IntroBot: Exploring the Use of Chatbot-assisted Familiarization in Online Collaborative Groups

Donghoon Shin
dhoon@uw.edu
University of Washington
Seattle, WA, USA

Soomin Kim
soominkim@snu.ac.kr
Seoul National University
Seoul, Korea

Ruoxi Shang
rxshang@uw.edu
University of Washington
Seattle, WA, USA

Joonhwan Lee
joonhwan@snu.ac.kr
Seoul National University
Seoul, Korea

Gary Hsieh
garyhs@uw.edu
University of Washington
Seattle, WA, USA

ABSTRACT

Many people gather online and form teams with strangers to collaborate on tasks. However, while intrateam trust and cohesion are critical for team performance, such characteristics take time to establish and are harder to build up through computer-mediated communication. Building on prior research that has shown that enhancing familiarity between members can help, we hypothesized that the use of a chatbot to support the familiarization of ad hoc teammates can help their collaboration. As such, we designed IntroBot, a chatbot that builds on an online discussion facilitator framework and leverages the social media data of users to assist their familiarization process. Through a between-subjects study (N=60), we found that participants who used IntroBot reported higher levels of trust, cohesion, and interaction quality, as well as generated more ideas in a collaborative brainstorming task. We discuss insights gained from our study, and present opportunities for the future of chatbot-assisted collaboration.

CCS CONCEPTS

• **Human-centered computing** → **Interactive systems and tools**; **Collaborative and social computing systems and tools**.

KEYWORDS

chatbot, familiarization, collaboration, computer-mediated communication

ACM Reference Format:
Donghoon Shin, Soomin Kim, Ruoxi Shang, Joonhwan Lee, and Gary Hsieh. 2023. IntroBot: Exploring the Use of Chatbot-assisted Familiarization in Online Collaborative Groups. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (CHI '23), April 23–28, 2023, Hamburg, Germany*. ACM, New York, NY, USA, 13 pages. https://doi.org/10.1145/3544548.3580930

CHI '23, April 23–28, 2023, Hamburg, Germany

ACM ISBN 978-1-4503-9421-5/23/04.
https://doi.org/10.1145/3544548.3580930

1 INTRODUCTION

Online ad hoc collaboration has become an integral part of our lives. This type of collaboration can exist in many forms, such as work-related short-term distributed teams within global organizations [38], or algorithmically matched multiplayer teams for non-work purposes [21]. It can vary in size, duration, and types of tasks. Such remote collaborations have also become more prevalent over the past few years due to the global pandemic, and are expected to remain an important part of the future of work [8, 65].

Despite its prevalence and importance, one of the main challenges with remote collaboration remains the need to establish strong social relationships between team members. A myriad of studies has shown that trust and cohesion between team members are critical to the functioning of the team, and that these factors mediate performance in collaborations [11, 22, 47, 62]. However, building trust and cohesion amongst strangers is challenging, and takes longer to achieve in online settings due to the lack of opportunity and diminished cues [44]. Indeed, such lack of social relationship qualities has been suggested as a hurdle for many types of remote collaborations, such as telesurgury [26], online collaborative learning [19], and online gaming [40].

Prior research suggests that trust and cohesion are developed through frequent and meaningful interactions. By discussing common interests and increasing familiarity between members, they can then perform as a team more effectively [30, 34, 36, 57, 63]. Thus, studies of interpersonal teams have often manipulated trust and cohesion by facilitating dialogues and fostering perceptions of similarity among group members [1, 39]. However, despite the importance of these team-building exercises, such strategies often require a facilitator, thus hard to scale up in ad hoc collaborative contexts for online teams [3, 25, 60].

One potential solution is to use chatbots that perform the role of the facilitator to support relationship building between new team members. As chatbots are always available and easy to scale, recent studies have proposed and explored the use of chatbots in various polyadic interaction settings, such as facilitating brainstorming [49] and supporting group decision making [49, 50]. Similarly, some chatbot prototypes have also been proposed to support relationship-building [58, 76]. However, what is missing from these studies is that they have not yet explored the efficacy of these chatbots in supporting intrateam trust and cohesion. As such, it is still unclear whether and how chatbots can be effectively designed to serve the role of improving online collaboration between ad hoc teammates.

Thus, we designed and developed IntroBot, a chatbot that facilitates a structured and efficient familiarization process for strangers collaborating in ad hoc teams. Designed to familiarize strangers in a short period of time, IntroBot supports the familiarization process of a collaborative team through three key roles outlined in the online discussion facilitator framework [82]: intellectual (e.g., topic suggestion based on user's social media), managerial (e.g., detecting & supporting the recovery of dying talk, time management), and social support (e.g., encouraging the sharing of relevant photos related to the topic). (Figure 1)

To evaluate the efficacy of IntroBot, we conducted a between-subjects, mixed-methods study (N = 60) and examined if the chatbot-driven familiarization processes can affect relationship qualities, interaction quality, and ultimately team performance in an online collaborative task. Through our study, we found that participants who used IntroBot chatted more during the familiarization process, compared to participants who familiarized themselves by chatting without IntroBot facilitation. In addition, participants who used IntroBot reported higher trust, cohesion, and interaction quality, and performed better in the collaborative task (i.e., idea-generation task) compared to participants who chatted freely and those in the baseline condition who did not get to chat prior to working on the collaborative task. Through our analyses of their qualitative responses, we found that IntroBot's facilitation strategy of utilizing social media data helped teammates identify and discuss the mutual topic of interest, and enhanced intimacy through the sharing of photos. IntroBot-facilitated familiarization was thus able to foster a more comfortable environment, leading users to be more creative and productive during their collaboration.

To summarize, our study contributes:

(1) IntroBot, a chatbot designed to support the familiarization process for online ad hoc teams by guiding and managing discussions between teammates
(2) Empirical data demonstrating the efficacy of chatbot-driven familiarization process in enhanced trust and cohesion between strangers, and improved interaction quality and task performance
(3) Insights and future opportunities for using chatbots to help strengthen social relationships in online collaborative teams

2 RELATED WORK

2.1 Social Relationships, Group Performance, and Computer-Mediated Communication

Trust [22] and cohesion [47] have long been identified as important relationship qualities for the successful collaboration of a team. Interpersonal trust is defined as the willingness to "accept vulnerability based upon positive expectations of the intentions or behavior of another" [72], and intrateam trust is the aggregation of trust the team members have in each other [52]. Trust has been found to positively influence team performance in many different ways, such as reducing uncertainty and vulnerability so they may work more efficiently and effectively [23], and helping focus the team on team goals as opposed to focusing on personal interests [45]. In a meta-analysis of 112 studies, it was found that trust among team members positively affects team performance, even when controlling various factors (e.g., trust in the leader, past team performance) and regardless of the degree of task interdependence and skill differentiation [22].

Similarly, studies of group cohesiveness have also shown it is a strong predictor of the group's collaboration process and outcome. Cohesion is defined as "an individual's sense of belonging to a particular group and his or her feelings of morale associated with membership in the group" [9]. Studies suggest that more cohesive groups can use their resources more efficiently because they know each other better, and are more motivated to complete the task [2]. In addition, higher cohesiveness among members is known to reduce the social loafing effect and yield better performance outcomes from the collaborative process [47]. The moderate effect of cohesion on performance has been consistently shown across meta-analyses [32].

Given the rise in technology-mediated remote and online collaboration, studies of trust and cohesion have been extended to the online context. Similar to the in-person contexts, these factors have also been found to be important in online settings [39, 41]. However, much research has explored how the virtuality of virtual teams (e.g., the level to which interactions occur through computer mediation) can affect the formation of trust and cohesion [66] (e.g., [13, 68]). Due to the reduced cues through computer-mediated communication, and the lack of shared physical and local contexts, many have hypothesized and found that the formation of these online social relationships will be hindered [10, 35]. Walther's Social Information Processing Theory [80] suggests the problem is with the rate of transfer, and empirical studies have shown that trust and cohesion take longer to develop over computer-mediated communication [81]. These prior studies highlight the additional need to effectively support social relationships for online collaborations.

2.2 Technology-Mediated Support of Familiarization

A number of different types of solutions have been explored to support relationship building in the online context. One set of solutions explored ways to enhance social presence in computer-mediated communication in hopes of overcoming the reduced social cues due to virtuality. Defined as a "sense of being with others" [7], social presence has been suggested as a key factor in empowering interpersonal relationships in online spaces. This set of research includes studies of Media Spaces (real-time visual and acoustic areas that span distributed spaces) that can help enhance trust across sites [37, 42, 77]. Others have also explored making computer-mediated communication richer by augmenting these channels with additional contextual cues [4, 5]. However, most of these works are intended to support established collaborators, instead of assisting new ad hoc teams. In fact, the success of these tools requires an existing level of trust; otherwise, users may not be comfortable with sharing the additional social and contextual information [59]. The use of virtual and augmented reality has also been explored to help enhance social presence and trust [64]. Social presence may be increased by immersing users in the medium and offering contextual factors. However, these solutions would require VR/AR devices and restrict the online collaborators to interact through VR/AR, thus may not offer a scalable solution for online collaborative groups.

