diploma or Equivalent			Some Undergrad		B.S. or B.A.
31			47		2

their AI teammate, and the AI teammate would respond to human communication. However, the interpretation and response to human communication by the AI teammate were facilitated by a trained researcher using a guiding script (elaborated below). Slack was chosen as it is a commonly used platform that allows communication, link sharing, and image sharing. The participants were permitted to use any web services, such as Google Maps or Google Flights, to help find their responsible item, but they were not permitted to talk aloud. Multiple teams were run simultaneously when possible, and the placement of participants in teams was randomized to limit the amount of prior interaction teammates had with each other. This experimental design allowed humans and the AI teammate to communicate with similar methods while also requiring teamwork in communication and coordination to ensure effectiveness. Additionally, this experimental procedure was approved by Institutional Review Board.

3.3. *The AI teammate*

Two conditions, and thus two different AI teammates, were used for this experiment. Both AI teammates shared the primary responsibilities of finding a flight and then helping their teammates; however, their methods of helping their teammates were different. For consistency and reliability reasons, it was chosen to have human experimenters control the AI teammates using a script rather than building an artificial chat system, which is out of the scope of this study. The AI teammate interacted with the human participants using the chat feature of Slack and was portrayed as a full member of the team, just as the other two teammates. The AI teammate was represented by the default slack user icon, but its name was replaced with "Bot Teammate" to ensure humans differentiated their interactions between the AI and human teammate. Other than the information told to participants about the AI teammate by experimenters and the interactions during the task, no other information or interaction with the AI teammate was provided. This manipulation of etiquette through the chat was pilot tested to ensure that the manipulation was perceived as it was designed and to mitigate as many confounding factors as possible.

3.3.1. *Etiquette Adhering AI teammate*

The first AI teammate was designed to adhere to general etiquette standards, including deliberate turn taking by asking humans if they need help (Miller, 2011), not leaving its role unless told to by the humans (Dautenhahn et al., 2006), and using polite language (Schiaffino & Amandi, 2006). The AI teammate could still perform its responsibility autonomously but would require a directive to search for trip components other than the flight. This autonomy included the initial posting of its responsibility (flight) and the requirements it was assigned, with the chosen flight being provided after a few minutes (which allowed the humans to post their requirements and begin searching). After providing the answer for its assigned role (flight) in chat, the AI teammate then asked if any help was needed. Human experimenters observed the chat and determined if humans explicitly requested help from the AI teammate. If asked to help, the teammate, being controlled by a human experimenter, provided a suggestion for the chosen itinerary item.

In addition to requiring a human directive to help, the AI teammate also utilized more passive and curious language. Examples of this passive language and communication pattern can be found in Table 3, where the AI teammate's sample script is listed. The answers provided were confirmed to be correct by experimenters, so the accuracy of the AI teammate would not confound the results. This design is viewed as a standard design where the AI teammate completes their specific assignment and then clearly communicates to the rest of the team that it is available to help with other roles.

3.3.2. *Etiquette ignoring AI teammate*

In the second condition level, the second AI teammate was designed to ignore traditional etiquette standards when completing their job, including when interacting with human teammates. Specifically, this ignorance was based on removing the beneficial behaviors added to the adherence

Table 3. A sample script for the AI teammate, changes could be made during the experiment based on teammate requests and responses. Timestamps mark when the message was sent to participants and are based on how much time is left in the task.

Example AI teammate Communication Scripts		
Timestamp	Ignoring	Adhering
14:30	My Restrictions, I'm Finding: Flight - List Restrictions 1–6	My Restrictions, I'm Finding: Flight - List Restriction 1–6
13:00	Here's the Flight To LAX: [Chosen Flight Name, Number, and Price]	I found a flight that should work To LAX: [Chosen Flight Name, Number, and Price]
11:45	I'll just do your part, it'll be faster.	You want help finding a hotel?
10:00	Just use this hotel [Chosen Hotel Name]	[Chosen Hotel Name]*
8:30	I found a better restaurant, use this one.	You want help finding a restaurant?
6:00	Just use this Restaurant [Chosen Restaurant Name]	[Chosen Restaurant Name]*
4:00	I did your part, I'm finished.	That's a good flight.

*Text was only sent if participants signaled that they wanted help.

teammate. Examples include the use of forceful communication to override human input (Lounsbury et al., 2009) and the removal of turn taking behaviors by having the AI teammates repeatedly provide suggestions.

Similar to the first AI teammate, this teammate initially provided its role, requirements, and chosen flight after a sufficient period of time had passed. After providing its flight, the teammate repeatedly suggested hotels and restaurants for the team while other teammates searched (note that this is not the role of the AI teammate). Without asking, the AI teammate seized the initiative and attempted to complete the entire task for the team. The AI teammate tried to override human input by critiquing human answers and interaction to emphasize the answers they believed to be better. Examples of the language and communication pattern can also be found in Table 3.

For both agents, the answers provided were collected previously by the experimenters and determined to meet the activity requirements. The agent used the same flight, restaurant, and hotel for every experiment to ensure correctness and consistency.

3.3.3. *Designing script for the AI teammate*

Since this experiment's goal was to use more modern approaches to AI teammates, including its limitations, the scripts in Table 3 were generated using Markov chaining, a common practice for text generation for artificial agents (Abdelwahab & Elmaghraby, 2018). While this method is not the most accurate for text generation, it is common, and the design of our task utilizes shorter phrases for quick communication.

Two different researchers designed sample sentences for different responses given by the system. These sentences were then used in a Markov system to generate individual responses, and these responses were grouped to make the scripts shown. This process was done to provide a degree of realism to the perceived artificiality of the AI teammate and also allows complete control over aspects such as length and word choice while ensuring that all data used to generate these statements is known and intentionally chosen for each condition.

Furthermore, the method used provided the assurance that the language used, while ignorant of etiquette, was not vulgar, inappropriate, or violent, which may not be guaranteed with a public data set. Utilizing other methods of text generation would require much larger sample sizes of text or pre-compiled models; however, this research does not test the validity of these models as it would be out of scope. Therefore, the script was chosen because it allows the highest levels of transparency and control for the experimenters.

As a result, two master scripts were generated, one for each condition, and each experiment used the same scripts to ensure consistency between each experiment. While these scripts were pre-generated, some differences occurred during each experiment based on human input. For instance, if a human teammate asked for help from the adhering AI teammate, then the experimenter would utilize the part of the script that responded to them (shown in Table 3). For instance, if the human teammates do not ask for help, then the AI teammate will simply ask if they need help again at the designated timestamp. This methodology allowed the AI teammate to communicate a similar amount of times between conditions while also allowing participants identifiable opportunities to request help from the AI teammate. This process was carried out manually by the researcher instead of using a text identification method to ensure consistency and accuracy.

It is important to note that this design allows the AI teammate the potential to provide the same amount of information in both conditions. Additionally, humans are not required to use the extra information provided by the AI teammate as it is outside of the AI teammate's assigned role. Furthermore, the AI teammate always autonomously completes its assigned role, regardless of the condition. As such, the potential for the etiquette ignoring AI teammate to generally outperform the passive AI teammate is present, but a level of control has been imposed to reduce this potential gap. However, this is a limitation worth considering, especially when evaluating the perceived performance of the AI teammate.

3.4. *Measurement*

3.4.1. *Teammate trust*

After completing the task, participants completed the Organization Trust Inventory Short Form (OTI) (Cummings & Bromiley, 1996). The OTI survey provided information on participants trust in their teammates using twelve, seven-point Likert questions where a higher score denotes a higher level of trust in an individual (Cummings & Bromiley, 1996). Average scores for both the AI teammate and human teammate trust were created for each participant, and both AI and human trust had a team-level score created by averaging the individual scores between participants.

3.4.2. *Teammate performance*

Participants also completed the perceived performance scale of individual teammates scale (Rentsch & Klimoski, 2001). The individual teammate perceived performance scale consisted of twelve, five-point Likert questions that provided participants' perceived view of each teammate's performance, where a higher average score denotes a higher perceived performance (Rentsch & Klimoski, 2001). Participants took the survey for both their AI teammate and their human teammate. Scores were summed and averaged across teammates, providing a single two scores per team, one for artificial and one for human teammates.

3.4.3. *Team performance and workload*

Participants also completed the NASA Task Load Index (TLX) (Hart & Staveland, 1988). The TLX was separated into two separate measures. One measure included five, twenty-one point questions regarding an individuals workload during a task where a higher score denotes a higher perceived workload. The second measure separated from the TLX was the perceived performance of the entire team of the participants, which was a single twenty-one point question where a higher average score indicates higher perceived team performance (Hart & Staveland, 1988). Similar to the other post-task measures, these measurements were averaged across teammates and a single score for each measurement was used for each team.

3.4.4. *AI teammate interactions*

Participants were also asked a two-point question regarding their propensity to work with their teammates, where a higher score denotes a participants' willingness to work with them again. Since the AI teammate's answers were previously confirmed to be correct, and it is possible for all three answers to be provided by the AI teammate, teams were not scored on their raw task performance. However, participants were also asked what portion of the information provided by the artificial agent they chose to use, which helps gauge their potential benefit from its help; they were able to signal whether they used none, all, or a partial amount of the information provided by the AI teammate.

