# "I Hear You, I Feel You": Encouraging Deep Self-disclosure through a Chatbot

**Conference Paper** · April 2020

DOI: 10.1145/3313831.3376175



**4 authors**, including:

Yi-Chieh Lee
National University of Singapore


Naomi Yamashita
NTT


Yun Huang
University of Illinois, Urbana-Champaign

# "I Hear You, I Feel You": Encouraging Deep Self-disclosure through a Chatbot

**Yi-Chieh Lee**[1 2], **Naomi Yamashita**[2], **Yun Huang**[1], **Wai Fu**[1]
[1]University of Illinois Urbana-Champaign
[2]NTT Communication Science Laboratories
ylee267@illinois.edu, naomiy@acm.org, yunhuang@illinois.edu, wfu.uiuc@gmail.com

## ABSTRACT
Chatbots have great potential to serve as a low-cost, effective tool to support people's self-disclosure. Prior work has shown that reciprocity occurs in human-machine dialog; however, whether reciprocity can be leveraged to promote and sustain deep self-disclosure over time has not been systematically studied. In this work, we design, implement and evaluate a chatbot that has self-disclosure features when it performs small talk with people. We ran a study with 47 participants and divided them into three groups to use different chatting styles of the chatbot for three weeks. We found that chatbot self-disclosure had a reciprocal effect on promoting deeper participant self-disclosure that lasted over the study period, which the other chat styles without self-disclosure features failed to deliver. Chatbot self-disclosure also had a positive effect on improving participants' perceived intimacy and enjoyment over the study period. Finally, we reflect on the design implications of chatbots where deep self-disclosure is needed over time.

## Author Keywords
Chatbot; Self-disclosure, Mental well-being;

## CCS Concepts
•**Applied computing** → *Psychology;* •**Human-centered computing** → **User studies;**

## INTRODUCTION
Self-disclosure is a process in which a person reveals personal or sensitive information to others [23, 1] and is crucial for developing a strong interpersonal relationship [1]. The advancement of computing technologies has enabled new ways for people to self-disclose [47, 33]. 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 [10, 3]. Interviewees may disclose themselves more openly in an interview session when using virtual agents [33, 12]. The challenge is that people naturally avoid revealing their vulnerabilities to others [8, 29].

Chatbots (also called conversational agents) have great potential to create breakthroughs in self-disclosure research [33, 41], 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 [26]. 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 [17]. Similarly, several works demonstrated the potential benefits of using chatbots for mental wellbeing [48, 25, 6]. Recently, Ravichander et al. also shared their findings that reciprocity could occur in human-machine dialog [41]. 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, [39, 27], require sustained self-disclosure of sensitive topics over a period of time.

In this work, we design, implement and evaluate a chatbot that has self-disclosure features when it performs small talk with people. We 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). We 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.

Our work makes the following contributions to the HCI community: 1) we explore how varied levels of a chatbot's self-disclosure influence the depth of people's self-disclosure, 2) we contribute new understandings of how time plays a role in chatbot and people's self-disclosure interactions, and 3) our findings also provide new implications into designing and using chatbots where deep self-disclosure is needed.



*CHI'20,* April 25–30, 2020, Honolulu, HI, USA



ISBN 978-1-4503-6708-0/20/04. . . $15.00

DOI: 10.1145/3313831.3376175

## RELATED WORK

*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 [23, 1]. A leading explanation of the self-disclosure process is social penetration theory (SPT) [1], 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. As such, self-disclosure can be evaluated from two main dimensions, breadth and depth [38]. The breadth dimension denotes wide-ranging discussion of multiple topics such as music preferences and food, whereas the depth dimension comprises more personal details and intimate topics such as sexual relationship and self-perceived failures.

Self-disclosure plays an important role in a wide range of settings, including mental well-being [39], customer service [36], and employment [33, 12]; thus extensive research has been conducted on self-disclosure's relationships to various constructs including trust [49], intimacy [9], gender [19], and personality [9]. 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 [11]. 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.