Another set of solutions seeks to facilitate conversation to increase familiarity amongst teammates. Conversations are important to help reduce uncertainty [81], help create a shared future [39], and increase overall familiarity [78]. Familiarity specifically is known to be one of the important antecedents of trust [78]. Studies have shown that the higher the familiarity with others, the lower the social distance and the higher the cohesiveness [15, 28]. Prior experiments on trust and cohesion have often manipulated trust and cohesion through an introductory exercise or some discussion of a common topic [1]. This strategy has been applied in the online context to enhance social relationships for online group members. For example, in the online learning context, facilitators have been used to support ice-breaking activities to enhance closeness between online learners [25]. McGrath et al. also describe a number of virtual tools that could be used to facilitate this ice-breaking process to engage students [60]. However, such techniques are limited in that they require a human facilitator during the familiarization process, which may be costly, difficult to coordinate, and hard to scale up. Instead, in this study, we focus on the use of chatbots, which are always accessible, scalable to serve unlimited number of users, and easily attachable to text-based medium without having to include additional devices or human facilitators [29].

2.3 Chatbots in Polyadic Conversations

Due to their scalability and availability, chatbots have been applied to serve roles in various real-world interactions. Particularly, one specific area in which chatbots have recently shown promise is in polyadic conversational settings [84]. Defined as a context where a chatbot supports the interaction of multiple people, polyadic chatbots have been devised and investigated due to the extensive application of third party add-ons in popular collaboration tools (e.g., Slack, Teams) [6]. For example, GroupFeedBot is designed as a chat facilitator that moderates group decision making by improving efficiency and ensuring active participation [49]. In a learning discussion setting, Dyke et al. explored the role of chat agents in supporting the collaborative reasoning processes of learners toward learning objects [27]. These studies revealed the efficacy of chatbots in enhancing goal-oriented discussions (e.g., reaching a consensus [49, 50], reasoning [27]) or ensuring the engagement of members (e.g., inducing even participation of members [49, 74]).

The idea of using polyadic chatbot has also been explored for relationship-building among strangers [84]. One example is Grätzelbot that is designed for onboarding of college students [58]. Aside from answering new students' questions, it also offers scavenger hunt that if solved collaboratively resulted in more points – thus indirectly facilitating relationship building. Helper Agent [43] and BlahBlahBot [76] instead more directly mediates the conversation between strangers by suggesting discussion topics to spark conversations. However, despite the potential of polyadic chatbots in supporting familiarization of strangers, as noted in a recent review paper, their affects on social perceptions are still understudied [84]. For example, the Helper Agent [43] study found that their bot only enhanced trust for some participants but not others. Further, prior systems primarily focus on topic recommendation as the main feature for relationship building [43, 76]. According to the online discussion facilitator framework [82], such an intellectual functionality is only one of the roles of the facilitator. How might additional roles of a facilitator be built into the chatbot, and how might these roles affect intrateam trust and cohesion? To realize the potential for these chatbots to support collaborative work, additional research is needed.

3 DESIGN OF INTROBOT

To explore the use of chatbots to enhance trust and cohesion for ad hoc online collaborations, we designed and implemented a chatbot: IntroBot. In this section, we describe the overall structure and design process of IntroBot.

3.1 Guiding Framework for Design

In order to systematically support the familiarization process of ad hoc teammates, we used an online discussion facilitator framework [82] to guide the design of our IntroBot. This framework suggests that an effective facilitator performs four key roles: *intellectual*, *social*, *managerial*, and *technical*. *Intellectual* role is focused on helping participants with their learning objectives (in this case, learning about each other); *managerial* role is about making sure the discussions stay on track and go smoothly; *social* role is to create a friendly and interactive environment; and *technical* role is targeted to providing technical guidance on how to interact via the mediated platform.

Of the four roles, we decided to scope out the *technical* facilitation and instead focused our design to ensure that the chatbot is intuitive to use without the need for explicit usage guidance from the chatbot. We believe this more closely adheres to existing chatbot design guidelines [51]. Thus, we focused our design on the *intellectual*, *social*, and *managerial* features.

3.2 Design and Pilot Testing of IntroBot's Intellectual Role

The key goal of IntroBot is to facilitate familiarization. Therefore, we first focused on the chatbot's intellectual role – how to help users actually get to know each other. In interpersonal communication settings, sharing common interests is frequently used as an effective way of building familiarity, as individuals can start building understandings on their communication partner by discussing common interests and gain the sense of being close [24, 31, 33, 73, 79]. As such, in this work, our idea for supporting familiarization is to utilize people's social media posts to explore topical interests [17] and then to automatically recommend topics of mutual interest for the chat.

For our design, the system first collected every keyword from each user's Instagram posts. It then embedded [61], compared, and calculated pairwise cosine similarity of keyword sets between users, and presented the top 20 keywords from this list. Then, before initiating the conversation, users are asked to choose at least three keywords that they want to talk about from the presented list to ensure them a chance to drop unpleasant topics. Overlapped selections are then used as discussion topics by the chatbot; otherwise, topics were randomly selected from the list of the keywords that two people chose.

To check if this topic recommendation based on social media data is feasible and can support familiarization for online ad hoc teams, we conducted a pilot study. 18 participants were paired up and randomly split into one of three groups. In one group, pairs of participants were simply asked by the chatbot to chat freely without topical guidance. In another group, participants were asked to converse using one of the four predefined topics (e.g., *"What are you planning to do this winter"*). Finally, in the third group, participants were presented with the aforementioned personalized topics based on their social media posts.

We found that, even with just 3 pairs of participants per condition, our discussion topic facilitation resulted in significantly or marginally better perceived conversational quality (M = 5.83 compared to M = 4.50 and M = 4.17) and perceived closeness to the partner compared with the random topics (M = 5.56 compared to M =3.83 and M = 3.70) compared to the other conditions (out of a 7 point Likert scale, where t = 1.38, p < 0.1 and t = 1.98, p < 0.05 when comparing between personalized topics vs. random topics, and t = 1.99, p < 0.05 and t = 1.71, p = 0.06 when comparing between personalized topics vs. free chatting). Participants also provided various positive feedback on the topic recommendation of our prototype system: *"I really like the way this bot helped us get familiarized."* (P2); *"By easily identifying common interests with topic recommendations, we were able to quickly get closer and the conversation went smoothly."* (P6)

From the results, we could identify the feasibility of our proposed method for the chatbot's intellectual role. Then, we focused on designing the other key features of IntroBot, as described in the next section.

3.3 Key IntroBot Features

IntroBot is designed to begin by first introducing itself, and asking the users to start by introducing themselves. It then recommends topics based on the social media of the users and encourages and helps them to share photos from their social media posts to enhance their conversations. It further moderates the conversation by intervening when a conversation is dying. Below, we present the key features of IntroBot organized by its three major roles.

3.3.1 Intellectual.

- *Selection of the topic of common interests.* As discussed, the main feature of IntroBot is to identify and recommend the topic of users' mutual interest, using their Instagram posts. (Figure 1-1a)
- *Initiating chat.* Based on the curated topic from the users, IntroBot initiates the conversation by asking users to discuss the recommended topic. (Figure 1-1b)

3.3.2 Managerial.

- *Structuring the conversation.* IntroBot is designed to guide users through the familiarization process. This includes prompts to introduce themselves (Figure 1-2a), discussions about the topic, and telling users to wrap up their chat and say goodbye (Figure 1-2b).
- *Detecting and recovering dying talk.* It is important to ensure active conversation when users are familiarizing themselves and building rapport [55]. Therefore, IntroBot is also designed to detect when the conversation is dying. Once detected, IntroBot encourages the user who has talked less so far during the chat, and asks the user to share more about the topic they are currently discussing. (Figure 1-2c)
- *Time management.* Time management during the online discussion is considered crucial when ensuring that users are on track [56]. On such an account, we designed the system to let users know the remaining time when the 30 seconds are left for the introducing/wrapping-up step and 1 minute for the topic discussion step. (Figure 1-2d)
  In addition, the system keeps showing up the remaining time in the top-right bar, letting users see the remaining time of each phase whenever they want to check. (Figure 1-2e)

3.3.3 Social.

- *Encouragement.* IntroBot uses utterances and languages that encourage participation. For example, IntroBot provides positive feedback to users and lets them know that they are having an active and interesting chat (i.e., *"You both seem to have a very interesting chat"*), which is followed by an encouragement to the less active user to chat more. (Figure 1-3a)
- *Recommending users to share relevant photos.* To encourage more social information sharing, IntroBot also supports the sharing of photos from social media, specifically by using the photo that corresponds to the recommended topic. To minimize potential privacy concerns, we designed IntroBot to get users' permission via to show a private chatbot dialogue before sharing the photo. (Figure 1-3b)

3.4 Implementation of IntroBot

There are several options for implementing the interface of a chatbot system, such as Telegram, SMS, web-based implementation, or native app. Among these, we chose to implement IntroBot as a native iOS application, as it easily allows us to detect whether the user is typing in real-time, thus enabling our feature of managing when the conversation is dying.

For the intellectual features (i.e., topic recommendation), IntroBot uses posts from people's social media accounts to infer topical interests and recommend discussion topics, enabled by using Instagram API to collect users' post data. The Python server then calculates and suggests mutual keywords of interest through free morphemes using KoNLPy [67], each of which is embedded by Word2Vec [61]. Then, by comparing every possible keyword pair using cosine similarity, the server chooses a random keyword between two keywords from each pair and sends the recommended keywords to each user's app.

