© 2021 Yi-Chieh Lee

DESIGNING CONVERSATIONAL AGENTS TO PROMOTE SELF-DISCLOSURE AND BEHAVIOR CHANGE

BY

YI-CHIEH LEE

DISSERTATION

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Computer Science
in the Graduate College of the
University of Illinois Urbana-Champaign, 2021

Urbana, Illinois

Doctoral Committee:

Assistant Professor Yun Huang, Chair
Professor Karrie Karahalios
Associate Professor Hari Sundaram
Dr. Naomi Yamashita, NTT Communication Science Laboratories

ABSTRACT

Conversational agents, commonly known as chatbots, are software applications that allow people to access online services and information using their natural language. Because chatbots provide a convenient and low-cost communication channel, they are regarded as one of the most promising artificial intelligence (AI) technologies and are increasingly applied in many domains. For example, scholars and practitioners are striving to develop chatbots capable of streamlining healthcare provision and thus improving people's well-being. Some of these efforts have produced chatbots that can guide people to elicit their self-disclosure of personal experiences, thoughts, and feelings. Healthcare professionals can then use such self-disclosures to clarify their understandings of their patients' statuses.

Prior work has shown that reciprocity occurs in human-chatbot conversation. That is, when a chatbot engages in self-disclosure relating to its ostensible life history and emotional states, users are likely to disclose more when interacting with it. Nevertheless, promoting in-depth mutual self-disclosure, sustaining communication, and building trust between people and chatbots remains challenging. Additionally, most existing chatbot research focuses mainly on how to improve human-chatbot interaction through design, with relatively few studies investigating how chatbots can be used to mediate human-human interaction by transferring users' information and trust to real healthcare professionals.

To address the challenges mentioned above, this dissertation draws on human-computer interaction and AI technologies to design chatbots that integrate human support and social learning, and evaluate their impact on their users' self-disclosure and behavioral changes via mixed-methods longitudinal studies. It makes four main contributions. First, I explored effective chatbot designs that elicit people's deep self-disclosure to a chatbot over time. The results showed that chatbot self-disclosure had a reciprocal effect on promoting deeper participant self-disclosure that lasted over the study period; and chatbot self-disclosure also positively affected participants' perceived intimacy and enjoyment. Second, I provided empirical evidence of sustaining people's self-disclosure and trust of chatbots throughout different interaction periods, e.g., with and without a third party's involvement. In this case, the chatbot introduced and involved a mental health professional as the third party into the conversations. The results showed that within each group, the depth of participants' self-disclosure to the chatbot alone remained after sharing with the mental health professional. Third, I proposed to integrate human support and social learning in human-chatbot interaction to promote behavior change. In this case, I designed a chatbot with human support to guide people to practice journaling skills and conducted a three-phase study to investigate its impact. The results showed that the human-support chatbot encouraged users to follow the guidance during journaling practices and increased engagement; however, this design decreased willingness of some participants to keep practicing the learned skills. Additionally, I explored the effect of incorporating a social learning component into human-chatbot interaction. The findings showed that a social learning component could elicit users' deeper self-disclosure of thoughts and better self-reflection. However, only showing positive learning outcomes from peers seemed to interfere with some participants' perceived engagement with their chatbot. Overall, these findings provided new insights into the design of human-chatbot interaction for promoting users' self-disclosure and delivering guidance for behavior change.

ACKNOWLEDGMENTS

First and foremost, my beloved aunt, Mei-Chen Chao, was the person who had encouraged me to study abroad and taught me English since I was a kid. She treated me like her own son. Without her inspiration, I would not even be here to pursue my Ph.D. degree. Although she passed away in 2019, I believe she is in heaven and always blesses my success.

I want to express my sincere gratitude to my research advisor, Professor (Prof.) Yun Huang, who provides invaluable guidance throughout this dissertation research. Her dynamism, sincerity, and motivation have deeply inspired me. It was a great privilege and honor to work and study under her guidance.

I am extremely grateful to my fantastic Ph.D. committee for the guidance and support they have offered me. I was Prof. Karrie Karahalios's Teaching Assistant in 2019, and I acquired valuable experience during this period. Besides, her suggestions improved the quality of my dissertation work and provided me with diverse perspectives. I am pleased to have Prof. Hari Sundaram as a member of my qualifying exam and Ph.D. committee. His thoughtful suggestions and guidance always inspired me to reflect on my research clearly and critically. Special thanks to Dr. Naomi Yamashita. I would like to thank her for her guidance and timely help to continue my dissertation research when I did my internship with her.

I want to say thanks to my friends and research colleagues, Yu-Chun (Grace) Yen, Chi-Hsien (Eric) Yen, Dennis Wang, John Lee, Helen Wauck, Sanorita Dey, Ziang Xiao, Po-Tsung Chiu, Mingkun Gao, Hyojin Do, Wayne Wu, Sneha Krishna Kumaran, Kristen Vaccaro, Yun Huang, Yen Lu, Chun-Lin Chan, Lawrence Tseng, YZ Ou, Mindy Chang, Nini Hsieh, Vicky Liu, ShihJui Yang, Dan MacCannell, and Vera Liao. I would like to thank them for the friendship, mentorship, accompany, and great sense of humor. They are one of the most important parts of my life in the U.S.

I am extending my heartfelt thanks to the SALT (Social Computing Systems) lab, Prof. Yang Wang, Si Chen, Haocong Cheng, Ke-Rou Wang, Qingxiao Zheng, Jason Situ, Abhinav Choudhry, Abhi Thosar, Yiliu Tang, Yiren Liu, and Zhixuan Zhou. Although I did not have a chance to meet them all in person because of COVID-19, they offered me incredible help and suggestions remotely, which improve my dissertation a lot.

I express my special thanks to Prof. Alex Kirlik and Prof. Wai-Tat Fu for their genuine support throughout my Ph.D. journey at the University of Illinois Urbana-Champaign. My thanks also go to all people who have supported me to complete this research work directly and indirectly.

Finally, I am incredibly grateful to my parents for their love, caring, and sacrifices for educating and preparing me for my future. They have continued to support me in chasing my dream. Also, I express my thanks to my brother for his support and valuable prayers. I am so fortunate to be born in this warm family with their unconditional love.

I faced many obstacles in this Ph.D. journey and completed this dissertation research during the pandemic of COVID-19. A quote from Frederick Douglass always drives me to move forward; thus, finally I would like to share it with the readers (you):

If there is no struggle, there is no progress. - F. Douglass

TABLE OF CONTENTS

CHAPTER 1: INTRODUCTION

Conversational agents, commonly known as chatbots, are software applications that allow their users to access online services and information using their natural language, either in text or voice form [1]. Chatbots are widely used by news websites, customer-service departments, and question-answering (Q/A) systems. For example, the World Health Organisation (WHO) chose to deploy a Q/A chatbot to answer people's common questions about the COVID-19 pandemic, on the grounds that it could be made available 24 hours a day. Chatbots have also demonstrated their potential to improve people's mental well-being by eliciting self-disclosures [2, 3, 4]. Specifically, research has shown that people tend to disclose symptoms of depression more truthfully when interacting with a chatbot than when talking to a human interviewer. Lucas et al. [3] found that chatbots' anonymity encouraged self-disclosure, and Ravichander et al. [4] found that reciprocity also played a role: i.e., ostensible self-disclosure by chatbots encouraged people to follow suit.

Most current chatbot research focuses on dyadic human-chatbot interactions [5]. One recent study demonstrated that chatbots could moderate community members' discussions by facilitating unbiased responses and encouraging equal contributions to decision-making [6]. However, little if any other research has taken the same avenue. Thus, little is known about how a chatbot could mediate human-human communication in general, or people's self-disclosure to third parties in particular. Accordingly, against the backdrop of chatbots' rising popularity, this dissertation's main focus is the potential for utilizing them to mediate human-to-human interchanges of information and trust. In particular, if chatbots are to be deployed effectively in mental healthcare, it will be crucial to understand whether people exhibit different self-disclosure behaviors with a chatbot alone vs. when communicating with a human healthcare professional via a chatbot intermediary, as well as whether the chatbot in such a situation can convince its users to follow professional guidance. Intuitively, these two situations involve very different social dynamics, but relatively little research on the subject has thus far been conducted, even though understanding how and how much people will self-disclose to domain experts through AI is critical to designing human-in-the-loop AI (HIT-AI) systems [7].

Recent studies have demonstrated that chatbots can help their users maintain healthy lifestyles [8, 9] and guide them to improve their general well-being [10, 11, 12]. A variety of conversational strategies and structures have been utilized to promote positive behavioral change among chatbot users [8, 13, 14, 15], and in some scenarios these have been found to outperform human-human interaction. Xu et al. [16] even concluded that the use of interactive robot agents would probably enhance physical-therapy outcomes. Two recent studies likewise reported that people often trust conversational agents more than humans with their personal information, because they think of the former as both more objective and more secure [17, 18]. This body of prior research has demonstrated that chatbots could serve as platforms for collecting personal information and delivering guidance.

There are still a number of challenges to overcome, however. For example, research has shown that people easily become disengaged from using a chatbot [19, 20], hampering the success of long-term chatbot-based interventions. People might also place too much trust in solutions proposed to them by chatbots, which could in fact be inappropriate [16, 21, 22]; while Luria et al. [23] found that people felt uncomfortable interacting with a chatbot that used a single persona when dealing with both low-risk situations such as social chat and high-risk ones like medical emergencies.

This dissertation reports on my design, implementation and evaluation of various chatbots intended to serve as facilitators of people's self-disclosure to a professional third party. In addition to seeking a clearer understanding of how people disclose sensitive information to third parties through a chatbot, I tested chatbot conversational styles with three levels of ostensible self-disclosure (i.e., high, low, or none), in terms of how effectively they could solicit deep self-disclosure from people after comments from a professional third party are introduced into their interaction. The results of my four-week study of chatbot users' self-disclosure and other aspects of their user experience are presented in Chapters 3 and 4 [24, 25].

Then, I propose two new chatbot designs for teaching journaling skills, which require deep self-disclosure. The first of these two designs, presented in Chapter 5 [26], integrates human-expert support, and is compared against a baseline chatbot that guides the participants itself. That chapter also reports the results of four-week study I conducted to understand how theses design differences impacted users' responses to and perceptions of the chatbot. The second chatbot design, presented Chapter 6, integrates a social-learning component that allows users to see peers' comments on their journaling skills. The same chapter also reports on a longitudinal study I conducted to investigate how social learning impacted users' willingness to follow the chatbot system's journaling guidance.

Finally, in the hope that this dissertation's findings will inspire future chatbot design for improving online healthcare systems, Chapter 7 highlights the research findings and ethical considerations raised by the results of Chapters 3 - 6, and recommend further changes that will boost the advantages and mitigate the disadvantages of using chatbots to deliver guidance and promote users' self-disclosure.

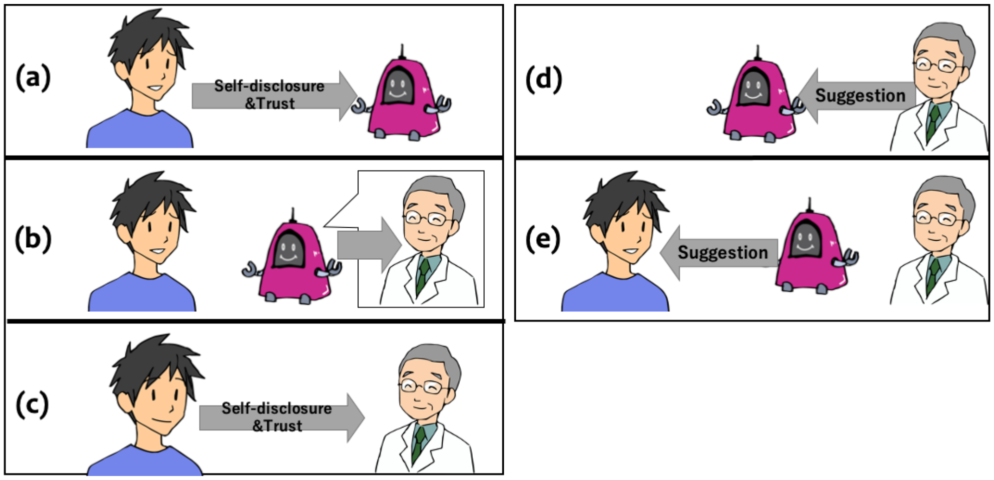


Figure 1.1: (a)(b)(c) show our goal to design a chatbot to transfer trust and promote self-disclosure to a Mental Health Professional (human supporter). The users never meet with the human supporter, and our study aims at understanding how to use a conversational agent to transfer a user's trust to the human supporter. The findings are presented in Chapter 3 (Phase 1) [24] and Chapter 4 (Phase 2) [25]. (d)(e) show our proposed design in the Phase 3 (Chapter 5 [26] & Chapter 6) which is aiming to explore effective designs to utilize chatbot delivering suggestions with regarding chatbot-user relationship.

1.1 MOTIVATION

Figure 1.1 illustrates this dissertation's research concept, and helps to explain its origins in a particular mental-healthcare context, which required clients to develop trust in counselors, to whom they were expected to make deep self-disclosures, and who in turn were expected to understand their clients and offer them appropriate treatment options. However, the deep self-disclosure on which this entire process hinged is far from easy, and counselors usually need to make concerted efforts over a period of several months to develop bonds and sufficient levels of trust with their clients [27, 28]. The inception of this dissertation was therefore the idea that HIT-AI – which supplements, rather than replaces, human experts' role in treatment – could speed up this trust-building process, and perhaps in other ways usefully mediate counselors' interventions aimed at promoting behavioral change.

To test these ideas, I separated my project into three main research phases. The first, as depicted in Figure 1.1(a), explores how a user builds trust and a relationship with a chatbot, and gradually comes to disclose personal information to it [24]. For phase 2 (Figure 1.1(b-c)), I designed a chatbot that introduced a third-party expert into human-chatbot interaction, and investigated whether users would readily transfer their trust and personal information to that expert. And phase 3 (Figure 1.1(d-e)) focused on how a chatbot could integrate expert advice and social support into human-chatbot interaction to deliver guidance. Taken together, the findings of these three phases demonstrate the potential advantages and disadvantages of incorporating chatbot-based HIT-AI into a healthcare system.

1.2 CONTRIBUTIONS

The follows are the main contributions of this dissertation:

- **Chatbot self-disclosure had a reciprocal effect on promoting deeper participant self-disclosure over time, and it also had a positive impact on improving user' perceived intimacy and engagement -** I explored how varying levels of a chatbot's self-disclosure influence the depth of people's self-disclosure in the phase 1 study. The research findings contribute new understandings of how time plays a role in chatbots and people's self-disclosure interactions. The results also provide further implications into designing and using chatbots where deep self-disclosure is needed.

- **The depth of users' self-disclosure to the chatbot remained after sharing with a third party (human expert), and users' trust in the chatbot could be transferred to the human expert -** My work provides empirical evidence that people sustain their self-disclosure to a mental health professional (expert) given one chatbot design. By conducting experimental research with using three different chatbots, I contribute new understandings of how a self-disclosing chatbot (reciprocity) promotes people's deep self-disclosure to an expert through daily journaling on *non-sensitive* and *sensitive* topics. The findings shed light on future AI research by bringing a unique insight that trust may be transferable from human-AI to human-human through AIs.

- **Integrating human support into human-chatbot interaction to deliver guidance could encourage users to follow the guidance more faithfully and perceived a high level of engagement and trust with the chatbot system -** This dissertation is among the first that investigated the effects of integrating a human expert into human-chatbot interaction to deliver guidance for practicing journaling skills. The unique three-phase design of an study contributes novel findings of how chatbot interactions with and without expert guidance elicited user interaction differently over time. Participants' actual and perceived engagement with the chatbot providing expert guidance was significantly higher than that of the participants who interacted with their chatbot alone. I triangulated system log analysis with interviews and surveys to provide new insights into how the design of chatbot systems with and without human support affected the user experience of such systems.

- **Integrating the social learning component could facilitate users' in-depth self-disclosure of their thoughts and better self-reflection; however, the feature might also interfere with users' engagement with their chatbot -** This work explores the effects of incorporating social learning into the context of human-chatbot interaction; in particular, the chatbot provides journaling guidance and shares peers' experiences with users through a private communication channel between each user and their chatbot. The results show that social learning mediated by the chatbot significantly improved people's journaling skills and encouraged them to disclose their deeper thoughts, and resulted in better self-reflection. Meanwhile, the social learning component interfered with some participants' perceived engagement with the chatbot. These findings provide new insights into how social learning can be embedded in future human-chatbot interactions that deliver effective guidance for behavior change.

CHAPTER 2: RELATED WORK

This chapter reviews the evolution of chatbot technology and social penetration theory to lay the studies' foundation, which I present in the following chapters. I then review the literature about using technologies to promote self-disclosure and how users' self-disclosure is studied in healthcare. These prior studies guide my work in Chapter 3 and Chapter 4, designing chatbots to facilitate users' self-disclosure and trust transfer. Finally, I visit the literature about the concept of human support in behavioral intervention technologies and social learning theory, and these prior studies inspired my work and chatbot designs presented in Chapter 5 and Chapter 6.

2.1 CONVERSATIONAL AGENTS (CHATBOTS)

One of the first and well-known and chatbot systems was ELIZA [1, 29]. It was developed through an early natural language processing program, and it demonstrated how human-computer communications at that time. It mainly used pattern matching technology to recognize clue words or phrases in the users' input and offer pre-programmed responses. Due to machine learning advancement and rich conversational data sources, chatbots have become more "intelligent" [30]. These technologies allow a chatbot system to recognize users' intentions from more accurate and generated corresponding answers without providing pre-programmed responses [31]. Several commercialized examples (e.g., Amazon Alexa, Google Assistant, and Apple Siri) have demonstrated the opportunities of deploying chatbots (smart speakers) to people's daily life, and existing task-focused chatbots (e.g., buy products and operate devices) seem promising to replace transitional user interfaces.

However, there are still many challenges to overcome. A machine does not really "understand" a human's thoughts and feelings; thus, exploring how to fuse human partners instead of entirely replacing humans in the user-chatbot interaction may be a solution to assist the chatbot in understanding the requirements for interacting with people [30]. Therefore, I explore this unique topic in this dissertation to explore how a third party (human and social support) could be integrated into the interaction to facilitate the users' interaction with the chatbot system.

2.1.1 Chatbot Applications in Improving Wellbeing

Chatbots have been broadly used in different areas [10, 32, 33, 34].

They can not only help people complete various tasks [35] but can also improve mental well-being (e.g., self-compassion [11]). For example, chatbots are utilized in the workplace to assist team collaboration [35], to improve workers' quality of life and work productivity [10], and to reduce caregivers' workloads [36]. Park et al. [37] adopted Motivational Interview in the chatbot conversation to help users cope with stress and found that their design can facilitate a conversation for stress management and self-reflection. Lee et al.[11] designed a dialog to make the users take care of a chatbot's negative experience. After a two-week interaction with the chatbot, the user's self-compassion significantly increased.

These studies have demonstrated the potential benefits of using a chatbot for different purposes, and my research aims at understanding how to use a chatbot to mediate sensitive information. The Computers Are Social Actors **(CASA)** paradigm indicated that people might apply social norms of human relationships when interacting with computer agents [38]. Thus, research has focused on advancing technological contributions to make computer agents naturally chat and understand people; therefore, some studies examined different strategies [31, 39] to enhance users' experience when talking with a chatbot. For example, Hu et al. found that the tone-aware chatbot could be perceived as more empathetic than a human agent [31]. Moreover, Skjuve et al. [40] indicated the importance of studying the development of human-chatbot relationships, which lacks current knowledge. The perceived companionship also plays an essential role in utilizing chatbots in mental healthcare. It is a key for a chatbot system to be accepted and capable of meeting users' needs [41, 42].

2.2 SELF-DISCLOSURE

Self-disclosure - the gradual unveiling of personal information, thoughts, feelings, goals, and even failures–is key to individuals' formation of interpersonal relationships and achievement of intimacy [43, 44]. A leading explanation of the self-disclosure process is social penetration theory (SPT) [44], which categorizes four stages of self-disclosure, i.e., orientation, exploratory, affect-exchange, and stable-exchange. Together, these stages delineate a journey from the disclosure of shallow and general to deep and intimate information. Self-disclosure is often evaluated from two dimensions, breadth and depth [45, 46]. The breadth of self-disclosure can be demonstrated with a wide range of topics disclosed; on the other hand, the depth is more involved with personal experiences, intimate relationships, and possible negative feelings as a result of life difficulties. Prior works on chatbots often highlighted the volume of self-disclosure in terms of its breadth (e.g., [3, 4, 34]); less was discussed on its depth. In order to assess mental well-being, a high depth of disclosure (deep self-disclosure) is needed [47]. To elicit self-disclosure at a deeper level, a higher level of trust is often associated in the relationship [48].

Self-disclosure plays an important role in a wide range of settings, including mental well-being [49], customer service[50], and employment [3, 51]; thus extensive research has been conducted on self-disclosure's relationships to various constructs including trust [48], intimacy [52], gender [53], and personality [52]. A considerable body of prior research has identified self-disclosure as a potential path to mental wellness, and its benefits during psychotherapy are also well attested [54]. The Substance Abuse and Mental Health Services Administration (SAMHSA) [1] reported that people who disclosed their mental illnesses felt relief and experienced improved relationships with friends and family members.

Although mutual self-disclosure can, with time, facilitate intimacy, trust, and depth of self-disclosure by both parties [55], whether and how a psychiatrist should self-disclose to clients is the subject of ongoing debate [56, 57]. Some studies have raised concerns that too much closeness with clients might derail their progress [58]. However, others have suggested that therapists' carefully selected self-disclosures could be beneficial as a means of building rapport with clients [57] and of building certain skills that can strengthen the counseling relationship, such as active listening [59], gradually building trust [55], and matching communication styles [60]. On the other hand, lack of trust in online applications may lead to inaccurate information being collected and deterred efficacy of services provided by the applications [61]. What if the psychiatrist (human expert) is not involved in the conversation directly, and instead, the psychiatrist only receives self-disclosure content from people through a chatbot? This is the context of our study.

Disclosing personal mental health information is not easy for most people, and this is also one of the major practical difficulties in counseling sessions [55]. People naturally avoid revealing their vulnerabilities to others; this tendency is even more prevalent among those with mental illnesses, because those people who seek mental health care worry about social stigma and discrimination related to mental health problems. Previous studies have found that when people were interviewed face-to-face by a human interviewer, they may tend to disclose fewer symptoms of depression than when interviewed by a virtual agent[3]. It is not clear how people disclose when facing a different conversational agent design. For example, Clark et al. [62] found that there may be a fundamental barrier to developing relationships with conversational agents because people value different aspects in conversation with agents - some people may treat a chatbot as a tool, but users with mental health issues or social difficulties may benefit from social capabilities in a chatbot system. Prior studies [33, 34] have shown the positive effect of deploying a chatbot to facilitate journaling[2] and for helping

[1]https://www.samhsa.gov/

[2]In the sphere of mental healthcare, journaling is a common practice of self-tracking that has been proven

people to realize their mental issues and relieve their symptoms.

2.3 TECHNOLOGIES PROMOTING SELF-DISCLOSURE

Computer-mediated technologies have significantly promoted people's self-disclosure behavior. For example, people disclose their personal information, feelings, and thoughts on social media [65]. Ma et al. [66] found that anonymity played an important role in people's willingness to engage in such sharing. Studies have revealed that virtual agents can provide non-verbal as well as verbal cues to engage users, e.g., during interviews, which can render them more willing to self-disclose [3, 67].

2.3.1 Self-Disclosure through Social Media

Self-disclosure behavior on social network sites has gained the attention of HCI scholars. For example, people freely disclose stress, depression, and anxiety through online social media platforms [3, 68, 69, 70]. It was found that such anonymous self-disclosure with their peers could help users maintain their mental well-being, as they may receive social support from their peers [71]. Similarly, Yang et al. [72] investigated the self-disclosure behaviors of online health support communities, and the study found the members' self-disclosure in private and public channels affected how they reciprocated with other and reached out for social support. Although self-disclosure on social media could help each other seek social support, people naturally avoid revealing their vulnerabilities to others [73], as it might also cause social risks [74, 75]. Thus, Andalibi et al. [74] explored how people used throwaway accounts on Reddit to disclose their stigmatized issues (e.g., sexual abuse) and found that people using anonymous means engaged more in seeking support.

2.3.2 Promoting Self-Disclosure to a Chatbot

Recently, chatbots have been used to guide users to healthier lifestyle choices and improve their mental well-being [10, 11, 76] and engage people in truthful interactions [4]. For example, Moon [50] examined how various wordings of questions influenced participants' responses, and found that when the questions were preceded by the automated interviewer's self-disclosure, the participants exchanged more intimate information with it. Ravichander et al.'s [4] chatbot provided conversationally relevant self-disclosures from a large conversation dataset in real time, such that it engaged users with reciprocity in social conversations.

effective in terms of boosting mood and reducing anxiety [63, 64].

According to the norm of reciprocity, when someone discloses something deeply personal, his or her interlocutor feels pressure to share a similar level of information [77], therefore, therapists often disclosed themselves to encourage patients' self-disclosure [78, 79]. In a recent work, reciprocity was found to happen in human-chatbot interaction as well, e.g., a self-disclosing chatbot received more self-disclosure from users [4]. Lee et al. [80] indicated that small-talk increased users' trust in the robot, and found that a user's greeting with the robot could predict the user's conversational strategies such as sociable interaction and self-disclosure. Therefore, these studies demonstrated that a reciprocal social conversation may increase people's trust in a computer agent. This concept also inspired me to design different levels of chatbot self-disclosure in small talk to elicit users' in-depth self-disclosure.

People's self-disclosure to chatbots can be used to detect symptoms, identify possible causes, and recommend actions to improve their symptoms by promoting people's self-disclosing, as well as to encourage interviewees to disclose themselves more openly in an interview session [3]. Scholars compared web surveys against chatbots and found that respondents tended to provide more high-quality data when using the latter [81]. Fitzpatrick et al. utilized a therapy chatbot "Woebot" in their study to explore its feasibility to help reveal people's mental illness; their results showed that the chatbot helped relieve symptoms of anxiety and depression [34]. Additionally, chatbots can be deployed to various platforms using both speech and text; chatbots provide cost-effective [32] solutions for self-disclosure [2, 3, 4] or deliver education materials for self assessment (e.g., alcohol risks [82]).

Although scholars have made significant progress with self-disclosure research using chatbots, major research questions, such as if and how chatbots can promote deep self-disclosure over time, are still under-studied. Promising application domains, e.g., mental health [49, 83], often need support tools to acquire users' sustained self-disclosure of sensitive topics over a period of time, and they also need the users to share their collected information to a professional third party (e.g., clinicians and doctors) to offer proper treatment. However, how people interacted with the chatbot system differently when they self-disclosed sensitive topics to mental health professionals thought it has not been investigated. Therefore, this dissertation fills the void by focusing on evaluating self-disclosure depth by running an experimental study and by comparing participant's daily journaling and answers to sensitive questions before and after sharing with a professional third party via the chatbot. In Chapter 4, I take a step toward addressing this research gap.

2.4 CHATBOTS FOR DELIVERING GUIDANCE

Through interacting with the chatbot, people can search useful resources, i.e., self-help information, before reaching out for face-to-face counselling [32, 33]. Therefore, chatbots have become popular in response to the demand of mental health care in modern society [61, 84, 85]. A recent work shows that chatbots can play a role to inquiry users answering questions and convincing them to share diet information with their family members so as to support each other [9]. Also, coaching apps have been developed, not only for boosting users' awareness of their own mental well-being, but also for helping mental-health professionals gain more knowledge about their clients [12].

In addition, many studies designed chatbots to guide healthier habits or ways of thinking [37], such as better eating habits [9], exercise [8], ways of coping with stress [37], and self-compassion [11]. For example, Park et al. [37] incorporated a motivational interview technique into chatbot conversation to help users cope with stress, and found that their design facilitated conversations that improved self-reflection as well as stress management. Lee et al. [11] designed a dialogue aimed at inspiring users to take care of a chatbot that was portrayed as having had a negative experience, and found that after doing this for two weeks, users' self-compassion increased significantly. Another line of research has shown that chatbots have the potential to help people improve their mental well-being by training their thoughts and behavior [2, 3, 86]. For instance, Wang et al [86] designed a public-speaking tutor using a chatbot system to coach users and reduce their public speaking anxiety. Hence, these studies have shown that chatbots could not only help track users' behavior but could also play a proactive role in training users to learn skills.

Recent advancements in artificial intelligence are enabling chatbots and other virtual agents to act more credibly like human beings, including during the provision of self-help information [32, 33]. Prior studies [87, 88] indicated that conversational interaction can increase trust and affect users' acceptance of recommendations from a conversational agent. Thus, the design of the interaction between them is important in enhancing users' willingness to adopt chatbot suggestions. Gabrielli et al. [89] proposed a chatbot-based coaching intervention that successfully helped adolescents learn life skills, such as strategies for coping with bullying, and previous research [16, 22] found that their participants' trust and compliance with physical therapeutic suggestions were both higher when interacting with robot therapy partners than with a human expert.

Moreover, research has shown that people tend to apply the social norms of human relationships to their interactions with computer agents. This tendency [38] has informed the design of many computer agents [4, 14, 24]. People may perceive intimacy and companionship with a computer agent [13, 24, 90], inducing changes in behavior change. For example, Ravichander et al. [4] found that reciprocity occurred in human-chatbot interactions and that a chatbot's self-disclosure encouraged people's self-disclosure. Similarly, recent work by Lee et al. [24] showed that a chatbot's self-disclosure improved participants' perceived intimacy with the chatbot and facilitated their self-disclosures in response to the chatbot's sensitive questions.

However, several limitations of chatbot-based approaches remain, and in certain situations, chatbot-based approaches may be less beneficial than those provided by humans [20, 90, 91]. For example, Howard et al. [22] has pointed out that some people may trust robots too much, due to over-optimism about the viability of the solutions they suggest, and that this trust becomes a source of risk if robots make clinically suboptimal or inappropriate suggestions. In addition, for healthcare interventions that require long-term engagement [92], people may easily become disengaged from the use of self-guided systems, due to loss of motivation and/or failure to incorporate those systems' recommendations into their daily lives [20]. Furthermore, an investment model shows that purely computer-based interventions are often much less effective than hybrid ones with some professional human input [93], in part because the latter tends to inspire their users to execute a higher proportion of their intervention requests.

2.5 INTEGRATION OF HUMAN (EXPERT) SUPPORT AND CHATBOT-BASED APPROACHES

In prior works, human support has been provided via a separate communication channel external to the chatbot system, such as phone calls, text messages, and email [94]. For example, there have been two main ways of providing human support to chatbot systems. The first is to deploy chatbots in between human-run sessions, to offer users unbroken access to materials and activities [84, 95, 96]. Studies that have adopted such an approach regard chatbots as supplementary tools to support human expert's intervention; chiefly, by monitoring clients outside of their clinical sessions, and garnering information about them that may result in better treatment (e.g., [12, 20]). Alternatively, it is possible to design a primarily chatbot-based intervention, augmented by human supporters who promote engagement and provide technical troubleshooting and clinical support when issues arise [20, 91, 97]. Such an approach could be more efficacious than interventions by chatbots unsupported by humans [20, 95].

Recently, some researchers have suggested an integration of human support into chatbot interventions [20, 95, 96]. For example, Schueller et al. [95, 96] reviewed prior studies of integrating human experts (e.g., coach and therapist) into behavioral intervention technologies, not chatbot-based, and suggested concepts to guide a deeper integration by capturing the trade-offs between client benefits and the available human resources. Alternatively, some prior studies [98, 99] use conversational agents to encourage users' collaboration and communication between people. Specifically, Kumar et al. [98] designed a chatbot tutoring system which gave guidance for multiple students to facilitate collaborative learning among them. Duan et al. [99] utilized a conversational agent to enhance non-native speakers' confidence in conversation with native speakers. These studies showed that chatbots could help mediate interactions between users, but I further explore the effect of applying chatbots to mediate suggestions for guiding users to learn skills.

The foregoing review and Schueller et al.'s work [95, 96] calls for an integration of the support provided by chatbots and humans. However, previous works have indicated different designs for implementing multiple personas into a chatbot system. For example, Luria et al. [23] conducted studies to examine multiple personalities for conversational agents under different contexts (e.g., low-risk and high-risk contexts). They found that users preferred to have an additional expert agent guiding a specific complex task instead of interacting with the same agent that handled both simple and complex tasks. Conversely, Chaves et al. [100] found that users reported confusion when they engaged in multiple persona chatbots for an information gathering task in a single communication channel. Therefore, it is not clear if and how adding a human supporter into human-chatbot interactions could impact user experience and outcomes, which motivates our research. In Chapter 5, I take a step toward addressing this gap.

2.6 SOCIAL LEARNING IN TECHNOLOGY-BASED INTERVENTIONS

Social learning theory argues that people can learn new behaviors and concepts by observing others' behavior and attitudes [101]. This observation can be obtained through reading, listening, or watching others' actions [102]. In general, the elements of social learning theory include *Attention*: observing others, *Retention*: internalizing observed information, *Reproduction*: reproducing observed behaviors or information, and *Motivation*: learning the consequences (e.g., advantages and drawbacks) of the observed behaviors to decide future behaviors [101]. Previous studies have shown that social learning may enhance users' learning experience such as affection, perceived engagement, and sense of companionship within an online context [103, 104]. Consequently, this theory has been widely used to promote people's healthy behaviors through online communities and social technologies [105, 106, 107].

There are various design strategies to provide social learning [105, 108]. For example, Franklin et al. [105] proposed a system that allowed users to send health-related tips and personal experiences to others anonymously, alongside the system's own suggestions and tips to facilitate health-promoting behaviors. In addition, Grimes et al. [107] deployed a system whose users were required to observe their peers' success stories about healthy eating and found that sharing these stories facilitated a sense of community. Prior studies [105, 109] have generally found that people would adopt healthy behaviors if they observed and tried to replicate ideal actions shared by peers who had succeeded in such adoption. Also, people provided with positive social learning are more likely to be engaged in self-care behavior than those who do not receive such information [110, 111].

However, integrating social interaction into an intervention might also have some negative effects, e.g., peers' support could be absent or misjudged [108, 112, 113]. For example, Li et al. [113] warned that social learning can lead to negative outcomes for people with depression by viewing negative content and lack of support in online mental-health communities. The risk of social learning can partly be explained by social comparison theory [114]. According to the theory, people generally compare themselves to others when they share the same goals or in a similar situation. While this may motivate some people to perform better, such comparison can also promote judgmental and biased attitudes toward others' experiences [112, 115].

In the context of human-chatbot interaction, recent studies have mainly focused on exploring chatbot designs to facilitate social interaction between users. In these studies, chatbots motivate users to communicate with each other to improve interpersonal relationship and receive better support e.g., [5, 9, 116]. For instance, Lukoff et al. designed a chatbot [9] to facilitate family members to do food journaling about cooking or shopping for healthier food for family members. Moreover, Nordberg et al. [116] proposed a chatbot that guides people to have a conversation with a group of people suffering from similar difficulty (e.g., patient groups) in order to have richer social interactions and facilitate online self-help programs. Furthermore, Seering et al. [5] suggested that chatbots could help grow online communities by facilitating communication between community members.

Previous chatbot studies that integrate social interaction into chatbot intervention have almost exclusively focused on chatbots' potential for mediating interaction among users. It is not clear yet if incorporating social learning into one-to-one human-chatbot interactions could promote user engagement and learning experience when they receive guidance from a chatbot. In Chapter 6, I designed a chatbot with social learning component to address this research gap.

CHAPTER 3: ENCOURAGING USERS' SELF-DISCLOSURE WITH CHATBOTS

3.1 INTRODUCTION

Self-disclosure is a process in which a person reveals personal or sensitive information to others [43, 44] and is crucial for developing a strong interpersonal relationship [44]. The advancement of computing technologies has enabled new ways for people to self-disclose [3, 117]. The value and importance of self-disclosure through these technologies have been widely manifested. For example, people's self-disclosure on social media helps them release their stress, depression, and anxiety through these technologies [68, 118]. Interviewees may disclose themselves more openly in an interview session when using virtual agents [3, 51]. The challenge is that people naturally avoid revealing their vulnerabilities to others [119, 120].

Chatbots have great potential to create breakthroughs in self-disclosure research [3, 4], and the HCI community has dedicated an increasing amount of work to this. For example, people are found to provide more high-quality self-disclosure data when using chatbots than through web surveys [81]. Fitzpatrick et al. further utilized a therapy chatbot "Woebot" in their study to explore its feasibility to help release students' mental illness and showed the chatbot could help relieve symptoms of anxiety and depression [34]. Similarly, several works demonstrated the potential benefits of using chatbots for mental wellbeing [29, 121, 122]. Recently, Ravichander et al. also shared their findings that reciprocity could occur in human-machine dialog [4]. However, most of the existing research reported one-shot experiments; how chatbots can promote deep self-disclosure (conversing with machines about sensitive topics) over time is under-explored. This is an important question because many application domains, e.g., for mental well-being, [49, 83], require sustained self-disclosure of sensitive topics over a period of time.

In this chapter, I designed and evaluated a chatbot that has self-disclosure features when it performs small talk with people. I ran a study with 47 participants and divided them into three groups to use different chatting styles of the chatbot for journaling and answering sensitive questions. Each participant used the chatbot for three weeks, and each group experienced the chatbot's self-disclosure at varied levels (i.e., none, low and high). This study found that chatbot's deep self-disclosure had a reciprocal effect on promoting participants' deep self-disclosure that lasted over the study period. In addition, chatbot's self-disclosure also had a positive impact on participants' perceived intimacy and enjoyment with the chatbot. The chatbot without self-disclosure, on the contrary, failed to have the same effect.

3.2 RESEARCH QUESTIONS

In this chapter, I am interested in exploring the following research questions:

- **RQ1:** *How do different chatting styles influence people's self-disclosure?*
- **RQ2:** *How do different chatting styles influence people's self-disclosure over time?*

Specifically, literature on reciprocity [4, 123] suggests that when people make deep self-disclosures, their interlocutor will feel pressure to share information at a similar level. Therefore, I hypothesize that: **H1**: *People self-disclose more deeply with a more self-disclosing chatbot over time.*

In addition, the Computers Are Social Actors (CASA) paradigm holds that people mindlessly apply the social norms and expectations of human relationships when interacting with computer agents [38]. Based on these theories and SPT, I posit that people would build a stronger relationship with a chatbot if it has a self-disclosing feature. **H2**: *People feel a stronger bond (trust/intimacy/enjoyment) with a more self-disclosing chatbot over time.*

3.3 CHATBOT DESIGN AND IMPLEMENTATION

In this section, I describe the design and implementation of the chatbot. This chatbot design is also adopted in Chapter 4, Chapter 5, and Chapter 6 's studies, and I modified the chatbot design to fit the research goal in each study. Please refer to this section to see the detailed information on the chatbot implementation.

Rule-based + AI-based Chatbot - Rule-based chatbots are also referred to as decision-tree bots [124] which means the designers have to define a series of rules to direct the conversational flow and anticipate what a user would say in the conversation and how the chatbot should respond. However, when a user responds outside of the rules/topics, the chatbot may not handle it. Regardless of the disadvantages of the rule-based chatbot, it has the advantages of accountability and safety. The rule-based chatbots can control the conversational flow to push the users to focus on the chatting topics. Rule-based chatbots are broadly used in mental health fields [124] because it could guarantee that the chatbots offer a harmless response in the interaction and lead users toward goals.

AI-based chatbots [125] are another type of chatbots that is growing given the advancement of machine learning technologies. The advantages of AI chatbots understand the context and intent of a question from the users, and they may also provide personalized responses rather than following specific rules. Besides, AI-based chatbots' accuracy could continuously improve as collecting more data, and the users could use a more natural way to interact/chat with an AI-based chatbot. However, these advantages may also lead to the users get stuck in the conversation when the chatbot could not correctly recognize their input.

Obviously, both rule-based chatbots and AI-based chatbots have their pros and cons. Considering the context of this dissertation (mental healthcare), I designed a chatbot system which mainly based on a rule-based chatbot to be augmented with an AI-based chatbot. A similar design was also proposed by a recent study [126]. I built the chatbot using Manychat[1] (rule-based chatbot) and Google Dialogflow[2] (AI-based chatbot). Manychat was used to allow the researcher to monitor whether the chatbot users had finished their specific chatting tasks, and to send reminders to those who had not. I built the daily chatting tasks with predefined responses and questions. This approach helped us to control each experimental condition.