However, disclosing personal mental health information is not easy for most people, and this is also one of the major practical difficulties in counseling sessions [18]. 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 [33]. It is not clear how people disclose when facing a different conversational agent design. For example, Clark et al. [7] 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.

### 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 [20]. Ma et al. [34] 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 [33, 32].

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

Recently, conversational agents have been used to guide users to healthier lifestyle choices and improve their mental well-being [50, 30, 14] and engage people in truthful interactions [41]. For example, Moon [36] 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 [41] chatbot provided conversationally relevant self-disclosures from a large conversation dataset in real time, such that it engaged users with reciprocity in social conversations.

Though scholars have made significant progress with self-disclosure research using chatbots, major research questions, such as how chatbots can promote deep self-disclosure over time, are still under-studied. Promising application domains, e.g., mental health [39, 27] often need support tools to acquire people's sustained self-disclosure of sensitive topics over a period of time. Thus, in our work, we are interested in exploring: **RQ1: How do different chatting styles influence people's self-disclosure?** and **RQ2: How do different chatting styles influence people's self-disclosure over time?** Specifically, literature on reciprocity [35, 41] suggests that when people make deep self-disclosures, their interlocutor will feel pressure to share information at a similar level. Therefore, we 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 [37]. Based on these theories and SPT, we 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.**

## METHOD

### Chatbot Design

We built our chatbot using Manychat and Google Dialogflow. Manychat was used to allow the researchers to monitor whether the chatbot users had finished their specific chatting tasks, and to send reminders to those who had not. We 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, we 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.

In addition, Dialogflow helped to handle some exceptional questions. In any experiment of this kind, participants in-

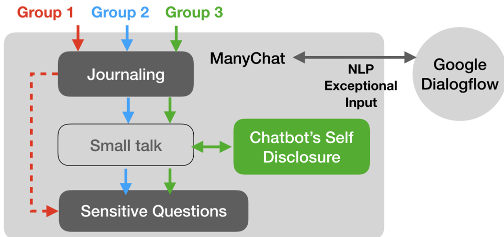


**Figure 1. Illustration of the study design. Standard questions are given to users during two sessions, i.e., *Journaling* and *Sensitive Questions*, and the chatbot does not self-disclose and only gives general responses in these two sessions. During *Small Talk* session, the chatbot gives low (high) self-disclosure to participants from group 2 (3).**

evitably 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, 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.

Instead of defining the gender and appearance of the chatbot, we 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. 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 we felt 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 other times, it would only give them simple replies to prevent users' other chatting behaviors influencing their impression of the chatbot.

### Chat Sessions

We designed and conducted a study, where we divide 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 1.

#### *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, we 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 [45] 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 [5] and to facilitate user self-disclosure. To explore our research questions, we designed two types of dialogue for Groups 2 and 3 specifically: one for LD, and the other for HD (Figure 2). 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 2). For conversational topics, two topics were randomly picked from [2] for each day. 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.

We based our sensitive questions on those used in prior studies [34, 36, 24]. 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 [22]. Thus, we adapted questions relevant to these six areas from the Jourard Sixty-item Self-disclosure Questionnaire (JSDQ) [24] and two other studies [34, 36], 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 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.

| | Small Talk Question 1 | Small Talk Question 2 |
|---|---|---|
| **Group 2 (Low Level Self-disclosure)** | CB (asking Q1 to S22): If you could only celebrate one holiday, what would it be?<br>S22: thanksgiving!<br>CB: I prefer to celebrate Christmas because it is a beautiful day. Everyone will decorate their house and prepare traditional foods for celebration.<br>S22: I like that because it always has big discount | CB (asking Q2 to S30): When was the last time you went to the zoo?<br>S30: Three month ago<br>CB: It was 2 years ago. I went their with my classmates. I think everyone likes zoo, right?<br>S30: No really, I am not a fan for that |
| **Group 3 (High Level Self-disclosure)** | CB (asking Q1 to S43): If you could only celebrate one holiday, what would it be?<br>S43: It would be new year's eve<br>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.<br>S43: Sounds nice! I also like to spend time with my friends and family to celebrate holiday. | CB (asking Q2 to S39): When was the last time you went to the zoo?<br>S39: maybe 1 year ago<br>CB: It was 2 years ago. I really enjoyed the time to handout with my classmates. It was a pleasant memory. How about you?<br>S39: I like panda, they are so cute. Yes, I agree zoo is a good place to go with friends. |