One of our main managerial features is the detection when a conversation is dying. To implement it, we designed an adaptive algorithm that takes into account both pre-collected data from the pilot study and real-time user chat logs. First, IntroBot sets the time interval of greater than 16.28 seconds between chats (95th percentile of the time interval between two consecutive chats collected from the pilot study) as an indicator that the conversation is dying. Then, during the first 2 minutes, IntroBot app client keeps updating the pre-collected time interval array by appending new time intervals from the chat log database, adjusting the threshold in real-time by

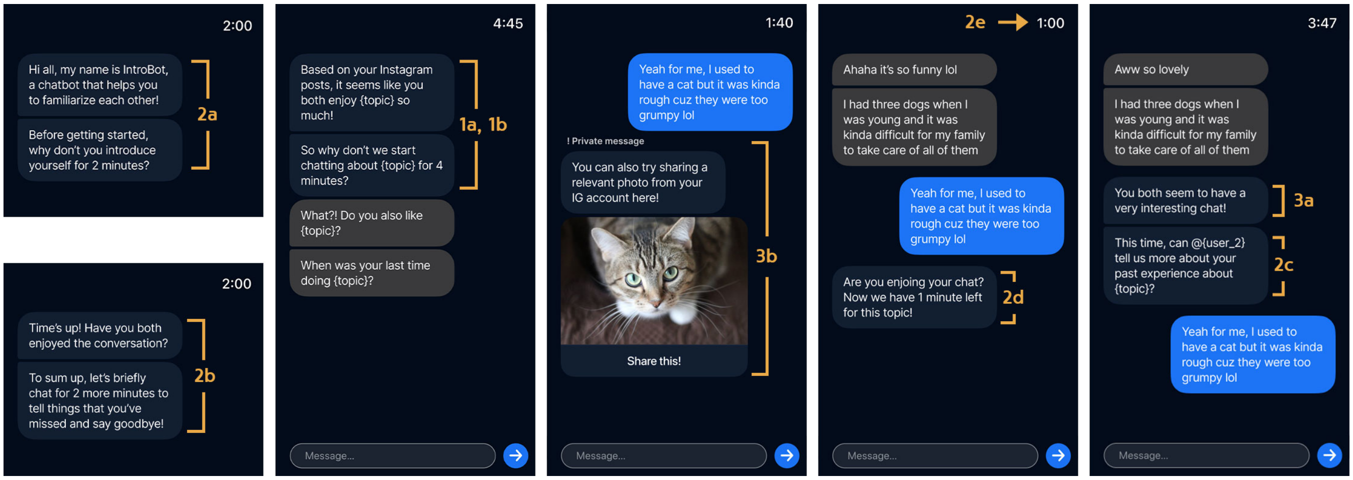

Figure 1: Key screens of our system. Through intellectual (1), managerial (2), and social (3) supports, IntroBot aims to help familiarization process of online ad hoc team

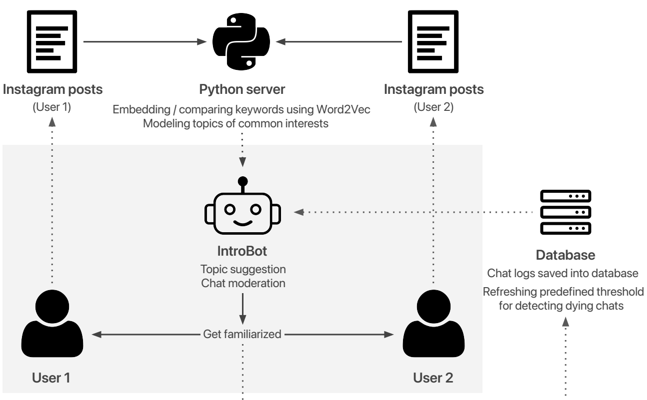

Figure 2: Overview of the system components for IntroBot's chat facilitation

calculating the new 95th percentile of the array. In other words, the threshold may be reduced/increased if the average time interval for the users has been lower/higher during their first 2 minutes of conversation.

Every chat log created during the chatting is collected in a NoSQL-based database so that user's app is able to read data and populate it in their chatting interface promptly. To ensure the privacy of users' social media data, we decided not to save posts and recommended topics of users in our database. Similarly, all the photos recommended for sharing are invalidated once recommended and shared, which is informed to users via app interface. The overview of the system components of IntroBot is illustrated in Figure 2.

4 STUDY

To understand the potential role and impact of IntroBot in facilitating online collaborative work, we conducted an online between-subjects study (N = 60). Participants were randomly paired up with a stranger-partner, and were randomly assigned into one of three conditions: **IntroBot-facilitated chatting** (C1), **free chatting** (C2), and the **no chatting** (C3) baseline.

4.1 Hypotheses

We designed IntroBot to facilitate the discussion between strangers in ad hoc online collaborative teams. Given the IntroBot's ability to find a common interest and manage conversations between strangers, we first hypothesized that, by comparing participants using IntroBot to those not using the system, IntroBot users will chat more:

> **H1**. Participants using IntroBot will chat more during the familiarization process

Prior literature suggests that discussing common interests and increasing familiarity between strangers can enhance their sense of trust and cohesion [71]. Thus, we also hypothesized that our system would enhance trust and cohesion between members of a collaborative team:

> **H2**. Participants using IntroBot will have higher trust
> **H3**. Participants using IntroBot will have higher cohesion

Improving the social relationships between teammates is known to affect their interactions during the collaborative task [20, 22]. Thus, we also hypothesized that IntroBot facilitation will result in higher perceived interaction quality during the collaborative task and can result in higher task performance:

> **H4**. Participants using IntroBot will have higher interaction quality
> **H5**. Participants using IntroBot will have higher performance

4.2 Study Setup

To test our hypotheses, we designed a between-subjects study consisting of 3 conditions. One is the intervention condition, where participants participated in **IntroBot-facilitated chatting** prior

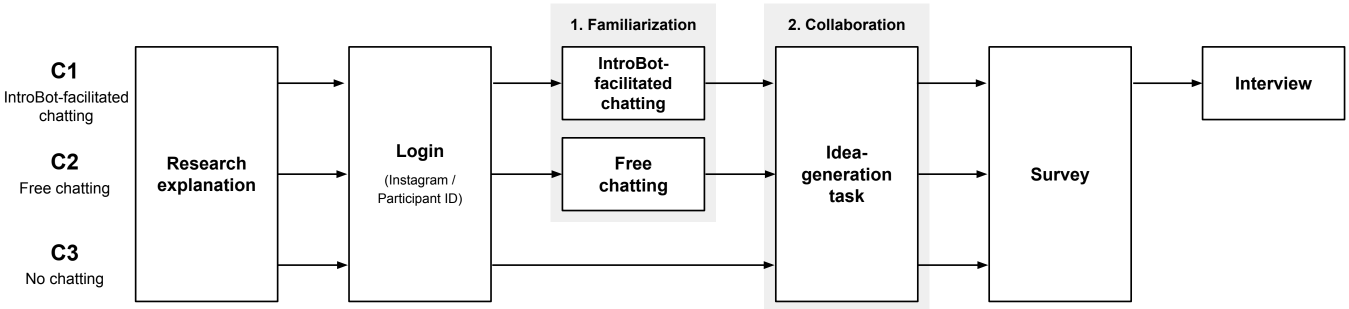

Figure 3: Protocol of our study. Participants in C1 (IntroBot-facilitated chatting) and C2 (free chatting) had time for familiarization before getting into the collaborative task, while participants in C3 (no chatting) did not have any time to familiarize themselves prior to the task

to the collaborative task (C1). The second is a **free chatting** condition (C2), representing the type of manipulation that is commonly used in prior work where participants chat without any structured facilitation [49, 50]. The third condition is **no chatting** condition (C3), the baseline condition where participants did not participate in familiarization prior to the task. This allowed us to examine the impact of IntroBot-facilitation, and to replicate the importance of team familiarization in our online collaborative task.

Similar to lots of prior studies that evaluated the group collaborative processes at a dyadic level [47, 48], we focused our study on dyadic teams (i.e., pairs) as opposed to larger teams. Participants who registered for the study were randomly assigned a partner, and the pair were randomly assigned to one of the 3 conditions. We then invited all the participants who are assigned to join at the same time to a Zoom audio call, in which the participants were told to join with an anonymous ID and their audio/video muted to prevent getting to know each other beforehand. In this call, we explained the research goal and procedure, as well as guided each participant to install an iOS application for participating in our study.

At the start of the study, participants who were assigned with IntroBot (C1) and free chatting (C2) conditions were first asked to familiarize themselves with their partner for 9 minutes - under the moderation of IntroBot (C1) and without any intervention (C2), respectively. Unlike C1 and C2, participants of C3 were not given any time to get to know their partners.

Once completed, participants were directed to the collaboration screen (Figure 4), where they collaborated on a specific task with their partner. We chose the idea generation task (i.e., *"come up with as many possible uses of a knife as possible"*) as the collaborative task. This task has been frequently used in prior works to study group work [16, 47, 83] because (i) effort could be directly related to performance and (ii) task can be presented in a meaningful way to the participants [83]. Following these studies, participants in our study were also instructed to come up with as many possible uses of knives as possible during 5 minutes. They were informed that, when they came up with an idea and want to submit it, they would submit it by clicking the `@` button and prepending `@IntroBot` tag to their idea.