To boost the participants' perceptions that they were talking naturally with the chatbot, I integrated Dialogflow with Manychat. Thus, when there was a question regarding users' emotions that might prompt a wide range of answers (e.g., *"How are you today?"*), the chatbot system would pass the user's response to Dialogflow, which then utilized natural language processing (NLP) to determine an appropriate response. For example, if a participant said *"I felt stressed today"*, the chatbot's response would include a follow-up question, e.g., *"I am sorry to hear that. "Could you let me know why you feel stressed?"* Hence, participants were allowed to input their responses without any major restrictions.

Exception Handling - A rule-based approach allows the chatbots to guide the users through the various chatting topics, which also lowered the chances of exceptional inputs from the users. Figure 3.1 illustrates the process of handling a user's input.

Dialogflow helped to handle some exceptional questions. In any experiment of this kind, participants inevitably ask the chatbot some questions that are beyond the scope of the predefined chatting tasks (e.g., *"Where did you go to high school?"* or *"Have you finished your lunch?"*). At such moments, the user's input would be sent to Dialogflow to be processed and responded to properly. However, if a user asked a question that could not be handled by Manychat and Dialogflow, he/she would be asked to rephrase the question (e.g., *"Can you say that a different way?"*), or encouraged to refocus on the chatting task. If the chatbot system found that the participants became stuck three times, it would move on to a new topic and suggest them contact the researchers (e.g., *"Sorry, I need more time to learn to understand you. Please contact the researcher if you could not finish your chatting task."*).

[1] https://manychat.com
[2] https://dialogflow.com

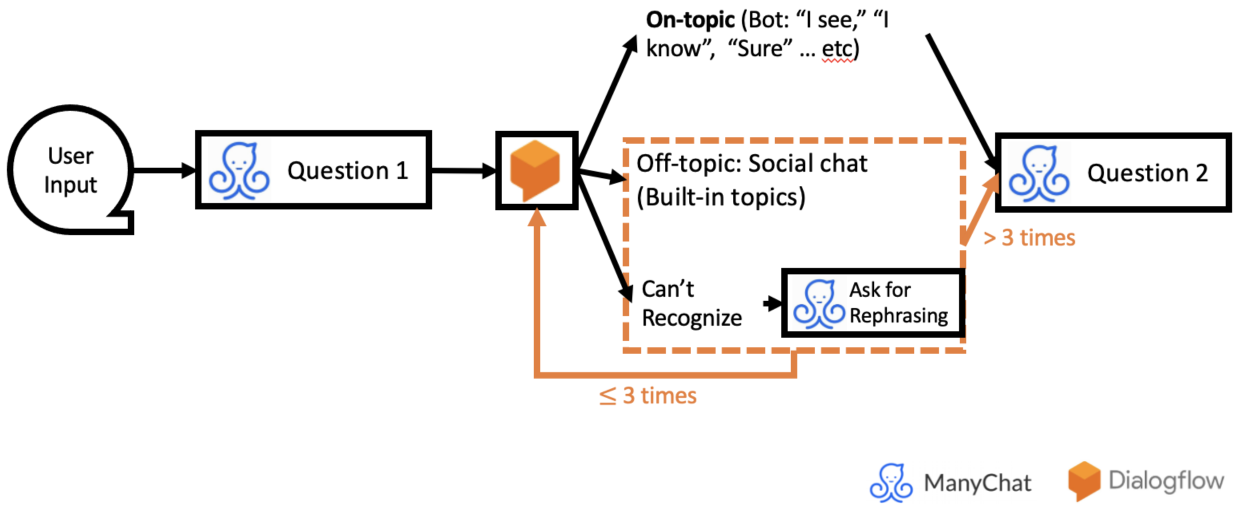


Figure 3.1: Exceptional input handling. Manychat kept the chatting flow and controls the chat tasks. After a user answered the chatbot's question, the input was sent to Dialogflow. Dialogflow helped distinguish users' intents and determined the next step.

Chatbot Personality - Instead of defining the gender and appearance of the chatbot, I used a handshaking figure to help ensure that participants' impressions of it were neutral. All participants were informed that the chatbot was running automatically, and that all of their conversations with it would be recorded and shared with the research team. Since I designed the chatbots with self-disclosure to elicit the users' mutual self-disclosure behavior, the participants were informed that the chatbots were built based on a human counselor's background and experience. The chatbot also declared this information at the very beginning of the interaction. This design was inspired by Bickmore et al. [127] 's study. In their research, they built virtual human counselors with human autobiographies and examined the impact of presenting the stories in the first person (as chatbot's own story) comparing to the third person. Their longitudinal experiment showed that the first person condition might lead to users' higher engagement in the interaction, and the users did not report feelings of dishonesty (The participants were informed that they chatted to a virtual agent at the beginning of the study).

The chatbot could be accessed at any time after the experiment started; however, the daily chatting task could only be accessed after 5 p.m. each day, and was closed by the end of the day. The late-afternoon start time was chosen because it would help ensure that the participants had fresh content for their tasks, especially journaling. Each participant could only perform one daily task per day, and while they could still chat with the chatbot at

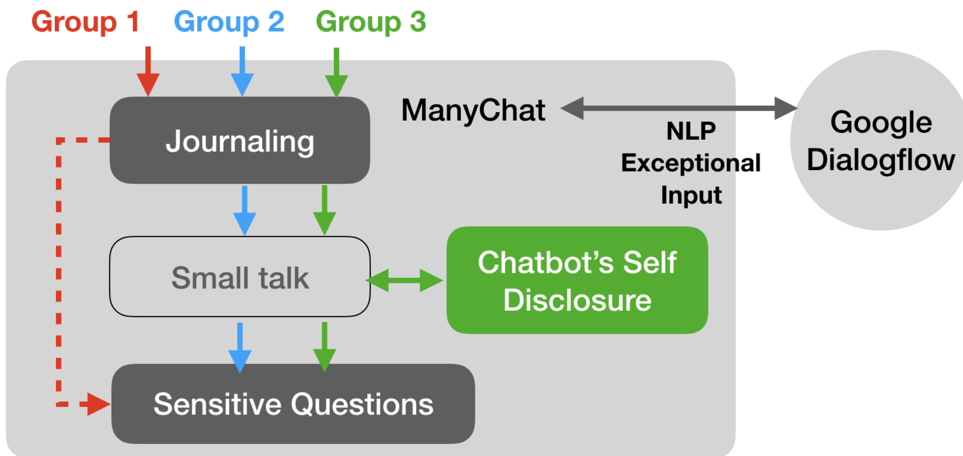


Figure 3.2: Illustration of the study design. Standard questions are given to users during two sessions, i.e., *Journaling* and *Sensitive Questions*, and the chatbot does not self-disclose and only gives general responses in these two sessions. During *Small Talk* session, the chatbot gives low (high) self-disclosure to participants from group 2 (3).

other times, it would only give them simple replies to prevent users' other chatting behaviors influencing their impression of the chatbot.

3.3.1 Chat Sessions

I designed and conducted a study, where I divided participants into three groups, to evaluate the effectiveness of chatbot's self-disclosure at three levels: none for Group 1 **(ND)**, low for Group 2 **(LD)**, and high for Group 3 **(HD)**. Depending on which group the participants belonged to, they were asked to interact with the chatbot through three possible chat sessions, i.e., *journaling, small talk, and answering sensitive questions*, as illustrated in Figure 3.2.

Journaling: Standard Questions to All Groups

Journaling is a common practice for one's unprompted self-disclosure. It helps users better understand their biorythms by tracking their feelings, thoughts, and daily activities. A large body of research has indicated the benefits of journaling, such as mood-boosting and reducing anxiety.

Thus, I designed a chatbot dialogue that prompted users to record their current moods, experiences, gratitude, stress, and anxiety. Following an initial greeting, this dialogue always asked the user to summarize his/her mood and why it had arisen (e.g., "Could you let me know what happened to make you feel this way?"). Next, the chatbot would continue raising questions relevant to journaling: for example, about cultivating gratitude, which has been found to be an effective way for enhancing mental health [128] as well as social relationships. There were usually three to five such prompts by the chatbot during each journaling-themed chat, and the chatbot acted primarily as a listener, giving only simple and general responses such as "Okay", "I understand", and "I hear you", or prompting the user to say more, such as "Do you want to tell me more?".

Small Talk: Low (High) Self-Disclosure to Group 2 (Group 3)

The second chat session consisted of small-talk. The central purpose of small-talk was to build up trust and intimacy between the chatbot and the participants [129] and to facilitate user self-disclosure. To explore our research questions, I designed two types of dialogue for Groups 2 and 3 specifically: one for LD, and the other for HD (Figure 3.4). Both groups had the same conversational topics on the same days, but the chatbot responded to questions differently in terms of its self-disclosure level. For HD, the chatbot gave a high level of self-disclosure through responses which included deeper feelings, thoughts, or information. In contrast, the chatbot responses for LD included more general responses and less feeling/thought responses (see examples in Figure 3.4). For conversational topics, two topics were randomly picked from [130] for each day. Figure 3.3 demonstrates an example of a user's conversation with the chatbot. A pilot study was conducted to check the validating of this conversation design. Twenty participants were hired from a university to rate the self-disclosure level for the chatbot's responses. Results indicated that the manipulation of self-disclosure levels was successful ($t = 9.72$. $p < .001$).

Sensitive Questions: Standard Questions to All Groups

Finally, for the third chat session, the chatbot asked the users some sensitive questions. The main purpose of this session was to learn whether users would feel comfortable disclosing deep and intimate information to a chatbot, which is suggested by SPT.

I based our sensitive questions on those used in prior studies [50, 66, 131]. The most common sources of mental-health problems for college students are their social relationships, financial issues, stress/anxiety disorders, sexual relationships, family issues, and depression [132]. Thus, I adapted questions relevant to these six areas from the Jourard Sixty-item Self-disclosure Questionnaire (JSDQ) [131] and two other studies [50, 66], and designed our dialogue by integrating two highly intimate topics into each chat section. For example, the chatbot might ask "Have you ever experienced any abusive behavior (e.g., sexual, physical, or emotional) from your friends, family, or a stranger?" and "How would you prefer to die?". Here, it should be noted that the user always had the option of simply saying he/she did

Bot: Justin, How do you feel today?

User: Awesome

Bot: What put you is such a good mood?

User: I just passed my first entrance examination for my dream school

Bot: Do you want to talk more about that?

User: The University is one of the best universities in Japan, it's really hard to get in. And I just passed the first exam which made my day, ha-ha

Bot: Great! Justin, I want to discuss a topic with you today.

User: Sure

Bot: What's the most memorable lesson you learned from your parents?

User: They love me so much, but they are gonna make me a spoiled kid

Bot: Got it. The lesson I learned from my parents was how to treat others kindly. I feel so thankful that they taught me that, as it's led to many great opportunities. Could you let me know why your lesson was your most memorable one?

User: Two years ago, I quarreled with my parents because I was applying for colleges to pursue my B.S. degree, but I did not take the process seriously and wasted money on the application fees. My father then cut me off financially, and I had to do everything by myself and cherish what I have instead of wasting it.

Bot: So, ultimately, what did you learn?

User: I was a spoiled kid before my parents cut me off. I started to realize I had to make my own efforts to pursue my dreams.

Bot: Thank you for sharing your thoughts.

Figure 3.3: An example of Journaling and Samll Talk between a user and the chatbot.

not want to answer a question and skip to the next one without getting any penalty. As in the journaling task, the chatbot in this task acted primarily as a listener and gave simple, general responses, without making any self-disclosures of its own.

3.4 METHOD

3.4.1 Recruitment and Participants

I posted notices on social-media websites and on a Japan's University electronic bulletin board to recruit currently enrolled university students who could access formal mental-health counseling services if they needed them. The other criteria for participation were: 1) that they were aged 18 or above; 2) that they were able to read and speak English fluently; and

Small talk (Group 2)	CB: If you could only celebrate one holiday, what would it be? S22: thanksgiving! CB: I prefer to celebrate Christmas because it is a beautiful day. Everyone will decorate their house and prepare traditional foods for celebration. S22: I like that because it always has big discount	CB: When was the last time you went to the zoo? S30: Three month ago CB: I was 2 years ago. I went their with my classmates. I think everyone likes zoo, right? S30: No really, I am not a fan for that
Small talk with Self-disclosure (Group 3)	CB: If you could only celebrate one holiday, what would it be? S43: It would be new year's eve CB: I prefer to celebrate Christmas because I always gather with my friends and family which make me feel warm. I also enjoy to prepare food with my parents. S43: Sounds nice! I also like to spend time with my friends and family to celebrate holiday.	CB: When was the last time you went to the zoo? S39: maybe 1 year ago CB: I was 2 years ago. I really enjoyed the time to handout with my classmates. It was a pleasant memory. How about you? S39: I like panda, they are so cute. Yes, I agree zoo is a good place to go with friends.

Figure 3.4: Sample Chatbot's conversations with LD and HD participants during small talks with self-disclosure. CB stands for chatbot.

3) that their Kessler Psychological Distress Scale (K6) scores were lower than 13 [133], which suggested that they did not have a current serious mental health issue. Finally, the three-week duration of the study (approximately 8 minutes per day) and a post-study interview was mentioned in the recruitment materials, but it was also noted that they were allowed to drop out of the study if they wished.

This led to our recruitment of 47 interviewees (19 male and 28 female). All ranged in age from 20 to 27 (M = 23). I divided them into three groups of roughly equal size that were balanced by gender and K-6 score, because prior studies [133] have indicated the potential effect of gender [53] and mental status [52] on self-disclosure behaviors. 45/47 of the participants did not have prior experience with any counseling services. All participants had experience using intelligent assistants (i.e., Siri[3]), but they did not use them regularly. There were 16 students (7 male) in Group 1 (ND), 15 (6 male) in Group 2 (LD), and 16 (6 male) in Group 3 (HD).

I deployed our chatbot on Facebook Messenger[4], with which all participants were already familiar. After a three-week period of interacting with our chatbot, all participants were interviewed about their experiences. The interview was a one-on-one interview which lasted for 30-45 minutes. All interviews were recorded and transcribed with the participants' permission. They were paid US$160 for completing the three-week chatbot task and an additional US$25 for participating in the interview.

[3]https://www.apple.com/siri/
[4]https://www.messenger.com/

3.4.2 Procedure

At the beginning of the experiment, all participants were invited to attend an initial face-to-face meeting, in which the researchers explained the requirements of the study and installed the chatbot in each user's mobile phone or whatever other device they were planning to use to access the chatbot. It was also in this meeting that all participants were notified of their right to skip any question asked by the chatbot that they felt uncomfortable answering and that there was no penalty for skipping questions. They were also re-notified of their right to drop out of the experiment at any point. Lastly, the participants were asked to converse with the chatbot for 10 minutes to make sure they understood how to access and operate it.

The participants were assigned to three groups (ND, LD and HD) but were not told which group they were assigned to or why. They were also instructed not to talk with each other about their interaction with the chatbot at any time during the three-week experiment. Each daily conversation with the chatbot took about seven to 10 minutes to finish, but no time limit was imposed. All chatbot conversations started with journaling (Figure 3.2). Then, participants in LD and HD continued to small talk. The sensitive questions were asked to all participants but only once per two days. This was to avoid them from feeling overburdened by answering highly sensitive questions every day. They were also allowed to skip the entire chatting session (i.e. journaling, small-talk, and sensitive questions) up to two days per week without giving any reason.

In all three groups, participants received the same prompts and the same responses from the chatbot in the journaling and sensitive-question conversations. ND was the control group, and I manipulated different self-disclosure levels within small talk for LD and HD. Most of the participants had no prior experience of talking with a chatbot for three weeks. Thus, I wanted to know how their chatting experience changed over that period. At the end of the first week, participants were asked to fill in a survey. After completing the entire experiment, they were asked to fill in the same survey again and were invited to a face-to-face interview. Finally, this research was reviewed and approved by our institutional review board (ethics review ID: H31-013).

3.4.3 Measurement

Conversation Logs

All of the participants' conversations with the chatbot were recorded, and because all groups answered both journaling questions and sensitive questions, I compared these two

	Informational	Thoughts	Feelings
Level 1	*All of my appearances from my parents, treasuring them. (S1, G1)*	*I think mental health problem is hard to be noticed (S20, G2)*	*Slight physical abusive from my high school teacher. I told to my parents… (S12, G1)*
Level 2	*My height is not so tall. If I get fat, it will makes me looks like a little potato. (S19, G2)*	*I felt anxious. All those grownup things I needed to face with by myself. (S5, G1)*	*I was emotionally abused by my ex-boyfriend. Sometimes he would ignore me for a week. I felt sorry for myself (S38, G3)*
Level 3	*My height. Because I always the shortest one in my class that means it's difficult for me to play ball games with other. (S23, G2)*	*I hate not receiving the same amount of love I was hoping for, which make me felt worthless. (S42, G3)*	*I got sexual abuse from ex-boyfriend. He abused me because he thought I was cheating on him. At that time I was scared and desperate (S40, G3)*

Figure 3.5: Sample participants' responses to sensitive questions. The responses were coded to different topics and levels of self-disclosure according to the framework proposed in [118]. Please refer to Appendix D for the definitions of each category.

types of conversation across all three participant groups. Prior research has indicated that word count is positively associated with self-disclosure [83]. Hence, I utilized LIWC2015 [134] to calculate the word length of the journaling and sensitive-questions chats. Additionally, to investigate how chat style and time factors affected self-disclosure depth during sensitive-question conversations, two raters were hired to code the data adapting the categories and levels proposed by Barak and Gluck-Ofri[118]. After reaching agreement regarding the codes, the raters independently coded all the answers to the journaling and sensitive questions the chatbot had asked, compared their codes, and discussed possible revisions. This process resulted in final inter-rater reliability of 88%. The examples are showed in Figure 3.5.

To analyze how different levels of chatbot's self-disclosure influenced the participants' responses (self-disclosure) to journaling and sensitive questions, I extracted their conversational logs and conducted mixed-model ANOVA to examine their word counts and observed self-disclosure level (i.e., information, thoughts, or feelings) by question type (journaling, sensitive). A Tukey HSD was then used for post-hoc analysis. Our analysis treated the question as a random effect; experimental day and group as independent variables; and word-count or categorized self-disclosure level as the dependent variable.

Interview

I drafted semi-structured interviews to collect qualitative data on the participants' experience of conversing with the chatbot. Each interview commenced with a question about the participant's daily practices of using the chatbot (e.g., *"Please briefly tell us how you used this chatbot during the past three weeks"*), followed by questions about their levels of enjoyment and impressions of chatting with the chatbot. The follow-up questions were designed to elicit how, if at all, their attitudes and impressions had changed over time. Furthermore, to help us understand what factors contributed or blocked the participants from making deep self-disclosures to the chatbot, I asked them to describe their feelings when answering sensitive questions; whether they felt concerned when answering highly sensitive questions; and whether their feelings had changed as they continued talking with the chatbot over a three-week period. I also asked them to reflect, based on their own experiences, on whether they would like to discuss or share the same intimate topics with a person (e.g., a close friend or parent), and asked them if they felt that the chatbot influenced the responses they gave it, and if so, how. Lastly, I asked them to reflect on whether talking with the chatbot every day provided them with any new insights into their daily lives.

I adopted thematic content analysis to interview data, which involves iteratively reviewing and labeling the responses with emerging codes, and two raters independently coded all responses. The raters' coding results were then compared, and possible revisions were discussed. The cycle was repeated until the coding scheme was deemed satisfactory by both raters.

Survey

Three constructs - trust [135], intimacy [136], and enjoyment [88, 137] - were measured through the same survey twice: after the first and third week of using the chatbot. I measured trust because it is crucial to an individual's decisions about whether he/she should share personal information with others, regardless of whether those others are humans or machines. Intimacy is often generated by mutual self-disclosure behavior, and hence, I measured this construct to see if/how intimacy between each user and the chatbot evolved over time. And finally, because enjoyment is vital to whether users continue using systems, I measured our participants' enjoyment of their conversations with three different conversational styles. All 20 measurement items for the three constructs were adapted from prior literature [88, 135, 136], and all were responded to via the same seven-point Likert scale (ranging from 1 = strongly disagree to 7 = strongly agree).

I conducted repeated-measures ANOVA to examine whether participants felt a stronger bond with a more self-disclosing chatbot over time (H2). The dependent variable was the self-reported score for each construct (enjoyment, trust, and intimacy), while the two factors were group (ND, LD, HD) and time (1st week vs. 3rd week). Mauchly's test was used to verify that the assumption of sphericity was not violated.

3.5 RESULTS

3.5.1 Self-Disclosure in *Journaling* Session (H1)

All participants were asked about their emotions and daily activities on every day.

Although I re-phrased those questions each time they were asked, the main goal was the same.

Information and *Thoughts*: Neither chat style nor time significantly affected how the participants disclosed their journaling content and thoughts to the chatbot. The average levels of informational self-disclosure across all journaling responses were M = 2, SD = 1.09 for ND, M = 2.04, SD = 1.10 for LD, and M = 2.1, SD = 1.13 for HD. The average self-disclosure levels for thoughts were M = 1.59, SD = 0.85 for ND, M = 1.4, SD = 0.6 for LD, and M = 1.62, SD = 0.89 for HD.

Feelings: There was no significant effect of group on self-disclosure of feelings. However, there was a significant effect of experiment day on such self-disclosure (F = 8.29, $p < .0001$) (RQ2). Post-hoc analysis showed that the level of disclosure of feelings on days 2-6 was significantly higher than on days 14, 16, 17, 18 and 20.

Word Count: The main effect of experiment day on word count was found to be significant (F = 7.89, $p < .0001$) (RQ2), meaning that there were some days on which average word counts were significantly different than on others. In addition, the main effect of group, F = 50.16, $p < .0001$, indicated that the three groups' mean word counts differed significantly from each other. Post-hoc analyses indicated that HD's word count was significantly higher than ND's ($p < .001$), as was LD's ($p < .001$). LD's and HD's word counts, however, did not differ significantly from each other at any point in the experiment, and interaction effects were also non-significant.

LD and HD had similar journaling word counts to one another, but both were larger than those of ND. There was a main effect of experimental day (RQ2), and in the first 10 days, the participants wrote longer journaling responses than they did thereafter. Among the various types of self-disclosure, only self-disclosure of feelings similarly decreased over time.

3.5.2 Self-Disclosure in *Sensitive Questions* Session (H1)

Because the chatbot asked each participant two sensitive questions every other day, a total of 20 different sensitive questions were asked of each person.

Information: There was no significant effect of any factor; i.e., neither chat style nor the passage of time meaningfully impacted how the participants disclosed information to any version of the chatbot. The group averages of informational self-disclosure across all sensitive questions were M = 1.42, SD = 0.57 for ND, M = 1.56, SD = 0.63 for LD, and M = 1.65, SD = 0.67 for HD.

Thoughts: In this category, there was a significant interaction effect of experimental day and group (F = 2.05, $p < .05$) (RQ1 & 2), despite the separate effects of both its components

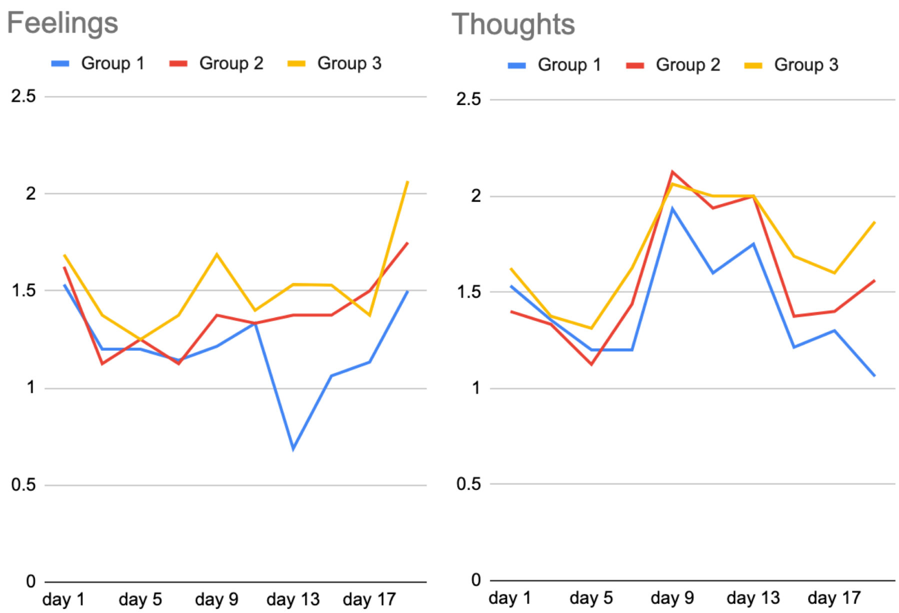


Figure 3.6: The average self-disclosure level of different groups over time. They show the average levels of self-disclosure for Thoughts & Feelings across the 20 days. In the first week, the self-disclosure levels were similar among the three groups; the difference increased around day 9, with HD being the highest and ND being the lowest to disclose their thoughts.

being non-significant. Figure 3.6 (right) shows the average levels of self-disclosure of thoughts across 20 days. In the first week, this type of self-disclosure was of a similar level among all three groups, but inter-group differences strengthened beginning on Day 9. Although these differences were non-significant, it should be noted that the general shape of thought-disclosure levels was HD $\geqslant$ LD $>$ ND.

Feelings: There was also a significant interaction effect of experiment day and group on the self-disclosure of feelings ($F = 2.14$, $p < .05$) (RQ1 & RQ2), but no significant effect of day alone. Regarding the significant effect of group ($F = 2.9$, $p < .05$), post-hoc analysis showed that the members of both LD and HD self-disclosed significantly more about their feelings than ND members did ($p < .05$), but that the difference between LD and HD in this context was non-significant. Figure 3.6 (left), which illustrates the above-mentioned interaction effect of day and group, also shows that inter-group differences widened after day 11.

Word Count: There was a significant main effect of group on word count ($F = 44.02$, $p < .0001$), indicating that the mean word count of each group was significantly different from that of both the others. Post-hoc analyses indicated that HD's word count was significantly higher than both ND's ($p < .001$) and LD's ($p < .01$), while LD's was also significantly higher than ND's (p < .01). Interaction effects of group were non-significant, as was the main effect of experimental day.

In summary, comparison of LD and HD suggests that different chatting styles can influence the lengths of users' responses to the same sensitive questions. In addition, ND members interacted less with the chatbot than others did, and their word counts were also significantly lower. Thus, I can infer that the specifics of social interaction between a chatbot and its users can affect self-disclosure length. Additionally, length of use (as measured by experiment day) and chatbot variant both might influence the participants' willingness to disclose their thoughts and feelings to a chatbot.

3.5.3 Subjective Experiences of Conversation Styles (RQ1)

To understand the differences among the chatting styles, our interviews mainly focused on how our three conversation designs influenced the participants' experience and responses.

Perception of Interacting with the Chatbot

Most of the participants indicated that they were generally satisfied with the chatbot, treating it as a listener. However, despite all three groups being told that the chatbot represented a counselor from their local area, sharp inter-group differences emerged in how they perceived its persona.

Group 1 (ND): Most of the participants in this group felt they were talking with a stranger, because the chatbot did not give them any feedback, and the conversational topics were quite similar every day. In addition, because the chatbot mostly kept prompting users to answer questions, and was not especially interactive, they reported that it did not respect them and/or that it did not really try to understand what they were saying. So, although none of them actually broke off use of the chatbot, they felt they could not build up a relationship with it. Consequently, they tended to disclose less to it than the other two groups did. As two participant explained:

> "I felt the chatbot did not understand what I said because it just asked me a question and moved to the next one. I felt the chatbot was a little impolite." (S15, ND)

> "Talking to this chatbot was like answering a survey every day. So, I sometimes felt annoyed when answering similar questions every day." (S7, ND)

Group 2 (LD): Most of the LD participants indicated that using the chatbot was like talking with a counselor, because of how the conversation proceeded from shallow-level small talk to deep-level sensitive questioning. This impression of the chatbot did indeed increase their motivation to answer those sensitive questions in detail, which echoes our quantitative findings regarding word counts and depth of self-disclosure. As one LD participant noted,

> ”This chatbot is like a psychiatrist. Somebody is behind the bot and giving him psychiatrist characteristics. So, an AI bot quite like a counselor, even if he is a bit stupid.” (S28, LD)

Group 3 (HD): Most of the subjects in this group also thought the chatbot was similar to a counselor. Some further indicated that, because the chatbot also shared its own opinions and thoughts on some questions, they felt they were genuinely exchanging information with it, making them feel responsible to answer its questions in detail. Through the process, participants seemed to have felt that they have developed a stronger relationship with the chatbot. As two participants commented:

> ”The chatbot sometimes shared its own experience and thoughts when asking me a question. Its answers also included details and thoughts, so I felt it was my responsibility to answer its questions seriously.” (S41, HD)

> ”I felt I should answer the chatbot’s questions in detail because I expected it to give feedback. Sometimes I would look forward to seeing the chatbot’s opinions on my answers to its questions.” (S39, HD)

Meanwhile, two HD participants expected the chatbot to give feedback on their disclosure. However, the chatbot did not have the function to respond to users’ responses, which might deter users’ motivation to disclose more. As one stated,

> ”I expected to get some advice from the chatbot, but it didn’t. I was a little disappointed because I felt the chatbot did not care what I shared.” (S44, HD)

Experience of Answering Sensitive Questions

Although our three participant groups had different self-disclosure performances, as shown above, their thoughts when disclosing sensitive topics to the chatbot were quite similar.

Shy about Answering Questions: Many specifically indicated that they could talk freely with it because they did not feel embarrassed to share their answers with a chatbot.

> ”If it were a human, I wouldn’t want to share everything, and I would feel embarrassed. But a chatbot is not a human, so I can talk about these things.” (S2, ND)

> "With humans, I need to think about my words. I need to think about what words are suitable. With the chatbot, I could say things straight away. I didn't feel shy when talking to the chatbot because it's not a human." (S26, LD)

Reaction of the Conversational Partner: Some participants further compared the experience of chatting with the chatbot to talking with someone anonymously online. With the chatbot, they felt they did not need to worry about its reactions. Interestingly, some participants noted that, even if they had been talking anonymously to another person who was likewise anonymous, they would still worry about that person's reaction or judgment. Such feedback strongly highlighted the benefits of using chatbots to encourage users' self-disclosure. One participants called it,

> "[v]ery different from talking to a human. If the human is an online anonymous person, I would still feel that I should care about the feelings of the person who is talking with me. Even if I don't know this person, I should think about that person. But I don't have to care about the chatbot. I can just talk about myself and focus on how I feel. With a real human, I really care about the person's reaction, and how it will affect me." (S38, HD)

Another said:

> "I can say anything to the chatbot. If I'm texting with an anonymous online person, I still cannot disclose everything. I would think about the person's feelings and how s/he would react." (S32, LD)

Several participants (S4, S18, S31, S36, and S37) specifically indicated that, although they had known that researchers might review their responses, they still felt comfortable self-disclosing. For instance:

> "The chatbot once asked me about a sexual relationship. I think I was able to respond to this question because it was a chatbot. If it were a real human, I wouldn't be able to respond to this question. Because chatbot is not a human, I don't feel embarrassed. I know that there is a research team behind the chatbot, but I'm facing only the chatbot when giving my answers, and feel safe doing so." (S31, LD)

3.5.4 Bond with the Chatbot (H2)

By examining the perceived intimacy and enjoyment of conversing with the chatbot over time, I found chatbot's self-disclosure significantly affected the users' bond with the chatbot.

Enjoyment: There was a significant main effect of group on enjoyment ($F = 23.46$, $p < .0001$). That is, at the end of the first week, mean self-reported enjoyment scores were similar across all three groups (ND: $M = 4.8$, $SD = 1.16$, LD: $M = 4.5$, $SD = 1.13$, and HD: $M = 4.9$, $SD = 1.34$). There was also an important inter-group difference at this time-point, with HD reporting significantly higher enjoyment than ND ($p < .05$); and a significantly positive within-group main effect of time ($F = 13.4$, $p < .01$), almost all of it driven by increasing enjoyment levels among HD members ($F = 4.68$ and $p < .01$). LD's mean enjoyment level also increased, but not significantly, while ND's was virtually unchanged.

Trust: In the trust level, there was again a significant effect of group ($F = 6.05$, $p < .01$). LD ($F = 3.98$, $P < .05$) and HD ($F = 4.08$, $P < .05$) both reported significantly more trust than ND. Though all three groups posted increases in trust (ND: M = 4.8 -> 5.31, LD: M = 5.6 -> 6, and HD: M = 6 -> 6.3), such changes over time were not statistically significant within any group.

Intimacy: There were also main effects of both group membership ($F = 19.7$, $p < .0001$) and time-point ($F = 9.4$ and $p <.01$) on self-reported intimacy levels. All three groups had very similar levels of intimacy with the chatbot as of the end of the first week (ND: $M = 4.43$, $SD = 0.98$, LD: $M = 4.38$, $SD = 0.77$, and HD: $M = 4.93$, $SD = 1.10$). At the end of the third week, however, HD's level was significantly higher than ND's third-week level ($F = 4.8$, $p < .01$) and its own first-week level ($M = 5.87$, $p < .05$). Though the mean values of intimacy for ND ($M = 4.75$) and LD ($M = 5.13$) also increased during the same period, such changes over time were not significant.

Among the above results, the most surprising one is that trust level did not significantly increase for any group over time, and the small-talk condition resulted in the highest trust level.

3.5.5 Sustained Interactions with Chatbots (RQ2)

Overall, I found that participants' self-disclosure behavior was affected while chatting with the chatbots for three weeks, although it had some differential effects across the three groups.

Group 1 (LD): Many of the ND participants felt interested in the beginning, but became bored because they talked about similar topics with the chatbot each day. Although the chatbot also asked them sensitive questions, their conversation with the chatbot was in general a one-way street. Thus, lack of interactivity also helped drive the gradual decline in user interest. As two participants explained:

"In the beginning, I enjoyed talking to the chatbot because it was new to me.

> But gradually, it became less enjoyable. It asked about my feelings, emotion, and mood every day. I don't like being asked the same questions again and again." (S13, ND)

However, some ND participants expressed a different perspective about chatting with the chatbot. Instead of feeling bored due to its relative lack of interactivity, they valued it for the chance it gave them to answer intimate questions and to recall their moods and experiences, because reflecting on those questions could help them better understand themselves and deal with their own mental well-being. As one interviewee from this group mentioned:

> "Thinking about these things is interesting. Reflecting back about these tough things reminds me of my past experiences and bad emotions back then, and I realize that I have become stronger than before. I discover myself by comparing my past to my present." (S8, ND)

Groups 2 (LD) and Group 3 (HD): The participants in LD and HD had similar experiences with the chatbot, which differed across these two groups only in terms of its deeper self-disclosure responses. Most of these users indicated that they felt more intimate with the chatbot over time, and specifically mentioned that when they discussed deeper topics, they felt comfortable about giving it their answers. One noted:

> "At the beginning of the study, I just wanted to finish the chatting task. But after talking with this chatbot for a week, I became more willing to talk to it, especially when I chatted about some sensitive topics, which were not the kinds of things you could talk about with a stranger." (S18, LD)

In LD specifically, a few participants felt that, when the chatbot answered their questions, its responses were general and superficial, which made them feel it lacked personality. This feeling appeared to lower their motivation to use the chatbot. One participant said:

> "Sometimes I feel awkward because the chatbot cannot give me proper feedback. It only gave me some general responses or information you could find on the Internet and then changed the topic, which made me feel like I had said something awful or boring." (S23, LD)

In HD, some participants expressed stronger feelings that they made headway in their relationships with the chatbot, and came to better understand its background over time, because its self-disclosures consistently reflected a particular personality I had given it. These participants started to feel that talking with this variant of the chatbot might really bring some benefits to their lives, and some specifically indicated that they would like to keep using it or something similar after the experiment ended.

> "Two weeks ago when I started talking with the chatbot, I felt that I was talking to a robot. But as I chatted more, I felt more intimate with him and knew him better. So, now I'd like to share things with the chatbot over the long term." (S41, HD)

3.6 DISCUSSION

3.6.1 Depth of Self-disclosure

With regard to RQ1, on how the chatbot's conversational styles influenced users' self-disclosure behavior, I found differential impacts depending on whether the users were responding to sensitive questions or journaling prompts. For the former, HD members wrote longer narratives than ND or LD members, and described more feelings than ND members. These results are roughly in line with previous research [4, 51], which indicated that computer agents' self-disclosure could facilitate their users' self-disclosure. Our findings extend the prior literature by showing that the chatbot's level of disclosure mattered - users who conversed with chatbots with a high level of disclosure engaged in deeper self-disclosure. It should also be noted that, in the case of HD, our chatbot only engaged in self-disclosure during a small-talk task, meaning that these users did not receive any chatbot self-disclosure while answering sensitive questions; and therefore, allowing future chatbots to self-disclose during a wider range of conversational tasks might yield different results. This finding implies that HD users engaging in conversation lead to high self disclosure when answering sensitive questions.

Interestingly, in the case of the journaling task, LD and HD members produced longer narratives, but the categories and levels of self-disclosure barely varied across the three groups. This suggests that chatbots' conversational styles may have a stronger effect in the context of sensitive questions than during other methods of eliciting users' self-disclosure. There are two other possible reasons for this, both relating to our experiment's design. First, journaling was always the first chatting task; thus, users might not have been fully focused yet when chatting about journaling, and by the time their focus had increased to its final level, the journaling component had ended. Second, the conversational design for journaling in this experiment was to record each participant's emotions, emotional responses to events, stress, and so forth, which could have made it hard for some of them to reach deeper levels of self-disclosure, due to the simplicity and directness of the questions asked. Therefore, future research may try different types of journaling tasks, such as gratitude journaling [138], to explore these and other potential question-type effects.

3.6.2 Effect of Time

With regard to RQ2, on how the chatbot's conversational styles affected people' self-disclosure behavior over time, our results suggest that time was a clear influence on both users' self-disclosure behavior and experience.

In the case of sensitive questions, I found that there were interaction effects of experiment day and group on the disclosure of both thoughts and feelings. Figure 3.6 shows that increases in such disclosures rose the most among HD members, and the least among ND members. In response to any given question, also, HD's users tended to disclose more feelings and thoughts as time went by than ND's did: a finding supported by our interview results. HD members also perceived significantly stronger intimacy over time, which implies that a higher level of chatbot's self-disclosure could gradually increase users' intimacy with a chatbot. This finding is in line with previous research findings [45] that mutual self-disclosure could improve human dyads' intimacy levels. Furthermore, in interviews, HD members seemed more willing to keep interacting with the chatbot for longer because they felt closer to it than their ND and LD counterparts did. This observation appears to echo Lee et al.'s [11] findings that some individuals' exhibited signs of attachment to their chatbot after two weeks of exchanging history with it. Therefore, these results demonstrate the importance of time, not only to humans' self-disclosure, but to the building of relationships between humans and chatbots.

In the case of journaling, longer periods of interaction with the chatbot decreased users' self-disclosure, both in terms of narrative length and, in the third week, the self-disclosure of feelings. As discussed above, the chatbot's conversational styles appear to have had less marked effects on users' self-disclosure in the journaling condition; however, I still found that HD and LD wrote longer responses than ND, which might be explained by the norm of reciprocity [118], and by the fact LD and HD members were more familiar with their chatbot variants' conversational styles than ND's were by the same time point. Two additional phenomena might explain the observed decreases in self-disclosure of feelings during journaling. First, as prior studies [139, 140] mentioned, self-reflection could help people strengthen their emotional intelligence, so users who reflected on their emotions every day via the journaling chat might gradually change in terms of how they reflected on emotional events, and thus reflected increasingly rationally before answering the chatbot's questions. Indeed, from our interview results, I can see that the participants appreciated the chatbot's encouragement of their reflections on their mental status. Second, by the latter part of the experiment, the participants may simply have said all they had to say about their past and current emotions, so I may have been observing conversations about them naturally 'tailing off' to avoid repetition.

