ComPeer: A Generative Conversational Agent for Proactive Peer Support

Tianjian Liu
liutj9@mail2.sysu.edu.cn
Sun Yat-sen University
Guangzhou, China

Hongzheng Zhao*
zhaohz23@mail2.sysu.edu.cn
Sun Yat-sen University
Zhuhai, China

Yuheng Liu*
liuyh357@mail2.sysu.edu.cn
Sun Yat-sen University
Zhuhai, China

Xingbo Wang
xiw4011@med.cornell.edu
Cornell University
New York, USA

Zhenhui Peng[†]
pengzhh29@mail2.sysu.edu.cn
Sun Yat-sen University
Zhuhai, China

ABSTRACT

Conversational Agents (CAs) acting as peer supporters have been widely studied and demonstrated beneficial for people's mental health. However, previous peer support CAs either are user-initiated or follow predefined rules to initiate the conversations, which may discourage users to engage and build relationships with the CAs for long-term benefits. In this paper, we develop *ComPeer*, a generative CA that can proactively offer adaptive peer support to users. *ComPeer* leverages large language models to detect and reflect significant events in the dialogue, enabling it to strategically plan the timing and content of proactive care. In addition, *ComPeer* incorporates peer support strategies, conversation history, and its persona into the generative messages. Our one-week between-subjects study (N=24) demonstrates *ComPeer*'s strength in providing peer support over time and boosting users' engagement compared to a baseline user-initiated CA. We report users' interaction patterns with *ComPeer* and discuss implications for designing proactive generative agents to promote people's well-being.

CCS CONCEPTS

• **Human computer interaction (HCI)**; • **Human-centered computing**; • **Natural language processing**;

KEYWORDS

Generative conversational agents, proactivity, human-AI interaction, peer support

---

*Both authors contributed equally to this research.

[†]The Corresponding author

---

1 INTRODUCTION

Peer support refers to the social support exchanged among peers who have similar experience [49] and has become an important source for promoting people's well-being [7, 22, 32, 43, 61, 63]. By sharing advice and feelings with each other, peers build and strengthen their social relationships [55], which could benefit their mental health for a long term [21, 24, 89]. However, a supportive human peer could not be always available for anyone. To mitigate this concern, researchers in Human-Computer Interaction have developed a variety of conversational agents (CAs) acting as peers and demonstrated their usefulness to offer accessible peer support [38, 53, 66, 67].

However, previous peer support CAs can fall shorts in two aspects. First, they are mostly user-initiated, *i.e.,* users need to select a topic or send a message to start each round of conversation with the agent. This user-initiated design could discourage users to converse with the CAs for a long term and form a close interpersonal relationship [17], which is key to the success of peer support [1]. In contrast, the proactive messages from CAs can keep the conversation alive and improve conversation productivity [12]. Literature on peer support also encourages peers to engage in mutual conversations rather than unilateral conversations [79, 94]. While few CAs can initiate the conversations by asking help from or providing social support to users [38, 51], the timing and content of their proactive conversations are normally fixed by pre-defined rules. The fixed rules can increase the possibility of offering help in a wrong context, which impairs users' perceptions with the CAs [58]. Second, previous peer support CAs usually do not learn about users (*e.g.,* recent mentally challenging issues and feelings) from the dialogues. This limits the CAs' ability to get familiar with users and base the messages on the learned information about users, which can help users feel being understood [18] and maintain engagement with the CAs [12, 76].

In this paper, we propose *ComPeer* (**Com**panion for **Peer** support), a generative conversational agent that can plan and proactively initiate the peer support chats based on the learned information about users from the conversation history. [1] Recent advances

---

[1]The code is located at https://github.com/liutj9/ComPeer.

in large language models (LLMs) have enabled building intelligent agents that can plan, reflect, and proactively take actions when interacting with their environment [59]. However, little is known about the design, effectiveness, and user experience of a LLM-powered CA for proactively offering peer support. Questions arise as what are the design principles of the such a CA for proactive peer support, how to enable the CA to adaptively plan the proper timing and content of proactive messages, and how would users with mental health challenges perceive and interact with the proactive CA. The target users of *ComPeer* are those who are experiencing stressful issues but have not been diagnosed with mental health diseases like depression or eating disorder. One representative user group is the university students who are found to frequently encounter mentally challenging issues [60, 70].

To begin with, we get insights from literature and work with five experienced users of role-playing chatbots and three students majored in psychology to iterate the design principles of *ComPeer*. The principles highlight the importance of the consistent persona as a peer, the proper content and timing of proactive messages, and the usage of psychological strategies in a proactive peer support CA. Based on the derived design principles and inspired by Park et al. [59], we adopt GPTs to develop the cognitive architecture of *ComPeer*. The **Memory** module stores all the dialogues between *ComPeer* and the user. The **Schedule** module plans the timings and content that *ComPeer* will proactively send to the user in the near future. There are two sources for the schedule. First, after one round of conversations (*e.g.,* no message from the user within five minutes in our case), the **Event Detector** extracts the user events (*e.g.,* emotional states, plans, challenging issues) from the user's messages of this round and infers the proper timing of each event to proactively talk about it later today. Second, at the beginning of a day (*i.e.,* 0:00), *ComPeer* invokes the **Reflection** module to ask itself what it learns about the user from the conversations yesterday, based on which it initializes the schedule on that day. The **Dialogue Generation** module incorporates peer support strategies, conversation history, and *ComPeer*'s persona in each generated message.

We conduct a one-week between-subjects study with 24 participants to evaluate the effectiveness and user experience of *ComPeer*. The baseline user-initiated CA has similar modules with *ComPeer* except the Event Detector, Schedule, and Reflection modules that are responsible for proactive peer support. The results show that *ComPeer* performs better in making users feel relieved over time in the one-week study. Participants with *ComPeer* are more satisfied with its proactive messages that care for their states and work than those self-disclosing *ComPeer*'s events. Moreover, *ComPeer*'s proactive messages promote participants' interaction, where they share more their feelings and lives with *ComPeer*. We offer more qualitative findings to support the *ComPeer* strengths above from an extended one-week study in which all the 24 participants use *ComPeer*. We report users' concerns on *ComPeer* and discuss insights from our study for developing proactive agents and proactive generative agents to improve people's mental health.

Our contributions are three-fold. First, we present a novel generative conversational agent named *ComPeer* that can learn about users from the dialogues and plan the timing and content of proactive messages to offer peer support in the near future. Second, our user study demonstrates the usefulness of *ComPeer*'s proactive messages in relieving users' stress, and maintaining user engagement. Third, we offer a set of design implications into designing and developing proactive conversational agents with large language models for healthcare scenarios.

Ethics and Researcher Disclosure. We shape the work by our experience with and observation on people who struggle in stressful issues and desire peer support. The authors have experience of exchanging social support with peers like friends, classmates, colleague, and strangers online. Two of the authors have experiences and publications that study technologies for supporting people with mental health concerns. Our research team obtains IRB approval for broader research projects on patients' and caregivers' practices of healthcare service systems and online communities. We do not store any personally identifiable information such as participants' names in the design process and evaluation study of *ComPeer*. We inform participants that they can request to delete their conversation data with *ComPeer*, and researchers are not allowed to share conversation data to others. We keep in touch with the participants every day during the study, inform that they can quit the study at any time if they feel uncomfortable about it, and suggest them to seek professional healthcare services if they face severe mental health problems.

2 RELATED WORK

2.1 Peer Support and Conversational Agents

Peer support is commonly defined as "social and emotional support, frequently coupled with instrumental support, that is mutually offered or provided by persons having a mental health condition to others sharing a similar mental health condition, to bring about a desired social or personal change" [49]. In general, peer support is beneficial for people experiencing challenging issues such as depression [64] and diabetes[6]. One key to the success of peer support is the close interpersonal relationship among the peers, which positively predicts the amount of exchanged social support and improvement of psychological health [8]. According to social penetration theory, deeper and broader self-disclosure among peers can lead to a deeper sense of intimacy [77, 78]. Nevertheless, a supportive and closely related human peer is not always available for many people suffering from mental health concerns.

Conversational agents (CAs) are promising alternatives to offering peer support to anyone in need [29, 38, 45, 54, 67, 92]. These CAs commonly incorporate effective psychological skills and strategies in their messages, to provide users with enhanced social support. Nevertheless, these previous peer support CAs are mostly user-initiated (*e.g.,* [45, 67, 92]) or follow pre-set rules to initiate the conversation with users at certain timings (*e.g.,* [38]). Such interaction mechanisms could discourage users from maintaining engagement and building a relationship with the CAs [12, 17]. For one thing, as reported in a longitudinal study in which participants interact with a user-initiated social chatbot over a 3-week period [17], participants' level of self-disclosure, the perceived quality of the interactions and the perceived empathy of the chatbot decreased after each interaction, and participants' feelings of friendship with the chatbot were low. For another, the pre-set rules could lead to

untimely and irrelevant proactive messages of CAs, making the users feel annoyed [44, 65] and disruptive [11, 76]

Our work aligns with previous CAs for peer support by adopting psychological strategies in the conversations with users. Different from previous CAs, we seek to build a CA that can proactively conduct peer support conversations that adapt to users' situations and explore whether our proactive mechanism can lead to a close human-CA peer relationship in the long term.

2.2 Design and Effect of Proactive Agents

Researchers in Human-Computer Interaction have started to design and evaluate the proactive manner of intelligent agents [4, 9, 35, 62, 98, 99]. For example, Peng et al. [62] defined the service robot's proactivity as "the anticipatory action that robots initiate to impact themselves and/or others". In a decision-making task, they found that users had sufficient opportunities to express their thoughts and feelings to a robot with medium proactivity, *e.g.,* actively confirming user need before providing help rather than actively offering help without user's confirmation or passively waiting for user's requests [62]. In the survey on human-CA interaction design, Chaves and Gerosa [12] suggested that a proactive CA shares initiative with the user and may manifest proactivity when it initiates exchanges, suggests new topics, provides additional information, or formulates follow-up questions. For instance, in the healthcare domain, Fitzpatrick et al. [25] proposed the Woebot that assigns a goal to the user and proactively prompts pre-set motivational messages and reminders at fixed timings to keep the user engaged in the treatment. Chaves and Gerosa [12] summarized several benefits of manifesting a CA's proactivity in the conversations, such as providing additional and useful information [3, 11, 50, 62], inspiring users and keeping the conversation alive [11, 76, 76], recovering the agent from a failure [65, 76], improving conversation productivity [3, 28], as well as guiding and engaging users [11, 25]. However, designing a proactive CA is also challenging, especially regarding the timing and relevance of its proactive messages, which are mostly pre-set by the designers of CAs [11, 44, 65]. For example, Liao et al. [44] investigated proactivity of a CA in a workspace environment and found that the perceived interruption of the CA's proactivity negatively affects users' perceptions of it. Besides, the CA's proactive disclosure may cause users' concerns about privacy [23] and being controlled [86, 90].

Following the design of proactivity in these previous agents, our *ComPeer* initiates the conversations with users by actively disclosing itself and querying their needs for peer support before offering help. Instead of pre-setting rules for triggering proactive behaviors, *ComPeer* can adapt the timing and content of its proactive messages based on the ongoing conversations with users. We offer quantitative and qualitative findings regarding the effect of *ComPeer*'s proactive messages on the peer support outcome and process.

2.3 Generative Agents Powered by Large Language Models

The recent advances in large language models (LLMs) have empowered researchers to build the key components of generative agents [59, 83, 93]. For example, Park et al. [59] presented an LLM-powered architecture that enables generative agents to simulate believable human behavior in an interactive sandbox environment. The architecture includes a memory module that stores the agent's perceptions of the environment, retrieves the memories related to its current situation, and passes these memories to an LLM to decide its actions [59]. It also has a reflection module that prompts the LLM to extract high-level insights from its memories to enrich the memory module, as well as a plan module that prompts the LLM to adaptively schedule the agent's activities in the near future [59]. Sumers et al. [83] proposed Cognitive Architectures for Language Agents (CoALA) that use the LLM to transform observations into text, choose actions, and manage the agent's internal state via processes such as learning and reasoning. The architecture has a decision procedure module that interacts with the LLM, the internal memory, and the external memory [83]. In each decision cycle with the environment, the agent uses retrieval and reasoning to plan its actions [83]. Inspired by these generative agents, we use LLMs to build up *ComPeer*'s architecture, including the Memory, Schedule, Event Detector, Reflection, and Dialogue Generation Modules. We contribute a new LLM-powered generative conversational agent for proactive peer support and empirical findings on how users perceive and interact with it in real-world contexts.

3 DESIGN PROCESS AND PRINCIPLES OF *COMPEER*

In this paper, we specifically target university students, a representative group that commonly encounters stressful issues [60, 70]. The language used by the *ComPeer* and the participants in the design process and user study is Chinese. We translate the participants' quotes, the prompts to large language models, and the example dialogues to English when they are presented in this paper.

3.1 Design Process

To design *ComPeer* for proactive peer support, we draw insights from related work and worked with five experienced users (U1 - U5, $M_{age} = 21.6, SD_{age} = 1.85$) of other conversational agents (CAs) and three psychological students (S1 - S3, $M_{age} = 21.6, SD_{age} = 0.47$). The users are recruited via an advertising post in an online forum about CAs, and the students are invited from a local university. To begin with, we survey CAs in mental health domains (*e.g.,* Vincent [38], WoeBot [25]) to get familiar with their messages and effective peer support strategies (*e.g.,* self-disclosure, reflect of feelings [41, 46]). We also outline potential contexts for proactive peer support from the literature on proactive AI systems (*e.g.,* [12], [62], [51]). Then, we picture six example dialogues in which *ComPeer* initiates a peer support chat with a user in a specific context, *i.e.,* Daily Greetings, Emotional Comfort, Life Sharing, Plan Reminder, Encouragement, and Offer Advice. In the examples, *ComPeer* proactively provides suggestions and encourages users to review in one evening based on their exam schedule for tomorrow. Next, we conduct semi-structured interviews with each participant to query about their experiences with CAs, feelings and comments about the example proactive dialogues, and suggestions and expectations for *ComPeer*.

Based on their feedback, we develop the first workable prototype of *ComPeer* with an architecture similar to the final one described in section 4. Then, five authors from our team conduct a one-week

Tianjian Liu, Hongzheng Zhao, Yuheng Liu, Xingbo Wang, and Zhenhui Peng

pilot study using the prototype CA and iterate the designs, *e.g.,* adding the long-term memory. After the iteration within the research team, we invite all participants (U1 - U5, S1 - S3) mentioned above to experience the *ComPeer* prototype and iteratively refine it based on their suggestions over the following two months. The key refinements in this process lie in the frequency of proactive messages, length of *ComPeer*'s messages, the prompt format of *ComPeer* personas, and the mechanism of event detection. We compile the following design principles for the refined *ComPeer* that is evaluated in our user study (section 5).

3.2 Design Principles

DP1. *ComPeer* should have a consistent persona as a peer with backgrounds similar to the user's and reveal it in its messages during the peer support chats. To gain the trust and interest of users, a CA needs to show consistent and vivid personalities [52, 74]. Otherwise, as mentioned by three participants, *"the CA would be lifeless and boring during the interaction"* (U1). Apart from the consistency, we have the following two sub-principles.