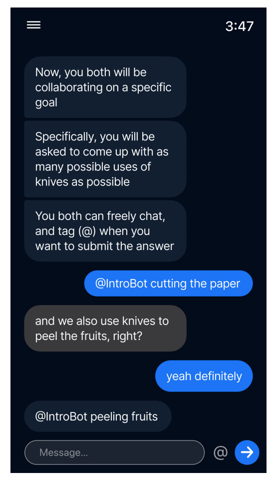

Figure 4: A chat interface where the participants took part in the collaborative task (i.e., idea-generation task)

After 5 minutes of working on the task, the app switched to a survey screen and asked participants to fill out a survey. Specifically, participants were asked to self-report survey questionnaires on trust [53], cohesion [18], and interaction quality [75] in a 7-point Likert scale to understand their perceptions of their partner and their interactions. In addition, to gain richer insights on overall experience with IntroBot, as well as potential improvements, participants in C1 (IntroBot-facilitated chatting) were asked to fill out additional open-ended questions (Section 4.4.3). The overall study procedure for each condition group is illustrated in Figure 3.

After completing the study, each participant was compensated with an approximate value of 7.5 USD (10,000 KRW) for their participation. The overall procedure of our study was conducted after obtaining IRB approval from the university human subjects division.

4.3 Recruitment & Participants

Participants were recruited from two universities' online communities. We posted a recruitment message that contains a link where participants could sign up to participate. Because our chatbot system was built as an iOS application, we screened participants based on whether they use iOS as their mobile operating system. In addition, as IntroBot requires Instagram post data for the topic recommendation, we also required participants to be Instagram user whose account contains at least 10 posts.

We started by scheduling 60 participants for the study. When participants did not show up to their scheduled sessions, we would schedule additional sessions with new participants from our pool. The 60 people who participated in our study were evenly split across our conditions (3 study conditions × 10 pairs × 2 members). Their average age is 25 (SD_{age} = 4.4), and 35 of them were female, 24 male, and 1 non-binary. All the participants were South Korean, who used Korean as their main language.

4.4 Measurement

To test our hypotheses, we collected three types of data: chat logs, self-reported measures, and interview responses.

4.4.1 Chat logs. To assess the performance of each team (H5), we counted the number of ideas generated from the chat logs saved on our server. Specifically, since our system asks participants to submit the idea in a chat interface by clicking `@` button and prepending `@IntroBot` tag, we first listed up all the submissions that contain `@IntroBot` tag from each pair's chat log. Then, we checked if each submission of `@IntroBot` tag contains an idea to check the validity of each submission. Then, we counted the total valid submissions as the performance of each pair.

In addition to the performance, we also counted the number of chat messages generated during the familiarization session for C1 and C2 to check if there is a difference in the level of chat activity between these groups. (H1)

4.4.2 Self-Reported Measures. To assess trust, cohesion, and interaction quality (H2, H3, H4), we used the dyadic trust scale [53], small group cohesion scale [18], and quality of interaction scale [75], respectively.

4.4.3 Interview. To gain richer insights on the effects and efficacy of IntroBot, we asked participants who used IntroBot (C1) to fill in questionnaires that consist of the following questions: *general experience of IntroBot-facilitated conversation* and *effects of IntroBot-facilitated conversation in their performance task.*

4.5 Analyses

For our quantitative measures, we conducted several statistical tests to assess whether there were significant differences across the groups:

For the number of messages exchanged across conditions, we ran a t-test to check if there is any significant difference between the two groups who spent time in the familiarization phase (C1, C2). For the self-reported measures of trust, cohesion, and quality of interaction, each of which averaged within each pair, we first ran a one-way ANOVA analysis to check if there exist significant differences across groups. Once we observed the differences, we ran post-hoc pairwise comparisons with Holm-Bonferroni correction (p = 0.05) to check if there was a difference between each pair of conditions.

For the performance measure (i.e., the number of ideas generated), we used a negative binomial regression to test the relationship between the primary outcome variable *number of ideas generated per pair* against the main predictor variable (*study conditions*).

To analyze interview responses, we conducted a thematic analysis [12]. During the analysis, two researchers independently coded the notable themes from the responses and discussed them three times, until mutually agreeable themes were formed. To be specific, we used the bottom-up approach with the following procedure: (i) Two authors read qualitative responses from 20 participants to get familiarized themselves with the data. (ii) Then, the authors identified meaningful responses that provide insights for user experience on our system. Through this process, the initial codes were generated. (iii) This process was repeated three times until the themes were finalized by reviewing the themes together. (iv) Finally, we surfaced four major themes from the clustered keywords and sentences. The themes and their example quotes are described in Table 1, and each participant from C1 whose qualitative responses were analyzed is noted as P1 - P20 in the result description section (Section 5.2).

5 RESULT

In this section, we describe the result of our analysis on quantitative results (chat logs and quantitative survey), followed by the results of qualitative survey analysis.

5.1 Quantitative Analysis

5.1.1 The number of chats during the familiarization process. During the study, participants in C1 and C2 had time with their partner to familiarize themselves. Analyzing the messages exchanged, we observed that IntroBot induced users to chat more, compared to the participants who chatted freely. Using IntroBot, 10 pairs of participants created 84.20 messages on average (SD = 33.17), which is significantly higher than that from participant pairs that chatted freely (M = 48.80, SD = 18.65; t = 2.94, $p < 0.01$). (H1 supported)

5.1.2 Relationship measures.

(i) Trust. Participants from C1 reported the highest level of trust (M = 5.84, SD = 0.58), followed by C2 (M = 4.96, SD = 0.86) and C3 (M = 3.73, SD = 0.62). ANOVA revealed that such a difference is significant across groups ($F(2,27)$ = 23.01, $p < 0.001$). Post-hoc analyses showed that the trust of participants from C1 is significantly higher compared to both C2 ($p < 0.05$) and C3 ($p < 0.001$), and there was also a significant difference in trust between C2 − C3 ($p < 0.01$). (H2 supported)

(ii) Cohesion. Similar to trust, participants reported the highest level of cohesion in C1 (M = 5.36, SD = 0.88), which is followed by C2 (M = 4.56, SD = 0.75) and C3 (M = 2.72, SD = 1.28). The result from ANOVA analysis indicates that there was a significant difference in cohesion across the three conditions ($F(2,27)$ = 18.40, $p < 0.001$). Through the post-hoc pairwise analysis, we found that cohesion was shown significantly different between C1 − C2 ($p < 0.05$), C2 − C3 ($p < 0.01$), and C1 − C3 ($p < 0.001$). (H3 supported)

5.1.3 Interaction quality. In terms of quality of interaction, we also found that C1 resulted in the highest interaction quality (M = 6.17,

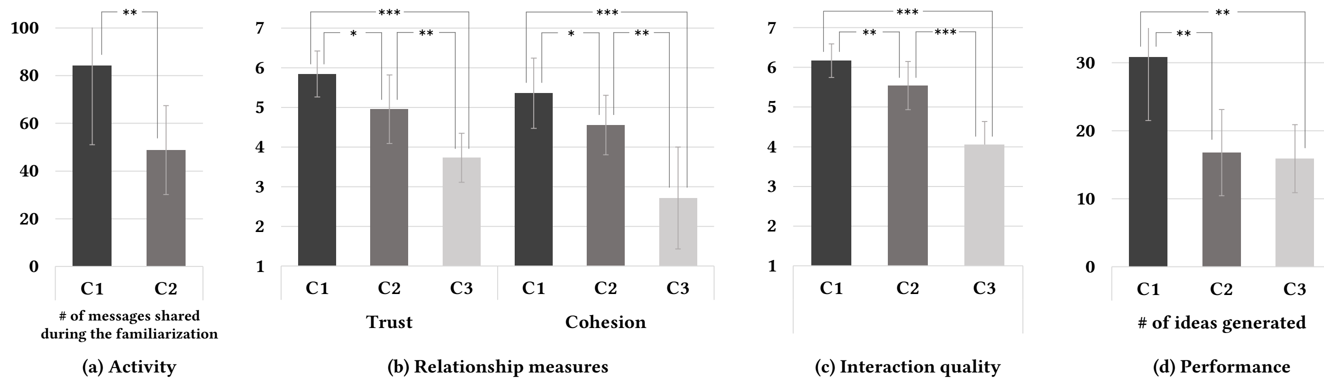

Figure 5: Average (a) Familiarization activity, (b) Relationship measures, (c) Interaction quality, and (d) Performance of pairs from each condition. The number of ideas generated and the number of messages shared during the familiarization was collected directly from the chat log, while the others were measured from the survey (7-point Likert scale). There was a group-wise difference in average number for every measure that compares three groups (b, c, and d), and stars indicate the significance of pairwise comparison (*: p < 0.05, **: p < 0.01, ***: p < 0.001)

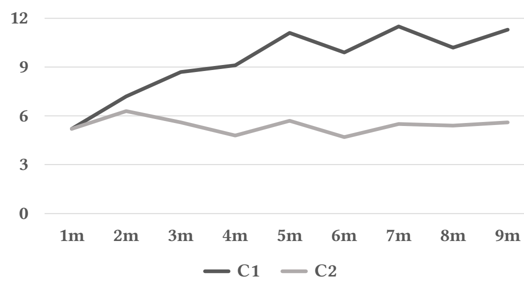

Figure 6: Average number of chats exchanged during the familiarization process over time