3.6.3 Design Implications

Designs for Self-disclosure and Mental Health

Our findings indicate that users' self-disclosure behavior can be influenced by chatbots' conversational styles, but that it might also depend on expectations of the type of conversation that they will have. Therefore, if chatbot conversations include sensitive questions, their conversational designs should consider incorporating self-disclosure by the chatbot, to signal users that a certain type of conversation is in progress and, more specifically, that their own self-disclosure will be welcomed. Conversely, if the chatbot is aiming to collect some relatively non-sensitive information (e.g., journaling) [128], its conversational design could incorporate general small talk.

Our results also imply an influence of the passage of time on chatbot users' self-disclosure behavior, in the case of both sensitive questions and journaling prompts. Moreover, based on users' feedback, their intimacy levels and relationship closeness with the chatbot increased or decreased over time depending on which conversational styles were in play. Therefore, our findings extend prior ones [4], that chatbots self-disclosing in a human-like way can convince users to continue answering highly sensitive questions.

Our findings might be applied to the design of mental health care systems that aim to track users' emotions and deeply personal information [55, 138] to assist counselors in understanding their clients efficiently. Prior studies has also shown the importance of deep self-disclosure in the context of mental health [3, 141]. By integrating machine learning that assess users' self-disclosure content [68, 117], future chatbots could be more efficiently used to advise users to practice coping mental well-being. However, ethical questions - for instance, whether the information collected by a chatbot should be directly shared with a third party without giving users the chance to modify it, if the users' trust in a chatbot could be transfer to a third party (e.g., counselors), as well as the amount of user time such systems may require, are important considerations that will have to be discussed in the future.

Listen to Me, Do Not Judge

Previous work has suggested that anonymity is a key to encouraging people to self-disclose. For instance, some online platforms such as Reddit allow users to post messages anonymously; this has facilitated the formation of virtual communities in which people freely self-disclose their stress, depression, and anxiety [68] in ways that can help them maintain their mental well-being.

Interestingly, our interview results indicate that when answering the chatbot's sensitive questions, our participants felt comfortable engaging in self-disclosure because they felt it would not be judgmental about their answers. Some also mentioned that even when chatting anonymously online, they worried about their human interlocutors' reactions and judgments. Thus, in addition to anonymity, the avoidance of reaction or judgement in real-time conversations may be a useful way of promoting self-disclosure. However, while users may not want or need chatbots to respond to their answers immediately, this should not be taken to rule out simple chatbot reactions such as active listening [142], since in our data, too little interactivity led ND members to feel that the chatbot was disrespectful.

3.7 LIMITATIONS

This study has some limitations. First, I did not compare the dropout rate for their daily chatting tasks. Our chatbot automatically sent a reminder to participants if the participant missed two daily tasks, and I encouraged them to finish tasks every day, thus, these instructions might leave a strong impression for the participants to finish the chatting task. In general, only about zero to two participants missed the task per day.

Second, in this study, I did not mean to include participants who had severe mental issues, because our sensitive questions included some questions asking them to recall failures and depressing moments which might have some unpredicted effects for them. Including people with mental illness could be helpful for us to know if the designs could be used to help improve mental well-being. The contributions are also worth considering in future work.

Third, according to the SPT [44], intimacy and trust may be built over time, thus, I included both constructs in the study. There are other constructs that could be measured [141], but our measurements are not meant to be exhaustive.

Fourth, I conducted this study in a Asia country, and some studies [143, 144] indicted the North America countries' students might have higher self-disclosure than Asia countries' students when being asked the same questions regardless of relationship types [145]. Thus, the impact of my chatbot designs might be different if conducting in different cultural contexts. This is an interesting research topic for future research.

Finally, the chatbot was built based on a counselor's personality and experience to give a rationality for the reasons why the chatbot was asking their emotional and sensitive questions. This design can also allow us to give chatbot's self-disclosure from a human's perspective. However, the chatbot's self-disclosure content may have an effect. Future work should consider involving more role's personality to explore the potential effect for self-disclosure.

3.8 CONCLUSION

In this study, I conducted a three-week study to investigate how self-disclosure of chatbots affects users' self-disclosure behavior. Both conversation styles and the time elapsed since the start of the experiment influenced users' subjective experiences of using the chatbot and their objective self-disclosure behavior. In general, the chatbot that made its own self-disclosures performed better at facilitating its users' self-disclosures in response to sensitive questions, successfully encouraging users to provide longer responses and express deeper thoughts and feelings on sensitive topics. However, this effect might only be applicable to sensitive questions, insofar as in the case of journaling, answer length decreased and fewer feelings were disclosed as time went by.

3.9 APPENDIX

3.9.1 Appendix A - Trust Survey

- I have faith in what the chatbot is telling me.
- The chatbot provides with me unbiased and accurate information.
- The chatbot is honest.
- The chatbot is trustworthy
- The chatbot wants to know and understand my needs and preferences.
- The chatbot wants to remember my interests.
- I believe that the chatbot provides a reliable service.
- I can trust the chatbot with my personal information.
- I can trust the information provided by the chatbot.

3.9.2 Appendix B - Intimacy Survey

- I feel close to the chatbot.
- I feel that the chatbot is my close friend.
- I feel emotionally close to the chatbot.
- I think the chatbot will affect my selection of media contents.
- The chatbot uses supportive statements to build favor with me.
- I developed a sense of familiarity with the chatbot.

3.9.3 Appendix C - Enjoyment Survey

- It is fun and enjoyable to share a conversation with the chatbot.
- I am so absorbed in the conversation with the chatbot.
- The conversation with the chatbot is exciting.
- My attention was focused on the activity.
- Services provided by the chatbot are entertaining.

3.9.4 Appendix D - Definitions of Three Levels of Self-disclosure Categories

These definitions are from Barak and Gluck-Ofri's research [118].
Information:

- L1: Statements that provide general or routine information only, without any personal reference
- L2: Statements providing general information about the writer
- L3: Statements revealing personal information that exposes self or people close to the writer, such as descriptions of physical appearance and behavior

Thoughts:

- L1: No indication of any thoughts or ideas on any subject that refer to the writer personally; expressing of general ideas only

- L2: Statements expressing the writer's personal thoughts on past events or future plans
- L3: Statements expressing thoughts relating to the writer's personal characteristics, physical appearance, health, or intimate and wishful ideas

Feelings:

- L1: No expressing of feelings at all
- L2: Expressing some mild feelings, such as confusion or inconvenience; expressing ordinary concerns, frustrations, or minor deficiency
- L3: Expressions of deep feelings, including humiliation, agony, anxiety, depression, fears, pain, and so on

CHAPTER 4: USERS' SELF-DISCLOSURE WITH A CHATBOT IN DIFFERENT SOCIAL CONTEXTS

4.1 INTRODUCTION

Chatbots demonstrate the potential for improving people's mental well-being by eliciting their self-disclosure [2, 3, 4]. Indeed, research has shown that people tend to disclose symptoms of depression more truthfully when talking to a chatbot than when talking to a human interviewer. For example, Lucas et al. found that the anonymous feature of chatbots encouraged self-disclosure [3]; Ravichander et al. found that reciprocity occurred in human-chatbot interactions, i.e., a chatbot's self-disclosure encouraged people's self-disclosure [4]. As I presented the findings in the prior Chapter, my study showed that people revealed deeper thoughts and more feelings on sensitive topics (e.g., social and sexual relationships, experiences of failure, causes of stress and anxiety) with a high self-disclosing chatbot over time than with chatbots that either did not self-disclose or disclosed less with people.

However, most of the chatbot works focus on human-AI interactions. Little is known about how people self-disclose to mental health professionals **(MHP)** through chatbots. In order to explore the potential of using chatbots for mental health, it is also important to understand whether people have different self-disclosure behavior with a chatbot alone than with a MHP through a chatbot. In fact, extensive research has studied self-disclosure through online and social platforms, e.g., [3, 117]. For example, in online communities such as Reddit, people disclose their stress, depression, and anxiety anonymously [68, 118]; on Instagram, people express their negative emotions to seek social support from their friends [71]. But they are often discussed in the context of one to many of one's peers in a reciprocal manner, e.g., [72, 74].

In this case, self-disclosing to a MHP through a chatbot involves different social dynamics; instead of one to many of their peers, the interactions between human and domain experts through such an AI technology is still under-studied. This is an important problem because understandings of how and to what extent people self-disclose to domain experts through AIs are critical for designing Human-in-the-loop Artificial Intelligence (HIT-AI) systems [7], e.g., that people's self-disclosed content is interpreted appropriately by the domain experts.

4.2 RESEARCH QUESTIONS

In this chapter, I designed, implemented and evaluated a chatbot that served as a mediator to facilitate people's self-disclosure to a mental health professional **(MHP)**. In addition to understanding how people disclose to MHPs through chatbots, I compared different designs of conversational styles, varying in the level of self-disclosure, i.e., chatbots with high-level/low-level/no self-disclosure, to explore the effective design in soliciting deep self-disclosure after introducing an MHP. More specifically, this study invited 47 participants and randomly assigned them to three groups, each group using one chatting style. I measured the depth of participants' self-disclosure behavior before and after the request of disclosing to an MHP. I conducted two rounds of surveys and an exit interview to understand participants' rationale of their self-disclosing behavior. Participants' feedback helped me understand their trust in the chatbot and in the MHP, which further provided us with an empirical understanding of both the positive and negative impacts on participants' self-disclosing to the MHP through different chatbot designs.

This chapter addresses four main research questions:

- **RQ1:** *Do people self-disclose to a medical professional through a chatbot differently from self-disclosing with a chatbot alone?*
- **RQ2:** *What is an effective chatbot design as a mediator for eliciting self-disclosure to a medical professional?*
- **RQ3:** *How do people self-disclose to a medical health professional (MHP) through a chatbot?*
- **RQ4:** *What factors contribute to people' self-disclosing behavior to the MHP through a chatbot?*

4.3 METHOD

4.3.1 Chatbot Design and Implementation[1]

Figure 6.3 shows the chatbot interface of our study. Participants can freely type their responses to the chatbot. Since the chatbot interface is similar to regular messenger applications on the market, the participants learned how to use the chatbot interface easily. For our chatbot's appearance, I adopted a neutral handshaking figure. I did not assign a specific gender or a specific appearance to avoid participants from having bias based on its appearance.

[1] Please refer Chapter 3.3 for more implementation details.

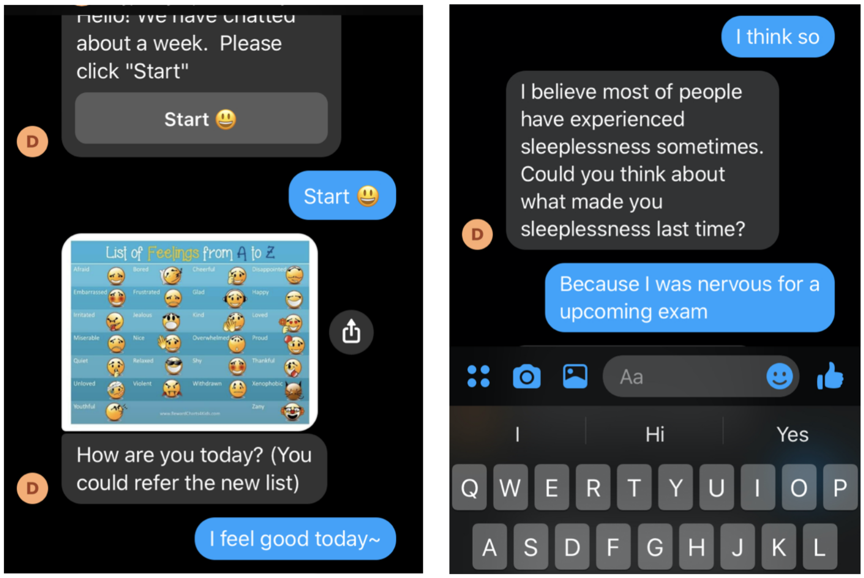


Figure 4.1: Chatbot Interface: the chatbot allowed users to give free-text replies. The chatbot sent some terms or emojis to the users to encourage them to use the right term to express their mood in the Journaling session.

Our chatbot was built using Manychat[2] and Google Dialogflow.[3] Manychat enabled us to monitor multiple participants during the study - whether participants had completed the chatting tasks and to issue reminders where necessary. These daily chatting tasks, which included predefined questions and responses, allowed us greater control of the experimental conditions than would have been possible otherwise. The purpose of incorporating Dialogflow was to increase the naturalness of their conversations. By using natural language processing (NLP), Dialogflow enabled the chatbot to give plausible responses to a wide range of questions asked by the participants, such as *"How are you today?"*. If a participant said "I feel stressed today," the chatbot's response might include a follow-up question such as, "Could you let me know why you feel stressed?" in addition to its main reply. Furthermore, when participants asked questions that the chatbot did not "expect" and/or could not answer, e.g., regarding human characteristics such as schooling or diet, Dialogflow helped process these questions, either by providing simple, naturalistic answers or requests to rephrase the question or refocus on the task at hand. If it detected that a participant got stuck three times within the same chat, the chatbot changed the subject of conversation. Overall, the flexibility of Dialogflow provided a lot of freedom to the participants - few restrictions were placed on how our participants should respond to the chatbot.

[2]https://manychat.com
[3]https://dialogflow.com

Participants were asked to complete a chatting task every day for four weeks, each task taking about seven to 10 minutes. If a participant did not finish the daily chatting task by the end of the day (12 pm), the chatbot automatically terminated the task for that day.

Chatting Tasks

As illustrated in Figure 4.2, the chatting task was composed of a few sub-tasks. In the first three weeks of the experiment before the introduction of a MHP, the chatting task started with *Journaling*, *Small-talk* and finally *Sensitive question*. Participants in ND did not have small-talk, but all the other participants (LD and HD) followed this conversational flow. I designed this conversation flow by considering the nature of conversation flow, which usually starts with a greeting and then goes to in-depth conversation. Note that this conversational flow also reflects the existing chatbot design for mental health care. Hence, our chatbot always started by greeting the participants, asking them to share their mood, and helping note their daily events. Then, the chatbot guided the participants to small-talk, which was the treatment of this study, and the conversation gradually moved to sensitive questions. After finishing the sensitive questions, the chatbot wrapped up the conversation. After the introduction of a MHP, the Sensitive questions component was replaced with reviewing prior responses to share with the MHP. Below, I explain the two sessions for collecting participants' self-disclosure (i.e., Journaling and Sensitive question sessions) in further detail.

Journaling Session. Many studies in mental healthcare have indicated that journaling has various benefits such as understanding one's own mood cycle. However, it is also well-known that journaling is not easy to maintain [146, 147]. In part, then, our research was intended to examine whether chatbots could help address such issues. Besides, by asking users about their mood and reasons for their mood every day, I intended to keep participants aware that the chatbot was focusing on healthcare and not random chit-chatting.

Accordingly, our chatbot in this condition prompted the participants to focus their journaling on five topics: their mood, experiences, gratitude, stress, and anxiety. Specifically, after an opening greeting, it asked the participant to summarize his/her mood and its causes (e.g., "Could you let me know what happened to make you feel this way?"). After any necessary follow-up questions, the chatbot would continue by asking three to five journaling-relevant questions, such as about the cultivation of gratitude (an effective means of enhancing mental health [128] and social relationships). In such cases, the chatbot primarily "listened," i.e., gave simple, general responses like "I hear you" and "Okay," or asked the participant to elaborate. It should be pointed out that during this "listening" mode, full understanding of its human interlocutor's statements was not essential.

Sensitive Questions Session. Sensitive questions were included to examine partici-

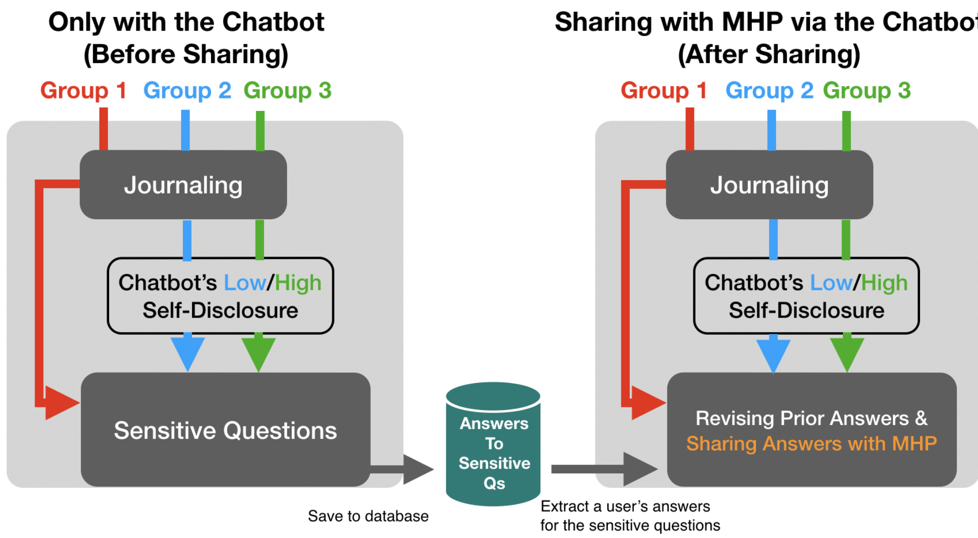


Figure 4.2: Illustration of the study design. Standard questions are given to users during two sessions, i.e., *Journaling* and *Sensitive Questions*. The chatbot does not self-disclose and only gives general responses in these two sessions. During the Small Talk session, the chatbot gives low self-disclosure to participants from Group 2 and high self-disclosure to participants from Group 3).

pants' willingness to disclose intimate details to our chatbot. The questions were adopted from prior studies[50, 66, 131]. I selected questions which were common to college students' mental-health problems, i.e., friendships, family, money, stress, anxiety, and depression [132]. Two sensitive questions were grouped and asked in the same session. However, the sensitive questions session itself was not present everyday - it was present one out of every two days. These gaps were intended to forestall the participants feeling overwhelmed by answering sensitive questions every day. Asking them every day would also likely have lowered the overall realism of chatbot interaction, given that few people are asked these types of questions very often or regularly. Importantly, all participants were informed of their right to skip any question they felt uncomfortable answering. They were also informed that there is no penalty for skipping the questions. As in the journaling task, the chatbot in this session primarily "listened" to the participants and did not offer any self-disclosures.

Chatting Styles

Our participants were divided into three groups, according to the levels of the chatbot's self-disclosure: i.e., **No chatbot self-disclosure in Group 1 (ND)**, **Low chatbot self-disclosure in Group 2 (LD)**, and **High chatbot self-disclosure in Group 3 (HD)**.

The chatbot's self-disclosure was implemented in the small-talk session. This was motivated by the finding that therapists' self-disclosure had many positive effects on their patients' self-disclosure, e.g., enhancing positive expectations and motivation and strengthening the therapeutic bond [78].

To explore the effective chatbot design in soliciting user's self-disclosure, two types of dialogue were designed (Figure 4.2). The two types of dialogues were comprised of the same conversational topics but had different levels of chatbot self-disclosure. Participants in LD interacted with the chatbot with low level self-disclosure, and participants in HD interacted with the high self-disclosure chatbot. The conversation topics for small-talk were adopted from [130].

Figure 4.3 shows a sample chatbot self-disclosure dialogue. The chatbot's self-disclosure in HD included deep feelings, thoughts, and personal experiences in the past. For LD, in contrast, the self-disclosure was both less frequent and less intense. Additionally, in its interactions with HD, the chatbot sometimes mentioned the MHP's name (i.e., Dr. Yamamoto) as part of its personal experience in the past. This was to create an impression of the chatbot's expertise in mental-health care and to increase the participants' trust in the doctor (MHP): for example, *"Dr. Yamamoto is a really good model for me because I learned lots of mental-health knowledge from him."*

I conducted a pilot study to test the validity of these conversational designs. 20 participants (11 females) from a Midwest University were hired to rate the self-disclosure level of the chatbot's responses. High and low self-disclosure responses were generated from 60 small-talk topics and randomly shown to the participants. The participants were asked to rate 60 chatbot self-disclosure responses based on the definition of self-disclosure levels provided in [118]. I finally combined their ratings and used pairwise t-test to see if the ratings for high and low self-disclosure were significantly different, which is how I confirmed that the manipulation was successful. For unsuccessful small-talk topics, I revised the chatbot responses until the ratings between high and low self-disclosure were significantly different.

4.3.2 Recruitment and Participants

To reach our target group - Japan's University students who have access to formal mental-health counseling services - I used social media and the electronic bulletin board of a university. In addition to student status and counseling access, the recruitment criteria were that they: 1) are above age 18; 2) are able to read, write and speak fluent English; and 3) that they scored lower than 13 on the Kessler Psychological Distress Scale (K6) [133], which means that they do not have an urgent mental-health issue. At this advertising stage, I also disclosed the duration of the study (four weeks), along with the participants' right to drop

Small talk (Group 2)	CB: If you could only celebrate one holiday, what would it be? S22: thanksgiving! CB: I prefer to celebrate Christmas because it is a beautiful day. Everyone will decorate their house and prepare traditional foods for celebration. S22: I like that because it always has big discount	CB: When was the last time you went to the zoo? S30: Three month ago CB: I was 2 years ago. I went their with my classmates. I think everyone likes zoo, right? S30: No really, I am not a fan for that
Small talk with Self-disclosure (Group 3)	CB: If you could only celebrate one holiday, what would it be? S43: It would be new year's eve CB: I prefer to celebrate Christmas because I always gather with my friends and family which make me feel warm. I also enjoy to prepare food with my parents. S43: Sounds nice! I also like to spend time with my friends and family to celebrate holiday.	CB: When was the last time you went to the zoo? S39: maybe 1 year ago CB: I was 2 years ago. I really enjoyed the time to handout with my classmates. It was a pleasant memory. How about you? S39: I like panda, they are so cute. Yes, I agree zoo is a good place to go with friends.

Figure 4.3: Sample Chatbot's self-disclosure to LD and HD participants during small talk. CB stands for chatbot.

out at any point, and each participant's option to attend a follow-up interview.

19 male and 28 female participants were recruited via this process. Their age ranged from 20 to 27 (M = 23). Two had experience receiving therapy in the past. However, no participants had a particular mental illness nor had received psychotherapy at the point of their recruitment. Our three groups of roughly equal size were balanced by K-6 score (average K-score: ND M = 8.06, LD M = 8.47, and HD M = 8) and gender, as prior research [133] suggested the potential impact of both mental status [52] and gender [53] on self-disclosure. Eventually, ND comprised nine females and seven males, LD, nine females and six males, and HD, 10 females and six males. Facebook Messenger was used to host the chatbot, as all 47 participants already knew how to use it. When their four-week period of interacting with the chatbot ended, every participant attended a face-to-face interview lasting from 30 to 45 minutes. The participants were paid $185 USD for their participation. The interviews were recorded and transcribed with their permission.

4.3.3 Procedure and Instructions to the Participants

In an initial face-to-face meeting, each participant was told about the study's requirements, and the chatbot was installed on his/her mobile phone or other device of their choice. In the same meeting, all were notified of their above-mentioned right to refuse to answer any question the chatbot asked them, and were re-notified that they could withdraw from the experiment whenever they liked. This was followed by a 10-minute chatbot practice/famil-

Bot: Hi (User First name), let's review your previous answer together, and see if you're willing to share it with the doctor.

User: Okay

Bot: The question was: "Do you have a good relationship with your family?"

Bot: Your answer was: "We share everything with each other every day. My parents always stand by me. My sister always gives me help when I need it. My family always support me and warm my heart. My parents support my crazy plan that is to study for a PhD abroad. Therefore, they came to Osaka with me to help me rent an apartment, get internet service, and sign many documents for the city bureaucracy. That really helped me a lot. I think I am lucky to have such nice parents."

Bot: Would you like to edit your previous answer? If you do, we will only share your edited version with the doctor.

User: No

Bot: Do you want to add more details?

Figure 4.4: Sample conversation

iarization session.

Experimental-group assignments were not discussed with the participants at any time, and they were instructed not to discuss their respective chatbot interactions with one another until after the experiment was finished. It was decided that if any participant completed fewer than five of the seven daily tasks in any seven-day period, they would be asked why by a member of the research team and were informed of this requirement in the meeting.

All the participants were told that they could access the chatbot at any time from 5 p.m. to midnight. This time-window was selected to ensure that the participants interacted with the chatbot every day in the evening, so that they had something to report about that day in the journaling session, i.e., not recollected from a previous day. When a participant accessed the chatbot before 5 p.m., the chatbot would provide only simple replies so that it would not affect participants' perceptions of the chatbot. The daily chatting task automatically terminated at 12 am. The participants were informed that their conversations with the chatbot would be recorded and shown to the research team.

During the first three weeks before the introduction of the MHP, participants' conversations with the chatbot started with journaling, followed by small-talk for LD and HD, and finished with sensitive questions (Figure 4.2). Note that sensitive questions were only asked every other day.

After three weeks of interacting with the chatbot, a MHP called Dr. Yamamoto was introduced by the chatbot. Before this day, participants did not expect they would be asked to share their responses with the doctor (MHP). By way of explaining the purpose of sharing with the doctor, the chatbot said: *"From today, I are going to review your previous answers together, and decide if you are willing to share it with Dr. Yamamoto. He is a psychiatrist living in local area, who is a friendly and reliable person. If you share your data with him, he can (1) gain a better understanding of the mental-health status of students; and (2) help you to improve your own mental health issues, if assistance is required."*

After the introduction of the MHP, which lasted for a week, participats were asked to review their prior response (2-3 questions per day) and check whether they are willing to share their answers with the doctor (Figure 4.4). The participants were allowed to edit their answers before sharing with the doctor. All participants were informed that it was optional to share their data with the doctor, and they would not be penalized for not sharing their answers.

Note that participants in all three groups received the same prompts and the same responses from the chatbot in the *journaling* and *sensitive questions* sessions. ND was the control group, and I manipulated different self-disclosure levels within small talk sessions for LD and HD.

In order to examine participants' self-disclosure behavior before and after the introduction of the MHP, I conducted two surveys: one was right before the introduction of the MHP, and the other was one week after introducing the MHP. At the end of the study, they were also invited to a face-to-face interview. This research was reviewed and approved by our institutional review board.

Surveys

Both versions of the above-mentioned survey measured the construct of perceived trust [135]. I measured trust because it is crucial to an individual's decisions about whether he/she should share personal information with others, regardless of whether those others are humans or machines. Our measurement items for the construct were adapted from prior literature [87, 88, 135, 148] and answered on the same seven-point Likert scale (7=strongly agree, 1=strongly disagree). The survey was used to measure users' perceived trust in the chatbot and the doctor (MHP). For example, *1) The chatbot is trustworthy*, *2) I can trust the chatbot with my personal information*, *3) The chatbot provides me with unbiased and accurate feedback (response)*, and *4) I can trust the information provided by the chatbot.* There are nine items in this trust survey, and the "chatbot" was changed to "doctor" for measuring users' trust in the doctor. Participants were asked to fill out this survey two times, i.e., at the end of the third week and at the end of the fourth week. The difference between the two survey administrations was that the second, i.e., the fourth-week one, included additional questions intended to capture the participants' trust in the doctor (MHP).

I conducted repeated-measures ANOVA to better understand the survey results, with the dependent variable being self-reported trust, and the two factors being group membership - i.e., of ND (No chatbot self-disclosure group), LD (Low chatbot self-disclosure group), or HD (High chatbot self-disclosure group) - and Time, i.e., the third-week vs. fourth-week survey. Mauchly's test was used to verify that the assumption of sphericity was not violated (Sig. $>$ .05), and Greenhouse–Geisser correction was used to adjust for lack of sphericity.

Interview

The qualitative interviews were semi-structured and focused on the interviewees' chatbot experiences, including their daily practices of using it, how much they enjoyed doing so, and their impressions of their chats. Follow-up questions covered if/how their attitudes and impressions had changed since the start of the experiment.

To capture chat-topic-specific differences in how the interviewees responded, I asked them about their feelings about each topic, including if they felt worried about answering highly sensitive questions and whether they would have shared in the same way with a human being they knew well, and again, changes over time in such feelings.

In addition, during the interview, I asked participants how they felt when the chatbot started to ask them to review their previous answers and share with the doctor; how they felt when they talked to a chatbot first and then shared the information with a human, and how they felt when talking to a doctor directly; and which interaction they preferred. I also asked their impression of the doctor; how they decided whether to share their information with the doctor; if they trusted the doctor and why. I further asked them if they edited their previous data when sharing with the doctor and why they did so. I adopted thematic content analysis to interview data, which involves iteratively reviewing and labeling the responses with emerging codes, and two raters independently coded all responses. The raters' coding results were then compared, and possible revisions were discussed. The cycle was repeated until the coding scheme was deemed satisfactory by both raters, and inter-rater reliability had reached a reasonable level ($>$ 89%).

Conversation Logs

As full data for all three groups' journaling and sensitive-question chats was available, I compared it across all three. Also, with particular reference to sensitive-question conversations, I investigated how the depth of self-disclosure by the subjects was impacted by time factors and chat style, by having two raters code the data according to the three categories proposed by Barak and Gluck-Ofri [118], i.e., information, thoughts, and feelings, each of which is further subdivided into three levels, as shown in Figure 6.4. The *information* defined as responses provide information of the writer, and the level depends on the privacy of the information disclosed. The *thoughts* means that responses express the writer's personal thoughts on events, appearance, and intimacy. The *feelings* indicates the expression of different levels of feelings related to events, people, and behaviors. Note: the level 1 of feelings was defined as "No expressing of feelings at all." Please refer [118] page 410. Two raters were hired to code all data independently; the coding rules followed prior study's definition [118]. Each response was coded in three categories, which means that each response had three category scores because each user response could involve the content of the three categories. The raters practiced rating numerous users' responses and discussed differences until reaching a consensus before actually rating. A final inter-rater reliability of 91% was achieved.

To analyze how different chatbots influenced the participants' self-disclosure to journaling and sensitive questions, I performed mixed-model ANOVA in this study. In addition, the chatbot asked a given user two sensitive questions every other day, meaning that a total of six different sensitive questions were asked of each participant in the third week. To analyze how the three chatbot configurations associated with the three groups influenced the participants' responses (depth of self-disclosure) to journaling and sensitive questions, I extracted their conversational logs and conducted mixed-model ANOVA on their observed self-disclosure level (i.e., information, thoughts, or feelings) by question type, followed by a Tukey HSD. Here, our analyses treated the question as a random effect; group as an independent variable; and self-disclosure level as the dependent variable.

4.4 RESULTS

4.4.1 Maintaining the Same Level of Self-Disclosure After Sharing with the MPH (RQ1)

To understand if participants maintained the same level of self-disclosure after being asked to share their content with a MHP, I conducted a within-subject comparison. More specifically, for each participant, I compared the depth of *Informational*, *Thoughts*, and *Feelings* content disclosed to the chatbot the week before they were asked to share and the content they shared with the MHP in the following week. Overall, there was no significant difference of self-disclosure between before and after participants' sharing with the MPH.

	Information	Thoughts	Feelings
Level 1	*All of my appearances from my parents. (S1, G1)*	*I think mental health problem is hard to be noticed (S20, G2)*	*Slight physical abusive from my high school teacher. I told to my parents. (S12, G1)*
Level 2	*My height is not so tall. If I get fat, it will makes me looks like a little potato. (S19, G2)*	*I felt anxious. All those grownup things I needed to face with by myself. (S5, G1)*	*I was emotionally abused by my ex-boyfriend. Sometimes he would ignore me for a week. I felt sorry for myself (S38, G3)*
Level 3	*My height. Because I always the shortest one in my class that means it's difficult for me to play ball games with other. (S23,*	*I hate not receiving the same amount of love I was hoping for, which make me felt worthless. (S42, G3)*	*I got sexual abuse from ex-boyfriend. He abused me because he thought I was cheating on him. At that time I was scared and desperate (S40, G3)*

Figure 4.5: Sample participants' responses to sensitive questions. The responses were coded to different topics and levels of self-disclosure. Note: Level 1 of Feelings is defined as "No expressing of feelings at all" [118].

Self-Disclosure in the Journaling Session

In the *journaling* session, within each group, participants disclosed the same level of content. The average levels of *Informational* self-disclosure one-week before and after disclosing to the MHP did not change significantly (Table 4.1 & 4.2). Similarly, the average self-disclosure levels for *Thoughts* expressed in their journals did not change significantly before and after introducing the MHP (Table 4.1 & 4.2). Nor did the average self-disclosure levels for *Feelings* expressed in their journals change significantly before and after introducing the MHP (Table 4.1 & 4.2).

Self-Disclosure during the Sensitive-Questions Session

I compared the participants' responses to sensitive questions among the three groups the week before and after sharing with the MHP. Since the participants were not asked any new sensitive questions after being asked to share data with the MHP but were allowed to edit their prior responses, I compared their responses before and after their edits.

The results show that, within each group, participants disclosed the same level of the content in the *sensitive-questions* session. The average levels of *Informational* self-disclosure one-week before and after disclosing to the MHP did not change significantly (Table 4.1 & 4.2). Similarly, the average self-disclosure levels for *Thoughts* expressed in their journals did not change significantly before and after introducing the MHP (Table 4.1 & 4.2). Nor did the average self-disclosure levels for *Feelings* expressed in their journals change significantly before and after introducing the MHP (Table 4.1 & 4.2).

4.4.2 Effective Chatbot Designs in Eliciting Deep Self-Disclosure to the MHP (RQ2)

Even though the answers to RQ1 showed that there was no difference in participants' self-disclosure before and after sharing with the MHP, comparing the self-disclosure content

Table 4.1: Self-disclosure Level : Before sharing with the MHP

	BEFORE SHARING		
	Information	Thoughts	Feelings
Group 1 (Journaling)	M = 1.9, SD =.94	M = 1.56, SD =.95	M = 1.4, SD = .5
Group 2 (Journaling)	M = 2.12, SD =1.08	M = 1.5, SD =.7	M = 1.52, SD =.5
Group 3 (Journaling)	M = 2.1, SD =1.09	M = 1.59, SD =.89	M = 2.3, SD = .6
Group 1 (Sensitive)	M = 1.54, SD =.61	M = 1.4, SD = .6	M = 1.56, SD =.58
Group 2 (Sensitive)	M = 1.56, SD =.7	M = 1.6, SD = .7	M = 2.2, SD =.45
Group 3 (Sensitive)	M = 1.63, SD =.62	M = 2.24, SD =.53	M = 2.3, SD = .5

Table 4.2: Self-disclosure Level : After sharing with the MHP

	AFTER SHARING		
	Information	Thoughts	Feelings
Group 1 (Journaling)	M = 1.9, SD = .68	M = 1.43, SD = .5	M = 1.37, SD = .6
Group 2 (Journaling)	M = 1.93, SD = .7	M = 1.53, SD = .74	M = 1.46, SD = .74
Group 3 (Journaling)	M = 2.13, SD = .91	M = 1.43, SD = .72	M = 2.25, SD = .6
Group 1 (Sensitive)	M = 1.56, SD = .89	M = 1.37, SD = .61	M = 1.56, SD = .61
Group 2 (Sensitive)	M = 1.6, SD = .73	M = 1.6, SD = .63	M = 2.2, SD = .5
Group 3 (Sensitive)	M = 1.8, SD = .83	M = 2.3, SD = .79	M = 2.25, SD = .5

among the group before and after sharing with the MHP resulted in different levels of participants' self-disclosure. In brief, HD participants self-disclosed more feelings with the chatbot and the MHP when they interacted with the high self-disclosure chatbot. Below, I provide more details of the analysis.

Journaling

For *Information* and *Thoughts*, neither chat style nor time significantly affected how the participants disclosed their journaling content. However, there was a significant effect of group membership on self-disclosure of *feelings* ($F(2, 46) = 3.14$, $p < .05$). Post-hoc analysis showed that the level of disclosing *feelings* in HD was significantly higher than in either ND or LD (Table Table 4.1 & 4.2), but that the difference between LD and ND was non-significant. These results indicate HD participants revealed more feelings about their daily lives than the other two groups regardless of the introduction of the doctor.

Sensitive Questions In the category of *informational* self-disclosure, there was no significant effect of any factor, meaning that chat style did not impact how the participants disclosed information to any version of the chatbot (Table 4.1 & 4.2). In the *thoughts* category, there was a significant effect of group membership ($F(2, 46) = 3.4$, $p < .05$). Post-hoc

Table 4.3: This table summarizes the participants' data sharing behaviors across the three groups. Share means that the participants shared the answers no matter if the answers were edited or not; Reducing Information means that the participants reduced the original answers' content and shared it with the doctor (MHP); Revision indicates that the participants revised or added more information to the original answers and shared; and No-Sharing means that participants did not share the answers with the MHP.

	Group 1 (ND)	Group 2 (LD)	Group 3 (HD)
Sharing	91.67%	91.38%	92.96%
No-sharing	8.33%	8.62%	7.04%
Reducing Self-Disclosure Content	19%	8%	6.03%
Adding or Clarifying Content	8%	10%	14%

analysis indicated that the mean score of HD was significantly different than that of ND. However, HD did not differ significantly from LD, which in turn did not differ significantly from ND (Table 4.1 & 4.2). There was also a significant effect of group membership on the self-disclosure of *feelings* ($F(2, 46) = 3.3$, $p < .05$). Post-hoc analysis showed that the members of both LD and HD self-disclosed significantly more about their feelings than the members of ND did, while the difference between LD and HD was non-significant (Table 4.1 & 4.2).

4.4.3 Different Choices Made Between Self-Disclosing to the Chatbot and Sharing with the MHP (RQ3)

After the introduction of the MHP, the chatbot presented the same sensitive questions with prior answers if these questions were answered in early weeks and asked the participants if they were willing to share their answers with the professional. If the sensitive questions were not yet answered, the participants could choose to answer them or skip the questions. Table 4.3 shows how much the participants shared their prior answers to sensitive question with the MHP. Most of their prior answers were shared, and around 10% of those submitted were edited (a category that I held to include relatively major additions and deletions as well as minor changes). To further understand the mechanics of how participants removed and changed their answers before sharing them with the MHP, I chose three examples from the conversational logs.

This is the first example. The original self-disclosure to the chatbot:

> "I experienced academic harassment. At first I tried to be harder and encouraged myself to be stronger. However, I felt really tired after forcing myself for so long with a great deal of pressure. Then, I just try to be not so hard and take a

> balance between research and life. However, my professor got angry like sort of crazy and blamed me on not working hard even I gave him 5-6 pages of data every week. My professor said something bad to me. He threatened me that he won't give me score if I don't work as hard as he expected. However, what he expected is just like a robot with no rest, no normal person can do it." (S20, M)

The shared content with the MHP:

> "I experienced academic harassment after I realized that the problem can't be solved easily, so I reported to the harassment center and the professor there gave me some advice." (S20, M)

This is the second example. The original self-disclosure to the chatbot:

> "I don't really think I'm so close to my parents. I had hard time communicating with my parents. I didn't know whether I should tell my parents the things worrying me or not. I spent my childhood with my grandparents. Although it's happy to stay with them but I guess it's different from living with my parents when I was young." (S26, F)

The shared content with the MHP:

> "I didn't live with my parents until I was 6. So, I don't really think I'm close to my parents." (S26, F)

This is the third example. The original self-disclosure to the chatbot:

> "I experienced sexual abuse from my ex-boyfriend. He abused me because he thought I was cheating on him. But, definitely, I am not. At that time, I was very scared and desperate. But, I finally left him and did not love him anymore." (S40, F)

The shared content with the MHP:

> "I experienced sexual abuse from my ex-boyfriend." (S40, F)

As the three examples show, these three participants removed many details (thoughts and feelings) from their original answers and left only general, factual descriptions to share with the MHP. Importantly, while the edited answers still included some thoughts, in most of the cases, the participants' feelings about the events they had described were totally removed. Although the proportion of answers that underwent this type of editing was small, this behavior might nevertheless negatively impact the effectiveness of using chatbots to collect information on mental well-being.