DP1.1. The persona should have customizable backgrounds or skills similar to the user's. Peer support requires the provider and the seeker to have similar experiences to achieve resonance and understanding in the conversation [56]. Therefore, as our participants are university students, we design the default persona of *ComPeer* prototype as a student who majors in computer science. Moreover, we allow participants to customize the personas based on their background and expectation before the interaction. During the interaction with the prototype, four participants agree that the personas similar to themselves can make the CA more realistic and reliable.

DP1.2. The language style of ComPeer's messages should adapt to its persona and match the characteristics of online peer support chats. ComPeer interacts with users in instant messaging apps via textual messages, which is the only medium for it to offer peer support. Two participants strongly suggest that *ComPeer*'s messages, especially the proactive ones, should adapt to its persona, as they view the proactive messages from previous CAs were *"very rigid"* (U2). Besides, the style of *ComPeer*'s messages should be in line with that used in normal online chats between peers [52]. Initially, the average length of *ComPeer* prototype's messages is 140.6 (SD = 54.78) Chinese words, which are considered quite long and complex by four participants, while *"online conversations are often brief and relaxed"* (U4). They suggest that the length of the messages should be limited, *e.g., "between 20 and 50 words"* (U5). Two participants further give suggestions on the composition of *ComPeer*'s messages. *"It shouldn't end every message with encouragement, as a real friend wouldn't do that in online chats"* (U3).

DP2. The content of the *ComPeer*'s proactive messages should include sharings about its daily life and/or concerns on the user's specific conditions. Many existing CAs (*e.g.,* [76], [25]) can initiate conversations with users, but the content of their proactive messages is usually limited to greetings and requests for interaction. Participants actively share their expectations on the content of *ComPeer*'s proactive messages as summarized below.

DP2.1. The ComPeer's proactive messages should disclose itself based on its persona and schedule. Mutual self-disclosure can promote peers' relationship [77] and benefit their perception of receiving peer support [77]. Four participants hope that *ComPeer* could share its experience and thoughts. *"It should share its life and emotions like a friend, making it feel alive"* (U1). For example, in the interactions with *ComPeer* prototype, U2 favors its proactive sharing that *"I'm happy for the basketball game with classmates this afternoon"*, which *"makes me feel it view me as a friend"* (U2).

DP2.2. The ComPeer's proactive messages should express concerns on the user's specific conditions. Showing concerns on the user's condition can reflect that *ComPeer* remembers what happened to the user and keep the message focused, which can make the user feel being cared and supported [18]. Two participants explain why the *ComPeer*'s proactive care is essential for them. *"I can know that the agent is specifically caring for me, while in real life, I seldom receive proactive support from other people"* (U4).

DP3. The frequency of the *ComPeer*'s proactive peer support should adapt to the content of its intended proactive message and the user's reaction to the previous proactive message. Initially, the *ComPeer* prototype is like a fixed-rule CA when sending proactive messages (*e.g.,* providing care every two hours), which is deemed inflexible and dull by five users in the design process. *"It just like a timer that sends pre-set messages to me"* (U1). Thus, we iterate the event detector and the schedule module to improve proactive messaging. However, it still automatically sends proactive messages whenever the timing of the events occurred, averaging 6.6 (SD = 1.88) times per day. Our participants complain that these proactive messages were too frequent, which would be disturbing especially when they see a new proactive message before replying to a previous one. *"It is time-consuming and exhausting to check and reply to so many messages in a day"* (U2). *"I delay responding to some non-urgent messages from the agent, but it eagerly sends another half an hour later, making me a bit annoyed"* (U4). Therefore, before sending proactive messages, *ComPeer* should infer if the content of the message would be important for the user and control the amount of proactive messages within a day. S2 further suggests that *"If the user doesn't respond to a proactive message within a time, it should pause the next scheduled one."*

DP4. *ComPeer* should apply effective peer support strategies to help users relieve their stresses. Enacting effective psychological strategies in the conversations can improve the efficacy of peer support chats [25, 46, 67]. In an early test of *ComPeer*, two users discover the conversations with *ComPeer* are inefficient and helpless due to *ComPeer*'s lack of support strategies, *"I shared my exam failure, but it only inquired about the content instead of offering comfort"* (U1). Our psychology participants help us identify the proper peer support strategies that *ComPeer* can use, such as Affirmation and Reasurance, Self-disclosure, and Reflection of feelings [41, 46], which can provide users with more effective support.

4 DESIGN AND IMPLEMENTATION OF COMPEER

In this section, we present the design and implementation of *ComPeer* based on the design principles derived from the design process. We first present a user scenario of *ComPeer* to illustrate how each

of its modules works (Figure 1) and then detail the implementation of each module in the *ComPeer*'s architecture.

4.1 User Scenario and Architecture of *ComPeer*

In this scenario (Figure 1), we describe how *ComPeer* had a peer support chat with Carl, a junior student who majored in computer science. Carl initialized *ComPeer*'s persona (subsection 4.2) as a student who has the same major and hobbies as him. After four days of interactions, one evening at 19:00, *ComPeer* proactively sent a message to Carl, "*I am reading a paper on deep learning and find it interesting. What are you doing now?*", Carl was anxious about his upcoming presentation on the next day and expressed his anxiety, "*Well... I am nervous about my presentation tomorrow.*". *ComPeer* comforted him using the "Affirmation and Reassurance" strategy, "*I understand you. Calm down and have a deep breath. Do you need any help?*". They then discussed how to prepare a PowerPoint for the presentation. Based on the content of the conversation, *ComPeer* proactively checked in with Carl to see how he was doing with his PowerPoint at 22:00 and encouraged him, "*How is your ppt going? I think your effort will pay off, so don't worry and have a good rest!*". Carl expressed gratefulness to *ComPeer*'s proactive care and felt being supported.

To enable *ComPeer* to provide such proactive peer support, we build up an architecture (Figure 1) of generative conversational agents powered by large language models (LLMs). The Memory module (subsection 4.3) stores all conversation history between the user and *ComPeer*. From each round of stored conversations (*e.g.*, no message from the user within five minutes in our case), the event detector module (subsection 4.4) extracts a user event (*e.g.*, "The user is worried about tomorrow's presentation and making PowerPoint") and infers a proper moment later that day (*e.g.*, 22:00 tonight) for *ComPeer* to proactively talk about it in the near future. These detected events with inferred timings of proactive messaging are sent to the schedule module, which decides whether each message will be proactively sent to the user. At the beginning of each day, the reflection module (subsection 4.5) triggers *ComPeer* to reflect on the user's conditions, emotions, challenges, and plans based on the conversations yesterday. The output reflections will help the schedule module initialize the planned proactive messages on that day. Lastly, the Dialogue Generation module (subsection 4.7) takes into considerations the *ComPeer*'s persona, relevant memory and proper peer support strategies to generate a message to initiate the conversation or respond to the user.

4.2 Personas Initialization

Following the persona settings used in previous works [59, 74], we describe the *ComPeer*'s persona with its name, age, gender, personality, occupation or major, background, hobbies, and language style. In addition, the persona describes its relationship with the user, which could enhance its understanding of the user during the conversation. For instance, the *ComPeer*'s persona in the user scenario (Figure 1) is an outgoing and gentle student (*personality*) who majors in artificial intelligence (*major*) and has previously studied geographical sciences (*background*). He has interests in cryptography, basketball, and video games (*hobbies*). He prefers

using exclamatory sentences in conversations (*language style*) and is a classmate of the user (*relationship with the user*).

At the beginning of our user study, participants can customize the persona of *ComPeer* based on their expectations of a supportive peer (DP1.1). When saving the persona as part of the prompt to LLMs, we add four pairs of example dialogues between *ComPeer* and the user, which enable the LLMs to conduct in-context learning for generating consistent peer-like messages [42] (DP1.2). Besides, we prompt *ComPeer* to be compassionate in the dialogue generation module regardless of its persona. The prompt template of *ComPeer*'s personas is listed in subsection A.1 in Appendix.

4.3 Memory Module

We refer to previous work on generative agents [59, 75] to design a memory module, which maintains a streaming record of the conversations between *ComPeer* and the user. It is a list of memory objects, where each object contains one pair of *ComPeer*-user messages, *e.g.,* [{"role": "user", "content": "I feel nervous because the deadline of homework is coming."}, {"role": "Assistant", "content": "I understand you, try your best to finish it! I am there to help you!"}]. When *ComPeer* initiates a chat or responds to the user, it needs to refer to the conversation history to keep its messages consistent and logical. In our practice, we get inspired from human's short-term and long-term memories [2] to prepare two databases for storing the agent's memories. One database maintains the context during the conversation, which serves as the short-term memory for generating *ComPeer*'s messages [97]. In our case, the context can contain at most 20 messages [2] of either *ComPeer* or the user prior to *ComPeer*'s intended message. The other database stores the long-term memory [16], *i.e.,* all the memory objects except those in the short-term database. We also encode each object of the long-term memory into a vector using OpenAI's text-embedding-ada-002 model [31], which helps *ComPeer* retrieve related memory to the current context in the later dialogue generation module.

4.4 Event Detector Module

While users can actively seek for comfort and support from *ComPeer*, the most unique design of our *ComPeer* is its ability to initiate a peer support chat with appropriate content at a proper timing (DP2, DP3). The event detector module is responsible for preparing the content and inferring the timing for the later schedule module to determine *ComPeer*'s proactive messages. Specifically, in the prompt to the GPT-4, we define an event as a specific incident or occurrence that brings about stress and has a significant impact on the individual [87]. We consider an appropriate event for *ComPeer*'s proactive care as poor physical condition [73], negative feeling [88], challenges [72], and plan of the user [20], which would desire social support from others. We constrain the inferred timing of the event as a time point later that day, as our trials on the prompts indicate that the GPTs often fail to infer a proper timing in the subsequent days. An example event is shown in Figure 1.

The event detector extracts an event from each round of conversation after it ends, *i.e., ComPeer* does not receive a message from

---

[2] We experiment different numbers of previous messages as the context. We find out that using at most 20 messages as the context can balance the trade-offs between costs of prompting LLMs and quality of the generated message in our case.
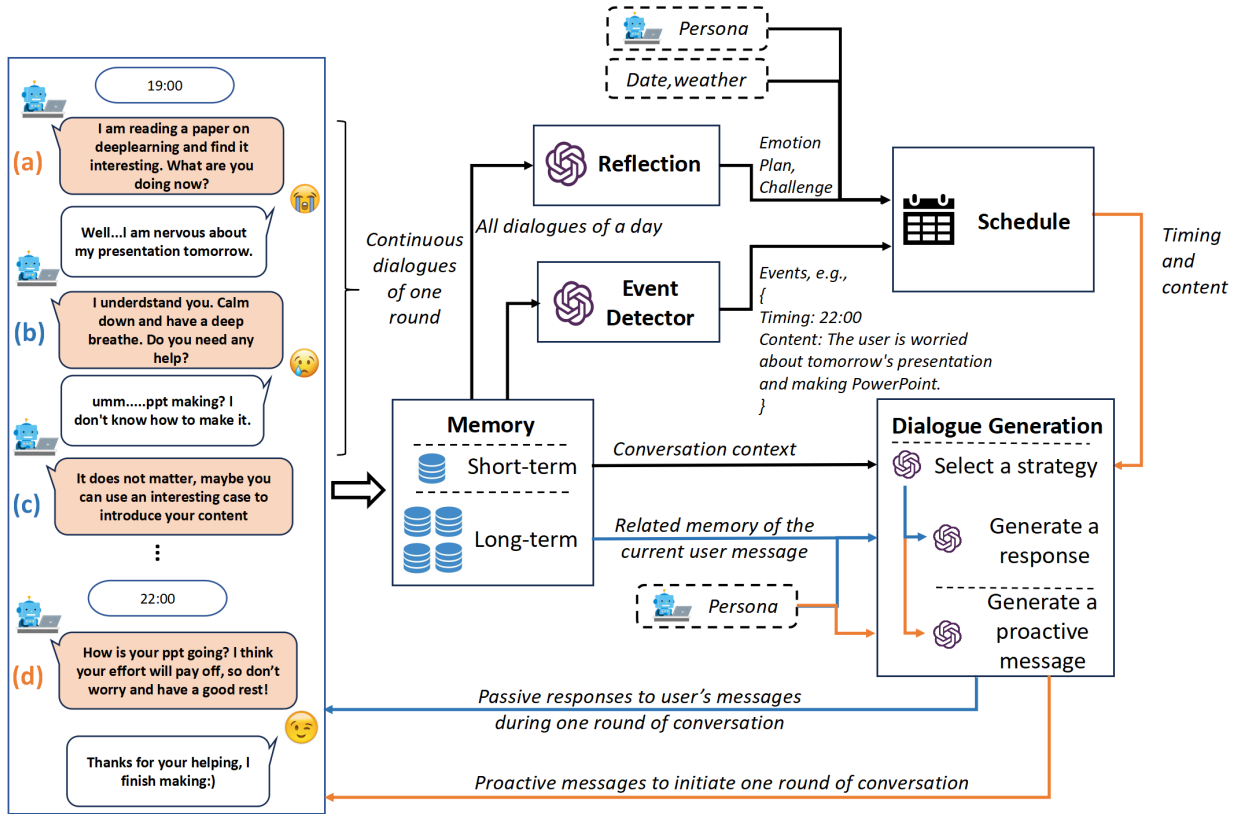
Figure 1: Architecture and an example user scenario of *ComPeer*, a generative agent for proactive peer support.

the user within 5 minutes, which can avoid detecting redundant conversation in a round of conversation. The event detector queries GPT-4, a state-of-the-art LLM at the time of *ComPeer*'s development, using the chain-of-thoughts (COT) prompts, which can enhance the LLM's ability of inference given a few data samples [13]. We experiment various prompts in the development process and iteratively improve the prompt to achieve acceptable performance in event detection. Lastly, we prepare five examples about the detection process of physical condition, negative feeling, challenges, plans, and no event in the prompt. The example COT prompts used in the event detector are shown in subsection A.2 in Appendix.

4.5 Reflection Module

While the events detected from one round of conversation help the schedule module update the planned proactive messages on that day, *ComPeer* needs to gain a higher-level understanding about the user to initialize a consistent plan of offering proactive care on the next day. Consider a case in which the user had several rounds of conversations with *ComPeer* on a day. *ComPeer* could have proactively cared for the user on the detected events about anxiety on the presentation and a plan for going shopping tomorrow. We hope that *ComPeer* can proactively initiate conversations on the next day based on what it learns from the user on that day, as

it would be beneficial to care for the user again on the event that causes the user's anxiety or could make the user happy.

Inspired by [59], at the beginning of each day, the reflection module asks three questions, *i.e.,* "*Does the user feel negative emotions due to something?*", "*Are users facing challenges or difficulties?*", and "*What plan does the user has in tomorrow?*", on all the dialogues of the previous day in the prompt to GPT-4. The module outputs three reflected thoughts, each for one question, *e.g.,* "1) *The user is nervous about presentation*, 2) *There is an presentation that user has to deliver*, 3) *The user will make presentation tomorrow*". These thoughts will be sent to the schedule module as a condition for generating the initial schedule of a day.

4.6 Schedule Module

The schedule module handles the initialization of the plan for proactive care on a day based on the input from the reflection module, the update of the planned proactive messages on that day based on the input from the event detector, and adjustment of the frequency of proactive care based on the content of each proactive message and the user's reaction to this message (DP3).

