Assessing Cognitive and Social Awareness among Group Members in AI-assisted Collaboration

Sander de Jong
Aalborg University
Aalborg, Denmark
sanderdj@cs.aau.dk

Joel Wester
Aalborg University
Aalborg, Denmark
joelw@cs.aau.dk

Tim Schrills
University of Lübeck
Lübeck, Germany
tim.schrills@uni-luebeck.de

Kristina S. Secher
Aalborg University
Aalborg, Denmark
kristinaskjodt@gmail.com

Carla F. Griggio
Aalborg University
Copenhagen, Denmark
cfg@cs.aau.dk

Niels van Berkel
Aalborg University
Aalborg, Denmark
nielsvanberkel@cs.aau.dk

Abstract

Successful collaboration in computer-mediated teams requires awareness among group members of each other's knowledge, skills, and goals. Large Language Models (LLMs) can play a mediating role in establishing and maintaining this awareness among group members. In an in-situ study, we explored the impact of an LLM-based chatbot on cognitive and social group awareness through a distributed text-based group task. We instructed participants ($N = 48$) to complete a travel-planning task in sixteen groups of three, with each member given conflicting goals. Each chat was complemented by a chatbot that could be asked for assistance. Through a survey and semi-structured interview, we gained insight into participants' deliberations on the task and the chatbot's role. We found that the chatbot's presence helped increase group awareness as users are forced to clearly and transparently formulate their intentions when prompting the chatbot. The chatbot's ability to provide suggestions that compromise between user goals based on the chat history helped participants reach a consensus. We present implications for the design of chatbots for collaborative settings.

CCS Concepts

• **Human-centered computing → HCI design and evaluation methods**; **Collaborative interaction**.

Keywords

human-AI teams, artificial intelligence, Large Language Models, collaboration, group awareness

ACM Reference Format:
Sander de Jong, Joel Wester, Tim Schrills, Kristina S. Secher, Carla F. Griggio, and Niels van Berkel. 2024. Assessing Cognitive and Social Awareness among Group Members in AI-assisted Collaboration. In *International Conference on Mobile and Ubiquitous Multimedia (MUM '24), December 01–04, 2024, Stockholm, Sweden.* ACM, New York, NY, USA, 13 pages. https://doi.org/10.1145/3701571.3701582

MUM '24, December 01–04, 2024, Stockholm, Sweden

ACM ISBN 979-8-4007-1283-8/24/12
https://doi.org/10.1145/3701571.3701582

1 Introduction

Digital tools play an increasing role in the completion of cooperative tasks. Following the 'Computers Are Social Actors' paradigm [54], humans react socially to computers and consider them team members [53]. Therefore, the use of novel tools—such as the recent trend of large language models (LLMs)—raises questions about how group dynamics are changed and what characteristics of LLMs change established concepts in collaboration, including end-users' mental models of collaboration between team members. To aid team members in effective collaborative decision-making, establishing a shared mental model between team members is critical [16], but LLM features may be perceived differently and can hamper forming shared mental models. Following Fransen et al.'s framework on shared expert team decision-making, a shared mental model can be achieved through '*group awareness*', i.e., by the group members' acquisition and interpretation of information from the environment in order to update and monitor team performance and expertise [27]. Group awareness is a multidimensional construct including, for example, cognitive, and social group awareness [10], representing an individual's acquisition, interpretation, and projection of information on their team members' activities, knowledge, and skills, as well as the functioning of the whole group [39].

While the amount of group awareness information can vary between collaborative settings, computer-mediated collaborations can exhibit features that complicate group awareness acquisition, such as asynchronous activity or limited access to cues typical in face-to-face interactions [32]. Currently, the effects of the design of LLM features on reaching a common understanding of a problem have become an important research objective, as the usage of LLM in collaborative settings rises [41]. As aligning group awareness through computer-mediated communication requires additional support [11, 22], optimal design of LLMs in collaborative settings should be informed by a deeper understanding of their effects on group awareness. Group awareness in computer-mediated collaborations is further affected by the presence of digital agents that act as team members. For example, Kim et al. found that rule-based chatbots can play a moderating role in online group discussions, leading to more diverse opinions and better-perceived consensus forming [42]. In a human-AI teaming study, participants sought out a rule-based AI team member's opinions on conflicts with other human team members [73], indicating its potential as a conflict mediator. The capabilities of LLM-based chatbots to keep track of the context of a conversation may be promising features for mediation in decision-making scenarios. By combining team members' input, chatbots can utilise a group's conversation to contextualise its replies and adequately respond to conflicts. If chatbots can reconcile the team members' viewpoints, they can be used as tools to reach a consensus or provide alternatives the team has not yet considered. For the potential use of LLMs as an enabler in collaborative scenarios, it is important to understand which novel features of LLMs allow for harmonised and fair team processes through improved group awareness and which may be insufficient to support group processes [67].

We are interested in understanding how LLM-based chatbots affect cognitive and social awareness among group members in a collaborative, remote text-based task and to address the existing research gap on LLM features in relation to group awareness. To this end, we conduct an in-situ qualitative observation study in which groups of three physically distributed participants plan a week-long holiday via Discord, each setting out with distinct goals to achieve. Participants have access to a chatbot to provide on-demand assistance. Following the task completion, we interview participants to gain insight into their opinions about the chatbot. We analyse the data deductively and inductively informed by group awareness concepts derived from Janssen and Bodemer's framework [38].

Our results indicate that the presence of an LLM-powered chatbot can increase users' group awareness and help mediate between conflicting user goals in two distinct ways. First, with LLMs typically restating a summary of user requests, they make the different users' goals explicit in a shared chat, increasing the group's awareness of each other's preferences. LLM-based chatbots hereby support the general acquisition of cognitive group awareness by explicitly stating individual goals, making them obtainable for the group. Second, LLM responses that provide a compromise between aforementioned user preferences helped participants reach a consensus on the final decisions. Participants expressed that their preferences were taken into account in the chatbot suggestions. As LLM-based chatbots reiterate information to formulate an answer (e.g., compromises), the general knowledge awareness of the group is improved. Consequently, group members may feel more represented in the proposed consensus. Further, participants attributed authority to the chatbot and generally accepted its explanations to settle disputes. This initial competence attribution waned in cases where participants observed clear mistakes or when their requests were denied.

Our findings inform the design of AI system features that support text-based group discussions and collaboration (e.g., restating a summary of user requests in a collaborative context). While we found that the chatbot performed well in aligning user goals and was generally appreciated by participants in this mediating role, additional explanations would have helped to manage participant expectations. Informing users about the best use cases and approaches for using the bot would further help to improve their awareness of the chatbot's skills. Additionally, participants suggested that a proactive chatbot would result in the chatbot being more part of the group, and could play a useful role in settling disagreements by providing potential solutions. Critically, such a proactive configuration should be designed with user agency in mind, as participants also expressed that proactive prompts could be intrusive and negatively impact user initiative.

2 Related work

2.1 Group Awareness

Computer-based communication in collaborative contexts can limit the amount of information on different aspects of team members, for example, their opinions, skills, activity, or knowledge. Yet, obtaining and utilising information on such aspects of team members, also referred to as group awareness [61], is crucial for effective collaboration. Group awareness (or team awareness) can, for example, support collaborative learning, as it guides different group members on pacing [45], enable efficient communication by revealing (complementary) knowledge gaps, or establishing otherwise opaque rules or norms on processes between team members [58]. For the context of our study, group awareness, as defined by [61], offers a more extended framework to explore human behaviour in interaction with LLM-based chatbots.

The effectiveness of teams collaborating on a shared task is often discussed in terms of shared mental models [57]. Mental models are subjective, context-dependent knowledge structures applied by humans (e.g., to interact with complex systems [16]). In team contexts, so-called shared mental models are used to describe the congruence of mental models among team members. Shared mental models are crucial in determining the quality of collaboration, as they support the avoidance of misunderstandings. Highly congruent shared mental models are based on information team members have about the nature of the task and are able to support individual understanding and confidence about their roles [57].

In comparison to shared mental models, group awareness explicitly describes an individual's state and perception of a group and its individuals rather than the overlap or alignment of multiple perspectives. While group awareness potentially supports the alignment of mental models [27], it is dependent on user perception and attention, while mental models represent a knowledge structure. As such, group awareness allows the examination of individual psychological processes that may lead to improved shared mental models.

For the present study, we differentiate cognitive and social group awareness, which reflect an individual's awareness of team members' information as well as team members' roles and intentions within a collaborative task as based on Ghadirian et al. [28].

2.1.1 *Cognitive Group Awareness.*

Similar to shared mental models, group awareness is usually not depicted as a unidimensional concept, but addresses multiple levels of awareness about other team members' states. For example, awareness about cognitive processes in others can influence an individual's communication. Cognitive group awareness is the awareness that results from information about group members' knowledge, the information they possess, or the opinions they hold, all of which can be used to coordinate collaborative activities [38]. For example, cognitive group awareness was facilitated previously by initial knowledge quizzes and depiction of the results [23]. It was shown that highlighting knowledge gaps can guide team members when prioritising activities. In accordance, fostering cognitive group awareness could be a beneficial feature of LLM chatbots (e.g., by pointing out missing information or summarising existing information).