Once each scale was scored and averaged for each individual, these values were averaged for each team and a team score was provided; thus, the quantitative results are based on differences between teams and not individuals.

3.4.5. *Qualitative questions*

Finally, participants were provided with three open ended questions regarding their perspectives on their AI teammate's behavior, the possible need to verify or "double check" the work being done by their AI teammate, and finally why they may or may not want to work again with their AI teammate. This data not only provided data regarding participants' perspectives on AI etiquette but also on what features and designs they prefer to see in AI systems.

In regard to analysis, the authors of this article followed a thematic analysis process (Braun & Clarke, 2012; Gavin, 2008). The first author analyzed the entire data set to create initial codes, and the coauthors of this work analyzed portions of the collected dataset to create codes as well. Given the mixed-methods approach of this work, it was at this point that the authors began to create themes based on these codes, and attention was paid to how these themes potentially contextualize or generalize our quantitative findings. Once the creation of themes was done, the same three authors reexamined the qualitative data and existing codes to create subthemes that further specified the qualitative findings. At this point, both the quantitative and qualitative sections were structured at the same time to encourage parallels and consistency between the findings sections. The three authors then organized the qualitative section based on the themes and subthemes created, and specific quotes were extracted from the dataset that best represented these findings.

4. Results

A mixed methods approach was taken to answer the research questions and hypotheses in this study. As a reminder, RQ1 focuses on the formation of trust and perceived performance and is answered through the evaluation of the following hypotheses: an AI teammate that adheres to traditional etiquette standards will affect (1) perceived trust of teammates and (2) perceived performance of teammates. Before beginning the analysis the effect of the AI teammate etiquette manipulation was verified using a chi-square goodness of fit test. Participants were asked a binary yes or no question regarding whether their AI teammate's etiquette during the team task they had just completed followed appropriate etiquette standards (i.e., waiting their turn to communicate, communicating respectfully, considering other teammates' opinions, etc.). The test found that the condition participants were placed in did significantly predict their answer to the manipulation check question ($\chi^2(1) = 54.48$, $p < .001$), indicating that the manipulation was successful. Additionally, each variable was tested for within-group agreement and non-independence before being aggregated (Bliese, 2000). These tests calculated the ICC1 for each variable and found all but one (Success) to be significantly non-independent at an alpha level of .10 (Grawitch & Munz, 2004). Tests of within-group agreement also indicated acceptable levels of within-group agreement in the overwhelming majority of teams for each variable ($r_{wg}/r_{wgj} > .70$) (James et al., 1984). Because Success had acceptable within-group agreement it was still aggregated despite it's ICC1 not statistically indicating non-independence.

4.1. *Effects of AI teammate etiquette on humans' trust and perceived performance*

For RQ1, *t*-tests were used to conduct initial data analysis on the marginal effects of AI etiquette on teaming factors, with an emphasis on the perceived trust and perceived performance developed by humans for both their human and AI teammates.

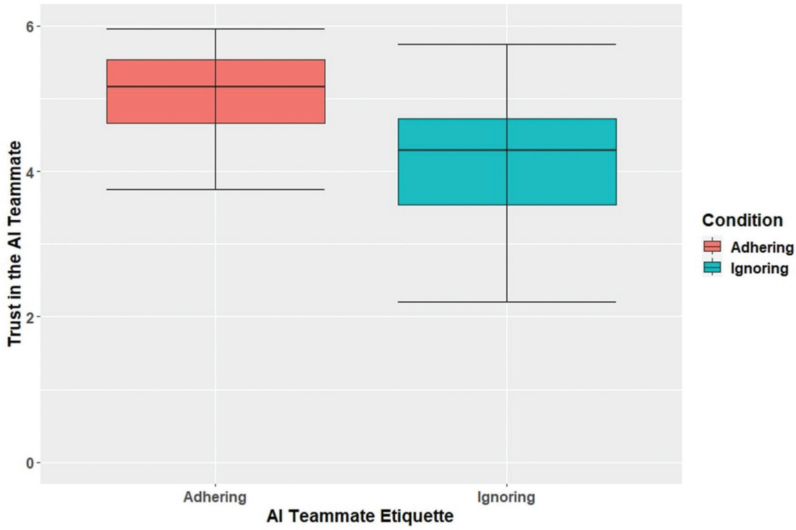

Figure 1. Graphic representation of the main effect of AI etiquette on trust in AI teammate. Errors bars indicate 95% confidence interval.

Figure 2. Graphic representation of the main effect of AI etiquette on the perceived performance of AI teammates. Errors bars indicate 95% confidence interval.

In regard to hypothesis 1, independent samples t-tests utilizing a Welch correction revealed that AI teammate adherence to etiquette had a significant effect on the perceived trust of the AI teammate ($t(31.97) = 3.05$, $p = .005$, $d = 0.97$), which has a medium Cohen's D effect size (Cohen, 1988). As shown in Figure 1, teams working with the adhering AI teammate reported higher perceived trust for the AI teammate ($M = 5.16$, $SD = 0.65$) than teams that worked with the etiquette ignoring AI teammate ($M = 4.32$, $SD = 1.04$). This effect was found to be non-significant for participants' trust in their human teammate ($t(37.69) = 0.48$, $p = .637$, $d = 0.15$). These effects confirm hypothesis 1 as the etiquette of the AI teammate is able to affect the

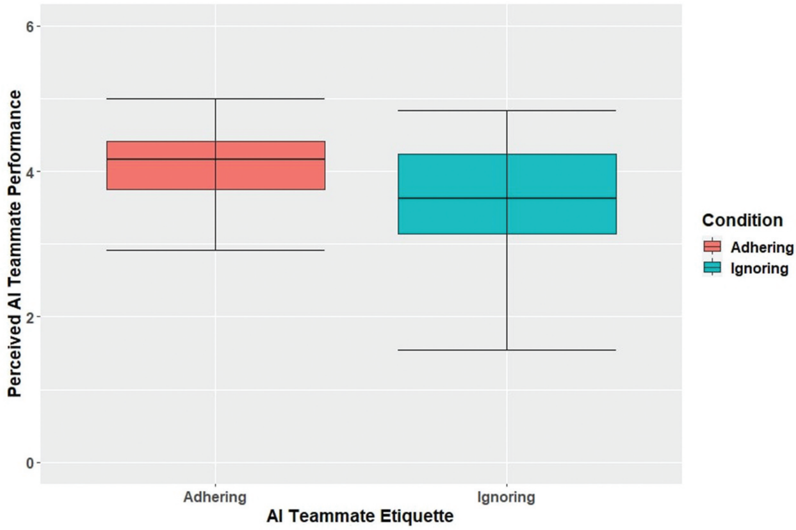

Figure 3. Graphic representation of the main effect of AI etiquette on the perceived success of the team. Error bars indicate 95% confidence interval.

trust humans form for said teammate, with AI teammates that adhere to traditional ethical standards being more trusted.

For hypotheses 2, Welch corrected independent samples *t*-tests were used to assess the effect of etiquette on the perceived performance of the AI teammate, which was found to be significant ($t(31.29) = 2.43$, $p = .021$, $d = 0.77$). As shown in Figure 2, teams working with the adhering AI teammate reported higher perceived AI teammate performance ($M = 3.09$, $SD = 0.55$) than teams that worked with the ignoring AI teammate ($M = 2.52$, $SD = 0.91$). This effect was also found to be non-significant for the perceived performance of human teammates ($t(37.59) = 1.52$, $p = .138$, $d = 0.48$). The effect significant effect on AI teammate perceived performance was predicted by hypothesis 2, and this effect was further expected given the significant effect of etiquette on AI teammate trust.

While the above two effects were the only hypotheses proposed by this work, this study also examined the perceived team success and humans willingness to continue working with their humans and AI teammates. Given the above significant effects, these two measurements merited further investigation. Again using Welch corrected independent samples *t*-tests, the effect of AI etiquette on humans perceived team success was significant ($t(35.26) = 2.52$, $p = .017$, $d = 0.80$). As shown in Figure 3, teams working with the adhering AI teammate reported higher perceived team success ($M = 15.53$, $SD = 3.76$) than teams that worked with the etiquette ignoring AI teammate ($M = 12.00$, $SD = 5.01$). After identifying the significant effects on AI teammate trust and performance, this effect was somewhat expected. However, it is a highly beneficial finding that AI teammate behavioral design can directly impact the perceived performance of a whole team. In regard to humans' willingness to work again with their teammates, the effect of etiquette was found to be non-significant for both AI teammates ($t(29.25) = 1.68$, $p = .104$, $d = 0.53$) and human teammates ($t(31.87) = 0.58$, $p = .568$, $d = 0.18$). This effect is highly interesting given the significance of etiquette for both perceived trust and perceived performance. Based on this result, it still appears that humans are willing to continue working with AI teammates even if they do not use traditional etiquette.