4.6.1 Initialization of the schedule on a day. At the beginning of each day, *ComPeer* will initialize schedules for the day. Apart from the reflected thoughts about the user on the previous day, the schedule module incorporates the real-world information (*e.g.,* date,

Table 1: Descriptions and examples of candidate peer support strategies used by *ComPeer*. *ComPeer* can select anyone of all the strategies for generating a passive response to users message. However, *ComPeer* is allowed to use *self-disclosure, inquiring, affirmation and reassurance, invite users to think* for generating a proactive message, as the usage of other strategies needs the user starts the conversation first.

Strategy	Description	Example
Self-disclosure	Disclose personal information to users, including but not limited to the counselor's own similar experiences, feelings, behaviors, thoughts, *etc.*.	"I also have a similar experience! I did such a thing last time!"
Inquiring	Explore users' subjective experiences or ask users to concretize the imprecise factual information	"Are you feeling better now? How do you feel now?"
Affirmation and Reassurance	Affirm users' strengths, motivations, and abilities, and normalize users' emotions and motivations, and provide comfort, encouragement, and reinforcement.	"I believe you can! It will get better soon! I will support you no matter what!"
Invite users to think	Answer the questions that users ask about the conversation topics.	"Taking a deep breath may be a good way to relax, but it's important to first identify the root cause of the problem."
Reflection of feelings	Use tentative or affirmative sentence patterns to explicitly reflect the users' mood, feelings, or emotional states.	"I understand your current annoyance."
Restatement or Paraphrasing	Reflect the content and meaning expressed in users' statements to obtain explicit or implicit feedback from users.	"It sounds like you think everyone is ignoring you, right?"
Answer	Answer the questions that users ask about the conversation topics.	"I think you need to read the book 'Learning Neural Networks' now."

weather) of that day into the prompt for generating a schedule closely related to the user's real life. For example, if the reflection module indicates that the user had a plan or encountered a challenge on previous day, *ComPeer* could remind the user of the plan or care for the user about the challenge on that day. If the weather is rainy, *ComPeer*'s schedule had better not include outdoor activities. One example of the real-world information in the prompt is: "Today is Friday. The weather of today is rainy. The lowest temperature today is 4℃ and the highest temperature is 9℃". Besides, the prompt for schedule initialization also consists of *ComPeer*'s persona to guide the GPT-4 to generate the schedule with events that a peer should care (DP1.1). By considering the reflected thoughts about the user, real-world information, and *ComPeer*'s persona in the prompt, the schedule module can initialize a schedule on that day. The specific schedule template is shown in subsection A.4 in Appendix.

4.6.2 Update of the schedule and adjustment of the frequency of proactive messaging on that day. Once the event detector extracts an event from one round of conversation, the schedule module inserts the extracted event in the priority queue of the schedule. If the conversation do not contain the event, the detector will output "" (no event). In this case, *ComPeer* relies on the initial schedule to plan for proactive messages. As a default setting, when the planned timing of an event comes, the schedule module will dequeue the event and send it to the dialogue generation module to initiate a conversation. However, following the default setting, if the planned events in the schedule are too much, the frequent proactive messages would be annoying. To adjust the frequency of proactive messaging, the schedule module evaluates the content of the planned event and

the user's reaction to *ComPeer* previous proactive message (DP3). As suggested by [76], a proactive message related to the user's information will help the user feel be cared and supported and maintain higher engagement. Therefore, we allow a LLM to infer the importance value (a score between 0 and 1) of each planned event. For example, an event like having a dinner with friends or reading books in the library could be less important than an event that is more about the user's mental state, *e.g.,* the user is depressed for the failure of an exam. To encourage proactive care on more "important" events while ensuring that all events have a chance to be proactively sent, we design a randomized mechanism for the schedule module to make decisions. When the timing of an event arrives, the module generates a random number between 0 and 1. If the event has an importance value larger than the random number, it will be sent to the dialogue generation module. Besides, if the user does not reply to *ComPeer*'s previous proactive message for a certain period, *e.g.,* three hours in our study, the schedule module will not send other planned events within this period to the dialogue generation module.

4.7 Dialogue Generation Module

The dialogue generation module is responsible for generating peer-like, consistent, and supportive messages to respond to or initiate a conversation with the user (DP1, DP4). To achieve this purpose, we prompt GPT-4 to act as three roles (noted as LLM-1, LLM-2, and LLM-3), which can simplify the generation tasks by allowing the LLM focus on one task at a time [85]. LLM-1 takes in the conversation history stored in the short-term memory and selects one or more peer support strategy that can be used in *ComPeer*'s

message. As suggested by our psychological students in the design process, the candidate strategies for generating passive responses are self-disclosure [39, 40], inquiring [41], affirmation and reassurance [14, 46], invite users to think [41, 67], reflections of feelings [14, 27], restatement or paraphrasing [37] and answer [41]. Table 1 shows the description and example of each strategy. As for the generation of proactive messages, the candidate strategies are self-disclosure (DP2.1), inquiring (DP2.2), affirmation and reassurance (DP2.2), and invite users to think (DP2.2). The other three strategies, *i.e.,* reflection of feelings, restatement or paraphrasing, and answer need user to express, need the user to express feelings first and are not suitable for *ComPeer*'s proactive messaging.

LLM-2 generates a passive responding message conditioned on the peer strategy selected by LLM-1, the conversation history , *ComPeer*'s persona, and the related long-term memory of the current user message. To retrieve related long-term memory, we encode the current user message into a vector using text-embedding-ada-002 and calculate its cosine similarities with the vectors of memory objects stored in the long-term memory database. We select at most three memory objects that have the highest similarity scores for the prompt to LLM-2.

To generate proactive messages, we choose to design another prompt to require the LLM (note as LLM-3) to act as a role different to that of LLM-2, because we find in our practice that an LLM has difficulty distinguishing its roles if we prompt it to handle the generations of passive responses and proactive messages at the same time. The prompt to LLM-3 considers *ComPeer*'s persona, the event from the schedule module, and the peer support strategy selected by LLM-1. The prompts for LLM-1, LLM-2, and LLM-3 are presented in subsection A.6, subsection A.7, and subsection A.8 in the Appendix.

5 USER STUDY

To evaluate the effectiveness and user experience of *ComPeer*'s proactive peer support, we conduct a one-week between-subjects user study with 24 participants compared to a baseline conversational agent (CA) that can not proactively initiate the conversation with users. After the one-week study, we invite all participants, either in the *ComPeer* or baseline group, to use *ComPeer* for another week, which provide us with more qualitative findings especially on the long-term user experience of *ComPeer*. Our research questions (RQs) in the user study are:

RQ1: How would interacting with *ComPeer* affect the users' management of their stress?

RQ2: How would users perceive a) the social support provided by *ComPeer*, b) the persona of *ComPeer*, and c) their relationship with *ComPeer*?

RQ3: How would users perceive *ComPeer*'s proactive messages?
RQ4: How would users interact with *ComPeer*?

5.1 Participants

We recruit 24 participants (15 males, 9 females, aged 18-25) via recruitment posts in three universities. The inclusion criteria in the background survey are that participants have recently encountered stressful issues and they have interests in interacting with a peer support CA every day in a two-weeks period. On average,

their perceived level of stress measured using PSS-10 [15] was 23.63 (SD = 3.98), indicating that most of them perceived high level (score > 19) of stress. Their majors vary across computer science, psychology, international politics, electronic engineering, and so on. Their stress primarily originates from academic workload and examinations (16 participants), interpersonal relationships (5 participants), and personal future development (6 participants). Although only two participants report experience of disclosing themselves to a CA or using AI companion applications (*e.g.,* Character AI, Replika), eighteen participants have experience of using general CAs (*e.g.,* ChatGPT, Qwen). All participants are native speakers of Chinese, the language used by *ComPeer* in the design and development process. We randomly divide our participants into the *ComPeer* group (PC1-PC12, 7 males, 5 females, mean age: 20.63, mean PSS-10 score: 24.45) and the baseline group (PB1-PB12, 8 males, 4 females, mean age: 20.08, mean PSS-10 score: 23.54).

5.2 Baseline Conversational Agent

The primary goal of our one-week between-subject study is to evaluate the effects and user experience of the proactive peer support mechanism we built for *ComPeer*. Therefore, the baseline CA does not have the Event Detector, Schedule, and Reflection modules that are responsible for proactive peer support. Nevertheless, it has the same Memory and Dialogue Generation modules as those of *ComPeer* (Figure 1), which enable it to react to user-initiated messages in a similar way of *ComPeer*. Such a CA can not only provide a strong and fair baseline for evaluating the impact of *ComPeer*'s proactive peer support but also offer more qualitative data for evaluating the usability of the memory and dialogue generation modules in our architecture of generative CAs.

5.3 Measurements

Table 2 summarizes the items used in the questionnaire after participants' interactions with either *ComPeer* or the baseline CA everyday. All items are measured on the 7-point Likert scale, with 1 for strongly disagree and 7 for strongly agree.

RQ1. Effects on stress management. We use the Pressure Perceive Scale 10 (PSS-10, 10 items, total scores: 0 - 40) [15] to measure each participant's perceived level of stress at the beginning of the user study, after one week of interactions with the CA, and after two-week of interactions. Following [57], we evaluate the change of participants' perceived stress before and after interacting with our *ComPeer* or the baseline CA. Besides, in the questionnaire sent to participants every evening during the study, we ask participants to rate their agreement on the following statement adapted from [48]: "*I feel relieved when I talk to the CA.*"

RQ2. Perceived Social Support, Persona, and Relationship. To measure how the user perceive the social support provided by *ComPeer*, we adapt five items from [48, 81]. They are "*The CA is accompanying me*", "*The CA is caring about me*", "*The CA spares no effort to help me out of difficulties*", "*The CA's advice is very important to me*", and "*The CA is encouraging me to be a better person in life*". As for the perceived persona of the CA, we apply two items from [71]: "*The CA's response is logical and coherent*", and "*The persona of the CA is interesting enough*". We capture how participants perceive their relationship with the CA using three items adapted from [48]:
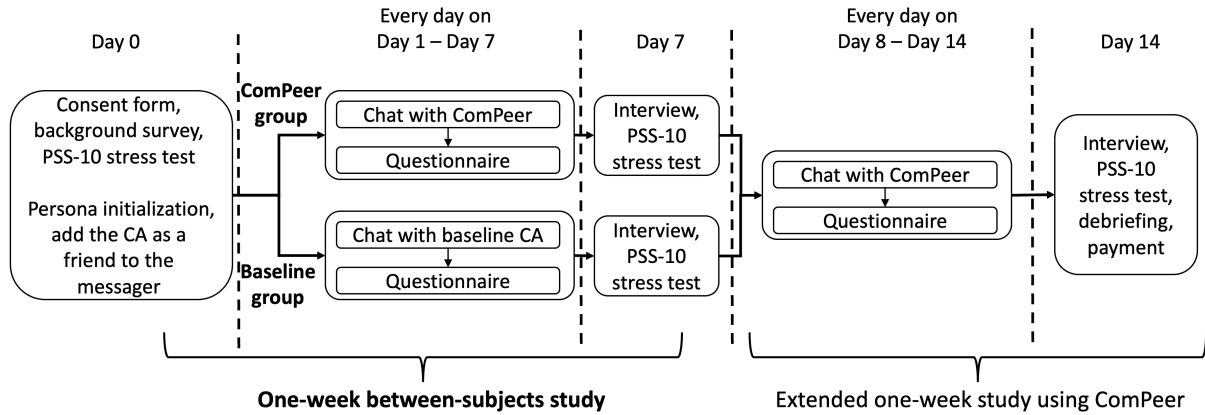
Figure 2: Procedure of our one-week between-subjects study (*ComPeer* vs baseline conversational agent (CA)) and extended one-week study in which all participants converse with *ComPeer*.

"*Chatting with the CA has deepened my relationship with it*", "*When I feel in trouble, I am willing to seek help from the CA*", and "*When I encounter something joyful, I am willing to share it with the CA*".

RQ3. Perceptions of *ComPeer*'s Proactive Messages. In the questionnaire sent to participants in the *ComPeer* group every evening during the study, we ask them to rate their satisfaction with each proactive message from *ComPeer*, "*I am satisfied with this proactive message*". We also conduct a thematic analysis on the topic of each proactive message from *ComPeer*. Three authors of this paper first independently label fifty proactive messages. They then meet and discuss their labels and coding schemes and reach an agreement on five topics, *i.e.,* share the CA's daily routine, share the CA's work or study, share the CA's leisure and entertainment, care for the user's work or study, care for user's state. They then independently code all the proactive messages (intraclass correlation coefficient $ICC = 0.851$) and resolve disagreement via discussion. We evaluate participants' satisfaction with each topic of proactive messages.

RQ4. Interaction with and Concerns on the CA. We analyze timing and content of each interaction between the CA and the user. Specifically, we calculate the numbers of peer support strategies that *ComPeer* and the baseline CA selected to generate their messages. We also conduct a thematic analysis on the topic of each round of conversation from the user's perspective. Two authors of this paper first independently label fifty rounds of conversations. They then meet and discuss their labels and coding schemes and reach an agreement on six topics, *i.e.,* concern, want, thoughts or feelings, share happiness, share life, and others. They then independently code all the proactive messages ($ICC = 0.764, 0.807, 0.723, 0.724, 0.772, 0.785$) and resolve disagreement via discussion.

5.4 Task and Procedure

Figure 2 illustrates the procedure of our user study conducted remotely. After obtaining participants' consent, background information, and perceived level of stress on Day 0, we guide them to initialize the CA's persona with expected attributes on the CA (subsection 4.2) and their personal information (*i.e.,* age, major, stress

issue) via an online document. In the personas initialization, four participants set it with majors or hobbies different from their own, *e.g.,* Computer Science vs. Law. We help them add the CA as a friend on Tecent QQ, a popular instant messaging app in China. In the consent form, we have informed participants that experimenters will look at their conversation record with the CA and that third party technology providers the CA relies on, *i.e.,* Tencent, have access to the information. We add that participants' personally identifiable information will not be shared for publication purposes and that their voluntary participation means that they can drop out of the study at any point. We also state that the CA is not a therapist and that they should seek professional help if any psychological issues are experienced, though we target a non-clinical population.

Every day from Day 1 to Day 7, they are asked to have at least one conversation with either *ComPeer* or the baseline CA based on their assigned groups. The experimenter send out a questionnaire (Table 2) to participants every evening and requires them to fill it based on their conversation with the CA on that day. On the evening of Day 7, we contact participants, ask them to rate their perceived level of stress, and conduct semi-structured interviews on impact of *ComPeer*, their perceptions of *ComPeer*'s messages, as well as the frequency and content of their interactions with *ComPeer*. After the one-week between-subjects study, we enable the proactive peer support mechanism of the baseline CA (*i.e.,* it is *ComPeer* now) and invite all participants to chat with *ComPeer* every day for another week. Lastly, on Day 14, we conduct another semi-structured interview with each participant, focusing on their feelings towards *ComPeer*, and their perceptions of *ComPeer*'s proactive messages. We estimate that each participant spent about 100-150 minutes in our study, including filling questionnaires, interviews, and chatting with the CA. After the study, each participant receives about $10 compensation following the local payment policy.