SD = 0.42), compared to that from C2 (M = 5.54, SD = 0.61) and C3 (M = 4.06, SD = 0.58). The difference was significant ($F(2,27)$ = 39.91, p < 0.001). Like cohesion and trust, post-hoc analyses showed that interaction quality was significantly different between C1 – C2 (p < 0.01) and C1 - C3 (p < 0.001). Similarly, we could observe a significant difference between C2 – C3 (p < 0.001). (H4 supported)

5.1.4 Performance (i.e., # of ideas generated). Participants in C1 produced the most ideas among three conditions (M = 15.40, SD = 5.70), followed by C2 (M = 8.40, SD = 3.56) and C3 (M = 7.95, SD = 2.82). Using negative binomial regressions, we found that the difference between C1 – C2 and C1 – C3 was significant (z = 4.5, p < 0.01; z = 4.8, p < 0.01), but there was no significant difference between C2 and C3 (z = 0.38, p = 0.71). The coefficients suggest that participants using IntroBot (C1) generated 1.8 times and 1.93 times more ideas compared to those in C2 and C3, respectively. (H5 supported)

5.2 Qualitative Analysis

5.2.1 Topic recommendation of IntroBot helped ice-break. We found that the topic recommendations by IntroBot (i.e., the *Intellectual* feature) helped team members get to know each other in two key ways. First, IntroBot's topic recommendation helped by reducing potential awkwardness between users. Participants reported that, if it were not for the prompt topic recommendations from IntroBot, there would have been a long awkward silence: *"I loved the chatbot's feature of automatically finding common interests based on Instagram data. I think there would have been a long silence if we had to find things to talk about."* (P3); *"It was much easier to have a conversation with someone I didn't know at all because the chatbot suggested the topic to talk about."* (P8) Second, it acted as a stepping stone for exploring further common interests between members, leading to the more active conversation: *"The beach was presented as a topic, which led to another topic of travel and it actually helped for ice-breaking."* (P18)

5.2.2 Photo sharing between members enhanced social cues and trust. In addition to topic recommendation, IntroBot brings up a relevant photo from the Instagram account and lets users share it during the conversation. Participants mentioned such a photo-sharing feature helped them to feel intimate quickly: *"Sharing the photos was so cool. The picture evoked memories of a particular city, so we could have more related conversation."* (P1); *"Conversation that could have been superficial were enriched with photos."* (P4); *"It was nice enriching the conversation with photos. Photos made us get closer to each other more quickly."* (P15)

Specifically, such intimacy was made possible by strengthening their social cues by seeing the photos. Even though participants were in an ad hoc setting where they lack prior knowledge of their partner, participants responded that photos served as a social cue during their conversation by helping them to imagine each other better: *"I didn't know anything about my partner beforehand. However, I could imagine my conversation partner in detail by seeing the picture they shared, which made me feel much closer to my partner."* (P10); *"(...) once the partner shared the photo from Jeju island (a popular tourist attraction in South Korea), it reminded me of the days when I traveled there before, and the partner seemed to be 'more lively' and closer."* (P13)

Ultimately, sharing a photo was reported to have them build mutual trust during the conversation, making their chat richer. Letting them perceive social cues stemming from the photo-sharing with their partner, participants were able to build mutual trust during their conversation: *"Seeing each other's photo and getting to know more about the other, I guess we both were able to believe in each other."* (P9) As part of our design, IntroBot asked participants whether they would like to share the recommended photo with their partner. Participants noted that such a design helped minimize privacy concerns: *"Once I saw the chatbot's private message asking if I'd like to share a photo, I thought that photo had automatically been shared (...) but soon I realized it was asking me for the consent, which made me feel that the chatbot cared my privacy."* (P11); *"As I was able to choose to share it (photo) or not, I didn't have to worry about potential privacy issues."* (P10)

5.2.3 IntroBot facilitated social interaction by managing conversation. Participants perceived that IntroBot effectively performed its *managerial* role of moderating their discussion by structuring the overall conversation process. For example, participants described the role of chatbot as a *moderator*, *facilitator*, *mediator*, *helper*, or *manager*. In particular, participants pointed out that IntroBot's feature of structuring the discussion and time management helped them to have a systematic and efficient conversation, without having to worry about wasting time: *"The chatbot has well structured the overall collaboration process. It set a topic for our conversation, so we could have a fun time. Also, we could have a systematic discussion because the chatbot did the time-check."* (P2); *"The chatbot's role was satisfactory. First of all, I liked its function of checking and managing time because we could have an efficient conversation by sharing and developing ideas within a limited time."* (P7)

Moreover, the chatbot was reported to have helped their conversation to be continuous with timely assistance through our intervention techniques, such as dying chat detection and photo recommendation for sharing. Such interventions are reported to have led the conversation flow to be seamless and natural, keeping the chat alive over time (Figure 6): *"When there was an awkward silence for a moment, the chatbot proactively induced me to chat, making it easy to continue our chat."* (P19); *"The chatbot recommended the other person's photo at the very right moment, so the conversation could go on smoothly without interruption."* (P10)

5.2.4 Enhanced understanding between members improved task experience and performance. Our participants also described how mutual understanding can positively affect trust and interaction quality, thus improving the collaboration process and output. In particular, understanding each other helped the collaborative process by creating a more comfortable environment and allowing *collaborative* relationship, rather than being *individualistic*. In addition, overcoming the limitations of anonymity by disclosing themselves with their existing social media data (e.g., topic of mutual interest, photo sharing), while at the same time with their own agency of determining how much to share, also had a positive effect on members during the collaboration.

First, a deepened familiarity with their partner created an environment in which participants could share ideas more easily and freely. Participants in our study reported that the warm environment offered by IntroBot-assisted familiarization continued, making the collaborative task less stiff and more casual: *"I can talk about strange ideas without walking on eggshells because we had a conversation and got to know each other."* (P11); *"It was much easier for me to speak random ideas because we had everyday conversations. It was nice that I didn't have to worry about 'Is this idea too ridiculous?'."* (P4) Furthermore, such a casual process of sharing opinions enabled participants to generate more creative and novel ideas: *"We could exercise our creativity in a free atmosphere."* (P6)

Second, enhanced mutual understanding between team members made them more open to/focused on each other's way of thinking. It subsequently enabled participants to perform collaborative tasks together, rather than being independent in a task-performing process, generating synergy: *"(during the collaboration,) Based on the way of my partner's generating ideas, I could think of ideas in a new way and add new ideas."* (P18) Also, some participants mentioned an enjoyable experience working on a task together: *"Rather than coming up with an idea individually, I enjoyed the task while laughing and talking with my partner."* (P16)

Lastly, increased trust led by taking away the sense of anonymity through identifying commonalities continued until the collaborative process, helping participants lubricate the collaborative process. This was reported to ultimately empower the performance of a collaborative task: *"Sharing our photos made each other more trustworthy. We were able to rely on each other during the task."* (P12); *"The image of my partner was somewhat pictured in my mind. The thought of working together with that partner made me more focused on the task and take responsibility."* (P9); *We talked about 'rabbit' (...) and I realized that the partner and I both like rabbits, which made me think that the partner was close enough and become easier to talk to the partner later during the collaboration."* (P7)

6 DISCUSSION

Overall, our results suggest that a chat facilitator can successfully play the role of supporting the team-building process in collaborative ad hoc settings. Specifically, we identified that participants who familiarized themselves using IntroBot chatted more, showed higher trust and stronger cohesion. Prior research suggested that polyadic chatbots can enhance the outcomes and engagement of people in online discussions [49, 50, 74], and that they can potentially support familiarization processes among strangers [58, 76]. Building on this line of research, our study offers empirical evidence that chatbots can support the familiarization process, enhance social relationships between strangers, and ultimately boost collaboration (i.e., interaction quality and performance) through the teammate familiarization process. Previous literature offered grounds that increased familiarity can positively affect outcomes of collaborative tasks [30, 34, 36, 46, 57, 63], yet existing online team-building supports mostly require human facilitators with limited scalability [25, 60]. From the study, we found that a chatbot can create a pleasant chatting atmosphere by easing conversations between online ad hoc teammates, increasing relationship qualities, and ultimately boosting creativity in teamwork. Participants indeed reported that the process of exchanging ideas became more casual after the IntroBot-facilitated chat. This finding is also consistent with previous research that collaboration in a pleasant and cheerful environment leads to increased creativity [14].

Though we did not explicitly test for it, one of the underlying hypotheses behind our design intervention is that enabling conversations between online ad hoc teammates can help enhance social

Table 1: Main themes from the qualitative analysis

Main theme	Sub theme	Quote example	Related facilitator role
(1) Topic recommendation	Reducing awkwardness	*"(...) I think there would have been a long silence if we had to find things to talk about."*	Intellectual
	Supporting ice-breaking	*"Icebreaking such as a common interest recommendation function helped a lot."*	
(2) Photo sharing	Increasing intimacy	*"It was nice enriching the conversation with photos. Photos made us get closer to each other more quickly."*	Social
	Enhancing social cues	*"(...) I could imagine my conversation partner in detail by seeing the picture they shared, which made me feel much closer to my partner."*	
	Building mutual trust	*"Seeing the photos made me think that my partner is more engaged."*	
(3) Conversation management	Supporting efficient conversation	*"(...) we could have a systematic discussion because the chatbot did the time-check."*	Managerial
	Enabling seamless conversation flow	*"When there was an awkward silence for a moment, the chatbot proactively induced me to chat, making it easy to continue our chat."*	
(4) Mutual understanding	Providing a comfortable environment for conversation	*"It was much easier for me to speak random ideas because we had everyday conversations."*	Managerial / Social
	Enhancing collaboration and enjoyment	*"Based on the way of my partner's generating ideas, I could think of ideas in a new way and add new ideas."*	

relationships. This is based on prior work's finding that spending time to get to know each other can help build trust and cohesion [1]. One of our findings shows that those who were provided time to free chat (C2) resulted in higher trust and cohesion than those in the baseline no chatting condition (C3). This again speaks to the importance of supporting conversation amongst team members, especially in a computer-mediated context, to strengthen social relationships.