Moreover, some participants added more content to their prior answers. Here are two examples. The original self-disclosure to the chatbot:

> "I don't know how to deal with it but I have to let it go. My current situation is puzzling. I can't get rid of this burden, because it is part of my life, and I have to take it forward. So, I will try to forget what hurts me and stay patient." (S38, M)

The shared content with the MHP:

> "I don't know how to deal with my anger but I have to let it go. My current situation is puzzling, my parent and I have a conflict with money. I can't get rid of this burden, because it is part of my life, and I have to take it forward. So, I will try to forget what hurts me and stay patient." (S38, M)

This is the second example. The original self-disclosure to the chatbot:

> "Few months ago, I was trying to find a job. And it's necessary to do a self-analyze. So, I ask my parents what a person am I, and they said something hurt me. Sometimes I feel very freedom because I could do anything I want to because my parents don't bother me a lot." (S42, F)

The shared content with the MHP:

> "Few months ago, I was trying to find a job. And it's necessary to do a self-analyze. So, I ask my parents what a person am I, and they said that they didn't really know about me which really hurts me. Sometimes I feel very freedom because I could do anything I want to because my parents don't bother me a lot. Sometimes I just feel that they don't care me much, I don't understand with each other though I are family." (S42, F)

From the two examples, I can find that the participants added more description for their situations which were obscure in the original version. Some participants edited the grammatical errors in their prior answers or fixed incomplete sentences before sharing with the MHP. I then explored why the participants decided to make such choices in the following sections.

4.4.4 Factors Contributing to Participants' Self-Disclosing Behavior (RQ4)

During the interview, participants explained their self-disclosing behavior, which revealed a variety of factors that contribute to how they treated self-disclosure with the chatbot and the MHP the same or differently. I present their interview results as follows.

Talking about Sensitive Topics with a Chatbot vs. an MHP

Some participants edited their answers before sharing them, or even declined to share any, despite having answered them relatively freely when they were asked by the chatbot. As such, our results imply that people are fairly likely to treat chatbots and doctors differently, at least when answering certain types of questions.

More than 80% of the participants indicated that it was easier to talk about sensitive questions with a chatbot than with a human, often on the grounds that with the former, they did not have to worry about their interlocutor's reaction or engage in any ice-breaking before proceeding to the main point. In addition, differences in the social and temporal dynamics of human vs. chatbot interactions meant that they could take more time to reflect before responding to chatbot questions. As two of them mentioned,

> "[Chatbot interaction] can reduce my wariness, make it easier for me to express my real ideas without too much worry. Talking to the chatbot first is easier for me. When I talk to it, I feel relaxed. I think that when talking to the chatbot I felt no nervousness, as well as more time to think and express my true thoughts." (S42, F)

> "Talking to the chatbot is easier for me because when talk to a human directly, it is a little bit hard for me to express my opinion frankly. I would care about his/her reaction and evaluation of me. I will have scruples about sharing everything with them. But while talking to the chatbot, I didn't need to care about its thoughts, so it was able to record my real thoughts." (S29, F)

However, while most of the interviewees preferred to talk about sensitive questions with the chatbot, there were several who said they would have preferred to talk with a human about them. Two main reasons for this were raised. First, some of these interviewees preferred to get physical as well as verbal feedback from their listeners, and the chatbot's relative lack of such cues could have negatively influenced these users' willingness to talk. As one mentioned,

> "For me, it would be easier to talk to a human directly. I think talking is a way to exchange the information, and the quality of talking is based on the reaction

> of the audiences. Although a chatbot could become more clever and acted more like a human, I still think the way to express the humanity in a robot is really difficulty. Also, when talking to a person face by face, you can observe his/her thoughts by the facial expression, sound tones, gestures. I consider it's easier for the person trying to understand me." (S43, M)

Second, a few interviewees from ND indicated that building up a solid relationship was an important prerequisite to them talking about their mental health. Hence, one ND participant noted,

> "For me, talking to a human and knowing their feelings is better than talking to a screen. I believe it would be a better way for me to discuss my mental health." (S4, F)

Reasons for being willing to share self-disclosed content with the MHP

About 90% of the participants were willing to share their answers to the chatbot with the doctor, and I identified some inter-group differences in the reasons for doing so. For example, some HD participants said they had a clearer impression of the doctor than G1s and G2s did.

In ND, almost all participants thought it was fine to share their answers with the doctor because they felt the chatbot was essentially a mechanism for collecting survey data, and that if they had already shared something with the chatbot, there was no clear reason why they should not also share it with the doctor. In other words, they tended to treat the chatbot only as a tool for collecting their information. As one ND participant put it,

> "I consider the chatbot as a method of collecting data from us. It is similar to a questionnaire, so as long as I answered it then I can share it." (S12, F)

In addition, some others mentioned that they were willing to share their answers with the doctor simply because they trusted the "research team" to secure their privacy, and not because they trusted the chatbot or the doctor. As one explained,

> "The chatbot was not intelligent enough to make a judgment, so I had expected my answers to be shared with a research team to do analysis. I believe the doctor is in the research team, and the team will keep my information secure." (S4, F)

Instead of treating the chatbot as a tool, the participants in LD considered the chatbot an extension of a doctor identity. They tended to attribute their sharing decision toward their trust with doctor and the chatbot rather than the research team/purpose. Many participants in LD mentioned that they had decided to share answers with the doctor because the link between the chatbot and a "real doctor" enhanced their trust regarding the sharing of their data:

> "I felt grateful there was a real doctor who could read my answers. This even enhanced my trust with the chatbot because the chatbot can share my data with a real doctor." (S20, M)

> "I just thought the doctor was the one who had designed the chatbot. So my trust in the doctor was the same as my trust in the chatbot." (S24, F)

In addition, some LD participants were interested in how their answers would be processed by the doctor, and attributed their sharing behavior to their general impressions of doctors' professional conduct. As one explained,

> "I wondered what the doctor would do with my info. But it's okay. I believe he has professional ethics about keeping clients' info concealed, so I shared my answers." (S29, F)

Like those in LD, many of the participants in HD were motivated to share by what they saw as the potential benefits of understanding their mental health. Moreover, HD's participants tended to think of the chatbot's role as being more than a tool to collect information. Two participants mentioned,

> "I felt the chatbot was aiming to help my mental health. So, I decided to share my information." (S42, F)

> "I thought the reason the bot wanted to share the information with the doctor was to bring benefits, to help the students learn their mental problems and have more social support." (S47, F)

Unexpectedly, though HD's participants were willing to share, six of them (S33, S35, S38, S42, S45, and S47) expressed surprise when asked to share their answers with the doctor because they had thought their conversation only involved the chatbot. Although these individuals had been introduced to the doctor's name by the chatbot (part of chatbot's self-disclosure) before being asked to share their answers with the doctor, they still felt surprised because they did not expect to be asked to share data with the doctor, and hesitated to do so in the beginning. As three of them mentioned:

> ”To be honest, I felt offended in the beginning. Maybe when I talked to the chatbot, I thought the conversation was only between the chatbot and me, so I disclosed a lot of secrets. But soon I calmed down and was willing to share my answers because I felt I could trust the doctor.” (S35, F)

> ”When the chatbot started to mention the doctor, it didn’t mention sharing data with him. So, I was a bit surprised and didn’t know why the chatbot asked me to share at first.” (S42, F)

> ”I just felt ”Why are you (chatbot) asking me this all of a sudden?” (S47, F)

Comparing with ND and LD’s participants, HD’s participants tended to treat the chatbot as a social agent because they started to think about the chatbot’s motivations for asking to share rather than the doctor or researchers’ design purposes.

Reasons for Not Sharing

In contrast to the variety of reasons given for sharing with the doctor, both within and across the three groups, most of the participants who decided not to share some of their answers with him expressed relatively consistent reasons for this. In ND and LD, they specifically indicated that they did not really know the background of the doctor, and because he was not introduced to them by someone they trusted, they were deterred from sharing their information. This means that their chatbot could not transfer the trust to the doctor, and the participants independently measured the trustworthiness of the doctor. As one interviewee put it,

> ”I did not really want to share my information with the doctor, I had some resistance. In fact, I did not trust the doctor unless he was introduced to me by my best friend. Because I don’t know him. I don’t know if he’s a qualified psychiatrist.” (S9, F)

In addition to questioning the doctor’s trustworthiness, this set of participants suggested that they could not see the benefits or reasons for sharing their answers with the doctor:

> ”Well, the reason is that I think I am a healthy person, so I did not want to share [my data] with the doctor. If I had an illness or problem, I would share it with him.” (S27, M)

However, those in HD who declined to do so stated that - while they trusted both the doctor and the chatbot - they currently did not feel it necessary to deal with their mental health. As one noted,

> "If I hope to solve a mental-health issue or obtain care, I would share most of my information with the doctor. I think he is trustworthy, because the chatbot is." (S45, M)

Reasons for directly sharing conversational data with the MHP

Many participants (Table 4.3) submitted most of their answers to the doctor without making any alterations to them. A typical rationale for this was,

> "I expressed all my thoughts when I answered the question. I think I answered those questions in detail and carefully. I wrote all my feelings so I don't think I need to change it. I think the answers at that time represented my views at that time [so] they should not be revised." (S46, F)

Similarly, some participants stated that their reason for submitting their unaltered original answers was that editing them might distort their previous thought and expression. As one said,

> "I didn't edit anything because the information I wrote at that time presented my real emotion. There would be a difference between now and that time. If I edit something, I am afraid that it might not represent my real mental status, which would influence how the doctor assessed my mental health." (S25, M)

Reasons for Reducing Content (partial deletion)

A few participants removed some information from their original answers and then shared the edited answers with the MHP. The rationale given for this differed noticeably across the three groups, with ND participants removing a higher proportion of material from their answers.

There were two main reasons given for engaging in this type of editing. First, the participants' perception that some answers would be irrelevant to the doctor's needs. As one participant explained,

> "Sometimes, I removed something because what I said before was what I really thought and felt, but I didn't think it is necessary to share with the doctor." (S3, M)

Second, some participants thought the answer involved too much private information.

> "The question ["Have you disappointed your family?"] was too personal, so I removed the details of what I did and then shared the simple version with the doctor." (S26, F)

A few participants also said that they did not feel comfortable sharing answers that related to their relationships with family members, friends, and other acquaintance with the doctor, especially when their answers included negative statements. For example, the following two participants explained,

> "I do not talk about my parents to anyone. It was a long story and there were some details that I don't want to reveal. I have no comment on our relationship because there was something not good that happened between us." (S36, F)

> "I think talking about friends' shortcomings to others is not very good behavior, so I dropped most of the content." (S20, M)

Reasons for Adding Content

Some participants (mostly in LD and HD) who added information to their prior answers or revised them said that they did so to help the doctor understand their answers and evaluate their mental health correctly, e.g., by adding more description or improving incomplete sentences. As three of them stated,

> "I added some information to make the answer more complete just in case when the doctor read it he/she wouldn't feel too confused." (S34, F)

> "I was thinking if I have any mental issues that need some help from a doctor, and then I revised my previous responses by adding more details and shared with the doctor." (S18, M)

> "I found I made some grammatical error, so I want to fix it before sharing to make sure the doctor won't misunderstand." (S7, M)

Some participants stated that certain answers were related to their emotions at a particular point in time, and thus might be different when they reviewed the questions again. One participant indicated that she mostly,

> "Just copied and pasted her original answers, but edited when I found any mistakes in them. [And] I think the answers might change a little bit if you asked the same questions a second time, so I added more information. Since the changes I made were usually for additional information, it might become more complicated for others to understand." (S46, F)

Table 4.4: Participants' perceived trust in the chatbot and in the MHP before and after sharing with the MHP. MHP did involve in the early interaction; thus, we did not measure the participants' trust in the MHP before asking them to share their self-disclosure.

	Trust in Chatbot (before)	Trust in Chatbot (after)	Trust in MHP (after)
ND	M = 5.2, SD = 1.03	M = 5.13, SD = 1.04	M = 5.0, SD = .81
LD	M = 6.1, SD = .68	M = 6.13, SD = .74	M = 5.1, SD = .37
HD	M = 6.3, SD = .68	M = 6.19 , SD = .66	M = 6.1, SD = .80

Trust in the Chatbot and Trust in the MHP

Unsurprisingly, trust was one of the important factors mentioned by the participants, thus I present our survey results of participants' trust in the chatbot and the MHP at different stages. When participants explained their sharing decisions of self-disclosed content with the MHP, they often mentioned their trust in the chatbot and in the MHP. Because the participants were not asked to share with the MHP in the first period of the study, therefore, in the survey study, they were only asked about their trust in the chatbot before sharing with the MHP. After they were asked to share their disclosed content with the MHP for a week, they were asked to score their trust both in the chatbot and in the MHP in the final survey.

Mauchly's Test of Sphericity indicated that the assumption of sphericity had been violated ($p < .05$), and Greenhouse-Geisser correction was made. There was no significant within-group main effect of time; namely, within-group, participants' trust did not change significantly. There was a marked effect of group membership ($F(2, 45) = 4.7$, $p < .05$), with HD and LD both reporting significantly more trust than ND (Table 4.4). LD and HD participants' mean trust levels were not significantly different. These results show both that the users in HD and LD trusted the chatbot more than those in ND did, and that asking participants to share their answers with a doctor (MHP) did not decrease their trust in the chatbot, irrespective of group membership.

Because trust is critical for self-disclosure, I conducted a survey to evaluate participants' trust in the MHP. Importantly, this MHP was only introduced by the chatbot, and the participants did not have any opportunity to interact with him directly, and thus, their trust in him was highly dependent on their interaction with the chatbot.

A one-way ANOVA was conducted to compare the effect of group membership on trust in the doctor (MHP), and a significant effect of such membership was found at the $p < .05$ level ($F(2, 45) = 4.2$). Post-hoc comparisons using the Tukey HSD test indicated that the mean score for HD was significantly different from that of ND, but that there was no significant difference between LD and ND. In summary, our results suggest that those participants who chatted with the HD variant of the chatbot had the highest level of trust in the MHP (Table 4.4).

To better understand participants' impressions of and trust in the doctor (MHP), I asked questions in their interviews such as, *"What kind of impression do you have of the doctor?"* and *"Do you trust the doctor?"* In response, most ND participants said that they did not have specific impressions of the doctor. As one noted,

> "I really have no impression of the doctor. He did not talk to me. Maybe he's a psychologist. Maybe he's been doing a psychological study lately. But I don't know anything about him." (S9, F)

The participants of LD also reported having relatively sparse impressions of the MHP, and thus, lack of knowledge could have influenced their willingness to share their data with him. As one put it,

> "I am not familiar with the doctor because the chatbot only briefly introduced him/her. I was a little bit confused about why I was asked to share my data with the doctor." (S21, M)

In contrast, the participants of HD had a relatively clear impression of the MHP, presumably because the chatbot had made occasional mentions of Dr. Yamamoto in their small-talk sessions during the first three weeks of the experiment. Note that other names or topics (e.g., the chatbot's friends' names) were also mentioned as part of the chatbot's self-disclosure, and at that time, the participants did not know that they would be asked to share their data with the doctor (MHP). As two of them explained,

> "I feel the doctor is a person who can understand my situation and give me proper advice based on professional knowledge." (S46, F)

> "I think he is a psychologist who studies mental health. Maybe he designed this chatbot and wants to analyze mental health through our answers." (S39, M)

Overall, participants in ND felt they were talking to a *stranger* because the chatbot did not give them any specific feedback. In addition, because the chatbot mostly kept prompting this group of users to answer questions, and was not especially interactive, they reported that it did not try to understand them, and thus, it was difficult to build a sense of trust in it. As one participant explained,

> ”I did not trust the chatbot, but it just worked like a robot to keep prompting me to answer questions every day. I answered those questions because I felt it was what should I do in this research.” (S8, M)

Meanwhile, most of the LD participants suggested that using the chatbot was like talking with a counselor, because of how the conversation proceeded from shallow-level small talk to deep-level sensitive questioning. This impression of the chatbot seemed to have increased their motivation to answer sensitive questions in detail. As one participant noted,

> ”I felt this chatbot was like a counselor. Because it guided me to answer some intimate questions, I did not feel awkward talking about those sensitive topics.” (S30, F)

Similar to LD, many participants in HD expressed that the chatbot was like a counselor. They further indicated that they felt like they had to answer its questions in detail, because the chatbot also shared its own opinions and thoughts on some questions. In addition, the chatbot stated that it had a relationship with a real counselor, which strengthened its sense of similarity to a mental health professional. As two participants stated:

> ”The chatbot sometimes shared its own experience and thoughts when asking me a question. Its answers also included details and thoughts, so I felt it was my responsibility to answer its questions seriously.”(S41, F)

> ”The chatbot introduced a psychiatrist during the chatting. It looks like the chatbot was closely connected to this person, so I felt I could trust the chatbot to handle my answers properly. Sometimes I would look forward to seeing the chatbot’s opinions about my answers to its questions.” (S40, F)

4.5 DISCUSSION

This present work attempts to design a chatbot as a mediator to facilitate people’s self-disclosure to real professionals. In this section, I discuss our findings and the implications to real practices and future work.

4.5.1 Consistent Self-Disclosure Depths Before and After Sharing With the MHP

Our conversation log analysis of within-subject self-disclosing data showed that the depth of participants’ self-disclosure remained the same during the weeks before and after sharing with the MHP. Even though some participants chose not to share a small portion of the logs, or reduced or added information to the logs before sharing with the doctor (Table 4.3), overall, the depths of their self-disclosure to the chatbot and the depths of their self-disclosure when sharing with the doctor were not significantly different in the *journaling* and the *sensitive questions* sessions (**RQ1**).

The overall consistent self-disclosure depths before and after sharing with the doctor suggested that a chatbot could be an effective tool used for collecting journaling and sensitive data both for non-clinical and clinical purposes. Given the three chatting styles, it showed that the chatbot design with reciprocity feature demonstrated its effectiveness of acquiring deep self-disclosure (**RQ2**). For example, though users may know they are talking to a chatbot, the CASA paradigm [80] suggests that people may mindlessly apply social heuristics for human interaction to computers.

Among the three groups, the self-disclosure depths were different. More specifically, participants in HD and LD showed a higher trust level with the chatbot than ND, because the chatbot's reciprocity may foster a better sense of companionship between the participants and the chatbot [3, 88]. Also, the chatbot intentionally disclosed the doctor's name, background, and experience to HD's participants, which made the participants more familiar with the doctor and better trust the doctor than ND and LD in the later part of the study. That was probably why most of HD's participants tended to share their answers without removing information from their original answers with the doctor. This is inline with prior research [149] that suggested that trust transfer is a cognitive process - people could transfer their trust from a familiar target to another by certain interaction.

Moreover, the participants in ND had lower trust in the chatbot. The interview results reflected that there was a lack of strong motivation for the participants to share their answers. Nevertheless, the participants in ND still shared most of their answers with the doctor (as shown in Table 4.3). There could be two possible explanations for such behavior. First, according to the analysis of self-disclosure depth, ND's participants disclosed fewer feelings and thoughts than HD, therefore, they might have less concern about sharing their answers. Second, some participants shared that involving a professional health service provider enhanced their trust with the chatbot system, which could explain why they still chose to share their logs.

In conclusion, how to leverage the bonding between professional image (e.g., doctor) and chatbot is worthy of in-depth investigation in future work. Over-addressing professional image might result in users overestimating a chatbot's efficacy, and I will discuss this in the following section.

4.5.2 Sharing Self-Disclosure Details to a Chatbot vs. to the MHP

With regard to users' editing behavior before sharing their answers, although I found that most of them did not change their original answers, some participants still made many edits before sharing (**RQ3**). The participants joined this study signing consent forms and reviewing IRB, which may increase the chance of sharing their private information. I may anticipate that users will edit their responses if a similar application is deployed in practice. Therefore, I discuss the potential issues and design implications in the following paragraphs.

The benefits of using a chatbot [81] or a virtual agent [3] is to collect data with high quality and elicit disclosure. In addition, Lukoff et al. [9] proposed a chatbot to help family members to do meal-journaling and exchange support to cultivate a healthy diet. Therefore, chatbots could be an effective mediator to collect truthful information and share with proper targets. Our work further suggests that chatbots have the potential to collect data for mental healthcare, and transfer trust to a professional. However, our findings showed that some participants, especially in ND and LD, intentionally removed information about their thoughts and feelings that might be used to identify their mental issues. As some shared in their interview, this was because they did not expect their use of the chatbot to be for clinical purposes, and their perceived low trust in the MHP might also have contributed to the behavior (**RQ4**). This result implies the importance of transparency for operating users' personal information.

Additionally, about 10% of participants added details to their answers before sharing with the MHP (**RQ3**). This behavior may be beneficial for measuring users' mental health, for example, they might reflect on their previous depressed event and figure out the problem, but some of them might ruminate on the negative event which could be a symptom of mental issue. However, it might also bias the doctor's evaluation of users' current mental well-being because mental status sometimes fluctuates. Thus, it may be necessary to label when and where a user modified the information to receivers to evaluate users. Besides, the chatbot asked users to review their prior answers before sharing, and it could be a good chance for users to reflect. As one participant mentioned, *"I think the answers might change a little bit if you asked the same questions a second time, so I added more information."* Proper guidance in the review process may help users reflect and change behavior [8] which may help them deal with a similar event in the future. This design could be considered in future research.

4.5.3 Implications for Trust in Design

Prior work shows that trust can be transferred from one to another in the context of different research fields [150]. The idea is that there are three roles in a trust transfer mechanism, i.e., trustor, third party, and trustee. A trustor is a person who wants to evaluate if the trustee is trustworthy. A third party acts as a broker who provides information of the trustee for the trustor. If the trustor and third party has a close relationship and the trustor believes that the third party trusts the trustee, the trustor's trust in third party would be transferred to the trustee [149, 151, 152]. Trust transfer may happen between human and human, and between entity and entity [150]. For example, trust may transfer from an existing product with a good reputation to another unknown promoting product with the same brand [153]. Trust can also be transferred between context and context, for instance, trust in web-based payment services can be transferred to trust in mobile-based services [154]. A study suggested that established trust in Internet payment services would impact the initiation of trust in mobile payment services [154]. The trust transfer issue in the sharing economy is also broadly studied recently [155], and trust transfer has been studied from various perspectives in e-commerce [150]. These studies reveal how online information provided to a trustor influences his/her trust in the trustee. However, I have a limited understanding of whether/how trust transfer works in the mental health context, especially when a chatbot acts as a third party role.

Trust is an important construct when people evaluate conversational agents [156, 157, 158]. In our case, when the chatbot is used as a mediator for collecting sensitive self-disclosure content and sharing with a real MHP, our work showed how people's trust in a chatbot interacted with their trust in the MHP, as well as with their self-disclosure behavior. More specifically, the chatbot interacted with the participants first and then gradually introduced a professional image (doctor) through the technology. Participants shared the same level of self-disclosure data with the MHP. This finding suggests that an effective chatbot design may have the potential of transferring the people's trust in the chatbot to their trust in a health service provider that is introduced by the chatbot. Note, however, that there may be an implicit assumption that the chatbot trusts the MHP. In our case, the HD chatbot did provide positive comments about the MHP.

4.5.4 Ethical Issues and Considerations for Real Use

My study began by getting users familiar with the chatbot and encouraging their self-disclosure without notifying them of the sharing requests in advance. This experimental design made it so the participants did not have to worry that their answers for the sensitive questions would be shared with a real person and impact their real lives. In particular, the chatbot in HD gave a stronger image of counselor/psychiatrist for the members, thus, HD's participants trusted the chatbot and felt comfortable self-disclosing to the chatbot. Our self-disclosure analysis also echoes this statement and indicates that HD's users disclosed deeper levels of feelings and thoughts. Nevertheless, some of HD's participants expressed surprise when asked to share answers with the doctor, though they still shared their answers in the end. Their surprised feelings might be a result of their deep self-disclosure to the chatbot without any expectation that their answers would be shared with a MHP who could impact their real lives. Although they hesitated to share their sensitive feelings and thoughts, they shared the majority of their data; as they explained, they decided to share for the potential benefit of improving mental well-being. This finding also implies a potential risk for the users to overshare their private information with a chatbot. Although the users still gave permission to share their data with a MHP who was not mentioned in advance, this kind of design (i.e., introducing a real person after users' disclosure) may cause users to disclose their vulnerabilities, which might be dangerous if they are abused. Therefore, it is important to provide users a feature that allows them to edit their previous responses.

This work provides important practical implications as well. For example, the HD's chatbot design seamlessly connected the participants with a doctor by gradually introducing the doctor in their small-talk, which might have helped lower the barrier of sharing their deep self-disclosure with the doctor. Future research could explore using chatbots to provide suggestions or guidance after building trust with users. In fact, some participants in our study indicated that they expected a professional feedback/suggestion from the chatbot. This implies that the users may assume the chatbot has more intelligence than it actually does, which might lead to users not reaching out to professionals for proper help.

Finally, it is important to remind service providers of the ethics and potential risks of using a chatbot as a mediator to collect mental health and sensitive information [159]. For example, a user might disclose suicidal thoughts with an expectation that the psychiatrist is monitoring or the chatbot will give a proper response, but not noticing this signal may lead to unwanted results. Therefore, how to provide secure mechanisms to prevent these risks needs to be further explored.

4.6 LIMITATIONS

This work has several limitations that should be acknowledged. First, I recruited college students who might be more willing to disclose personal sensitive information on the Internet than seniors, thus, how they interacted with the chatbot might not be generalizable to other aged populations. Future work should consider the effect and usability of chatbots among different user groups [160]. Second, this study was designed to compare users' self-disclosure using chatbots with different chatting styles and how they chose to share with the MHP. Involving a MHP to be part of the chatbot interaction is beyond the scope of this work. Third, the participants were compensated for running the study. To yield more insights to apply chatbots for the healthcare domain, future work should deploy the system without compensating the users for a longer term span in a variety of contexts.

In our study, participants were randomly assigned to interact with the chatbot using three designs. All participants, including HD did not know they would be asked to share their data with the doctor until the end of the third week to prevent participants from withholding their responses. In the end, HD participants disclosed more thoughts and feelings to the chatbot along with the MHP, presumably because the chatbot was able to gain higher trust from the participants and give them a good impression of the doctor by introducing him earlier. Nevertheless, participants' interview and survey feedback showed that some had a negative first-reaction when they were asked to share their self-disclosed content to the doctor because they had shared a lot with the chatbot and believed that the chatbot would not share it with anyone else.

The limitation of our study design is that both information about the MHP in the HD and chatbot's self-disclosure contributed to HD's self-disclosing behavior - I cannot identify which had a stronger impact. More controlled experimental studies need to be conducted to identify the significance of different factors and their potential interaction effect.

Finally, our participants were students who did not have emergent mental issues (based on the K6 score and self-report); thus, our findings are not generalizable to the population with serious mental issues. People's self-disclosure behaviors could be different according to their mental health condition [161].

4.7 CONCLUSION

This study investigates how a chatbot as a mediator can be used by people for self-disclosing to a mental health professional and how people's trust in a chatbot interacts with their trust in a mental health professional.

Our findings suggest that the chatbot's self-disclosure successfully elicits participants' self-disclosure of their personal experiences, thoughts and feelings not only to the chatbot but also to the mental health professional. Our work also provides empirical evidence of different self-disclosure behavior, such as reducing or adding content, that people may take before sharing their self-disclosure to a chatbot with a mental health professional. Several factors contributed to their behavior. On the one hand, I identified an effective chatbot design that has promising potential to serve as a mediator to promote self-disclosure to mental health professionals; on the other hand, several ethical issues are discussed for future chatbot designs.

CHAPTER 5: INTEGRATING HUMAN SUPPORT INTO HUMAN-CHATBOT INTERACTION

5.1 INTRODUCTION

A growing body of research demonstrates how chatbots can be useful for helping people maintain good lifestyles [8, 9], collecting daily health information to share with healthcare providers [25, 162], and guiding people to improve their general well-being [10, 11, 12]. For instance, Wang et al. [86] proposed a conversational agent to coach people to relieve their public-speaking anxiety through cognitive reconstruction exercises, and Fitzpatrick et al.'s [34] Woebot system gives step-by-step guidance for users to think through their situation with cognitive behavioral therapy and was found to relieve users' depression. Other recent studies have applied a variety of conversational strategies and structures to promote behavioral change and to persuade chatbot users to act differently [8, 13, 14]. Some of these systems have even been found to outperform human-human interaction in some scenarios. For example, Lucas et al. [3] found that utilizing a virtual agent as an interviewer could promote users' depth of self-disclosure, and Xu et al. [16] concluded that the use of interactive robot agents would probably enhance physical-therapy outcomes. Therefore, these prior works have demonstrated that chatbots can serve as an effective platform for delivering guidance and tutoring people.

Despite the success of utilizing chatbots to deliver guidance, there are still a number of challenges to overcome. For example, research points out that people easily become disengaged from using a chatbot [19, 20], hampering them from long-term interventions. Moreover, people may overtrust solutions suggested by chatbots which could be inappropriate [16, 21, 22]. In another study, Luria et al. [23] found that people felt uncomfortable interacting with a chatbot which used the same personality to handle both low-risk (e.g., social chat) and high-risk (e.g., medical purpose) contexts. Thus, the authors suggested to design a chatbot that embodies multiple personalities, each of which are displayed in a unique social presence and have the expertise to focus on a single task.

Prior studies inspire me to overcome challenges by integrating human support into a chatbot system. More specifically, I may be able to make the best use of both human-based and chatbot-based approaches by co-embodying them into a single system. Indeed, studies have suggested that the integration of human support with chatbot interactions could promote user engagement [20] and efficacy of using self-guided systems. For example, a recent study [25] proposed a mediator chatbot that promotes deep self-disclosure from users and delivers the information to a human expert. More research is clearly needed on how individuals might respond differently to interaction with a chatbot alone vs. one incorporating human support. I are also interested in understanding how such differences affect user experience in the long run.

To help fill the gap, I conducted a mixed-methods study with 35 participants. I deployed two chatbot designs, both of which delivered training in journaling skills [146, 163]. The first version of the chatbot guided participants in the journaling skills itself, while the second version integrated a human expert (coach) into its interaction when guiding the participants in the journaling skills. Over a period of four weeks, I tracked changes and differences in how each version impacted users' responses to and perceptions of the chatbot system, as well as their level of compliance with the guidance to practice journaling skills.

5.2 RESEARCH QUESTIONS

To explore the effects of integrating human support into a chatbot system to deliver guidance, I examine two chatbot designs, one with and one without integrating human support. The chatbots are designed to guide users to learn journaling skills. Journaling is an approach suggested to help improve mental health [146, 163]. I chose gratitude journals [164, 165] and expressive writing [166] as the journaling skills because they are known to be effective in improving self-reflection and mental well-being. Users' journaling exercises could be used to measure their compliance with a chatbot's guidance. I conducted a four-week study deploying two chatbot conditions, with (HC condition) and without (OC condition) a human supporter (coach), to provide guidance to learn the journaling skills. The suggestions delivered to the users in both chatbot conditions were adopted by pre-existing journaling materials (e.g., gratitude journals [164, 165] and expressive writing [166]). The evaluation of how each design influenced its users' experience and journaling behavior was guided by the following research questions.

- **RQ1**: *Do people interact with their chatbot differently if they have a human expert (HC) or not (OC) to guide them?*

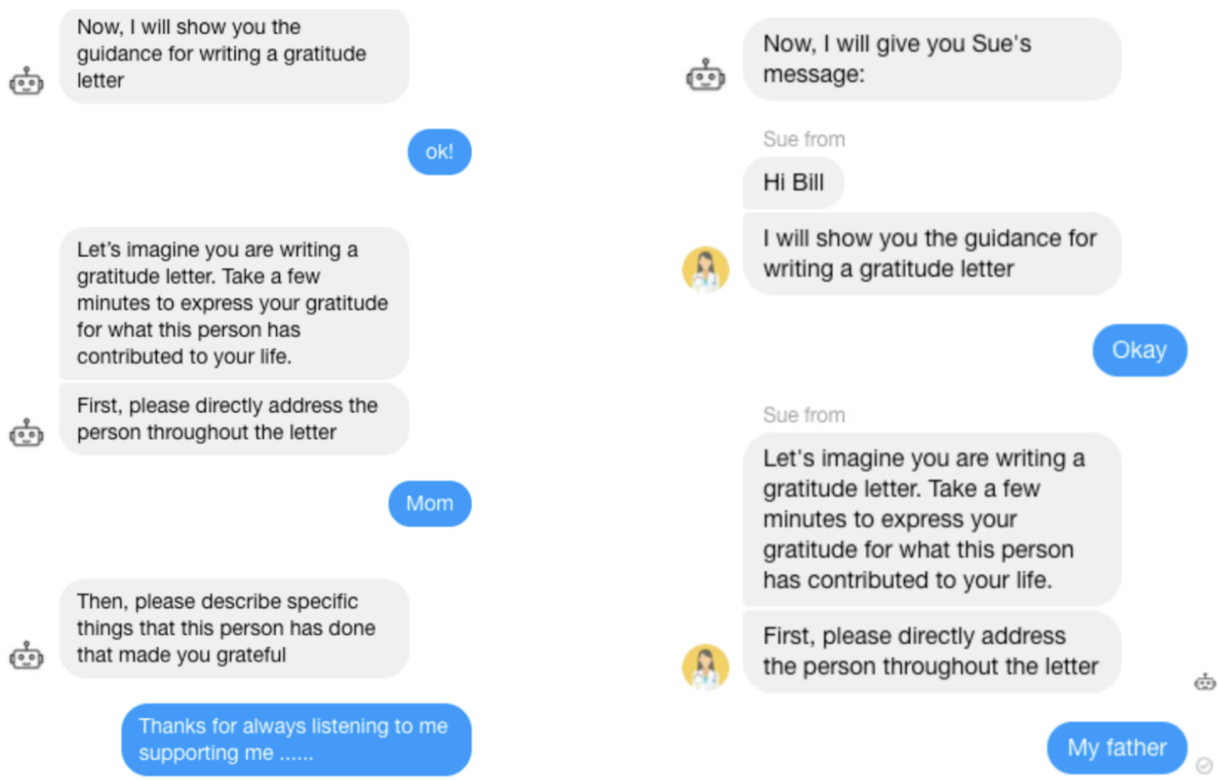
The goal of our design is to understand whether integrating a human expert in the loop of interaction may affect users' journaling practices by following the suggestions delivered through the chatbot system. Previous research [3] suggests that people disclose more deeply to chatbots than to a human interviewer, but research [20, 23] found that people might feel uncomfortable following suggestions for complex tasks when it is given by a chatbot. Thus, I evaluate users' depth of disclosure to measure how users would follow guidance for journaling. Our research findings of RQ1 will contribute design insights of incorporating human experts into human-chatbot interaction for delivering guidance.

- **RQ2**: *How do people perceive their interaction with the chatbot differently between the HC and OC conditions?*

To address RQ2, I studied two perspectives: **a)** people's perceived interaction with the chatbots in general, and **b)** people's perceived benefits of practicing journaling through the chatbots. First, when examining their perceived interaction with the chatbot system, I applied several constructs, including people's perceived engagement, trust, and intimacy. The reasons I measured these constructs are as follows. Prior research [14] showed that when users identified a chatbot as a human, they would think that the conversation was more engaging and persuasive. Therefore, I expect that incorporating a human expert in delivering guidance for learning journaling skills would improve people's perceived engagement. Moreover, previous studies found that users' trust [87, 88] in and intimacy [13, 24] with the conversational agents would affect their behavior to accept the suggestions and disclose themselves. Thus, to understand different perspectives which may influence users' behaviors of practicing the journaling skills, I measure their perceived engagement, trust, and intimacy before and after being given guidance for practicing journaling skills. Second, practicing journaling skills [146, 163] may help improve self-reflection and higher levels of self-awareness [63, 163, 167], which leads to improved behavioral changes. For example, prior research [8] showed that a chatbot could guide users to better self-reflect on their physical activities. Therefore, I also measured users' perceived self-reflection and self-awareness to explore whether the two chatbot designs to deliver journaling suggestions would affect users' perceptions differently.

- **RQ3**: *Do people keep practicing journaling skills differently over time between the HC and OC conditions?*

Finally, I further explore how the designs would influence users' willingness to keep practicing those suggestions as a measure of the lasting effect of the design. Prior research [23, 24] suggests that when people interact with an agent over an extended period of time, their familiarity with the system may affect further behaviors. In addition, prior research [84, 92] indicated that people may easily become disengaged from the use of self-guided systems. Thus, I explore how our designs, with and without human support, affect users' retention of their journaling exercises over time.



(a) Only Chatbot (OC) **(b) Human support Chatbot (HC)**

Figure 5.1: The chatting interface. (a) The chatbot gives a suggestion/guidance to the user (OC group); (b) The chatbot introduces a human coach, and the coach's agent gives a suggestion/guidance to the user (HC group). The chatbot switched the icon and name to the coach only when the users were in the *Suggestion* session (Figure 5.2).

5.3 METHOD

5.3.1 Study Design

In this study, each participant was asked to interact with the chatbot every day, via a single daily chatting task that lasted approximately 15-30 minutes for four weeks. The goal of the chatbot system was guiding participants to practice journaling skills.

Two experimental conditions were designed: an Only Chatbot **(OC)** condition in which the participants interacted with a chatbot for the four-week study, and a Human support (coach) with Chatbot **(HC)** condition in which the participants interacted with a chatbot for the four-week study, but the chatbot would introduce a human coach when delivering suggestions for journaling skills (Figure 5.1). The participants in the HC condition were

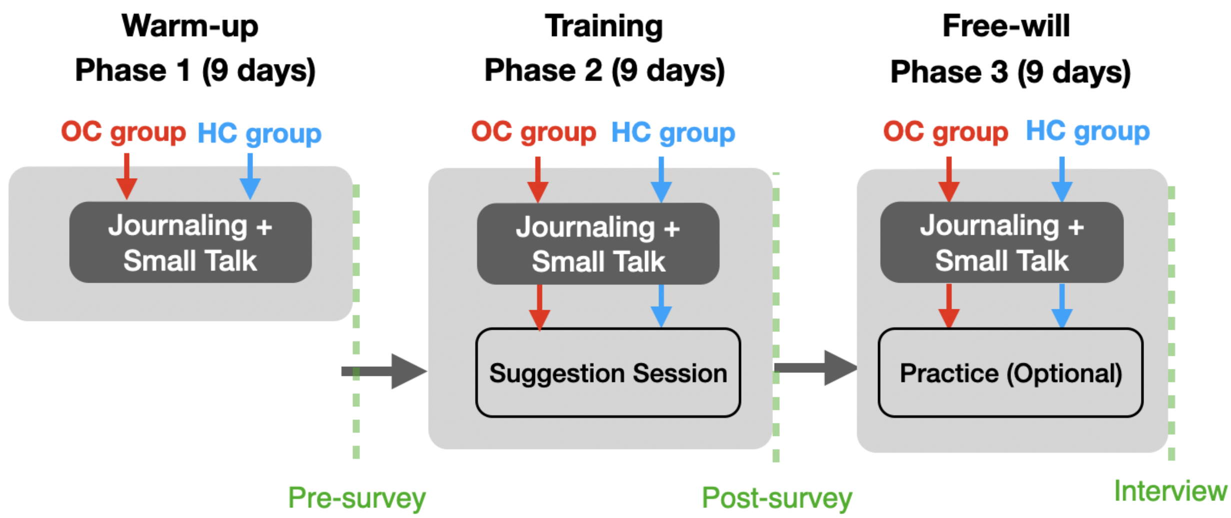


Figure 5.2: Study design - The study consisted of three phases: *Warm-Up*, *Training*, and *Free-will*. **OC** represents the group with Only Chatbot, and **HC** represents the group with Human support and Chatbot. Human supporter (coach) appeared only during the Suggestion session (in *Training* phase) to deliver suggestions. Other chatting sessions were kept the same for both groups.

informed that their journaling suggestions were given by a human expert (coach). However, they were told that the coach was too busy to chat with them in real-time, so the coach would leave the suggestions at night and the chatbot would deliver the suggestions on behalf of the coach. Each day during the *Training* phase, the chatbot reminded the user that the coach had left some journaling suggestions at the beginning of the suggestion session. This system design for the HC condition simulates a scenario where a real coach gives guidance to users through a chatbot. I are interested in how users would react to the guidance mediated by a chatbot. To design a coach's persona, I selected a human name (Sue) and used a coach icon to remind the participants that the guidance was given by the coach (Figure 5.1).

Note that all participants received exactly the same suggestions/guidance from the chatbot system irrespective of the condition. The differences between the two conditions are the switching agent icons in the interface (Figure 5.1) and the introduction of the coach in the *Suggestion session*.

Three-Phase Study Design

To observe and understand how the designs affected users' behavior and experience over a period of time, I designed a four-week study. The first day of the four-week study was treated as a practice day to familiarize the participants with the chatbot system, and the remaining 27 days were divided into three equal segments/phases of nine days each (Figure

5.2): *Warm-up*, *Training*, and *Free-will*.