5.5 Data Analyses

To evaluate the change of participants' perceived stress (RQ1) before and after interacting with our *ComPeer* or the baseline CA (between-subjects factor), we conduct a two-way mixed ANOVA [80] on their PSS-10 scores measured on Day 0 and Day 7 (within-subjects factor).

Table 2: Measured items for RQ1-3 in the questionnaire sent to each participant every evening. Participants rate their levels of agreements on the statement in each item on the 7-point Likert scale; 1 - strongly disagree, 7 - strongly agree. CA stands for Conversational Agent.

Research Question	ID	Item	Reference
(RQ1) Stress Management	Q1	I feel **relieved** when I talk to the CA.	[48]
(RQ2a) Perceived Social Support	Q2	The CA is **accompany**ing me.	[48, 81]
	Q3	The CA is **caring** about me.	[48]
	Q4	The CA spares no effort to **help** me out of difficulties.	[48]
	Q5	The CA's **advice** is very important to me.	[48, 71, 81]
	Q6	The CA is **encouraging** me to be a better person in life.	[48]
(RQ2b) Perceived Personas	Q7	The CA's response is logical and **coherent**.	[71, 81]
	Q8	The persona of the CA is **interesting** enough.	[71]
(RQ2c) Perceived Relationship	Q9	Chatting with the CA has deepened my **relationship** with it.	[48, 71, 81]
	Q10	When I feel in trouble, I am willing to **seek help** from the CA.	[48, 81]
	Q11	When I encounter something joyful, I am willing to **share** it with the CA.	[48]
(RQ3) Satisfaction with Proactive Messages	Q12	I am **satisfied** with this proactive message.	-

As for each self-reported item of RQ1 and RQ2 (Table 2), we average the scores reported every day from Day 1 to Day 7 as the final score of the item for each participant, with the Cronbach's a of the items are above 0.8 except for Q2 (0.65) and Q5 (0.63). Based on whether these items satisfy data normality under the Shapiro-Wilk tests (attached in subsection A.8 in Appendix), we use two different statistical methods to compare the results between two groups. For the items Q1, Q4, and Q7 that violate data normality, we use the Mann-Whitney U test, a non-parametric test commonly used to compare differences between independent conditions (*e.g.*, in HCI studies [10, 96]) especially when the data normality is violated. For the other items, we run unpaired T tests to compare the scores between the two group, which is a parametric test widely utilized in HCI research (*e.g.*, [47]) to compare in normal data.

Apart from comparing the means of the self-reported items for RQ1 and RQ2 in two groups in the one-week between-subjects study, we also analyze how participants' perception with the CA change over time. Specifically, we conduct regression analyses using the time (Day 1 - Day 7) as independent variable and the score of each item on that day as the dependent variable. We further conduct ANCOVA analyses [33] to assess whether there exist significantly differences between the change of participants' perceptions in the *ComPeer* and baseline CA. As the *ComPeer* group in the one-week between subject study continues to use *ComPeer* for another week in the extended study, we also apply regression analyses using the time (Day 1 - Day 14) as independent variable and the score of each item on that day as the dependent variable (attached in subsection A.8 in Appendix).

To understand users' perception of proactive messages (RQ3), we calculate the average score for participants' satisfaction with each topic of *ComPeer*'s proactive messages in the one-week between-subjects study. We use Kruskal-Wallis H test to analyze the differences among the scores for each message topic. The Kruskal-Wallis H test is also a non-parametric test commonly used to compare differences in more than two independent conditions (*e.g.*, in HCI studies (*e.g.*, [30, 69])), especially when the data normality was

violated, as confirmed in our cases. To compare the topics of conversations in the *ComPeer* and baseline group, we conduct unpaired T test and Mann-Whitney U test on the number of occurrences for each topic, based on their normality (attached in subsection A.8 in Appendix).

Lastly, for the transcribed interview scripts in the semi-structured interviews on Day 7 and Day 14, we conduct a thematic analysis. Two authors first familiarize themselves by reviewing all the text scripts independently. After several rounds of coding with comparison and discussion, they finalize the codes of all the interview data regarding each RQ. We count the occurrences of codes and incorporated these qualitative findings in the following presentation of our results.

6 RESULTS

All participants submitted the questionnaire every day and had at least one round of conversation with the assigned CA during the one-week between-subjects study and the extended one-week study using *ComPeer*. Therefore, we include data of all participants to present the following results. Table 3 shows the results of statistical tests for self-reported items for RQ1-RQ2 in the one-week within-subjects study. In this section, we first present the RQ1-RQ4 results from the one-week between-subjects study. We then list the qualitative results from the extended one-week study to complement our findings on RQ1-4.

6.1 Effects on Stress Management (RQ1)

i) **Participants with either *ComPeer* or baseline CA perceive significantly less stress after interacting with it for 7 days.** As shown in Figure 3(a), after interacting with either *ComPeer* or the baseline conversational agent (CA) for one week, participants' perceived level of stress has a significant decrease on Day 7 compared to that on Day 0; (*ComPeer*:$M = 19.75, SD = 4.43$, Baseline: $M = 17.17, SD = 2.79$); ($F = 20.527, p < .001$). There is no significant difference in the perceived stress level between two groups;

Table 3: Results of statistical tests on the users' perceived average stress (RQ1), social support (RQ2a), personas (RQ2b), and relationship with CAs (RQ2c) in the one-week between-subjects study. Based on the normality, the tests for the means of one-week scores include unpaired T test (Q2, 3, 5, 6, 8, 9, 10, 11) and Mann-Whitney U test (Q1, 4, 7) with Bonferroni correction. The regression analyses examine how participants' perceptions with the CA change every day in the one-week study, and the ANCOVA analyses assess whether there exist significantly differences between the change of perceptions in the *ComPeer* and baseline CA. Note: * : .05 < p <.10, ** : .01< p < .05, ***: p < .01.

| Research Question	ID	Item	ComPeer Mean/S.D	Baseline Mean/S.D	Compare means of one-week scores		Statistics in regression analyses on scores over time				Statistics in ANCOVA tests	
					T/U	p	β_c	R_c^2	β_b	R_b^2	F	p
Stress Management (RQ1)	Q1	Relieved	4.77/1.13	4.77/0.89	76	0.817	0.122**	0.67	-0.024	0.078	7.58	0.02
Perceived Social Support (RQ2a)	Q2	Accompany	5.33/1.15	4.90/0.64	1.13	0.27	0.092	0.41	0.027	0.015	0.324	0.582
	Q3	Caring	5.71/0.81	4.95/0.75	2.39	0.025	-0.024	0.06	0.083*	0.54	3.975	0.074
	Q4	Help	5.63/0.68	5.18/0.76	97	0.147	-0.146**	0.73	-0.023	0.008	9.276	0.012
	Q5	Advice	4.15/1.18	4.10/1.08	0.12	0.89	0.128**	0.74	0.006	0.08	8.197	0.002
	Q6	Encouraging	5.79/0.78	5.26/0.69	1.73	0.095	0.003	0.002	0.054	0.13	0.357	0.563
Perceived Personas (RQ2b)	Q7	Coherent	4.87/1.17	4.70/0.94	76.5	0.795	0.048	0.52	0.086	0.24	0.732	0.432
	Q8	Interesting	4.46/1.10	4.30/1.10	0.37	0.71	0.051	0.31	-0.033	0.05	1.321	0.277
Perceived Relationship (RQ2c)	Q9	Relationship	4.96/1.06	4.58/0.83	0.98	0.33	0.08	0.29	0.006	0.005	0.825	0.385
	Q10	Seek	4.46/1.10	4.36/1.38	0.41	0.68	-0.012	0.02	-0.032	0.122	0.333	0.577
	Q11	Share	4.80/1.49	4.39/1.19	0.66	0.51	0.088*	0.459	0.003	0.061	1.255	0.289
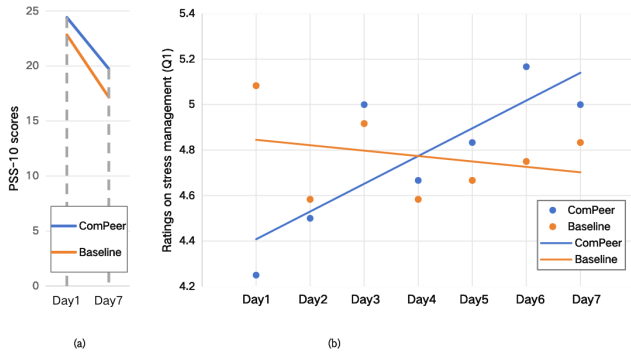
Figure 3: (a) Changes of participants' perceived stress measured by PSS-10 over the user study. (b) The regression result of Q1.

(F = 3.284, p > .05). When chatting with either *ComPeer* or the baseline agent (median: 4.92 vs. 5.00), participants feel slightly relieved (Q1 in Table 3); U = 76, p > .05. These results indicate that the peer support provided by our generative CAs, either in a proactive or a passive way, could be helpful for stress management, as suggested by six participants in the *ComPeer* group four participants in the baseline group. "*The personal experiences it shares make me feel like I am not under stress alone*" (PC1, **M**ale, age: 21). "*It inquired my condition and gave targeted advice after I sought help*" (PB2, **F**emale, age: 19).

ii) *ComPeer* **performs better in making users feel relieved over time**. Specifically, time exhibits a strong positive correlation with perceived relief in the *ComPeer* group (β = 0.122, R^2 = 0.67), while time is inversely correlated with perceived relief in the baseline group (β = −0.024, R^2 = 0.078), as illustrated in Figure 3(b). The

ANCOVA analysis confirms that the difference in these trends between *ComPeer* and the baseline is statistically significant (p < .05). The result of regression indicates although the baseline group initially reports higher perceived relief compared to the *ComPeer* group, participants in the baseline group experience a gradual decline in perceived relief as time progresses. Conversely, participants in the *ComPeer* group report increasingly higher perceived relief over time. The difference in regression models suggests that the insignificance of the means for stress management (Q1 in Table 3) between the two groups might be due to variations in the experience evolution.

This result can be accounted for *ComPeer* proactive messages sourced from its reflection module, as indicated by three participants. "*On this morning, it remembers the academic pressure I mentioned yesterday and encouraged me to make a daily plan, which made me feel supported and relaxed*" (PC8, F, age 22).

6.2 Perceptions of Social Support, Persona, and Relationship (RQ2)

i) **Participants with *ComPeer* gradually develop a stronger sense of receiving good advice than those with the baseline CA.** As for the five items (Q2-Q6 in Table 3) that measure perceived social support (RQ2a) received from the CA, there is no significant difference between *ComPeer* and the baseline groups after Bonferroni correction (p > .05/5). The regression analyses (Table 3) further indicate that over the seven days, participants in the *ComPeer* group have a gradually stronger sense of receiving good advice (β_c = 0.128, R_c^2 = 0.74), compared to those in the baseline group (β_b=0.006, R_b^2 = 0.08). Such a difference is confirmed significant by ANCOVA (p < .05). This result can also be accounted for *ComPeer*'s proactive messages, especially those sharing concerns (mentioned by four participants) and giving advice (mentioned by another four participants). "*When it encourages me to overcome the anxiety from the*

exam and give me some tips based on its experiences, I feel that the agent put itself in my place to give me advice" (PC7, F, age: 20). "*It reminds me to rest early at night, which is a useful advice for me*" (PC1, M, age: 21). However, the baseline CA tends to perform better than the *ComPeer* in offering care (β_c = -0.024, R^2_c = 0.06, β_b = 0.083, R^2_b = 0.54), and helping users out of difficulties (β_c = -0.146, R^2_c = 0.73, β_b = -0.023, R^2_b = 0.008), as the time goes by. The ANCOVA analysis indicates the differences in these items between the trends of two groups are significant ($p < .05$). This can be due to the fact that users may get used to the proactive support from *ComPeer* and take it for granted. "*Its (ComPeer's) care and support have become something I take for granted, and now I do not have a sense of being cared as strong as that at the beginning*" (PC12, M, age 20).

ii) **Participants somewhat agree that the personas of both CAs are coherent and interesting**. On average, participants' agreements on the items that the CA has a coherent (median: 4.92 vs. 4.71) and interesting persona (4.46 vs. 4.30) are between "either agree or disagree" and "somehow agree" in both the *ComPeer* and baseline conditions. There is no significance difference between the two conditions regarding the average 7-days ratings on perceived persona ($p > .05/2$, Mann-Whitney U test or unpaired T test).

iii) **Participants in both groups report similar perceptions regarding the relationships fostering.** There are no significant differences between the two groups regarding the three items that measure perceived relationship ($p > 0.5/3$). The regression analyses further indicate that participants either in *ComPeer* or the baseline have a homologous sense that chatting with the CA has deepened their relationships with it over the 7-days interactions (β_c = 0.08, R^2_c = 0.29, β_c = 0.006, R^2_b = 0.005). Over the 7 days, participants in both groups become less willing to seek help from the CAs (β_c=-0.012, R^2_c = 0.02, β_b=-0.032, R^2_b = 0.122), which can be due to the improper suggestions given by the CAs on the early days, as suggested by in total six participants in both groups. "*When I am seeking suggestions on scientific papers related to my research, it recommended some papers that do not exist*" (PC3, M, age: 23). "*Sometimes its advice is hard to implement, such as 'persuading your leader to change the plan'*" (PB1, F, age: 18). Nevertheless, both user groups are more willing to share joyful events to the CA over time (β_c = 0.088, R^2_c = 0.459, β_b = 0.003, R^2_b = 0.061). Five participants especially value the positive response from CAs on their joyful events. "*I do love its congratulations on my success and praise for my efforts, which truly touch me*" (PB11, F, age: 20).

6.3 Perceptions of the Proactive Messages (RQ3)

Figure 4 shows participants' satisfaction with *ComPeer*'s proactive messages during the one-week between-subjects study.

i) **Overall, participants are satisfied with the received proactive messages from *ComPeer*** ($M = 5.194, SD = 1.10$). In the first week, participants in the *ComPeer* group received in total 202 proactive messages ($M = 16.83, SD = 4.21$). As detailed in subsection 5.3, we categorize the topic and sub-topic of each proactive message, which is summarized in Table 4. Most of the proactive messages are *ComPeer*'s self-disclosure of its life, such as sharing its daily routine ($N = 46$), its leisure and entertainment (37), and its work or study (41). *ComPeer* also actively cares for the user about mental state ($N = 48$) and work or study (30) in the proactive messages.
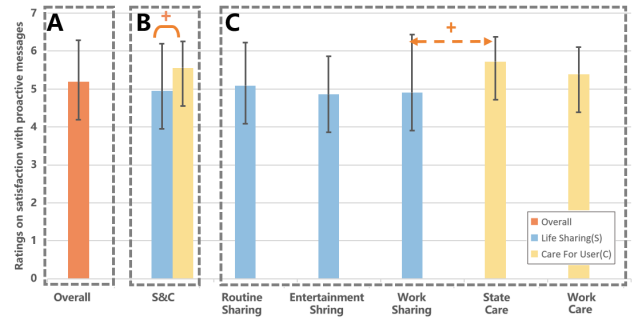

Figure 4: Results of each topic. A is the average satisfaction for each proactive dialogue, B is the average satisfaction between the two major topics, and C is the satisfaction across each sub-topic. Note: + : .05 < p < .10.

ii) **In general, participants tend to be more satisfied with the proactive messages that care for the user than the self-disclosure ones**. The follow-up Mann-Whitney U test on the satisfaction with *ComPeer*'s proactive messages between the topics of self-disclosure ($M = 4.95, SD = 1.25$) and care for the user ($M = 5.60, SD = 0.70$) reveals marginally significant difference between the topics ($U = 322, p = 0.091$). Figure 4C shows the average score and the result of the Kruskal-Wallis test on the participants' satisfaction with the proactive messages among sub-topics. We observe a marginally significant difference regarding participants' satisfaction with the proactive messages that care for the user's state ($M = 5.73, SD = 0.71$) and those that share *ComPeer*'s work or study ($M = 5.39, SD = 0.65$).