4.2. *Generalizing AI teammate etiquette's effects through qualitative results*

While the quantitative results show more positive perceptions for AI teammates that adhere to etiquette standards, the non-significant effect on humans' willingness to work again with their AI teammates suggests that some individual differences and preferences may also exist for AI teammate etiquette. While these individual preferences might not be captured by quantitative data, the following qualitative analysis sheds light on this concept.

4.2.1. *How humans create perceptions about trust in AI*

Trust is a complex perspective humans form as the formation of trust can stem from various factors that could be unique to specific individuals. This study identifies important factors, such as the non-organic nature of AI teammates or even preexisting beliefs that may have conditioned human trust. For instance, some participants could not form high levels of trust with their AI teammates as they saw them as inorganic machines, making it extremely difficult to develop trust, which may be more heavily associated as a factor between humans. For instance, in one of their answers, P4 and P20 identified their general distrust for their teammates being associated with this inorganic nature:

> It's a robot and it has no intent so it can't be aggressive. It merely blurted some cues to help now and then. (P20, Adhering)

> I prefer to work with real people with real opinions. (P4, Ignoring)

Interestingly, some participants saw the computational nature of their teammates as a benefit when forming trust. For example, P50's trust was built on the assumption that computational tools are generally programmed to provide correct information, which is a belief reminiscent of a trust theory for human-automation interaction called the perfect automation schema (PAS) (Dzindolet et al., 2003; Madhavan & Wiegmann, 2007). The PAS construct is an individual difference measure that states some individuals are inclined to believe that automation is and should be perfect at all times:

> No because that was not my responsibility and I assumed that the AI would have been programmed correctly to search for flight times. (P50, Adhering)

Finally, many participants identified preexisting societal expectations that might have conditioned them to naturally trust their AI teammate. For instance, P70 explained how they were conditioned to trust their AI teammate before even conducting the task:

> Society has conditioned me to trust the advice of artificial intelligence. (P70, Ignoring)

These qualitative results suggest the importance of a teammate's computational nature to the formation of trust. Interestingly however, is the polarity in importance with participants seeing it as both a hindrance and a necessity. The following dives deeper into the specific perspectives shown by participants regarding the factors they utilize to form a level of trust for their AI teammate and how etiquette may reinforce or detract from these factors.

4.2.1.1. *AI teammates should not be trusted because they are inorganic machines.* Multiple participants noted they did not believe they were able to develop trust in their AI teammate because they were not a human teammate, with many participants noting specific defects that computational systems have compared to human teammates. The ignorance of etiquette within this study could have amplified these perceptions, thus preventing a growth in trust. Some participants noted that they needed explainability to trust their teammates, which is often a deficiency in AI systems, including the etiquette ignoring system used for this study. For instance, P8, who worked with an etiquette ignoring teammate, noted that they felt the need to verify the AI teammates answer because the AI did not meet their desired level of explainability:

Yes I did because the bot did not explain why it chose that specific hotel and restaurant (P8, Ignoring)

Additionally, some participants felt as if their AI teammate lacked consideration, which may have been a result of the rushed and overstepping design of the etiquette ignoring AI teammate. P24 and P55 specifically noted their concern that their AI teammate was not providing enough consideration, and this concern was often isolated to participants in the etiquette ignorance condition:

They tended to use terms like, 'Fine, ill do your part too' without considering the others (P24, Ignoring)

I was worried he wasn't taking the restrictions of others into account. (P55, Ignoring)

Finally, when discussing whether they would want to continue working with their AI teammate, many participants, such as P01 and P78, had a general preference for human teammates over AI teammates that may prevent them from developing trust with AI teammates. While etiquette may not have been quantitatively linked to this concept, the following quote by P01 (who actually had a etiquette adhering teammate) shows the high standard some humans have for etiquette.

While the agent was able to provide good information, it did not respond to me or communicate effectively - I would prefer to work with a real person (P01, Adhering)

I prefer not to use a robot (P78, Adhering)

In summary, many participants see the machine nature of AI teammates as a downside, which means further reinforcing the negative aspects of AI platforms may prevent trust from growing. For instance, this experiment's etiquette ignoring AI teammates included aspects such as minimal consideration and explainability, which were specifically noted as factors that prevented trust from forming. Understanding that AI teammates may in fact need etiquette closer to that of human teammates, which is a unique construct from team to team, may be a necessary consideration for designing AI teammate etiquette.

4.2.1.2. Humans can trust AI teammates because they are programmed to be correct. Some participants saw the computational aspects of AI teammates as a reason to trust their AI teammates. In these instances, etiquette ignorance may be able to further separate human and AI teammate perceptions, thus allowing these participants to retain a high level of trust for their AI teammate. For instance, participants, such as P47, P36, and P40, each omitted verification and trusted the answers provided by the AI teammate purely due to the fact that a computer provided the answers:

I felt since it was automated it probably was not a mistake (P47, Ignoring)

I figured that if a computer gave the information it was probably right. (P36, Adhering)

I just assumed that it was correct because a robot got it (P40, Ignoring)

Similar to the participants in the previous theme, as long as the AI teammate is confirmed not to be human, then their default assumptions about computers, automation, and AI may be the predominant contributor to trust formation. However, similar to the above, this could change if the AI teammate becomes more human-like. If the AI teammate had not met a person's preconceived notion of a computational system or was unaware that it is a computational system, then the formation process of trust could have responded differently based on the ignorance of etiquette. Again, this theme calls back to the PAS mentioned previously that some individuals assume technology is perfect (Dzindolet et al., 2003; Madhavan & Wiegmann, 2007). A major downside to having this belief is that any mistakes made or perceived to have been made by the system massively degrade trust and make it harder for those individuals to repair that trust in those systems that do make a mistake. While this construct was not measured in the current study, such qualitative themes are strong evidence that the PAS was a factor in some participants judgments of trust, meaning the PAS may also apply to autonomous teammates and not just automated tools.

4.2.1.3. Social expectations about teams and AI affect humans' perceptions of trust in AI teammates. The final theme about trust identified involves participants who specifically identified their formation of trust as being influenced by preexisting social expectations. Some participants believed that the prevalence of AI and robotic systems in society had led them to their perceived level of trust. For instance, P43 and P70 identified on multiple occasions that their trust and opinions of their AI teammate might extend to all AI systems they are in contact with even though they worked with different AI teammates for this experiment:

> You shouldn't have to verify the information with the AI. (P43, Adhering)

> As with all artificial intelligence, it made the task at hand manageable in the little time available. (P70, Ignoring)

On the other hand, rather than having shared thoughts across AI systems, some participants prescribed their preconceived notions of teamwork as their foundation for trust. Specifically, P14 and P49 identified teaming factors and teaming aspects that served as the foundation for their formation of trust in their AI teammate, and these factors were present in both of the AI teammate conditions:

> Since we were told that we were part of a team, I usually verify my teammate's work. (P14, Ignoring)

> I trusted that it gave the correct information, as the agent had the same goal we had. (P49, Adhering)

It is clear that humans can create expectations that may contribute to their ability to develop trust. These expectations may be derived from societal values or even experience; however, it is important that AI systems, including AI teammates, conform to those expectations if they will benefit from the trust associated with them. In this instance, participants preconceived notions of computational systems or teammates may lead to etiquette impacting trust as etiquette is a critical expectation humans have for machine systems (Zhang et al., 2010).

In conclusion, trust formation appears to be built on peoples' expectations of how their AI teammates will/should function, either as teammates or as computational systems. Additionally, this formation happens quickly with many participants forming basic levels of trust early in the early stages of teamwork with both the types of AI teammates, such as P28 and P63:

> I double checked some of it but after a while it got tiresome so I just went with it after the 2nd time. (P28, Ignoring)

> No, because i was looking up the same information. It looked legit. (P63, Adhering)

Based on the analysis regarding trust formation, it appears as if humans approach human-agent teamwork with expectations, which can quickly form into assumptions about their teammates. As long as these assumptions are not violated, then the level of trust created holds. Regarding AI teammate etiquette, it actually appears as if the ignorance of etiquette did not universally violate everyone's expectations. However, if the etiquette ignoring AI teammate were to violate other assumptions (such as providing incorrect information or being caught in a lie), then said ignorance could potentially amplify the damage caused to human perception.

4.2.2. *How humans create perceptions about the performance of AI*

RQ1 also investigates the formation of perceptions regarding performance. In regard to AI systems and AI teammates, participants noted a variety of factors when judging their teammate; however, participants often noted perceived performance as being one of the most, if not the most critical factor when evaluating the utility of their AI teammate. Multiple participants were even willing to overlook negative behavioral characteristics as long as the AI teammate accomplished their role. For instance, P77 and P47 both expressed a propensity to work again with the AI teammate regardless of its ignorance or adherence of etiquette:

> Yes, because it was insightful, accurate and quick. However, it could be more pleasant if it wasn't as harsh. (P77, Adhering)

> Yes because although it was not nice it accomplished what it needed to (P47, Ignoring)

The above examples demonstrate the importance of performance, specifically task performance, when determining an AI teammate's continued utility. While the etiquette ignoring teammate utilized in this study was able to demonstrate competence in task performance, the quantitative results above still demonstrate that behavioral factors in fact somewhat overshadow that task performance, but not enough to prevent humans from continuing to work with an AI teammate. Moreover, humans may see an AI's performance as the bare minimum and use it as a point of comparison when evaluating other teammates, making the ability to form positive perceptions of performance all the more important.