2.1.2 Social Group Awareness. Social group awareness deals with individual activities and team composition (e.g., the current position in a shared document or the current task in a collaborative project). Accordingly, social group awareness not only refers to the perception of the other as a 'real' person but also to awareness about what group members are doing, with whom they are communicating, how they are contributing to the common group goal, or how they view and behave in their respective roles [38]. For example, previous research demonstrated that visualising the flow of interactions between group members supported knowledge construction within a team, where only demonstrating the amount of contribution had no effect on group processes [47].

2.2 Computer-Mediated Support for Group Awareness in Distributed Collaboration

Designing support for group awareness has been a longstanding focus within the HCI and CSCW communities, with foundational works dating back decades [1, 20, 21, 52, 64]. For example, Dourish and Belotti explored group awareness and coordination in relation to a shared editor [20], highlighting passive awareness about the activities of others as a key mechanism to support collaboration. Such awareness tools are nowadays commonplace in collaborative software. For example, colour-coded or labelled cursors can help team members locate each other in shared collaborative environments such as text authoring tools [46], programming environments [12], or digital whiteboards (e.g., Miro); notifications alert team members of changes or annotations done by others in their absence [14] (e.g., new content or comments on a Google Doc), and histories of changes often label traces of past activity in the shared workspace.

Tailored support for group awareness has improved team performance in a range of domains and collaborative tasks, such as sensemaking [30], searching [51, 55], and text authoring [68]. For example, Hong et al. [35] showed that communication costs in decision-making tasks decrease when visualising team members' preferences on diverse criteria (e.g., showing each member's filter ranges on the same query). Grønbæk et al. [31] designed a video conferencing system with draggable, resizable video feeds for hybrid meetings, which increased awareness of group members' focus and activities between remote and collocated collaborators. Teams performing sensemaking tasks with visual analytics can benefit from increased group awareness by sharing visualisations of each other's data selections and queries [34] as well as each member's hypotheses and collected evidence [29]. In collaborative learning environments, visualising one's own perception of social behaviour within a group (e.g., addressing others' requests) [56], as well as peers' prior knowledge on a topic [59], can enhance social performance and learning outcomes. Wijenayake et al. found that increased group awareness can also negatively impact individual behaviour, showing an increase in social conformity as a result of social presence indicators [70].

Especially relevant to our work is prior research on ways to align group awareness using text-based communication. Gutwin et al. found that software developers were able to maintain group awareness through primarily text messages [33]. A culture of sharing questions in public channels helped developers identify the person they needed for specific tasks. This culture of 'overhearing' conversations in public channels helped align group awareness. In the context of information searching, Shah et al. found that having an overview of saved search results and queries used reduced the time group members spent on coordination [62]. Similarly, groups interacting with search chatbots integrated into a Slack chat had greater awareness of group members' activities and found it easier to share information and reach a consensus [6]. Group awareness was further improved by restricting participants to the shared Slack chat, as compared to allowing them to copy results from web searches [5].

In this paper, we focus on group awareness in chat settings that include artificial conversation partners. As group awareness can unfold in many ways (e.g., being aware of each other's activity vs. each other's goals), a conceptualisation of group awareness in LLM-mediated contexts is key towards the adoption of this technology in collaborative contexts.

2.3 Human-AI Teaming

Collaboration between humans and technology is an enduring challenge. In the early 2000s, Klien et al. outlined open challenges for enabling systems that are good team players [44]. Part of these are related to aligning awareness between the system and human team members. For example, intelligent agents must be able to adequately model the other participants' intentions and actions and, likewise, be able to make their status and intentions obvious to their teammates.

In recent years, human-agent teaming challenges have been approached from various domains such as HCI, HRI, and other neighbouring domains [8, 25]. Recently, Demir et al. investigated how human-robot teams can take inspiration from human-human teams, suggesting *shared cognition* as an important factor in human-robot teaming [18]. However, their results also indicate that team predictability is only desirable to some degree, which challenges Klien et al.'s point that team members must be mutually predictable.

A common theme in the research on human-AI teaming is the composition of these teams. For example, McNeese et al. compared four different team compositions: human-only, human-human-AI, human-AI-AI, and AI-only in terms of performance, situation awareness, and perceived team cognition [49]. Their results indicated lower scores when teams were composed of human-only team members and higher scores of mixed human-AI teams, including the AI-only composition. Similarly, Schelbe et al. explored the effects of trust, team performance, and team cognition within human-agent teams depending on team dynamics, suggesting that action-related communication and shared goals are beneficial to team cognition, although team cognition decreased with agent teammates compared to human ones [60]. Moreover, human-agent teams' trust decreased with agent teammates when working with only agents and no humans.

Do et al. investigated imperfect interventions in a collaborative environment and how conversational agents can act in such situations [19]. Participants were sent private messages that they were under-contributing to the discussion and had to assess their agreement with this evaluation. Participants who were falsely detected as under-contributing by the CA increased their participation. Under-contributing participants who were falsely detected as contributing equally did not increase their participation but reduced their awareness of their and other group members' participation.

Suh et al. studied collaborative music composition between pairs with and without the help of an AI team member [65]. The AI system played a mediating role in the social dynamics during the collaborative composition as participants used the AI's generated output to find common ground. The AI's suggestions throughout the session also opened up new ideas in various ways. It provided initial content as a starting point, narrowed down abundant options to concrete ideas, generated alternatives when participants were stuck, and filled in unfinished composition parts. When participants disagreed, the AI composer could mediate by creating alternative ideas for the composition's continuation. Zheng et al. studied human-AI teams using an essay scoring task [73]. Participants were asked to collectively rate essays in collaboration with an AI system. Participants appreciated the critical questions the AI system asked and suggested that future critical AI systems could take the role of a *conflict mediator* or *challenger*.

Building on these works on teams collaborating with AI systems, we explore how aspects of a chatbot's cognitive and social group awareness impact user behaviour in a collaborative task setting.

3 Method

We conducted an in-situ, qualitative observation study to investigate the impact of an LLM-based chatbot on group dynamics in the context of a distributed collaborative task. The study was designed to investigate the impact of LLM-based chatbots from different perspectives. For example, we explored how groups collaboratively discover opportunities for leveraging LLMs in their conversations, studied user perceptions of the chatbot as a tool or a group member, and elicited speculative opinions on alternative chatbot behaviours (e.g., participating in the conversation without explicit prompts). To this end, we focus on understanding how groups adopt LLM-based chatbots to influence and leverage group awareness. The observational study consisted of groups of three participants chatting on the Discord messaging app[1] during a collaborative travel-planning task.

3.1 Apparatus

We created a Discord chatbot that connects with the LLM (GPT-3.5-turbo, May 2023 version). We open a new channel on the Discord channel for each group and hide the other channels. All messages sent in the Discord channel are forwarded through the OpenAI API as individual prompts in the same session in order for the LLM to have access to the entire conversation history. The bot only shows a LLM-generated response in the chat when a participant adds the prefix 'ASK' before a message so that the chat does not overflow with automated, unwanted LLM responses and participants feel full agency over when to prompt the chatbot. Keeping an open conversation with the LLM allows us to explore how it may use previous messages to inform its responses to explicit participant prompts, and to observe whether this influences group awareness. We tested the apparatus in a pilot study to ensure the task was clear and to verify whether the conversation would be properly tracked and would not be hindered by hallucinations.

[1]https://discord.com/

Table 1: The description of the persona goals as communicated to participants.

Personas	Goals
Persona 1	Go on a city trip See historical landmarks and interesting architecture Visit museums Do a lot of sightseeing
Persona 2	Do not spend too much money on the trip Arrange some free activities
Persona 3	Relax on a beach holiday Enjoy lots of sun Go swimming Limit the number of activities

3.2 Procedure

Groups are asked to plan a trip on a Discord channel in which our chatbot is active. Travel planning tasks require group coordination to align individual preferences and are therefore often used in collaborative information retrieval studies [24, 71]. Participants in our study can similarly use the chatbot as an information retrieval tool but can additionally use its reasoning capabilities to summarise and contextualise the acquired information. We do not instruct participants on how to use the chatbot to prevent biasing their behaviour towards the chatbot. The entire study takes around 45 minutes per group, with 15 minutes dedicated to the travel-planning task.

3.2.1 Group Setup and Instructions. Each group member is assigned the role of a persona that focuses on either the destination, activities, or budget of the trip (Table 1). We intentionally designed the personas to conflict with each other to motivate participants to explicitly seek agreement. Participants are instructed to plan a trip together trying to achieve their assigned goals without directly revealing them to the other members. They are also informed about the availability of a travel assistant chatbot that can answer questions when they type 'ASK' before their messages. After obtaining informed consent from participants, we separate them into different rooms and ask them to communicate exclusively through our Discord channel and avoid using the audio channel.

3.2.2 Demographics and Profile Questionnaire. When participants join the Discord channel, the study facilitator asks them to complete a questionnaire dedicated to describing the participant sample, including questions about:

- Demographic information: gender, age, and education.
- Experience with chatbots on a 5-point Likert scale ('none at all' - 'a great deal').
- Experience with LLMs on a 5-point Likert scale ('none at all' - 'a great deal').