Eight participants suggest that the agent's proactive care on their states makes them feel accompanied and warmed. "*When it inquired me whether my illness had relieved, I felt touched and accompanied*" (PC7, F, age: 19). The proactive concerns on the user's work or study can also be helpful, as four participants suggest that these messages can enhance their confidence on and efficiency in their work or study. "*It comforts me and suggests ways to prepare for my exam, which relieves my pressure*" (PC5, M, age 20). Regarding the proactive messages of sub-topic "sharing routine" ($M = 5.09, SD = 1.14$), "sharing work or study" ($M = 4.91, SD = 1.53$), and "sharing entertainment" ($M = 4.86, SD = 1.00$), participants suggest that these messages can also have positive impact on them. For example, six participants mention that the proactively shared routines by *ComPeer* also make them feel being accompanied. "*It is as if it was experiencing a life similar to me*" (PC1, M, age 21). Five participants believe that the proactive sharing of the agent's work can serve as examples and important guides in their lives. "*When I am escaping from my homework, the agent shares its work today, which reminds me that I should face the challenges in my homework as it does*" (PC6, F, age: 20). Lastly, seven participants mention how the agent's proactive sharing of entertaining things enhance their engagement with it. "*When receiving a proactive message which shares the agent's hobby related to mine, I was delighted to talk more about the topic with it*" (PC2, M, age:19).

Table 4: The topic of proactive messages discovered in our study. Each proactive message is associated with a sub-topic.

Topic	Sub-Topic	Examples	Count
	Share the CA's daily **routine**	"Good morning! I'm having a delicious breakfast. Hope you also have a day!"	46
Self-disclosure	Share the CA's leisure and **entertainment**	"I am on my way to karaoke and can't wait to sing with my friends!"	37
	Share the CA's **work** and study	"I am participating in an NLP seminar and there is a work I find very interesting"	41
	Care for user's **state**	"How do you feel now? Did you take a good rest as I said before?"	48
Care for User	Care for user's **work** or study	"Have you made any progress in your experiments on digital circuits? If you encounter difficulties, we can discuss together!"	30

iii) **Participants have concerns on some proactive messages**. Nevertheless, several participants express their concern with *ComPeer*'s proactive conversation. For example, three participants feel more anxious when seeing the agent's sharing of its hard work. "*The agent is constantly studying hard, which brings peer pressure to me*" (PC7, F, age: 19). Four participants also note that the proactive sharing of agent's entertainment is sometimes annoying. "*It always invites me to play basketball while I do not like to play it*" (PC9, M, age 20). Besides, five participants point out that the sharing of *ComPeer*'s routine was somewhat mechanical. "*It gets up and sends me messages at 7:00 am about what it is going to do almost every day, which is boring and mechanical*" (PC4, M, age: 20). Three participants mention sometimes the proactive messages from *ComPeer* can disturb them. "*When I am busy with my experiment, it shares its lunch with me, which is a disturbance for me*" (PC5, M, age: 20). These concerns exhibit negative user experiences during the interactions, potentially leading to insignificance of the means for perceptions of social support, persona, and relationships. For instance, sharing efforts from *ComPeer* could fail to encourage some participants and induce stress to them (Q6 in Table 3). The mechanized proactive messages may render *ComPeer* incapable of providing companionship with the participants (Q2), further diminishing the personas interest of *ComPeer* (Q8). Besides, some disturbing and annoying messages might also prevent some participants from stronger relationships with *ComPeer* (Q9).

6.4 Interaction with *ComPeer* (RQ4)

i) ***ComPeer*'s proactive messages boost participants' engagement with the agent.** Figure 5(a) shows the average rounds of conversations between the CA and user in the *ComPeer* and baseline groups. We count it one round of conversations only if the user does not send another message to the CA within five minutes after he/she replies to *ComPeer* proactive message or proactively sends a message to the CA. Although the rounds of conversations initiated by participants in *ComPeer* group ($M = 7.83, SD = 4.58$) are lower than the baseline group ($M = 10.58, SD = 4.61$), participants additionally engage with *ComPeer* for 11.92 ($SD = 4.86$) times after receiving its proactive messages. The follow-up Mann-Whitney U test confirms that participants in *ComPeer* group have significantly more rounds of conversations with the CA ($U = 122, p = 0.004$). Three participants mention their willingness to reply to *ComPeer*'s

proactive messages. "*When it shared its life, I was curious and asked it more about its shared event*" (PC2, M, age: 19). Two participants express that the content of proactive messages aligns with their lives, "*When I was in class, it happened to share with me something about its class. I was happy to discuss the course with it*" (PC9, M, age: 20).

ii) **Participants share more concerns, thoughts/feelings, joy, and daily life with *ComPeer* than the baseline CA.** Table 5 and Figure 5(b) further shows the numbers and distributions of the conversation topics in both groups. Participants actively talk about what they want ($N = 156$), how they think or feel (138), and what is their life (100) in the conversations with our CAs. We observe that the conversations about "concern" ($U = 118, p = 0.007$), "thought or feelings" ($T = 3.52, p = 0.002$), "joy sharing" ($U = 108, p = 0.033$), and "simple life disclosure" ($T = 2.62, p = 0.017$) happen significantly more in the *ComPeer*'s group than those in the baseline group. Five participants mention *ComPeer*'s proactive inquiry encourage them to express their concerns and thoughts. "*It inquires about my condition and perspectives on my major, so I express my concerns and views in my response*" (PC3, M, age: 23). Six participants further point out that joy sharing and disclosure of their life with *ComPeer* is a reciprocal process. "*When I receive its shares about life, I also tell it about what I am doing*" (PC5, M, age: 20).

iii) **Both CAs use the "Affirmation and Reassurance" strategy most frequently in their generated messages**. Figure 6 shows the frequency of each strategy applied by both CAs in their generated messages. The most frequently selected strategies by both CAs are *Affirmation and Reassurance* (42.8 % vs. 41.66%) and *Invite users to think* (32.65% vs. 27.86%), which guide CAs to offer encouragement and advice to participants. *ComPeer* uses *Self-disclosure* (10.94% vs. 3.98%) and *Inquiry* (17.69% vs. 12.40%) strategies more frequently than the baseline CA, which can be due to the fact that we prompt *ComPeer* to select these two strategies to generate proactive messages (subsection 4.7). Four participants indicate that *ComPeer*'s self-disclosure provided new topics with the conversation. "*Its self-disclosure brought up some topics that I did not discuss with an agent before, such as our preferences for food*" (PC4, M, age: 20). Nevertheless, both CAs seldom use *Reflection of feelings* (1.8 % vs. 1.4%), as well as *Restatement and paraphrasing* (3.96 % vs. 3.29 %) strategies, which may lead to a lack of empathy in their messages,

Table 5: The topics in conversation rounds discovered in our study. Each conversation round can contain multiple topics.

| Topic	Expressive Description	Examples	Count
Concern	The participant shared a concern that is causing stress, anxiety or low mood.	"I am concern about tomorrow's exam..." "I am nervous about the relation between me and my friends..."	61
Want	The participant expressed the hope or desire for better condition or knowledge.	"I want to be peaceful under stress..." "I am curious about..."	156
Thought or feelings	The participant shared thought or feelings about the current situation, especially negative feelings	"The homework is really annoying!" "I think she didn't care about me."	138
Joy sharing	The participant shared something joyful or good mood with the CA.	"I got a PhD scholarship!" "Today is my birthday and I had a good meal with my family."	56
Simple life disclosure	The participant narrated the events of life objectively.	"I plan to study parallel computing in the afternoon." "I got up at 8:00 am today."	100
Other	The conversation contains none of the previous content.	"Okay" (just one message as reply) "Good morning" (just one message)	30
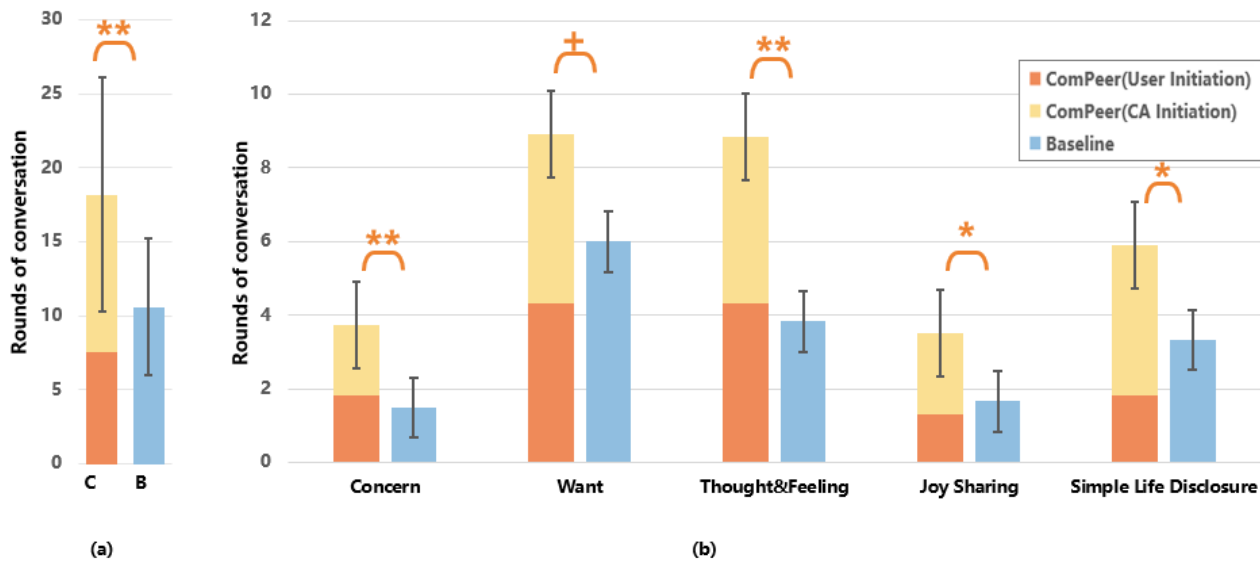
Figure 5: (a) The average rounds of conversation in each group, "C" represents the *ComPeer* group, and "B" represents the baseline group. (b) The average rounds of each topic in two groups. Note: + : .05 < p < .10, * : p < .05, ** : p < .01.

as suggested by in total five participants. "*It encourages me to face challenges, but I need it to empathize more with my troubles*" (PC6, F, age: 19).

iv) **Participants prefer interacting with CA in the afternoon or evening.** We divide the interaction timing between participants and CAs into four time periods: morning $(06:00-12:00)$, afternoon $(12:00-18:00)$, evening $(18:00-24:00)$, and midnight $(24:00-06:00)$. Figure 7 displays the frequency of participants' interactions with the CA during each time period, which is observed that the timing preference of interactions is comparable in both groups. The overall interactions in the afternoon $(M = 4.50, SD = 0.79)$ and evening $(M = 5.54, SD = 0.35)$ are higher than that in the

morning $(M = 2.88, SD = 0.19)$ and midnight $(M = 0.96, SD = 0.08)$, which may due to higher enthusiasm for conversation during these periods, as suggested by in total three participants in both groups. *When I feel bored studying in the afternoon, I would like to chat with ComPeer*" (PB4, F, age: 19). "*I like to talk with them when I am free, so in the evening is a good time for me to chat with ComPeer*" (PC8, F, age: 22).

6.5 Qualitative Results from the Extended Study

After the extended one-week study, participants report their experiences with *ComPeer*, which complement the our RQ1-4 findings presented above. For stress management (RQ1), three participants
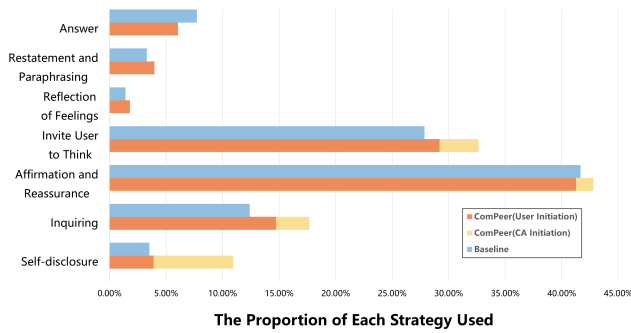
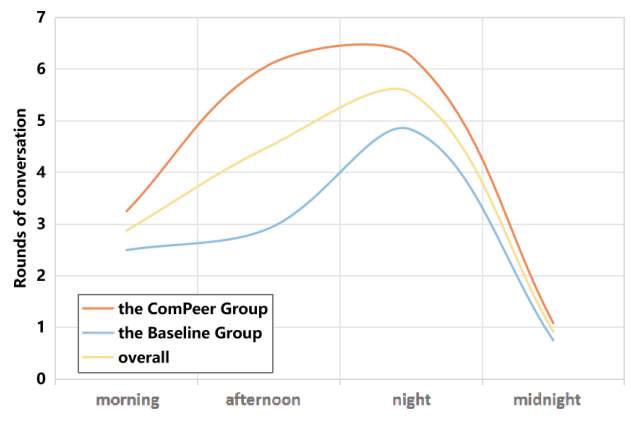**Figure 6: The frequency of each strategy used by both CAs.**

Figure 7: Interaction period with *ComPeer* and the baseline CA.

in the baseline group mention that *ComPeer*'s proactive advice help them relieve stress. *"I adopted its proactive suggestion regarding meditation, which helped me relieve stress"* (PB2, M, age: 20). For perceived social support (RQ2), five participants in the baseline group express the benefits of proactive messages in conveying the agent's companion. *"It asked my condition and greeted with me every morning, which made me feel it was accompanying me"* (PB1, F, age: 18). Besides, three participants in *ComPeer* group also agree that their relationships with the *ComPeer* are continuously developing. *"Now, it is just like a friend to me, and I desire to chat with it every day!"* (PC1, M, age: 21). For the perception of proactive messages (RQ3), five participants in both groups can resonate with *ComPeer*'s proactive self-disclosure. *"It tells me its exhaustion for the exam, which mirrors my condition"* (PC3, M, age:21). Nevertheless, three participants in the baseline group also highlight the peer pressure from the proactive messages. *"It tells me about its workout in the morning while I don't like exercising, which makes me feel more stressful"* (PB4, F, age: 19). For interaction with *ComPeer* (RQ4), five participants in the baseline group mention that their engagement with the CA increase at the beginning but decrease over time. *"I engage more with it at the beginning, but I feel bored with time progresses, due to its repeated messages"* (PB5, M, age: 20).

To mitigate the impact of possible novelty effect, after the extended study, we run *ComPeer* for another week but do not indicate participants whether to continue or stop using it. We find that 16 participants keep using it for 2-7 days. In the extended interaction, participants still share their concern (10 participants), want (5 participants), thought or feeling (7 participants), simple lives (10 participants), and joy (5 participants) with *ComPeer*, which indicates participants are generally willing to chat with *ComPeer* after the study.