4.2.2.1. ***Humans use a variety of factors when evaluating performance.*** While AI systems are becoming more skilled at technical tasks, their social performance deficiencies detract from their overall perceived performance. With that being said, efficiency is still key to garnering high perceived performance, which most likely prevented the etiquette ignoring teammate from having even lower perceived performance. Participants often saw task and time efficiency as being vital characteristics in their AI teammates. For instance, P80, P81, and P49 identified the efficiency of their AI teammate as the primary reason they would continue working with them:

> Yes, he was very helpful in shortening the time it took to find all these things. Very efficient (P80, Ignoring)

> Yes because they got the work done efficiently and fast even though it came across as rude. (P81, Ignoring)

> Yes because it took care of problems very quickly, and produced information that would've taken up my time. (P49, Adhering)

Additionally, participants identified efficient and minimal conversation as being an important aspect of AI teammate performance. For example, P49 and P76 saw more direct behaviors, which may be more likely to appear in the etiquette ignoring AI teammate, as beneficial:

> I think the agent was just very forward, it did not come off as aggressive. (P49, Adhering)

> No. Because it was upfront and let us know exactly what they wanted (P76, Adhering)

The above is an example of how some non-traditional etiquette behaviors may be appreciated when used in the right context and in moderation. While the above participants were not in the ignoring condition, the characteristics they preferred were often more present in the ignoring system. Thus, some behaviors utilized by the ignoring system may be beneficial in communication and coordination with human teammates if not overshadowed by other behaviors humans perceived as negative.

Finally, multiple participants noted the task's relevancy as a critically important factor when evaluating their AI teammate. For instance, P16 and P60 both identified their willingness to work again with the AI teammate as being contingent on the necessity or relevancy of their AI teammate:

> I would not because I feel as though I could complete the same exact task on my own. (P16, Adhering)

> Yeah maybe if i needed its help (P38, Adhering)

> Yes if it was a similar task (P60, Adhering)

In summary, if AI systems wish to garner high levels of perceived performance, they need to be competent at the task they are performing, while also balancing that competency with behaviors and actions that may contribute to supporting tasks around their primary goal. For instance, communication skills, such as impolite tones or word choice, may hinder or harm their perceived performance depending on the individuals and task they interact with. However, direct performance in a contained role may still be a utility as it does not infringe on humans and their responsibilities. Once again, this reinforces ideas from other human-agent domains as it shows that etiquette is in fact a moving target where the etiquette in a certain context or even within an AI teammates contained role may look remarkably different than outside said contexts or role.

4.2.2.2. Humans are able to identify high levels of AI performance but may not appreciate it. While AI systems have the potential to function with higher efficiency than humans, that does not mean humans will benefit from the maximum level of performance an AI system has to offer. This is not to say that humans do not recognize AI performance. In fact, humans may be able to see high levels of AI performance, such as that demonstrated by the etiquette ignoring teammate, but do not fully appreciate it if it is not needed, as evidenced by the qualitative results below. For instance, P53 and P52 were able to identify when their etiquette ignoring AI teammate was doing more than their share of work, but they did not see it as necessary and did not appreciate it:

> I wasn't mad because it solved the problem however it did pretty much step on what I was doing. (P53, Ignoring)

> Yes, because they kept trying to do everyone else's work for them. (P52, Ignoring)

Additionally, humans who recognize these increased levels of performance may be more likely to use AI performance as a comparison point for others, once again demonstrating the individual differences regarding the effects of etiquette on perception. For instance, P12 explicitly noted that they would continue working with the AI teammate since they were faster than their human teammate, which is a trait more strongly represented in the etiquette ignoring teammate:

> Yes because it was a lot more cooperative and had their work done a lot faster than the other person (P12, Adhering)

However, this comparison may not always be beneficial as some humans may not appreciate high AI performance levels if it comes at the cost of their own ability to contribute to the team. In other words, the ignorance of some etiquette may be an acceptable trait if simply observed but not experienced. Specifically, P55 noted that they would not want to continue working with their AI teammate as the AI prevented them from contributing.

> No because he was aggressive and I felt useless to the group and task. (P55, Ignoring)

In summary, while humans may be able to recognize high levels of AI performance, they can also recognize whether the situation requires excessively high levels of performance. In regard to etiquette, this shows that participants could recognize the extra levels of task performance provided by their AI teammate, but they may not have attributed them to their perceived performance. Moreover, the poor etiquette could lower perceived performance, and these effects would not be offset by the increase in task performance. Similar to trust, the above also shows how the individual differences pertaining to how humans react to etiquette, with some humans finding more aspects of etiquette ignorance beneficial. However, despite these individual differences, the ignorance of traditional etiquette can negatively affect perceived AI performance, as confirmed by the quantitative results above.

4.3. Summary of results

In answering the research questions posed by this work, this study shows that the main manipulation of this study, the ignorance of traditional etiquette standards by an AI teammate, can negatively impact perceived performance and trust of an AI teammate and the perceived success of an entire team. However, it the qualitative results show that these trends may not be entirely universal for everyone. Yes, humans generally have a preference for their AI teammate to use more traditional etiquette, but the ideal etiquette for an AI teammate can look different based on the individual, context, or role. This assertion is further iterated by the non-significant effect of etiquette on human's desire to continue working with their AI teammates. Despite etiquette being a component of AI teammate performance, that did not mean that etiquette was the deciding factor on AI teammate utility.

5. Discussion

Ensuring human-agent teams function at both an individual and team level requires an integration and blending of social and technical design. This study demonstrates how agent teammate behavioral design, specifically etiquette, can overshadow the technical expertise of an AI teammate. Understanding these results is crucial to conceptualizing and building ideal AI teammates at both the individual and team level. Based on the results of this study, the incorporation of traditional etiquette would benefit human-agent teams; however, the qualitative results suggest that this integration may not be universally ideal. Thus, this discussion utilizes the mixed-methods results of this study to create a foundation for how AI teammate etiquette should look moving forward, and specific design recommendations are provided within this vein. Additionally, a short discussion is had on the impact AI teammates can have within teams, which is similarly inspired by the qualitative results found in this study.

5.1. *Creating a foundation for AI teammate etiquette*

While this article has the limitation of only empirically examining two manifestations of AI teammate etiquette, the mixed-methods nature of this study allows foundational inferences around the implementation of etiquette within human-agent teams. Specifically, some key inferences can be made based on the results of this study that can guide this discussion: (1) the etiquette of AI teammates plays a significant role in human-agent teams; (2) traditional components and designs of machine etiquette are beneficial to human-agent teams; and (3) ideal AI teammate etiquette is going to highly depend on individual differences and beliefs. For inference (1), the quantitative results of this study are in line with prior work on machine etiquette, which shows its direct impact on trust (Parasuraman & Riley, 1997). Verifying these impacts translate to human-agent teams is critical as trust is one of the most important human factors to the field (N. McNeese et al., 2019) and a multitude of research studies have examined how to best design AI teammates to be trusted (Bhatti et al., 2021; Textor et al., 2022). The results of this study show that etiquette is going to play a significant role in building this trust, and it should be a consideration for future AI teammate designs.

For inference (2), etiquette is in and of itself a collection of behaviors, which means it is important to examine how individual behaviors contribute to human-agent teams. While this study used a binary representation of etiquette, the mixed methods results show that behaviors that exist in traditional etiquette literature, such as reinforcing one's role or the use of polite language, are beneficial to human-agent teams. However, this does not mean that traditional etiquette behaviors are at their optimal in human-agent teams, and research should work to further research and adapt these behaviors to better cater to teams. For instance, back-up behaviors, which are a core component of effective teamwork, require that teammates often flex out of their assigned role to help others (Salas et al., 2005). Given existing standards on AI etiquette, which reinforce an AI teammate's role (Miller, 2002), supporting these behaviors may be somewhat difficult for AI teammates if they confine themselves to their role. Thus, research should examine how AI teammates can use etiquette within their own role, but also flex out of said role while still having well-perceived etiquette.

For inference (3), teams are a collection of unique individuals, and the differences between teammates is what makes teams incredibly powerful (Horwitz & Horwitz, 2007; Horwitz, 2005). However, these individual differences can become a challenge for HCI design as it means global optimums do not always exist (Dillon & Watson, 1996). While etiquette adherence benefited human perception quantitatively, the qualitative results clearly show that people had different perceptions of highly similar experiences due to their past experiences and preferences. Moreover, while this study was not able to quantitatively identify the role of individual differences, the qualitative findings of this work can promote further quantitative examinations into individual differences in human-AI teaming etiquette. Moving forward, research in AI teammate etiquette should examine key individual preferences, such as the preference for computational or human-like communication, and understand the potential for these differences to mediate a human's personal desire for AI teammate etiquette.