- The medium where participants normally plan a holiday (single choice: in group chats online, in person, or other with the option to specify other).
- The 9-item ATI scale for technical affinity, designed to assess the participant's tendency to engage or avoid technological interaction [26].

3.2.3 Travel Planning Chat Session. Once all group members finish answering the questionnaire, the study facilitator writes a message in the chat announcing that they can start planning the trip. Participants are given complete freedom about how to manage their conversation while planning the trip according to their personas' goals. While participants plan their trip, the facilitator and another researcher observe the chat and take notes on relevant interactions between participants as well as between participants and the chatbot. To ensure a similar level of engagement with the chatbot between participants, we limit the task duration to 15 minutes from the moment the first participant message is sent.

3.2.4 Post-Study Questionnaire. After the chat session ends, participants complete a second survey rating their experience on a list of 5-point Likert-scale questions:

- How useful was the travel assistant chatbot? ('not at all' - 'extremely')
- How likely are you to use a chatbot again for travel planning? ('extremely unlikely' - 'extremely likely')
- I was able to achieve the goals of my persona. ('strongly disagree' - 'strongly agree')
- The chatbot helped in achieving the goals of my persona. ('strongly disagree' - 'strongly agree')
- The group members contributed equally to the conversation. ('strongly disagree' - 'strongly agree')
- The group members prompted the chatbot equally. ('strongly disagree' - 'strongly agree')
- The chatbot was part of the group. ('strongly disagree' - 'strongly agree')

3.2.5 Post-Study Group Interviews. The study concludes with a group, face-to-face interview as an additional opportunity to observe the group's dynamics while discussing the chatbot's role in their decision-making process. Two interviewers are present during each interview following an agreed-upon semi-structured interview guide, mainly probing ways in which they perceived the chatbot to be useful and the role it played in their consensus-seeking process. Additionally, the interviewers ask groups to explain and reflect together about particular episodes of their chat session as captured in their observational notes.

3.3 Data Collection

We collect data from the following sources: (1) observational notes by the study facilitators during the chat sessions, (2) demographics and post-study questionnaires, (3) the transcript of chat messages from the study, and (4) audio-recorded group interviews and transcripts. Chat messages are pseudonymised by using aliases for participants' Discord users (e.g., Participant 1, Participant 2).

3.4 Participants

We recruited 48 participants (30 men, 18 women), with an average age of 24.52 (ranging from 21 to 31, $SD = 2.31$), whom we equally distributed among 16 groups, each consisting of three members. We recruited participants through convenience sampling, distributing the call for volunteers in student groups within our institution. This helped us recruit participants who already knew each other, allowing us to form groups with some level of existing group awareness. In this way, we seek to increase the internal validity and reduce the required time for the travel planning task. Moreover, we decided on a group size of three to limit the effect of social loafing (i.e., the tendency of people to put less effort into a collective task than an individual or co-operative task) as this is more prominent in larger groups [40].

3.5 Thematic Analysis

We conduct a thematic analysis [13] on all qualitative data. We apply a deductive and latent analysis approach to identify characteristics of cognitive and social group awareness in participants' interactions and perceptions of the chatbot. We adopt a constructivist perspective when coding and creating themes, meaning that the results of our analysis reflect our interpretation of the participants' statements and our observations of their behaviour, which are informed by our underlying interest in identifying opportunities for the design of chatbots to support mediated collaboration. Two of the paper's authors start by independently selecting potential codes from Janssen et al.'s categorisation of group awareness [38], for example, concepts related to cognitive group awareness (e.g., awareness of others' knowledge, skills, and opinions) and social group awareness (e.g., awareness of what others are doing, their roles, and their contributions to a common goal). We then use these codes to inform the first step of our deductive analysis, focusing on patterns in the data that illustrate instances of interactions among the participants and between them and the chatbot relevant to different aspects of group awareness. During the coding phase, we also generate our own codes with more specific descriptions of patterns in the data, such as different roles assigned to the chatbot (e.g., the authority figure) and barriers to activating its capabilities (e.g., novelty effects). Last, we iteratively construct themes around the most representative codes in the data, reflecting on how different patterns in the data point to interesting roles of the chatbot in acquiring group awareness and to what extent participants treated the chatbot as a collaborator.

4 Results

We present participant characteristics based on survey data and subsequently outline the results of our thematic analysis, structured as patterns of cognitive and social group awareness we found.

The participants' affinity for technology interaction (ATI: $M = 3.12$, $SD = 0.59$) was close to 3.5, which is the assumed average of the population based on quota sampling [26]. Males ($M = 3.29$ $SD = 0.52$) scored higher on ATI as compared to females ($M = 2.83$, $SD = 0.60$), which is a significant difference ($t(32) = 2.71, p = 0.010$). We asked the participants to self-report their prior experience with chatbots and LLMs, and rate the usefulness of the chatbot. The participants had moderate experience with chatbots (M = 2.46, SD = 0.68) and LLMs (M = 2.56, SD = 0.82). Of the participants, 75% expressed that they plan their holidays at least partly online.

The participants found the chatbot useful for the travel planning task (M = 3.35, SD = 0.91) and would generally use it again for travel planning (M = 3.50, SD = 1.01). They expressed achieving the goals of their persona (M = 4.04, SD = 0.74), indicating that the chatbot helped reach their goals (M = 3.92, SD = 0.90). Group members expressed that they contributed equally to the conversation (M = 4.15, SD = 0.97) and prompted the chatbot an equal amount of times (M = 3.44, SD = 1.30). Participants generally felt that the chatbot was part of the group, but answers were dispersed (M = 3.10, SD = 1.36). We did not find significant differences between genders in terms of chatbot perceptions.

Next, we present our thematic analysis of the qualitative data, which consists of observational notes, transcripts from the chat logs, and participant interviews. We organise this analysis in accordance with the two overarching categories of group awareness: *cognitive group awareness* and *social group awareness*. The identifier PX.Y refers to the participant in group X with number Y in order to distinguish between groups and individual members within the groups.

4.1 Cognitive Group Awareness

Our results concerning cognitive group awareness highlight the effect of *transparency of information-seeking* on group awareness, negative effects of *information getting lost in the mix of messages and prompts*, and *activating the chatbot's capabilities*.

4.1.1 Expecting Transparency in Individual Information-Seeking Processes. Compared to a scenario where each member would look for recommendations on their own (e.g., with search engines or their own chatbot interfaces) and then share the outcomes with the group, the study required participants to prompt the chatbot from inside the group chat. This supported the acquisition of cognitive group awareness regarding the information-seeking processes of all members, making knowledge about the chatbot recommendations and the prompts that triggered them easily accessible.

> *"Normally, when we do something like this in a group chat to plan a holiday, it's always like somebody is using maybe an hour or something looking things up, and then they have to convert what they just found out to the group chat, while here, all the information is always in the group chat. The people have the same knowledge as the other people and you don't have to convey so much as when you just research on your own."* (P2.3)

Many participants believed that having the chatbot recommendations visible to everyone removed potential advantages for individuals. For example, one group mentioned that sharing personal preferences through prompts in a group chat contributed to a more equal knowledge distribution compared to selectively presenting individual search results to the group, thereby leading to an open discussion about the suggested recommendations. This brought a sense of transparency to all travel recommendations, ensuring that nobody would bias their recommendations in favour of their own preferences.

> *"If I was the only one that was able to see what it said, I could just say it recommended doing all the things I wanted to do, and no one would know."* (P13.3)

These examples highlight the importance participants place on their cognitive awareness of the source of the recommendations, wanting to be sure that they have access to the same information to trust that the outcome of the final decision would be fair towards all group members.

4.1.2 Getting Lost in the Mix of Group Chat and Prompting. While interacting with the chatbot within the group chat can transparently communicate each member's prompts, it can also hinder cognitive awareness if multiple members prompt simultaneously or too frequently. Participants wanted to avoid 'flooding' the conversation with prompts to the chatbot, as an excessive number of responses could cause them to miss others' contributions and relevant chatbot recommendations.

> *"You had one point where two of us asked kind of the same, and then it drops two prompts of the same. Information can easily drown in the chat. When it comes with like a huge box of text, and you're writing, it just disappears."* (P2.1)

> *"I think that I waited because I don't want to get a lot of prompts in the chat all of a sudden."* (P9.2)

This behaviour led some of the groups to suggest the separation of the group chat and the chatbot's answers.

> *"Maybe if we were writing in one text channel and have a separate text channel only for the bot we could flood that channel with everything that we want to ask about while still kind of knowing where we are in the conversation."* (P4.2)

A participant described hesitance towards starting a parallel conversation with the chatbot, highlighting that the ongoing discussion about an idea can dominate the conversation and prevent other inquiries.

> *"Yeah, it was also because you saw the person was writing. And if he's asked the first question about the flight, I just assumed that he would continue"* (P2.2)

There is a clear tension with Section 4.1.1, highlighting the challenge of providing an open channel for communication between participants and chatbot while preventing messages from getting lost. The potential for attaining cognitive group awareness by having an overview of all messages is lost when participants are not able to keep up with the conversation.