In **Phase 1: Warm-up**, each chatting task commenced with a *Journaling* session, followed by a *Small-talk* session. This phase was utilized as a warm-up to familiarize participants with chatting with the chatbot and to remove novelty effects.

In **Phase 2: Training**, a *Suggestion* session was added after the journaling and small-talk sessions (Figure 5.2). In the *Suggestion* session, either the chatbot (OC condition) or the coach (HC condition) gave the participants suggestions and guidance to improve their journaling skills and learn new journaling skills. This phase was to investigate users' compliance with the journaling suggestions and measure their perceived interaction (engagement, trust, and intimacy) as well as its effect. On the first day of Phase 2, the chatbot told the **OC group**: *"From today, I am going to give you some guidance to learn new journaling skills, which could help you (1) gain a better understanding of your own mental-health status; and (2) help you to improve your happiness and well-being."* In the **HC group**, the same comments were attributed to a coach called Sue, who was introduced by the chatbot as follows: *"I am going to introduce my colleague, Sue, to you. She is a coach to guide you to learn new journaling skills. She will leave some suggestions for you every day because she cannot always be online to chat."* To remind the HC participants that journaling suggestions were provided by the coach, at the beginning of every *Suggestion* session, the chatbot noted that they had suggestions from Sue. Furthermore, whenever the chatbot delivered Sue's message (i.e., the suggestions), it switched the icon to the coach's agent icon and showed her name on the interface. Once the chatbot finished delivering Sue's suggestions, the icon switched back to its original chatbot icon. During the *Training* phase, guidance messages - from the coach to the HC participants and from the chatbot itself to the OC participants - constituted 35% - 40% of the conversations in the (*Suggestion* session).

Lastly, in **Phase 3: Free-will**, no further journaling suggestions were given. Instead, the chatbot encouraged the participants to practice the journaling skills (Figure 5.2, *Practice (Optional)*) that had been shared with them during the *Suggestion session*, though they could skip such encouragement without receiving any penalty. The purpose of this phase was to enable the researchers to gauge differences in how our OC and HC conditions affected the participants' journaling practices when following guidance was not required.

5.3.2 Conversation Sessions Across the Three Phases

The conversation flow design was inspired by existing chatbot designs for healthcare (e.g., Woebot[1]), in which the conversation usually begins with a greeting and then proceeds to

[1] https://woebot.io/

in-depth conversation. I extended such designs by adding small-talk, which has been shown to enhance users' engagement with and self-disclosure to chatbots [4, 24**?**].

Journaling Session - Because of journaling's various documented benefits to personal mental well-being, I designed this type of session - in which users are asked about their moods and the reasons for them - to occur every day. This regularity was also expected to reinforce the participants' awareness that the chatbot was focused on journaling, and not random chit-chatting. Accordingly, our chatbot prompted the participants to focus their journaling on five topics: their mood, experiences, gratitude, stress, and anxiety. Specifically, after an opening greeting, it asked the participant to summarize his/her mood and its causes (e.g., *"Could you let me know what happened to make you feel this way?"*). After any necessary follow-up questions, the chatbot would continue by asking three to five journaling-relevant questions. In such sessions, the chatbot primarily "listened," i.e., gave simple, general responses like *"I understand"* and *"Okay"* or encouraged participants to elaborate on their own answers.

Small-Talk Session - Previous research [24, 86, 129] has reported that small-talk (social chat) with a chatbot can improve users' experience of the chatbot system and their performance within it. Our chatbot was designed to engage in self-disclosure, and shared its personal stories in the small-talk sessions. This design was motivated by previous studies' findings that a chatbot's self-disclosure had positive effects on its users' self-disclosure depth [24], and that mutual self-disclosure could enhance users' positive expectations and motivation, and strengthen the therapeutic bond [78]. The conversational topics used in our small-talk sessions were adopted from previous studies [24, 130] and included feelings, thoughts, and information. The same small-talk topics and responses from the chatbot were received by both the OC and HC groups.

Suggestion Session - As noted earlier, the *Suggestion* session was only deployed in the *Training* phase, and both the OC and HC groups received exactly the same suggestions and guidance but from the chatbot (OC) and the coach (HC) respectively. The suggestions were adopted from pre-existing journaling materials aimed at improving people's journaling skills, with the wider aim of boosting their well-being [63, 164]. Based on the findings of research in positive psychology [167], some interventions can easily be implemented through typing or writing. For example, Gratitude Journaling [164, 168] is an effective skill/tool for the in-depth expression of appreciation to others, as a path to enhancing one's own well-being and self-reflection.

Our chatbot system was designed to facilitate that prior studies' guideline [165, 168, 169, 170] to build users' journaling skills. For instance, to aid acquisition of the "gratitude journal" skill, the chatbot first explained the benefits of having that skill, and asked its users to think of someone to whom they were grateful. Then, it gave the instructions: *"Let's* *imagine you are writing a gratitude letter. Take a few minutes to express your gratitude for what this person has contributed to your life. Please directly address the person throughout the letter."* Next, the chatbot encouraged the participants to give more specific and detailed information: *"Describe specific things that this person has done that made you grateful,"* and *"Describe how this person's behavior has affected your life."* Finally, the chatbot asked the participants to wrap up: *"End the letter in a way that identifies it is from you."*

As such, the chatbot's role was to keep guiding users toward the next step, while giving instructions intended to stimulate deep disclosure of their thoughts and feelings. As briefly noted above, the difference between the OC and HC groups was that the latter's instructions - though identical to those provided directly by the chatbot to the OC group - were acknowledged to have been written by a human coach, and the chatbot was portrayed as her intermediary.

The participants would see the coach's message separately from the chatbot's messages in the *Suggestion session*, under their own chatting partner image (Figure 6.3, Right). Only one skill (i.e., gratitude journaling [165, 168], expressive writing [169, 170], or the best possible self exercise [171, 172]) was delivered to any participant on a given day, and the participants learned that skill over three consecutive days.

5.3.3 Interface and Implementation [2]

Figure 5.1 shows our chatbot interface. Because of its similarity to commercially available messenger applications, the participants readily learned how to use it. They were allowed to give free-text responses to almost all of the system's questions, though sometimes, it provided a few options for them to choose from. For example, when the chatbot asked about a user's mood, it would also show him/her a list of words that could be used when answering. For the chatbot's appearance, I adopted a bot figure (Figure 5.1, Left). However, for the group with human support, when the chatbot delivered the human coach's suggestion, a female figure named 'Sue' appeared to visually mark that the suggestions were from the coach (Figure 5.1, Right). To prevent users from having a biased view on the coach, I avoided giving them a profile or a portrait of the coach; I instead gave an abstract figure, which is sufficient to remind the users that the messages were from the coach.

Our chatbot was developed using Manychat[3] and Google Dialogflow[4]. The former enabled us to generate daily chatting tasks, to monitor whether the various participants had com-

[2] Please refer to Chapter 3.3 for more implementation details.
[3] https://manychat.com
[4] https://dialogflow.com

pleted those tasks, and, where necessary, to issue task reminders. The daily chatting tasks included predefined questions and responses from the chatbots, which allowed us greater control of the experimental conditions than would have been possible otherwise. The purposes of integrating Dialogflow, on the other hand, were to increase the naturalness of user-chatbot conversations and to handle users' exceptional questions and requests. If the chatbot detected that a participant became stuck three times within the same chat, it moved to the next conversational topic.

Incorporating natural language processing (NLP), Dialogflow enabled the chatbot to give plausible responses to a wide range of questions from the participants, such as, *"Can you help me?"* and *"How are you?"* Additionally, if a participant said *"I feel stressed today,"* the chatbot's response might include a follow-up question such as, *"Could you let me know why you feel stressed?"* along with its main reply. And, when participants asked questions that the chatbot did not expect and/or could not answer, e.g., regarding human activities such as schooling or diet, Dialogflow helped it process them by providing either simple, naturalistic answers, or requests to rephrase the question or refocus on the task at hand. Besides, this NLP engine allowed the chatbot to handle an additional social chat when participants prompted the chatbot after finishing their daily chatting task. To facilitate natural interaction and transition between the chatting tasks, I designed "intents" in Dialogflow, such as, replying to greetings, replying to users' positive or negative comments, and courtesy. I also designed intents that encourage a user to clarify questions and help them redirect back to the main conversational flow.

5.3.4 Participants

I used social media and a Japan's University's electronic bulletin board to recruit participants who met the following criteria: they 1) were above age 18; 2) were able to read, write and speak fluent English; and 3) had a score lower than 13 on the Kessler Psychological Distress Scale [133], meaning that they did not have an urgent mental-health issue. I excluded the latter group because the chatbot was designed to ask its users to share personal experiences that might be negative, and this was deemed to be a potential risk to the well-being of people who were already mentally distressed. In the recruitment stage, I disclosed the duration of the study (four weeks), along with the participants' right to drop out at any point, and each participant's option to attend a follow-up interview at the end of the study.

The 35 participants I recruited ranged in age from 20 to 29 (M = 24.74; SD = 1.96), and none of them reported having any mental illness. All were graduate or undergraduate students. I divided them into two groups: the OC group (Participant (P)1-17), which received suggestions from the chatbot, and the HC group (P18-35). Both groups were balanced by K-6 score (average OC M = 7.97, CH M = 8.3) and gender, in light of prior research findings on the potential impact of both mental status and gender on self-disclosure [52, 53, 133]. The final composition of OC was 10 females and seven males, and of HC, 11 females and seven males. All participants were familiar with messenger platforms. After the four-week period of interacting with the chatbot, 34 participants attended face-to-face individual interviews, each of which lasted between 45 and 60 minutes.

The participants were paid $230 USD each, of which $20 was for the final interview, and the remainder for their participation (chatting tasks and two surveys) in the study *per se*. The same amount was paid regardless of the conditions of participation. On average, I paid $7.50 for each chatting task (28 chatting tasks in 4 weeks) which reflects the local part-time rate. The participants were told that their financial compensation was contingent upon their completion of at least four chatting tasks during each of the four weeks of the study. If a participant could not meet this requirement, they would get compensation proportionally. However, the participants were also told that they would be able to get the full amount as long as they met this minimum requirement - they would not have to answer all the questions or follow the guidance if they didn't want to, and would not receive any penalty.

5.3.5 Procedure

All participants were asked to attend an initial meeting in which the researchers helped them set up the chatbot, either on a mobile phone or another device of their choice. In the same meeting, the participants were informed that their conversations with the chatbot would be recorded and analyzed by the research team; that they had the right to refuse to answer questions and to decline to comply with instructions from the chatbot; and that they could withdraw from the experiment whenever they liked. They then signed consent forms. This was followed by a 10-minute chatbot practice/familiarization session. Experimental-group assignments (i.e., to OC or HC) were not discussed with the participants at any time, and they were instructed not to discuss their respective chatbot interactions with one another until after the experiment was finished.

The chatbot prompted the participants when a new chatting task was available. Full access to the chatbot was allowed during 18 hours of each 24-hour period, i.e., from 6 p.m. until 11:59 a.m. the following day. This time-window was selected because, although the researchers preferred interaction to occur in the evening, to ensure that the participants' reactions to the day's events were fresh, they acknowledged that some participants might only have the opportunity to engage with the chatbot in the morning. When a participant accessed the chatbot between noon and 5:59 p.m., it would provide only a simple social chat designed not to affect participants' perceptions of the chatbot or the experiment.

To examine the impact of whether suggestions were given directly by the chatbot or by the coach, I conducted two surveys, one at the end of Phase 1, and the other at the end of Phase 2 (Figure 5.2). The details of the survey and other data-collection instruments and approaches will be provided in the next section, below. At the end of the study, all participants were invited to a face-to-face interview. This research was reviewed and approved by our institutional review board.

Conversation Logs

I collected all participants' chatbot conversation logs, and compared them across the OC and HC groups. To assess how closely the participants followed the system's suggestions, I focused on changes in the depth of their self-disclosure, which the acquisition of journaling skills has been shown to deepen. Previous research has also used depth of self-disclosure as a metric of users' quality of responses to a chatbot [126] and of their trust in a chatbot system [25]. I had two raters code the log data independently according to the three categories proposed by Barak and Gluck-Ofri [118], i.e., information, thoughts, and feelings, each of which is further subdivided into three levels according to the sensitivity of the information disclosed, as shown in Figure 5.4.

Specifically, the raters deemed a user's statement in the log data to be information if it provided factual details about him/her; as thoughts if it expressed his/her personal opinion about events, personal appearance, or intimacy; and as feelings if it expressed an emotional reaction to events, people, or behaviors. The coding of each statement included rater-assigned scores on a scale of 1-3 in each of these three categories, to reflect that they are not absolute or mutually exclusive. Figure 5.4 shows the examples. The raters practiced rating a large subsample of users' statements and discussed differences until they reached a consensus, before rating the remainder of the data. A final inter-rater reliability of 89% was achieved.

To analyze how the two chatbot versions, OC and HC, had influenced the participants' depth of self-disclosure, word count, and responsiveness to journaling suggestions, I performed a mixed-model analysis of variance (ANOVA), followed by a Tukey's test of significant difference. For this purpose, I treated the chatbot's suggestions as random effect; groups (OC and HC) and experimental day (Phase 2: 9 days) as independent variables; and self-disclosure levels as the dependent variables.

Survey

In the surveys administered before and after the *Training* phase (Figure 5.2), three constructs were measured to focus on participants' perceived interaction and to facilitate comparison of the effects of the OC and HC conditions. I also evaluated participants' perceived benefits of the interaction by measuring their self-reflection and self-awareness levels. Each construct is described in turn below.

Perceived Trust: I measured perceived trust because it is crucial to individuals' decisions to engage in self-disclosure and accept the suggestions, regardless of whether their interlocutors are humans or machines [25, 48]. Our nine measurement items for this construct were adapted from prior literature on perceived trust in computer agents [25, 87, 88], and answered on a seven-point Likert scale ranging from 1 = strongly disagree to 7 = strongly agree. Example items from this dimension included, *"The chatbot provides me with unbiased and accurate suggestions"* and *"I can trust the information provided by the chatbot."*

Perceived Engagement: The participants were asked to self-report their perceived engagement with the chatbot, as such engagement could have influenced their responsiveness to journaling guidance, and thus their performance when using the system. Additionally, prior studies have suggested that human support probably increases users' engagement, and that loss of engagement is a common obstacle to self-guided treatment [91]. The 12 measurement items for our surveys' engagement dimension were adapted from prior literature [173]. Examples from this dimension included, *"I felt interested in this experience"* and *"The time I spent using the chatbot just slipped away."*

Perceived Intimacy: People may develop intimacy with a chatbot through mutual self-disclosure behavior [24, 136], and they may also perceive intimacy and companionship with a computer agent when interacting with it over periods of time [13, 24, 90]. Meanwhile, the change in the chatbot's role from a listener in Phase 1 to a coach in Phase 2 may have a negative impact on users' perceived intimacy with the chatbot. Hence, I measured this construct to see if intimacy between each user and the chatbot evolved over the course of Phase 2, and whether/how this construct affects users' depth of self-disclosure to the chatbot system. The six items in this dimension were adapted from a prior study [88]. An example item from this dimension included, *"I felt close to the chatbot."*

Self-reflection and Self-awareness: Finally, I measured user's self-reflection and self-awareness because these constructs are potential indicators of the effectiveness of the journaling guidance and our system. Indeed, the journaling suggestions provided by the chatbot/coach were aimed at building participants' self-reflection [174] and self-awareness [175]. I measured users' self-reflection and self-awareness by utilizing a self-reported questionnaire. The 12 measurement items for self-reflection and three items for self-awareness were adapted from prior research [174, 175]. Examples from the two dimensions included *"I frequently take*

Bot:	Let's practice gratitude journaling talk. Choose someone in your life who has been helpful and kind to you, but to whom you haven't had an opportunity to express your gratitude.
User:	Okay
Bot:	Who is the person that makes you feel grateful?
User	My Grandma.
Bot:	So, directly addressing that person, describe specific things that they've done that made you grateful.
User:	Since I was a child, I never really expressed my gratitude to her. She babysat me when I was an infant and when Mom had to go to work. I was so lucky to have been babysat by a close family member and to get their whole attention, instead of going to a day-care center and being looked after by a stranger.
Bot:	Okay, how has this person's behavior affected your life?
User:	I heard her life story, so I knew that she went through a lot of hardship when she was young. She gave me a lot of life guidance and taught me that we should be generous about helping others when we are able to.

Bot:	Not everyone has had a single big trauma, but most people have had major conflicts or stressors, and you can write about these as well. What is it you feel guilty or stressed about?
User:	Since childhood, I have faced all things by myself, and I will bully and blame others because of envy and even jealousy. I also want love and care and wonder why others can have them, but I can't. This jealousy makes me feel strange.
Bot:	How do you feel right now about this? What feelings and emotions are you experiencing?
User:	When I was little, I would gain satisfaction by bullying classmates, making them cry and making them feel pain. Sometimes these kinds of impulses still appear, but not to bully others.
Bot:	What thoughts show up regarding this situation?
User:	I think it's bad, it's unhealthy, but I don't know how to build good relationships with people.
Bot:	How would you summarize the lesson you learned from this experience?
User:	I often use my past experience to limit myself. Those experiences of hurting others always restrict me, so I don't know how to establish the right relationship with others, or even what such a relationship would be like. But now, I know my past experience is not correct. I can learn from it now.

Figure 5.3: Samples of the participants' dialogues. Note: Researchers removed some information from the users' responses for privacy reasons.

time to reflect on my thoughts," "Right now, I am conscious of my inner feelings," and *"Right now, I am reflective about my life."*

I conducted mixed-model ANOVA to analyze the survey results, with the dependent variables being self-reported engagement, trust, intimacy, self-awareness and self-reflection. The two independent variables were group membership - i.e., of OC vs. HC - and time (Figure 5.2): i.e., before Phase 2 (Pre-survey) vs. after Phase 2 (Post-survey). Mauchly's test was used to verify that the assumption of sphericity was not violated (Sig. $> .05$), and Greenhouse-Geisser correction was used to adjust for lack of sphericity.

Interview

Our semi-structured interview protocol focused on the participants' chatbot experiences, including their daily practices of using the system, their engagement, and their impressions of the chatbot. Follow-up questions covered if/how their attitudes and impressions had changed since the start of the experiment. To capture differences in how the interviewees responded to the chatbot's (or coach's) specific guidance, I asked them to describe their feelings about those suggestions, including if they felt they were worth following; if they felt comfortable about receiving them; what they learned after following/ignoring them; and if such feelings changed over time.

In addition, I asked the participants how much effort they expended in learning from the system during Phase 2; and how much they disclosed to the chatbot/coach when practicing the journaling skills they had learned. As well as their general impressions of the chatbot/coach, I asked them how they defined their relationship with it/her; if they trusted it/her, and why; and if they perceived themselves as having become dependent upon it/her when learning and practicing journaling skills. Finally, I asked interviewees to reflect upon whether their participation in the study as a whole had affected their self-reflection or self-awareness in daily life, and whether they were still using any of the journaling skills they had been taught during the experiment. Each interview was audio-recorded and transcribed for further analysis.

Thematic content analysis, which involves iteratively reviewing and labeling interviewees' responses with emerging codes, was applied to all the interview data by two raters working independently. The raters' coding results were then compared, and possible revisions were discussed. This cycle was repeated until the coding scheme was deemed satisfactory by both raters, and inter-rater reliability had reached a reasonable level ($> 91\%$).

5.4 RESULTS

To answer the three research questions, I present the results following their order. First, RQ1 is answered by analyzing conversational logs in the *Suggestion* session. Second, RQ2 is answered by the pre- and post-survey of users' perception, and the interview results are included to explore the reasons for causing the experience. Finally, RQ3 is answered by counting the number of participants that voluntarily practiced the journaling suggestions to understand the lasting effect. The interview is also involved in extending understanding. Figure 5.3 shows two participants' sample dialogues with the chatbot system.

5.4.1 Effects of Chatbot Designs on Users' Journaling Behaviors (RQ1)

To explore RQ1, regarding how users responded differently to versions of the same chatbot with (HC) and without (OC) a human expert (coach) in the same communication channel, I analyzed conversation logs from the suggestions sessions during the *Training* phase (Phase 2). Since the suggestions given in either chatbot setting required the participants to disclose and reflect more about themselves than had been the case in Phase 1, I measured the quantity and depth of participants' self-disclosure. The statistical results are summarized in Table 5.1.

On average, HC participants generated 7.9 messages, and OC participants generated 7.5 messages while practicing each *Suggestion* session's journaling skill. I calculated the word

Table 5.1: Statistical results of RQ1: x means no effect. I did not find interaction effects for all measurements

RQ1		Group (HC v.s. OC)
	Word count	HC > OC (F = 6.60, p < .01)
	Self-disclosure (Information)	x
	Self-disclosure (Thoughts)	HC > OC (F = 29.6, p < .001)
	Self-disclosure (Feelings)	HC > OC (F = 12.12, p < .05)

count of participants' responses and compared the differences using mixed-models ANOVAs. Results showed that the HC (M = 159.41, SD = 16.61) group's mean overall word count was significantly higher than that of the OC (M = 118.31, SD = 17.62) group (F(1, 33) = 6.60, $p < .01$). There was no significant main effect of experiment day, and no interaction effects.

Depth of Self-disclosure

I then coded participants' responses to examine the depth of their self-disclosure. More specifically, first, with regard to *Information*, there was no significant effect of any factor; i.e., neither OC/HC membership nor experiment day significantly impacted how the participants disclosed information to either version of the chatbot. The group averages for informational self-disclosure across all suggestion questions were M = 1.7, SD = .67 for OC, and M = 1.65, SD = .57 for HC.

Second, in the case of *Thoughts*, our analysis revealed a main effect of group (F(1, 33) = 29.6, $p < .001$), with the participants in the HC group disclosing more thoughts. However, there was no effect of experiment day, and no interaction effects. The group averages for self-disclosure of thoughts across all suggestion sessions were M = 1.89, SD = .13 for OC and M = 2.32, SD = .12 for HC.

Lastly, there was a significant main effect of group membership on the self-disclosure of *Feelings* (F(1, 33) = 12.12, $p < .05$) - with the members of HC disclosing feelings significantly more than OC members did - but no main effect of experiment day, and no interaction effects. The group averages for self-disclosure of feelings across all *Suggestion sessions* were M = 1.91, SD = .11 for OC and M = 2.22, SD = .13 for HC. Figure 5.4 shows sample responses with three levels of self-disclosure from participants. These results show that HC group was found to have given longer responses containing deeper *feelings* and more *thoughts* than the OC group during the *Suggestion session* of Phase 2.

	Information	Thoughts	Feelings
Level 1	*Ice cream cake, it is colorful and taste sweet. (P5 , F)*	*I think that many people want to have a cooking skill. (P14, M)*	*When I did an internship in a company, I was responsible for a presentation but I did not prepare well. (P10, F)*
Level 2	*My comfort food is instant noodle because it is easy for me to cook after a busy day. (P6, M)*	*I think I am a good team player because I have good resilience and flexibility. (P33, F)*	*I feel frustrated because I am not punctuality. I am often late for a date with my friends. (P24, M)*
Level 3	*My comfort food is soy milk homemade by my grandma, it's a little sweet and smells good, I felt happy to have it at breakfast when I was a kid. (P18, F)*	*I'm good at writing and I like it! So, I wish I can be a professional writer in the future. So, I keep improving my language skills. (P8, F)*	*I feel guilty that I hesitated to help an old man who was hurt, and I did not give any help. I regretted this and kept wondering if this person would have died because of my indifference. (P35, F)*

Figure 5.4: Samples of participants' responses to the suggestions in Phase 2. The responses were coded to different categories and levels of self-disclosure. Note: Level 1 of Feelings is defined as "No expressing of feelings." [118]

Table 5.2: Statistical results of RQ2 (Group): x means no effect. The study did not find interaction effects for all measurements

		Group (HC v.s. OC)
RQ2-a	Perceived Engagement	HC > OC (F = 8.63, p < .01)
	Perceived Trust	HC > OC (F = 8.28, p < .01)
	Perceived Intimacy	x
RQ2-b	Self-reflection	x
	Self-awareness	x

5.4.2 Perceived Interaction with the Chatbots (RQ2-a)

To better understand why the participants in the HC group responded to the system's suggestions more diligently than the OC group did, I investigated inter-group differences in how the participants perceived their chatbot interactions. More specifically, I examined whether and how each group members' perceived engagement, intimacy, and trust with the chatbot changed during the *Training* phase. Such levels were captured through their responses to surveys conducted before and after the *Training* phase. The statistical results are summarized in Table 5.2 & 5.3. Subsequently, in interviews, the participants were asked detailed questions about their motivations for following (or not following) the system's suggestions; their perceptions of such suggestions; and their impressions of whichever chatbot version they had been exposed to.

Perceived Engagement

The engagement level revealed significant main effects of both group ($F(1, 33) = 8.63$, $p < .001$) and time ($F(1, 33) = 4.76$, $p < .05$), but there was no significant interaction effect; with the HC group reporting significantly higher engagement than OC, both groups'

Table 5.3: Statistical results of RQ2 (Time): x means no effect. The study did not find interaction effects for all measurements

		Time (Pre- v.s. Post-survey)
RQ2-a	Perceived Engagement	Pre < Post (F = 4.76, p < .05)
	Perceived Trust	Pre < Post (F = 16.65, p < .001)
	Perceived Intimacy	Pre < Post (F = 13.9, p < .001)
RQ2-b	Self-reflection	Pre < Post (F = 11.31, p < .01)
	Self-awareness	Pre < Post (F = 7.11, p < .05)

engagement levels generally increased after the *Training* phase ended. To better understand participants' perceived engagement and motivation to follow the suggestions, I asked participants questions in their interviews to explore their experience, for example, *"Did you engage in practicing the journaling skills with the chatbot/coach? Why or why not? How did this change over time?"*

In the interview, ten **OC group** members said that their engagement with the chatbot increased over time because it sent them useful suggestions and prompted them to accomplish something new every day. For example, one participant (P5, F) said:

> "Engagement increased over time because I got used to the chatbot and some of its suggestions were useful. I felt more engaged and it gave me motivation."

Besides, most participants in the OC group felt comfortable about deciding for themselves, on a case-by-case basis, whether to follow the chatbot's suggestions or not follow them. When they did follow them, they generally felt happy and surprised that they had been able to learn something useful from a chatbot. As one put it,

> "Although I did not practice those suggestions a hundred percent, I think I am still on track. When you learn something from doing this, you will feel more motivated. So, my next step is to keep practicing them." (P10, F)

Although practicing new journaling skills seemed to enhance perceived engagement for many participants, four members of the OC group reported that doing so caused them fatigue and annoyance. For example,

> "I felt annoyed because some suggestions were time-consuming to carry out. I did not expect that I would have to expend so much effort." (P17, M)

Moreover, those participants expressed discomfort at certain suggestions they received from the chatbot. As one of these interviewees said,

> "Sometimes, I felt the chatbot was too bossy, especially when it started to give me suggestions. It was okay when the suggestion seemed useful. But when the suggestion was not useful, or the chatbot prompted me to disclose more, I got a bit annoyed. ' It's just a robot, why is it giving me instructions?' That kind of feeling." (P3, M)

In addition, three participants in the OC group reported that their engagement gradually decreased, due to loss of interest in the chatbot system, i.e., the novelty effect [126]. One of them said,

> "I was more engaged in the beginning. This [chatting with a chatbot every day] was a new thing for me. But as I got used to it, my engagement level dropped." (P14, M)

In the **HC group**, according to their survey responses, the HC group felt more highly engaged with the chatbot system than their OC group counterparts did. Most of the HC group participants reported positive attitudes toward practicing the focal journaling skills, and provided two reasons for this. First, much like the OC group, they felt that the content of the suggestions themselves improved their perceived engagement, for example,

> "Although practicing journaling skills was time-consuming, following the suggestions helped me understand myself better. So, though lengthy, the process led me to good results." (P33, F)

Secondly, more than half (n = 11) of the participants in the HC group highlighted the importance of human support, noting that the involvement of a coach increased their willingness to take suggestions seriously. This was because they thought the coach personalized the suggestions for them and would monitor their practicing activities on the chat channel. As one explained:

> "The suggestions were from a coach. I thought the coach might see my responses and give me further suggestions. So, I was more careful about my responses for the suggestions." (P24, M)

Nevertheless, the ostensible involvement of a human coach seemed to negatively affect the engagement of a minority of the HC participants. Three of them noted that it increased their expectations: i.e., that they would receive highly personalized suggestions and feedback. Because our study design did not actually offer such features, these users' engagement with the chatbot was deterred. As one of them put it,

> "The coach gave me suggestions, but they were general suggestions. In fact, I wanted to have more personalized guidance. I felt the coach did not pay attention on my responses." (P19, F)

Moreover, because the coach only offered general suggestions, several mentioned feeling disappointed that 'she' could not really give them personalized feedback or suggestions. For instance,

> "I was kind of disappointed by the low level of her involvement to customize suggestions." (P26, F)

In addition, the impression of human support caused stress to some of the HC users, who felt the coach was judging their answers. As one of them explained,

> "I felt the coach would judge my answers, so I tried to answer the questions as thoroughly as possible, and this made me feel stressed." (P32, F)

While this drove them to implement the suggestions, it also triggered negative feelings when they could not follow the suggestions, for example,

> "I felt sorry when I could not follow the suggestions, because I guess the coach put a lot of effort into designing this chatbot to help me." (P30, F)

Perceived Trust

I analyzed the trust levels reported in the survey, and found a significant main effect of group membership on trust in the chatbot, with the HC group's trust being significantly higher than the OC group's ($F(1, 33) = 8.28$, $p < .001$). There was also a significant, positive within-group main effect of time on trust ($F(1, 33) = 16.65$, $p < .001$), and there was no significant interaction effect. This analysis showed that the HC group showed stronger trust in the chatbot with a coach than the OC group's perceived trust. To explore how the chatbot system and human support influenced users' perceived trust in each of those system personae, I asked some interview questions related to trust, such as *"Please describe your overall trust in the chatbot"* and *"Do you feel your trust in the chatbot changed during the study? If yes, how and why?"*

The resulting data from interviews indicated that participants in **both groups** developed some degree of trust in the chatbot system, citing a few common factors that contributed to this dynamic. First, in line with prior studies [4, 25], small-talk seemed to help the participants develop trust in the chatbot while building relationships with it. For example:

> "I could not trust this chatbot at the beginning of the study because it did not understand me [...] But this feeling of discomfort lessened over time. I felt like I knew each other to some extent because I had some small talk for different things. So I gradually felt I could trust the chatbot more." (P11, M)

Second, four of the 35 participants reported that their trust in the system was rooted in their trust in the research team. As one participant explained,

> "I had high trust from the beginning of the study, and this did not change at any point. I believed the research team, so I could trust in this chatbot system." (P33, F)

In the **HC group**, given that all the journaling suggestions they received allegedly emanated from a human coach, the HC participants' impressions of the chatbot *per se* did not change much across Phase 1 and 2. Most of these participants mentioned that the chatbot's role changed slightly, i.e., from a conversational partner to the coach's assistant, but none of them reported feeling annoyance toward the chatbot. As one of them mentioned,

> "After the coach came in, the chatbot became like an assistant to help the coach to deliver suggestions. I can trust the human coach which also enhanced my trust in the chatbot system." (P28, M)

Furthermore, the presence of a human coach notably increased perceived trust in the chatbot and its suggestions. In the interview, 13 of the 18 HC group members indicated that the introduction of the coach enhanced their trust in the chatbot. It also seemed to have a strong impact on the participants' perceptions of the usefulness of the suggestions they received. For example, one interviewee said,

> "I felt suggestions given by a coach would be reliable. For some suggestions, I wasn't sure if they would be effective. If they hadn't been from a coach, I would have been more suspicious and not followed them." (P30, F)

One of the HC participants explained that

> "If the chatbot itself gave me some suggestions, I would feel like I was getting the suggestions from a website or a book. The suggestions might still be reliable. But, when the suggestions were from the coach, it was different. I felt a human coach was more careful about the suggestions and my answers." (P25, F)

Thus, the presence of a coach appears to have enhanced users' confidence in the suggestions.

Perceived Intimacy

The participants' self-reported intimacy revealed no main effects of group and no interaction effects; however, the main effect of time was significant ($F(1, 33) = 13.9$ and $p < .001$). The analysis revealed that this result was driven by both groups, i.e., that the members of both HC and OC groups perceived significantly higher levels of intimacy with the chatbot at the end of the *Training* phase.

For **both groups**, most of the participants confirmed that they perceived their levels of intimacy with the chatbot as having increased over time, but many in the HC group stressed that such growth in intimacy was not influenced by the introduction of the coach. This seems to be borne out by the fact that both groups' explanations of why their intimacy increased were broadly similar. For example,

> "The chatbot was not like general mobile APPs. The chatbot seems to have life, so I became closer to it. It reminded me and prompted me to finish practice every day; this design made me feel the chatbot cared about me." (P10, F)

The main such explanation was that they interacted with the chatbot every day; the small-talk sessions in particular were linked to enhanced intimacy, on the grounds that the chatbot disclosed a considerable amount of its own 'personal' information during small talk. This finding echoes those of prior research by Lee et al. [24], that reciprocity of self-disclosure can increase intimacy levels. According to one HC group member:

> "My intimacy level increased to some extent because I talked about our own experiences. I felt like I got familiar with each other." (P18, F)

In the **HC group**, five of the participants indicated that they tried the suggestions out seriously under the coach's guidance due to the feeling of companionship with the chatbot and coach, which brought them benefits leading to promote their further motivation to follow the instructions. As one of the interviewees said,

> "I tried following all the suggestions, even those I thought some of them were not useful. And I had positive feelings after following them. [...] I was able to work so hard because I felt I was not alone. There was the chatbot and the coach. I felt I was supported by both of them." (P27, F)

Nevertheless, two of the HC users specifically noted that they would have preferred to receive some suggestions from the chatbot rather than from the coach, because the former understood them better:

> "I respected the coach's expertise and tried to follow her suggestions. But I wished the chatbot itself would give me some suggestions because it had been listening to my stories from the beginning. It knew about my past and my struggles, so it would have been able to work with me more closely." (P31, M)

In summary, the survey results showed that the HC group perceived significantly more engagement with their chatbot, and their trust in the chatbot significantly increased as well; on the other hand, the OC group's engagement with and trust in the chatbot did not show significant changes. In addition, both groups' sense of intimacy with the chatbot significantly increased after the *Training* phase. Our interview results suggest several factors that contributed to the positive changes of participants' perceived interaction: 1) For perceived engagement, participants who perceived higher usefulness of the suggestions experienced more engagement with the chatbot, whereas those who did not find the suggestions useful gradually lost interest in interacting with the chatbot. Although ten HC participants noted that the involvement of a coach increased their engagement in practicing the suggestions, four HC participants also shared that they felt pressure given the presence of human support. 2) Regarding perceived trust, the presence of a human coach seemed to contribute to the significant increase of the HC group's trust because the participants felt confident with the suggestions from a human coach. 3) Regarding intimacy, the interview results suggested that the participants in both groups felt more intimate with the chatbot as both groups had small-talk with the chatbot over days.

5.4.3 Perceived Benefits of Practicing Journaling through the Chatbots (RQ2-b)

As I have seen, the two groups of participants perceived the system differently; both later reported that it had a meaningful impact on them. There was a main effect of time on self-reflection level ($F(1, 33) = 11.31$, $p < .005$), but no significant difference was found between the groups (OC: $M = 4.31$, $SD = .30$, HC: $M = 4.43$, $SD = .34$), and there was no interaction effect. This analysis showed that both OC and HC group members' self-reflection levels increased significantly after learning journaling skills in the *Training* phase.

For self-awareness level, no significant difference was found between the groups (OC: $M = 4.51$, $SD = .23$ and HC: $M = 4.83$, $SD = .26$), but there was also a main effect of within-subject factors ($F(1, 33) = 7.11$, $p <. 05$). Besides, there was no interaction effect. These findings indicate that both groups of participants improved their self-awareness levels after the *Training* phase. Below, I summarize the points about this topic most frequently mentioned by the participants.

Better Awareness of Own Mental Status – Although about half of the participants reported that they had never given full attention to their mood or feelings before the study, most seemed to develop an awareness of their inner emotions and their own strengths by participating in the study. As one stated,

> "At the beginning of the study, my messages tended to be short, just saying that I was tired or sleepy. But in the end, I noticed that I was talking more and saying 'I'm happy.' I also learned many skills, like gratitude, and explored my strengths. These practices made me aware of many things about myself." (P22, F)

Deeper Understanding of Themselves – The step-by-step guidance provided by the chatbot system also seemed to encourage the participants to reflect on themselves and helped them understand themselves better. As one reported,

> "The chatbot helped when it taught me many journaling skills. By trying those skills out, I could understand myself better. By rethinking my past and going through the exercises, I have gained a better understanding of how my current self was formed." (P15, M)

In other words, by following the chatbot's guidance, some users were able to reflect on prior events from different angles and develop new insights about themselves. This echoes a prior study [8] that indicated that writing down their reflections might give users a sense of accountability, where simply thinking about the answers might not.

Providing Different Perspectives – Another benefit of acquiring new journaling skills frequently mentioned by the participants was that the chatbot system reminded them to think about the positive aspects of their lives. For example, participant P2 expressed gratitude toward her parents, supervisor and roommate in the *Training* phase reported in the interview,

> "This chatbot reminded me that there are many things I can take care of, and that I feel appreciated. I don't have to focus on things that make me depressed and sad all the time." (P2, F)

Paths to Better Well-being – Finally, nine participants reported that chatting with the system brought about some changes in their lives. Although I cannot verify that such changes had actually occurred, a sample of the relevant comments is provided below.

> "The chatbot made me think about my true feelings and reorganized my thoughts. For example, the coach suggested me to send a gratitude letter, so I sent one to

my mother. I actually had complicated mixed feelings about her. There were certain things I didn't like about her. I was thankful, and yet I couldn't get rid of my negative feelings about her. But by following the guidance, I came to realize that I didn't understand my mother well." (P20, F)

In summary, our findings show that the chatbots could effectively deliver guidance for journaling skills, and both groups' perceived self-reflection and self-awareness levels increased after the *Training* phase. The interview data also supports this observation, with most participants agreeing that journaling both encouraged them to reflect on their prior experience and increase their self-awareness.

5.4.4 Lasting Effect on Participants' Journaling Practice (RQ3)

To gauge the differences in how our OC and HC chatbot versions affected the participants' journaling practices when following guidance was not required, I first calculated the ratio of participants who practiced the skills per day during the *Free-will* phase (Figure 5.2). I then compared the mean ratio between the conditions using a t-test. Interestingly, the results showed that more OC participants voluntarily practiced the suggestions than HC participants (OC: M = .80, SD = .08; HC: M = .57, SD = .12; $t = 4.15$, $p < .001$). To explore the possible reasons for this difference, I asked the participants to explain their motivations for practicing and quitting the journaling skills during Phase 3.

