Collective Attention in Human-AI Teams

Josie Zvelebilova[1], Saiph Savage[1], and Christoph Riedl[1,*]

[1]Northeastern University, Boston, MA
[*]c.riedl@northeastern.edu

ABSTRACT

How does the presence of an AI assistant affect the collective attention of a team? We study 20 human teams of 3-4 individuals paired with one voice-only AI assistant during a challenging puzzle task. Teams are randomly assigned to an AI assistant with a human- or robotic-sounding voice that provides either helpful or misleading information about the task. Treating each individual AI interjection as a treatment intervention, we identify the causal effects of the AI on dynamic group processes involving language use. Our findings demonstrate that the AI significantly affects *what* teams discuss, *how* they discuss it, and the alignment of their mental models. Teams adopt AI-introduced language for both terms directly related to the task and for peripheral terms, even when they (a) recognize the unhelpful nature of the AI, (b) do not consider the AI a genuine team member, and (c) do not trust the AI. The process of language adaptation appears to be automatic, despite doubts about the AI's competence. The presence of an AI assistant significantly impacts team collective attention by modulating various aspects of shared cognition. This study contributes to human-AI teaming research by highlighting collective attention as a central mechanism through which AI systems in team settings influence team performance. Understanding this mechanism will help CSCW researchers design AI systems that enhance team collective intelligence by optimizing collective attention.

1 Introduction

The coordinated attention of team members is fundamental to effective collaboration [1]. Collective attention is manifested in various forms of cognitive alignment that together define how teams cooperatively focus, divide, sustain, and alternate their attention to team-related activities to accomplish team goals. It encompasses the alignment of both the timing and content of attention [2]. For example, alignment on shared terminology within a team reflects coordinated attentional priorities that facilitate the synchronization of collective focus.

With the rise of human-AI teaming, a critical question arises: How does the presence of an AI assistant affect cognitive alignment in human teams? While cognitive alignment is widely recognized as a key driver of team performance [3], it may fail to materialize in human-AI teams due to fundamental differences between human-human and human-AI interactions [4–6]. Despite their growing importance, the dynamics and implications of collaborative processes in human-AI teams remain under-explored [6–8]. Understanding the impact of AI systems on team processes beyond that of a single human-AI dyad is vital for informing computer-supported cooperative work (CSCW) research on the design of human-AI systems that facilitate productive collaboration and enhance the collective intelligence of human teams [1, 8–10].

Cognitive alignment in teams is a multi-faceted concept closely intertwined with alignment on shared language [11, 12]. Convergence on common terminology signifies that groups have established efficient meta-cognition, a prerequisite for effective group processes. Through this lexical alignment, team members can effectively direct each other's attention [13]. Alignment on a shared language also affects what is salient and attended to within a team [14]. A shared language is not only an indicator of cognitive alignment, but may in fact be the primary mechanism through which humans achieve cognitive alignment [12] and thus coordinate collective attention.

AI systems assume increasingly prominent roles in high-stakes collaborative contexts, such as emergency response and medical diagnosis. However, research on the impact of AI on team performance has yielded inconsistent results, with some studies finding increased performance and others finding decreased performance [e.g., 15–17]. To explain why AI systems are sometimes helpful and sometimes not—and, ultimately, to design more effective human-AI collaborative systems—it is crucial to better understand how AI affects known antecedents of team performance. Specifically, we need to examine how AI influences secondary outcomes like collective cognition (including shared language) that shape teams' collective attention [13, 18]. If AI systems provide quality input but undermine collective attention, they may fail to achieve their goals of enhancing team performance. Without a deeper understanding of how AI affects fundamental team processes of collective attention (and related components of collective memory and collective reasoning; [2]), it will be difficult to predict the impact of AI on team performance. Insight into these interactions can guide the design of AI systems that better support human teams, enhancing their effectiveness in complex tasks [9].

A distinguishing aspect of our study is its focus on human-human communication within AI-assisted teams. We examine groups of multiple interacting humans paired with a single team-based AI assistant, as opposed to individual human-AI dyads. This study design reflects the complexity and dynamics of real-world collaborative team environments [3], introducing additional layers of communication, coordination, and decision-making. By studying teams with multiple human members, we aim to paint a more comprehensive picture of how the presence of an AI assistant affects team communication dynamics (for studies with a similar focus on human-human interaction see, e.g., [19]). Throughout this work, we use the term "human-AI teams" to refer to groups of two or more humans collaborating with the help of an AI assistant. This is in contrast with other work that uses the term "human-AI teaming" to describe dyadic interactions between a single human and an AI system [e.g., 20–23].

Our study analyzes data from an IRB approved, randomized controlled trial of 20 teams of 3-4 individuals with 69 total human subjects. During a 40 minute group task, teams were joined by a voice-only AI assistant. We employed a 2×2 between-subject design, randomly assigning teams to four conditions that manipulated the AI assistant's *quality* (helpful vs. wrong) and its *voice* (human vs. robotic). Our investigation focuses on three aspects of collective attention spanning different timescales: (1) dynamic patterns in the timing of *what* teams discuss, (2) the shared language teams develop (i.e., *how* they discuss it), and (3) a static, task-level measure of shared mental model alignment. To shed light on the underlying mechanisms of cognitive alignment, we incorporate additional data sources and measures, including trust in AI, perception of the AI assistant as a teammate, and pronoun use.

We address the following research questions:

RQ1 Does the presence of an AI assistant affect the collective attention of human teams? (i.e., impact the timing of *what* teams talk about)

RQ2 Will teams adopt the language introduced by the AI? (i.e., impact *how* teams talk about objects)

RQ3 Does the effect of the AI differ based on its (a) quality and (b) humanness?

RQ4 What mechanisms can explain the observed effects?

We find strong causal evidence that the AI assistant shifts the collective attention of teams by shaping *what* they discuss and *how* they discuss it. Humans are 1.5 times as likely to discuss objects mentioned by the AI assistant compared to objects not mentioned by the AI. Humans also align their language use with the AI terms for core referents, which are the focus of task milestones. Unsurprisingly, the alignment is stronger when the AI is of higher quality and human-sounding. When the AI is of low quality, humans nonetheless adopt the AI's language around peripheral referents (neutral items for which the low quality of the AI's input is less obvious). This alignment happens despite teams being aware of the AI's low quality, despite teams not considering the AI a genuine team member, and despite having decided to not to trust it. Similar effects of the AI on cognitive alignment are reflected in team members' retrospective and reified perceptions of their own team processes. Overall, a consistent pattern emerges in which the "worst" AI (low quality and robotic sounding) has a surprising positive effect on cognitive alignment: being perceived as an "other," it is easy to exclude from the team, strengthening collective identity among the human teammates and thus facilitating cognitive alignment. Conversely, the "better" AI (low quality but human sounding) is difficult to exclude, causing confusion and doubt in teams, which results in poorly aligned mental models.

Our findings present intriguing insights into the dynamics of human interaction in team settings when they are influenced by an AI assistant. The AI assistant demonstrates a significant impact on team communication, even though team members do not directly interact with it. This suggests that the presence of an AI agent can direct the focus of collective attention in human teams, potentially steering discussions towards priorities that it identifies. Such influence on cognitive alignment indicates that the quality and character of the AI can substantially affect how humans adjust their communicative behavior to align with it, potentially enhancing team communication efficiency. The observed variations underscore how the perceived reliability and utility of the AI influence not only task-specific discussions but also the general communicative patterns and social dynamics within the team.

These results enhance our understanding of the way AI influences human team dynamics and performance in collaborative contexts. Our research demonstrates that the causal effect of AI interjections on lexical alignment in human teams is a complex, context-dependent phenomenon. This process is influenced by the perceived humanness and reliability of the AI, the relevance of its contributions, and the social expectations and communicative goals of the human teammates. Importantly, our findings indicate that humans may subconsciously adopt the language of an AI agent even while consciously doubting its competence. This tendency could potentially lead teams astray if the AI's contributions are unreliable, giving weight to the critical need for AI systems that are both accurate and contextually appropriate. The interplay between team members' lexical adjustments and their perceptions of the AI suggests that successful integration of AI into human teams requires a careful balance between the AI's capabilities and its presentation.

This paper contributes to the growing body of research on the social dynamics of human-AI interaction and highlights the need for further interdisciplinary work to elucidate the cognitive and communicative mechanisms underlying coordination in mixed human-machine teams. This research informs the design of AI systems by identifying collective attention as a central mechanism through which AI systems in team settings influence team performance. In summary, this paper makes three main contributions to the field of CSCW, specifically focusing on the dynamics of team-level interactions in human-AI collaboration:

1. We provide a detailed analysis of lexical, conceptual, and attentional alignment among humans in the presence of an AI agent. This study moves beyond commonly researched dyadic interactions (human-AI or human-human) to examine the influence of AI on cognitive alignment within entire human teams.
2. We investigate the influence of AI quality, voice anthropomorphism, and trustworthiness on the adoption of AI terminology, demonstrating that poor AI quality and mistrust in the AI do not preclude lexical alignment.
3. We assess lexical alignment on two distinct classes of words and demonstrate that human adoption of AI terminology diverges from expected patterns in referents that are less central to the task at hand. This distinction highlights the selective influence of AI on human lexical choices, depending on the perceived task relevance of the terms.

2 Background

This study investigates the impact of diverging AI designs on two critical aspects of human teamwork: cognitive alignment of attention and lexical alignment. The following section elucidates these concepts, highlighting their significance and relevance to our study in the context of computer-supported collaborative work and artificial intelligence.

Human Attention. Human attention is the ability to actively process specific information in the environment while tuning out other, less relevant details [24]. Attention is limited in terms of both capacity and duration [25]. Managing this limited resource effectively is critical to achieving high task performance [26].

In group settings, individuals usually align their attention with the attention of other group members, which includes what individuals perceive and remember [27]. Effective group work requires coordination of attention in order to achieve coordination of actions [28]. Groups are said to have a high level of collective attention when individuals are co-attending to an object or event with others [29]. This results in a shared emotional and psychological experience with implications for motivation, judgment, and behavior [1]. For example, collective attention in a shared social context on the same target allows group members to prioritize and coordinate their team-related activities [1]. As a result, collective attention is strongly associated with collective intelligence and team performance [28, 30–32].

Collective attention is sometimes operationalized as the temporal focus of a group on specific information or events, such as viewing or sharing content on social media [33], various forms of interactional synchrony such as facial expression and vocal alignment [1], and language use [12]. Language, in particular, plays a crucial role in coordinating beliefs among humans and aligning neural processes that govern collective attention [34, 35]. Humans may even be predisposed to align their mental states with those of other individuals [36].

Why Does Human Attention Matter? Higher levels of collective attention not only increase the pool of attentional resources available for achieving team goals, but also actively shape the approaches and strategies used in problem-solving [37]. That is, the object in focus and the shared language used to talk about it can shape abstract intentions, plans, and proposals [38]. The resulting heightened focus enhances the team's ability to devise innovative solutions and respond effectively to challenges [39].