4.1.3 Activating the Chatbot's Capabilities. If users want to treat the chatbot as a group member, it is essential that the group members collectively know how to activate its capabilities. Many participants were surprised by the chatbot's capabilities and were not able to leverage its full potential.

The lack of participants' cognitive awareness of the chatbot is exemplified by the fact that many participants were surprised that the chatbot considered the wider conversational context. Participants generally perceived this as a positive surprise as the answers provided were more personalised.

> *"Yeah, I was kind of surprised on how well it works for plane tickets, choosing which months to travel and what's the cheapest option as well as activities when you Google like what to do in Bali. I don't expect things like that to just show up like some of them are."* (P7.2)

Specifically, participants were pleasantly surprised by the chatbot's capabilities to develop specific suggestions for locations and activities.

> *"It was a lot smarter than I thought, so I didn't think you would come up with that specific ideas. And so that was impressive."* (P2.2)

Apart from the surprise when it worked, participants also demonstrated their lack of awareness by requesting features the chatbot was already capable of, such as giving an overview of everything the group members discussed or providing answers based on the chat history.

> *"Yeah. Also, if it could remember previous questions and use that for the answer. For example, when I say, 'I want to go see museums', and some other guy says, 'I want to see nature', then it can propose in nature, but then say next also, 'If you want to see a museum, there's this place in the mountains that supports both'. I think that would be nice. Instead of it's always one of the other, and then you have to ask it directly: 'Can you combine these?'"* (P12.1)

In conclusion, participants were insufficiently able to form cognitive awareness of the chatbot through their prompts, leading to unused potential and participant surprise. This surprise could either be positive when the chatbot provided answers beyond participants' expectations or negative when answers to their requests were unsatisfactory or denied.

4.2 Social Group Awareness

We observed patterns of social group awareness due to the chatbot's presence, and participants elaborated on these in the interviews. In particular, our results focus on how *prompting aligns individual and common goals*, the authoritative role of the chatbot in *mediating consensus*, and how participants want to *reach consensus before prompting* the chatbot.

4.2.1 Aligning Individual and Common Goals. To obtain quality answers from the chatbot, participants expressed their travel planning preferences explicitly and unambiguously in their prompts. This not only made prompts clear to the chatbot but also to others reading the chat, thus increasing the group's cognitive awareness of each other's preferences. In some instances, participants used this knowledge to keep track of their assigned roles, hereby increasing their social group awareness.

For example, in group 16, the group member with the persona preferring a beach holiday and wanting to limit the number of activities was able to realise that the group had taken on too many activities. One of the group members asked for an overview of things to do in the city they wanted to visit and all group members adjusted this initial list by asking the chatbot for suggestions on activities in certain categories. The participant tasked with limiting activities saw that the itinerary became very long, asked the other participants to wait a moment, and prompted:

```
P16.2: ASK make a 7-day long vacation to Lisbon where
       the first half consists of visiting historical
       landmarks and museums. Incorporate night out
       activities. The other half of the vacation should
       include relaxing at the beach. The activities must
       be cheap.
```

The prompt indicates that the participant identified all three personal goals from the conversation and used these goals to ask the chatbot for a compromise while incorporating the relaxing activities missing from the previous itinerary. All participants were happy with the list and moved on to other aspects of the planning.

Multiple groups mentioned that reading each other's prompts gave them insights into others' deliberations and motivations, generating ideas for approaching the group's common goals. For example, P7.1 preferred a warm destination but knew that the other two members preferred to go skiing. At one point, P7.1 prompted the chatbot: *"ASK is there a place in Asia where you can go skiing and also enjoy the sun?"*. They continue deliberating ideas, and P7.3 asks if they should go hiking, to which P7.2 responds *"Yes, as long as there's a swimming pool for P7.1"*, and ends up prompting the chatbot about hiking locations with swimming pools.

> *"I think the chatbot gives some kind of discussion. It opens up for the other participants. I can see what they want to do and then I can sit down and tell myself, 'Do I want to do that? I want to follow that example. Do I want to come with my input? I like to do these things. Perhaps they want to do that as well.'"* (P7.3)

Keeping track of group members' roles and preferences through the chatbot prompts and answers improved participants' social group awareness, stimulating an open discussion on good compromises between all preferences and inspiring new ideas.

4.2.2 Mediating Consensus. When groups faced disagreements about their travel plans, they often assigned the role of a 'judge' to the chatbot rather than a member, treating it as an impartial source of knowledge or 'truth'. We observed this in situations where participants' interactions with the chatbot shifted from treating it as a mere tool to temporarily elevating it to the authority figure in the group to settle a conflict.

```
   P1.3: I'm pretty sure plane tickets are the cheapest
         option to go to anywhere that isn't three hours
         away.
   P1.3: ASK what is cheapest? plane or train to Malaga?
Chatbot: [...] In general, taking a plane can be cheaper
         than taking a train, especially if you book your
         flights in advance and take advantage of discounts
         and promotions. [...].
   P1.3: There we go, the plane is cheapest.
```

In the subsequent interview, one of the group members reflected on this conversation:

> *"We used it to settle a disagreement. What is cheaper: planes or trains? So, we kind of use it as a judge, I guess."* (P1.1)

Another participant used the chatbot to mediate a disagreement among the two other group members on the destination. By asking the chatbot for a list of advantages and disadvantages for each location and incorporating the group members' preferred activities, the group better understood the options and settled the disagreement.

This role as authority waned when participants were convinced that the chatbot provided incorrect information, reducing their trust in its subsequent answers and urging them to fact-check the chatbot's answers:

> *"Yeah, I mean, it was very interesting because it said that there were no trains to Greece, which I certainly know there are. I became a bit more like, I definitely would have to go online and fact check afterwards what it was saying because I would not just trust price range and such because it might also come with false information there."* (P9.1)

Multiple groups ascribed parental roles to the chatbot, calling it the mother of the group or comparing prompting the chatbot to going to their parents for advice.

> *"The chatbot was like a middleman in trying to figure out where to go, because we could have spent hours, the three of us just willing where to go, let alone getting forward from where we're going. So that helps. I think it helped us mediate between each other so we could just come to a quicker consensus and get to the fun part."* (P6.1)

For some groups, the role of the chatbot as an objective mediator was such a valuable asset that they even wished for an alternative, proactive design where the chatbot chimes in to suggest a compromise between members' preferences and help settle conflicts:

> *"Rather than just waiting for us to request info that if it sees that this is what we keep referring back to because we spent a lot of time in deciding where we would like to go, then maybe if it could jump in being like, 'I can see you're deciding between these two. If neither of them appeals, here's a third option.' Or, 'Here is a quick summary since we already had the pros and cons.'"* (P6.1)

4.2.3 Seeking Agreement Before Prompting. Several participants mentioned wanting to reach an agreement on what to ask before prompting the chatbot to not forgo the opinions of others. The mere option of asking a chatbot, and not the answers, therefore invoked discussions between participants and facilitated knowledge exchange. The participants hesitated to ask the chatbot too quickly as it would be a definitive answer shared amongst all group members and added to the collective understanding of the situation. In multiple groups, this resulted in the group members asking "Should we ask the chatbot X?" and waiting for group confirmation before each prompt.

> *"I think since everybody can see the response, there's this element of 'I've decided this for all of us now' because then you've probably already asked it, and now everybody can see this is the answer."* (P3.2)

An example of this can be seen in the transcripts from another group;

```
P4.3: Okay, so we have to find a place to go to, how
      to get there, something to do during the day and
      something to do at night.
P4.3: Let's start with where.
P4.3: We could maybe ask where some of the cheapest
      places to travel to are.
P4.2: Yeah. I'll ask it.
P4.2: ASK What are some cheap places to travel to?
```

Similarly, another group described prompting the chatbot as very definitive compared to a verbal discussion, requiring deliberation before committing to it:

> *"When you speak, you can say something without it being caught. But when you write, the bot or they [other participants] will definitely see it and remember it. Words become less powerful when we verbally speak because they get covered up by many different sentences, but when we ask the chatbot, that's the sentence. It is quite definitive"* (P16.2)

Public prompting implicitly involved everybody in consensus-seeking behaviour, as the visibility of the prompts might have made them look inconsiderate when starting to prompt directly. Participants hereby aligned their social group awareness about the process before committing to it. Groups where one person took the lead in prompting the bot considered less diverse opinions and seemed to be affected by effects akin to 'groupthink' [17] (i.e., reaching a consensus before exploring potential other options). This pattern seemed to be broken when the other participants started parallel conversations with the chatbot. However, participants expressed hesitance towards starting parallel conversations as it would risk overflooding the groupchat with requests (see Section 4.1.2).

5 Discussion

In this study, we explored how a chatbot's presence can align knowledge and goals and impact cognitive and social group awareness in collaborative tasks. Table 2 provides an overview of the takeaways of our thematic analysis, divided into themes related to cognitive and social group awareness. We link the themes to group behaviour and to the LLM characteristics that affected the observed behaviour. We outline implications for aligning group awareness using chatbots.