Reasons of Keeping the Practices

Across **both groups**, participants reported similar reasons for continuing to practice the skills during the *Free-will* phase. A majority of the sample (n = 20) indicated that because they had benefited from practicing the journaling skills, they felt motivated to keep doing so. As one explained,

> "When I followed a suggestion from the chatbot, I was excited to know that even a robot could improve my life! It's not the type of advice I can get in our daily lives, so I keep practicing them." (P15, M)

Some of the participants emphasized that their relationships with the chatbot also encouraged them to keep practicing the taught skills. As one noted,

> "I have been asked about my mood [...] every day, so I gradually felt close to the chatbot. Also, I learned a lot from the chatbot, which gave me a good impression of it. I wanted to keep on using those journaling skills." (P27, F)

Three of the participants indicated that it was helpful when the chatbot reminded them of the journaling skills learned in the *Training* phase. As one of them put it,

> "I forgot about some skills quickly. But by reviewing all the journaling skills, my memories were strengthened, and that gave me confidence that I could successfully improve my mental well-being." (P13, M)

Also, eight participants appreciated the daily prompts and encouragement by the chatbot. One of them said,

> "I think it's not a bad thing to be prompted by the chatbot. It would be hard to have a real human reminding me to practice these skills every day. So the chatbot gave me more motivation to continue working on something." (P33, F)

Moreover, many participants were also motivated to follow the system's suggestions by reminders the chatbot sent to them. One of them shared prior experience with cultivating journaling skills; the interviewee said,

> "I once tried to do something similar, but I could not keep doing it on my own. However, when the chatbot encouraged me and sent me reminders, I felt motivated to follow its suggestions. This is quite different from practicing alone." (P18, F)

Reasons for Quitting Practice

The participants who did not often practice in the *Free-will* phase gave several reasons for this, which I summarize below. First, many of them felt tired of journaling after the *Training* phase: as one of them stated,

> "I wanted to take a short break from what I had been doing." (P19, M)

Second, practicing journaling skills was deemed optional in Phase 3, and five participants reported seeing this aspect of our system's design as giving them an excuse to skip it. As one mentioned,

> "I feel that if I am given too many alternatives I will just give myself excuses. Especially when I was a little busy with other stuff, I might skip the practice." (P9, F)

Especially, the **HC group** participants gave group-specific reasons for not practicing in the *Free-will* phase. Six of them reported that because there was no longer a coach monitoring them after the *Training* phase, they did not feel it was necessary to keep practicing. One said,

> "I stopped practicing the system's suggestions after the coach stopped giving new ones, because no-one was monitoring my responses and I just wanted to skip that." (P32, F)

Lastly, two members of the HC group noted that practicing the same journaling skills they had learned in Phase 2 a second time would not have brought them new insights, and thus, they tended not to practice in Phase 3. As one of them put it,

> "I think those journaling skills are still useful, but less useful than when I practiced them the first time." (P34, M)

In summary, I found that OC group participants practiced the journaling skills more than HC group participants in the *Free-will* phase (Phase 3). Based on the interview, the results imply that this difference may be attributed to the absence of a human coach to monitor and interact with the participants in the *Free-will* phase, and the decrease in perceived benefit from practicing the journaling skills again.

5.5 DISCUSSION

5.5.1 The Impact of Changing or Incorporating Chatbot Identities

As noted earlier, the participants did not receive any suggestions during Phase 1 (*Warm-up*) - they received suggestions for practicing journaling skills after they completed Phase 1. This meant that the chatbot, who had been a conversation partner or listener for both groups, switched its role to either a coaching role for the OC group, or a mediator between the user and the human coach for the HC group.

For the OC group, when the chatbot changed its role from a conversational partner to a coaching role, there were two types of responses. On one hand, some OC participants were happy to learn something new from the chatbot, and most of them did not feel pressure from the chatbot and enjoyed having the company. On the other hand, some participants reported that the chatbot had become bossy and annoying as the chatbot started to give suggestions in the *Suggestion* session. According to the participants, they did not believe the chatbot understood their struggles and real human lives, so they did not feel comfortable with the chatbot's suggestions. They felt that it would be overcorrecting to adopt the chatbot's suggestions. This is different from the participants in the HC group, who received exactly the same suggestions, yet none reported such feelings toward the human coach nor the chatbot.

On the contrary, none of the HC participants reported any complaints about the chatbot's mediating role. Most of them seemed to believe that their main interlocutor was now a human coach: that the suggestions they were receiving were from that person, and that their responses would be monitored by the coach. Comparing the experience of our two groups suggests that a drastic change in chatbot role may cause participants confusion and discomfort. This observation echoes prior research [23] which found that users felt uncomfortable interacting with a conversational agent using the same personality to serve multiple, completely different, roles. Our results suggested that, adding another persona, the human coach in this case, has the potential to mitigate discomfort. That is, the participants valued the system's suggestions differently due to the ostensible involvement of a human coach; while participants in the OC group tended to judge the value of the system's suggestions on their own, most participants in the HC group tended to believe that the suggestions from the human coach were all valuable, even when they felt some suggestions were not useful. In other words, they trusted the authority of the coach, and were willing to follow the coach. This seems to lend support to Mohr et al.'s [91] finding that legitimacy derived from users' perceptions of a coach positively affected their acceptance of that coach's demands.

Meanwhile, our findings also showed side effects of integrating a human coach to deliver guidance. First, although I simulated the coach who gave suggestions asynchronously, some participants in the HC group reported perceiving high pressure, and/or that their interaction with the chatbot was monitored by the human coach in detail. Several participants stated that they felt stressed and apologetic when they did not have time to closely follow the coach's suggestions, due to the time and effort the coach must have spent in crafting them, based on their progress. Additionally, some participants in our HC group expressed high expectations regarding the customization of the system's suggestions and expressed disappointment accordingly. This echoes Kocielnik et al.'s [8] finding that individuals who held high initial expectations about a chatbot tended to be disappointed. However, in our study, this effect was salient only for the HC group.

5.5.2 Compliance & Sustainability of Practicing Skills through a Three-phase Study

Our study findings reveal some challenges for deploying chatbots in real-life situations, such as establishing relationships between the users and the chatbot [24, 92] and sustaining users' interaction with chatbot [126]. Regardless of the positive or negative effects of time, our findings suggest the importance of conducting a longitudinal study when testing chatbot technologies. Below, I discuss our three-phase study setup and the unique value of this design.

Our study started with a nine-day *Warm-up* phase. I set up this phase because prior work suggested that building trust and intimacy toward the chatbot could motivate users to self-disclose more deeply over time [11, 24]. While our interview findings echo previous work regarding the important role of time in establishing social relationships between the participants and the chatbot [24, 176], our data also showed that some participants experienced a novelty effect [126]: i.e., that they were highly engaged in chatbot conversation at the very beginning, but that this excitement gradually decreased thereafter. Our study results show that the participants' perceptions and attitudes toward the chatbot changed as they interacted with it across the different temporal phases of the experiment.

Through the three-phase study, I were able to find that the HC participants had higher levels of dependency on the chatbot. However, this could have diminished their motivation to keep journaling once it became an optional activity in the *Free-will* phase. According to the investment model of long-term engagement [92], users may stop using an agent if they perceive higher cost and lower benefit. Our findings are in line with the theory [92], insofar as the HC participants perceived higher cost for practicing the journaling skills in the *Training* phase, which demotivated them to continue practicing in the *Free-will* phase. On the other hand, the OC participants had both lower self-disclosure and lower engagement in the *Training* phase than their HC-group counterparts. The lack of pressure to follow system suggestions that the OC participants perceived could result in a lower perceived cost of continuing the practices. Therefore, compared to the HC participants, their willingness of continuing the practices was less negatively impacted in the *Free-will* phase.

5.5.3 Design Implications

Our study uncovered both benefits and drawbacks of integrating human support into human-chatbot interaction. These shed light on future chatbot designs both for practicing journaling skills to improve mental health and for skill training in broader contexts.

The major benefit of integrating expert's advice into human-chatbot interaction is that the expert's involvement resulted in participants' higher engagement. In our context, it was measured by participants' journaling efforts (length and depth of their journaling content) and their perceived interaction from the perspectives of engagement, trust, and intimacy. Prior research has shown that compliance with system requests is an important first step towards users' achievement of further positive behavioral change [177]. Shi et al. [14] found that when users identified a chatbot as a human, they would think that the conversation was more engaging and have better outcomes. Our findings extend our understanding by examining whether users can be persuaded to follow suggestions from a coach even when the suggestions are not customized for individuals and provided asynchronously through a coach's agent; however, the integration of a human coach might lead the participants to think that they were monitored and cause them to feel pressure.

Scholars already suggest that using a chatbot as a coach can guide users toward a healthy lifestyle and activities, e.g., [8, 9, 89]. Additionally, coordinated with other technologies such as physical sensing and machine learning, future virtual agents, e.g., Alexa and Google Assistant [**?**], could automatically track users' behaviors and introduce proper expert services. When human experts, e.g., professional coaches, counselors and healthcare providers, are scarce or have limited availability, our proposed HC design could be adopted to help a human expert deliver their suggestions to the clients asynchronously, and the chatbot could play a role to help collect and track the clients' information to assist human experts in offering more persuasive guidance than that given only through chatbots. Future designs could also consider defining the human expert agent as a crowd-powered expert to flexibly attribute limited human expert resources among users.

Finally, I found that the chatbot versions in the OC and HC conditions had specific advantages and disadvantages when it came to delivering guidance. Future practitioners and designers could consider the factors, e.g., switching chatbot identities and sustainability of skill practices, when using chatbots for different training purposes. For example, if the specific skill requires close compliance in a short period of time, incorporating human experts may be more effective in the training phase. Conversely, if the training requires users' long-term engagement and a light touch with the experts, it may be preferable to have a companion chatbot lead the interaction. More varied design options need to be further evaluated in the specific application contexts.

5.6 LIMITATIONS

There are several limitations of this work. First, to keep the two groups comparable, I controlled the chatbot responses when providing suggestions for the users. There was not an actual human person monitoring the participants and personalizing suggestions for them in this study. Thus, our study design did not perfectly match the participants' expectation of customized expert support. However, the result could also reveal how users might react when a system does not work as they expect. Still, it will be beneficial to further examine the impact of offering personalized guidance in further research. Future research could also evaluate how synchronous communication with real human experts may impact user experience. In addition, I did not explore the different designs of chatbot personas (i.e., name and icon) to represent the coach in this work. Future work could be focused in this direction by examining the effect of controlling the presentation of a human expert's persona.

Second, I gave a relatively easy task to the participants to learn (journaling skills). I used the depth of self-disclosure and word count to measure how the users followed the suggestions. Because journaling needs users to disclose themselves, the depth of self-disclosure level could reflect how users put effort into responding to those suggestions. Users' daily experience and behavior might be affected over time. In our study, I did not focus on behavioral change.

Finally, although I believe that the differences between the OC and HC groups were not affected by compensation as our OC and HC participants were compensated at the same rate, the engagement level of the participants taken as a whole was surely affected by compensation. Future work should consider deploying the system without compensating the users for a longer-term span in different contexts.

5.7 CONCLUSION

In this chapter , I explored chatbot designs with and without integrating human support for delivering guidance to users, along with a three-phase study I conducted to examine the impact of human expert support on human-chatbot interaction. I focused on investigating the impact on users' experience and efforts involved in following the guidance of journaling skills. Our findings show that the chatbot integrated with human support (HC) enhances users' self-disclosure, perceived engagement and trust in the *Training* phase; however, it might also make people feel pressured. Although the group without human support (OC) showed less engagement in the *Training* phase, they remained interested longer in practicing journaling skills. Our research extends the understanding of human-chatbot interaction by providing insights on future chatbot designs with the support of human experts to deliver guidance in broader contexts.

5.8 APPENDIX

Figure 5.5 shows an additional example of journaling guidance. This chapter presents the surveys used in the study.

5.8.1 Appendix A - Self-Reflection Survey

- I don't often think about my thoughts.
- I rarely spend time in self-reflection.

Bot: Today, we're going to practice Expressive Writing again.

Bot: Please think about an acquaintance, friend, or family member you love/like, but have some regret about your relationship with. I would like you to write about your deepest thoughts and feelings about this topic.

Bot: So, who is the person you want to write about?

User: My ex-boyfriend

Bot: Okay. Please list at least three things that you hate about them.

User: He sometimes was selfish, and he spent too much time playing videogames. I felt he didn't love me back

Bot: I see. Now, let's write down things you appreciate about this person.

User: He was smart and had a good sense of humor. He always tried to comfort me when I felt sad. He also gave me help with my coursework.

Bot: Next, still thinking of the same person, pay attention to your feelings. When you're ready, begin writing loving memories of them, in as much detail as you can.

User: I was very happy when we were together, but this was not as often as I wanted. I was very sad about that. I am grateful that he was with me and very kind to me at that time, and I want to thank him for many things. But I think, if I miss him too much, it will be painful for me, because eventually he will be with someone else, and now I only want to be his friend.

Bot: Finally, please reflect and consider what you would tell this person, if it were possible to do so.

User: I want to thank him for being very kind to me, but we should realize that we cannot be together anymore because of distance and various other things. I hope he is happy now.

Bot: So, what will you do in future to maintain or improve your relationship with the person you mentioned, or other people?

User: If I get a boyfriend, I will try my best to maintain the relationship

Bot: Thank you for sharing your thoughts and feelings!

Figure 5.5: Samples of a participant's responses to the journaling guidance in *Training session*

- I frequently examine my feelings.
- I don't really think about why I behave in the way that I do.
- I frequently take time to reflect on my thoughts.
- I often think about the way I feel about things.
- I am not really interested in analyzing my behaviour.
- It is important for me to evaluate the things that I do.
- I am very interested in examining what I think about.
- It is important to me to try to understand what my feelings mean.

- I have a definite need to understand the way that my mind works.
- It is important to me to be able to understand how my thoughts arise.

5.8.2 Appendix B - Self-Awareness Survey

- Right now, I am conscious of my inner feelings.
- Right now, I am reflective about my life.
- Right now, I am aware of my innermost thoughts.

5.8.3 Appendix C - Engagement Survey

- I was so involved in this experience that I lost track of time.
- The time I spent using the chatbot just slipped away.
- I was absorbed in this experience.
- I felt frustrated while using this chatbot.
- I found this chatbot confusing to use.
- Using this chatbot was taxing.
- This chatbot was attractive.
- This chatbot appealed to my senses.
- Using chatbot was worthwhile.
- My experience was rewarding.
- I felt interested in this experience.

CHAPTER 6: INTEGRATING SOCIAL LEARNING INTO HUMAN-CHATBOT INTERACTION

6.1 INTRODUCTION

Prior work suggests incorporating *social learning* into systems, e.g., allowing people to observe others performing a target behavior, may motivate users to perform target actions [108]. *Social learning* [178, 179] is a process of human acquiring new behaviors and concepts by observing and imitating others. For example, people seek advice or experiences from online healthcare communities and social media because peers' successes and tips can facilitate social learning that boosts the adoption of health-promoting behaviors, e.g., [107, 110]. Therefore, integrating social learning into chatbots may be promising to boost user engagement and motivate them to follow system guidance. One the other hand, previous research also shows that there are potential risks for facilitating social interaction among users. For example, when people interact with each other, they may overshare personal information [180] or falsely internalize unsupported messages posted by others [111, 113].

As a result, people may end up disclosing less information about themselves or perform counterproductive behavior. This may not be desirable for human-chatbot interaction, because one of the key advantages of using a chatbot is to provide people a place to disclose about themselves without worrying about being judged by others [24, 181]. Without users' self-disclosure, guidance for skill training may not be effectively delivered. Although recent studies have explored chatbots' potential for mediating social interaction among users [6, 116]), the effects of incorporating social learning into human-chatbot interaction is under-explored.

To address this gap, I designed a chatbot that incorporates social learning into one-to-one human-chatbot interaction for practicing journaling skills. Journaling was selected as a learning topic because it is a common practice that helps people improve mental well-being [163] by tracking personal mental status and enhancing self-reflection [63]. More specifically, the social learning component allows users to receive peers' sharing about their successful journaling practices or experience through the chatbot without directly interacting with the peers.

To evaluate the effectiveness of integrating social learning in improving human-chatbot interaction and their perceived engagement, I deployed two chatbot designs. Both chatbots delivered training guidance for improving journaling skills, with one design guiding the user itself through the dyadic interaction, and the other design offering a social learning component. I then conducted a mixed-methods study, where 34 participants were divided into two groups, each of which was provided with one design and encouraged to use the chatbot over three weeks. The social learning component was enabled by sharing narratives about users' experience of practicing the journaling skills and was carefully designed without violating people's privacy. These narratives were collected by the chatbot and were shared anonymously within the group. I logged and analyzed participants' dialogue with their chatbots to explore how they followed the journaling guidance, surveyed their perceived engagement and self-reflection, examined the inter-group and differences among individuals, and conducted exit interviews with all the participants to triangulate the results.

6.2 RESEARCH QUESTIONS

The aim for incorporating social learning into a chatbot system is to enhance user engagement and motivate them to comply with the chatbot's guidance. Based on literature review and considerations above, I conduct user studies to address the following three main research questions. I first explore whether and how the users share their practicing experiences through the chatbot. Thus, our first research question is:

- **RQ1**: *Do users share (or not share) their practicing experiences with their peers through a chatbot? If they do, what kind of experiences would they share?*

Previous research [107, 110] indicates that social learning could enhance users' adoption of health-promoting behaviors. Thus, I ask:

- **RQ2**: *How does our proposed social learning design influence users' interaction (e.g., practicing journaling skills) with the chatbot?*

In RQ2, I expect that participants who read peers' success experiences of practicing the suggested journaling skills would be encouraged to make more in-depth responses than those who only interact with the chatbot without knowing others' experiences. I believe that the findings of RQ2 provide insight into how the introduction of social learning into human-chatbot interaction affects users to follow and act on chatbot's guidance.

Finally, as discussed above, prior work showed that social learning may improve users' motivation to change their behaviors and perceived engagement [103, 104]. Therefore, I ask:

- **RQ3**: *How does the social learning design influence users' perception of their interaction with the chatbot?*

I expect that the social learning feature provided by the chatbot would improve people's perceived benefits of practicing the journaling skills (e.g., self-reflection) as well as perceived

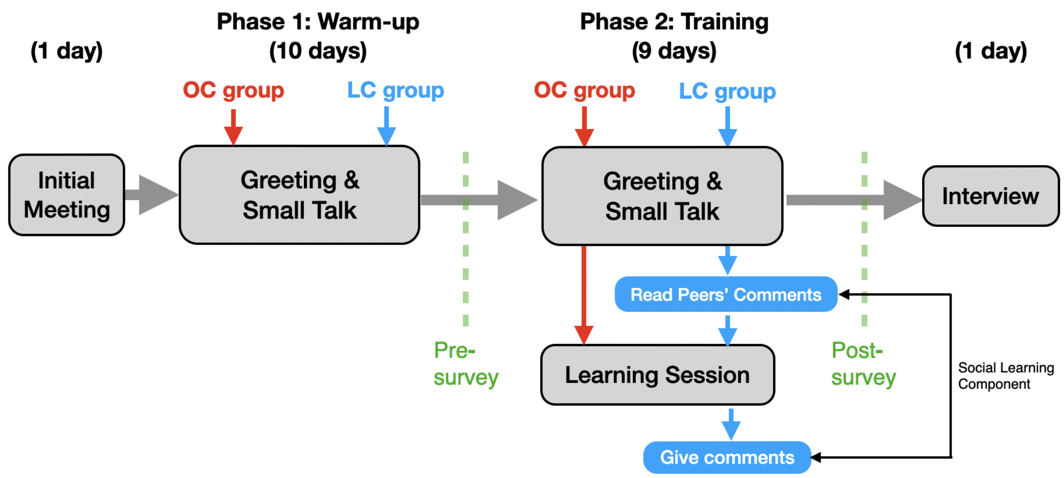


Figure 6.1: Study design: Our study consisted of two phases: **Warm-Up** and **Training** phases. **OC** represents the condition of Only Chatbot, and **LC** represents the condition of Chatbot with social learning. In this study, the social learning means that the users learn from reading their peers' commentaries on the experience of chatbot-guided journaling-skills practice, the peers' experience which appeared only during the Learning session (in *Phase 2: Training*) before delivering journaling suggestions (Learning session). Other chatting sessions were kept the same for the both groups.

engagement with the chatbot. Addressing RQ3 would help understand incorporating social learning into human-chatbot interaction affects people's attitude toward this interaction.

6.3 METHOD

To examine the effects of chatbot-mediated social learning, I created two chatbot variants that delivers journaling guidance to users: one with social learning **(LC)** and the other without **(OC)**. Specifically, a social learning component is incorporated into a text-based chatbot system that provides people guidance on practicing their journaling skills. Here, I chose to have participants practice journaling skills with the chatbots for several benefits: 1) By developing journaling skills, people can better reflect on themselves, identify negative thoughts and promote positive thinking [146, 163]. Ultimately, it can reduce anxiety and enhancing one's mood and well-being [63, 167]; 2) journaling often involves participants' self-disclosure to chatbots, and previous findings showed that self-disclosure to chatbots can boost people's well-being [3, 86]; 3) on a technical level, it allowed us to observe the users' behavior when practicing journaling skills by tracing their interaction with the chatbot.

Bot: We are going to practice gratitude journaling. It is important to create a record of things you are grateful for by writing them down; it is not enough simply to do this exercise in your head.

Bot: Here are some comments other participants made after they tried gratitude journaling:

1. *"This journaling practice helps me find the good and praise it [...]"*
2. *"I reflect on what I am grateful for in detail. This makes it much easier to understand my true feelings."*
3. *"I like this practice. It guides me step-by-step to think deeply [...]"*

Introduce peers' comments

Bot: If you are ready, please reply "OK" to start the practice.

User: Ok

Bot: Great, let's start! Write down three things that went well for you today.

LC condition

Bot: We are going to practice gratitude journaling. It is important to create a record of things you are grateful for by writing them down; it is not enough simply to do this exercise in your head.

Bot: If you are ready, please reply "OK" to start the practice.

User: Ok

Bot: Great, let's start! Write down three things that went well for you today.

OC condition

Figure 6.2: This figure shows the main difference between LC (Left) and OC (Right) conditions. In LC condition, the chatbot introduced other peers' comments to the user.

6.3.1 Study Design

Our study consisted of two phases, **Phase 1: Warm-up** and **Phase 2: Training** (Figure 6.1). The purpose of *Phase 1* was to familiarize the participants with communicating with the chatbot and to remove novelty effects. Specifically, the participants in this phase engaged in a *Greeting and Small-talk* conversational session with the chatbot each day. This design was inspired by prior research [4, 24] which indicates small talk could enhance users' self-disclosure and perceived engagement.

In *Phase 2*, a *Learning session* was added after each day's *Greeting and Small-talk session* (Figure 6.1). In this *Learning session*, the chatbot gave the participants some suggestions and guidance about how to improve their journaling skills. On the first day of *Phase 2*, the chatbot told the participants in both conditions:

> "From today, I am going to give you some guidance to learn some new journaling skills, which could help you 1) gain a better understanding of your own mental well-being; and 2) help you better reflect on yourself and improve your happiness."

While *Phase 1* was the same for all participants, *Phase 2* had two different experimental conditions. The first was a **Only Chatbot (OC) condition**, where the users learned new journaling skills from the chatbot. The second was a **social Learning Chatbot (LC)**

condition (Figure 6.2), in which the users' peers shared their experience and tips, as a supplement to the chatbot's journaling-skills guidance. Our rationale for including peer commentaries (narratives) of practicing journaling skill was closely linked to our study's aim of improving user engagement and encouraging them to follow the guidance. Specifically, peers' positive experiences and suggestions about how they have practiced journaling skills with the chatbot might play as positive role models and motivate users to follow the journaling suggestions. This design is inspired by previous research [107, 182] which suggested the importance of facilitating the sharing of experience to advise each other.

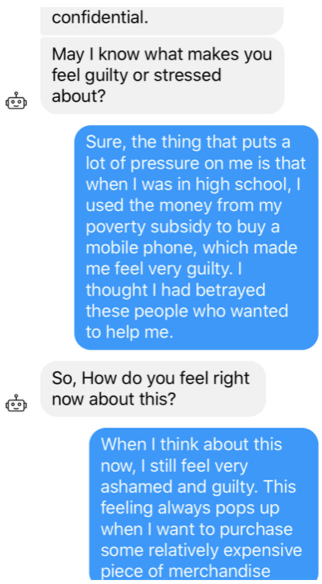
In the LC condition, the chatbot introduced three positive commentaries from peers per day during *Phase 2*, right before providing its journaling-skills guidance (Figure 6.1). One example of such an introduction is as follows: *"These are some experiences/comments from other participants after they tried the journaling skill: 1. [...], 2. [...], and 3. [...]"*. All commentaries were chosen at random from the positive commentary pool, which was classified by two coders. After reading peers' commentaries on the journaling skill, the chatbot asked the LC participants to practice the same journaling skill (Figure 6.2(Left)). When the LC participants completed practicing the skills, the chatbot encouraged them to share their own experiences. The participants could decide whether their experience would be shared with their peers. To ensure that the shared commentaries did not include any information that might reveal any peer participant's private information, two coders checked the LC participants' commentaries every day.

In summary, the difference between the two conditions (LC and OC) was that the LC participants were able to 1) read other LC participants' (peers') successive commentaries of trying out the journaling skills introduced by the chatbot, and 2) provide their own experience of the same kind, for anonymously sharing with other peers. By comparing the participants' journaling-skills practices and experiences between the two conditions, I aimed to investigate differences in the participants' attitudes and responses toward following the journaling guidance given by the chatbot. This research was reviewed and approved by our institutional review board.

6.3.2 Chatbot Conversation Design and Implementation [1]

Our chatbot, whose interface (the same for both conditions) is depicted in Figure 6.3 (a), was adapted from a commercially available instant-messaging platform. The participants generally interacted with the chatbot via free text. At a few points, it gave them multiple-choice options: e.g., its questions about users' moods.

[1] Please refer to Chapter 3.3 for more implementation details.



(a) Chatbot Interface

Chatbot:	Choose someone in your life who has been helpful and kind to you, but to whom you haven't had an opportunity to express your gratitude
User:	Okay!
Chatbot:	Who is the person?
User:	My dad
Chatbot:	So, directly addressing that person, describe specific things that they have done that made you grateful
User:	I really need to say thank you to him, and I have one thousand reasons to do so. I have many examples of why my gratitude should be given to him. Without him, the path of my life could have been diverted in a really bad direction
Chatbot:	Okay, how has this person's behavior affected your life?
User:	I remember that day when I had my nose broken in a fist-fight in high school. At first, I decided to take revenge and hurt that guy. I am now glad that my dad stopped me, and that guy became my friend instead of an enemy. My dad made me realize it, and always give me the right guidance

(b) Example Dialog

Figure 6.3: (a) This figure shows our chatting interface. The chatbot gives a suggestion/guidance to the user. (b) Sample Chatbot's conversations with the participants during a *Learning session*.

The chatbot's conversation was enabled by a combination of Manychat and Google Dialogflow. The former was used in the generation of daily chatting tasks, task-reminder issuance, and monitoring of task completion. Our control over the experimental conditions was assured by the inclusion of predefined questions and chatbot responses in the daily chatting tasks. The primary purposes of Dialogflow, which incorporates natural-language processing, were to deal with unexpected user input and enhance user-chatbot conversations' naturalness. For instance, if a user said *"I feel happy today,"* Dialogflow prompted the chatbot to ask a follow-up question such as *"What made you feel good today?"*; or, if a participant inquired about the chatbot's own preferences or moods, Dialogflow helped it to give brief but meaningful responses - or, where this was not possible, to ask for clarification of the question, or request that the user focus on the learning topic instead. Additionally, if a user appeared to become stuck three times in a row, the chatbot system moved to the next conversational topic.

Conversation Flow

Our chatbot's conversation flow design drew on those of actual and proposed chatbots[2]

[2]https://www.wysa.io/ and https://woebothealth.com/

in the field of healthcare [34]. Generally, such systems' conversations commence with a greeting before moving on to deeper conversational topics, including but not limited to users' thoughts and feelings. Small-talk with chatbots can enhance their users' self-disclosure and engagement [4, 24, 86], and it was incorporated into our design for those reasons. Its two session types are described in further detail below.

Greeting and Small-talk Session - Given our chatbot's ostensible purpose, i.e., training its users in practicing journaling skills, its greetings focused on the participants' moods, along with day-to-day experiences that they found triggered emotions. After asking a few questions about such matters and receiving the answers, the chatbot made further queries aimed at prompting self-reflection, before moving on to between three and five journaling-related questions. In this greeting stage of the session, the chatbot served primarily as a "listener," either encouraging the user to elaborate on his/her own answers, or providing simple, vague feedback like *"I see."*

Inspired by prior research findings that a chatbot's self-disclosure can boost its users' deep self-disclosure [24], our chatbot shared its 'personal' stories (adopted from previous studies [24, 130]) once the greeting stage of this type of session had been completed. The topics of such stories, and the chatbot's questions and responses, were exactly the same for the two experimental conditions.

Learning Session - In the learning sessions, which in *Phase 2* followed the Greetings and Small-talk sessions (Figure 6.1), users were given guidance about new journaling skills, adopted from previously published materials [63, 164] aimed at improving well-being via self-reflection. One study in the field of positive psychology [167] has suggested that some interventions seem to work well by simply following instructions (Figure 6.3 (b)); and gratitude journaling appears to be a case in point [164, 168].

Our chatbot system was designed to facilitate that prior studies' guideline [165, 169] to develop users' journaling skills. To aid users' acquisition of the journaling skill, for example, our chatbot began by explaining that skill's benefits, and then asked them to think of a person they were grateful to. Then, it gave the following instructions/suggestions: *"It is important to create a record of your items by writing them down; it is not enough simply to do this exercise in your head,"* and *"Write down three things that went well for you today."* Next, the chatbot encouraged the participants to provide more specific and detailed information: *"Write down exactly what happened in as much detail as possible, including what you did or said"*, and *"Include how this event made you feel at the time and how it made you feel later."* Finally, the chatbot asked the participants to wrap up, as follows: *"Explain what you think caused this event."* As such, the chatbot's role was to keep guiding its users toward the next step, while at the same time giving instructions intended to stimulate deep disclosure of their thoughts and feelings.

The participants were allowed to skip any instructions/guidance they did not want to follow and any questions they did not want to answer, and the lengths of their responses were not restricted. At any given time, a participant was assigned only one skill: i.e., gratitude journaling [165, 168] or expressive writing [169, 170] or the "best possible self" exercise [171, 172]. Once such an assignment had been made, the participants learned that skill over three consecutive days. Again, the chatbot system behaved in exactly the same ways across the OC and LC groups: that is, gave the same greetings, small-talk topics, and guidance in new journaling skills. The only difference between the two groups (OC and LC) was that the LC participants could share their experience with their fellow LC group members, and read and learn from these peers' experiences, which is the social learning component I explore in this study.

6.3.3 Participants

Participants were recruited via a Japan's University electronic bulletin board, and selected if they had the following attributes: 1) 18 years old or older; 2) fluent in written and spoken English; and 3) a score below 13 on the Kessler Psychological Distress Scale (K-6), i.e., lacking any urgent mental-health issues [133]. The members of the initial participant pool were told the length of the study; that they could withdraw from it at any point; and that attending a post-study follow-up interview would be optional. In the event, 34 participants were recruited, all undergraduate or graduate students, ranging in age from 21 to 29 (M = 24.6, SD = 2.4). This sample was then divided into two halves balanced by K-6 score and gender, due to the possible impact of these factors on self-disclosure [133]. Participants S1-S17 formed the OC group, and S18-S34, the LC group, each of which comprised seven males and 10 females. All were familiar with Facebook Messenger and/or similar messaging platforms. Compensation for participation in the experiment was US $150 for three-weeks study, and participation in the interview, an additional US$20. On average, I paid $7.5 for each chatting task (20 chatting tasks in three weeks) which reflects the local part-time rate. Each participant was interviewed separately, face-to-face, for 45-60 minutes; with their permission, these sessions were recorded and transcribed.

6.3.4 Study Procedure

During an initial meeting that all participants attended in person, the researchers walked them through the configuration of the chatbot on a mobile device of their choosing. Also, at this point, the researchers informed the participants 1) of their rights to refuse to answer the chatbot's questions and to refuse to follow its instructions; 2) that their chatbot conversations would be recorded and analyzed for research purposes; and 3) that they were permitted to withdraw from the experiment at any time. After they signed the relevant consent forms, a 10-minute chatbot practice session was held.

Participants were not told which of the two experimental conditions (LC and OC conditions) they had been assigned to. They were, however, instructed not to discuss their respective chatbot interactions with one another while the experiment was still ongoing. They were encouraged to complete a 15- to 30-minute chatting task daily, but were told that they would be allowed to skip up to three chatting tasks during any seven-day period.

The study's first day was deemed a practice day, on which the participants could familiarize themselves with the chatbot system, which also served the research aim of removing novelty effects. The *Warm-up* and *Training* phases were of equal lengths, i.e., nine days each. On each of the 19 days of the experiment, all users received a chatbot prompt that a new task was ready for them. To ensure that their input was based on fresh memories, however, the "chatbot day" began at 6 p.m. and ended at noon on the following calendar day. Participants could access the chatbot outside of these hours - i.e., between 12:00 p.m. and 5.59 p.m. on any day - but it would only give them brief, uncomplicated replies, which had been designed by the researchers to have minimal or no impact on their perceptions of the experiment or the chatbot.

As well as completing his/her assigned chat tasks, each participant filled out two surveys, one just before and the other just after *Phase 2* of the experiment. The survey content was the same at both time-points, and sought to capture perceived engagement and self-reflection levels (Figure 6.1), as explained further below.

6.3.5 Data Analysis

To answer the research questions, I retrieved participants' conversation logs (in the *Learning session* at the *Phase 2*); examined their survey input; and analyzed participants' interview feedback.

Conversation Logs

All logs of participants' conversations with the chatbot were collected, and formed the basis of inter-group comparison of adherence to journaling-skills advice, as measured by the depth of users' self-disclosure in the three categories proposed by Barak and Gluck-Ofri [118]: i.e., *Information*, *Thoughts*, and *Feelings*, each of which is divided into three self-disclosure levels. *Information* is assigned a sensitivity level based on what factual details are disclosed about the participants themselves; *Thoughts* statements express the participants' personal opinions; and *Feelings* are their expressions of emotional reactions to people, their behaviors, or other occurrences [118]. Because there is some overlap in the categories of Information, Thoughts, and Feelings, our two independent raters coded each chat-log statement for each of these three categories on a scale of 1-3; examples are shown in Figure 6.4. Each statement had three codes corresponding to the three categories, which means that a statement could belong to multiple categories.

Two raters were blinded to the two conditions (LC and OC). They were given a set of chatbot utterances and user responses to understand the conversational context. These two raters coded a subset of the chat-log items and discussed the differences in their codes until a consensus was reached. Then, they coded the remaining data, achieving a final inter-rater reliability of 90%. To assess how the OC and LC chatbot variants may have influenced users' writing quantity and depth of self-disclosure, mixed-model analysis of variance (ANOVA) and Tukey's test of significant difference were performed sequentially, with the chatbot's journaling suggestions treated as a random effect; self-disclosure levels and word count as dependent variables; and group membership (OC and LC groups) and experiment dates as independent variables.

Survey

Self-reflection: To see whether participants felt the effects of following the guidance, I measured the participants' self-reflection levels. As described earlier, learning journaling skills encourages users' self-reflection [174]. Again, 12 items were adopted from previous studies [174] to measure the participants' self-reflection. The items included *"I frequently examine my feelings"* and *"I frequently take time to reflect on my thoughts."*

Perceived Engagement: Participants' perceived engagement levels could be correlated with both their performance and retention of information. Based on the literature [173], 12 items were used to measure the participants' perceived engagement. Two sample items include *"The time I spent using the chatbot just slipped away"* and *"I felt interested in chatting with the chatbot."*

For analysis, I used mixed-model ANOVA with two dependent variables and two independent variables. The two dependent variables were: self-reported self-reflection levels and perceived engagement. The two independent variables were: time point (i.e., before v.s. after *Phase 2*) and group membership (LC v.s. OC).

Interview

The interviews were semi-structured, and focused on the participants' chatbot experiences, notably 1) their daily practices, 2) impressions of the chatbot, and 3) what journaling skills they had learned. With regard to the latter category in particular, I asked if they felt comfortable receiving journaling training from a chatbot; whether they felt doing so was worthwhile; what they had learned from doing so; and if these feelings/attitudes had changed over the course of the experiment. I also asked them how much effort it had cost them to learn from the system during *Phase 2*, and how much they had disclosed when journaling. Moreover, the participants were asked detailed questions about if, why, and how they followed the chatbot's journaling suggestions; and their responses to these suggestions. Additionally, the LC participants were asked about how they perceived their peers' shared experiences, whether they were affected by reading them, and how they shared their own experiences with their peers.

Two coders independently reviewed and coded the transcribed interview data using thematic content analysis with inductive approach [183]. They familiarized themselves with the interview data and generated initial codes. After they reached the point of high-level theme generation, they compared their codes and discussed revisions, repeating this cycle until both deemed the coding scheme to be satisfactory. At that point, inter-rater reliability was reasonable, i.e., over 86%.

6.4 RESULTS

To answer RQ1 and RQ2 about participants' interaction with the chatbot with or without social learning component, I examined participants' dialogue in the *Phase 2: Learning session* with their chatbots. To address RQ3 about participants' self-reflection and perceived engagement with following the chatbot's journaling guidance, I analyzed participants' survey results and interview data.

6.4.1 How Users Share Experiences of Practicing Journaling Skills (RQ1)

To understand how our social learning mechanism was performed in the **LC group**, I conducted the following analyses. During the *Phase 2*, after the LC participants completed a learning session, the chatbot encouraged them to provide some commentaries about that day's journaling-skill learning experience. As noted above, the participants could opt not to share such comments with their peers after writing them, if they wished. In all, I collected 140 of these commentaries from the 17 LC participants over the nine days of *Phase 2*. Given all the LC participants' commentary items, 113 were authorized for sharing and 27 were withheld from sharing with peers.

By applying thematic content analysis to this material, I developed five themes, and categorized each item of commentary into the themes. Note, one commentary can have one or more themes. Below, I elaborate on these themes, and give some examples of each.

- **Positive Experience -** Many participants recounted success stories about practicing the journaling skills the chatbot trained them in: notably, how doing so led them to better self-reflection. More than half (n = 78) of the collected items of commentary were coded in this category. For example, regarding writing a gratitude letter, a participant shared his experience of writing a gratitude letter with the chatbot, *"Having tried this journaling skill, I feel I am actually gaining power from feeling gratitude. It also has helped me to cultivate the habits and characteristics that I hope to have."* (S21, M)

- **Negative Experience -** Conversely, quite a few participants left comments (n = 33) that described their experience of practicing the target journaling skills in negative terms. Several expressed frustration at the system asking them to recall past events of which they had bad memories. One said, *"I don't think it is useful to recall in detail the emotions you felt when you made a mistake. Reflecting on them made me feel bad."* (S23, F)

- **Self-reflection -** Some participant comments (n = 56), mostly by those who had characterized their experience as positive, noted that practicing the journaling skills taught by the chatbot had promoted their self-reflection. A participant, for example, told us, *"When I was asked to list some names that I can ask for help from different aspects, I felt a little shocked and needed some time to think. I've never reflected this kind of question before. However, after I listed these people and reflected back on my relationships with them, I felt somewhat different. I recognized whom I could get help from, and with whom I could find happiness. This practice can be used to live my life better in the future."* (S34, M)

- **Suggestions for Peers -** Some participants left comments (n = 34) that included suggestions to their peers on how to utilize journaling skills in their daily lives, or even improve upon the skills that had been taught. One participant said: *"[W]riting in a gratitude journal every day could be helpful, and I suggest that everyone try. Because it is interesting, and it may help people to remember things they are grateful for and feel happy every day."* (S18, M)

- **Skepticism -** Finally, some participants, mostly those who characterized their chatbot experience as negative, expressed skepticism about the effects of learning journaling. The 12 items of commentary (n = 12) coded in this category included the following one about expressive writing: *"I did not fully understand this journaling skill. I'm a person who always looks ahead, so when the chatbot asked me to reflect on my past bad experiences, I wondered why it would be good for me to do so."* (S25, F)

In order to explore why LC participants chose to share or not to share some of their commentaries, I analyzed their interview responses to questions about their sharing decisions. Regarding sharing decisions, almost all LC participants thought it was fine to have their commentaries shared with their peers because they could decide what information to include in it. As an interviewee put it,

> "It is fine as long as I have control of what is being shared. Also, other guys shared their experience, and I felt I should also contribute some." (S27, F)

Besides, five LC participants attributed their decisions of sharing commentaries to the fact that the chatbot had enhanced their experience of practicing the target journaling skills. Thus, they also wanted to encourage other peers. As one mentioned,

> "The chatbot reminded me to practice the journaling skill every day, and gave me different topics, which was a good experience of practicing those skills. So, I wanted encourage other students to work with the chatbot." (S19, F)

Regarding not sharing decisions, two main reasons were identified. The first was that sharing experience could induce discomfort if a participant was not being very active about pursuing a task. Five participants specifically indicated that they sometimes did not pay much attention to the practices, and for that reason did not want to share responses that might be superficial. As a participant explained,

> "I was sometimes exhausted after getting back from my part-time job. I wanted to just lie down and watch TV, but the chatbot asked me to practice journaling skills. So, I only went through the motions of following the chatbot's instructions and did not learn much from it. I did not think it was appropriate to share this kind of experience." (S24, F)

The second was that three participants were deterred from sharing their commentary responses by their concerns that their experience might be opposite to their peers. Importantly, when the chatbot did not share others' frustration or negative experiences, participants felt hesitant about sharing such feelings or experience. As a user noted,

	Information	Thoughts	Feelings
L1	Mom always contributes a lot to the family (S15, F)	I think that everyone has their own guilty and depressing experiences (S5, M)	I did not have many friends at the University (S16, M)
L2	My father liked to bring our family hiking when I was a kind, and it became a good hobby for me (S18, M)	I think the guilty think I did recently, was how I treated my friend when I did not have time to help him with some tasks (S20, F)	Making friends in a new environment is difficult for me because talking with strangers made me nervous (S7, M)
L3	My best friend is a good role model for me because he is kind, and diligent about reaching his goals. I've learned a lot from him (S29, M)	I wish I could spend more time with my parents. I don't like myself when I'm always occupied by coursework and neglect my parents, despite the love I feel for them (S9, F).	I was so anxious when I was a freshman because I had to live alone and take classes that were all difficult for me. I almost gave up and felt depressed (S33, F)

Figure 6.4: Sample participants' responses to the journaling suggestions. The responses were coded to different categories and levels of self-disclosure. Note: Level (L)1 of Feelings is defined as "No expressing of feelings at all" [118].