**Figure 2. Sample Chatbot's conversations with LD and HD participants during small talks with self-disclosure. CB stands for chatbot.**

## Recruitment and Participants

We posted notices on social-media websites and on a university's 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 3) that their Kessler Psychological Distress Scale (K6) scores were lower than 13 [40], 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). We divided them into three groups of roughly equal size that were balanced by gender and K-6 score, because prior studies [40] have indicated the potential effect of gender [19] and mental status [9] 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), 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). We deployed our chatbot on Facebook Messenger, 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.

## 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 1). 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 we 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, we 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).

## Measurement

### *Conversation Logs*

All of the participants' conversations with the chatbot were recorded, and because all groups answered both journaling questions and sensitive questions, we compared these two types of conversation across all three participant groups. Prior research has indicated that word count is positively associated with self-disclosure [27]. Hence, we utilized LIWC2015 [44] 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

| | 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. Sample participants' responses to sensitive questions. The responses were coded to different topics and levels of self-disclosure according to the framework proposed in [3]**

adapting the categories and levels proposed by Barak and Gluck-Ofri [3]. 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.

To analyze how different levels of chatbot's self-disclosure influenced the participants' responses (self-disclosure) to journaling and sensitive questions, we 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*

We 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, we 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. We 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, we asked them to reflect on whether talking with the chatbot every day provided them with any new insights into their daily lives.

We 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 [13], intimacy [4], and enjoyment [31, 46] - were measured through the same survey twice: after the first and third week of using the chatbot. We 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, we 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, we 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 [13, 4, 31], and all were responded to via the same seven-point Likert scale (ranging from 1=strongly disagree to 7=strongly agree).

We 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.

## RESULTS

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

All participants were asked about their emotions and daily activities on every day of the experiment. Although we rephrased 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)

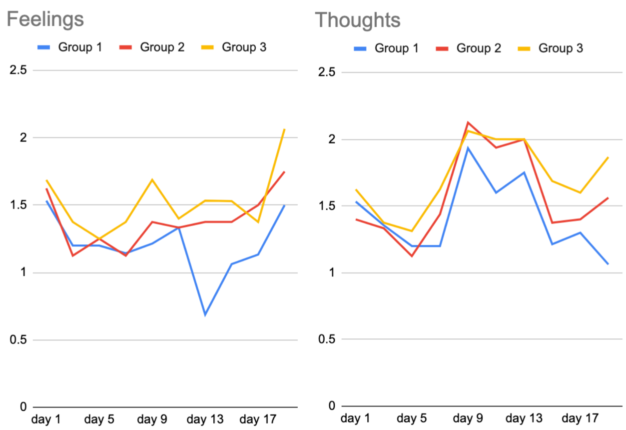


**Figure 4. The average self-disclosure level of different groups over time. They show the average levels of self-disclosure for Thoughts & Feelings across the 20 days. In the first week, the self-disclosure levels were similar among the three groups; the difference increased around day 9, with HD being the highest and ND being the lowest to disclose their thoughts.**

(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.

### 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 being non-significant. Figure 4 (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 4 (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, we 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.

### 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:** 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:** 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:** 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)*

### Bond with the Chatbot (H2)

By examining the perceived intimacy and enjoyment of conversing with the chatbot over time, we 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.