5.1 Mediating Consensus

Our results highlight multiple factors affected by the presence of the chatbot in terms of participants' group awareness. These are the mediating factors of the chatbot's contextual suggestions and increased transparency of group members' knowledge and goals by prompting the chatbot unambiguously in a shared chat context.

The main mediating factor in the group's contextual knowledge was the chatbot's ability to formulate answers based on a contextual understanding. Participants valued the chatbot's ability to reconcile opposing viewpoints and used these contextual suggestions to negotiate their individual goals or find inspiration when they were stuck. Suh et al. found a similar dynamic in collaborative music composition [65]. Here, their AI system was able to mediate the social dynamics between collaborators by providing suggestions for a starting point or by filling in gaps in the composition.

Table 2: Summary of findings for cognitive and social group awareness in relation to group behaviour and the LLM configuration.

	Theme	Key observations in group behaviour	Influencing factors of LLM-based chatbot
Cognitive Group Awareness	Expecting transparency in individual information-seeking processes.	Seeing others' prompts brings transparency to individual information-seeking processes.	Prompts of all group members are visible in the group chat; The chatbot needs specific and precise prompts.
		Chatbot may reduce asymmetric relationships and avoid selective sharing of individual search results.	Group members can only prompt from within the group chat.
	Getting lost in the mix of group chat and prompting.	Prompting selectively.	Group members can only prompt from within the group chat.
		Losing track of the conversation among group members.	Long answers within the group chat; answers persist in the conversation history.
	Activating the chatbot's capabilities.	Feeling surprised about the chatbot's use of the group's conversation history.	Implicit addition of all conversation history in the chatbot context.
		Identifying chatbot answers that need fact-checking.	No access to real-time information.
Social Group Awareness	Aligning individual and common goals.	Users learn about each other's deliberations and motivations from the prompts.	Prompts of all group members are visible in the group chat; The chatbot needs specific and precise prompts.
		Prompting as a process is used to mediate between goals and supports goal awareness.	The prompts need to be complete and contain information from all parties to be effective.
	Mediating consensus.	Users perceive the chatbot as an authoritative figure with the knowledge to settle conflicts.	Correcting or improving group members' statements while remaining impartial.
	Seeking agreement before prompting.	Perceiving prompts as a definitive action that affected the group's knowledge.	Long answers within the group chat; answers persist in the conversation history.
		Exchanging knowledge about goals and preferences to agree on how to prompt.	Prompts need to be complete and contain input from all parties to be effective.

The process leading up to the chatbot's generation of contextual prompts also contributed to shared participant understanding in several ways. Participants expressed that having the entire conversation visible in one chat context improved their cognitive group awareness as compared to dispersed communication. In online travel planning, individuals in the group typically browse and evaluate options outside of the shared context, after which they bring selected suggestions to a group meeting. Prior work shows that users describe this process as collaborative [50]. However, this curation of information by individual group members is likely to (inadvertently) skew the results in favour of an individual's preferences and excludes other group members from observing the selection process and the discarded options. Planning the trip in a shared chat enables group members to follow all intermediate steps each group member takes in their search and decision process. Increased cognitive group awareness through centralised communication has precedent in HCI and CSCW research, for example, in collaborative search environments [6, 51] and in collaborative assessment tasks [69].

Knowledge sharing is made even more explicit by the (preparation of) chatbot prompts. Participants expressed wanting to reach a consensus before prompting the chatbot, as prompting the chatbot was perceived as a definitive action. Since all prompts are kept in the chatbot's memory, they become part of a shared knowledge base. Participants typically did not want to forego the opinions of their group members by initiating a prompt directly. The discussions that took place prior to formulating the prompts informed group members about each other's goals and thus increased social group awareness. In some groups, one individual prompted the bot without discussing it first. As participants generally treated the chatbot as an authoritative figure (see Section 5.2), chatbot answers played an influential role in the group discussion and steered the conversation. As LLM prompts need to be unambiguous to give the desired results, the prompt formulation contributed to knowledge awareness within the group. Participants took this overview of each other's preferences into account in subsequent decisions, positively impacting goal negotiation. Faez et al. found that dyads in creative collaboration improved their creativity more by explaining their ideas to each other than by chatting about their work [2]. Chatbot prompting similarly forces participants to explain the reasoning behind their ideas.

5.2 Chatbot as Authority

Participants generally found the chatbot useful for the travel planning task and were often positively surprised by its suggestions. Skjuve et al. categorised reactions to LLM suggestions using a pragmatic-hedonic framework [63]. The reactions in our study can similarly be split into primarily pragmatic (e.g., providing useful information and offloading tasks to it) and some hedonic (e.g., being surprised by the chatbot's contextual understanding) reactions.

The chatbot could mediate the collaborative task as participants generally accepted the validity of its answers and attributed authority to it. The positive stance towards the mediating function of the chatbot is in line with earlier work evaluating AI roles. Kim et al. defined AI roles based on participant ratings and found that laypeople appreciated a mediating role most [43]. They define this role as having high AI autonomy and human involvement and argue that participants appreciate the shared control between AI systems and humans. We found a similar pattern in our interviews, seeing appreciation drop when the chatbot got too much influence, for example, when the chat got overwhelmed by prompt answers that steered the discussion.

This authoritative role waned when participants noticed the chatbot communicating false information or denials. As misinformation may convincingly be presented as fact by LLMs [9], users need to take this into account when evaluating chatbot output. The lack of awareness of the LLM's knowledge may have caused participants to overlook mistakes and have an overly positive view of the chatbot. The convincing presentation of misinformation by the chatbot made participants doubt their own judgement. After recognising clear misinformation, participants expressed a need to fact-check all subsequent suggestions. As LLMs are still a novel technology, users' limited awareness of LLM-based chatbots may have resulted in an overestimation of their abilities. An overestimation of the interpretation capabilities of AI team members has been shown to negatively affect collaboration results [4].

5.3 Implications for LLM Design and Research

Our results inform the design of chatbot interactions in collaborative tasks to support group members' cognitive and social awareness. We outline three design recommendations to improve group awareness.

Maintain a shared context between human and chatbot responses, but visually separate chatbot answers: Presenting the conversation and prompting in the same context guided participants to communicate their goals between each other prior to prompting, as well as reflect their goals in the prompts themselves. However, lengthy chatbot replies can overcrowd the chat, overshadowing the conversation and exerting a strong influence on the chat's direction. Methods to forcefully limit the length of the chatbot's response can be undesirably in the context of collaborative tasks, as participants deemed longer overviews useful. A more suitable solution would, therefore, be to visually separate the user chat and chatbot prompts into distinct panes but within the same view to maintain a shared context. Rather than presenting the chatbot output in a chronological, linear text format, information could be organised spatially, as suggested by Suh et al. [66]. Especially for tasks with multiple subtasks (such as the travel planning in this study), Suh et al. found that structuring information hierarchically helps users with exploration.

Introduce chatbot capabilities: The main cause for confusion among participants was a lack of awareness about the chatbot's skills and knowledge. This could be addressed by letting the chatbot introduce itself and provide examples of its capabilities to manage user expectations. Jain et al. found that informing new users on a chatbot's capabilities indeed helps to manage expectations [37]. However, Yeh et al. explored different ways of guiding users and concluded that instructions presented during onboarding are either ignored or directly mimicked without considering other uses of the options [72]. The authors suggest informing users at the start of each task to couple task descriptions and chatbot instructions. Since LLMs can, to a limited degree, comprehend the current state of a conversation, future work should assess the feasibility of providing instructions on a chatbot's capabilities based on the conversational context.

Proactive cues to increase group awareness: Correctly timed cues could help users better understand the LLM's skills and knowledge. Participants asked for features which the chatbot was already capable of, such as giving an overview of the decisions made. Other groups did request an overview and said it gave them a better understanding of the itinerary and requested alternatives for the parts they did not like. Making these features more approachable through proactive cues in the conversation could increase users' cognitive awareness of the chatbot's capabilities. However, proactive prompts can also have negative consequences. Avula et al. found that proactive suggestions can be disruptive, require unanticipated effort, and can cause a perceived loss of privacy and agency [7]. Our participants similarly raised privacy concerns when discussing proactive cues – despite the fact that the chatbot was already 'listening in' on their conversation. Apart from instructing users on the system's knowledge of chat history, a possible solution would be to add a private mode, as proposed by Amershi et al. [3]. Timing the presentation of the proactive cues is critical to prevent disruption [6]. Interventions are less disruptive when users are transitioning to the next task [36], and contextual understanding of the group's efforts is, therefore, key. For example, following a major group decision would be an opportune moment to present the current itinerary.

5.4 Future Work and Limitations

We limited the chat session to 15 minutes as a constraint to encourage participants to stay focused on the task goals. However, this may have also limited opportunities to observe more instances of team members exhibiting or leveraging group awareness. Human team performance is improved by synchronising team members' mental models [48]. In the 15-minute chat session, participants may have had insufficient time to align mental models. We tried to minimise the overhead time due to introductions by recruiting groups of participants who knew each other. Travel planning generally happens over multiple sessions and at least partly asynchronously, which is not reflected in the experiment. Future work should explore the integration of chatbots in long-term collaborative tasks, which often are of asynchronous nature, to evaluate the chatbot's impact on group awareness in a more common approach to distributed collaboration.