A critical question for research on human-AI teams is how AI agents might influence the collective attention of a team [8, 40]. Understanding this impact is essential for optimizing collaboration and enhancing team performance [41]. Importantly, AI may affect not only the overall level of collective attention but also the object and manner of that attention [42], both of which may affect the problem solving approaches employed by the team [43, 44].

Alignment on Shared Language. The study of linguistic convergence has progressed significantly since its inception, shifting from early explorations of interactions between large language groups and dialects to contemporary explorations of fine-grained language coordination in interpersonal conversations [45]. Lexical alignment, a key focus in this domain [46], examines the progressive adoption of similar vocabulary by interlocutors in dialogue [47]. This phenomenon exemplifies the formation of shared conceptual frameworks, which emerge as speakers subtly negotiate and align their terminology [48]. Within this paradigm, the principle of input-output coordination describes a fundamental interaction mechanism [49]: each participant's utterances propose a conceptual framework that shapes the next speaker's response. These interactions often solidify into "conceptual pacts"—persistent, partner-specific shared understandings [50]. The importance of these pacts is underscored by the processing costs that arise when a speaker unexpectedly deviates from them [51]. A shared language thus enhances a group's collective attention by freeing up cognitive resources, facilitating more efficient communication and cognition.

Open Questions How could the presence of an AI assistant affect the formation of a shared lexicon? Several gaps remain in our understanding of collective attention in human-AI teams. Recent research has shown that lexical alignment may be affected by the perceived competence of conversational partners [52]. Humans have been shown to align more towards a computer partner than a human partner, even when the linguistic behavior is identical [21]. This supports a "communicative design" account of alignment, where speakers deliberately align more with partners they perceive as less communicatively competent in order to facilitate successful communication.

Human-AI teaming studies often focus on dyads of one human and one AI agent [e.g., 53, 54] leaving unanswered questions regarding how AI affects the interaction of teams in which multiple humans work together on cognitive tasks through cooperative communication. Consequently, our understanding of how humans and AI interact in team settings is limited [55]. The importance of this area is underscored by the fact that cognitive alignment within a team is not a product of individual interactions, but emerges from complex, multi-agent dynamics involving team members and AI [56, 57]. The effects of this cognitive alignment are often unpredictable [58]; for example, one team member's successful alignment with the AI might be counteracted by an other's misalignment, impacting the emergent collaborative process and overall team performance in ways that are difficult to anticipate [59]. As AI systems become more prevalent in workplace settings, particularly through the rise of generative AI, developing a deeper understanding of these team dynamics becomes crucial for fostering effective collaboration and enhancing collective intelligence in human-AI teams.

While previous research has investigated the influence of perceived partner competence on alignment [60, 61], it remains unclear whether human groups will collectively align their language with that of an AI assistant. Furthermore, little is known about how other attributes of AI systems, such as perceived trustworthiness and helpfulness, affect cognitive alignment in teams [62, 63]. In particular, the dynamics of lexical choice negotiation within mixed human-AI teams during collaborative problem-solving tasks warrant deeper investigation [64, 65]. To realize the vision of designing AI systems that contribute to the collective intelligence of human groups—and thus use AI responsibly [66]—it is crucial to gain a more comprehensive understanding of the impact of AI presence on the diverse forms of cognitive alignment underlying collective attention.

3 Related Work

Our research is most closely related to work on (a) human-AI teaming and (b) lexical alignment between humans and artificial systems.

3.1 Advancing Team Dynamics: The Role of AI Design in Enhancing Communication and Collaboration

Previous research has explored the design of human-AI interfaces to improve team dynamics. These interfaces enhance collaboration by sharing control, improving understanding, employing social-cognitive strategies, building mutual trust, and boosting overall team performance [20, 67–71]. For instance, sharing control in human-robot teams leverages complementary skills, which significantly enhances decision-making and task efficiency [67]. This aligns with findings that emphasize the importance of understanding AI capabilities, particularly an AI system's error boundaries, in boosting team performance by complementing human skills [20]. Other research has focused on developing AI agents capable of constructing mental models of human teammates in order to steer team communication by alleviating information overload and supporting metacognitive processes, such as building shared mental models of "who knows what" [8]. Moreover, research indicates that establishing bi-directional trust in human-AI teams not only involves dynamic relationship management and adaptive systems, but also necessitates co-discovery learning and algorithmic transparency [71]. Other studies reinforce this finding by demonstrating that combining explicit AI explanations with implicit manipulations significantly improves trust dynamics within these teams [72].

Building upon the foundation laid by previous research, the present study investigates the impact of specific AI designs on team communication and collaboration. Inspired by prior work exploring the dynamics of human-AI interaction, we advance into the nuances of AI influence on teamwork. By analyzing the effect of different AI demeanors on team processes, we aim to expand the current state of the art in this field. Our research addresses a notable gap by focusing on the direct effects of AI behavioral characteristics on the efficiency and quality of team interactions. This approach not only enhances our knowledge of effective AI integration into collaborative settings, but also sets new benchmarks for designing AI systems that are more adept at facilitating human teamwork. Through this work, we contribute to a deeper understanding of AI deployment in team-based environments.

3.2 Lexical Alignment in Human-AI Interaction

Lexical alignment is a complex phenomenon [21, 49] influenced by various factors, including the perceived competence of the interlocutor, the interaction history between partners, and specific task demands [73–75]. While several studies have investigated lexical alignment in human-AI interaction [21, 23, 76–78], they predominantly focus on dyadic interactions between one human and one AI agent [76], leaving the dynamics of human-human-AI interactions largely unexplored.

Evidence suggests that humans align differently with AI partners than with human partners [79, 80]. For example, research on text-based and speech-based interfaces indicates that humans are at least as likely to adopt terms used by a computer system as they are to adopt terms used by human partners [21]. A more recent study suggests humans align more strongly with other humans than with artificial agents [23].

While linguistic alignment is observed in both human-human and human-AI interactions, the underlying processes may differ significantly. With human partners, linguistic alignment emerges from a collaborative, negotiated process of establishing shared conceptualizations. With a computer, individuals may adopt the system's terms more strategically in order to avoid errors and accommodate a partner perceived as less flexible [21]. This distinction raises questions about the dynamics in teams with multiple humans and an AI agent. If alignment with the AI's language is strategically motivated, its success would depend on all human teammates making similar strategic adjustments. This would indicate a complicated, multi-step reasoning process that is more likely to fail and may not consistently result in successful cognitive alignment in human-AI teams.

4 Data

We ran a randomized controlled trial with 69 human subjects as an IRB-approved lab study. Teams of three or four people collaborated to solve a puzzle task using online video conferencing software. During the task, the team was assisted by an audio-only AI team assistant. We used a 2×2 between-subjects design which manipulated the quality (helpful vs. wrong) and voice (human-sounding vs. robotic-sounding) of the AI assistant. Following the puzzle task, participants completed a post-experiment questionnaire. For their time, which lasted up to 90 minutes, they received compensation in the form of Amazon gift cards at a rate of $20 per hour. We recruited participants through flyers, email lists, and word of mouth from the Northeast United States to participate in this study. Team size varied between three and four people based on convenience. While we scheduled four participants for each session, we proceeded if at least three were present. The participant pool was predominantly young, with a mean age of 21.7 years, and skewed female (Table 1).

Treatment	Teams	Participants	Female	Mean Age (SD)	Native & Full Proficiency
Human/Helpful	5	17	65%	22.2 (4.56)	100%
Human/Wrong	5	17	47%	21.9 (3.37)	100%
Robotic/Helpful	5	18	56%	20.9 (2.78)	83%
Robotic/Wrong	5	17	71%	21.8 (3.62)	100%
All	20	69	59%	21.7 (3.67)	96%

Table 1. Participant demographics. English fluency is self-reported as Native Proficiency, Full Proficiency, and Proficiency. No participants reported a lower fluency.

4.1 Experimental Task

The experimental task is a pattern recognition challenge framed as a "Cursed Treasure Puzzle." Teams have 40 minutes to reach five milestones. The milestones require teams to uncover patterns among the gems in different treasure chests to determine the rules associated with each curse (Fig. 1). For example, the "octopus curse" is present when none of the gems touch each other. The task is particularly suitable for studying the development of shared language because it presents participants with several unfamiliar symbols for which they do not have established terminology. Since the symbols can be described using a variety of words, teams must develop a shared vocabulary to effectively distinguish them in order to direct each other's attention and communicate precisely [81, 82]. Several other studies investigating the development of shared language have relied on similar symbol tasks [81, 82]. The pattern recognition task is designed to be challenging; most teams fail to reach all milestones. To help them, we introduce an AI assistant called the "AI Puzzle Master" that dispenses one clue for each milestone.

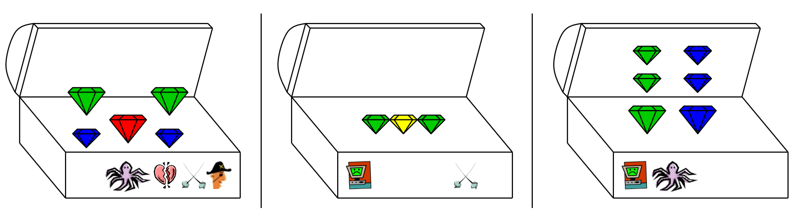

Figure 1. A sample of three chests in the Cursed Treasure puzzle. Notice that the octopus curse relates to the separation of gems.

4.2 AI Assistant

The "AI Puzzle Master" was presented as a disembodied entity, represented by a black screen displaying its name in the online video conferencing software. At the experiment's outset, an experimenter read a script introducing the Puzzle Master as an AI assistant capable of offering helpful information. The AI assistant's interaction was one-sided, communicating with teams only via pre-recorded audio messages played at predetermined intervals. We created four different AI assistant recordings, one for each treatment condition, to ensure uniformity in message content, sequence, and timing. Each statement delivered by the AI assistant is an intervention driving the causal identification of effects in our statistical analysis. Our agent does not meet the definition of agency required for most definitions of human-agent teams [83]; however, this trade-off allows us to precisely attribute changes in teams' cognitive alignment to the AI assistant's interventions. By controlling the timing and content of the AI assistant's interjections, we limit the confounding effects of endogenous processes that would arise if the AI possessed agency [e.g., 84, 85].

We manipulated the content of the AI assistant messages (helpful or unhelpful) and the AI assistant's voice (human or robotic sounding). The helpful AI assistant provided accurate information, such as "the octopus rule relates to separation between the gems," while unhelpful agents provided false information, like "the octopus rule has to do with the sizes of the gems." These clues were designed to encourage teams to explore patterns associated with the octopus milestone. For the human voice condition, we hired a female voice actor to record the audio. We used a text-to-speech program to generate the robotic voice, which had a female timbre but lacked human prosody and paralanguage.