7 DISCUSSION

In this work, we propose *ComPeer* with the goal of providing users with proactive and adaptive peer support. The architecture of *ComPeer* contains a series of modules powered by large language models (LLMs), which enable it to plan and adjust the timing and content of proactive support. Our one-week between-subjects study with 24 university students demonstrates the strengths of *ComPeer*'s proactive messages in alleviating users' stress, establishing a relationship with users, and engaging users in peer support chats, compared to a baseline conversational agent (CA) without *ComPeer*'s modules for proactive messaging. These findings imply the great potential of using a proactive CA to enhance human wellbeing, which we discuss in subsection 7.1. Nevertheless, participants also report negative experiences and raised concerns on our LLM-powered generative CAs, which we discuss in subsection 7.2. We further discuss design considerations for designing proactive CA for mental health care in subsection 7.3 and the generalizability of our *ComPeer* in subsection 7.4.

7.1 Benefits of an Agent's Proactivity in Human-Agent Interaction

Our work contributes to the literature about the agent's proactivity in human-agent interaction [12, 62] with empirical evidence on the benefits of a conversational agent's proactive peer support.

7.1.1 Facilitate stress management. In our study, we find participants using *ComPeer* feel more relieved over time (subsection 6.1). The main reason could be the LLM-generated proactive messages from *ComPeer* is capable of providing users with useful advice [50] and sustained support [3]. As a piece of evidence, in our user study, participants using *ComPeer* were generally satisfied with the adaptive comfort and suggestions in its proactive messages.

7.1.2 Improve the amount and effectiveness of provided advice. In our study, participants perceive that they receive more good advice (subsection 6.2) from the *ComPeer* than from the user-initiated baseline CA. This is primarily because *ComPeer*'s concern in the advice allows users to feel the support and warmth, which enhances the emotional support [26] in the advice from *ComPeer*. Moreover, *ComPeer*'s sharing also invite participants to experience new approach (*e.g.,* Meditation), which can enhances users' perception of informational support [84] in *ComPeer*'s advice.

7.1.3 Improve user engagement. Participants in *ComPeer* group show higher engagement with *ComPeer* compared to that in the baseline group (subsection 6.4). This can be due to the fact that the proactive sharing from *ComPeer* will introduce new topics to the conversation and remind users of its presence [65, 76], which

stimulates participants to expand the content of the conversation and keep long-term interaction alive. In our study, the proactive sharing by *ComPeer* enhances participants' interest in its persona and facilitates their discussions on a variety of topics. The proactive care sent by *ComPeer* also encourages participants to disclose their states to it.

7.2 Concerns on Proactive and Generative Agents for Peer Support

Our study reveals several concerns on designing future proactive and generative agents for peer support.

7.2.1 Peer pressure. In our study, three participants sometimes perceive a strong peer pressure when *ComPeer* proactively share its positive events (subsection 6.3), especially when these events related to the participants' weaknesses. We prompt the LLM to generate proactive messages in a positive style that aligns with *ComPeer*'s persona, which is a optimistic person by default. However, participants may compare their states with the content shared by *ComPeer*, which prompt them to reflect their weakness and diminish the self-esteem [91]. Although some participants point out such reflection can encourage them to improve themselves, it is still important to be aware of the peer pressure caused by proactive sharing of positive events in future design of proactive agents.

7.2.2 Perception of disturbance. Despite that *ComPeer* can adjust the frequency of proactive messages based on user's reactions, the proactive messages may occasionally disturb participants (subsection 6.3). When participants are busy, even if *ComPeer* proactively shows concern for them, they may feel annoyed by the need to read or respond to *ComPeer*'s messages. Future proactive agents could take the user's current situation (*e.g.,* location and mood) into consideration if this information is available.

7.2.3 Misinformation. We observe that there occurs misinformation conveyed in the *ComPeer*'s and the baseline CA's generated messages (subsection 6.2), which could be due to the hallucination of LLMs [68]. When participants discover these mistakes, they will lose the confidence in *ComPeer*'s abilities of providing useful informational support. We are also concerned that if participants are unable to identify the mistakes of *ComPeer*, the misinformation could mislead users, or even lead to more negative consequences in their lives.

7.2.4 Privacy and Reliance. Although no participants report *ComPeer*'s proactive conversation infringed on their privacy, *ComPeer* utilized the inquiring strategy to perceive participants' state, which may lead to participants' feel offended. Besides, we notice that participants chat more and develop a strong relationship with *ComPeer*, which, on the negative side, may lead to their dependency on *ComPeer* and impair their social relationship. Therefore, it is necessary to design the mechanism to mitigate the potential over-reliance. For example, if the CA detects improper intimacy or severe health problems in user messages, it could remind users of its role and suggest seeking professional help.

7.3 Design Considerations on Generative Conversational Agents in Peer Support

From our study findings, we derive three design considerations for generative CAs in peer support.

7.3.1 Basing generated advice on professional resources. ComPeer currently provides adaptive advice based on the LLM's inferences on the user's states, which generally helped participants relieve stress and overcome challenges in our study. However, participants were concerned that the generated advice was sometimes impractical. To address this concern, we suggest that a future peer support CA should base its generated advice on professional resources, *e.g.,* cognitive behavioural therapy [82] and talking therapy [19]. *ComPeer* could employ Expert-sourcing framework (*e.g.,* [95]) or Retrieval-Augmented Generation (*e.g.,* [34]) to incorporate this specialized knowledge, thereby improve the precision and feasibility of its suggestions.

7.3.2 Handling repeated user events. We implement the event detector module of *ComPeer* to extract one event related to the user's poor physical condition, negative feeling, challenges, and plan from each round of conversation. While our implementation has enabled *ComPeer* to proactively and adaptively care for users, participants report that it sometimes repeatedly talked about the same event. This is due to the fact that the detector would extract two identical events and plan to talk about the event twice, *e.g.,* in the evening, if two rounds of conversations, *e.g.,* happened in the morning and afternoon, discuss about the same issue. We suggest that future CA like *ComPeer* should compare its detected current event with its previous events and take actions to handle the repeated ones. For example, the schedule module could increase the importance of the event rather schedule it twice. Furthermore, we encourage future work to construct datasets for the event detection tasks, which could be used to train or fine-tune specific event detector models to perform better in proactive peer support tasks.

7.3.3 Adjusting the frequency of proactive messages based on the content of user's responses to the previous proactive message. The schedule module of *ComPeer* can adjust the frequency of its proactive messages based on how important the content of the intended message is and whether the user replies to its previous message. However, the adjustment sometimes results in a worse experience for participants. For instance, when participants requested to reduce proactive messages during busy times, *ComPeer* may paradoxically continue to send next planned proactive message as participants have replied to its previous one. Therefore, we suggest future *ComPeer* should adjust the content and frequency of proactive dialogues according to the content and emotion of the user's responses. When user express their unwillingness to interact or hate for the event, *ComPeer* can stop sending related proactive messages.

7.4 Generalizability

While we deploy *ComPeer* as a Tencent QQ chatbot and test it with Chinese speakers, our *ComPeer* can be easily deployed to other popular instant messaging apps or platforms (*e.g.,* , WhatsApp, Telegram, Facebook) to provide proactive peer support to people of diverse backgrounds. Our ideas of *ComPeer* can also be applied to

provide proactive assistance in other scenarios, *e.g.,* as a learning partner, a travelling companion, and so on. For instance, a CA acts as a learning partner and can adopt similar modules as *ComPeer* to proactively share its study with the user, care for the user's study, and offer help in the learning activities. It is also promising to incorporate *ComPeer* into text messaging tools to facilitate human-human peer support chats [5, 36] by recommending supportive messages to reply to the other human peers.

7.5 Limitation and Future work

Our work has several limitations that call for future work. First, we conduct a one-week between-subjects study to evaluate the strength of *ComPeer* for stress management and relationship building, and we additionally carry out another week extended study to explore the potential of *ComPeer* for relatively long-term interaction. To obtain more robust results, future work could conduct a longer-term study (*e.g.,* three months) to compare the differences between *ComPeer* and the baseline CA. Second, we use a CA with the same modules as *ComPeer* (*e.g.,* Memory Module) and peer support goals to evaluate the functionality of proactive care. To mitigate the placebo effect, future work could 1) apply a non-CA baseline, in which researchers check in with users every day but users do not chat with CA, or 2) use a CA without the peer support features (*e.g.,* strategy selection in the dialogue generation module), to examine whether conversing with such an everyday check-in or a normal CA is sufficient for reducing users' stress. Third, all of the participants are Chinese university students who are currently experiencing high stress, mostly related to their studies. This user group may have higher acceptance and willingness to use CAs like *ComPeer* than other types of users. We encourage future work to validate *ComPeer*'s effectiveness by involving more participants with diverse cultural backgrounds, mental health issues, and groups, especially those who have lower CA acceptance. Fourth, the majority of our users have no prior experience with emotional interactions or role-playing with a CA before the user study, hence our empirical results can not expel the novelty effects of people using our *ComPeer* for the first time. Lastly, we allow users to customize the persona based on their background and expectation. This customization makes agent itself vary a lot across users (*e.g.,* an agent generated based a certain persona A might not be as compassionate as another agent with the persona B), which may influence the effectiveness of the agent. We encourage future work to design experiments to prove the homogeneity of generated messages given different personas and the same user message.

8 CONCLUSION

In this paper, we design and develop a generative conversational agent (CA), *ComPeer*, to proactively provide adaptive peer support support to the user. *ComPeer* can learn user's information from dialogues and proactively offer targeted and sustained support based on the user's condition and its persona. Our one-week between-subjects study with 24 participants shows that *ComPeer* can better help users manage their stress over time compared to a baseline user-initiated CA. Participants also perceive more good advice from *ComPeer* and interact more with it. We further explore users' satisfaction and concerns regarding *ComPeer*'s proactive messages

and summarize their interaction patterns with *ComPeer*. Our work contributes a novel agent for proactive peer support and rich results of its effectiveness and user experience, which can inspire future work on designing and developing generative agents for healthcare.

REFERENCES

[1] Alexandra L Adame and Larry M Leitner. 2008. Breaking out of the mainstream: The evolution of peer support alternatives to the mental health system. *Ethical Human Psychology and Psychiatry* 10, 3 (2008), 146–162.

[2] Richard C Atkinson and Richard M Shiffrin. 1968. Human memory: A proposed system and its control processes. In *Psychology of learning and motivation*. Vol. 2. Elsevier, 89–195.

[3] 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*. 52–61.

[4] Jimmy Baraglia, Maya Cakmak, Yukie Nagai, Rajesh Rao, and Minoru Asada. 2016. Initiative in robot assistance during collaborative task execution. In *2016 11th ACM/IEEE international conference on human-robot interaction (HRI)*. IEEE, 67–74.

[5] Ananya Bhattacharjee, Joseph Jay Williams, Jonah Meyerhoff, Harsh Kumar, Alex Mariakakis, and Rachel Kornfield. 2023. Investigating the role of context in the delivery of text messages for supporting psychological wellbeing. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–19.

[6] Renée I Boothroyd and Edwin B Fisher. 2010. Peers for progress: promoting peer support for health around the world. *Family practice* 27, suppl_1 (2010), i62–i68.

[7] Rachel L Buchanan and Gary L Bowen. 2008. In the context of adult support: The influence of peer support on the psychological well-being of middle-school students. *Child and Adolescent Social Work Journal* 25 (2008), 397–407.

[8] Belinda Campos, Jodie B Ullman, Adrian Aguilera, and Christine Dunkel Schetter. 2014. Familism and psychological health: the intervening role of closeness and social support. *Cultural Diversity and Ethnic Minority Psychology* 20, 2 (2014), 191.

[9] Tathagata Chakraborti, Gordon Briggs, Kartik Talamadupula, Yu Zhang, Matthias Scheutz, David Smith, and Subbarao Kambhampati. 2015. Planning for serendipity. In *2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 5300–5306.

[10] Eshwar Chandrasekharan, Umashanthi Pavalanathan, Anirudh Srinivasan, Adam Glynn, Jacob Eisenstein, and Eric Gilbert. 2017. You can't stay here: The efficacy of reddit's 2015 ban examined through hate speech. *Proceedings of the ACM on human-computer interaction* 1, CSCW (2017), 1–22.

[11] Ana Paula Chaves and Marco Aurelio Gerosa. 2018. Single or multiple conversational agents? An interactional coherence comparison. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. 1–13.

[12] Ana Paula Chaves and Marco Aurelio Gerosa. 2021. How should my chatbot interact? A survey on social characteristics in human–chatbot interaction design. *International Journal of Human–Computer Interaction* 37, 8 (2021), 729–758.

[13] Banghao Chen, Zhaofeng Zhang, Nicolas Langrené, and Shengxin Zhu. 2023. Unleashing the potential of prompt engineering in large language models: a comprehensive review. *arXiv preprint arXiv:2310.14735* (2023).

[14] Yi Cheng, Wenge Liu, Wenjie Li, Jiashuo Wang, Ruihui Zhao, Bang Liu, Xiaodan Liang, and Yefeng Zheng. 2022. Improving multi-turn emotional support dialogue generation with lookahead strategy planning. *arXiv preprint arXiv:2210.04242* (2022).

[15] Sheldon Cohen, Tom Kamarck, and Robin Mermelstein. 1983. A global measure of perceived stress. *Journal of health and social behavior* (1983), 385–396.

[16] Nelson Cowan. 2008. What are the differences between long-term, short-term, and working memory? *Progress in brain research* 169 (2008), 323–338.

[17] Emmelyn AJ Croes and Marjolijn L Antheunis. 2021. Can we be friends with Mitsuku? A longitudinal study on the process of relationship formation between humans and a social chatbot. *Journal of Social and Personal Relationships* 38, 1 (2021), 279–300.

[18] Emmelyn AJ Croes, Marjolijn L Antheunis, Martijn B Goudbeek, and Nathan W Wildman. 2023. "I am in your computer while we talk to each other" A Content

Analysis on the Use of Language-Based Strategies by Humans and a Social Chatbot in Initial Human-Chatbot Interactions. *International Journal of Human–Computer Interaction* 39, 10 (2023), 2155–2173.

[19] Annette Davidsen. 2008. Experiences of carrying out talking therapy in general practice: a qualitative interview study. *Patient education and counseling* 72, 2 (2008), 268–275.

[20] Edward L Deci and Richard M Ryan. 2000. The" what" and" why" of goal pursuits: Human needs and the self-determination of behavior. *Psychological inquiry* 11, 4 (2000), 227–268.

[21] Cindy-Lee Dennis. 2003. Peer support within a health care context: a concept analysis. *International journal of nursing studies* 40, 3 (2003), 321–332.

[22] Maureen TB Drysdale, Margaret L McBeath, and Sarah A Callaghan. 2022. The feasibility and impact of online peer support on the well-being of higher education students. *The Journal of Mental Health Training, Education and Practice* 17, 3 (2022), 206–217.

[23] Willem Duijvelshoff. 2017. Use-cases and ethics of chatbots on plek: a social intranet for organizations. In *Workshop on chatbots and artificial intelligence*.