Moreover, it should be noted that our participant sample was considerably large for a qualitative observational study, but it consisted solely of university students, limiting the external validity of the results in terms of age, technological expertise, and other factors. Despite this, participants' self-reported technical affinity was only moderate, as well as their self-reported experience with LLMs and chatbots.

Further, we evaluated only one collaborative task – travel planning. Although proxy tasks may not translate to actual decision-making tasks [15], we argue that travel planning using online chat groups offers a realistic scenario as 75% of the participants expressed that they plan their holidays at least partly online. However, participants mentioned that it would be harder to reach a consensus in real-world decision-making as they would push back harder on a suggestion if they had personal requirements as compared to the ones set by their persona.

6 Conclusion

Aligning group members' awareness of each other's knowledge and skills is crucial for computer-mediated collaboration. We explored the impact of an LLM-based chatbot on group awareness in a collaborative group task. Our findings indicate that LLM-based chatbots can help foster group awareness in multiple ways. Group members can learn about each other's deliberations by formulating unambiguous prompts in a shared context, which the chatbot summarises in its subsequent answers, hereby increasing cognitive group awareness. The chatbot suggestions contribute to goal negotiation by offering a compromise to conflicting goals. Participants perceived the chatbot as an authoritative figure, accepting its opinion to settle conflicts. However, this authoritative perception waned when the chatbot presented incorrect information or was unable to fulfil requests. This highlights the importance of ensuring group awareness; not only between group members but also between the group members and the chatbot. As our results indicate that LLM-based chatbots can be useful in mediating conflicts and providing information, we invite the HCI community to investigate new ways of bridging the gaps in cognitive and social awareness in AI-assisted collaborative tasks.

Acknowledgments

This work is supported by the Carlsberg Foundation, grant CF21-0159.

References

[1] Mark S. Ackerman and Brian Starr. 1995. Social Activity Indicators: Interface Components for CSCW Systems. In *Proceedings of the 8th Annual ACM Symposium on User Interface and Software Technology* (Pittsburgh, Pennsylvania, USA) *(UIST '95)*. Association for Computing Machinery, New York, NY, USA, 159–168. https://doi.org/10.1145/215585.215969

[2] Faez Ahmed, Nischal Reddy Chandra, Mark Fuge, and Steven Dow. 2019. Structuring Online Dyads: Explanations Improve Creativity, Chats Lead to Convergence. In *Proceedings of the 2019 Conference on Creativity and Cognition* (San Diego, CA, USA) *(C&C '19)*. Association for Computing Machinery, New York, NY, USA, 306–318. https://doi.org/10.1145/3325480.3325486

[3] Saleema Amershi, Dan Weld, Mihaela Vorvoreanu, Adam Fourney, Besmira Nushi, Penny Collisson, Jina Suh, Shamsi Iqbal, Paul N. Bennett, Kori Inkpen, Jaime Teevan, Ruth Kikin-Gil, and Eric Horvitz. 2019. Guidelines for Human-AI Interaction. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3290605.3300233

[4] Zahra Ashktorab, Casey Dugan, James Johnson, Qian Pan, Wei Zhang, Sadhana Kumaravel, and Murray Campbell. 2021. Effects of Communication Directionality and AI Agent Differences in Human-AI Interaction. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems (CHI '21)*. Association for Computing Machinery, New York, NY, USA, Article 238, 15 pages. https://doi.org/10.1145/3411764.3445256

[5] Sandeep Avula, Jaime Arguello, Robert Capra, Jordan Dodson, Yuhui Huang, and Filip Radlinski. 2019. Embedding Search into a Conversational Platform to Support Collaborative Search. In *Proceedings of the 2019 Conference on Human Information Interaction and Retrieval* (Glasgow, Scotland UK) *(CHIIR '19)*. Association for Computing Machinery, New York, NY, USA, 15–23. https://doi.org/10.1145/3295750.3298928

[6] Sandeep Avula, Gordon Chadwick, Jaime Arguello, and Robert Capra. 2018. SearchBots: User Engagement with ChatBots during Collaborative Search. In *Proceedings of the 2018 Conference on Human Information Interaction & Retrieval* (New Brunswick, NJ, USA) *(CHIIR '18)*. Association for Computing Machinery, New York, NY, USA, 52–61. https://doi.org/10.1145/3176349.3176380

[7] Sandeep Avula, Bogeum Choi, and Jaime Arguello. 2022. The Effects of System Initiative during Conversational Collaborative Search. *Proc. ACM Hum.-Comput. Interact.* 6, CSCW1, Article 66 (2022), 30 pages. https://doi.org/10.1145/3512913

[8] Gagan Bansal, Besmira Nushi, Ece Kamar, Daniel S. Weld, Walter S. Lasecki, and Eric Horvitz. 2019. Updates in Human-AI Teams: Understanding and Addressing the Performance/Compatibility Tradeoff. *Proceedings of the AAAI Conference on Artificial Intelligence* 33, 01 (2019), 2429–2437. https://doi.org/10.1609/aaai.v33i01.33012429

[9] Emily M. Bender, Timnit Gebru, Angelina McMillan-Major, and Shmargaret Shmitchell. 2021. On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?. In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency* (Virtual Event, Canada) *(FAccT '21)*. Association for Computing Machinery, New York, NY, USA, 610–623. https://doi.org/10.1145/3442188.3445922

[10] Daniel Bodemer and Jessica Dehler. 2011. Group awareness in CSCL environments. *Computers in Human Behavior* 27, 3 (2011), 1043–1045. https://doi.org/10.1016/j.chb.2010.07.014

[11] Daniel Bodemer, Jeroen Janssen, and Laura Schnaubert. 2018. Group Awareness Tools for Computer-Supported Collaborative Learning. In *International Handbook of the Learning Sciences*, Frank Fischer, Cindy E. Hmelo-Silver, Susan R. Goldman, and Peter Reimann (Eds.). Routledge, 351–358. https://doi.org/10.4324/9781315617572

[12] Marcel Borowski, Bjarke V. Fog, Carla F. Griggio, James R. Eagan, and Clemens N. Klokmose. 2023. Between Principle and Pragmatism: Reflections on Prototyping Computational Media with Webstrates. *ACM Trans. Comput.-Hum. Interact.* 30, 4, Article 61 (2023), 53 pages. https://doi.org/10.1145/3569895

[13] Virginia Braun and Victoria Clarke. 2006. Using thematic analysis in psychology. *Qualitative Research in Psychology* 3, 2 (2006), 77–101. https://doi.org/10.1191/1478088706qp063oa

[14] A. J. Bernheim Brush, David Bargeron, Jonathan Grudin, and Anoop Gupta. 2002. Notification for Shared Annotation of Digital Documents. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Minneapolis, Minnesota, USA) *(CHI '02)*. Association for Computing Machinery, New York, NY, USA, 89–96. https://doi.org/10.1145/503376.503393

[15] Zana Buçinca, Phoebe Lin, Krzysztof Z. Gajos, and Elena L. Glassman. 2020. Proxy Tasks and Subjective Measures Can Be Misleading in Evaluating Explainable AI Systems. In *Proceedings of the 25th International Conference on Intelligent User Interfaces* (Cagliari, Italy) *(IUI '20)*. Association for Computing Machinery, New York, NY, USA, 454–464. https://doi.org/10.1145/3377325.3377498

[16] Janis A. Cannon-Bowers, Eduardo Salas, and Sharolyn Converse. 1993. Shared Mental Models in Expert Team Decision Making. In *Individual and Group Decision Making: Current Issues*. Lawrence Erlbaum Associates, Inc, Hillsdale, NJ, US, 221–246. PsycInfo Database Record (c) 2020 APA, all rights reserved.

[17] Barry E Collins and Harold Steere Guetzkow. 1964. *A social psychology of group processes for decision-making*. New York: Wiley.