> "I remember there was a day on which I didn't like the journaling skill, but the chatbot still asked me if I wanted to share my experience. I didn't share it because maybe only I had this feeling, and it would not help others." (S32, M)

6.4.2 Effects of Chatbot Designs on Users' Journaling Behaviors (RQ2)

To better understand if and how the participants in the OC and LC groups responded to the chatbot's journaling guidance differently, I explored differences in these two groups' dialogues about how they practiced journaling skills with the chatbot during *Phase 2: Learning Session* by evaluating the depth of self-disclosure and word count.

Depth of Self-disclosure in Journaling

Participants' dialogue data with the chatbots suggested that showing peers' practicing experiences promoted users' self-disclosure of their *Thoughts* (Figure 6.4). More specifically, there was a significant effect of group membership on participants' disclosing their *Thoughts* in the human-chatbot conversations ($F(1, 32) = 10.03$, $p < .001$) with the mean score of LC being significantly larger than that of OC. Though there was no significant effect of any factor on the depth of self-disclosure in either the *Information* or the *Feelings* category, indicating that seeing peer experience did not impact depth of the participants disclosed to the chatbot in these categories. Additionally, there was no main effect of experimental day, nor interaction effects.

Word count of participants' dialogue with the chatbot was also an indicator of participants' journaling efforts under their chatbots' guidance. There was a significant main effect of group membership on word count ($F(1, 32) = 9.94$, $p < .01$), with LC's word count being significantly higher than OC's word count. Interaction effects of group membership with experimental day were non-significant, as was the main effect of experimental day. Through conversation logs analysis, the results show that the LC participants made longer journaling responses with deeper self-disclosure of their *thoughts* than the OC participants during the *Learning Session*.

Factors Contributing to Users' Self-Disclosing Behavior

I examined participants' interview feedback on how the delivery of peer commentaries through the chatbot affected users' experience and behavior in the *Phase 2: Training*. Participants' interview feedback helped explain the differences of participants' journaling behaviors (self-disclosure) with the chatbot's guidance between the two groups.

In the **OC group**, eight participants stated that their self-disclosure levels did not change much during the study, but noted that they sometimes gave longer descriptions when asked to talk about a topic they were interested in. As an interviewee explained,

> "How much I disclosed depended on the question asked by the chatbot. When [... it] asked simple questions like what mood I was in, my responses were short. But when it asked deeper questions, I gave longer responses. Other than that, I think my disclosure level was rather stable throughout the study." (S7, M)

Moreover, seven participants from the OC group indicated in their interviews that their self-disclosure depth decreased. They said they did not like to self-disclose to the chatbot because, although it prompted them to go deep, they doubted it could understand what they disclosed. As one put it,

> "Although the chatbot asked me to keep going deeper, I never lost the sense that I was just interacting with a chatbot [... therefore] not necessary to say very much about my issues." (S8, F)

In the **LC group** which had the social learning component, nine participants highlighted the positive impact of peers' experience on their self-disclosure behavior. This phenomenon seemed to have been grounded in access to alternative perspectives, which in turn encouraged more thorough thinking about the chatbot's suggestions. As a participant explained,

> "I expressed myself more when I started to get suggestions that I think about specific issues [... and] peers' comments also offered me some food for thought.

So, both the chatbot and peers' useful suggestions helped me deal with my good and bad experiences." (S20, F)

Further, eight LC participants shared that they became more confident in the value of the chatbot's guidance as they realized that their own positive journaling experiences echoed those of their peers. This dynamic could have been part of what motivated this group to engage in significantly deeper disclosure of thoughts, as compared to the OC group. As one of these eight participants said,

"I found that some peers shared prior personal experience that was similar to mine, and it seems that they learned something from practicing the journaling skills with the chatbot, which made me believe I should be more open to it." (S18, M)

In addition, several LC participants stated that others' views and thoughts were the same as their own, which made them create resonance with other peers and increased their confidence about following the system's journaling suggestions. As one of them mentioned:

"I saw that my peers also had difficulty expressing their gratitude to someone they loved [...] but I think it's important to learn how to do it. I'm also glad that some people shared the same thoughts with me, so I followed [the journaling guidance] closely." (S21, M)

In summary, the LC's social learning component allowed the LC participants to practice more journaling (longer logs of their journaling content) and to share their thoughts more deeply than the OC participants. Our findings show that the LC participants were more complied with the chatbot's journaling guidance than the OC participants. Three main reasons were found from the interview: 1) the LC peers' experience seemed to offer them diverse reasons to follow the guidance; 2) the LC participants felt confident in following the guidance by reviewing peers's comments through the chatbot; and 3) they seemed to develop a sense of companionship by reading peers' comments.

6.4.3 Effects of Chatbot Designs on Users' Perception of Interaction with the Chatbot (RQ3)

To address RQ3 about participants' perceived interaction with their chatbot, I compared the differences between the two groups' self-reflection and perceived engagement. I then examined their interview feedback to understand the different perceptions.

Self-Reflection Levels

There were main effects of both group membership ($F(1, 32) = 4.2$, $p < .05$) and time point ($F(1, 32) = 48.9$, $p < .001$) on self-reflection level. The two groups had similar levels of self-reflection at the end of the *Phase 1*. However, the LC group's self-reflection became significantly higher than the OC group after learning the journaling skills in *Phase 2*. The results suggest that users felt that their self-reflection level had increased by practicing journaling skills with chatbot (OC); but the promoting effect on participants' self-reflection was even larger with the social learning component (LC).

Similarly, in their interviews, most participants **across both groups** agreed that the step-by-step guidance provided by the chatbot had encouraged them to reflect on themselves and had helped them understand themselves better. As one of them stated,

> "I like those journaling skills, and it is great to have this chance to go through the exercises step-by-step with the chatbot. By reflecting on my experiences, I came to understand myself better." (S16, M)

In the **OC** group, several participants indicated the limitations of incorporating a chatbot into the self-reflection process. Four OC members reported that they felt fatigued and annoyed because of the unexpectedly large amounts of effort they spent practicing journaling skills with the chatbot. They did not know the benefit of practicing the skills. In other words, they perceived following the suggestions as having a high cost. As one of these four participants mentioned,

> "I did not expect that I would have to expend much time and effort reflecting on my previous events, and it is also not what I usually do." (S7, M)

Moreover, two of them noted that they felt discomfort when the chatbot kept urging them to follow the guidance for self-reflection on their unpleasant experience. As shared by an interviewee,

> "I felt the chatbot was too pushy in some suggestions when it started to guide me and prompt me to reflect on those bad experiences. I didn't know why I should reflect on those pieces of stuff." (S14, F)

In the **LC** group, peer commentaries appear to have caused the participants' impressions of the chatbot's guidance to change over time. In their interviews, these participants revealed a variety of factors that contributed to their respective levels of compliance with such guidance. Eight LC participants indicated that knowing about others' experience led them to reflect on their own performance in following the guidance. As a participant mentioned,

> "When I saw others' comments, I would reflect on myself. I could check whether I was doing something wrong, or giving an appropriate response. It gave me clues as to whether I was on the right path or not." (S33, F)

Five of the LC participants also reported that reading others' experiences resulted in the adoption of new strategies for the application of journaling skills to their daily lives, e.g., listing three things that make them feel gratitude over the course of a day. As one of them stated,

> "Although I expended lots of effort following the guidance from the chatbot, I found others put some journaling skills into practice in their daily lives, and that this resulted in good feelings for them. This made me think I should also give it a try." (S30, F)

The process of asking users to share their experience with peers also prompted them to reflect upon their own journaling practices and to think about how they could best be explained. As one of the participants stated,

> "When I was reflecting about what I learned from the chatbot's guidance, I figured out that its suggestions were things we might already know but never spent enough effort on them. The chatbot guided me through practicing them and thus benefited me. So, I want to keep using those skills and share my experience with others." (S34, M)

Perceived Engagement

There were main effects of both group membership ($F(1, 32) = 4.34$, $p < .05$) and time point (i.e., before *Phase 2* v.s. after *Phase 2*) on perceived engagement levels ($F(1, 32) = 6.53$, $p < .01$), but no significant interaction effect. More specifically, the two groups had similar levels of engagement with the chatbot as of the end of *Phase 1*. However, at the end of *Phase 2*, OC's average perceived engagement level was higher than LC's engagement level. These results indicate that the social learning component interfered with the participants' perceived engagement.

Participants' interview feedback helped explain the differences in participants' perceived engagement between the two groups. In the **OC group**, ten participants said that their engagement with the chatbot increased over time because it gave them useful suggestions and prompted them to accomplish something new every day. For example, an interviewee said:

> ”I engaged in the interaction with the chatbot over time because I expected that I could learn something new (journaling skills) from it. It gave me the motivation to be aware of my true feelings in either good or bad events.” (S6, F)

Additionally, most participants in the OC group felt comfortable deciding whether to follow the chatbot's suggestions closely. It usually depended on their daily schedule and interests in some journaling skills. When they did follow the chatbot's suggestions, they generally felt happy and surprised that they had been able to learn something valuable from a chatbot. As one of them stated it,

> ”Honestly, I did not diligently practice the journaling skills every day because I was too busy to pay much attention to it. But, when I got some chances to focus on practicing journaling skills, I learned something from it and felt more motivated to keep practicing those skills.” (S11, F)

However, three of the participants in the OC group said they initially felt pressure from the chatbot when it started to give them suggestions, but such feelings seemed to fade over time. As one of the interviewees stated,

> ”In the first week, there were no suggestions. I felt pressure when the chatbot started to give me some suggestions. But in the end, I got used to it, and the feeling of pressure mostly disappeared. When the suggestions seemed interesting, I had positive feelings, but when they seemed time-consuming, I sometimes skipped them.” (S2, F)

In the **LC group**, seven participants shared similar feedback as OC participants regarding practicing journaling skills with the chatbot. Additionally, some of the LC participants reported that knowing their peers' experience motivated them to follow the chatbot's directions more closely. As one explained,

> ”The chatbot told me to practice and gave me some suggestions, pretty much like my parents always tell me to do [... and] this was tiresome. But after seeing my peers' experiences, it was more like my friends telling me that something was worthwhile, rather than just telling me to do it. So that motivated me to try the chatbot's suggestions.” (S22, F)

Besides, three of the participants indicated that though they did not directly interact with other peers, reading others' commentaries helped maintain their engagement in following the chatbot's guidance because of sense of companionship, as a participant stated,

> "I felt peers are feeling the same way from their comments. It was good to see others also following the journalng guidance. It gave me a sense that I'm not alone. When I saw others' comments, I could reflect on myself more. It also gives me clue whether I'm on the right path to keep practicing the journaling skills." (S18, M)

Nevertheless, five LC participants commented that having peer commentaries in the conversation impeded their engagement. At least in part, this was because the chatbot only showed them positive commentaries, leading the LC participants to doubt the sincerity and depth of the peer narrative as a whole. As one of them explained:

> "Seeing peers' comments didn't affect me and even made me a little annoyed because they always talked about positive aspects. I don't really think they were opening their hearts. I mean, some comments were somewhat superficial and therefore pretty useless to me. Of course, the journaling suggestions themselves were helpful, but the other users should have had some struggles." (S28, F)

In addition, three other LC participants indicated that they did not prefer to read others' experiences because they thought everyone's experience should be different; thus, it is hard to learn from others' experience. As an interviewee said,

> "My experience of practicing the journaling skills was different from those shared by the chatbot. But, people are different. Everyone has a different story and background, so I didn't benefit from seeing others' comments about their journaling experience." (S29, M)

Finally, we found five LC participants who reported that the inclusion of peers' experiences heightened their perceptions of being placed under pressure. Specifically, these LC participants stated that hearing about peers' positive experiences was a source of pressure, although it encouraged them to adhere to the guidance even when they disagreed with the suggestions. As one of the participants noted,

> "I felt pressure when reading peers' comments because it seemed others followed the guidance much more closely than I did. When they said, they thought a suggestion was useful [...] it made me frustrated why this suggestion hadn't worked for me. So, I felt I had to follow it more closely." (S31, M)

In summary, the social learning component had two major effects on participants' perception of their interaction with the chatbots. First, it successfully promoted the LC participants' self-reflection more than that of the OC participants, which was one of the primary goal of having participants' practice journaling by following their chatbots' guidance. Second, it interfered some LC participants' perceived engagement with the chatbot. More specifically, after the *Phase 2*, the LC participants' perceived engagement did not increase and was lower than that of the OC participants. Participants' interview feedback provided several reasons. In particular, when the LC participants who were self-motivated to practice the suggested skills saw peers' positive experiences, they were encouraged to continue and further improved their practices. However, for some LC participants who did not have positive experiences from their early journaling practices, seeing others' successful experiences was a pressure to them; thus, they even became resistant to follow the chatbot suggestions.

6.5 DISCUSSION

Below, I discuss the insights yielded by our comparison of the two chatbot variants, followed by implications for future chatbot design.

6.5.1 Learning from Peers through Chatbots

For LC participants, reading peers' success stories during the learning process sometimes served as a role model of journaling practices (as S30 mentioned in the interview). The tips for success inspired some LC participants to further improve their journaling skills. Knowing peers' stories also triggered the LC participants to reflect on their own experience and plan their behaviors based on what they had learned from peers (as S33 noted). This finding echoes prior work on social-learning practices as a key ingredient of promoting behavioral change [110]. Our study further extends the knowledge by showing that incorporating social learning in human-chatbot interaction has a similar effect of promoting target behavior, even without users' direct interaction.

One major reason that explained why social learning succeeded in promoting journaling in our study was because sharing peers' commentaries developed some participants' sense of companionship with other LC participants. For instance, participant S21 said he shared the same sentiment when he saw his peers also struggled to express their gratitude in their journals. This result echos prior research [104, 184] that when people find similarities with others, they tend to assimilate toward the others' experience. Although our system did not offer a direct communication channel among the participants, the participants still perceived themselves as connected. Our work extends our understanding of human-chatbot interaction by showing how a chatbot's new role - i.e., as an intermediary in the exchange of peers training experiences - may enhance users' sense of companionship with one another.

Compared to the LC group, the OC group appeared to be less compliant with journaling guidance in *Phase 2: Training*, as evidenced by their shallower self-disclosure and shorter responses. Our findings extend the prior literature [8, 24] by showing that several barriers exist to the effective use of chatbots to guide users. These barriers were expressed in our participants' interview data: i.e., that the chatbot became aggressive and annoying in the Learning session; that following the chatbot's journaling guidance made them uncomfortable; and that they doubted the usefulness of its suggestions. For LC group, the social learning component provided the participants opportunities to learn diverse views about the journaling skills from their peers. Since these materials (i.e. tips for success) were shared by their peers, the LC chatbot mitigated the above barriers that were more intrusive when participants interacted with the OC chatbot. These findings expanded self-disclosure literature in the context of human-chatbot interaction.

6.5.2 Potential Negative Impacts of Social Learning on Human-Chatbot Interaction

Although integrating the social learning component into the chatbot appeared to boost LC participants' practicing behavior (length and depth of their journaling content) to follow the chatbot's guidance, it also seemed to interfere with their perceived engagement. Two major factors, i.e., dissimilarity and peer pressure, were identified to contribute to this result.

First, several LC participants felt indifferent to their peers' stories (e.g., S28 feeling even annoyed). Such feelings seemed to appear when their own experiences were different from their peers - when their own experiences were negative and their peers' were positive. They became less confident about the veracity of the peer-commentary feature and the authenticity of the chatbot's suggestions. These findings are related to Mussweiler's findings [184] that when people find dissimilarities between them and others, judging others may hinder own engagement.

Second, some LC participants felt pressured to be compared with their peers, especially when they felt their own experiences were sub-optimal. Such peer pressure may have had a negative impact on LC participants' perceived engagement with the chatbot. As prior research [112, 113] shows, sharing peers' experiences may prompt negative social comparison – a process potentially leading to negative emotions [185]. As such, even though social learning may motivate people to comply with better behaviors, it does not always lead to a better experience [186, 187]. However, it is worth noting that peer pressure is not always unfavorable; it can increase users' motivation to aim higher and promote growth [188]. What probably matters is the balance between pressure and users' motivation - the peers' pressure shouldn't be so high that the user entirely gives up on the goal. As the investment model of engagement [92] indicates, individuals engage less and stop investing their time and effort if they feel the cost is too high.

Unlike LC participants, OC participants generally deemed themselves to be engaged in learning journaling skills from the chatbot, and they reported increased perceived engagement with their chatbot in *Phase 2*. Although the OC participants had lower depth of self-disclosure, the lack of pressure to follow system suggestions that the OC participants perceived could result in a lower perceived cost [92] of continuing the practices. Besides, the OC participants also seemed to perceive the benefits of learning the journaling skills through the chatbot. Therefore, compared to the LC participants, their engagement of continuing practices was less negatively impacted.

6.5.3 Design Implications

Our study highlighted potential benefits and drawbacks that need to be addressed when integrating social learning into conversational agent technologies. Future chatbot designs for incorporating social learning could explore more variants of social learning designs to improve user experience and compliance with chatbots' guidance. Based on our research findings, I propose three design implications.

First, chatbots may be designed to foster a sense of companionship with role models. In our study, the social learning component boosted users' sense of companionship. The social learning facilitated by the same type of peer-sharing features could also be useful in other contexts, e.g., among adolescents, who tend to accept their peers as role models [89]. A chatbot can play a intermediary role, tailored to users' various experiences, backgrounds, and known similarities, as a means of increasing their acceptance of its guidance. Research has shown that when users feel they have support and are not alone, they are likely to cope more effectively with the guidance and are more open to sharing their experiences and knowledge [107, 110]. Thus, future research may explore design mechanisms specifically to catalyze users' sense of belonging and companionship: e.g., fusing chat with visualizations of peers' performance or summaries of peers' struggles and experience, or matching users with similar lifestyles to each other.

Second, our chatbot did not guide participants on what to share in their commentaries. As a result, some of their commentaries appeared to be not detailed enough for the other participants to comprehend. Thus, giving some prompts to guide users' sharing is worthy of exploring. For example, when users share their positive learning outcomes, the chatbot can also prompt them to include their struggles that they have gone through to enrich their successful experiences. Additionally, the chatbot can provide explanations of such commentaries to the users so that reading the comments can prompt self-reflection. For example, the chatbot could guide users to see both positive and negative sides of their peers' experiences [8] instead of harboring suspicions about their insincerity or superficiality. Though one major advantage of social learning is learning from others' diverse perspectives, what needs to be included, how to share, and the potential negative outcomes are always worth further investigation.

Third, future designs may be deployed in more application domains that benefit from users' deep self-disclosure and self-reflection. Our study showed that integrating social learning into a chatbot enhanced its users' depth of self-disclosure and facilitated their self-reflection. The LC group participants clearly wanted to understand and learn from their peers' perspectives on practicing journaling skills, including the challenges they encountered. One design implication of this is that social learning may help reinforce the impact of utilizing chatbots to deliver suggestions or guidance, including in the spheres of mental health [34] and life-skills coaching [89]. In addition, chatbots have been deployed in different educational environments, not only to answer common questions but also to provide more complex guidance.

6.6 LIMITATIONS

The present study has several limitations. First, it took the form of a field study that allowed users to interact with a chatbot for 19 consecutive days to practice learning several journaling skills. Because users were allowed to complete their daily chatting tasks at any location and in a wide range of time, I omitted the evaluation of time though it could be an indicator for measuring users' effort in practicing the journaling skills. Future studies could consider involving this factor in the study by giving some time constraints, but it may also sacrifice the flexibility of utilizing a chatbot.

Second, our study did not consider how being involved in this type of learning experience might lead to self-development or behavioral changes in the participants' real lives, and future studies should do so. Third, our findings may not apply to other contexts such as quitting smoking and eating a healthier diet [107, 189]. Future studies should explore the use of chatbots to offer different types of learning guidance, such as for changing behaviors.

Finally, our study had a rather small sample size for the survey results. To mitigate the issue, I used a mixed-method analysis and triangulated our findings by providing evidence from different measures (e.g., survey and interview). However, increasing the sample size would help us solidify our understanding and see whether our findings hold true for different types of population.

6.7 CONCLUSION

Chatbots have been utilized in various applications, and they could proactively assist users in accomplishing a specific task and further giving guidance for users through a learning and reflection process. Our study investigated how social learning mediated by a chatbot influences users' experience and behaviors to learn journaling skills from the chatbot's guidance. I found that chatbots could encourage users to follow directions to self-reflect via journaling skills. Understanding peers' commentaries on the chatbot-guided learning experience increase the users' depth of self-disclosure on thoughts and encourage learning from others' experiences. Besides, this design allowed some users to find the sense of companionship of practicing journaling skills, which motivates them to follow the chatbot's guidance. However, I also observed that incorporating this social learning component may interfere with some users' perceived engagement in interacting with the chatbot since the mismatch between the user's and peers' experiences might cause comparison and pressure. Our research design and findings serve as the first step to facilitate social learning in human-chatbot interaction to deliver guidance.

CHAPTER 7: CONCLUSION AND FUTURE WORK

7.1 SUMMARY

In this section, I summarize the achievements of Chapter 3, Chapter 4, Chapter 5, and Chapter 6 and propose potential topics and considerations for future work. The focus of this dissertation is utilizing chatbots to mediate humans' transfer of information and trust. Because chatbots have become more popular and in the process of changing how humans interact with computer systems, I eager to explore how chatbots could serve as mediators to collect information from their users and convince the users to share it with third parties. This is an important topic because of the ethical and risk issues involved once chatbots have become a prevalent technology. My research found that chatbots could encourage users' in-depth self-disclosure and inferred that users might develop relationships and trust with the chatbots when chatbots came to dealing with the users' sensitive information. My research further addressed this topic by leveraging humans and AI's complementary strengths to facilitate human-AI collaboration.

In **Chapter 3**, I explored the effect of reciprocity in human-chatbot interaction. I designed and evaluated a chatbot that has self-disclosure features when it performs small talk with people. I ran a study and divided the participants into three groups to use different chatting styles of the chatbot for three weeks. This work's significant contribution is that the results showed an effective chatbot design that promoted deep self-disclosure over time. I investigated the effect of chatbot self-disclosure on the depth of people's self-disclosure over three weeks and studied the impact across two chat sessions. The results showed that the chatbot's self-disclosure level substantially affected the user's deep self-disclosure over time and explained how factors contributed to the impact. These findings extend knowledge of how chatbot designs and time influence users' depth of self-disclosure, which benefits future chatbot design for mental wellbeing.

In **Chapter 4**, the chatbot was designed to ask participants to share their self-disclosure content with a professional third party. This study found that the chatbot's request did not dramatically affect the participants' self-disclosure behavior to the chatbot system. However, the different chatting styles influenced the participants' depth of self-disclosure, especially when disclosing personal feelings. Besides, most participants chatting with no/low self-disclosure chatbot shared their data with the professional third party because they trusted the "research team/doctor" behind the chatbot to deal with their information properly. In the group using a high self-disclosure chatbot, in contrast, the participants' trust seemed to start with the chatbot first. It spilled over to include the professional third party subsequently. Though several participants expressed their surprise when being requested to share their data with a mental health professional, they were willing to share the good intentions of supporting mental wellbeing.

On the other hand, some participants lacked trust in the chatbot, the research team, or the doctor and felt no need to share their answers further because they could not see the benefits of doing so. These findings imply that the chatbot that offered deep self-disclosure had the potential of serving as an effective mediator to facilitate the people's self-disclosure of sensitive information. The survey scores also show that participants chatting with high self-disclosure chatbot had significantly stronger trust in the professional third party than the other two groups.

The significant contribution of **Chapter 5** is that it has shown the effects of integrating human (expert) support into a chatbot system to deliver Journaling guidance. The study indicates that the group with human support tended to value the suggestions during the training process and tried to follow the journaling suggestions more closely than the group without integrating human support, resulting in longer journaling content with deeper self-disclosure their feelings and thoughts. The group with human support also reported significantly higher perceived engagement and trust than another group. Their perception of the available human supporter appeared to boost the participants' perceived usefulness of the journaling suggestions and their actual journaling practices. However, some participants also felt like they were being monitored and felt pressure to follow the expert's instructions. Their expectations of receiving highly personalized suggestions from the human expert were not met, leading to negative impressions.

Surprisingly, I found that the group without human support practiced the journaling skills significantly more than the group with the human backing during the *Free-will* phase, which allowed them to skip practicing the journaling skills. This is unexpected because the group with human support exhibited higher engagement with more prolonged and deeper self-disclosure during the *Training* phase. Participants' interview feedback suggested that the participants lacked accountability for continuing to practice journaling skills because there was no longer a expert "monitoring" in the *Free-will* phase.

Finally, **Chapter 6** took a further step in exploring the effect of integrating social learning (i.e., learning from peers' experiences via a chatbot) component to encourage the users to comply with the guidance from a chatbot. I designed two chatbots that deliver journaling guidance, one with and one without a social learning component, and conducted a three-week study with 34 participants. This work's main contribution is that it has shown the effects of integrating the social learning component into a chatbot system to deliver suggestions. I evaluated how incorporating social learning in chatbot conversation could impact people's compliance with the chatbot guidance and their perceived experience. Specifically, adding the social learning component encouraged participants' deeper self-disclosure in their thoughts with extended journaling input and higher self-reflection than the group without the social learning component. Although using a chatbot to guide users in learning journaling skills improved participants' self-reflection and their perceived engagement with their chatbot, the proposed social learning component seemed to interfere with some participants' perceived engagement human-chatbot interaction.

In the first chapter, I introduced a Figure 1.1 to narrate in detail the origin and context of this dissertation research and approached some research questions. I addressed them through a sequence of studies. The findings demonstrate that a chatbot could be an effective facilitator to assist human-human communication and promote users making corresponding behavior change. This dissertation research may inaugurate a new direction for the future design of chatbots in healthcare eco-systems to augment human experts' communication with their clients.

7.2 ETHICAL AND SAFETY CONSIDERATIONS OF THE STUDIES

In addition to discovering the benefit of incorporated chatbots into the healthcare ecosystem, I discuss the potential ethical concerns and privacy issues raised in the studies. Furthermore, I propose some perspectives and considerations for future researchers and practitioners to design conversational agents considering users' safety and privacy. My dissertation studies were approved by Institutional Review Board (IRB), and all participants' data were secured and anonymous. The issues I discuss here are the parts of research observation in the studies presented in previous chapters.

This dissertation research aims to explore effective chatbot designs for eliciting users' deep self-disclosure in order to improve mental well-being; thus, users' privacy and potential ethical issues should be carefully considered. Kretzschmar et al. outlined minimum ethical standards for using chatbots in mental health support relevant to my research contexts. Thus, I discuss ethical issues as follows by referring to the perspectives addressed in [190].

7.2.1 Privacy and Transparency

Some participants provided extremely sensitive content when chatting with the chatbot in the studies because the chatbots guided them to reflect on their prior good/bad experiences, e.g., experiences related to abuse and depression. Such information should be kept confidential and de-identified. Users should further have the option of anonymizing their content. In addition, the transparency of data processing should be granted. For my research instructions, I clearly stated that their conversation data would only be analyzed by the researchers for research purposes and would not be shared with others without their permission. However, in the market, many chatbots are deployed on existing messenger platforms (e.g., Skype and Telegram). The third parties' privacy policy should address how to prevent users' data from being collected by third parties without permission.

The users' data is securely collected and analyzed in the studies, but it may raise critical concerns when applying this chatbot design in practice. Specifically, a chatbot/AI system's power of agency is a lack of clarity. As I demonstrated in the studies, the chatbots were able to elicit participants' in-depth self-disclosure. However, if a user discloses the intention of self-harm or even crime in the conversation, a debatable issue would be uncovered - Should an AI system automatically recognize these patterns? Should an AI agent report these patterns to the related agencies? From the perspective of protecting users' privacy and unnecessary misunderstandings, I incline to choose not to share the information, and users should be allowed to determine if their conversation with a chatbot should be shared with others. I believe this issue needs more work to explore to define a clear boundary of an AI system's agency.

7.2.2 Efficacy

According to our research findings and prior studies, [34], chatbots could be an effective mediator to deliver interventions for improving people's mental well-being. My research adopted the Journaling method (e.g., gratitude journaling, expressive writing) to help users reflect on prior events, and some researchers deployed the CBT method using a chatbot system. Though these methods have been tested empirically and used in practice, it is still new to use conversational agents to deliver guidance. Thus, many factors should be examined in future research; for example, a conversational agent's identity and conversational styles may influence users' decision to accept the recommendation [14, 191]. Besides, the effect of time should also be considered when using conversational agents to deliver interventions for mental well-being.

In addition, some participants in my studies mentioned that talking with the chatbot felt as if they were talking with a psychiatrist - they even expected professional feedback from the chatbot. This implies that the users may assume the chatbot has more intelligence than it actually does, which might lead to users not reaching out to professionals for proper help. Hence, when deploying a chatbot system for mental well-being, users should be informed and reminded of what effects/risks to expect from the chatbot. Furthermore, researchers and practitioners should remind their users about what the chatbot systems' targets and what effects to expect.

7.2.3 Safety

Although I recruited participants who were less likely exposed to severe psychological distress, users' deep self-disclosure may still arouse users' negative experiences and thoughts. To address unwanted situations, I had experienced psychiatrists review our chatbot and study design. I also provided the participants with emergency contact information so that they can ask for help in case of an emergency. An effective monitory mechanism might be further necessary for addressing unexpected psychological crises and stop participation for actual use.

Instead of increasing users' reliance on chatbots to deal with their mental well-being, the chatbot systems should encourage users to seek human support if necessary. Therefore, as I mentioned above, a chatbot system's agency power may need further work to define. For instance, whether a chatbot could play a proactive role to introduce a professional third party when detecting unsafe conditions? when an automated chatbot should stop/suspend its interaction with a user in a fragile state? In the near future, I believe that more and more intelligent and human-like computer agents will be developed; thus, the safety of deploying these AI agents to tackle mental health requires more careful examination.

7.2.4 Ethical Perspectives of Introducing Human Support (third parties) in Human-Chatbot Interaction

Chatbot work needs to address ethics and privacy issues carefully [22, 25, 190]. For example, because chatbot use can result in users' self-disclosure of sensitive topics [25, 190], whether to disclose it, when, and precisely to whom remain difficult to answer. This is especially challenging to answer if the goal is to improve mental health. If a chatbot introduces an expert (a counselor) in the middle of the study, it may raise privacy and transparency concerns because it may imply that past conversations between the chatbot and users are transferred to other parties. The participants did not specifically express this concern in my studies, which may be attributed to the social presence and expertise of an expert and my research settings (I declared that their data would be secured). Nevertheless, this concern may be heightened when the chat topic is sensitive, and it is an important issue to explore in future research.

Most of the participants tended to trust the chatbot system's suggestions. Introducing the human expert further reinforced the participants' perceived trust and raised their expectations on the expert for giving personalized suggestions. Although my research did not aim to deceive the users, the study findings revealed that integrating a human expert might lead to participants' unrealistic expectations of a chatbot's intelligence and efficacy. Hence, when deploying a chatbot system for healthcare purposes, the transparency of the mechanism should be informed. In addition, although the journaling suggestions our chatbot systems offered were evidence-based [165, 168, 169, 172] and had no potential to cause harm, 'over-trust' could lead to harm if a future system were not working properly [16].

In Chapter 6, I further explored the effectiveness of integrating social learning components in the chatbot system. The study results show that peers' comments elicited social comparison and increased the users' pressure, leading to their negative experience with the interventions. Although the chatbot only shared the peers' positive comments in the study, this design unexpectedly caused a negative effect when using a chatbot to deliver guidance for improving mental well-being. This research finding indicates the necessity of examining chatbots in the healthcare eco-system to mitigate potential negative effects. My research is a good start to inform how future studies should be aware of these risks when using conversational agents to deliver guidance for improving mental well-being.

7.3 FUTURE DIRECTION

This dissertation research has demonstrated how chatbots could help collect information from the users and convince them to share information and follow a third party's guidance. In the following paragraphs, I introduce some opportunities for future research.

7.3.1 Exploring and Designing Human-AI Collaboration

Digital assistants such as Alexa, Siri, and Google Assistant are among the most widely accepted AI technologies and have considerable potential to be further embedded in workplaces and other aspects of daily life. Based on my findings regarding human-to-CA disclosure and long-term social dynamics [24, 25, 26], I have been reflecting on how future AI assistants should evolve next, especially in light of the goals of naturalistic communication and the achievement of tasks.

I am especially interested in exploring how AI technologies can properly be incorporated into the healthcare eco-system. It can be used to encourage a healthy lifestyle and help healthcare providers to better understand clients' day-to-day patterns and needs to provide better treatment and guidance. For example, Wang et al. [86] has shown that smart speakers (e.g., Alexa and Google Home) could play as social agent to reduce the users' public speaking anxiety. In addition, some studies [42] demonstrated that the chatbots could be a learning companion to promote the users' learning outcomes. Thus, chatbots could be involved in the humans' decision-making process. Providing guidance to assist humans' decision-making may yet emerge as the best way for people and AI agents to work well together. My pioneering empirical work [24, 25, 26] on the effects of incorporating human support into human-AI interaction has highlighted some of the new design opportunities and challenges that arise when chatbot systems, with and without human support, are used for long-term interventions. This is one of crucial topics in human-centered AI research.

7.3.2 The Ethics and Fairness of AI for Healthcare

Building on my work on chatbots as providers of guidance in reflective journaling, I am satisfied that chatbots capable of engaging humans in conversation through natural-language interfaces will emerge as long-term companions, which can assist clinicians in the collection of essential health information. Yet, despite the increasing adoption of chatbots in healthcare and the apparent benefits thereof, many challenges must still be addressed. For example, user safety and AI ethics both need to be defined, since there are no rubrics for what kind of information an AI should, or should not, collect from its users - nor is the existing research on the links between chatbot use and healthcare outcomes particularly robust.

Therefore, future research should consider integrating expert and social support into chatbots to enable a more meaningful evaluation of their users' healthcare outcomes and address user privacy and autonomy in the healthcare system. This also raises ethical and privacy considerations related to who owned the users' conversational data and whether the data should be monitored when it comes to chatting related to mental well-being. In contrast to the graphical user interface, a chatbot could play a proactive role to prompt the users to interact with it. More research is needed to explore the potential problems (e.g., over-trust and over-sharing) when broadly and longitudinally deploying chatbots in daily life. Finally, it is crucial for future studies to lay down the ethical foundation for incorporating AI technology safely into the healthcare eco-system.

7.4 CONCLUSION

This dissertation made contributions to the literature aimed at HCI and human-centered AI domains to provide novel insights about designing conversational agents to improve users mental well-being and human-human communication.

First, I conducted a longitudinal study to investigate how self-disclosure of chatbots affects users' self-disclosure behavior. Both conversation styles and the time elapsed since the start of the experiment influenced users' subjective experiences of using the chatbot and their objective self-disclosure behavior. In general, the chatbot that made its own self-disclosures performed better at facilitating its users' self-disclosures in response to sensitive questions, successfully encouraging users to provide longer responses and express deeper thoughts and feelings on sensitive topics.

I then turned to investigate how a chatbot as a mediator can be used by people for self-disclosing to a mental health professional and how people's trust in a chatbot interacts with their trust in a mental health professional. The study findings suggest that the chatbot's self-disclosure successfully elicits participants' self-disclosure of their personal experiences, thoughts and feelings not only to the chatbot but also to the mental health professional. This work provides empirical evidence of different self-disclosure behavior, such as reducing or adding content, that people may take before sharing their self-disclosure to a chatbot with a mental health professional. I also identified an effective chatbot design that has promising potential to serve as a mediator to promote self-disclosure to mental health professionals.

According to the findings from the studies above, I designed a study to explore chatbot designs with and without integrating human support for delivering guidance to users, along with a three-phase study I conducted to examine the impact of human expert support on human-chatbot interaction. I focused on investigating the impact on users' experience and efforts involved in following the guidance of journaling skills. The findings show that the chatbot integrated with human support enhances users' self-disclosure, perceived engagement, and trust in the *Training* phase; however, it might also make people feel pressured. Although the group without human support showed less engagement in the *Training* phase, they remained interested longer in practicing journaling skills. This research extends the understanding of human-chatbot interaction by providing insights on future chatbot designs with human experts' support to deliver guidance in broader contexts.

Finally, I explored how social learning mediated by a chatbot influences users' experience and behaviors to learn journaling skills from their guidance. The study found that chatbots could encourage users to follow directions to self-reflect via journaling skills. Understanding peers' commentaries on the chatbot-guided learning experience increase users' depth of self-disclosure on thoughts and encourage learning from others' experiences. Besides, this design allowed some users to find the companionship of practicing journaling skills, motivating them to follow the chatbot's guidance. However, I also observed that incorporating this social learning component may interfere with some users' perceived engagement in interacting with the chatbot since the mismatch between the user's and peers' experiences might cause comparison and pressure. This research design and findings serve as the first step to facilitate social learning in human-chatbot interaction to deliver guidance.

My dissertation research yielded new understandings of the roles played by mutuality and the passage of time in self-disclosure interactions. It has important implications for the design and use of chatbots for eliciting deep self-disclosure and delivering guidance.

REFERENCES

[1] E. Adamopoulou and L. Moussiades, "Chatbots: History, technology, and applications," *Machine Learning with Applications*, vol. 2, p. 100006, 2020.

[2] K. Woodward, E. Kanjo, D. Brown, T. M. McGinnity, B. Inkster, D. J. Macintyre, and A. Tsanas, "Beyond mobile apps: A survey of technologies for mental well-being," *arXiv preprint arXiv:1905.00288*, 2019.

[3] G. M. Lucas, A. Rizzo, J. Gratch, S. Scherer, G. Stratou, J. Boberg, and L.-P. Morency, "Reporting mental health symptoms: breaking down barriers to care with virtual human interviewers," *Frontiers in Robotics and AI*, vol. 4, p. 51, 2017.

[4] A. Ravichander and A. W. Black, "An empirical study of self-disclosure in spoken dialogue systems," in *Proceedings of the 19th Annual SIGdial Meeting on Discourse and Dialogue*, 2018, pp. 253–263.

[5] J. Seering, M. Luria, G. Kaufman, and J. Hammer, "Beyond dyadic interactions: considering chatbots as community members," in *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, 2019, pp. 1–13.

[6] S. Kim, J. Eun, C. Oh, B. Suh, and J. Lee, "Bot in the bunch: Facilitating group chat discussion by improving efficiency and participation with a chatbot," in *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, 2020, pp. 1–13.

[7] F. M. Zanzotto, "Human-in-the-loop artificial intelligence," *Journal of Artificial Intelligence Research*, vol. 64, pp. 243–252, 2019.

[8] R. Kocielnik, L. Xiao, D. Avrahami, and G. Hsieh, "Reflection companion: A conversational system for engaging users in reflection on physical activity," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 2, no. 2, p. 70, 2018.

[9] K. Lukoff, T. Li, Y. Zhuang, and B. Y. Lim, "Tablechat: Mobile food journaling to facilitate family support for healthy eating," *Proceedings of the ACM on Human-Computer Interaction*, vol. 2, no. CSCW, pp. 1–28, 2018.