### Sustained Interactions with Chatbots (RQ2)

Overall, we 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:** 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 and 3:** 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 we 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 we 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)*

## DISCUSSION

A major contribution of our work is that our results showed an effective chatbot design that promoted deep self-disclosure over time. We investigated the effect of the self-disclosing chatbot on the depth of people's self-disclosure over three weeks, and studied the effect across two chat sessions. Our results not only showed that the chatbot's self-disclosure level has a stronger effect on user's deep self-disclosure over time, but also explained how factors contributed to the effect. These findings extend knowledge of how chatbot designs and time influence users' depth of self-disclosure, which benefit future chatbot design for mental wellbeing.

### Depth of Self-disclosure

With regard to RQ1, on how the chatbot's conversational styles influenced users' self-disclosure behavior, we 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 [41, 12], 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 [15], to explore these and other potential question-type effects.

### 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, we found that there were interaction effects of experiment day and group on the disclosure of both thoughts and feelings. Figure 4 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 [38] 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 [30] 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, we still found that HD and LD wrote longer responses than ND, which might be explained by the norm of reciprocity [3], 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 [42, 21] 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, we 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 we may have been observing conversations about them naturally 'tailing off' to avoid repetition.

### Ethics

This work explores effective chatbot designs for eliciting users' deep self-disclosure, 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, which is relevant to our research contexts, thus we discuss ethical issues as follows by referring to the perspectives addressed in [28].

*Privacy and transparency:* Some participants provided extremely sensitive content when chatting with the chatbot, 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 our research, we 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, we find, 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 any permission.

*Efficacy:* Some participants mentioned in their interviews 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 what effects/risks to expect from the chatbot.

*Safety:* In this study, although we recruited participants who were less likely exposed to serious psychological distress, users' deep self-disclosure may still arouse users' negative experience and thoughts. To address unwanted situations, we had experienced psychiatrists review our chatbot and study design. We also provided the participants with emergency contact information so that they can ask for help in case of an emergency. For real use, an effective monitory mechanism might be further necessary for addressing unexpected psychological crises and stop the participation.

### 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) [45], 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 [41], 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 [15, 18] 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 [16, 33]. By integrating machine learning that assess users' self-disclosure content [47, 10], 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 [10] 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 [43], since in our data, too little interactivity led ND members to feel that the chatbot was disrespectful.

### Limitation and Future Work

This study has some limitations. First, we 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 we 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, we 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 [1], intimacy and trust may be built over time, thus, we included both constructs in the study. There are other constructs that could be measured [16], but our measurements are not meant to be exhaustive. 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.

## CONCLUSION

In this study, we 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.

## ACKNOWLEDGMENTS

This work is supported by Grant for Scientific Research (A) 17H00771 from Japan Society for the Promotion of Science (JSPS). We thank all reviewers' comments and suggestions to help polish this paper.

## REFERENCES


[1] Irwin Altman and Dalmas A Taylor. 1973. *Social penetration: The development of interpersonal relationships.* Holt, Rinehart & Winston.

[2] Arthur Aron, Edward Melinat, Elaine N Aron, Robert Darrin Vallone, and Renee J Bator. 1997. The experimental generation of interpersonal closeness: A procedure and some preliminary findings. *Personality and Social Psychology Bulletin* 23, 4 (1997), 363–377.

[3] Azy Barak and Orit Gluck-Ofri. 2007. Degree and reciprocity of self-disclosure in online forums. *CyberPsychology & Behavior* 10, 3 (2007), 407–417.

[4] Ellen Berscheid, Mark Snyder, and Allen M Omoto. 1989. The relationship closeness inventory: Assessing the closeness of interpersonal relationships. *Journal of personality and Social Psychology* 57, 5 (1989), 792.

[5] Timothy Bickmore and Justine Cassell. 1999. Small talk and conversational storytelling in embodied conversational interface agents. In *AAAI fall symposium on narrative intelligence*. 87–92.