To validate the treatment, we recruited 400 workers on Amazon Mechanical Turk to determine if the robotic voice was perceived as non-human. Eight variations of the robotic voice description were used: a machine, a computer, a robot, a digital assistant, an automated assistant, automation, and artificial intelligence. Fifty workers were assigned to each description. Workers listened to audio clips of the robotic agent's clues and rated on a seven-point Likert scale whether the voice was human or non-human. Each group received a question tailored to their assigned word (e.g., "Is this voice a machine or a human?"). We found that workers consistently perceived the robotic voice as non-human, regardless of the label used (e.g., at the high end (less significant), "robot": $z = 2.263$, $p = 0.024$; at the low end (more significant), "computer": $z = 3.677$, $p < 0.001$).

4.3 Tracking Linguistic Alignment

To measure the AI assistant's impact on team collective attention and linguistic alignment, we transcribed the audio recordings of team communications and identified each instance of a team member referring to one of the milestones (Fig. 2). Each milestone in the puzzle task is represented by a symbol corresponding to the nature of the "curse" that has been placed on the gems; for example, the "duel curse" milestone is represented by a pair of crossed swords. We compiled a complete list of terms used by participants to refer to each of the milestones, the symbols, and the gems, including different terms used to refer to the same thing. This list of terms formed our set of *tracked terms*. Each time a participant used one of the tracked terms to refer to a milestone, a symbol, or a gem, we recorded its use, forming the basis of our lexical alignment measure (see next section).

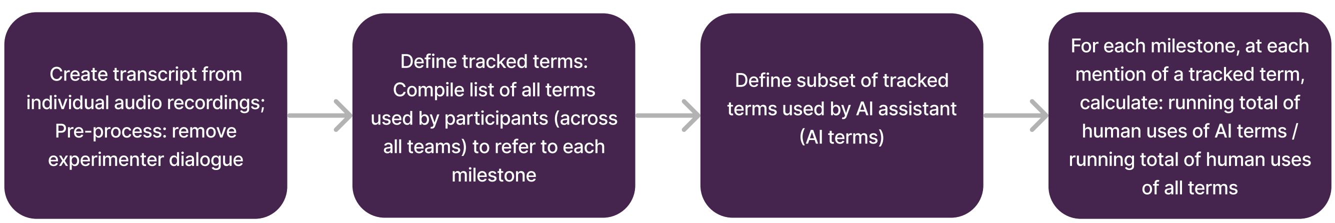

Figure 2. Steps of data pre-processing and variable construction.

The AI assistant consistently used the same language to refer to each of the milestones and to the gems. As a result, we were able to analyze the team's term usage to see whether they used the same terminology as the AI or different terminology. Every time we updated the running usage counter for one of the terms, we also updated a running lexical alignment score based on whether or not the term matched the AI's term for that referent. The alignment score increased if the term was an AI term and decreased if it was not (unless it was already at 100% or 0%, in which case it did not change).

Within the seven tracked referents, we distinguish between five core referents and two peripheral referents. The five referents correspond to the five milestones, each of which is the subject of one clue from the AI assistant. Since the core referents are each directly related to a milestone (a pattern that teams need to uncover), the AI assistant's interjection can be identified as high or low quality depending on whether it is helpful or intentionally misleading. The peripheral referents are "symbols" and "gems", which are relevant to the task but are not milestones themselves. These two referents can be considered neutral in that they are not the subject of a high or low quality interjection. Gems are mentioned in several AI interjections, while symbols are never mentioned, making the symbol referent an appropriate baseline from which changes in collective attention can be contrasted.

In summary, each referent corresponds to a set of tracked terms and each referent is mentioned by the AI assistant in one of its interjections. Following the interjection, teams may talk more (or less) about the mentioned referent, and they may do so using either the same term used by the AI or use a different term to refer to the referent (for example, "duel" as the term the AI assistant used or "swords" as some of the teams used).

4.4 Lexical Alignment Measure

Lexical consensus is calculated as a proportion: the usage of a particular term to refer to a given referent divided by the usage of any term to refer to the same referent. Lexical alignment is the direction of lexical consensus over time. Our interest is in alignment with the language used by the AI assistant, so the usage of the AI assistant's terminology is the numerator of the consensus ratio. The measure ranges between 0, indicating no use of the AI's terminology, and 1, indicating that all mentions use the AI term.

To illustrate the measure, Fig. 3 shows all uses of tracked terms in one of the teams. The very first mention of the gems referent uses the term "gem," starting the alignment measure at 1 (i.e., 100%) indicating that all mentions used the same term as the AI assistant. After this first reference, up to minute 14 and the first AI interjection, most references to the gem use a different term (e.g., "diamond") and the alignment measure decreases. Then the AI interjects, using the term "gem," and all future mentions in the team also use the AI term. The alignment measure increases, approaching 1 at the end.

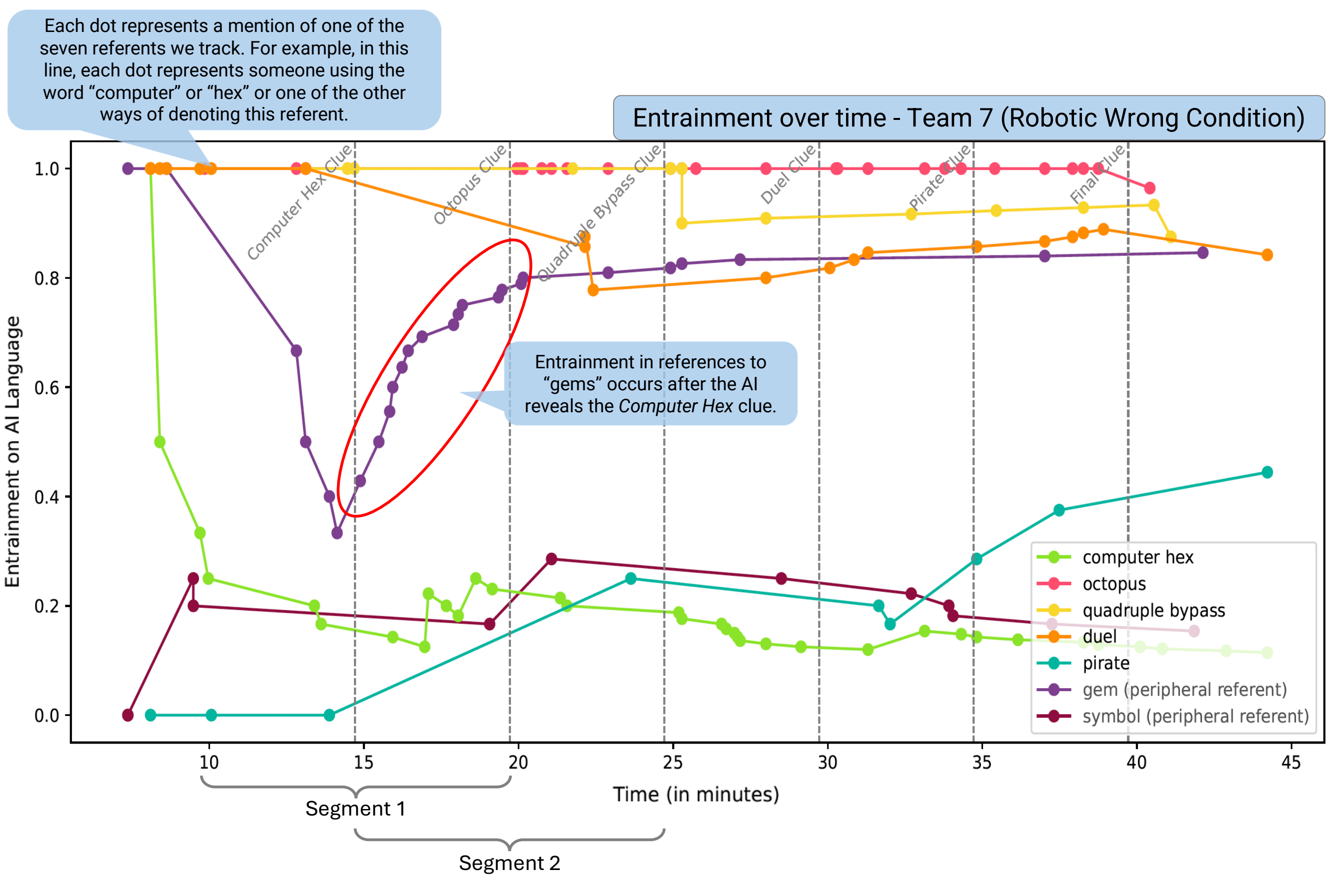

Figure 3. Illustration of lexical alignment on the AI assistant's language in one example team.

5 Method

Human communication is a dynamic process. A key challenge in our analysis is to *causally* attribute specific changes in this fluid team communication to the interventions of the AI assistant. To achieve this causal identification, we employ a panel regression using observations on the utterance level following a difference-in-difference (DD) framework [86]. Specifically, we consider term use (as captured by the lexical alignment measure) in segments right before the AI intervention and right after for each of the six AI intervention (e.g., Segment 1 surrounding the "computer hex" clue; see bottom of Fig. 3). The DD analysis then compares the before/after change in AI term usage during a specific segment to the same before/after change during the same segment of a term that was *not* used by the AI. For example, the before/after change in the usage of the term "octopus" during the segment surrounding the "pirate clue" serves as a no-treatment control (because the "pirate clue" did not mention "octopus") to the before/after change in term usage of "pirate" (which could have been affected by the AI's mention of "pirate"). Formally, we estimate the following regression equation

$$
\begin{aligned}
EntrainmentOnAI_{ogs} = & \beta_1 After_{ogs} + \beta_2 After_{ogs} \times AIQuality_g \\
& + \beta_3 ObjectMentions_{og} + \beta_4 ElapsedTime_{og} \\
& + \alpha_o + \alpha_g + \alpha_s + \varepsilon_{ogs}
\end{aligned}
\tag{1}
$$

where *After* is a dummy variable indicating whether the utterance related to object o happened before or after the AI intervention related to object o, *AIQuality* is a dummy indicator whether the team g was assigned to the high (vs. low) quality condition, α_o are object-level fixed effects, α_g are group-level fixed effects, α_s are segment-level fixed effects, and ε_{ogs} are error terms. The coefficient $\beta_1 After$ captures the causal effect of the AI intervention, and the coefficient β_2 for the interaction term of $After \times AIQuality$ captures the differential effect of the treatment condition. Estimates are driven by six before/after comparisons created by the six AI interventions, each one of them comparing five unaffected terms serves as control contrasted from one term affected by the intervention. We include the count of *ObjectMentions* and the *ElapsedTime* (in minutes) as control variables. We estimate variations of this equation where we substitute *Humanness* and *Treatment* as key variables of interest for *AIQuality*. We report standard errors clustered at the team-level. Notice the main effect of the treatment will drop out due to the team-level fixed effect and only the interaction term remains.