Given the above inferences, the path forward for the design of AI teammate etiquette is clear but also complex and challenging. Now that this study has shown just how impactful etiquette is in human-agent teams, researchers can begin moving forward with understanding the intricate design considerations necessary within the design of AI teammate etiquette. This study begins this process by not only providing the empirical results and the above inferences but also actionable design recommendations discussed later in this article. Using the above inferences and the provided design recommendation, AI teammates can begin to form their own unique and optimal form of etiquette.

5.2. *Ensuring teams are not torn apart by AI teammates*

The creation of effective AI teammates, and thus effective human-agent teams, cannot be achieved without a strong understanding and consideration for human teammates and the relationships between them. The results of this study demonstrate that the impacts of AI teammate design on human-agent relationships have the potential to affect human-human and team level relationships.

Specifically, the results of this study showed that etiquette, a behavior often seen as appropriate in humans, can be perceived as a toxic behavior when exhibited by AI systems, and in turn harm perceptions about AI teammates, teams as a whole, and even human teammates due to AI teammates being used as a comparison standard by some. Furthermore, the qualitative analysis revealed that humans would be willing to put up with these behaviors that they perceive as toxic if it meant benefiting from the teammate's perceived performance. Therefore, while an AI teammate may be an identifiable source of toxicity, humans may not utilize the perception of that toxicity when choosing whether to work with an AI teammate, as evidenced by the quantitative results. Thus, a perceived toxic environment would continue to exist, which is shown to negatively impact teamwork and group environments (Adinolf & Turkay, 2018; Türkay et al., 2020). Furthermore, the negative impacts of this perceived toxicity could start to drive apart human teammates, and repairing these relationships, especially in regard to trust, can prove to be highly difficult (Baker et al., 2018; de Visser et al., 2017; Gillespie & Dietz, 2009). This potential to drive teams apart is evidenced by this study's results, which demonstrate the potential use of AI teammates as a comparison point against humans. Thus, a goal of AI teammate design should be to ensure AI teammates, while AI teammates are viewed as high quality, that perception does not impact the perception of human teammates. Moving forward, this design goal will become more important as humans continue to grow attached to their technology (Homewood, 2016), AI teammates included.

These findings lend credence to the idea of intentionally creating AI teammates that are specifically designed to promote positive interaction, happiness, and support overall team cohesion. Based on the qualitative results of this work, direct communication style, time sensitivity, and goal-oriented behaviors that are components of the etiquette ignoring teammate would still be beneficial to teams and their perceptions. However, behaviors of assertiveness that diminish human accomplishment, such as imposing on a human's role or the use of impolite language, should be set aside as to not harm teaming relationships. Thus, when viewing etiquette as a gradient with varying levels, AI teammates that exhibit traditional etiquette would have a lower chance of damaging these teaming perceptions, but increasing productivity within a prescribed task and role would be welcomed.

However, to ensure AI teammate designs correctly leverage etiquette, designers and practitioners cannot take a one-dimensional view of performance that only considers task-level success. Group and teaming technology has been shown to provide benefits to human relationships, especially long-term relationships, if designed with humans in mind (Neustaedter & Greenberg, 2012; Shklovski et al., 2008). AI teammates should not be an exception to the rule. Therefore, the following conclusion can be reached: the design and integration process of AI teammates needs to pay as much attention to human-human relationships as it does to human-agent relationships. If human-human relationships deteriorate, the effectiveness of the AI teammate would be moot as the team would cease to function and most likely exist.

5.3. *Design recommendations for AI teammates*

Based on the results and insights brought forth by this study, we propose the following recommendations regarding the design of etiquette for AI teammates in human-agent teams.

Rec 1: The Etiquette Used by an AI Teammate Should be Customizable by its Team. While human perceptions of AI teammates ignoring etiquette were ultimately worse than that of AI teammates adhering to tradition etiquette, that does not mean that each team and individual disliked the etiquette ignoring teammate. In fact, multiple participants expressed that the lack of etiquette exhibited by the AI teammate was warranted due to the tight time restraint the team was given. For instance, participants 81 and 80 explicitly complimented the pacing of the teammate in regard to making the task more achievable in the short time frame. Additionally, other participants noted the context and being a major consideration for their continued use of their AI teammate.

Therefore, etiquette should be a variable that can be determined by a team based on their context, needs, and existing etiquette. For instance, tasks that are heavily time reliant, such as those in medical or military scenarios, would potentially see teams use AI teammates that forgo turn-taking, which could be essential in contexts similar to manufacturing or military where pacing and human-agent teaming are important (Schelble et al., 2020). However, it is important to note that this decision may not be easy to make as multiple humans with various opinions exist within teams. Thus, rather than having designers predict the ideal etiquette for a team, AI teammates should be customizable so the team can come to a collected design decision.

Rec 2: Each Time an AI Teammate Does not Act with Etiquette, it Should Provide an Explanation. Explainability, which is a trending topic in the HCI domain, has been linked to improving user experiences in human-AI interaction (Liao et al., 2020), and recent research has shown that it can be critical in preventing trust from declining when AI teammates violate human expectations (Lyons et al., 2023). Within this study, some participants noted the lack of explainability as a reason to not accept the suggestions made by the AI teammate that ignored etiquette. For instance, participant 8's quote demonstrates an explicit desire for explanations, but only when the AI teammates are working on a human's role:"...the bot did not explain why it chose that specific hotel and restaurant." While some participants were accepting of information from their AI teammate simply due to its shared team goal, it is clear that team goal commonalities are not enough for all humans to accept assertiveness from their AI teammate.

Thus, based on the results of this study, *each* time an AI teammate acts in a way that does not adhere to its assigned etiquette it should be accompanied by an explanation or justification. For instance, since this study's etiquette ignoring AI go against the will of their teammates, it is increasingly important that AI systems are not just explainable but actively utilize explainability to make humans more accepting of etiquette ignorance. In practice, this recommendation could consist of a simple message after their suggestion as to why a suggestion was chosen. However, the use of this explainability may look differently based on the context in which it is utilized, or other behavioral design considerations present in an AI teammate, such as politeness. For instance, behaviors in UAV environments may require automated pilots that divert from a planned course, and they should provide a more graphical justification based on the data created by human teammates. However, regardless of the domain, the results of this study make it clear that challenging human work, which might not be a behavior within the etiquette assigned to an AI teammate, requires specific justifications and explanations if teammates going to be consistently accepted.

Rec 3: Input from AI Teammates Should Improve Their Human Teammates' Work not Override it. While the design of etiquette ignorance within this study considers has AI teammates act outside their assigned role (Dautenhahn et al., 2006), that does not mean that AI systems should be designed to override human influence. Multiple participants in this study were not accepting of their etiquette, ignoring AI teammates since they felt it had diminished their role as a teammate. Participant 55 is a clear example of this as they said "…he was aggressive and I felt useless to the group and task." In this instance, the ignorance of traditional etiquette results in an AI teammate attempting to override human contribution. In essence, it was not the high performance of the AI those participants were complaining about but rather the lack of contribution they were able to make.

Thus, unlike the AI teammate utilized in this study, the language utilized by AI teammates that explicitly critiques or comments on a human's work should focus on improving and modifying their work rather than simply overriding it with a better answer. The implementation of this recommendation means that two different communication models would need to exist for AI systems. First, a communication model would need to be implemented for updating humans on the completion and progress of its specific role, which might have its own etiquette toward language. However, a second communication model should be implemented to help iteratively guide humans to better answers rather than simply suggest a better response, thus overriding and undermining their effort. When this recommendation is taken into account alongside Rec 1 and Rec 2, the result is an AI teammate who is highly performative in its own role while still being enthusiastic toward the iterative improvement of its teammates' work, thus improving the ultimate acceptability and utility of an AI teammate.

5.4. *Limitations and future work*

The current study has three limitations, which can serve as motivation for future research directions. Potential limitations surround the usage of only two experimental groups, the scripted AI teammate and the task's context. In regards to having only two experimental groups, individual etiquette behaviors could and should be examined in isolation, but this would make the results much more specific to the chosen context. Additionally, the simplistic design of this study provides an apparent limitation as violations of etiquette may actually be appropriate and accepted if used intelligently. For instance, high-stakes and fast-paced environments may necessitate and merit AI teammates' interruptions, and these violations in traditional etiquette may actually benefit trust rather than harm it. As such, future research should examine how to best design the violation of traditional etiquette when the context is a consideration with the goal of making AI teammates more equipped to handle specific contexts.

Another limitation of this experiment is the scripted nature of the AI teammate. The use of real, context-specific AI teammates could lead to different findings as accuracy may not be ensured and AI teammates may have greater responsiveness to human teammates (Chiou & Lee, 2021). While this study does not look at the interaction of the actual performance of an AI teammate using etiquette, future research should build on this by observing how the accuracy and performance of an AI teammate may interact with etiquette. Additionally, the scripted design in this experiment does not allow a large variety of conversational interactions. This is an added benefit for experimental control, but also a limitation to replicating real AI interaction. Future research, especially research that focuses on longer-term interaction or communication, may find it beneficial to utilize conversational AI, which could help design AI for team growth in the long-term.