In addition, we found that our chatbot-driven facilitation led to better social relationship outcomes than simply providing strangers a chance to chat. One of the key reasons, based on our data analysis, is the intellectual support that IntroBot provides. Our topic suggestion feature helped put people at ease talking to a stranger and led to more chatting in a limited time, as well as helping participants to skip the initial long awkward silence. In addition to topic suggestion, IntroBot's managerial support helped the familiarization by detecting and recovering dying conversations at critical times. Lastly, photo sharing further aided familiarization. Social media images provided social cues by allowing strangers to feel the presence of each other, and the richness of visual detail in photos also served as a stimulus for further conversation. As a result, such effects on relationship measures were made observable within a short time period (< 10 minutes) throughout IntroBot-facilitated chatting, which implies the future application of our system to contexts where the time for familiarization lacks.

We believe that our design can make the idea of IntroBot broadly applicable to a number of online collaborative contexts. First, IntroBot models the topics of users' mutual interests with Instagram post captions. This topic modeling approach can easily adapt to input from other social media sources, such as Twitter, Facebook, or LinkedIn feeds. We can further generalize this feature to include a broader range of social media users by combining these multiple social media data sources in order to diversify topics or overcome the lack of feed text from a single source. In addition, IntroBot is expected to be easily integrated into existing platforms (e.g., Slack, Teams, Discord) and workflows as a form of add-on feature, as well as other communication media (e.g., voice-based agents in videoconferencing) with simple text-to-speech synthesis tools. For example, an increasing number of workplaces (e.g., NHS [70], remote software developer teams [54]) are adopting chatbots mostly for altering managerial roles of facilitating team discussion, such as tracking work progress, identifying communication frictions, and reporting team issues. On top of these managerial supports, we envision that IntroBot's integration of intellectual and social supports may help boost their familiarization process to make such collaborations even more effective. Lastly, when relationships need to be built under time constraints, such as in online gaming, we believe that IntroBot would also play a role in accelerating the efficient familiarization process of users.

Still, in order for non-human chat facilitators like IntroBot to be scaled up, it would be necessary to take into account additional considerations for ensuring ethics and privacy. As we were aware of possible ethics and privacy risks regarding the IntroBot's use of social media and automated facilitators, we included precautionary steps to avoid potential issues (e.g., requiring user consent for sharing images, asking users to choose keywords to discuss from recommended topic lists). Yet, once IntroBot is scaled up and a wide range of users start to use it, we believe that further considerations must be considered. For instance, it becomes more difficult to detect and screen when users use offensive words during the chat, which may require a huge burden when IntroBot is adopted in a workflow. This suggests opportunities to augment IntroBot by integrating automated detecting tools to help moderate the conversation (e.g., pre-trained offensive language detection model [69]). In addition, although IntroBot requires users to consent prior to sharing users' photos, it is still possible that users fail to avoid sharing photos that contain their personally identifiable objects (e.g., faces). In this case, using a facial recognition system to detect and blur such personally identifiable objects would mitigate privacy issues. Lastly, even if the current version of IntroBot allows users to drop unwanted keywords by letting them choose topics out of the list, it is still possible that simply seeing improper keywords from the recommended keyword lists may make users feel embarrassed. We believe that curating and pre-defining a set of harmful stopwords and filtering them from the keyword recommendations may minimize such concerns.

7 LIMITATION & FUTURE WORK

The controlled experiment used in this study has both a strength and a weakness. Like prior research studying collaborative work often evaluated collaborative groups in controlled settings [34, 47, 48], which allowed us to test how IntroBot affects intrateam relationships and task performance, it does limit the realism and generalizability of our findings. Thus, additional studies need to be conducted to explore the use of chatbot facilitation for more complex collaborative settings, such as running a study with larger teams and examining in more complex discussion settings (e.g., collective planning task).

In addition, our participants were 25 years old on average, who use Korean as their main language, which may limit the generalizability of results. Still, as functionalities of IntroBot are designed to be adapted to diverse conditions (e.g., dying chat detection adaptive to slower keystroke & adaptability to social media sources whose main user group is older adults, using embeddings trained on different languages), we believe that the use of IntroBot can be successfully expanded to diverse age / language groups. Future studies may explore how IntroBot is used by pairs of participants with diverse backgrounds (e.g., main language, age).

8 CONCLUSION

In this paper, we designed and evaluated IntroBot, a chat facilitator that helps online ad hoc teammates get familiarized themselves with building on the online discussion framework and leveraging their existing social media data. Consisting of managerial, intellectual, and social roles, IntroBot supports the structured and efficient familiarization process prior to the collaboration between teammates. A user study with 60 participants revealed that IntroBot enhanced trust, cohesion, and interaction quality, as well as the collaborative performance of teammates in an idea-generation task.

REFERENCES

[1] Kirsten M Barbosa. 2006. Computer-mediated Communication and Group Cohesion. *Modern Psychological Studies* 12, 1 (2006), 6. https://scholar.utc.edu/mps/vol12/iss1/6

[2] Daniel J Beal, Robin R Cohen, Michael J Burke, and Christy L McLendon. 2003. Cohesion and Performance in Groups: A Meta-Analytic Clarification of Construct Relations. *Journal of Applied Psychology* 88, 6 (2003), 989. https://doi.org/10.1037/0021-9010.88.6.989

[3] S.A. Beebe and J.T. Masterson. 2015. *Communicating in Small Groups: Principles and Practices, Updated Edition.* Pearson Education.

[4] James" Bo" Begole, John C Tang, and Rosco Hill. 2003. Rhythm Modeling, Visualizations and Applications. In *Proceedings of the 16th Annual ACM Symposium on User Interface Software and Technology.* 11–20. https://doi.org/10.1145/964696.964698

[5] James" Bo" Begole, John C Tang, Randall B Smith, and Nicole Yankelovich. 2002. Work Rhythms: Analyzing Visualizations of Awareness Histories of Distributed Groups. In *Proceedings of the 2002 ACM Conference on Computer Supported Cooperative Work.* 334–343. https://doi.org/10.1145/587078.587125

[6] Ivo Benke, Michael Thomas Knierim, and Alexander Maedche. 2020. Chatbot-based Emotion Management for Distributed Teams: A Participatory Design Study. *Proceedings of the ACM on Human-Computer Interaction* 4, CSCW2 (2020), 1–30. https://doi.org/10.1145/3415189

[7] Frank Biocca, Chad Harms, and Judee K Burgoon. 2003. Toward a More Robust Theory and Measure of Social Presence: Review and Suggested Criteria. *Presence: Teleoperators & Virtual Environments* 12, 5 (2003), 456–480. https://doi.org/10.1162/105474603322761270

[8] Brodie Boland, Aaron De Smet, Rob Palter, and Aditya Sanghvi. 2020. Reimagining the office and work life after COVID-19. (2020). https://mck.co/3BUwhzg

[9] Kenneth A Bollen and Rick H Hoyle. 1990. Perceived Cohesion: A Conceptual and Empirical Examination. *Social Forces* 69, 2 (1990), 479–504. https://doi.org/10.1093/sf/69.2.479

[10] Nathan Bos, Judy Olson, Darren Gergle, Gary Olson, and Zach Wright. 2002. Effects of Four Computer-Mediated Communications Channels on Trust Development. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems.* 135–140. https://doi.org/10.1145/503376.503401

[11] Susanne Braun, Claudia Peus, Silke Weisweiler, and Dieter Frey. 2013. Transformational leadership, job satisfaction, and team performance: A multilevel mediation model of trust. *The Leadership Quarterly* 24, 1 (2013), 270–283. https://doi.org/10.1016/j.leaqua.2012.11.006

[12] 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

[13] Christina Breuer, Joachim Hüffmeier, and Guido Hertel. 2016. Does Trust Matter More in Virtual Teams? A Meta-Analysis of Trust and Team Effectiveness Considering Virtuality and Documentation as Moderators. *Journal of Applied Psychology* 101, 8 (2016), 1151. https://doi.org/10.1037/apl0000113

[14] Cheng-Ping Chang, Chih-Ting Hsu, and I-Jun Chen. 2013. The relationship between the playfulness climate in the classroom and student creativity. *Quality & Quantity* 47, 3 (2013), 1493–1510. https://doi.org/10.1007/s11135-011-9603-1

[15] John Chapin. 2001. Bridging the Gap Between Perception and Behavior: Psychological Distance in First-Person Perception. In *AEJMC Conference, Mass Communication and Society Division*, Vol. 2.