6 Results

This section presents five sets of results in a comprehensive analysis of the AI assistant's impact on team collective attention. First, we offer an anecdotal description of the experience of one specific team to showcase how teams adopt the AI's lexicon (addressing RQ1 & RQ2). Second, we offer causal evidence that the AI assistant affects *what* teams talk about (addressing RQ1). Third, we offer causal evidence that the AI assistant affects *how* teams talk about the referents we track (addressing RQ2) and how the alignment of language differs across the quality and humanness of the AI (addressing RQ3). Fourth, we show that these findings are reflected in the teams' retrospective and reified perceptions of their own team's processes (combining RQ1 & RQ2 into a higher-level construct). Finally, we draw on complementary data sources to explore the mechanism behind the influence of the AI on team collective attention (addressing RQ4).

6.1 How an AI Assistant Shapes Team Communication and Development of Shared Language

We begin our analysis with an anecdotal description of the experience of Team 18, a team of three in the *Human-Wrong* condition. After receiving their instructions, the team begins the task by looking for patterns among the gems. They refer to the gems as "diamonds." They point out various features of the treasure chests and identify some similarities between chests with the same curse. In the first fifteen minutes, they refer to the gems seven times using the word "diamond." Nobody uses the word "gem" until the AI assistant interjects at 15:07 with the first clue—in this case, a deliberately misleading clue about the computer hex milestone. This clue includes a mention of "gems."

Following the interjection, the team's attention shifts and they begin discussing the computer hex curse, which has so far not been addressed. They express confusion because they cannot see the pattern suggested by the clue (indicating an answer for our RQ1 about the AI's effect on the timing of *what* teams talk about). Nevertheless, they stop using the word "diamond" and instead adopt the term "gem" that was used by the AI. There are another eight references between the first and second AI interjection, all now using the term "gem." The team has adopted the AI's language for the peripheral gem referent (indicating an answer for our RQ2).

At 20:07, the AI provides the second clue. This clue states that the octopus milestone has to do with the sizes of the gems. At this point, the team grows suspicious of their AI assistant because this clue contradicts the pattern they have begun to identify. One participant suggests that the clues are intended to throw them off, but another team member says "we have to trust the Puzzle Master."

At 30:07, the AI provides the fourth clue, which is related to the duel milestone. Afterwards, a team member again suggests that the clues may be intended to mislead them. The team alternates between the term "duel" and "sword" to refer to the duel clue throughout the rest of the task, not fully adopting the language proposed by the AI assistant for this core referent. The team does not manage to complete the task in time.

A different team assigned to the *Robotic-Helpful* condition, Team 19, exhibits a contrasting pattern. Participants in this team demonstrated higher overall trust in their AI assistant, which provided them with helpful rather than misleading information. The initial two clues, pertaining to the "computer hex" and "octopus" milestones, employ terminology consistent with the group's usage at the experiment's outset, prompting no change in lexical choices. Both clues align with the group's observations, giving participants no reason to doubt the reliability of the AI. The third clue, introduced at 24:34, informs the group that the quadruple bypass milestone relates to the colors of the gems. At this point, a team member clarifies that the "quadruple bypass" milestone is synonymous with what the team had referred to as the "heartbreak" or "heart" milestone. Notably, the team persists in referring to this milestone as "heartbreak" or "heart," despite their trust in the AI. Similarly, the team maintains their use of "pirate" or "pirate eye" when discussing the "eye patch" milestone, even after the AI provides a helpful clue using the term "eye patch."

6.2 AI Assistant Shapes What Teams Talk About

The AI assistant significantly influenced the focus of team discussions, affecting *what* teams talked about and therefore what they paid attention to (addressing RQ1). Our analysis compares the frequency of tracked referent mentions during baseline segments to those immediately following AI interventions (Fig. 4). In every instance, team discussions of the referent increased following the AI's mention, suggesting that the AI intervention significantly shaped their collective focus of attention. Teams were significantly more likely to discuss the mentioned referent in the period immediately following the intervention compared to other referents (we estimate the same difference-in-difference model from 1 but use a binary indicator of whether a mentioned term corresponds to the AI intervention as the dependent variable; $\beta = 0.35$; $p < 0.001$). Overall, teams were 1.5 times more likely to discuss the referent in the segment following the AI's mention than in any other segment. The impact of the AI on team collective attention is not significantly different across the treatment groups, indicating that the mere mention of a referent by the AI influenced group attention, regardless of the quality or voice of the AI assistant. Recall that the AI in this study was not interactive; thus, the observed effect was not due to direct engagement with the AI, but to an influence on human-to-human communication.

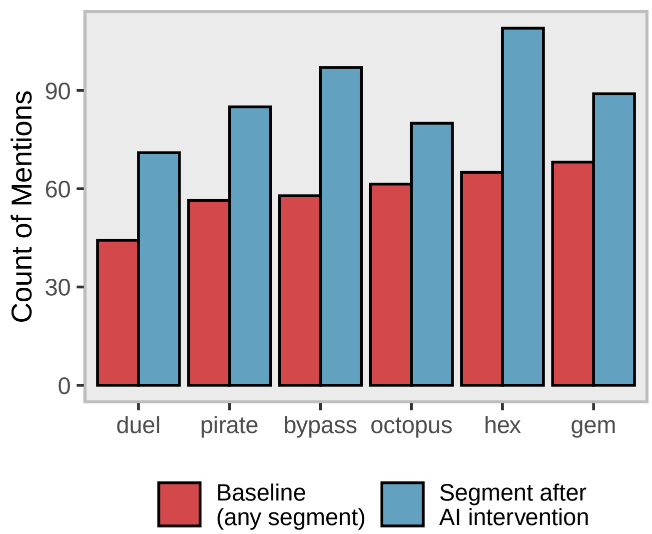

Figure 4. AI intervention affects collective attention of the team (sorted by mentions in baseline).

6.3 Team's Adoption of Terms Used by the AI: Lexical Alignment

The AI assistant not only shapes *what* teams talk about but also *how* they talk about it. We find significant signs that teams align their communication with the language introduced by the AI assistant (answering RQ2; Table 2). First, teams show strong patterns of alignment caused by the AI assistant, indicated by the positive and significant coefficient of the *After* dummy (Model 1, $\beta = 0.06$, $p < 0.01$). This overall coefficient masks important heterogeneity of an even stronger effect in some conditions and lower alignment in others. Alignment is especially high when the AI has a human sounding voice (Model 2, $\beta_{After} = 0.14$, $p < 0.001$) and teams align less on the AI's language when the AI has a robotic voice ($\beta_{robotic} = -0.11$, $p < 0.05$). Alignment is lower when the AI is of low quality (as compared with high quality) but the difference is not statistically significant (Model 3, $\beta_{wrong} = -0.08$, *ns*). Finally, we find significant interaction between the robotic voice and AI quality. Teams align with AI terminology most strongly in the *Human-Helpful* condition (Model 4, $\beta_{After} = 0.15$, $p < 0.001$). Compared against this baseline, alignment is significantly lower in the *Robotic-Helpful* condition ($\beta = -0.09$, $p < 0.01$), and lower still when the AI is both robotic and provides low-quality assistance ($\beta = -0.16$, $p < 0.05$). There is no significant difference between the *Robotic-Helpful* and *Robotic-Wrong* condition (linear hypothesis test: $\chi^2 = 1.40$; $p = 0.24$). Together, these results answer RQ3 regarding how the effect of the AI differs based on (a) its quality and (b) its humanness.

The lexical alignment patterns differ substantially between core referents and peripheral referents. Whereas the coefficient indicating *After* AI intervention is positive for referents with high relevance to the clues (β ranging from 0.06 to 0.15, p at least

< 0.01 across Models 1-4) the coefficient is negative and significant for peripheral referents (Model 5, $\beta = -0.16$, $p < 0.01$). The pattern for the effect of AI quality also reverses. Whereas the interaction effect between the AI's intervention and low AI quality is significant and negative for core referents in Models 3 and 4, it is positive and significant for peripheral referents (Model 5, $\beta = 0.27$, $p < 0.01$).

We theorize that this strong alignment on peripheral terms used by the low-quality AI assistant is driven by the collective attentional focus of the team. When teams are faced with low-quality AI input, they do not align with the AI's language for core terms where the low quality of the input is apparent; here, teams adopt their own language. For peripheral terms, however, humans might more readily align with the AI's terminology, as it is less obviously associated with misleading information. This could reflect a lowered barrier to alignment due to the reduced risk associated with following the AI's lead on less important elements of the interaction [87]. As the team's attention is focused on discerning (and confirming) the low-quality nature of the AI interjection, less attentional focus is available for other aspects of the task. Under the resulting cognitive burden, teams may align with AI terms that are only peripherally related to the misleading clue, overlooking their association with misinformation. This lack of shared attentional focus on peripheral referents may make teams more susceptible to unconsciously adopting peripheral terms used by the AI [12].

In summary, *all* teams adopted the AI's language: if not for core terms (in the high AI quality conditions), then for peripheral terms (in the low AI quality conditions). Teams show consistently strong lexical alignment with AI language for core terms when the AI quality is high and/or human sounding, but low alignment when the AI is low quality and/or robotic sounding. In contrast, for peripheral terms, teams tend to diverge from the AI terminology when AI quality is high (show *negative* alignment on AI terminology), but show very strong positive alignment on the AI assistant's language when the AI is low quality.

Dependent Variable:	Entrainment on AI Language				
	Core Terms				Peripheral Terms
	(1)	(2)	(3)	(4)	(5)
After AI Intervention	0.06^{**}	0.14^{***}	0.10^{**}	0.15^{***}	-0.16^{*}
	(0.02)	(0.03)	(0.03)	(0.03)	(0.07)
After × Robotic Voice		-0.11^{*}			
		(0.04)			
After × Wrong			-0.08		0.27^{**}
			(0.05)		(0.09)
After × Human Wrong				-0.04	
				(0.06)	
After × Robotic Helpful				-0.09^{**}	
				(0.03)	
After × Robotic Wrong				-0.16^{*}	
				(0.06)	
<u>Controls</u>					
Referent Mentions	$-0.01^{\dagger}$	$-0.00^{\dagger}$	$-0.01^{\dagger}$	$-0.01^{\dagger}$	0.00
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Elapsed Time	0.00	-0.00	0.00	-0.00	-0.00
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Num. obs.	1,995	1,995	1,995	1,995	713
Adj. R^2	0.70	0.70	0.70	0.70	0.79
Num. groups: Object	5	5	5	5	2
Num. groups: Team	20	20	20	20	20
Num. groups: Segment	7	7	7	7	7

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$; $^{\dagger}p < 0.1$

Standard errors clustered on team-level.

Table 2. Statistical models

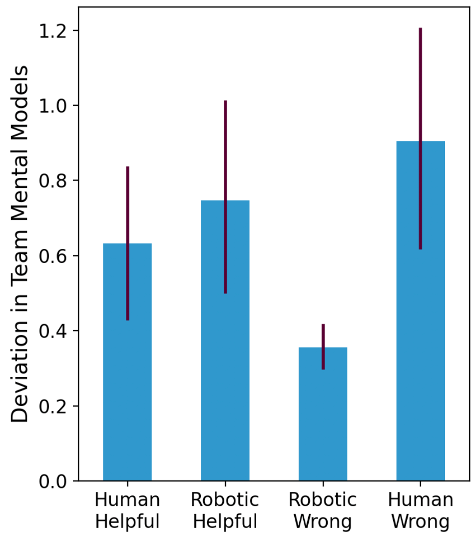

Figure 5. Alignment of mental model development points to exclusion of AI in the *Robotic-Wrong* condition. Bars show standard deviation across answers to shared mental model questions per treatment.