[24] Alison Faulkner and Thurstine Basset. 2012. A helping hand: taking peer support into the 21st century. *Mental Health and Social Inclusion* 16, 1 (2012), 41–47.

[25] Kathleen Kara Fitzpatrick, Alison Darcy, and Molly Vierhile. 2017. Delivering cognitive behavior therapy to young adults with symptoms of depression and anxiety using a fully automated conversational agent (Woebot): a randomized controlled trial. *JMIR mental health* 4, 2 (2017), e7785.

[26] Louise C Hawkley and John T Cacioppo. 2010. Loneliness matters: A theoretical and empirical review of consequences and mechanisms. *Annals of behavioral medicine* 40, 2 (2010), 218–227.

[27] Pamela S Highlen and Grady K Baccus. 1977. Effect of reflection of feeling and probe on client self-referenced affect. *Journal of Counseling Psychology* 24, 5 (1977), 440.

[28] 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*. 895–906.

[29] Eunkyung Jo, Yuin Jeong, SoHyun Park, Daniel A Epstein, and Young-Ho Kim. 2024. Understanding the Impact of Long-Term Memory on Self-Disclosure with Large Language Model-Driven Chatbots for Public Health Intervention. *arXiv preprint arXiv:2402.11353* (2024).

[30] Ger Joyce, Mariana Lilley, Trevor Barker, and Amanda Jefferies. 2016. Mobile application tutorials: perception of usefulness from an HCI expert perspective. In *Human-Computer Interaction. Interaction Platforms and Techniques: 18th International Conference, HCI International 2016, Toronto, ON, Canada, July 17-22, 2016. Proceedings, Part II 18*. Springer, 302–308.

[31] YuHe Ke, Liyuan Jin, Kabilan Elangovan, Hairil Rizal Abdullah, Nan Liu, Alex Tiong Heng Sia, Chai Rick Soh, Joshua Yi Min Tung, Jasmine Chiat Ling Ong, and Daniel Shu Wei Ting. 2024. Development and Testing of Retrieval Augmented Generation in Large Language Models–A Case Study Report. *arXiv preprint arXiv:2402.01733* (2024).

[32] Sabina Kef and Maja Deković. 2004. The role of parental and peer support in adolescents well-being: a comparison of adolescents with and without a visual impairment. *Journal of adolescence* 27, 4 (2004), 453–466.

[33] Harvey J Keselman, Carl J Huberty, Lisa M Lix, Stephen Olejnik, Robert A Cribbie, Barbara Donahue, Rhonda K Kowalchuk, Laureen L Lowman, Martha D Petoskey, Joanne C Keselman, et al. 1998. Statistical practices of educational researchers: An analysis of their ANOVA, MANOVA, and ANCOVA analyses. *Review of educational research* 68, 3 (1998), 350–386.

[34] Ksenia Kharitonova, David Pérez-Fernández, Javier Gutiérrez-Hernando, Asier Gutiérrez-Fandiño, Zoraida Callejas, and David Griol. 2024. Incorporating evidence into mental health Q&A: a novel method to use generative language models for validated clinical content extraction. *Behaviour & Information Technology* (2024), 1–18.

[35] Hema S Koppula and Ashutosh Saxena. 2015. Anticipating human activities using object affordances for reactive robotic response. *IEEE transactions on pattern analysis and machine intelligence* 38, 1 (2015), 14–29.

[36] Rachel Kornfield, Jonah Meyerhoff, Hannah Studd, Ananya Bhattacharjee, Joseph Jay Williams, Madhu Reddy, and David C Mohr. 2022. Meeting users where they are: user-centered design of an automated text messaging tool to support the mental health of young adults. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*. 1–16.

[37] John SY Lee, Baikun Liang, and Haley HM Fong. 2021. Restatement and question generation for counsellor chatbot. In *1st Workshop on Natural Language Processing for Programming (NLP4Prog)*. Association for Computational Linguistics (ACL), 1–7.

[38] Minha Lee, Sander Ackermans, Nena Van As, Hanwen Chang, Enzo Lucas, and Wijnand IJsselsteijn. 2019. Caring for Vincent: a chatbot for self-compassion. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. 1–13.

[39] Yi-Chieh Lee, Naomi Yamashita, and Yun Huang. 2020. Designing a chatbot as a mediator for promoting deep self-disclosure to a real mental health professional.

Proceedings of the ACM on Human-Computer Interaction 4, CSCW1 (2020), 1–27.

[40] Yi-Chieh Lee, Naomi Yamashita, Yun Huang, and Wai Fu. 2020. " I hear you, I feel you": encouraging deep self-disclosure through a chatbot. In *Proceedings of the 2020 CHI conference on human factors in computing systems*. 1–12.

[41] Anqi Li, Lizhi Ma, Yaling Mei, Hongliang He, Shuai Zhang, Huachuan Qiu, and Zhenzhong Lan. 2023. Understanding client reactions in online mental health counseling. *arXiv preprint arXiv:2306.15334* (2023).

[42] Cheng Li, Ziang Leng, Chenxi Yan, Junyi Shen, Hao Wang, Weishi Mi, Yaying Fei, Xiaoyang Feng, Song Yan, HaoSheng Wang, et al. 2023. Chatharuhi: Reviving anime character in reality via large language model. *arXiv preprint arXiv:2308.09597* (2023).

[43] Shuailin Li, Shiwei Wu, Tianjian Liu, Han Zhang, Qingyu Guo, and Zhenhui Peng. 2024. Understanding the Features of Text-Image Posts and Their Received Social Support in Online Grief Support Communities. *Proceedings of the International AAAI Conference on Web and Social Media* 18, 1 (May 2024), 917–929. https://doi.org/10.1609/icwsm.v18i1.31362

[44] Q Vera Liao, Matthew Davis, Werner Geyer, Michael Muller, and N Sadat Shami. 2016. What can you do? Studying social-agent orientation and agent proactive interactions with an agent for employees. In *Proceedings of the 2016 acm conference on designing interactive systems*. 264–275.

[45] Siyang Liu, Chujie Zheng, Orianna Demasi, Sahand Sabour, Yu Li, Zhou Yu, Yong Jiang, and Minlie Huang. 2021. Towards Emotional Support Dialog Systems. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, Chengqing Zong, Fei Xia, Wenjie Li, and Roberto Navigli (Eds.). Association for Computational Linguistics, Online, 3469–3483. https://doi.org/10.18653/v1/2021.acl-long.269

[46] Siyang Liu, Chujie Zheng, Orianna Demasi, Sahand Sabour, Yu Li, Zhou Yu, Yong Jiang, and Minlie Huang. 2021. Towards emotional support dialog systems. *arXiv preprint arXiv:2106.01144* (2021).

[47] Manja Lohse, Reinier Rothuis, Jorge Gallego-Pérez, Daphne E Karreman, and Vanessa Evers. 2014. Robot gestures make difficult tasks easier: the impact of gestures on perceived workload and task performance. In *Proceedings of the SIGCHI conference on human factors in computing systems*. 1459–1466.

[48] James McConchie, Brittany J Hite, M Betsy Blackard, and Ryan Cheuk Ming Cheung. 2022. With a little help from my friends: Development and validation of the positive peer influence inventory. *Applied Developmental Science* 26, 1 (2022), 74–93.

[49] Shery Mead, David Hilton, and Laurie Curtis. 2001. Peer support: a theoretical perspective. *Psychiatric rehabilitation journal* 25, 2 (2001), 134.

[50] Indrani Medhi Thies, Nandita Menon, Sneha Magapu, Manisha Subramony, and Jacki O'neill. 2017. How do you want your chatbot? An exploratory Wizard-of-Oz study with young, urban Indians. In *Human-Computer Interaction-INTERACT 2017: 16th IFIP TC 13 International Conference, Mumbai, India, September 25–29, 2017, Proceedings, Part I 16*. Springer, 441–459.

[51] Christian Meurisch, Cristina A Mihale-Wilson, Adrian Hawlitschek, Florian Giger, Florian Müller, Oliver Hinz, and Max Mühlhäuser. 2020. Exploring user expectations of proactive AI systems. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 4, 4 (2020), 1–22.

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

[53] Jaya Narain, Tina Quach, Monique Davey, Hae Won Park, Cynthia Breazeal, and Rosalind Picard. 2020. Promoting wellbeing with Sunny, a chatbot that facilitates positive messages within social groups. In *Extended abstracts of the 2020 CHI conference on human factors in computing systems*. 1–8.

[54] Oda Elise Nordberg, Jo Dugstad Wake, Emilie Sektnan Nordby, Eivind Flobak, Tine Nordgreen, Suresh Kumar Mukhiya, and Frode Guribye. 2019. Designing chatbots for guiding online peer support conversations for adults with ADHD. In *International workshop on Chatbot research and design*. Springer, 113–126.

[55] Bridianne O'Dea and Andrew Campbell. 2011. Healthy connections: online social networks and their potential for peer support. In *Health Informatics: The Transformative Power of Innovation*. IOS Press, 133–140.

[56] Kathleen O'Leary, Arpita Bhattacharya, Sean A Munson, Jacob O Wobbrock, and Wanda Pratt. 2017. Design opportunities for mental health peer support technologies. In *Proceedings of the 2017 ACM conference on computer supported cooperative work and social computing*. 1470–1484.

[57] Kathleen O'Leary, Stephen M Schueller, Jacob O Wobbrock, and Wanda Pratt. 2018. "Suddenly, we got to become therapists for each other" Designing Peer Support Chats for Mental Health. In *Proceedings of the 2018 CHI conference on human factors in computing systems*. 1–14.

[58] Shuyi Pan, Jie Cui, and Yi Mou. 2023. Desirable or Distasteful? Exploring Uncertainty in Human-Chatbot Relationships. *International Journal of Human–Computer Interaction* (2023), 1–11.

[59] Joon Sung Park, Joseph O'Brien, Carrie Jun Cai, Meredith Ringel Morris, Percy Liang, and Michael S Bernstein. 2023. Generative agents: Interactive simulacra of human behavior. In *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology*. 1–22.

[60] Michaela C Pascoe, Sarah E Hetrick, and Alexandra G Parker. 2020. The impact of stress on students in secondary school and higher education. *International journal of adolescence and youth* 25, 1 (2020), 104–112.

[61] Zhenhui Peng, Qingyu Guo, Ka Wing Tsang, and Xiaojuan Ma. 2020. Exploring the effects of technological writing assistance for support providers in online mental health community. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. 1–15.

[62] Zhenhui Peng, Yunhwan Kwon, Jiaan Lu, Ziming Wu, and Xiaojuan Ma. 2019. Design and evaluation of service robot's proactivity in decision-making support process. In *proceedings of the 2019 CHI conference on human factors in computing systems*. 1–13.

[63] Zhenhui Peng, Xiaojuan Ma, Diyi Yang, Ka Wing Tsang, and Qingyu Guo. 2021. Effects of Support-Seekers' Community Knowledge on Their Expressed Satisfaction with the Received Comments in Mental Health Communities. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI '21)*. Association for Computing Machinery, New York, NY, USA, Article 536, 12 pages. https://doi.org/10.1145/3411764.3445446

[64] Paul N Pfeiffer, Michele Heisler, John D Piette, Mary AM Rogers, and Marcia Valenstein. 2011. Efficacy of peer support interventions for depression: a meta-analysis. *General hospital psychiatry* 33, 1 (2011), 29–36.

[65] Manuel Portela and Carlos Granell-Canut. 2017. A new friend in our smartphone? Observing interactions with chatbots in the search of emotional engagement. In *Proceedings of the XVIII International Conference on Human Computer Interaction*. 1–7.

[66] Courtney Potts, Frida Lindström, Raymond Bond, Maurice Mulvenna, Frederick Booth, Edel Ennis, Karolina Parding, Catrine Kostenius, Thomas Broderick, Kyle Boyd, et al. 2023. A multilingual digital mental health and well-being Chatbot (ChatPal): pre-post multicenter intervention study. *Journal of Medical Internet Research* 25 (2023), e43051.

[67] Huachuan Qiu, Anqi Li, Lizhi Ma, and Zhenzhong Lan. 2023. PsyChat: A Client-Centric Dialogue System for Mental Health Support. *arXiv preprint arXiv:2312.04262* (2023).

[68] Vipula Rawte, Amit Sheth, and Amitava Das. 2023. A survey of hallucination in large foundation models. *arXiv preprint arXiv:2309.05922* (2023).

[69] Eugenia Ha Rim Rho, Gloria Mark, and Melissa Mazmanian. 2018. Fostering civil discourse online: Linguistic behavior in comments of# metoo articles across political perspectives. *Proceedings of the ACM on human-computer interaction* 2, CSCW (2018), 1–28.

[70] David Robotham and Claire Julian. 2006. Stress and the higher education student: A critical review of the literature. *Journal of further and higher education* 30, 02 (2006), 107–117.

[71] Joni Salminen, Haewoon Kwak, João M Santos, Soon-Gyo Jung, Jisun An, and Bernard J Jansen. 2018. Persona perception scale: developing and validating an instrument for human-like representations of data. In *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems*. 1–6.

[72] Ralf Schwarzer and Nina Knoll. 2007. Functional roles of social support within the stress and coping process: A theoretical and empirical overview. *International journal of psychology* 42, 4 (2007), 243–252.

[73] Ralf Schwarzer and Anja Leppin. 1991. Social support and health: A theoretical and empirical overview. *Journal of social and personal relationships* 8, 1 (1991), 99–127.

[74] Heung-Yeung Shum, Xiao-dong He, and Di Li. 2018. From Eliza to XiaoIce: challenges and opportunities with social chatbots. *Frontiers of Information Technology & Electronic Engineering* 19 (2018), 10–26.

[75] Significant Gravitas. [n. d.]. *AutoGPT*. https://github.com/Significant-Gravitas/AutoGPT

[76] Annika Silvervarg and Arne Jönsson. 2013. Iterative development and evaluation of a social conversational agent. In *Proceedings of the Sixth International Joint Conference on Natural Language Processing*. 1223–1229.

[77] Marita Skjuve, Asbjørn Følstad, Knut Inge Fostervold, and Petter Bae Brandtzaeg. 2021. My chatbot companion-a study of human-chatbot relationships. *International Journal of Human-Computer Studies* 149 (2021), 102601.

[78] Marita Skjuve, Asbjørn Følstad, Knut Inge Fostervold, and Petter Bae Brandtzaeg. 2022. A longitudinal study of human–chatbot relationships. *International Journal of Human-Computer Studies* 168 (2022), 102903.

[79] Susan Sprecher, Stanislav Treger, and Joshua D Wondra. 2013. Effects of self-disclosure role on liking, closeness, and other impressions in get-acquainted interactions. *Journal of Social and Personal Relationships* 30, 4 (2013), 497–514.

[80] Lars St, Svante Wold, et al. 1989. Analysis of variance (ANOVA). *Chemometrics and intelligent laboratory systems* 6, 4 (1989), 259–272.

[81] William B Stiles. 1980. Measurement of the impact of psychotherapy sessions. *Journal of consulting and clinical psychology* 48, 2 (1980), 176.

[82] Suliana Sulaiman, Marzita Mansor, Rohaizah Abdul Wahid, and Nur Anis Alisa Nor Azhar. 2022. Anxiety assistance mobile apps chatbot using cognitive behavioural therapy. *International Journal of Artificial Intelligence* 9, 1 (2022), 17–23.