[16] Emmanuelle Charbonnier, Pascal Huguet, Markus Brauer, and Jean-Marc Monteil. 1998. Social loafing and self-beliefs: People's collective effort depends on the extent to which they distinguish themselves as better than others. *Social Behavior and Personality* 26, 4 (1998), 329–340. https://doi.org/10.2224/sbp.1998.26.4.329

[17] Christy Cheung, Zach WY Lee, and Tommy KH Chan. 2015. Self-disclosure in social networking sites: the role of perceived cost, perceived benefits and social influence. *Internet Research* (2015). https://doi.org/10.1108/IntR-09-2013-0192

[18] Wynne W Chin, Wm David Salisbury, Allison W Pearson, and Matthew J Stollak. 1999. Perceived Cohesion in Small Groups: Adapting and Testing the Perceived Cohesion Scale in a Small-Group Setting. *Small Group Research* 30, 6 (1999), 751–766. https://doi.org/10.1177/104649649903000605

[19] Hyungshin Choi and Myunghee Kang. 2010. Applying an activity system to online collaborative group work analysis. *British Journal of Educational Technology* 41, 5 (2010), 776–795. https://doi.org/10.1111/j.1467-8535.2009.00978.x

[20] Ok-Kyu Choi and Erin Cho. 2019. The Mechanism of Trust Affecting Collaboration in Virtual Teams and the Moderating Roles of the Culture of Autonomy and Task Complexity. *Computers in Human Behavior* 91 (2019), 305–315. https://doi.org/10.1016/j.chb.2018.09.032

[21] Laura Dabbish, Robert Kraut, and Jordan Patton. 2012. Communication and Commitment in an Online Game Team. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems.* 879–888. https://doi.org/10.1145/2207676.2208529

[22] Bart A De Jong, Kurt T Dirks, and Nicole Gillespie. 2016. Trust and Team Performance: A Meta-Analysis of Main Effects, Moderators, and Covariates. *Journal of Applied Psychology* 101, 8 (2016), 1134. https://doi.org/10.1037/apl0000110

[23] Bart A De Jong and Tom Elfring. 2010. How Does Trust Affect the Performance of Ongoing Teams? The Mediating Role of Reflexivity, Monitoring, and Effort. *Academy of Management Journal* 53, 3 (2010), 535–549. https://doi.org/10.5465/amj.2010.51468649

[24] Rachel E Dianiska, Jessica K Swanner, Laure Brimbal, and Christian A Meissner. 2021. Using Disclosure, Common Ground, and Verification to Build Rapport and Elicit Information. *Psychology, Public Policy, and Law* 27, 3 (2021), 341. https://doi.org/10.1037/law0000313

[25] Julie Dixon, Heather Crooks, and Karen Henry. 2006. Breaking the ice: Supporting collaboration and the development of community online. *Canadian Journal of Learning and Technology/La revue canadienne de l'apprentissage et de la technologie* 32, 2 (2006). https://doi.org/10.21432/T25S3M

[26] Pieter Duysburgh, Shirley A Elprama, and An Jacobs. 2014. Exploring the Social-technological Gap in Telesurgery: Collaboration within Distributed OR Teams. In *Proceedings of the 17th ACM Conference on Computer Supported Cooperative Work & Social Computing*. 1537–1548. https://doi.org/10.1145/2531602.2531717

[27] Gregory Dyke, Iris Howley, David Adamson, Rohit Kumar, and Carolyn Penstein Rosé. 2013. Towards Academically Productive Talk Supported by Conversational Agents. In *Productive Multivocality in the Analysis of Group Interactions*. Springer, 459–476. https://doi.org/10.1007/978-1-4614-8960-3_25

[28] Steven M Edwards, Jin Kyun Lee, and Carrie La Ferle. 2009. Does Place Matter When Shopping Online? Perceptions of Similarity and Familiarity as Indicators of Psychological Distance. *Journal of Interactive Advertising* 10, 1 (2009), 35–50. https://doi.org/10.1080/15252019.2009.10722161

[29] Mingkun Gao, Xiaotong Liu, Anbang Xu, and Rama Akkiraju. 2022. Chat-XAI: A New Chatbot to Explain Artificial Intelligence. In *Proceedings of SAI Intelligent Systems Conference*. Springer, 125–134. https://doi.org/10.1007/978-3-030-82199-9_9

[30] Paul S Goodman and Dennis P Leyden. 1991. Familiarity and Group Productivity. *Journal of Applied Psychology* 76, 4 (1991), 578. https://doi.org/10.1037/0021-9010.76.4.578

[31] Jane Goodman-Delahunty and Loene M Howes. 2016. Social persuasion to develop rapport in high-stakes interviews: Qualitative analyses of Asian-Pacific practices. *Policing and society* 26, 3 (2016), 270–290. https://doi.org/10.1080/10439463.2014.942848

[32] Lindred L Greer. 2012. Group Cohesion: Then and Now. *Small Group Research* 43, 6 (2012), 655–661. https://doi.org/10.1177/1046496412461532

[33] Dwayne D Gremler and Kevin P Gwinner. 2008. Rapport-Building Behaviors Used by Retail Employees. *Journal of Retailing* 84, 3 (2008), 308–324. https://doi.org/10.1016/j.jretai.2008.07.001

[34] Deborah H Gruenfeld, Elizabeth A Mannix, Katherine Y Williams, and Margaret A Neale. 1996. Group Composition and Decision Making: How Member Familiarity and Information Distribution Affect Process and Performance. *Organizational Behavior and Human Decision Processes* 67, 1 (1996), 1–15. https://doi.org/10.1006/obhd.1996.0061

[35] Charlotte N Gunawardena and Frank J Zittle. 1997. Social Presence as a Predictor of Satisfaction within a Computer-mediated Conferencing Environment. *American Journal of Distance Education* 11, 3 (1997), 8–26. https://doi.org/10.1080/08923649709526970

[36] David A Harrison, Susan Mohammed, Joseph E McGrath, Anna T Florey, and Scott W Vanderstoep. 2003. Time Matters in Team Performance: Effects of Member Familiarity, Entrainment, and Task Discontinuity on Speed and Quality. *Personnel Psychology* 56, 3 (2003), 633–669. https://doi.org/10.1111/j.1744-6570.2003.tb00753.x

[37] Steve Harrison. 2009. *Media Space 20+ Years of Mediated Life*. Springer Science & Business Media. https://doi.org/10.1007/978-1-84882-483-6

[38] Pamela Hinds and Cathleen McGrath. 2006. Structures that Work: Social Structure, Work Structure and Coordination Ease in Geographically Distributed Teams. In *Proceedings of the 2006 20th Anniversary Conference on Computer Supported Cooperative Work*. 343–352. https://doi.org/10.1145/1180875.1180928

[39] Judith A Holton. 2001. Building trust and collaboration in a virtual team. *Team Performance Management* (2001). https://doi.org/10.1108/13527590110395621

[40] Matthew Hudson, Paul Cairns, and A Imran Nordin. 2015. Familiarity in Team-Based Online Games: The Interplay Between Player Familiarity and the Concepts of Social Presence, Team Trust, and Performance. In *International Visual Informatics Conference*. Springer, 140–151. https://doi.org/10.1007/978-3-319-25939-0_13

[41] Carol S Iacono and Suzanne Weisband. 1997. Developing Trust in Virtual Teams. In *Proceedings of the Thirtieth Hawaii International Conference on System Sciences*, Vol. 2. IEEE, 412–420. https://doi.org/10.1109/HICSS.1997.665615

[42] Ellen Isaacs, Alan Walendowski, and Dipti Ranganthan. 2002. Hubbub: A sound-enhanced mobile instant messenger that supports awareness and opportunistic interactions. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 179–186. https://doi.org/10.1145/503376.503409

[43] Katherine Isbister, Hideyuki Nakanishi, Toru Ishida, and Cliff Nass. 2000. Helper Agent: Designing an Assistant for Human-Human Interaction in a Virtual Meeting Space. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 57–64. https://doi.org/10.1145/332040.332407

[44] Stefanie K Johnson, Kenneth Bettenhausen, and Ellie Gibbons. 2009. Realities of Working in Virtual Teams: Affective and Attitudinal Outcomes of Using Computer-Mediated Communication. *Small Group Research* 40, 6 (2009), 623–649. https://doi.org/10.1177/1046496409346448

[45] Aparna Joshi, Mila B Lazarova, and Hui Liao. 2009. Getting Everyone on Board: The Role of Inspirational Leadership in Geographically Dispersed Teams. *Organization Science* 20, 1 (2009), 240–252. https://doi.org/10.1287/orsc.1080.0383

[46] Barbara G Kanki and H Clayton Foushee. 1989. Communication as Group Process Mediator of Aircrew Performance. *Aviation, Space, and Environmental Medicine* (1989).