6.4 Shared Mental Models

So far, we have shown evidence of the AI assistant's substantial impact on collective attention and shared cognition through various behavioral measures of alignment in team language use. In this section, we explore whether these findings are reflected in the participants' retrospective and reified perceptions of their own team processes (combining RQ1 & RQ2 into a higher-level construct). The combination of evidence from behavioral measures of group interaction and participants' self-reflection of their group's process offers particularly compelling support for the AI assistant's influence on collective attention [3, 11]. Specifically, we investigate the effect of AI on the emergent shared cognition through which groups create and use interpersonal understanding. This shared cognition represents the manner in which knowledge important to team functioning is mentally organized, represented, and distributed within the team [3]. Recognized as a central positive driver of team behavioral processes, motivational states, and team performance [3, 88], emergent collective cognition is especially beneficial in interdependent cognitive tasks where leveraging team members' expertise is crucial [89]. Teams with well developed shared mental models typically possess a common view of "what is happening, what is likely to happen next, and why it is happening" [88, p.879].

To assess the effect of the AI assistant on team cognition, we investigate whether the it affects the alignment—or similarity—of mental models within a team. We measure the alignment of mental models as the degree to which members' mental models are consistent [88]. First, we assess shared mental models using six items from [90]. Items propose statements like "my team knows specific strategies for completing various tasks" and participants are asked to respond on a Likert scale. With this data, we compute an intragroup agreement index within each item using standard deviation (to measure how consistently within-team responses overlap versus diverge), and then average the standard deviations across the six items.

We find a medium level of alignment in shared mental models in both helpful conditions, while the unhelpful conditions exhibit diverging patterns (Fig. 5). Mental models are less well aligned (i.e., have higher standard deviation) in the *Human-Wrong* condition, while we find the highest level of alignment in the *Robotic-Wrong* condition. A significant interaction exists between the two factors of our experimental design (one-way ANOVA $F(3,65) = 5.5$; $p = 0.041$). We theorize that this pattern is the result of an "us vs. them" dynamic, which we discuss in detail in the next section.

6.5 Exploring the Mechanism Behind AI's Influence on Team Collective Attention: A Closer Look at Complementary Data Sources

The above analyses reveal the substantial impact of the AI assistant on three aspects of collective attention: *what* teams talk about, *how* they talk about it, and the alignment of their mental models. In the following section, we provide further evidence for this significant effect on cognitive alignment within teams and explore the nuanced contrasting effects between core and peripheral terms (addressing RQ4). We investigate the role of trust in AI, the effect of the AI assistant on the collective identity of the team, and the alignment of mental models. By integrating the evidence from above with self-reported survey measures, we offer a more comprehensive account of cognitive alignment on a different timescale.

6.5.1 Trust in AI

Trust in AI plays a pivotal role in shaping the nature and quality of human-AI interactions[91]. We theorize that the the level of trust individuals place in AI not only influences direct human-AI interactions, but also impacts interpersonal dynamics among human team members when an AI assistant is present. To measure trust in the AI assistant and perceptions of the assistant's contribution quality, we use six items adapted from [92] (originally designed to measure trust in online avatars). Participants responded on a five-point Likert scale from Strongly Disagree to Strongly Agree to items like "The Puzzle Master gave helpful clues."

Humans rated the low-quality AI as significantly less trustworthy ($p < 0.001$) and less intelligent ($p < 0.001$) than the high-quality AI. Voice anthropomorphism did not have a significant impact on participant impressions of trustworthiness ($p = 0.570$), but the ratings were slightly lower in the Robotic treatments. The greater level of trust in the high-quality AI conditions may explain why participants aligned more with the AI assistant's language for core referents in these conditions. Humans are more likely to adopt recommendations by trustworthy agents that provide high-quality explanations [93]. High trust can also indicate the perception of higher status [94], which may explain why the AI assistant in in the high-quality conditions has more influence over the team's language dynamics. Conversely, lower trust and lower perceived status can make the AI assistant easier to ignore, thus giving it less influence over the lexical choices of team members.

Overall, a stark pattern emerges: even in low-quality conditions in which trust assessments of the AI assistant were significantly lower, the AI assistant directed team attention to the specific task aspects it mentioned (see Section 6.2) and influenced their lexical choices. This pattern demonstrates the power of the AI assistant to shape collective attention and shared language, even when teams are aware of its low quality and have decided not to trust it.

6.5.2 Is the AI Assistant a Part of the Team?

We asked individuals to report whether they felt that the AI assistant was part of their team or even a leader of their team. Humans did not consider the AI to be a part of their team, with 77% of participants responding "Strongly Disagree" or "Somewhat Disagree." Humans responded even more negatively when asked whether the AI was a leader of their team, with 93% of participants responding "Strongly Disagree" or "Somewhat Disagree." Ratings were especially low in the *Robotic-Wrong* condition, where 96% did not consider the AI assistant to be a leader of the team and 93% of participants did not consider it to be part of the team at all.

This suggests that the AI assistant's substantial influence over teams' collective attention and lexical alignment is not predicated on its inclusion as a team member, but rather occurs despite it being considered an outsider. This phenomenon indicates a path through which AI assistants can—intentionally or unintentionally—affect intra-team communication and cognitive alignment. By presenting as a non-human entity, AI assistants may strengthen the bond among human team members (the in-group) by serving as a contrasting out-group, giving rise to an "us vs. them" dynamic [95]. The degree to which AI assistants are (or can be) excluded from the team may further shape cognitive alignment among human teammates and enhance joint action. The next section provides additional evidence for this exclusionary dynamic through an analysis of pronoun use.

6.5.3 Pronoun Use

The alignment of shared cognition in teams is driven and reflected by language use at multiple levels [12]. In addition to aligning cognition through the use of shared terminology, language can also signal collective identity through the use of function words that require mutual understanding, such as pronouns [96]. Pronoun use relies on the existence of a shared understanding between interlocutors about the person, object, or concept being referenced; a speaker using a pronoun trusts that the listener will be able to infer the intended referent based on context and collective attention.

This suggests that in groups where pronoun use is more frequent, speakers have a higher level of trust in their teammates' ability to interpret intended meaning without the need for specific descriptors. We theorize that more frequent reliance on collective understanding, as signaled through pronoun use, can indicate a greater sense of unity and cognitive alignment among team members.

The implications of pronoun use for collective attention extend to group cognitive efficiency as well as alignment. Because pronouns serve as an economical way to refer to entities that are in the current focus of attention, they enable group members to manage and negotiate their collective attentional resources. The use of a pronoun signals to the addressees that the referent is salient and easily identifiable, reinforcing the group's collective focus. Pronoun use also allows speakers to minimize the effort expended on formulating referring expressions [97]. In essence, by reducing the collective effort needed for referential processes, pronouns free up cognitive resources that can be allocated to the actual problem-solving activities, potentially resulting in improved group performance.

To assess how the presence of an AI assistant shapes collective identity as expressed in pronoun use, we focus on first- and third-person pronouns (Fig. 6). The limited number of team-level observations limits the validity of statistical tests for differences, so we turn to descriptive analysis. Across a range of pronouns, the *Human-Helpful* and *Robotic-Helpful* conditions yield a moderate level of pronoun use (a mean of 281 and 269 pronouns per team, respectively). Teams in the

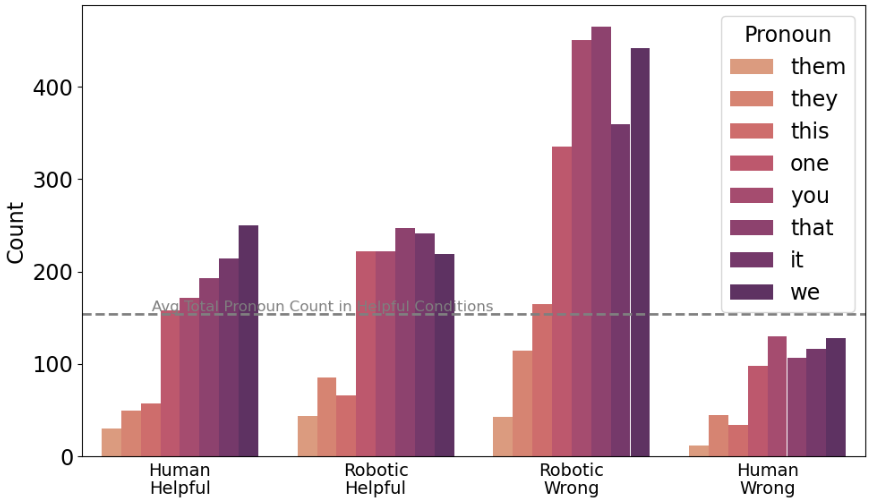

Figure 6. Pronoun use indicates exclusion of AI in *Robotic Wrong* condition. Bars show count of each pronoun across all teams in each treatment.

Robotic-Wrong treatment used considerably more pronouns (mean of 297 per team), while teams in the *Human-Wrong* treatment used significantly fewer (mean of 223 per team). In line with the insights previously discussed regarding whether participants consider the AI assistant a member of their team, our analysis reveals that 'we' is the most frequently used pronoun in the *Human-Helpful* condition but is used rarely in the *Human-Wrong* condition.

We theorize that an "us vs. them" dynamic between the human participants and the AI assistant explains this pattern, with the dynamic varying based on the quality and voice anthropomorphism of the AI. In both low-quality conditions, teams wish to exclude the disruptive and misleading AI assistant from the team. In the *Robotic-Wrong* condition, the unnatural voice of the AI makes it easier for the human participants to exclude it from their in-group. This leads to increased pronoun use among the human teammates as they place more faith in each other's abilities as compared to the AI. Conversely, in the *Human-Wrong* condition, the anthropomorphic voice makes it harder for the participants to exclude the AI assistant from the human in-group. The presence of a misleading influence in the in-group causes confusion and doubt, resulting in an overall decrease in pronoun use. This insight complements prior work that has noted that artificial agents that appear vulnerable can enhance human-human interaction [19].

This failure to exclude the disruptive AI agent explains the poor alignment of mental models (shown earlier in Fig. 5). In the *Human-Wrong* condition, mental model alignment is low because participants struggle to exclude the disruptive AI assistant from the team, likely due to its human-like voice. In contrast, the *Robotic-Wrong* condition shows high mental model alignment, as participants easily exclude the assistant due to its unnatural, artificial-sounding voice. We confirm through a correlation analysis that the pattern of pronoun use is similar to that of mental model alignment. Specifically, we find a marginal negative correlation between pronoun use and mental model alignment ($\rho = -0.363$; $p = 0.115$). Note that we measure alignment using standard deviation, where higher values indicate lower alignment.