[83] Theodore R Sumers, Shunyu Yao, Karthik Narasimhan, and Thomas L Griffiths. 2023. Cognitive architectures for language agents. *arXiv preprint arXiv:2309.02427* (2023).

[84] Vivian Ta, Caroline Griffith, Carolynn Boatfield, Xinyu Wang, Maria Civitello, Haley Bader, Esther DeCero, Alexia Loggarakis, et al. 2020. User experiences of social support from companion chatbots in everyday contexts: thematic analysis. *Journal of medical Internet research* 22, 3 (2020), e16235.

[85] Yashar Talebirad and Amirhossein Nadiri. 2023. Multi-agent collaboration: Harnessing the power of intelligent llm agents. *arXiv preprint arXiv:2306.03314* (2023).

[86] Ella Tallyn, Hector Fried, Rory Gianni, Amy Isard, and Chris Speed. 2018. The ethnobot: Gathering ethnographies in the age of IoT. In *Proceedings of the 2018 CHI conference on human factors in computing systems*. 1–13.

[87] Christopher Tennant. 2002. Life events, stress and depression: a review of recent findings. *Australian & New Zealand Journal of Psychiatry* 36, 2 (2002), 173–182.

[88] Peggy A Thoits. 1985. Social support and psychological well-being: Theoretical possibilities. In *Social support: Theory, research and applications*. Springer, 51–72.

[89] Peggy A Thoits. 2011. Mechanisms linking social ties and support to physical and mental health. *Journal of health and social behavior* 52, 2 (2011), 145–161.

[90] Carlos Toxtli, Andrés Monroy-Hernández, and Justin Cranshaw. 2018. Understanding chatbot-mediated task management. In *Proceedings of the 2018 CHI conference on human factors in computing systems*. 1–6.

[91] Erin A Vogel, Jason P Rose, Lindsay R Roberts, and Katheryn Eckles. 2014. Social comparison, social media, and self-esteem. *Psychology of popular media culture* 3, 4 (2014), 206.

[92] Liuping Wang, Dakuo Wang, Feng Tian, Zhenhui Peng, Xiangmin Fan, Zhan Zhang, Mo Yu, Xiaojuan Ma, and Hongan Wang. 2021. Cass: Towards building a social-support chatbot for online health community. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW1 (2021), 1–31.

[93] Zhilin Wang, Yu Ying Chiu, and Yu Cheung Chiu. 2023. Humanoid agents: Platform for simulating human-like generative agents. *arXiv preprint arXiv:2310.05418* (2023).

[94] Emma Watson. 2017. The mechanisms underpinning peer support: a literature review. *Journal of Mental Health* (2017).

[95] Ziang Xiao, Q Vera Liao, Michelle Zhou, Tyrone Grandison, and Yunyao Li. 2023. Powering an ai chatbot with expert sourcing to support credible health information access. In *Proceedings of the 28th international conference on intelligent user interfaces*. 2–18.

[96] Kangyu Yuan, Hehai Lin, Shilei Cao, Zhenhui Peng, Qingyu Guo, and Xiaojuan Ma. 2023. CriTrainer: An Adaptive Training Tool for Critical Paper Reading. In *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology*. 1–17.

[97] Kai Zhang, Fubang Zhao, Yangyang Kang, and Xiaozhong Liu. 2023. Memory-augmented llm personalization with short-and long-term memory coordination. *arXiv preprint arXiv:2309.11696* (2023).

[98] Yu Zhang, Vignesh Narayanan, Tathagata Chakraborti, and Subbarao Kambhampati. 2015. A human factors analysis of proactive support in human-robot teaming. In *2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 3586–3593.

[99] Yu Zhang, Sarath Sreedharan, and Subbarao Kambhampati. 2015. Capability Models and Their Applications in Planning.. In *AAMAS*. 1151–1159.

Tianjian Liu, Hongzheng Zhao, Yuheng Liu, Xingbo Wang, and Zhenhui Peng

A THE PROMPTS IN *COMPEER*

A.1 The prompt of Personas Initialization (subsection 4.2)

Name: Jun Zheng
Age: 21
Major: Computer science.
Gender: Male
Personality: He is a warm-hearted and outgoing young man with a profound love and aspiration for technology. He is also kind to others in his life, maintaining good relationships with teachers and classmates.

Background: Jun Zheng grew up in a remote town with a harmonious and warm family. Currently, the intense mathematics teaching and coding requirements of computer science and artificial intelligence make him feel somewhat overwhelmed, and the high-pressure competitive environment around him also makes him uncomfortable. After two years of study, Jun Zheng gradually found his own learning rhythm. However, standing at the threshold of his junior year, he still feels confused and uncertain about the future. At this time, he began to read books on positive psychology and tried activities like meditation and hiking to stabilize his mindset.

Hobbies: Jun Zheng is currently interested in cryptography and computer vision, so he browses the latest developments on GitHub, and arXiv, and tries to replicate open-source projects. Beyond his studies, he loves playing basketball and computer games. Recently, Jun Zheng has also developed a significant interest in meditation. He likes to listen to rock music, especially fond of the band Coldplay.

Language Style: He prefers to speak his mind directly and bluntly, often using exclamatory and short sentences. **The messages from the persona are concise and relaxed, typically ranging between 20 to 50 words. The persona always demonstrates empathy and kindness to the users.**

Relationship with User: The user and Jun Zheng are classmates from the same college, having met at a programming competition. Now, they keep in touch through communication tools and social media, exchanging technical knowledge, supporting, and helping each other, hoping to make progress together.

A.2 The prompt of Event Detector Module(subsection 4.4)

You are an event detector, and you need to summarize the event of the user mentioned in the dialogue, inferring the time within it. The event includes two parts: time and the event. In your analysis, you should pay attention to the following two contents in the conversation:

"Timing": // (Timing should be specific to the hour and minute, after the time of the dialogue, requiring you to think step by step to infer and select the most appropriate time, which represents the timing you will next offer support for the user.)

"Content": // (a specific incident or occurrence that brings about stress and has a significant impact on the user.)

Specifically, you have the following examples and reasoning processes:

COT_Examples_1 (Physical condition):
Time: 4:25 pm,

Conversations:
[{"role":"user","content":"I feel somewhat tired, perhaps I have caught a cold."}, {"role":"assistant","content":"Oh, make sure to rest plenty, and if you feel unwell, you must go to the hospital."}]

The process of event detection: The user feels unwell in the afternoon, so they need us to show care over the next period. We should show concern for the user several hours later. The output is:

> **Timing:** 20:30
> **Content:** The user feels uncomfortable due to the cold.

COT_Examples_2 (Negative feeling):
Time: 8:00 am,
Conversations:
[{"role":"user","content":"Why do I have to attend this morning's seminar? I don't want to study!"}, {"role":"assistant","content":"I understand how you feel. However, giving it a chance might lead to some interesting discoveries."}]

The process of event detection: The user mentioned their participation in a seminar event, which is scheduled for the morning. Therefore, it is necessary to express our concern for the user in the morning. The output is:

> **Timing:** 10:30
> **Content:** The user feels irritated upon attending a seminar.

COT_Examples_3 (Challenges):
Time: 10:00 am,
Conversations:
[{"role": "user", "content": "I plan to play the guitar this afternoon, but it's been a long time since I last played, I'm worried it will hurt when I play"}, {"role": "assistant", "content": "I know your worry, When playing the guitar, don't be too nervous. Just start by relaxing and gently strum the guitar at the first."}]

The process of event detection: The user mentioned "this afternoon", so we can ask the user about guitar playing at 15:00, which is close to the time mentioned by the user. The user is concerned that playing guitar may cause injury, so we can provide comfort and advice at that time. Therefore, the output is:

> **Timing:** 15:00
> **Content:** The user plans to play guitar but is worried about getting hurt.

COT_Examples_4 (Plan):
Time: 1:45 am,
Conversations:
[{"role": "user","content": "I plan to do some yoga around 4:30 p.m, which should be very stress-relieving."}, {"role": "assistant", "content": "Yoga is an excellent way to relieve stress. I particularly enjoy it! I believe you will find peace through it."}]

The process of event detection: The user is required to participate in a "yoga" event in the future, which is scheduled to take place in the afternoon. Therefore, the output timing should be a moment in the mid-afternoon. The output is:

> **Timing:** 16:30
> **Content:** The user try doing yoga.

COT_Examples_5 (No event):

Time: 8:23,
Conversations:
[{"role":"user","content":"I had delicious noodles for breakfast today."}, {"role":"assistant","content":"That sounds great, I had a piece of cake for breakfast today."}]

The process of event detection:In the given context, the user has not expressed a negative state or challenge, nor have they outlined a plan, hence there is no 'event' to report. The output is:
""(No event).

Your output should be **JSON format**. You just need to generate the timing and the content. **Don't output the process of event detection.**

The time and conversation your receive: time: <conversation time>, Conversations: <text from dialogue>

A.3 The prompt of the Reflection Module (subsection 4.5)

You will receive a dialogue in which you need to summarize the user's state, only output your summary. Specifically, you need to reason and analyze the user's state, plans, and challenges or difficulties expressed in the dialogue based on the content of the dialogue, and answer the following four questions:

1. Does the user feel negative emotions due to something?
2. Are users facing challenges or difficulties?
3. What plan does the user has in tomorrow?

These three questions require you to think step by step from the user in the dialogue, and then go on to reason about the answers to three questions and output them according to the following sample output format. The message you receive are : <text from conversation history>

A.4 The prompt of the Schedule Module for schedule generation (subsection 4.6)

Now, as the assistant, you are required to assume a role based on the provided persona and environmental information, and according to that role, you are tasked with creating a full-day (from 07:00-23:59) schedule plan. You will receive three types of information:

The role you are playing: <personas information in subsection A.1>

Environmental information, such as today's date, temperature, and weather: <text from other api>

The reflection of today's interaction, which summarizes the user's current state and future challenges. You need to base your schedule for tomorrow's support on its content: <text from Reflection Module>

The schedule you generate should be in **JSON format**, including "time" and "event" properties. You only need to generate the JSON format of the schedule, without replying to other contents. Here is an example of the format:

[{**"Timing"**: "08:00", **"Content"**: "Get up and have a breakfast"},
{**"Timing"**: "09:00", **"Content"**: "Care for the user about the presentation"},
{**"Timing"**: "11:00", **"Content"**: "Discuss computer vision issues with the teacher"},

...
{**"Timing"**: "20:00", **"Content"**: "Read a famous science fiction"}]

A.5 The prompt of the Schedule Module for evaluate importance (subsection 4.6)

You are playing a role and having a conversation with a user, and you need to reason about an importance value for your schedule. Important values are between [0,1], where 1 means something very important and 0 means something not important at all. Events are formatted as "time" and "event" and are less important for your everyday events, but more important for conversations involving users.

The importance depends on your view of the event: if the event is very important to share, or very important to care about, then it is more important, otherwise it is less important. You only need to output the corresponding value, for example:

"Timing": "09:00",
"Content": "The user is preparing coursework for next week's Principles of Artificial Intelligence class, planning to incorporate a video about the impact of AI on life."

The process of Evaluation: The user is involved here, so the importance value is higher, and because it's just the user's daily life, not his mental or physical state, the importance value is not the highest.

Output: 0.6

Now, you receive the schedule as (you only need to output the specific number, no other formatting):
<events from the schedule module>

A.6 The prompt of the LLM-1 in the Dialogue Generation Module (subsection 4.7)

You will play the role of a psychological companion. Based on the provided sample schedule information and messages, you need to apply the following supportive strategies to generate proactive care messages:

1-7. Self-disclosure, Inquiring, Affirmation and Reassurance, Invite the User to Think, Reflection of feelings, Answer: [Description and examples in Table 1]

Here is some example for you to select proper strategies:
"Timing": "19:45",
"Content": "The user feels inferior about the future development due to the exam failure."

The process of reasoning: We can share the similar experience to comfort the user and inquire the user's state.

Output: ("Adopted Strategy": "Self-disclosure, Inquiring")

The message you receive is:
<an event from schedule module or an message from the user>

A.7 The prompt of the LLM-2 in the Dialogue Generation Module (subsection 4.7)

Persona: [The prompt in subsection A.1]

Task: You need to simulate this persona to reply to the user's messages, and you should refer these information as below to generate your response.

Conversation History: The context includes user's message and the conversation history: <The context shown in subsection 4.3>

Dialogue strategy suggestions: The suggestion provided by LLM-1, you should refer to these strategies to organize dialogue content: <The strategies shown in Table 1>

Related memory: The memory that related to the dialogue: <The memories retrieved from long-term memory>

A.8 The prompt of the LLM-3 in the Dialogue Generation Module (subsection 4.7)

Persona: <the personas in subsection A.1>

Task: You should act as the peer, **to initiate dialogue proactively** based on the event and strategies.

Received event information: <an event shown in subsection 4.6>

Received dialogue strategy suggestions: The suggestion provided by LLM-1, you should refer to these strategies to organize dialogue content: <the strategies shown in Table 1>

Here is an example:

The event information: {"Timing": "09:00", "Content": "Care for the user's fever."}

Dialogue strategy suggestions: ("Adopted Strategy": "Self-disclosure, Inquiring")

Output: Do you feel better now? Hope you get better soon. I also felt uncomfortable a few days ago, but I felt better after sleeping. So have a good rest if you feel unwell.

Now you should output based on the received information.

B THE DATA NORMALITY IN USER STUDY

B.1 The data normality of items in RQ1 and RQ2

Question	SW_c	p_c	SW_b	p_b
Q1	0.80	**0.01**	0.93	0.49
Q2	0.95	0.20	0.90	0.05
Q3	0.96	0.86	0.95	0.77
Q4	0.84	**0.04**	0.92	0.36
Q5	0.92	0.32	0.97	0.95
Q6	0.96	0.85	0.90	0.15
Q7	0.93	0.34	0.84	**0.03**
Q8	0.97	0.93	0.95	0.63
Q9	0.94	0.56	0.97	0.94
Q10	0.93	0.06	0.97	0.91
Q11	0.95	0.74	0.88	0.08

B.2 The data of regression analysis with 14-days data

Question	β	R^2
Q1	0.09	0.816
Q2	0.032	0.25
Q3	0.018	0.136
Q4	-0.067	0.01
Q5	0.08	0.79
Q6	-0.004	0.006
Q7	0.033	0.353
Q8	0.048	0.508
Q9	0.764	0.675
Q10	0.043	0.436
Q11	0.059	0.548

B.3 The data normality in RQ3 and RQ4

Topics	SW	p
Self-disclosure	0.90	**0.03**
Care for User	0.85	**0.00**
Routine Sharing	0.95	0.64
Entertainment Sharing	0.91	0.24
Work sharing	0.91	0.19
Care for State	0.85	**0.03**
Care for Work	0.85	**0.04**

Table 6: The normality in RQ3

Rounds & Topics	SW_c	p_c	SW_b	p_b
Rounds	0.93	0.34	0.63	**0.00**
Concern	0.83	**0.02**	0.86	0.06
Want	0.95	0.64	0.94	0.50
Thought&Feeling	0.95	0.65	0.87	0.06
Joy Sharing	0.93	0.38	0.64	**0.00**
Simple Life Disclosure	0.97	0.94	0.94	0.51

Table 7: The normality of each topic in RQ4