[47] Steven J Karau and Jason W Hart. 1998. Group Cohesiveness and Social Loafing: Effects of a Social Interaction Manipulation on Individual Motivation Within Groups. *Group Dynamics: Theory, Research, and Practice* 2, 3 (1998), 185. https://doi.org/10.1037/1089-2699.2.3.185

[48] Steven J Karau and Kipling D Williams. 1997. The Effects of Group Cohesiveness on Social Loafing and Social Compensation. *Group Dynamics: Theory, Research, and Practice* 1, 2 (1997), 156. https://doi.org/10.1037/1089-2699.1.2.156

[49] Soomin Kim, Jinsu Eun, Changhoon Oh, Bongwon Suh, and Joonhwan Lee. 2020. Bot in the Bunch: Facilitating Group Chat Discussion by Improving Efficiency and Participation with a Chatbot. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. 1–13. https://doi.org/10.1145/3313831.3376785

[50] Soomin Kim, Jinsu Eun, Joseph Seering, and Joonhwan Lee. 2021. Moderator Chatbot for Deliberative Discussion: Effects of Discussion Structure and Discussant Facilitation. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW1 (2021), 1–26. https://doi.org/10.1145/3449161

[51] Raina Langevin, Ross J Lordon, Thi Avrahami, Benjamin R Cowan, Tad Hirsch, and Gary Hsieh. 2021. Heuristic Evaluation of Conversational Agents. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–15. https://doi.org/10.1145/3411764.3445312

[52] Claus W Langfred. 2004. Too Much of a Good Thing? Negative Effects of High Trust and Individual Autonomy in Self-Managing Teams. *Academy of management journal* 47, 3 (2004), 385–399. https://doi.org/10.5465/20159588

[53] Robert E Larzelere and Ted L Huston. 1980. The Dyadic Trust Scale: Toward Understanding Interpersonal Trust in Close Relationships. *Journal of Marriage and the Family* (1980), 595–604. https://doi.org/10.2307/351903

[54] Carlene Lebeuf, Margaret-Anne Storey, and Alexey Zagalsky. 2017. How Software Developers Mitigate Collaboration Friction with Chatbots. In *CSCW 2017 Workshop on Talking with Conversational Agents in Collaborative Action*.

[55] Moon-Hwan Lee, Yea-Kyung Row, Oosung Son, Uichin Lee, Jaejeung Kim, Jungi Jeong, Seungryoul Maeng, and Tek-Jin Nam. 2018. Flower-Pop: Facilitating Casual Group Conversations With Multiple Mobile Devices. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 1, 4 (2018), 1–24. https://doi.org/10.1145/3161170

[56] Sung-Chul Lee, Jaeyoon Song, Eun-Young Ko, Seongho Park, Jihee Kim, and Juho Kim. 2020. SolutionChat: Real-time Moderator Support for Chat-based Structured Discussion. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. 1–12. https://doi.org/10.1145/3313831.3376609

[57] Glenn Littlepage, William Robison, and Kelly Reddington. 1997. Effects of Task Experience and Group Experience on Group Performance, Member Ability, and Recognition of Expertise. *Organizational Behavior and Human Decision Processes* 69, 2 (1997), 133–147. https://doi.org/10.1006/obhd.1997.2677

[58] Christian Löw, Lukas Moshuber, and Albert Rafetseder. 2021. Grätzelbot: Social Companion Technology for Community Building among University Freshmen. In *International Workshop on Chatbot Research and Design*. Springer, 114–128. https://doi.org/10.1007/978-3-030-68288-0_8

[59] Wendy E Mackay. 1999. Media Spaces: Environments for Informal Multimedia Interaction. *Computer Supported Cooperative Work* 7 (1999), 55–82.

[60] Naomi McGrath, Sue Gregory, Helen Farley, and Pauline Roberts. 2014. Tools of the trade: Breaking the ice with virtual tools in online learning. In *Proceedings of the 31st Australasian Society for Computers in Learning in Tertiary Education Conference (ASCILITE 2014)*. Macquarie University, 470–474.

[61] Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013. Efficient Estimation of Word Representations in Vector Space. *arXiv preprint arXiv:1301.3781* (2013). https://doi.org/10.48550/arXiv.1301.3781

[62] Brian Mullen and Carolyn Copper. 1994. The Relation Between Group Cohesiveness and Performance: An Integration. *Psychological Bulletin* 115, 2 (1994), 210. https://doi.org/10.1037/0033-2909.115.2.210

[63] Gary M Olson and Judith S Olson. 2000. Distance Matters. *Human–Computer Interaction* 15, 2-3 (2000), 139–178. https://doi.org/10.1207/S15327051HCI1523_4

[64] Niklas Osmers, Michael Prilla, Oliver Blunk, Gordon George Brown, Marc Janßen, and Nicolas Kahrl. 2021. The Role of Social Presence for Cooperation in Augmented Reality on Head Mounted Devices: A Literature Review. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–17. https://doi.org/10.1145/3411764.3445633

[65] Adam Ozimek. 2020. The Future of Remote Work. *Available at SSRN 3638597* (2020). https://doi.org/10.2139/ssrn.3638597

[66] Niki Panteli and Mike Chiasson. 2008. Rethinking Virtuality. In *Exploring Virtuality within and beyond Organizations*. Springer, 1–17. https://doi.org/10.1057/9780230593978_1

[67] Eunjeong L Park and Sungzoon Cho. 2014. KoNLPy: Korean natural language processing in Python. In *Annual Conference on Human and Language Technology*. Human and Language Technology, 133–136. https://koreascience.kr/article/CFKO201408355727285.page

[68] Sara Jansen Perry, Natalia M Lorinkova, Emily M Hunter, Abigail Hubbard, and J Timothy McMahon. 2016. When Does Virtuality Really "Work"? Examining the Role of Work–Family and Virtuality in Social Loafing. *Journal of Management* 42, 2 (2016), 449–479. https://doi.org/10.1177/0149206313475814

[69] Amir H Razavi, Diana Inkpen, Sasha Uritsky, and Stan Matwin. 2010. Offensive Language Detection Using Multi-level Classification. In *Canadian Conference on Artificial Intelligence*. Springer, 16–27. https://doi.org/10.1007/978-3-642-13059-5_5

[70] Timothy Revell. 2017. The NHS is using a chatbot to do tedious corporate team-building. https://www.newscientist.com/article/2147634-the-nhs-is-using-a-chatbot-to-do-tedious-corporate-team-building/

[71] Tracey L Rockett and Gerardo A Okhuysen. 2002. Familiarity in groups: Exploring the relationship between inter-member familiarity and group behavior. *Toward Phenomenology of Groups and Group Membership* (2002). https://doi.org/10.1016/S1534-0856(02)04008-2

[72] Denise M Rousseau, Sim B Sitkin, Ronald S Burt, and Colin Camerer. 1998. Not So Different After All: A Cross-Discipline View Of Trust. *Academy of Management Review* 23, 3 (1998), 393–404. https://doi.org/10.5465/amr.1998.926617

[73] Melissa B Russano, Fadia M Narchet, Steven M Kleinman, and Christian A Meissner. 2014. Structured Interviews of Experienced HUMINT Interrogators. *Applied Cognitive Psychology* 28, 6 (2014), 847–859. https://doi.org/10.1002/acp.3069

[74] Joseph Seering, Michal Luria, Connie Ye, Geoff Kaufman, and Jessica Hammer. 2020. It Takes a Village: Integrating an Adaptive Chatbot into an Online Gaming Community. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. 1–13. https://doi.org/10.1145/3313831.3376708

[75] Leslie D Setlock, Susan R Fussell, and Christine Neuwirth. 2004. Taking It Out of Context: Collaborating within and across Cultures in Face-to-Face Settings and via Instant Messaging. In *Proceedings of the 2004 ACM Conference on Computer Supported Cooperative Work*. 604–613. https://doi.org/10.1145/1031607.1031712

[76] Donghoon Shin, Sangwon Yoon, Soomin Kim, and Joonhwan Lee. 2021. BlahBlahBot: Facilitating Conversation between Strangers using a Chatbot with ML-infused Personalized Topic Suggestion. In *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–6. https://doi.org/10.1145/3411763.3451771

[77] John C Tang, Nicole Yankelovich, James Begole, Max Van Kleek, Francis Li, and Janak Bhalodia. 2001. ConNexus to Awarenex: Extending awareness to mobile users. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 221–228. https://doi.org/10.1145/365024.365105

[78] Hung Wei Tseng and Hsin-Te Yeh. 2013. Team Members' Perceptions of Online Teamwork Learning Experiences and Building Teamwork Trust: A Qualitative Study. *Computers & Education* 63 (2013), 1–9. https://doi.org/10.1016/j.compedu.2012.11.013

[79] Jonathan P Vallano, Jacqueline R Evans, Nadja Schreiber Compo, and Jenna M Kieckhaefer. 2015. Rapport-Building During Witness and Suspect Interviews: A Survey of Law Enforcement. *Applied Cognitive Psychology* 29, 3 (2015), 369–380. https://doi.org/10.1002/acp.3115

[80] Joseph B Walther. 2008. Social information processing theory. *Engaging theories in interpersonal communication: Multiple perspectives* 391 (2008).

[81] Joseph B Walther and Ulla Bunz. 2005. The Rules of Virtual Groups: Trust, Liking, and Performance in Computer-Mediated Communication. *Journal of communication* 55, 4 (2005), 828–846. https://doi.org/10.1111/j.1460-2466.2005.tb03025.x

[82] Qiyun Wang. 2008. Student-facilitators' roles in moderating online discussions. *British Journal of Educational Technology* 39, 5 (2008), 859–874. https://doi.org/10.1111/j.1467-8535.2007.00781.x

[83] Kipling D Williams and Steven J Karau. 1991. Social Loafing and Social Compensation: The Effects of Expectations of Co-Worker Performance. *Journal of Personality and Social Psychology* 61, 4 (1991), 570. https://doi.org/10.1037/0022-3514.61.4.570

[84] Qingxiao Zheng, Yiliu Tang, Yiren Liu, Weizi Liu, and Yun Huang. 2022. UX Research on Conversational Human-AI Interaction: A Literature Review of the ACM Digital Library. In *CHI Conference on Human Factors in Computing Systems*. 1–24. https://doi.org/10.1145/3491102.3501855