7 Discussion

The findings from this study emphasize the substantial influence of artificial intelligence systems on team dynamics and communication, aligning with Theiner et al.'s insights on integrating AI into socially extended cognition [98, 99]. The AI assistant emerges as an active participant in the collaborative workspace, significantly shaping team communication, focus, dynamics, lexical alignment, cognitive alignment, and collective identity. These results underscore the critical role of thoughtful AI design in enhancing collaborative efficiency and cognitive processes in professional environments. Ultimately, this research calls for a reevaluation of the role of AI in professional settings, urging designers and organizations to consider the far-reaching implications of AI integration on team functioning and collective intelligence.

7.1 The Pervasive Influence of AI on Team Dynamics and Communication

Studies suggest that AI can enhance collective attention and decision-making in human teams by providing real-time data insights and strategic recommendations [100, 101]. AI systems, in particular, may be able to improve collective intelligence and performance by augmenting human capabilities with data-driven insights and predictive analytics [102]. However, beyond their intended direct contributions, AI systems may also subtly affect a variety of team processes that, together, shape the team's collective attention, shared memory formation, perceived coherence, and collaborative reasoning capabilities. In essence, AI may affect numerous antecedent processes that fundamentally mold the collective intelligence and performance of teams. This pervasive influence can lead to unintended consequences, including conflicts within the team, misdirected attention away from crucial aspects of tasks, and increased cognitive overhead. These outcomes potentially compromise team performance and hinder effective collaboration.

Our results confirm that the impact of AI on human teams persists regardless of its accuracy or direct relevance to team tasks, demonstrating the power of AI systems to shape decision-making processes and communication dynamics [103, 104]. The AI within our study altered team members' vocabulary and directed their focus, prompting discussions about objects it mentioned, irrespective of its helpfulness or trustworthiness. The ability to steer team attention and discussion enables the AI assistants to influence the mechanics of the collaborative environment.

Our study suggests that even a noticeably untrustworthy AI exerts a subtle yet significant influence on the attentional focus of human teammates. The AI's lexical choices play a critical role in shaping how team members allocate their attention, determining which aspects of a referent become prominent in perception and cognition. For instance, the terms "quadruple bypass" and "heartbreak" describe the same event but evoke different perspectives. "Quadruple bypass" emphasizes the medical and numerical aspects, while "heartbreak" invokes emotional and symbolic dimensions. This intricate link between language and attention highlights the impact of AI-driven lexical choices on shaping team dynamics and directing cognitive focus.

Our analysis reveals the significant influence of AI on lexical alignment. Teams naturally adopted the language used by their AI assistant; for instance, Team 18 switched from referring to "diamonds" to "gems" after the AI's intervention, even though they suspected the AI might be misleading them. This alignment was less pronounced in teams whose AI assistant did not have a human-sounding voice, even when the AI provided helpful clues. This highlights a vulnerability in human judgment: voice anthropomorphism is a stronger predictor of lexical alignment than the quality of the AI's assistance, even when the assistance quality is under suspicion. Moreover, teams demonstrated a strong inclination to align their communication patterns with AI assistants, particularly when the AI exhibited human-like qualities and provided high-quality assistance. Significant alignment was observed with high-quality AI featuring a human voice, whereas alignment was less pronounced with low-quality AI perceived as robotic. The study also revealed distinct alignment patterns between core and peripheral terms. Teams showed robust lexical alignment with core terms when interacting with high-quality AI, whereas alignment was more noticeable with peripheral terms when engaging with low-quality AI. These findings underscore how AI's quality influences its impact on team dynamics and communication patterns across different types of terms.

The relationship between AI quality and lexical alignment is more nuanced when core referents and peripheral referents are considered separately. In low-quality AI conditions, participants perceived the AI as less trustworthy and demonstrated reduced willingness to adopt its terminology for core referents. Interestingly, they were more likely to adopt the AI terminology for peripheral referents, suggesting that the AI's perceived quality impacts lexical alignment differently depending on whether the referent is central to the task. Across all conditions, the AI influenced collective attention and language use within teams; its impact was pervasive even when trust in the AI was low.

7.2 Lexical Alignment and Bias Propagation

The tendency for humans to align with the terminology used by an AI assistant even when they distrust it or suspect it of being intentionally misleading has implications for the amplification of social biases in human-AI interaction. Bias often manifests in lexical choices that reflect societal prejudices, and in AI systems, biased training data can lead to the generation of prejudiced language. Savvy users of AI are aware that language models may perpetuate biases and other inaccuracies in their training data, but our results suggest that these users may adopt the biased terminology regardless. The spread of biased terminology and conceptual frameworks could result, potentially influencing discourse and decision-making.

The dissociation between a user's explicit attitude (distrust of the AI) and their implicit behaviors (lexical alignment) could lead to situations where users inadvertently lend credibility to or perpetuate ideas they consciously oppose. This possibility is one more reason for designers of AI systems to maintain robust ethical guidelines and transparency about their models' potential biases.

The pronoun usage data provides additional insight into the effect of the AI assistant on collective identity within teams. In the *Human-Helpful* condition, frequent use of "we" indicated a strong sense of team unity. Conversely, in the *Human-Wrong* condition, there was a decrease in the use of "we" and in pronoun use overall, indicating uncertainty within the team. Interestingly, the *Robotic-Wrong* condition exhibited an increase in pronoun use, suggesting that the presence of the AI as an "outsider" may have strengthened intra-team bonds.

7.3 Cognitive Implications and Design Considerations

The concept of socially extended cognition posits that external elements, such as AI integrated into group processes, can significantly alter cognitive dynamics [105]. These fundamental differences between human-human and human-AI interactions could impact the emergence of cognitive alignment in mixed human-human and human-AI teams [4–6]. Our study illustrates this phenomenon by showing that AI not only directs team attention but can also detract from pertinent discussions, raising critical questions about AI system design and deployment in collaborative settings.

We examined how AI assistants influenced teams' shared mental models and cognitive alignment. Teams generally showed moderate alignment in helpful conditions but diverged notably in scenarios where the AI provided incorrect information. Surprisingly, the strongest mental model alignment occurred in the Robotic-Wrong condition, adding support to the "us vs. them" dynamic suggested by our analysis of pronoun use.

7.4 Practical Implications Across Professional Settings

While teamwork is common across many domains, our findings regarding the effect of AI assistants on cognitive alignment in teams may be particularly relevant in the following areas.

7.4.1 Offices

AI tools like virtual assistants and project management software streamline workflows and enhance productivity. They influence decision-making by prioritizing specific data points, potentially redirecting project trajectories without explicit human acknowledgment. This guidance can shift power dynamics and decision-making authority within teams, fostering dependencies on AI recommendations [106].

7.4.2 Newsrooms

In journalism, AI supports data analysis, content generation, and trend prediction, impacting editorial decisions and narrative framing. Journalists may adopt AI-suggested language and editorial styles, contributing to lexical alignment and standardizing content. This possibility raises ethical considerations regarding journalistic integrity and the diversity of viewpoints presented [107].

7.4.3 Gig Economy

As gig work shifts toward more collaborative team work [108], the influence of AI on cognitive alignment may be particularly problematic. The gig work environment is already rife with potential for exploitation [109]. Understanding these dynamics is crucial for optimizing the role of AI in promoting worker empowerment and autonomy in gig work environments.

7.5 Future work

Future research should prioritize the development of mechanisms to mitigate undesired AI-induced lexical alignment and explore the diverse impacts of different AI communication strategies on team collaboration. Continuous monitoring of AI's influence on team dynamics is essential for effectively managing and adapting to these impacts. Existing research underscores AI's positive impact on human teams, particularly in contexts like emergency response, where it enhances situational awareness and response times, thereby improving operational efficiency. However, challenges such as potential communication breakdowns and AI failures emphasize the critical need for developing robust AI integration strategies.

Additionally, it would be valuable to investigate the extent to which individuals are aware of the AI agent's influence on their language and thought processes. Studies might employ a combination of behavioral measures and self-report techniques to gauge participants' conscious awareness of linguistic and conceptual shifts. This metacognitive insight would advance our understanding of the subtleties of human-AI collaboration and enable more informed design decisions.

7.6 Conclusion

Our findings underscore the need to reconceptualize the role of AI in collaborative workspaces as not merely a passive instrument, but an active participant that shapes cognitive and communicative processes within teams. This paradigm shift not only carries significant implications for AI design but also advances the theoretical understanding of socially extended cognition, calling for ongoing empirical investigation. These insights contribute significantly to the fields of human-computer interaction and organizational behavior, illuminating the nuances of integrating AI into professional settings for enhanced collaborative efficacy. Recognizing the subtle yet substantial ways in which AI can shape team interactions enables designers and researchers to create AI systems that better support and augment human teamwork. This knowledge is crucial for leveraging AI to improve decision-making processes, enhance creativity, and foster more effective collaboration in professional environments.

In highlighting the correlation between AI's perceived humanness and its influence on teams, our study builds on existing research that suggests the impact of AI on human teams depends on its quality and perceived humanness [110]. Higher-quality AI systems capable of human-like interaction tend to enhance team performance and cohesion, while lower-quality AI or AI perceived as less human-like may provoke increased conflict and negative perceptions regarding the effectiveness and credibility of AI [111]. Furthermore, high-quality, human-like AI fosters greater linguistic alignment and more cohesive shared mental models, which are crucial for effective teamwork. These insights emphasize the importance of designing AI systems that are not only functional but also human-like in their interactions to enhance team cohesion and efficiency.

As AI agents increasingly permeate diverse professional environments, their subtle yet impactful ability to shape team dynamics is likely to persist, even without overt integration into team structures or trust frameworks. A clear grasp of these dynamics is vital to understanding the broader impact of AI on collaborative environments and decision-making processes across professional contexts.