[18] Mustafa Demir, Nathan J. McNeese, and Nancy J. Cooke. 2020. Understanding human-robot teams in light of all-human teams: Aspects of team interaction and shared cognition. *International Journal of Human-Computer Studies* 140 (2020), 102436. https://doi.org/10.1016/j.ijhcs.2020.102436

[19] Hyo Jin Do, Ha-Kyung Kong, Pooja Tetali, Jaewook Lee, and Brian P. Bailey. 2023. To Err is AI: Imperfect Interventions and Repair in a Conversational Agent Facilitating Group Chat Discussions. *Proc. ACM Hum.-Comput. Interact.* 7, CSCW1, Article 99 (2023), 23 pages. https://doi.org/10.1145/3579532

[20] Paul Dourish and Victoria Bellotti. 1992. Awareness and Coordination in Shared Workspaces. In *Proceedings of the 1992 ACM Conference on Computer-Supported Cooperative Work* (Toronto, Ontario, Canada) *(CSCW '92)*. Association for Computing Machinery, New York, NY, USA, 107–114. https://doi.org/10.1145/143457.143468

[21] Paul Dourish and Sara Bly. 1992. Portholes: Supporting Awareness in a Distributed Work Group. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Monterey, California, USA) *(CHI '92)*. Association for Computing Machinery, New York, NY, USA, 541–547. https://doi.org/10.1145/142750.142982

[22] Tanja Engelmann, Michail D. Kozlov, Richard Kolodziej, and Roy B. Clariana. 2014. Fostering group norm development and orientation while creating awareness contents for improving net-based collaborative problem solving. *Computers in Human Behavior* 37 (2014), 298–306. https://doi.org/10.1016/j.chb.2014.04.052

[23] Melanie Erkens, Daniel Bodemer, and H. Ulrich Hoppe. 2016. Improving collaborative learning in the classroom: Text mining based grouping and representing. *International Journal of Computer-Supported Collaborative Learning* 11, 4 (2016), 387–415. https://doi.org/10.1007/s11412-016-9243-5

[24] Jannatul Fardous, Jia Tina Du, Preben Hansen, Kim-Kwang Raymond Choo, and Songshan (Sam) Huang. 2021. Group trip planning and information seeking behaviours by mobile social media users: A study of tourists in Australia, Bangladesh and China. *Journal of Information Science* 47, 3 (2021), 323–339. https://doi.org/10.1177/0165551519890515

[25] Christopher Flathmann, Beau Schelble, Brock Tubre, Nathan McNeese, and Paige Rodeghero. 2020. Invoking Principles of Groupware to Develop and Evaluate Present and Future Human-Agent Teams. In *Proceedings of the 8th International Conference on Human-Agent Interaction* (Virtual Event, USA) *(HAI '20)*. Association for Computing Machinery, New York, NY, USA, 15–24. https://doi.org/10.1145/3406499.3415072

[26] Thomas Franke, Christiane Attig, and Daniel Wessel. 2019. A Personal Resource for Technology Interaction: Development and Validation of the Affinity for Technology Interaction (ATI) Scale. *International Journal of Human–Computer Interaction* 35, 6 (2019), 456–467. https://doi.org/10.1080/10447318.2018.1456150

[27] Jos Fransen, Paul A. Kirschner, and Gijsbert Erkens. 2011. Mediating team effectiveness in the context of collaborative learning: The importance of team and task awareness. *Computers in Human Behavior* 27, 3 (2011), 1103–1113. https://doi.org/10.1016/j.chb.2010.05.017

[28] Hajar Ghadirian, Ahmad fauzi Mohd Ayub, Abu Silong, Kamariah Abu Bakar, and Maryam Hassanzadeh. 2016. Group Awareness in Computer-Supported Collaborative Learning Environments. *International Education Studies* 9 (2016), 2016. https://doi.org/10.5539/ies.v9n2p120

[29] Nitesh Goyal and Susan R. Fussell. 2016. Effects of Sensemaking Translucence on Distributed Collaborative Analysis. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing* (San Francisco, California, USA) *(CSCW '16)*. Association for Computing Machinery, New York, NY, USA, 288–302. https://doi.org/10.1145/2818048.2820071

[30] Nitesh Goyal, Gilly Leshed, and Susan R. Fussell. 2013. Leveraging Partner's Insights for Distributed Collaborative Sensemaking. In *Proceedings of the 2013 Conference on Computer Supported Cooperative Work Companion* (San Antonio, Texas, USA) *(CSCW '13)*. Association for Computing Machinery, New York, NY, USA, 15–18. https://doi.org/10.1145/2441955.2441960

[31] Jens Emil Grønbæk, Banu Saatçi, Carla F. Griggio, and Clemens Nylandsted Klokmose. 2021. MirrorBlender: Supporting Hybrid Meetings with a Malleable Video-Conferencing System. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems (CHI '21)*. Association for Computing Machinery, New York, NY, USA, Article 451, 13 pages. https://doi.org/10.1145/3411764.3445698

[32] Carl Gutwin and Saul Greenberg. 2004. The importance of awareness for team cognition in distributed collaboration. (2004). https://doi.org/10.1037/10690-009

[33] Carl Gutwin, Reagan Penner, and Kevin Schneider. 2004. Group Awareness in Distributed Software Development. In *Proceedings of the 2004 ACM Conference on Computer Supported Cooperative Work* (Chicago, Illinois, USA) *(CSCW '04)*. Association for Computing Machinery, New York, NY, USA, 72–81. https://doi.org/10.1145/1031607.1031621

[34] Amir Hossein Hajizadeh, Melanie Tory, and Rock Leung. 2013. Supporting Awareness through Collaborative Brushing and Linking of Tabular Data. *IEEE Transactions on Visualization and Computer Graphics* 19, 12 (2013), 2189–2197. https://doi.org/10.1109/TVCG.2013.197

[35] Sungsoo (Ray) Hong, Minhyang (Mia) Suh, Nathalie Henry Riche, Jooyoung Lee, Juho Kim, and Mark Zachry. 2018. Collaborative Dynamic Queries: Supporting Distributed Small Group Decision-Making. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18)*. Association for Computing Machinery, New York, NY, USA, 1–12. https://doi.org/10.1145/3173574.3173640

[36] Shamsi T. Iqbal and Brian P. Bailey. 2008. Effects of Intelligent Notification Management on Users and Their Tasks. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Florence, Italy) *(CHI '08)*. Association for Computing Machinery, New York, NY, USA, 93–102. https://doi.org/10.1145/1357054.1357070

[37] Mohit Jain, Pratyush Kumar, Ramachandra Kota, and Shwetak N. Patel. 2018. Evaluating and Informing the Design of Chatbots. In *Proceedings of the 2018 Designing Interactive Systems Conference* (Hong Kong, China) *(DIS '18)*. Association for Computing Machinery, New York, NY, USA, 895–906. https://doi.org/10.1145/3196709.3196735

[38] Jeroen Janssen and Daniel Bodemer. 2013. Coordinated Computer-Supported Collaborative Learning: Awareness and Awareness Tools. *Educational Psychologist* 48, 1 (2013), 40–55. https://doi.org/10.1080/00461520.2012.749153

[39] Jeroen Janssen, Gijsbert Erkens, and Paul A. Kirschner. 2011. Group awareness tools: It's what you do with it that matters. *Computers in Human Behavior* 27, 3 (2011), 1046–1058. https://doi.org/10.1016/j.chb.2010.06.002

[40] Steven J. Karau and Kipling D. Williams. 1993. Social loafing: A meta-analytic review and theoretical integration. *Journal of Personality and Social Psychology* 65, 4 (1993), 681–706. https://doi.org/10.1037/0022-3514.65.4.681

[41] Callie Y. Kim, Christine P. Lee, and Bilge Mutlu. 2024. Understanding Large-Language Model (LLM)-powered Human-Robot Interaction. In *Proceedings of the 2024 ACM/IEEE International Conference on Human-Robot Interaction* (Boulder, CO, USA) *(HRI '24)*. Association for Computing Machinery, New York, NY, USA, 371–380. https://doi.org/10.1145/3610977.3634966

[42] Soomin Kim, Jinsu Eun, Joseph Seering, and Joonhwan Lee. 2021. Moderator Chatbot for Deliberative Discussion: Effects of Discussion Structure and Discussant Facilitation. *Proc. ACM Hum.-Comput. Interact.* 5, CSCW1, Article 87 (2021), 26 pages. https://doi.org/10.1145/3449161

[43] Taenyun Kim, Maria D. Molina, Minjin (MJ) Rheu, Emily S. Zhan, and Wei Peng. 2023. One AI Does Not Fit All: A Cluster Analysis of the Laypeople's Perception of AI Roles. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (CHI '23)*. Association for Computing Machinery, New York, NY, USA, Article 29, 20 pages. https://doi.org/10.1145/3544548.3581340

[44] G. Klien, David Woods, Jeffrey Bradshaw, Robert Hoffman, and Paul J. Feltovich. 2004. Ten Challenges for Making Automation a "Team Player" in Joint Human-Agent Activity. *Intelligent Systems, IEEE* 19 (2004), 91–95. https://doi.org/10.1109/MIS.2004.74

[45] Michail D. Kozlov, Tanja Engelmann, Jürgen Buder, and Friedrich W. Hesse. 2015. Is knowledge best shared or given to individuals? Expanding the Content-based Knowledge Awareness paradigm. *Computers in Human Behavior* 51 (2015), 15–23. https://doi.org/10.1016/j.chb.2015.04.029

[46] Ida Larsen-Ledet and Henrik Korsgaard. 2019. Territorial functioning in collaborative writing: fragmented exchanges and common outcomes. *Computer Supported Cooperative Work (CSCW)* 28 (2019), 391–433. https://doi.org/10.1007/s10606-019-09359-8

[47] Yanyan Li, Xiaoran Li, Yuan Zhang, and Xin Li. 2021. The effects of a group awareness tool on knowledge construction in computer-supported collaborative learning. *British Journal of Educational Technology* 52, 3 (2021), 1178–1196. https://doi.org/10.1111/bjet.13066