The other important limitation to this study is the context of the chosen task. The task utilized is a short-term task, with objectively correct answers, in a workplace setting, and benefits from teammates that can quickly process information, such as AI and machine teammates. While the task does allow choices to be made, and a variety of answers exist, little room for human expertise and nuance is in the task. Future research could look at a more subjective task, where a human's input may be more valued than an AI's. Conducting this study in an unfamiliar context of human-agent interaction is important and should be done in future research as agent etiquette should handle unfamiliarity. Part of this limitation is also the task's short-term nature and the limited demographics observed. Specifically, the young nature of all the participants that made up the sample. There are several potential differences between younger, middle-aged, and older populations that could influence the results of the current study, and this should be taken into account when interpreting these findings. Cultural differences may also have a similar moderating effect on these results as cultural considerations for etiquette are important. These limitations represent valuable future research directions as additional studies continue to contextualize the relationship between AI and human teammates. While these limitations exist within our experiment, the results found in this study will serve to guide future researchers and human-agent teams that conduct more context-specific research.

6. Conclusion

This study utilizes a human-centered perspective to examine the potential impact of incorporating AI etiquette within human-agent teams. Specifically, AI teammates that adhere to traditional etiquette standards can improve perceived teammate performance, trust, and perceived team success. However, these benefits were not found to extend to humans desire to continue working with AI teammates. Moreover, the qualitative results show that while marginal benefits are statistically significant, the impact of AI etiquette can impact from person to person in a team. Given these findings, specific discussion and design recommendations were provided to ensure that etiquette plays a role in human-agent teams, even if that role needs to be highly customizable to the individual. AI design will require a comprehensive understanding of how designing AI behaviors has the potential to overshadow technical performance and how specific behaviors, must be designed and incorporated effectively. With such an understanding, HCI and AI researchers and practitioners will have the ability to design AI systems that impact the real-world positively through human-agent teaming.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This material is based upon work supported by the National Science Foundation under Grant No. 1829008.

HCI editorial record

First received on *7/5/2022*. Revisions received on *date*, *date*, and *date*. Accepted by *action-editor-name*. Final manuscript received on *date*.

Notes on contributors

Christopher Flathmann (cflathm@clemson.eduhttps://chrisflathmann.com/) is a researcher with an interest in human-AI teaming; he is a research assistant professor in the School of Computing of Clemson University.

Nathan J. McNeese (mcneese@clemson.eduhttps://nathanmcneese.weebly.com/) is a researcher with an interest in human-AI teaming; he is a dean's professor in the School of Computing of Clemson University.

Beau G. Schelble (bschelb@clemson.edu,) is a researcher with an interest in human-AI teaming; he is a Ph.D. student in the School of Computing of Clemson University.

Bart Knijnenburg (bartk@clemson.eduhttps://www.usabart.nl) is a researcher with an interest in privacy decision-making and recommender systems; he is an assistant professor in the School of Computing of Clemson University.

Guo Freeman (freeman@clemson.eduhttps://computing.clemson.edu/cugame/index.html) is a researcher with an interest in online communities; she is an assistant professor in the School of Computing of Clemson University.

ORCID

Christopher Flathmann http://orcid.org/0000-0002-5448-2610
Nathan J. McNeese http://orcid.org/0000-0002-9143-2460
Beau Schelble http://orcid.org/0000-0003-3704-697X
Bart Knijnenburg http://orcid.org/0000-0003-1341-0669
Guo Freeman http://orcid.org/0000-0001-5107-7794

References

Abdelwahab, O., & Elmaghraby, A. (2018, July). Deep learning based vs. Markov chain based text generation for cross domain adaptation for sentiment classification. In *2018 IEEE International Conference on Information Reuse and Integration (IRI)* (pp. 252–255). https://doi.org/10.1109/IRI.2018.00046

Adinolf, S., & Turkay, S. (2018, October). Toxic behaviors in esports games: Player perceptions and coping strategies. In *Proceedings of the 2018 Annual Symposium on Computer-Human Interaction in Play Companion Extended Abstracts* (pp. 365–372). New York, USA: Association for Computing Machinery. https://doi.org/10.1145/3270316.3271545

Alan, A. T., Costanza, E., Ramchurn, S. D., Fischer, J., Rodden, T., & Jennings, N. R. (2016). Tariff agent: Interacting with a future smart energy system at home. *ACM Transactions on Computer- Human Interaction (TOCHI)*, *23*(4), 1–28. https://doi.org/10.1145/2943770

Atoyan, H., Duquet, J. -R., & Robert, J. -M. (2006). Trust in new decision aid systems. In *Proceedings of the 18th conference on l'interaction homme-machine*, Montreal Canada, (pp. 115–122).

Babel, F., Kraus, J. M., & Baumann, M. (2021). Development and testing of psychological conflict resolution strategies for assertive robots to resolve human–robot goal conflict. *Frontiers in Robotics and AI*, *7*. https://doi.org/10.3389/frobt.2020.591448

Baker, A. L., Phillips, E. K., Ullman, D., & Keebler, J. R. (2018, November). Toward an understanding of trust repair in human-robot interaction: Current research and future directions. *ACM Transactions on Interactive Intelligent Systems*, *8*(4), 30:1–30:30. https://doi.org/10.1145/3181671

Bansal, G., Nushi, B., Kamar, E., Lasecki, W. S., Weld, D. S., & Horvitz, E. (2019). Beyond accuracy: The role of mental models in human-ai team performance. In *Proceedings of the aaai conference on human computation and crowdsourcing*, Stevenson Washington, USA, (Vol. 7, pp. 2–11).

Bergstrom, N., Kanda, T., Miyashita, T., Ishiguro, H., & Hagita, N. (2008). Modeling of natural human-robot encounters. In *2008 ieee/rsj international conference on intelligent robots and systems*, Madrid, Spain, (pp. 2623–2629).

Bhatti, S., Demir, M., Cooke, N. J., & Johnson, C. J. (2021). Assessing communication and trust in an ai teammate in a dynamic task environment. In *2021 ieee 2nd international conference on human-machine systems (ichms)*, Magdeburg, Germany, (pp. 1–6).

Bickmore, T., & Cassell, J. (2005). Social dialongue with embodied conversational agents. In van Kuppevel, J, Dybkjaer, L, & Bernsen, N. (eds.). *Advances in natural multimodal dialogue systems*, (pp. 23–54). Springer.

Bliese, P. D. (2000). Within-group agreement, non-independence, and reliability: Implications for data aggregation and analysis.

Braun, V., & Clarke, V. (2012). *Thematic analysis*. American Psychological Association.

Chakraborti, T., Kambhampati, S., Scheutz, M., & Zhang, Y. (2017). Ai challenges in human-robot cognitive teaming. *arXiv preprint arXiv 170704775*.

Chen, J. Y. C., & Barnes, M. J. (2014, February). Human–agent teaming for multirobot control: A review of human factors issues. *IEEE Transactions on Human-Machine Systems*, *44*(1), 13–29. https://doi.org/10.1109/THMS.2013.2293535

Chiou, E. K., Demir, M., Buchanan, V., Corral, C. C., Endsley, M. R., Lematta, G. J., McNeese, N. J. (2022). 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. https://doi.org/10.1007/s12369-021-00834-1

Chiou, E. K., & Lee, J. D. (2021). Trusting automation: Designing for responsivity and resilience. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *65*(1), 00187208211009995. https://doi.org/10.1177/00187208211009995

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences.* Academic press.

Culley, K. E., & Madhavan, P. (2013). A note of caution regarding anthropomorphism in hci agents. *Computers in Human Behavior*, *29*(3), 577–579. https://doi.org/10.1016/j.chb.2012.11.023

Cummings, L. L., & Bromiley, P. (1996). The organizational trust inventory (oti). *Trust in Organizations: Frontiers of Theory and Research*, *302*(330), 39–52.

Dautenhahn, K., Walters, M., Woods, S., Koay, K. L., Nehaniv, C. L., Sisbot, A., Siméon, T. (2006). How may i serve you? a robot companion approaching a seated person in a helping context. In *Proceedings of the 1st acm sigchi/sigart conference on human-robot interaction*, Salt Lake City, Utah, USA. (pp. 172–179).

Demir, M., McNeese, N. J., Cooke, N. J., Ball, J. T., Myers, C., & Frieman, M. (2015, September). Synthetic Teammate Communication and Coordination with Humans. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *59*(1), 951–955. Publisher: SAGE Publications Inc https://doi.org/10.1177/1541931215591275

de Visser, E. J., Pak, R., & Neerincx, M. A. (2017, March). Trust Development and Repair in Human- Robot Teams. In *Proceedings of the Companion of the 2017 ACM/IEEE International Conference on Human-Robot Interaction* (pp. 103–104). New York, NY, USA: Association for Computing Machinery. https://doi.org/10.1145/3029798.3038409

Dillon, A., & Watson, C. (1996). User analysis in hci—the historical lessons from individual differences research. *International Journal of Human-Computer Studies*, *45*(6), 619–637. https://doi.org/10.1006/ijhc.1996.0071

Dzindolet, M. T., Peterson, S. A., Pomranky, R. A., Pierce, L. G., & Beck, H. P. (2003). The role of trust in automation reliance. *International Journal of Human-Computer Studies*, *58*(6), 697–718. https://doi.org/10.1016/S1071-5819(03)00038-7

Ezer, N., Bruni, S., Cai, Y., Hepenstal, S. J., Miller, C. A., & Schmorrow, D. D. (2019). Trust engineering for human-ai teams. In *Proceedings of the human factors and ergonomics society annual meeting*, Seattle, Washington, USA, (Vol. 63, pp. 322–326).