References

1. Woolley, A. W., Chow, R. M., Mayo, A. T., Riedl, C. & Chang, J. W. Collective attention and collective intelligence: The role of hierarchy and team gender composition. *Organization Science* **34,** 1315–1331 (2023).
2. Kommol, E., Riedl, C. & Woolley, A. W. *The Components of Collective Intelligence and their Predictors* in *Academy of Management Proceedings* **2023** (Academy of Management, Briarcliff Manor, 2023), 10519.
3. DeChurch, L. A. & Mesmer-Magnus, J. R. The cognitive underpinnings of effective teamwork: a meta-analysis. *Journal of applied psychology* **95,** 32 (2010).
4. Wang, Q., Saha, K., Gregori, E., Joyner, D. & Goel, A. *Towards mutual theory of mind in human-ai interaction: How language reflects what students perceive about a virtual teaching assistant* in *Proceedings of the CHI conference on human factors in computing systems* (Association for Computing Machinery, New York, NY, United States, 2021), 1–14.
5. Xie, W. *Two studies of human-AI interactions* PhD thesis (Hong Kong Polytechnic University, 11 Yuk Choi Rd, Hung Hom, Hong Kong, 2023).
6. Mahmud, B., Hong, G. & Fong, B. A Study of Human–AI Symbiosis for Creative Work: Recent Developments and Future Directions in Deep Learning. *ACM Transactions on Multimedia Computing, Communications and Applications* **20,** 1–21 (2023).
7. Terblanche, N., Molyn, J., de Haan, E. & Nilsson, V. O. Comparing artificial intelligence and human coaching goal attainment efficacy. *PLOS ONE* **17,** e0270255 (2022).
8. Westby, S. & Riedl, C. *Collective intelligence in human-AI teams: A Bayesian theory of mind approach* in *Proceedings of the AAAI Conference on Artificial Intelligence* **37** (AAAI Press, 101 Pennsylvania Ave, NW, Suite 300, Washington, DC 20004, 2023), 6119–6127.
9. Academies, N. *State-of-the-Art and Research Needs: Teaming, Human-AI* (National Academies Press, Washington, DC, 2022).
10. Riedl, C., Kim, Y. J., Gupta, P., Malone, T. W. & Woolley, A. W. Quantifying collective intelligence in human groups. *Proceedings of the National Academy of Sciences* **118,** e2005737118 (2021).
11. Lix, K., Goldberg, A., Srivastava, S. B. & Valentine, M. A. Aligning Differences: Discursive Diversity and Team Performance. *Management Science* **68,** 8430–8448 (2022).
12. Sievers, B., Welker, C., Hasson, U., Kleinbaum, A. M. & Wheatley, T. Consensus-building conversation leads to neural alignment. *Nature Communications* **15,** 3936 (2024).
13. Fusaroli, R. *et al.* Coming to terms: Quantifying the benefits of linguistic coordination. *Psychological Science* **23,** 931–939 (2012).
14. Fusaroli, R., Rączaszek-Leonardi, J. & Tylén, K. Dialog as interpersonal synergy. *New Ideas in Psychology* **32,** 147–157 (2014).
15. Autor, D. *Polanyi's paradox and the shape of employment growth* tech. rep. w20485 (National Bureau of Economic Research, 2014).
16. Fügener, A., Grahl, J., Gupta, A. & Ketter, W. Will humans-in-the-loop become borgs? Merits and pitfalls of working with AI. *Management Information Systems Quarterly* **45,** 1527–1556 (2021).
17. Dell'Acqua, F. *et al.* Navigating the jagged technological frontier: Field experimental evidence of the effects of AI on knowledge worker productivity and quality. *Harvard Business School Technology & Operations Mgt. Unit Working Paper* (2023).
18. Richardson, B. H. & Ball, L. J. Progressing the development of a collaborative metareasoning framework: Prospects and challenges. *Journal of Intelligence* **12,** 28 (2024).
19. Traeger, M. L., Strohkorb Sebo, S., Jung, M., Scassellati, B. & Christakis, N. A. Vulnerable robots positively shape human conversational dynamics in a human–robot team. *Proceedings of the National Academy of Sciences* **117,** 6370–6375 (2020).

20. Bansal, G. *et al. Beyond Accuracy: The Role of Mental Models in Human-AI Team Performance* in *Proceedings of the AAAI Conference on Human Computation and Crowdsourcing* (2019), 2–11.

21. Branigan, H. P., Pickering, M. J., Pearson, J., McLean, J. F. & Brown, A. The role of beliefs in lexical alignment: Evidence from dialogs with humans and computers. *Cognition* **121,** 41–57 (2011).

22. Cowan, B. R., Branigan, H. P., Obregón, M., Bugis, E. & Beale, R. Voice anthropomorphism, interlocutor modelling and alignment effects on syntactic choices in human- computer dialogue. *International Journal of Human-Computer Studies* **83,** 27–42 (2015).

23. Wudarczyk, O. A. *et al.* Robots facilitate human language production. *Scientific Reports* **11,** 16737 (2021).

24. Bundesen, C. A theory of visual attention. *Psychological review* **97,** 523 (1990).

25. Duncan, J. in *Attention and performance XI* 85–106 (Routledge, 5 Howick Place, London, SW1P 1WG, UK, 2016).

26. Duncan, J. The structure of cognition: attentional episodes in mind and brain. *Neuron* **80,** 35–50 (2013).

27. Kampis, D. & Southgate, V. Altercentric cognition: How others influence our cognitive processing. *Trends in Cognitive Sciences* **24,** 945–959 (2020).

28. Mayo, A. T. & Woolley, A. W. Variance in group ability to transform resources into performance, and the role of coordinated attention. *Academy of Management Discoveries* **7,** 225–246 (2021).

29. Shteynberg, G. Shared Attention. *Perspectives on Psychological Science* **10,** 579–590 (2015).

30. Chikersal, P., Tomprou, M., Kim, Y. J., Woolley, A. W. & Dabbish, L. *Deep structures of collaboration: Physiological correlates of collective intelligence and group satisfaction* in *Proceedings of the 2017 ACM conference on computer supported cooperative work and social computing* (Association for Computing Machinery, New York, NY, United States, 2017), 873–888.

31. Tomprou, M., Kim, Y. J., Chikersal, P., Woolley, A. W. & Dabbish, L. A. Speaking out of turn: How video conferencing reduces vocal synchrony and collective intelligence. *PLOS ONE* **16,** e0247655 (2021).

32. Riedl, C. & Woolley, A. W. Teams vs. crowds: A field test of the relative contribution of incentives, member ability, and emergent collaboration to crowd-based problem solving performance. *Academy of Management Discoveries* **3,** 382–403 (2017).

33. Weng, L., Flammini, A., Vespignani, A. & Menczer, F. Competition among memes in a world with limited attention. *Scientific reports* **2,** 335 (2012).

34. Nagai, Y., Hosoda, K., Morita, A. & Asada, M. A constructive model for the development of joint attention. *Connection Science* **15,** 211–229 (2003).

35. Kaplan, F. & Hafner, V. V. The challenges of joint attention. *Interaction Studies* **7,** 135–169 (2006).

36. Vasil, J., Badcock, P. B., Constant, A., Friston, K. & Ramstead, M. J. A world unto itself: human communication as active inference. *Frontiers in Psychology* **11,** 480375 (2020).

37. Bergenholtz, C., Vuculescu, O. & Amidi, A. Microfoundations of adaptive search in complex tasks: the role of cognitive abilities and styles. *Organization Science* **34,** 2043–2063 (2023).

38. Hawkins, R. D. *et al.* From partners to populations: A hierarchical Bayesian account of coordination and convention. *Psychological Review* **130,** 977 (2023).

39. Laureiro-Martinez, D., Arrieta, J. P. & Brusoni, S. Microfoundations of problem solving: Attentional engagement predicts problem-solving strategies. *Organization Science* **34,** 2207–2230 (2023).

40. Tambe, M. Towards Flexible Teamwork. *Journal of Artificial Intelligence ResResearch* **7,** 83–124 (1997).

41. Li, H. *et al.* Individualized Mutual Adaptation in Human-Agent Teams. *IEEE Transactions on Human-Machine Systems* **51,** 706–714 (2021).

42. Bendell, R., Williams, J., Fiore, S. & Jentsch, F. *Interventions by Artificial Socially Intelligent Agents in Collaborative Environments: Impacts on Team Performance and Knowledge Externalization* in *Artificial Intelligence, Social Computing and Wearable Technologies* (AHFE Conference, Honolulu, Hawaii, 2023), 165–175.

43. Williams, J., Bendell, R., Fiore, S. & Jentsch, F. The Role of Artificial Theory of Mind in Supporting Human-Agent Teaming Interactions. *Human Factors and Simulation* **83,** 46–56 (2023).

44. Pandya, R., Huang, S. H., Hadfield-Menell, D. & Dragan, A. *Human-AI Learning Performance in Multi-Armed Bandits* in *Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society* (Association for Computing Machinery, New York, NY, USA, 2018), 369–375.
45. Giles, H., Ogay, T., *et al.* in *Explaining communication: Contemporary theories and exemplars* 293–310 (Taylor & Francis, 4 Park Square, Milton Park, Abingdon, Oxfordshire, OX14 4RN, United Kingdom, 2007).
46. Sidnell, J. & Stivers, T. *The Handbook of Conversation Analysis* (John Wiley & Sons, 605 Third Ave. New York NY 10158 USA, 2012).
47. Wachsmuth, I., De Ruiter, J., Jaecks, P. & Kopp, S. *Alignment in Communication: Towards a New Theory of Communication* (John Benjamins Publishing, Amsterdam, the Netherlands, 2013).
48. Kalociński, D., Mostowski, M. & Gierasimczuk, N. Interactive semantic alignment model: Social influence and local transmission bottleneck. *Journal of Logic, Language and Information* **27,** 225–253 (2018).
49. Garrod, S. & Anderson, A. Saying what you mean in dialogue: A study in conceptual and semantic co-ordination. *Cognition* **27,** 181–218 (1987).
50. Brennan, S. E. & Clark, H. H. Conceptual pacts and lexical choice in conversation. *Journal of Experimental Psychology: Learning, Memory, and Cognition* **22,** 1482 (1996).
51. Metzing, C. & Brennan, S. E. When conceptual pacts are broken: Partner-specific effects on the comprehension of referring expressions. *Journal of Memory and Language* **49,** 201–213 (2003).
52. Hu, Z., Halberg, G., Jimenez, C. R. & Walker, M. in *Situated Dialog in Speech-Based Human-Computer Interaction* 151–164 (Springer Publishing Company, 11 W 42nd St, Ste 15, New York, NY, 2014).
53. Bansal, G. *et al. Updates in human-AI teams: Understanding and addressing the performance/compatibility tradeoff* in *Proceedings of the Thirty-Third AAAI Conference on Artificial Intelligence* (AAAI Press, Washington DC, 2019), 2429–2437.
54. Wang, D. *et al.* Human-AI collaboration in data science: Exploring data scientists' perceptions of automated AI. *Proceedings of the ACM on Human-Computer Interaction* **3,** 1–24 (2019).
55. Dennis, A. R., Lakhiwal, A. & Sachdeva, A. AI agents as team members: Effects on satisfaction, conflict, trustworthiness, and willingness to work with. *Journal of Management Information Systems* **40,** 307–337 (2023).
56. Demir, M., Mcneese, N. J. & Cooke, N. J. The Evolution of Human-Autonomy Teams in Remotely Piloted Aircraft Systems Operations. *Frontiers in Communication* **4,** unknown (2019).
57. Harrison, D., Mohammed, S., McGrath, J., Florey, A. & Vanderstoep, S. Time Matters in Team Performance: Effects of Member Familiarity, Entrainment, And Task Discontinuity On Speed And Quality. *Personnel Psychology* **56,** 633–669 (2003).
58. Balachandar, N., Dieter, J. & Ramachandran, G. S. Collaboration of AI Agents via Cooperative Multi-Agent Deep Reinforcement Learning. *ArXiv* **abs/1907.00327** (2019).
59. Reichel, U., Benus, S. & Mády, K. Entrainment profiles: Comparison by gender, role, and feature set. *Speech Communication* **100,** 46–57 (2018).
60. Reid, C. A., Green, J. D. & Davis, J. L. Attitude alignment increases trust, respect, and perceived reasoning ability to produce attraction. *Personal Relationships* **25,** 171–189 (2018).
61. Weatherholtz, K., Campbell-Kibler, K. & Jaeger, T. F. Socially-mediated syntactic alignment. *Language Variation and Change* **26,** 387–420 (2014).
62. Memmert, L. & Bittner, E. *Complex Problem Solving through Human-AI Collaboration: Literature Review on Research Contexts* in *Proceedings of the 55th Hawaii International Conference on System Sciences* (University of Hawai'i at Mānoa, Shidler College of Business, Information Technology Department, 2404 Maile Way, Honolulu, HI 96822, 2022), 0-0.
63. Andrews, R. W., Lilly, J. M., Srivastava, D. & Feigh, K. M. The role of shared mental models in human-AI teams: a theoretical review. *Theoretical Issues in Ergonomics Science* **24,** 129–175 (2023).
64. Erdogan, E., Dignum, F., Verbrugge, R. & Yolum, P. in *HHAI2022: Augmenting Human Intellect* 199–211 (IOS Press, 6751 Tepper Drive, Clifton, VA 20124, 2022).