[48] John E. Mathieu, Tonia S. Heffner, Gerald F. Goodwin, Eduardo Salas, and Janis A. Cannon-Bowers. 2000. The influence of shared mental models on team process and performance. *Journal of Applied Psychology* 85, 2 (2000), 273–283. https://doi.org/10.1037/0021-9010.85.2.273

[49] Nathan J. McNeese, Beau G. Schelble, Lorenzo Barberis Canonico, and Mustafa Demir. 2021. Who/What Is My Teammate? Team Composition Considerations in Human–AI Teaming. *IEEE Transactions on Human-Machine Systems* 51, 4 (2021), 288–299. https://doi.org/10.1109/THMS.2021.3086018

[50] Meredith Ringel Morris. 2008. A Survey of Collaborative Web Search Practices. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Florence, Italy) *(CHI '08)*. Association for Computing Machinery, New York, NY, USA, 1657–1660. https://doi.org/10.1145/1357054.1357312

[51] Meredith Ringel Morris and Eric Horvitz. 2007. SearchTogether: An Interface for Collaborative Web Search. In *Proceedings of the 20th Annual ACM Symposium on User Interface Software and Technology* (Newport, Rhode Island, USA) *(UIST '07)*. Association for Computing Machinery, New York, NY, USA, 3–12. https://doi.org/10.1145/1294211.1294215

[52] Bonnie A. Nardi, Steve Whittaker, and Erin Bradner. 2000. Interaction and Outeraction: Instant Messaging in Action. In *Proceedings of the 2000 ACM Conference on Computer Supported Cooperative Work* (Philadelphia, Pennsylvania, USA) *(CSCW '00)*. Association for Computing Machinery, New York, NY, USA, 79–88. https://doi.org/10.1145/358916.358975

[53] Clifford Nass, B.J. Fogg, and Youngme Moon. 1996. Can computers be teammates? *International Journal of Human-Computer Studies* 45, 6 (1996), 669–678. https://doi.org/10.1006/ijhc.1996.0073

[54] Clifford Nass, Jonathan Steuer, and Ellen R. Tauber. 1994. Computers Are Social Actors. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Boston, Massachusetts, USA) *(CHI '94)*. Association for Computing Machinery, New York, NY, USA, 72–78. https://doi.org/10.1145/191666.191703

[55] Sharoda A. Paul and Meredith Ringel Morris. 2009. CoSense: Enhancing Sensemaking for Collaborative Web Search. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Boston, MA, USA) *(CHI '09)*. Association for Computing Machinery, New York, NY, USA, 1771–1780. https://doi.org/10.1145/1518701.1518974

[56] Chris Phielix, Frans J. Prins, Paul A. Kirschner, Gijsbert Erkens, and Jos Jaspers. 2011. Group awareness of social and cognitive performance in a CSCL environment: Effects of a peer feedback and reflection tool. *Computers in Human Behavior* 27, 3 (2011), 1087–1102. https://doi.org/10.1016/j.chb.2010.06.024

[57] Divya Srivastava Robert W. Andrews, J. Mason Lilly and Karen M. Feigh. 2023. The role of shared mental models in human-AI teams: a theoretical review. *Theoretical Issues in Ergonomics Science* 24, 2 (2023), 129–175. https://doi.org/10.1080/1463922X.2022.2061080

[58] Matias Rojas, Miguel Nussbaum, Orlando Guerrero, Pablo Chiuminatto, Samuel Greiff, Rodrigo del Rio, and Danilo Alvares. 2022. Integrating a collaboration script and group awareness to support group regulation and emotions towards collaborative problem solving. *International Journal of Computer-Supported Collaborative Learning* 17 (2022). https://doi.org/10.1007/s11412-022-09362-0

[59] Mirweis Sangin, Gaëlle Molinari, Marc-Antoine Nüssli, and Pierre Dillenbourg. 2011. Facilitating peer knowledge modeling: Effects of a knowledge awareness tool on collaborative learning outcomes and processes. *Computers in Human Behavior* 27, 3 (2011), 1059–1067. https://doi.org/10.1016/j.chb.2010.05.032

[60] Beau G. Schelble, Christopher Flathmann, Nathan J. McNeese, Guo Freeman, and Rohit Mallick. 2022. Let's Think Together! Assessing Shared Mental Models, Performance, and Trust in Human-Agent Teams. *Proc. ACM Hum.-Comput. Interact.* 6, GROUP, Article 13 (2022), 29 pages. https://doi.org/10.1145/3492832

[61] Lenka Schnaubert and Daniel Bodemer. 2019. Providing different types of group awareness information to guide collaborative learning. *International Journal of Computer-Supported Collaborative Learning* 14 (2019), 7–51. https://doi.org/10.1007/s11412-018-9293-y

[62] Chirag Shah. 2013. Effects of awareness on coordination in collaborative information seeking. *Journal of the American Society for Information Science and Technology* 64, 6 (2013), 1122–1143. https://doi.org/10.1002/asi.22819

[63] Marita Skjuve, Asbjørn Følstad, and Petter Bae Brandtzaeg. 2023. The User Experience of ChatGPT: Findings from a Questionnaire Study of Early Users. In *Proceedings of the 5th International Conference on Conversational User Interfaces* (Eindhoven, Netherlands) *(CUI '23)*. Association for Computing Machinery, New York, NY, USA, Article 2, 10 pages. https://doi.org/10.1145/3571884.3597144

[64] M. Stefik, D. G. Bobrow, G. Foster, S. Lanning, and D. Tatar. 1987. WYSIWIS Revised: Early Experiences with Multiuser Interfaces. *ACM Trans. Inf. Syst.* 5, 2 (1987), 147–167. https://doi.org/10.1145/27636.28056

[65] Minhyang (Mia) Suh, Emily Youngblom, Michael Terry, and Carrie J Cai. 2021. AI as Social Glue: Uncovering the Roles of Deep Generative AI during Social Music Composition. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems (CHI '21)*. Association for Computing Machinery, New York, NY, USA, Article 582, 11 pages. https://doi.org/10.1145/3411764.3445219

[66] Sangho Suh, Bryan Min, Srishti Palani, and Haijun Xia. 2023. Sensecape: Enabling Multilevel Exploration and Sensemaking with Large Language Models. In *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology* (San Francisco, CA, USA) *(UIST '23)*. Association for Computing Machinery, New York, NY, USA, Article 1, 18 pages. https://doi.org/10.1145/3586183.3606756

[67] Wannapon Suraworachet, Jennifer Seon, and Mutlu Cukurova. 2024. Predicting challenge moments from students' discourse: A comparison of GPT-4 to two traditional natural language processing approaches. In *Proceedings of the 14th Learning Analytics and Knowledge Conference* (Kyoto, Japan) *(LAK '24)*. Association for Computing Machinery, New York, NY, USA, 473–485. https://doi.org/10.1145/3636555.3636905

[68] Minh Hong Tran, Gitesh K. Raikundalia, and Yun Yang. 2006. Using an experimental study to develop group awareness support for real-time distributed collaborative writing. *Information and Software Technology* 48, 11 (2006), 1006–1024. https://doi.org/10.1016/j.infsof.2005.12.009

[69] Niels van Berkel, Jorge Goncalves, Danula Hettiachchi, Senuri Wijenayake, Ryan M. Kelly, and Vassilis Kostakos. 2019. Crowdsourcing Perceptions of Fair Predictors for Machine Learning: A Recidivism Case Study. *Proc. ACM Hum.-Comput. Interact.* 3, CSCW, Article 28 (2019), 21 pages. https://doi.org/10.1145/3359130

[70] Senuri Wijenayake, Niels van Berkel, Vassilis Kostakos, and Jorge Goncalves. 2020. Quantifying the Effect of Social Presence on Online Social Conformity. *Proc. ACM Hum.-Comput. Interact.* 4, CSCW1, Article 55 (2020), 22 pages. https://doi.org/10.1145/3392863

[71] Edwin Mouda Ye, Jia Tina Du, Preben Hansen, Helen Ashman, Marianna Sigala, and Songshan (Sam) Huang. 2021. Understanding roles in collaborative information behaviour: a case of Chinese group travelling. *Information Processing & Management* 58, 4 (2021), 102581. https://doi.org/10.1016/j.ipm.2021.102581

[72] Su-Fang Yeh, Meng-Hsin Wu, Tze-Yu Chen, Yen-Chun Lin, XiJing Chang, You-Hsuan Chiang, and Yung-Ju Chang. 2022. How to Guide Task-Oriented Chatbot Users, and When: A Mixed-Methods Study of Combinations of Chatbot Guidance Types and Timings. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems (CHI '22)*. Association for Computing Machinery, New York, NY, USA, Article 488, 16 pages. https://doi.org/10.1145/3491102.3501941

[73] Chengbo Zheng, Yuheng Wu, Chuhan Shi, Shuai Ma, Jiehui Luo, and Xiaojuan Ma. 2023. Competent but Rigid: Identifying the Gap in Empowering AI to Participate Equally in Group Decision-Making. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (CHI '23)*. Association for Computing Machinery, New York, NY, USA, Article 351, 19 pages. https://doi.org/10.1145/3544548.3581131