[6] Timothy Bickmore and Amanda Gruber. 2010. Relational agents in clinical psychiatry. *Harvard review of psychiatry* 18, 2 (2010), 119–130.

[7] Leigh Clark, Nadia Pantidi, Orla Cooney, Philip Doyle, Diego Garaialde, Justin Edwards, Brendan Spillane, Emer Gilmartin, Christine Murad, Cosmin Munteanu, and others. 2019. 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, 475.

[8] Patrick Corrigan. 2004. How stigma interferes with mental health care. *American psychologist* 59, 7 (2004), 614.

[9] Paul C Cozby. 1973. Self-disclosure: a literature review. *Psychological bulletin* 79, 2 (1973), 73.

[10] Munmun De Choudhury and Sushovan De. 2014. Mental health discourse on reddit: Self-disclosure, social support, and anonymity. In *Eighth International AAAI Conference on Weblogs and Social Media*.

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

[12] David DeVault, Ron Artstein, Grace Benn, Teresa Dey, Ed Fast, Alesia Gainer, Kallirroi Georgila, Jon Gratch, Arno Hartholt, Margaux Lhommet, and others. 2014. 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, 1061–1068.

[13] Tamara Dinev and Paul Hart. 2006. Privacy concerns and levels of information exchange: An empirical investigation of intended e-services use. *E-Service* 4, 3 (2006), 25–60.

[14] Gavin Doherty, David Coyle, and Mark Matthews. 2010. Design and evaluation guidelines for mental health technologies. *Interacting with computers* 22, 4 (2010), 243–252.

[15] Robert A Emmons and Robin Stern. 2013. Gratitude as a psychotherapeutic intervention. *Journal of clinical psychology* 69, 8 (2013), 846–855.

[16] Barry Alan Farber. 2006. *Self-disclosure in psychotherapy*. Guilford Press.

[17] Kathleen Kara Fitzpatrick, Alison Darcy, and Molly Vierhile. 2017. Delivering cognitive behavior therapy to young adults with symptoms of depression and anxiety using a fully automated conversational agent (Woebot): a randomized controlled trial. *JMIR mental health* 4, 2 (2017), e19.

[18] Jean Hanson. 2005. Should your lips be zipped? How therapist self-disclosure and non-disclosure affects clients. *Counselling and Psychotherapy Research* 5, 2 (2005), 96–104.

[19] Charles T Hill and Donald E Stull. 1987. Gender and self-disclosure. In *Self-Disclosure*. Springer, 81–100.

[20] Hsin-Yi Huang. 2016. Examining the beneficial effects of individual's self-disclosure on the social network site. *Computers in human behavior* 57 (2016), 122–132.

[21] Yun Huang, Ying Tang, and Yang Wang. 2015. 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, 130–142.

[22] Justin Hunt and Daniel Eisenberg. 2010. Mental health problems and help-seeking behavior among college students. *Journal of adolescent health* 46, 1 (2010), 3–10.

[23] Emmi Ignatius and Marja Kokkonen. 2007. Factors contributing to verbal self-disclosure. *Nordic Psychology* 59, 4 (2007), 362–391.

[24] Sidney M Jourard and Paul Lasakow. 1958. Some factors in self-disclosure. *The Journal of Abnormal and Social Psychology* 56, 1 (1958), 91.

[25] Junhan Kim, Yoojung Kim, Byungjoon Kim, Sukyung Yun, Minjoon Kim, and Joongseek Lee. 2018. 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, LBW018.

[26] Soomin Kim, Joonhwan Lee, and Gahgene Gweon. 2019. 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 (CHI '19)*. ACM, New York, NY, USA, Article 86, 12 pages. DOI:http://dx.doi.org/10.1145/3290605.3300316

[27] Hamutal Kreiner and Yossi Levi-Belz. 2019. Self-Disclosure Here and Now: Combining Retrospective Perceived Assessment With Dynamic Behavioral Measures. *Frontiers in psychology* 10 (2019).