65. Erdogan, E., Dignum, F., Verbrugge, R. & Yolum, P. *Computational Theory of Mind for Human-Agent Coordination* in *International Workshop on Coordination, Organizations, Institutions, Norms, and Ethics for Governance of Multi-Agent Systems* (Springer, Cham, Switzerland, 2022), 92–108.
66. Blackman, R. *Ethical Machines: Your Concise Guide to Totally Unbiased, Transparent, and Respectful AI* (Harvard Business Press, Cambridge, MA, 2022).
67. Music, S. & Hirche, S. Control sharing in human-robot team interaction. *Annu. Rev. Control.* **44,** 342–354 (2017).
68. Haindl, P., Buchgeher, G., Khan, M. & Moser, B. *Towards a Reference Software Architecture for Human-AI Teaming in Smart Manufacturing* in *2022 IEEE/ACM 44th International Conference on Software Engineering: New Ideas and Emerging Results (ICSE-NIER)* (IEEE, 3 Park Avenue, 17th Floor, New York City, NY, 2022), 96–100.
69. Bendell, R., Williams, J., Fiore, S. & Jentsch, F. Supporting Social Interactions In Human-Ai Teams: Profiling Human Teammates From Sparse Data. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* **65,** 665–669 (2021).
70. Stowers, K., Brady, L. L., Maclellan, C., Wohleber, R. & Salas, E. Improving Teamwork Competencies in Human-Machine Teams: Perspectives From Team Science. *Frontiers in Psychology* **12,** 1–6 (2021).
71. Ezer, N. *et al.* Trust Engineering for Human-AI Teams. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* **63,** 322–326 (2019).
72. Sawant, S. *et al.* Balancing the Scales of Explainable and Transparent AI Agents within Human-Agent Teams. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* **67,** 2082–2087 (2023).
73. Coupland, J., Coupland, N. & Giles, H. in *Contexts of Accommodation* 1–68 (Cambridge University Press, Cambridge, MA, 1991).
74. Brennan, S. E. Conversation with and through computers. *User Modeling and User-Adapted Interaction* **1,** 67–86 (1991).
75. Pickering, M. J. & Garrod, S. Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences* **27,** 169–190 (2004).
76. Iio, T. *et al.* Lexical Entrainment in Human Robot Interaction. *International Journal of Social Robotics* **7,** 253–263 (2015).
77. Kejriwal, J. & vStefan Bevnuvs. *Relationship between auditory and semantic entrainment using Deep Neural Networks (DNN)* in *INTERSPEECH 2023* (ISCA, unknown, 2023), 2623–2627.
78. Frank, S. & Yang, J. Lexical representation explains cortical entrainment during speech comprehension. *PLOS ONE* **13,** unknown (2017).
79. Branigan, H. P., Pickering, M. J., Pearson, J., McLean, J. F. & Nass, C. *Syntactic alignment between computers and people: The role of belief about mental states* in *Proceedings of the 25th Annual Conference of the Cognitive Science Society* **31** (Taylor & Francis, Oxfordshire, UK, 2003), 186–191.
80. Branigan, H. P., Pickering, M. J., Pearson, J. & McLean, J. F. Linguistic alignment between people and computers. *Journal of Pragmatics* **42,** 2355–2368 (2010).
81. Clark, H. H. & Wilkes-Gibbs, D. Referring as a collaborative process. *Cognition* **22,** 1–39 (1986).
82. Reagans, R. E., Volvovsky, H. & Burt, R. S. Shared language in the team network-performance association: Reconciling conflicting views of the network centralization effect on team performance. *Collective Intelligence* **2,** 26339137231199739 (2023).
83. McNeese, N. J., Demir, M., Cooke, N. J. & Myers, C. Teaming with a synthetic teammate: Insights into human-autonomy teaming. *Human Factors* **60,** 262–273 (2018).
84. Schelble, B. G., Flathmann, C., McNeese, N. J., Freeman, G. & Mallick, R. Let's Think Together! Assessing Shared Mental Models, Performance, and Trust in Human-Agent Teams. *Proceedings of the ACM on Human-Computer Interaction* **6,** 1–29 (2022).
85. Zhang, R., McNeese, N. J., Freeman, G. & Musick, G. "An Ideal Human": Expectations of AI Teammates in Human-AI Teaming. *Proceedings of the ACM on Human-Computer Interaction* **4,** 1–25 (2021).
86. Angrist, J. D. & Pischke, J.-S. *Mostly harmless econometrics: An empiricist's companion* (Princeton University Press, Princeton, NJ, 2009).
87. Vasconcelos, H. *et al.* Explanations Can Reduce Overreliance on AI Systems During Decision-Making. *Proceedings of the ACM Human-Computer Interaction* **7** (Apr. 2023).

88. Mohammed, S., Ferzandi, L. & Hamilton, K. Metaphor no more: A 15-year review of the team mental model construct. *Journal of management* **36,** 876–910 (2010).
89. Lewis, K. Measuring transactive memory systems in the field: scale development and validation. *Journal of Applied Psychology* **88,** 587 (2003).
90. Johnson, T. E. *et al.* Measuring sharedness of team-related knowledge: Design and validation of a shared mental model instrument. *Human Resource Development International* **10,** 437–454 (2007).
91. Glikson, E. & Woolley, A. W. Human trust in artificial intelligence: Review of empirical research. *Academy of Management Annals* **14,** 627–660 (2020).
92. Pan, Y. & Steed, A. A comparison of avatar-, video-, and robot-mediated interaction on users' trust in expertise. *Frontiers in Robotics and AI* **3,** 12 (2016).
93. Wang, W. & Benbasat, I. Recommendation agents for electronic commerce: Effects of explanation facilities on trusting beliefs. *Journal of Management Information Systems* **23,** 217–246 (2007).
94. Portes, A. Social capital: Its origins and applications in modern sociology. *Knowledge and Social Capital* **24,** 43–67 (1998).
95. Gallotti, M. & Frith, C. D. Social cognition in the we-mode. *Trends in Cognitive Sciences* **17,** 160–165 (2013).
96. Swol, L. M. V., Ahn, P. H., Prahl, A. & Gong, Z. Language Use in Group Discourse and Its Relationship to Group Processes. *Sage Open* **11,** 1–7 (2021).
97. Garrod, S., Freudenthal, D. & Boyle, E. The Role of Different Types of Anaphor in the On-Line Resolution of Sentences in a Discourse. *Journal of Memory and Language* **33,** 39–68 (1994).
98. Theiner, G., Allen, C. & Goldstone, R. L. Recognizing group cognition. *Cognitive Systems Research* **11,** 378–395 (2010).
99. Goldstone, R. L. & Theiner, G. The multiple, interacting levels of cognitive systems (MILCS) perspective on group cognition. *Philosophical Psychology* **30,** 338–372 (2017).
100. Schoeffer, J., De-Arteaga, M. & Kühl, N. *Explanations, Fairness, and Appropriate Reliance in Human-AI Decision-Making* in *Proceedings of the CHI Conference on Human Factors in Computing Systems* (Association for Computing Machinery, 2024), 1–18. ISBN: 9798400703300.
101. Tolmeijer, S., Christen, M., Kandul, S., Kneer, M. & Bernstein, A. *Capable but Amoral? Comparing AI and Human Expert Collaboration in Ethical Decision Making* in *Proceedings of the CHI Conference on Human Factors in Computing Systems* (Association for Computing Machinery, New York, NY, USA, 2022), 1–17. ISBN: 9781450391573.
102. Liang, C., Proft, J., Andersen, E. & Knepper, R. A. *Implicit Communication of Actionable Information in Human-AI teams* in *Proceedings of the CHI Conference on Human Factors in Computing Systems* (Association for Computing Machinery, New York, NY, United States, 2019), 1–13.
103. Zhang, R. *et al.* Investigating AI Teammate Communication Strategies and Their Impact in Human-AI Teams for Effective Teamwork. *Proceedings of the ACM on Human-Computer Interaction* **7,** 1–31 (2023).
104. Valencia, S. *et al. "The less I type, the better": How AI Language Models can Enhance or Impede Communication for AAC Users* in *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Association for Computing Machinery, New York, NY, USA, 2023), 1–14. ISBN: 9781450394215.
105. Samadi, M. A. *et al. Minds and Machines Unite: Deciphering Social and Cognitive Dynamics in Collaborative Problem Solving with AI* in *LAK '24: Proceedings of the 14th Learning Analytics and Knowledge Conference* (Association for Computing Machinery, New York, NY, USA, 2024), 885–891. ISBN: 9798400716188.
106. Mieczkowski, H., Hancock, J. T., Naaman, M., Jung, M. F. & Hohenstein, J. AI-Mediated Communication. *Proceedings of the ACM on Human-Computer Interaction* **5,** 1–14 (2021).
107. Cheng, S. When Journalism Meets AI: Risk or Opportunity? *Digital Government: Research and Practice* (2024).
108. Fulker, Z. & Riedl, C. Cooperation in the Gig Economy: Insights from Upwork Freelancers. *Proceedings of the ACM on Human-Computer Interaction* **8,** 1–20 (2024).
109. Vallas, S. & Schor, J. B. What do platforms do? Understanding the gig economy. *Annual Review of Sociology* **46,** 273–294 (2020).
110. Schelble, B. G. *et al.* Investigating the Effects of Perceived Teammate Artificiality on Human Performance and Cognition. *International Journal of Human–Computer Interaction* **39,** 2686–2701 (2022).

111. Ashktorab, Z. *et al.* Human-AI Collaboration in a Cooperative Game Setting. *Proceedings of the ACM on Human-Computer Interaction* **4,** 1–20 (2020).