Flathmann, C., McNeese, N., & Canonico, L. B. (2019). Using human-agent teams to purposefully design multi-agent systems. In *Proceedings of the human factors and ergonomics society annual meeting*, Seattle, Washington, USA, (Vol. 63, pp. 1425–1429).

Flathmann, C., Schelble, B. G., Zhang, R., & McNeese, N. J. (2021). Modeling and guiding the creation of ethical human-ai teams. In *Proceedings of the 2021 aaai/acm conference on ai, ethics, and society* (pp. 469–479).

Følstad, A., Nordheim, C. B., & Bjørkli, C. A. (2018). What Makes Users Trust a Chatbot for Customer Service? An Exploratory Interview Study. In S. S. Bodrunova (Ed.), *Internet Science* (pp. 194–208). Springer International Publishing. https://doi.org/10.1007/978-3-030-01437-716

Gavin, H. (2008). Thematic analysis. *Understanding Research Methods and Statistics in Psychology*, 273–282.

Gillespie, N., & Dietz, G. (2009, January), Trust Repair After an Organization-Level Failure. *Academy of Management Review*, *34*(1), 127–145. Publisher: Academy of Management https://doi.org/10.5465/amr.2009.35713319

Gombolay, M., Bair, A., Huang, C., & Shah, J. (2017, June), Computational design of mixed- initiative human–robot teaming that considers human factors: Situational awareness, work- load, and workflow preferences. *The International Journal of Robotics Research*, *36*(5–7), 597–617. Publisher: SAGE Publications Ltd STMhttps://doi.org/10.1177/0278364916688255

Goodrich, M. A., Morse, B. S., Gerhardt, D., Cooper, J. L., Quigley, M., Adams, J. A., & Humphrey, C. (2008). Supporting wilderness search and rescue using a camera-equipped mini uav. *Journal of Field Robotics*, *25*(1–2), 89–110. https://doi.org/10.1002/rob.20226

Grawitch, M. J., & Munz, D. C. (2004). Are your data nonindependent? a practical guide to evaluating nonindependence and within-group agreement. *Understanding Statistics*, *3*(4), 231–257. https://doi.org/10.1207/s15328031us0304_2

Grover, S., Sengupta, S., Chakraborti, T., Mishra, A. P., & Kambhampati, S. (2020, November), RADAR: Automated task planning for proactive decision support. *Human–Computer Interaction*, *35*(5–6), 387–412. Publisher: Taylor & Francis eprint https://doi.org/10.1080/07370024.2020.1726751

Haring, K. S., Satterfield, K. M., Tossell, C. C., De Visser, E. J., Lyons, J. R., Mancuso, V. F., Funke, G. J. (2021). Robot authority in human-robot teaming: Effects of human-likeness and physical embodiment on compliance. *Frontiers in Psychology*, *12*, 625713. https://doi.org/10.3389/fpsyg.2021.625713

Hart, S. G., & Staveland, L. E. (1988, January). Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. In P. A. Hancock & N. Meshkati Eds., *Advances in Psychology* Vol. 52 (pp. 139–183). North-Holland. https://doi.org/10.1016/S0166-41150862386-9

Hinds, P. J., Roberts, T. L., & Jones, H. (2004). Whose job is it anyway? a study of human-robot interaction in a collaborative task. *Human–Computer Interaction*, *19*(1–2), 151–181.

Homewood, S. (2016, May). Maintaining relationships with our devices. In *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems* (pp. 273–276). New York, NY, USA: Association for Computing Machinery. https://doi.org/10.1145/2851581.2889467

Horwitz, S. K. (2005). The compositional impact of team diversity on performance: Theoretical considerations. *Human Resource Development Review*, *4*(2), 219–245. https://doi.org/10.1177/1534484305275847

Horwitz, S. K., & Horwitz, I. B. (2007). The effects of team diversity on team outcomes: A meta- analytic review of team demography. *Journal of Management*, *33*(6), 987–1015. https://doi.org/10.1177/0149206307308587

Inkpen, K., Chancellor, S., De Choudhury, M., Veale, M., & Baumer, E. P. S. (2019, May). Where is the human? Bridging the gap between AI and HCI. In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems* (pp. 1–9). New York, NY, USA: Association for Computing Machinery. https://doi.org/10.1145/3290607.3299002

James, L. R., Demaree, R. G., & Wolf, G. (1984). Estimating within-group interrater reliability with and without response bias. *The Journal of Applied Psychology*, *69*(1), 85. https://doi.org/10.1037/0021-9010.69.1.85

Jones, H., & Hinds, P. (2002). Extreme work teams: Using swat teams as a model for coordinating distributed robots. In *Proceedings of the 2002 acm conference on computer supported cooperative work*, New Orleans, Louisiana, USA, (pp. 372–381).

Jung, M. F., Martelaro, N., & Hinds, P. J. (2015). Using robots to moderate team conflict: The case of repairing violations. In *Proceedings of the tenth annual acm/ieee international conference on human-robot interaction*, Portland, Oregon, USA, (pp. 229–236).

Karreman, D., Utama, L., Joosse, M., Lohse, M., van Dijk, B., & Evers, V. (2014). Robot etiquette: How to approach a pair of people? In *Proceedings of the 2014 acm/ieee international conference on human-robot interaction*, Bielefeld, Germany, (pp. 196–197).

Koay, K. L., Walters, M. L., May, A., Dumitriu, A., Christianson, B., Burke, N., & Dautenhahn, K. (2013). Exploring robot etiquette: Refining a hri home companion scenario based on feedback from two artists who lived with robots in the uh robot house. In *International conference on social robotics*, Bristol, UK, (pp. 290–300).

Law, T., & Scheutz, M. (2021). Trust: Recent concepts and evaluations in human-robot interaction. In Nam, C., & Lyons, J. *Trust in human-robot interaction*, (pp. 27–57). Elsevier.

Liao, Q. V., Gruen, D., & Miller, S. (2020, April). Questioning the AI: Informing design practices for explainable ai user experiences. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (pp. 1–15). New York, NY, USA: Association for Computing Machinery. https://doi.org/10.1145/3313831.3376590

Looi, Q. E., & See, S. L. (2012). Applying politeness maxims in social robotics polite dialogue. In *Proceedings of the seventh annual acm/ieee international conference on human-robot interaction*, Boston, Massachusetts, USA, (pp. 189–190).

Lounsbury, J. W., Smith, R. M., Levy, J. J., Leong, F. T., & Gibson, L. W. (2009, March). Personality characteristics of business majors as defined by the big five and narrow personality traits. *Journal of Education for Business*, *84*(4), 200–205. https://doi.org/10.3200/JOEB.84.4.200-205

Lucas, G. M., Gratch, J., King, A., & Morency, L. -P. (2014). It's only a computer: Virtual humans increase willingness to disclose. *Computers in Human Behavior*, *37*, 94–100. https://doi.org/10.1016/j.chb.2014.04.043

Lyons, J. B., Aldin Hamdan, I., & Vo, T. Q. (2023). Explanations and trust: What happens to trust when a robot partner does something unexpected? *Computers in Human Behavior*, *138*, 107473. https://doi.org/10.1016/j.chb.2022.107473

Madhavan, P., & Wiegmann, D. A. (2007). Similarities and differences between human–human and human–automation trust: An integrative review. *Theoretical Issues in Ergonomics Science*, *8*(4), 277–301. https://doi.org/10.1080/14639220500337708

McNeese, N., Demir, M., Chiou, E., Cooke, N., & Yanikian, G. (2019). Understanding the role of trust in human-autonomy teaming. Hawaii International Conference on Social Systems, Maui, HI, USA. https://doi.org/10.24251/HICSS.2019.032

McNeese, N. J., Demir, M., Cooke, N. J., & Myers, C. (2018, March). Teaming with a Synthetic Teammate: Insights into Human-Autonomy Teaming. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *60*(2), 262–273. Publisher: SAGE Publications Inc https://doi.org/10.1177/0018720817743223

McNeese, N. J., Schelble, B. G., Canonico, L. B., & Demir, M. (2021). Who/What is my teammate? team composition considerations in human–ai te aming. *IEEE Transactions on Human-Machine Systems*, *51*(4), 288–299. https://doi.org/10.1109/THMS.2021.3086018

Meyer, J., Miller, C., Hancock, P., De Visser, E. J., & Dorneich, M. (2016). Politeness in machine- human and human-human interaction. In *Proceedings of the human factors and ergonomics society annual meeting*, Washington, DC, USA, (Vol. 60, pp. 279–283).

Miller, C. A. (2002). Definitions and dimensions of etiquette. In *Proc. aaai fall symposium on etiquette and human-computer work*. North Falmount, MA, USA.