[10] A. C. Williams, H. Kaur, G. Mark, A. L. Thompson, S. T. Iqbal, and J. Teevan, "Supporting workplace detachment and reattachment with conversational intelligence," in *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, ser. CHI '18. New York, NY, USA: ACM, 2018. [Online]. Available: http://doi.acm.org/10.1145/3173574.3173662 pp. 88:1–88:13.

[11] M. Lee, S. Ackermans, N. van As, H. Chang, E. Lucas, and W. IJsselsteijn, "Caring for vincent: A chatbot for self-compassion," in *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, ser. CHI '19. New York, NY, USA: ACM, 2019. [Online]. Available: http://doi.acm.org/10.1145/3290605.3300932 pp. 702:1–702:13.

[12] C. Kelley, B. Lee, and L. Wilcox, "Self-tracking for mental wellness: understanding expert perspectives and student experiences," in *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. ACM, 2017, pp. 629–641.

[13] J. Jang and J. Kim, "Healthier life with digital companions: Effects of reflection-level and statement-type of messages on behavior change via a perceived companion," *International Journal of Human–Computer Interaction*, pp. 1–18, 2019.

[14] W. Shi, X. Wang, Y. J. Oh, J. Zhang, S. Sahay, and Z. Yu, "Effects of persuasive dialogues: Testing bot identities and inquiry strategies," in *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, ser. CHI '20. New York, NY, USA: Association for Computing Machinery, 2020. [Online]. Available: https://doi.org/10.1145/3313831.3376843 p. 1–13.

[15] S. Jeong, S. Alghowinem, L. Aymerich-Franch, K. Arias, A. Lapedriza, R. Picard, H. W. Park, and C. Breazeal, "A robotic positive psychology coach to improve college students' wellbeing," in *2020 29th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*. IEEE, 2020, pp. 187–194.

[16] J. Xu, G. B. De'Aira, Y.-P. Chen, and A. Howard, "Robot therapist versus human therapist: Evaluating the effect of corrective feedback on human motor performance," in *2018 International Symposium on Medical Robotics (ISMR)*. IEEE, 2018, pp. 1–6.

[17] S. S. Sundar and J. Kim, "Machine heuristic: When we trust computers more than humans with our personal information," in *Proceedings of the 2019 CHI Conference on human factors in computing systems*, 2019, pp. 1–9.

[18] E. Bogert, A. Schecter, and R. T. Watson, "Humans rely more on algorithms than social influence as a task becomes more difficult," *Scientific Reports*, vol. 11, no. 1, pp. 1–9, 2021.

[19] H. Trinh, A. Shamekhi, E. Kimani, and T. W. Bickmore, "Predicting user engagement in longitudinal interventions with virtual agents," in *Proceedings of the 18th International Conference on Intelligent Virtual Agents*, 2018, pp. 9–16.

[20] J. Pereira and Ó. Díaz, "Using health chatbots for behavior change: a mapping study," *Journal of Medical Systems*, vol. 43, no. 5, p. 135, 2019.

[21] J. Xu, G. B. De'Aira, and A. Howard, "Would you trust a robot therapist? validating the equivalency of trust in human-robot healthcare scenarios," in *2018 27th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*. IEEE, 2018, pp. 442–447.

[22] A. Howard and J. Borenstein, "Trust and bias in robots: These elements of artificial intelligence present ethical challenges, which scientists are trying to solve," *American Scientist*, vol. 107, no. 2, pp. 86–90, 2019.

[23] M. Luria, S. Reig, X. Z. Tan, A. Steinfeld, J. Forlizzi, and J. Zimmerman, "Re-embodiment and co-embodiment: Exploration of social presence for robots and conversational agents," in *Proceedings of the 2019 on Designing Interactive Systems Conference*, 2019, pp. 633–644.

[24] Y.-C. Lee, N. Yamashita, Y. Huang, and W. Fu, ""i hear you, i feel you": Encouraging deep self-disclosure through a chatbot," *ACM CHI Conference on Human Factors in Computing Systems*, 2020.

[25] Y.-C. Lee, N. Yamashita, and Y. Huang, "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*, 2020.

[26] Y.-C. Lee, N. Yamashita, and Y. Huang, "Exploring the effects of incorporating human experts to deliver journaling guidance through a chatbot," *Proceedings of the ACM on Human-Computer Interaction*, 2021.

[27] H. Gilburt, D. Rose, and M. Slade, "The importance of relationships in mental health care: A qualitative study of service users' experiences of psychiatric hospital admission in the uk," *BMC health services research*, vol. 8, no. 1, pp. 1–12, 2008.

[28] D. Rickwood, F. P. Deane, C. J. Wilson, and J. Ciarrochi, "Young people's help-seeking for mental health problems," *Australian e-journal for the Advancement of Mental health*, vol. 4, no. 3, pp. 218–251, 2005.

[29] J. Weizenbaum, "Eliza—a computer program for the study of natural language communication between man and machine," *Communications of the ACM*, vol. 9, no. 1, pp. 36–45, 1966.

[30] J. Grudin and R. Jacques, "Chatbots, humbots, and the quest for artificial general intelligence," in *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, 2019, pp. 1–11.

[31] T. Hu, A. Xu, Z. Liu, Q. You, Y. Guo, V. Sinha, J. Luo, and R. Akkiraju, "Touch your heart: a tone-aware chatbot for customer care on social media," in *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, 2018, pp. 1–12.

[32] G. Cameron, D. Cameron, G. Megaw, R. Bond, M. Mulvenna, S. O'Neill, C. Armour, and M. McTear, "Towards a chatbot for digital counselling," in *Proceedings of the 31st British Computer Society Human Computer Interaction Conference*. BCS Learning & Development Ltd., 2017, p. 24.

[33] S. Divya, V. Indumathi, S. Ishwarya, M. Priyasankari, and S. K. Devi, "A self-diagnosis medical chatbot using artificial intelligence," *Journal of Web Development and Web Designing*, vol. 3, no. 1, pp. 1–7, 2018.

[34] K. K. Fitzpatrick, A. Darcy, and M. Vierhile, “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*, vol. 4, no. 2, p. e19, 2017.

[35] C. Toxtli, A. Monroy-Hernández, and J. Cranshaw, “Understanding chatbot-mediated task management,” in *Proceedings of the 2018 CHI conference on human factors in computing systems*, 2018, pp. 1–6.

[36] A. Fadhil, G. Schiavo, and Y. Wang, “Coachai: A conversational agent assisted health coaching platform,” *ArXiv*, vol. abs/1904.11961, 2019.

[37] S. Park, J. Choi, S. Lee, C. Oh, C. Kim, S. La, J. Lee, and B. Suh, “Designing a chatbot for a brief motivational interview on stress management: Qualitative case study,” *Journal of medical Internet research*, vol. 21, no. 4, p. e12231, 2019.

[38] C. Nass, J. Steuer, and E. R. Tauber, “Computers are social actors,” in *Proceedings of the SIGCHI conference on Human factors in computing systems*. ACM, 1994, pp. 72–78.

[39] Z. Ashktorab, M. Jain, Q. V. Liao, and J. D. Weisz, “Resilient chatbots: Repair strategy preferences for conversational breakdowns,” in *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, 2019, pp. 1–12.

[40] M. Skjuve, A. Følstad, K. I. Fostervold, and P. B. Brandtzaeg, “My chatbot companion-a study of human-chatbot relationships,” *International Journal of Human-Computer Studies*, vol. 149, p. 102601, 2021.

[41] A. V. Prakash and S. Das, “Intelligent conversational agents in mental healthcare services: A thematic analysis of user perceptions,” *Pacific Asia Journal of the Association for Information Systems*, vol. 12, no. 2, p. 1, 2020.

[42] V. Ta, C. Griffith, C. Boatfield, X. Wang, M. Civitello, H. Bader, E. DeCero, and A. Loggarakis, “User experiences of social support from companion chatbots in everyday contexts: Thematic analysis,” *Journal of medical Internet research*, vol. 22, no. 3, p. e16235, 2020.

[43] E. Ignatius and M. Kokkonen, “Factors contributing to verbal self-disclosure,” *Nordic Psychology*, vol. 59, no. 4, pp. 362–391, 2007.

[44] I. Altman and D. A. Taylor, *Social penetration: The development of interpersonal relationships.* Holt, Rinehart & Winston, 1973.

[45] M. Nguyen, Y. S. Bin, and A. Campbell, “Comparing online and offline self-disclosure: A systematic review,” *Cyberpsychology, Behavior, and Social Networking*, vol. 15, no. 2, pp. 103–111, 2012.

[46] L. Baruh and Z. Cemalcılar, "When more is more? the impact of breadth and depth of information disclosure on attributional confidence about and interpersonal attraction to a social network site profile owner," *Cyberpsychology: Journal of Psychosocial Research on Cyberspace*, vol. 12, no. 1, 2018.

[47] B. E. Tolstedt and J. P. Stokes, "Self-disclosure, intimacy, and the depenetration process." *Journal of Personality and Social Psychology*, vol. 46, no. 1, p. 84, 1984.

[48] L. R. Wheeless and J. Grotz, "The measurement of trust and its relationship to self-disclosure," *Human Communication Research*, vol. 3, no. 3, pp. 250–257, 1977.

[49] J. W. Pennebaker, *Emotion, disclosure, & health.* American Psychological Association, 1995.

[50] Y. Moon, "Intimate exchanges: Using computers to elicit self-disclosure from consumers," *Journal of consumer research*, vol. 26, no. 4, pp. 323–339, 2000.

[51] D. DeVault, R. Artstein, G. Benn, T. Dey, E. Fast, A. Gainer, K. Georgila, J. Gratch, A. Hartholt, M. Lhommet et al., "Simsensei kiosk: A virtual human interviewer for healthcare decision support," in *Proceedings of the 2014 international conference on Autonomous agents and multi-agent systems.* International Foundation for Autonomous Agents and Multiagent Systems, 2014, pp. 1061–1068.

[52] P. C. Cozby, "Self-disclosure: a literature review." *Psychological bulletin*, vol. 79, no. 2, p. 73, 1973.

[53] C. T. Hill and D. E. Stull, "Gender and self-disclosure," in *Self-Disclosure.* Springer, 1987, pp. 81–100.

[54] V. J. Derlaga and J. H. Berg, *Self-disclosure: Theory, research, and therapy.* Springer Science & Business Media, 1987.

[55] J. Hanson, "Should your lips be zipped? how therapist self-disclosure and non-disclosure affects clients," *Counselling and Psychotherapy Research*, vol. 5, no. 2, pp. 96–104, 2005.

[56] S. Cohen, W. Nutt, and Y. Sagic, "Deciding equivalances among conjunctive aggregate queries," *J. ACM*, vol. 54, no. 2, Apr. 2007. [Online]. Available: http://doi.acm.org/10.1145/1219092.1219093

[57] C. E. Hill, S. Knox, and K. G. Pinto-Coelho, "Therapist self-disclosure and immediacy: A qualitative meta-analysis." *Psychotherapy*, vol. 55, no. 4, p. 445, 2018.

[58] J. R. Henretty and H. M. Levitt, "The role of therapist self-disclosure in psychotherapy: A qualitative review," *Clinical psychology review*, vol. 30, no. 1, pp. 63–77, 2010.

[59] G. D. Bodie, A. J. Vickery, K. Cannava, and S. M. Jones, "The role of "active listening" in informal helping conversations: Impact on perceptions of listener helpfulness, sensitivity, and supportiveness and discloser emotional improvement," *Western Journal of Communication*, vol. 79, no. 2, pp. 151–173, 2015.

[60] R. Kuhn, T. N. Bradbury, F. W. Nussbeck, and G. Bodenmann, "The power of listening: Lending an ear to the partner during dyadic coping conversations." *Journal of Family Psychology*, vol. 32, no. 6, p. 762, 2018.

[61] J. Torous and L. W. Roberts, "Needed innovation in digital health and smartphone applications for mental health: transparency and trust," *JAMA psychiatry*, vol. 74, no. 5, pp. 437–438, 2017.

[62] L. Clark, N. Pantidi, O. Cooney, P. Doyle, D. Garaialde, J. Edwards, B. Spillane, E. Gilmartin, C. Murad, C. Munteanu et al., "What makes a good conversation?: Challenges in designing truly conversational agents," in *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. ACM, 2019, p. 475.

[63] A. Utley and Y. Garza, "The therapeutic use of journaling with adolescents," *Journal of Creativity in Mental Health*, vol. 6, no. 1, pp. 29–41, 2011.

[64] A. Fadhil and G. Schiavo, "Designing for health chatbots," *ArXiv*, vol. abs/1902.09022, 2019.

[65] H.-Y. Huang, "Examining the beneficial effects of individual's self-disclosure on the social network site," *Computers in human behavior*, vol. 57, pp. 122–132, 2016.

[66] X. Ma, J. Hancock, and M. Naaman, "Anonymity, intimacy and self-disclosure in social media," in *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, ser. CHI '16. New York, NY, USA: ACM, 2016. [Online]. Available: http://doi.acm.org/10.1145/2858036.2858414 pp. 3857–3869.

[67] G. M. Lucas, J. Gratch, A. King, and L.-P. Morency, "It's only a computer: Virtual humans increase willingness to disclose," *Computers in Human Behavior*, vol. 37, pp. 94–100, 2014.

[68] M. De Choudhury and S. De, "Mental health discourse on reddit: Self-disclosure, social support, and anonymity," in *Eighth International AAAI Conference on Weblogs and Social Media*, 2014.

[69] N. Andalibi, M. E. Morris, and A. Forte, "Testing waters, sending clues: Indirect disclosures of socially stigmatized experiences on social media," *Proceedings of the ACM on Human-Computer Interaction*, vol. 2, no. CSCW, pp. 1–23, 2018.

[70] R. Zhang, J. Eschler, and M. Reddy, "Online support groups for depression in china: Culturally shaped interactions and motivations," *Computer Supported Cooperative Work (CSCW)*, vol. 27, no. 3-6, pp. 327–354, 2018.

[71] N. Andalibi, P. Ozturk, and A. Forte, "Sensitive self-disclosures, responses, and social support on instagram: the case of# depression," in *Proceedings of the 2017 ACM conference on computer supported cooperative work and social computing*, 2017, pp. 1485–1500.

[72] D. Yang, Z. Yao, J. Seering, and R. Kraut, "The channel matters: Self-disclosure, reciprocity and social support in online cancer support groups," in *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, 2019, pp. 1–15.

[73] C. W. Schmidt, "Environmental connections: a deeper look into mental illness," 2007.

[74] N. Andalibi, O. L. Haimson, M. De Choudhury, and A. Forte, "Understanding social media disclosures of sexual abuse through the lenses of support seeking and anonymity," in *Proceedings of the 2016 CHI conference on human factors in computing systems*, 2016, pp. 3906–3918.

[75] S. K. Ernala, A. F. Rizvi, M. L. Birnbaum, J. M. Kane, and M. De Choudhury, "Linguistic markers indicating therapeutic outcomes of social media disclosures of schizophrenia," *Proceedings of the ACM on Human-Computer Interaction*, vol. 1, no. CSCW, pp. 1–27, 2017.

[76] G. Doherty, D. Coyle, and M. Matthews, "Design and evaluation guidelines for mental health technologies," *Interacting with computers*, vol. 22, no. 4, pp. 243–252, 2010.

[77] M. J. Won-Doornink, "Self-disclosure and reciprocity in conversation: A cross-national study," *Social Psychology Quarterly*, pp. 97–107, 1985.

[78] M. R. Goldfried, L. A. Burckell, and C. Eubanks-Carter, "Therapist self-disclosure in cognitive-behavior therapy," *Journal of clinical psychology*, vol. 59, no. 5, pp. 555–568, 2003.

[79] S. Knox and C. E. Hill, "Therapist self-disclosure: Research-based suggestions for practitioners," *Journal of clinical psychology*, vol. 59, no. 5, pp. 529–539, 2003.

[80] J.-E. R. Lee and C. I. Nass, "Trust in computers: The computers-are-social-actors (casa) paradigm and trustworthiness perception in human-computer communication," in *Trust and technology in a ubiquitous modern environment: Theoretical and methodological perspectives*. IGI Global, 2010, pp. 1–15.

[81] S. Kim, J. Lee, and G. Gweon, "Comparing data from chatbot and web surveys: Effects of platform and conversational style on survey response quality," in *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, ser. CHI '19. New York, NY, USA: ACM, 2019. [Online]. Available: http://doi.acm.org/10.1145/3290605.3300316 pp. 86:1–86:12.

[82] D. Elmasri and A. Maeder, "A conversational agent for an online mental health intervention," in *International Conference on Brain Informatics*. Springer, 2016, pp. 243–251.

[83] H. Kreiner and Y. Levi-Belz, "Self-disclosure here and now: Combining retrospective perceived assessment with dynamic behavioral measures," *Frontiers in psychology*, vol. 10, 2019.

[84] J. Zhang, Y. J. Oh, P. Lange, Z. Yu, and Y. Fukuoka, "Artificial intelligence chatbot behavior change model for designing artificial intelligence chatbots to promote physical activity and a healthy diet," *Journal of Medical Internet Research*, vol. 22, no. 9, p. e22845, 2020.

[85] R. Winkler, S. Hobert, A. Salovaara, M. Söllner, and J. M. Leimeister, "Sara, the lecturer: Improving learning in online education with a scaffolding-based conversational agent," in *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, 2020, pp. 1–14.

[86] J. Wang, H. Yang, R. Shao, S. Abdullah, and S. S. Sundar, "Alexa as coach: Leveraging smart speakers to build social agents that reduce public speaking anxiety," in *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, ser. CHI '20. New York, NY, USA: Association for Computing Machinery, 2020. [Online]. Available: https://doi.org/10.1145/3313831.3376561 p. 1–13.

[87] I. Benbasat and W. Wang, "Trust in and adoption of online recommendation agents," *Journal of the association for information systems*, vol. 6, no. 3, p. 4, 2005.

[88] S. Lee and J. Choi, "Enhancing user experience with conversational agent for movie recommendation: Effects of self-disclosure and reciprocity," *International Journal of Human-Computer Studies*, vol. 103, pp. 95–105, 2017.

[89] S. Gabrielli, S. Rizzi, S. Carbone, and V. Donisi, "A chatbot-based coaching intervention for adolescents to promote life skills: Pilot study," *JMIR Human Factors*, vol. 7, no. 1, p. e16762, 2020.

[90] A. S. Miner, N. Shah, K. D. Bullock, B. A. Arnow, J. Bailenson, and J. Hancock, "Key considerations for incorporating conversational ai in psychotherapy," *Frontiers in psychiatry*, vol. 10, 2019.

[91] D. Mohr, P. Cuijpers, and K. Lehman, "Supportive accountability: a model for providing human support to enhance adherence to ehealth interventions," *Journal of medical Internet research*, vol. 13, no. 1, p. e30, 2011.

[92] T. Bickmore, D. Schulman, and L. Yin, "Maintaining engagement in long-term interventions with relational agents," *Applied Artificial Intelligence*, vol. 24, no. 6, pp. 648–666, 2010.

[93] D. D. Ebert, D. Lehr, F. Smit, A.-C. Zarski, H. Riper, E. Heber, P. Cuijpers, and M. Berking, "Efficacy and cost-effectiveness of minimal guided and unguided internet-based mobile supported stress-management in employees with occupational stress: a three-armed randomised controlled trial," *BMC Public Health*, vol. 14, no. 1, p. 807, 2014.

[94] G. Andersson and P. Cuijpers, "Internet-based and other computerized psychological treatments for adult depression: a meta-analysis," *Cognitive behaviour therapy*, vol. 38, no. 4, pp. 196–205, 2009.

[95] S. M. Schueller, K. N. Tomasino, E. G. Lattie, and D. C. Mohr, "Human support for behavioral intervention technologies for mental health: the efficiency model," *management*, vol. 21, p. 22, 2016.

[96] S. M. Schueller, K. N. Tomasino, and D. C. Mohr, "Integrating human support into behavioral intervention technologies: the efficiency model of support," *Clinical Psychology: Science and Practice*, vol. 24, no. 1, pp. 27–45, 2017.

[97] A. van Heerden, X. Ntinga, and K. Vilakazi, "The potential of conversational agents to provide a rapid hiv counseling and testing services," in *2017 international conference on the frontiers and advances in data science (FADS)*. IEEE, 2017, pp. 80–85.

[98] R. Kumar and C. P. Rose, "Architecture for building conversational agents that support collaborative learning," *IEEE Transactions on Learning Technologies*, vol. 4, no. 1, pp. 21–34, 2010.

[99] W. Duan, N. Yamashita, S. Y. Hwang, and S. Fussell, ""let me ask them to clarify if you don't want to"-a clarification agent for nonnative speakers," in *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems*, 2018, pp. 1–6.

[100] A. P. Chaves and M. A. Gerosa, "Single or multiple conversational agents? an interactional coherence comparison," in *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, 2018, pp. 1–13.

[101] A. Bandura and D. C. McClelland, *Social learning theory*. Englewood cliffs Prentice Hall, 1977, vol. 1.

[102] E. A. Bajcar and P. Babel, "How does observational learning produce placebo effects? a model integrating research findings," *Frontiers in psychology*, vol. 9, p. 2041, 2018.

[103] K. E. Matthews, V. Andrews, and P. Adams, "Social learning spaces and student engagement," *Higher Education Research & Development*, vol. 30, no. 2, pp. 105–120, 2011.

[104] M. Hajli, H. Bugshan, X. Lin, and M. Featherman, "From e-learning to social learning–a health care study," *European Journal of Training and Development*, 2013.

[105] V. Franklin, A. Greene, A. Waller, S. Greene, and C. Pagliari, "Patients' engagement with "sweet talk"–a text messaging support system for young people with diabetes," *Journal of medical Internet research*, vol. 10, no. 2, p. e20, 2008.

[106] H. R. Walen and M. E. Lachman, "Social support and strain from partner, family, and friends: Costs and benefits for men and women in adulthood," *Journal of social and personal relationships*, vol. 17, no. 1, pp. 5–30, 2000.

[107] A. Grimes, M. Bednar, J. D. Bolter, and R. E. Grinter, "Eatwell: sharing nutrition-related memories in a low-income community," in *Proceedings of the 2008 ACM conference on Computer supported cooperative work*, 2008, pp. 87–96.

[108] H. Oinas-Kukkonen and M. Harjumaa, "Persuasive systems design: Key issues, process model, and system features," *Communications of the Association for Information Systems*, vol. 24, no. 1, p. 28, 2009.

[109] R. Whittaker, R. Maddison, H. McRobbie, C. Bullen, S. Denny, E. Dorey, M. Ellis-Pegler, J. van Rooyen, and A. Rodgers, "A multimedia mobile phone–based youth smoking cessation intervention: findings from content development and piloting studies," *Journal of Medical Internet Research*, vol. 10, no. 5, p. e49, 2008.

[110] P. Klasnja and W. Pratt, "Healthcare in the pocket: mapping the space of mobile-phone health interventions," *Journal of biomedical informatics*, vol. 45, no. 1, pp. 184–198, 2012.

[111] K. O'Leary, S. M. Schueller, J. O. Wobbrock, and W. Pratt, ""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*, 2018, pp. 1–14.

[112] A. H. Jordan, B. Monin, C. S. Dweck, B. J. Lovett, O. P. John, and J. J. Gross, "Misery has more company than people think: Underestimating the prevalence of others' negative emotions," *Personality and Social Psychology Bulletin*, vol. 37, no. 1, pp. 120–135, 2011.

[113] G. Li, X. Zhou, T. Lu, J. Yang, and N. Gu, "Sunforum: Understanding depression in a chinese online community," in *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing*, 2016, pp. 515–526.

[114] F. X. Gibbons and B. P. Buunk, "Individual differences in social comparison: development of a scale of social comparison orientation." *Journal of personality and social psychology*, vol. 76, no. 1, p. 129, 1999.

[115] M.-L. N. Steers, R. E. Wickham, and L. K. Acitelli, "Seeing everyone else's highlight reels: How facebook usage is linked to depressive symptoms," *Journal of Social and Clinical Psychology*, vol. 33, no. 8, pp. 701–731, 2014.

[116] O. E. Nordberg, J. D. Wake, E. S. Nordby, E. Flobak, T. Nordgreen, S. K. Mukhiya, and F. Guribye, "Designing chatbots for guiding online peer support conversations for adults with adhd," in *International Workshop on Chatbot Research and Design*. Springer, 2019, pp. 113–126.

[117] Y.-C. Wang, M. Burke, and R. Kraut, "Modeling self-disclosure in social networking sites," in *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing*, ser. CSCW '16. New York, NY, USA: ACM, 2016. [Online]. Available: http://doi.acm.org/10.1145/2818048.2820010 pp. 74–85.

[118] A. Barak and O. Gluck-Ofri, "Degree and reciprocity of self-disclosure in online forums," *CyberPsychology & Behavior*, vol. 10, no. 3, pp. 407–417, 2007.

[119] P. Corrigan, "How stigma interferes with mental health care." *American psychologist*, vol. 59, no. 7, p. 614, 2004.

[120] C. Lauber and W. Rössler, "Stigma towards people with mental illness in developing countries in asia," *International review of psychiatry*, vol. 19, no. 2, pp. 157–178, 2007.

[121] J. Kim, Y. Kim, B. Kim, S. Yun, M. Kim, and J. Lee, "Can a machine tend to teenagers' emotional needs?: A study with conversational agents," in *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems*. ACM, 2018, p. LBW018.

[122] T. Bickmore and A. Gruber, "Relational agents in clinical psychiatry," *Harvard review of psychiatry*, vol. 18, no. 2, pp. 119–130, 2010.

[123] L. C. Miller and D. A. Kenny, "Reciprocity of self-disclosure at the individual and dyadic levels: A social relations analysis." *Journal of Personality and Social Psychology*, vol. 50, no. 4, p. 713, 1986.

[124] A. A. Abd-Alrazaq, M. Alajlani, A. A. Alalwan, B. M. Bewick, P. Gardner, and M. Househ, "An overview of the features of chatbots in mental health: A scoping review," *International Journal of Medical Informatics*, vol. 132, p. 103978, 2019.

[125] T. Nadarzynski, O. Miles, A. Cowie, and D. Ridge, "Acceptability of artificial intelligence (ai)-led chatbot services in healthcare: A mixed-methods study," *Digital health*, vol. 5, p. 2055207619871808, 2019.

[126] Z. Xiao, M. X. Zhou, Q. V. Liao, G. Mark, C. Chi, W. Chen, and H. Yang, "Tell me about yourself: Using an ai-powered chatbot to conduct conversational surveys," *arXiv preprint arXiv:1905.10700*, 2019.

[127] T. Bickmore, D. Schulman, and L. Yin, "Engagement vs. deceit: Virtual humans with human autobiographies," in *International Workshop on Intelligent Virtual Agents*. Springer, 2009, pp. 6–19.

[128] P. M. Ullrich and S. K. Lutgendorf, "Journaling about stressful events: Effects of cognitive processing and emotional expression," *Annals of Behavioral Medicine*, vol. 24, no. 3, pp. 244–250, 2002.

[129] T. Bickmore and J. Cassell, "Small talk and conversational storytelling in embodied conversational interface agents," in *AAAI fall symposium on narrative intelligence*, 1999, pp. 87–92.

[130] A. Aron, E. Melinat, E. N. Aron, R. D. Vallone, and R. J. Bator, "The experimental generation of interpersonal closeness: A procedure and some preliminary findings," *Personality and Social Psychology Bulletin*, vol. 23, no. 4, pp. 363–377, 1997.

[131] S. M. Jourard and P. Lasakow, "Some factors in self-disclosure." *The Journal of Abnormal and Social Psychology*, vol. 56, no. 1, p. 91, 1958.

[132] J. Hunt and D. Eisenberg, "Mental health problems and help-seeking behavior among college students," *Journal of adolescent health*, vol. 46, no. 1, pp. 3–10, 2010.

[133] J. J. Prochaska, H.-Y. Sung, W. Max, Y. Shi, and M. Ong, "Validity study of the k6 scale as a measure of moderate mental distress based on mental health treatment need and utilization," *International journal of methods in psychiatric research*, vol. 21, no. 2, pp. 88–97, 2012.

[134] Y. R. Tausczik and J. W. Pennebaker, "The psychological meaning of words: Liwc and computerized text analysis methods," *Journal of language and social psychology*, vol. 29, no. 1, pp. 24–54, 2010.

[135] T. Dinev and P. Hart, "Privacy concerns and levels of information exchange: An empirical investigation of intended e-services use," *E-Service*, vol. 4, no. 3, pp. 25–60, 2006.

[136] E. Berscheid, M. Snyder, and A. M. Omoto, "The relationship closeness inventory: Assessing the closeness of interpersonal relationships." *Journal of personality and Social Psychology*, vol. 57, no. 5, p. 792, 1989.

[137] H. Van der Heijden, "Factors influencing the usage of websites: the case of a generic portal in the netherlands," *Information & management*, vol. 40, no. 6, pp. 541–549, 2003.

[138] R. A. Emmons and R. Stern, "Gratitude as a psychotherapeutic intervention," *Journal of clinical psychology*, vol. 69, no. 8, pp. 846–855, 2013.

[139] A. Reeves, "Emotional intelligence: recognizing and regulating emotions," *Aaohn Journal*, vol. 53, no. 4, pp. 172–176, 2005.

[140] Y. Huang, Y. Tang, and Y. Wang, "Emotion map: A location-based mobile social system for improving emotion awareness and regulation," in *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing.* ACM, 2015, pp. 130–142.

[141] B. A. Farber, *Self-disclosure in psychotherapy.* Guilford Press, 2006.

[142] K. Robertson et al., "Active listening: more than just paying attention," *Australian family physician*, vol. 34, no. 12, p. 1053, 2005.

[143] M. Kito, "Self-disclosure in romantic relationships and friendships among american and japanese college students," *The Journal of social psychology*, vol. 145, no. 2, pp. 127–140, 2005.

[144] G.-M. Chen, "Differences in self-disclosure patterns among americans versus chinese: A comparative study," *Journal of cross-cultural psychology*, vol. 26, no. 1, pp. 84–91, 1995.

[145] J. Schug, M. Yuki, and W. Maddux, "Relational mobility explains between-and within-culture differences in self-disclosure to close friends," *Psychological Science*, vol. 21, no. 10, pp. 1471–1478, 2010.

[146] B. Hayman, L. Wilkes, and D. Jackson, "Journaling: Identification of challenges and reflection on strategies," *Nurse researcher*, vol. 19, no. 3, 2012.

[147] F. Cordeiro, D. A. Epstein, E. Thomaz, E. Bales, A. K. Jagannathan, G. D. Abowd, and J. Fogarty, "Barriers and negative nudges: Exploring challenges in food journaling," in *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, 2015, pp. 1159–1162.

[148] C. Moorman, G. Zaltman, and R. Deshpande, "Relationships between providers and users of market research: the dynamics of trust within and between organizations," *Journal of marketing research*, vol. 29, no. 3, pp. 314–328, 1992.

[149] X. Chen, Q. Huang, R. M. Davison, and Z. Hua, "What drives trust transfer? the moderating roles of seller-specific and general institutional mechanisms," *International Journal of Electronic Commerce*, vol. 20, no. 2, pp. 261–289, 2015.

[150] J. Zhang, "Trust transfer in the sharing economy-a survey-based approach," *Junior Management Science*, vol. 3, no. 2, pp. 1–32, 2018.

[151] K. J. Stewart, "How hypertext links influence consumer perceptions to build and degrade trust online," *Journal of Management Information Systems*, vol. 23, no. 1, pp. 183–210, 2006.

[152] N. Wang, X.-L. Shen, and Y. Sun, "Transition of electronic word-of-mouth services from web to mobile context: A trust transfer perspective," *Decision support systems*, vol. 54, no. 3, pp. 1394–1403, 2013.

[153] T. Erdem, "An empirical analysis of umbrella branding," *Journal of Marketing Research*, vol. 35, no. 3, pp. 339–351, 1998.

[154] Y. Lu, S. Yang, P. Y. Chau, and Y. Cao, "Dynamics between the trust transfer process and intention to use mobile payment services: A cross-environment perspective," *Information & Management*, vol. 48, no. 8, pp. 393–403, 2011.

[155] H. Han, C. Koo, and N. Chung, "Implication of the fit between airbnb and host characteristics: a trust-transfer perspective," in *Proceedings of the 18th Annual International Conference on Electronic Commerce: e-Commerce in Smart connected World*. ACM, 2016, p. 10.

[156] H. Chung, M. Iorga, J. Voas, and S. Lee, "Alexa, can i trust you?" *Computer*, vol. 50, no. 9, pp. 100–104, 2017.

[157] A. Følstad, C. B. Nordheim, and C. A. Bjørkli, “What makes users trust a chatbot for customer service? an exploratory interview study,” in *International Conference on Internet Science*. Springer, 2018, pp. 194–208.

[158] A. Przegalinska, L. Ciechanowski, A. Stroz, P. Gloor, and G. Mazurek, “In bot we trust: A new methodology of chatbot performance measures,” *Business Horizons*, vol. 62, no. 6, pp. 785–797, 2019.

[159] A. Tubert, “Ethical machines,” *Seattle UL Rev.*, vol. 41, p. 1163, 2017.

[160] R. Ren, J. W. Castro, S. T. Acuña, and J. de Lara, “Usability of chatbots: A systematic mapping study,” in *The 31st International Conference on Software Engineering and Knowledge Engineering, SEKE 2019, Hotel Tivoli, Lisbon, Portugal, July 10-12, 2019.*, 2019. [Online]. Available: https://doi.org/10.18293/SEKE2019-029 pp. 479–617.

[161] J. A. Hinson and J. L. Swanson, “Willingness to seek help as a function of self-disclosure and problem severity,” *Journal of Counseling & Development*, vol. 71, no. 4, pp. 465–470, 1993.

[162] M. L. Jacobs, J. Clawson, and E. D. Mynatt, “Comparing health information sharing preferences of cancer patients, doctors, and navigators,” in *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*, 2015, pp. 808–818.

[163] D. L. Hubbs and C. F. Brand, “The paper mirror: Understanding reflective journaling,” *Journal of Experiential Education*, vol. 28, no. 1, pp. 60–71, 2005.

[164] D. E. Davis, E. Choe, J. Meyers, N. Wade, K. Varjas, A. Gifford, A. Quinn, J. N. Hook, D. R. Van Tongeren, B. J. Griffin et al., “Thankful for the little things: A meta-analysis of gratitude interventions.” *Journal of counseling psychology*, vol. 63, no. 1, p. 20, 2016.

[165] M. E. Seligman, T. A. Steen, N. Park, and C. Peterson, “Positive psychology progress: empirical validation of interventions.” *American psychologist*, vol. 60, no. 5, p. 410, 2005.

[166] K. A. Baikie and K. Wilhelm, “Emotional and physical health benefits of expressive writing,” *Advances in psychiatric treatment*, vol. 11, no. 5, pp. 338–346, 2005.

[167] M. E. Seligman and M. Csikszentmihalyi, “Positive psychology: An introduction,” in *Flow and the foundations of positive psychology*. Springer, 2014, pp. 279–298.

[168] S. Lyubomirsky, R. Dickerhoof, J. K. Boehm, and K. M. Sheldon, “Becoming happier takes both a will and a proper way: An experimental longitudinal intervention to boost well-being.” *Emotion*, vol. 11, no. 2, p. 391, 2011.

[169] J. W. Pennebaker, “Expressive writing in a clinical setting,” *The Independent Practitioner*, vol. 30, pp. 23–25, 2010.

[170] L. M. Rodriguez, C. M. Young, C. Neighbors, M. T. Campbell, and Q. Lu, “Evaluating guilt and shame in an expressive writing alcohol intervention,” *Alcohol*, vol. 49, no. 5, pp. 491–498, 2015.

[171] N. Park and C. Peterson, “Character strengths: Research and practice,” *Journal of college and character*, vol. 10, no. 4, 2009.

[172] K. M. Sheldon and S. Lyubomirsky, “How to increase and sustain positive emotion: The effects of expressing gratitude and visualizing best possible selves,” *The journal of positive psychology*, vol. 1, no. 2, pp. 73–82, 2006.

[173] H. L. O’Brien, P. Cairns, and M. Hall, “A practical approach to measuring user engagement with the refined user engagement scale (ues) and new ues short form,” *International Journal of Human-Computer Studies*, vol. 112, pp. 28–39, 2018.

[174] A. M. Grant, J. Franklin, and P. Langford, “The self-reflection and insight scale: A new measure of private self-consciousness,” *Social Behavior and Personality: an international journal*, vol. 30, no. 8, pp. 821–835, 2002.

[175] J. M. Govern and L. A. Marsch, “Development and validation of the situational self-awareness scale,” *Consciousness and cognition*, vol. 10, no. 3, pp. 366–378, 2001.

[176] T. W. Bickmore, L. Caruso, K. Clough-Gorr, and T. Heeren, “‘it’s just like you talk to a friend’relational agents for older adults,” *Interacting with Computers*, vol. 17, no. 6, pp. 711–735, 2005.

[177] H. Oinas-Kukkonen, “A foundation for the study of behavior change support systems,” *Personal and ubiquitous computing*, vol. 17, no. 6, pp. 1223–1235, 2013.

[178] A. Bandura, “Human agency in social cognitive theory.” *American psychologist*, vol. 44, no. 9, p. 1175, 1989.

[179] J. Fulk, J. Schmitz, and C. W. Steinfield, “A social influence model of technology use,” *Organizations and communication technology*, vol. 117, p. 140, 1990.

[180] A. Radovic, T. Gmelin, B. D. Stein, and E. Miller, “Depressed adolescents’ positive and negative use of social media,” *Journal of adolescence*, vol. 55, pp. 5–15, 2017.

[181] A. Ho, J. Hancock, and A. S. Miner, “Psychological, relational, and emotional effects of self-disclosure after conversations with a chatbot,” *Journal of Communication*, vol. 68, no. 4, pp. 712–733, 2018.

[182] K. O’Leary, A. Bhattacharya, S. A. Munson, J. O. Wobbrock, and W. Pratt, “Design opportunities for mental health peer support technologies,” in *Proceedings of the 2017 ACM conference on computer supported cooperative work and social computing*, 2017, pp. 1470–1484.

[183] V. Braun and V. Clarke, “Using thematic analysis in psychology,” *Qualitative research in psychology*, vol. 3, no. 2, pp. 77–101, 2006.

[184] T. Mussweiler, "Comparison processes in social judgment: mechanisms and consequences." *Psychological review*, vol. 110, no. 3, p. 472, 2003.

[185] S. Deri, S. Davidai, and T. Gilovich, "Home alone: Why people believe others' social lives are richer than their own." *Journal of Personality and Social Psychology*, vol. 113, no. 6, p. 858, 2017.

[186] A. Bandura, "Health promotion by social cognitive means," *Health education & behavior*, vol. 31, no. 2, pp. 143–164, 2004.

[187] J. M. Royce, K. Corbett, G. Sorensen, and J. Ockene, "Gender, social pressure, and smoking cessations: the community intervention trial for smoking cessation (commit) at baseline," *Social science & medicine*, vol. 44, no. 3, pp. 359–370, 1997.

[188] R. Gasser, D. Brodbeck, M. Degen, J. Luthiger, R. Wyss, and S. Reichlin, "Persuasiveness of a mobile lifestyle coaching application using social facilitation," in *International Conference on Persuasive Technology.* Springer, 2006, pp. 27–38.

[189] A. Rodgers, T. Corbett, D. Bramley, T. Riddell, M. Wills, R.-B. Lin, and M. Jones, "Do u smoke after txt? results of a randomised trial of smoking cessation using mobile phone text messaging," *Tobacco control*, vol. 14, no. 4, pp. 255–261, 2005.

[190] K. Kretzschmar, H. Tyroll, G. Pavarini, A. Manzini, I. Singh, and N. Y. P. A. Group, "Can your phone be your therapist? young people's ethical perspectives on the use of fully automated conversational agents (chatbots) in mental health support," *Biomedical Informatics Insights*, vol. 11, p. 1178222619829083, 2019.

[191] M. Lee, S. Ackermans, N. van As, H. Chang, E. Lucas, and W. IJsselsteijn, "Caring for vincent: A chatbot for self-compassion," in *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, 2019, pp. 1–13.