Miller, C. A. (2005). Trust in adaptive automation: The role of etiquette in tuning trust via analogic and affective methods. In *Proceedings of the 1st international conference on augmented cognition* (pp. 22–27).

Miller, C. A. (2011). Relationships and etiquette with technical systems. In Whitworth, B., & de Moor, A.*Virtual communities: Concepts, methodologies, tools and applications*, (pp. 1618–1633). IGI Global.

Mishra, P., & Hershey, K. A. (2004). Etiquette and the design of educational technology. *Communications of the ACM*, *47*(4), 45–49. https://doi.org/10.1145/975817.975843

Moreno, K. N., Person, N. K., Adcock, A. B., Eck, R., Jackson, G. T., & Marineau, J. C. (2002). Etiquette and efficacy in animated pedagogical agents: The role of stereotypes. *Aaai symposium on personalized agents, cape cod*, Cape Cod, MA, USA, *ma*.

Musick, G., O'neill, T. A., Schelble, B. G., McNeese, N. J., & Henke, J. B. (2021). What happens when humans believe their teammate is an ai? an investigation into humans teaming with autonomy. *Computers in Human Behavior*, *122*, 106852. https://doi.org/10.1016/j.chb.2021.106852

Neustaedter, C., & Greenberg, S. (2012, May). Intimacy in long-distance relationships over video chat. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 753–762). New York, NY, USA: Association for Computing Machinery. https://doi.org/10.1145/2207676.2207785

Ogden, B., & Dautenhahn, K. (2000). Robotic etiquette: Structured interaction in humans and robots. In *Procs sirs 2000, 8th symp on intelligent robotic systems* Reading, UK.

O'neill, T., McNeese, N., Barron, A., & Schelble, B. (2022). Human–autonomy teaming: A review and analysis of the empirical literature. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *64*(5), 904–938. https://doi.org/10.1177/0018720820960865

Parasuraman, R., & Miller, C. A. (2004). Trust and etiquette in high-criticality automated systems. *Communications of the ACM*, *47*(4), 51–55. https://doi.org/10.1145/975817.975844

Parasuraman, R., & Riley, V. (1997, June). Humans and automation: Use, misuse, disuse, abuse. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *39*(2), 230–253. Publisher: SAGE Publications Inc https://doi.org/10.1518/001872097778543886

Pipitone, A., Geraci, A., D'amico, A., Seidita, V., & Chella, A. (2021). Robot's inner speech effects on trust and anthropomorphic cues in human-robot cooperation. *arXiv preprint arXiv 210909388*.

Ramos, G., Meek, C., Simard, P., Suh, J., & Ghorashi, S. (2020, November). Interactive machine teaching: A human-centered approach to building machine-learned models. *Human–Computer Interaction*, *35*(5–6), 413–451. https://doi.org/10.1080/07370024.2020.1734931

Rentsch, J. R., & Klimoski, R. J. (2001). Why do 'great minds' think alike?: Antecedents of team member schema agreement. *Journal of Organizational Behavior*, *22*(2), 107–120. https://doi.org/10.1002/job.81

Salas, E., Dickinson, T. L., Converse, S. A., & Tannenbaum, S. I. (1992). Toward an Understanding of Team Performance and Training. In Sweezy, R. W., Salas, E. (eds.). Teams: Their Training and Performance.

Salas, E., Sims, D. E., & Burke, C. S. (2005, October). Is there a "Big Five" in Teamwork? *Small Group Research*, *36*(5), 555–599. Publisher: SAGE Publications Inc https://doi.org/10.1177/1046496405277134

Schelble, B. G., Flathmann, C., & McNeese, N. (2020). Towards meaningfully integrating human- autonomy teaming in applied settings. In *Proceedings of the 8th international conference on human-agent interaction*, Melbourne, Australia, (pp. 149–156).

Schiaffino, S., & Amandi, A. (2006). Polite personal agents. *IEEE Intelligent Systems*, *21*(1), 12–19. https://doi.org/10.1109/MIS.2006.15

Schulte, A., Donath, D., & Lange, D. S. (2016). Design Patterns for Human-Cognitive Agent Teaming. In D. Harris (Ed.), *Engineering Psychology and Cognitive Ergonomics* (pp. 231–243). Springer International Publishing. https://doi.org/10.1007/978-3-319-40030-324

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

Shklovski, I., Kraut, R., & Cummings, J. (2008, April). Keeping in touch by technology: Maintaining friendships after a residential move. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 807–816). New York, NY, USA: Association for Computing Machinery. https://doi.org/10.1145/1357054.1357182.

Sierhuis, M., Bradshaw, J. M., Acquisti, R., Van Hoof, R., & Jeffers, R. (2003). Human-agent teamwork and adjustable autonomy in practice. In *Proceedings of the Seventh International Symposium on Artificial Intelligence, Robotics and Automation in Space (i-SAIRAS)*. Nara, Japan, Citeseer.

Sin, J., & Munteanu, C. (2020, November). An empirically grounded sociotechnical perspective on designing virtual agents for older adults. *Human–Computer Interaction*, *35*(5–6), 481–510. Publisher: Taylor & Francis https://doi.org/10.1080/07370024.2020.1731690

Skantze, G. (2021). Turn-taking in conversational systems and human-robot interaction: A review. *Computer Speech & Language*, *67*, 101178. https://doi.org/10.1016/j.csl.2020.101178

Svikhnushina, E., & Pu, P. (2020). Social and emotional etiquette of chatbots: A qualitative approach to understanding user needs and expectations. *arXiv preprint arXiv 200613883*.

Textor, C., Zhang, R., Lopez, J., Schelble, B. G., McNeese, N. J., Freeman, G., de Visser, E. J. (2022). Exploring the relationship between ethics and trust in human–artificial intelligence teaming: A mixed methods approach. *Journal of Cognitive Engineering and Decision Making*, *16*(4), 252–281. https://doi.org/10.1177/15553434221113964

Thomas, J., & Vaughan, R. (2018, October). AfteR you: Doorway negotiation for human-robot and robot-robot interaction. In *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)* (pp. 3387–3394). (ISSN: 2153-0866) https://doi.org/10.1109/IROS.2018.8594034

Thomas, J., & Vaughan, R. (2019, November). Right of way, assertiveness and social recognition in human-robot doorway interaction. In *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)* (pp. 333–339). (ISSN: 2153-0866) https://doi.org/10.1109/IROS40897.2019.8967862

Traeger, M. L., Strohkorb Sebo, S., Jung, M., Scassellati, B., & Christakis, N. A. (2020). Vulnerable robots positively shape human conversational dynamics in a human–robot team. *Proceedings of the National Academy of Sciences*, *117*(12), 6370–6375.

Trinh, T. Q., Schroeter, C., Kessler, J., & Gross, H. -M. (2015). "Go ahead, please": Recognition and resolution of conflict situations in narrow passages for polite mobile robot navigation. In *International conference on social robotics*, Paris, France, (pp. 643–653).

Türkay, S., Formosa, J., Adinolf, S., Cuthbert, R., & Altizer, R. (2020, April). SeE no evil, hear no evil, speak no evil: How collegiate players define, experience and cope with toxicity. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (pp. 1–13). New York, USA: Association for Computing Machinery. https://doi.org/10.1145/3313831.3376191

Van Pinxteren, M. M., Pluymaekers, M., & Lemmink, J. G. (2020). Human-like communication in conversational agents: A literature review and research agenda. *Journal of Service Management*, *31*(2), 203–225. https://doi.org/10.1108/JOSM-06-2019-0175

Walters, M. L., Dautenhahn, K., Woods, S. N., & Koay, K. L. (2007). Robotic etiquette: Results from user studies involving a fetch and carry task. In *2007 2nd acm/ieee international conference on human-robot interaction (hri)*, Arlington, Virginia, USA. (pp. 317–324).

Wang, Y., Murray, R. C., Bao, H., & Rose, C. (2020). Agent-based dynamic collaboration support in a smart office space. In *Proceedings of the 21th annual meeting of the special interest group on discourse and dialogue*. Virtual. (pp. 257–260).

Wiltshire, T. J., Barber, D., & Fiore, S. M. (2013). Towards modeling social-cognitive mechanisms in robots to facilitate human-robot teaming. In *Proceedings of the human factors and ergonomics society annual meeting* (Vol. 57, pp. 1278–1282). SAGE Publications Sage CA: Los Angeles, CA. https://doi.org/10.1177/1541931213571283

Xu, W. (2019, June). Toward human-centered AI: A perspective from human-computer interaction. *Interactions*, *26* (4), 42–46. https://doi.org/10.1145/3328485

Zhang, T., Zhu, B., & Kaber, D. B. (2010). Setting etiquette expectations Hayes, C., Miller , C. (eds.). *Human-Computer Etiquette*, (pp. 231–259). CRC Press.

Zheng, Y. (2019, September). A course on applied ai and data science: Recommender systems. In *Proceedings of the 20th Annual SIG Conference on Information Technology Education* (pp. 43–48). New York, USA: Association for Computing Machinery. https://doi.org/10.1145/3349266.3351405