[28] Kira Kretzschmar, Holly Tyroll, Gabriela Pavarini, Arianna Manzini, Ilina Singh, and NeurOx Young PeopleâĂŹs Advisory Group. 2019. Can your phone be your therapist? Young peopleâĂŹs ethical perspectives on the use of fully automated conversational agents (Chatbots) in mental health support. *Biomedical informatics insights* 11 (2019), 1178222619829083.

[29] Christoph Lauber and Wulf Rössler. 2007. Stigma towards people with mental illness in developing countries in Asia. *International review of psychiatry* 19, 2 (2007), 157–178.

[30] Minha Lee, Sander Ackermans, Nena van As, Hanwen Chang, Enzo Lucas, and Wijnand IJsselsteijn. 2019. Caring for Vincent: A Chatbot for Self-Compassion. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19)*. ACM, New York, NY, USA, Article 702, 13 pages. DOI: http://dx.doi.org/10.1145/3290605.3300932

[31] SeoYoung Lee and Junho Choi. 2017. Enhancing user experience with conversational agent for movie recommendation: Effects of self-disclosure and reciprocity. *International Journal of Human-Computer Studies* 103 (2017), 95–105.

[32] Gale M Lucas, Jonathan Gratch, Aisha King, and Louis-Philippe Morency. 2014. ItâĂŹs only a computer: Virtual humans increase willingness to disclose. *Computers in Human Behavior* 37 (2014), 94–100.

[33] Gale M Lucas, Albert Rizzo, Jonathan Gratch, Stefan Scherer, Giota Stratou, Jill Boberg, and Louis-Philippe Morency. 2017. Reporting mental health symptoms: breaking down barriers to care with virtual human interviewers. *Frontiers in Robotics and AI* 4 (2017), 51.

[34] Xiao Ma, Jeff Hancock, and Mor Naaman. 2016. Anonymity, Intimacy and Self-Disclosure in Social Media. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI '16)*. ACM, New York, NY, USA, 3857–3869. DOI: http://dx.doi.org/10.1145/2858036.2858414

[35] Lynn C Miller and David A Kenny. 1986. Reciprocity of self-disclosure at the individual and dyadic levels: A social relations analysis. *Journal of Personality and Social Psychology* 50, 4 (1986), 713.

[36] Youngme Moon. 2000. Intimate exchanges: Using computers to elicit self-disclosure from consumers. *Journal of consumer research* 26, 4 (2000), 323–339.

[37] Clifford Nass, Jonathan Steuer, and Ellen R Tauber. 1994. Computers are social actors. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. ACM, 72–78.

[38] Melanie Nguyen, Yu Sun Bin, and Andrew Campbell. 2012. Comparing online and offline self-disclosure: A systematic review. *Cyberpsychology, Behavior, and Social Networking* 15, 2 (2012), 103–111.

[39] James W Pennebaker. 1995. *Emotion, disclosure, & health.* American Psychological Association.

[40] Judith J Prochaska, Hai-Yen Sung, Wendy Max, Yanling Shi, and Michael Ong. 2012. 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* 21, 2 (2012), 88–97.

[41] Abhilasha Ravichander and Alan W Black. 2018. An Empirical Study of Self-Disclosure in Spoken Dialogue Systems. In *Proceedings of the 19th Annual SIGdial Meeting on Discourse and Dialogue*. 253–263.

[42] Amy Reeves. 2005. Emotional intelligence: recognizing and regulating emotions. *Aaohn Journal* 53, 4 (2005), 172–176.

[43] Kathryn Robertson and others. 2005. Active listening: more than just paying attention. *Australian family physician* 34, 12 (2005), 1053.

[44] Yla R Tausczik and James W Pennebaker. 2010. The psychological meaning of words: LIWC and computerized text analysis methods. *Journal of language and social psychology* 29, 1 (2010), 24–54.

[45] Philip M Ullrich and Susan K Lutgendorf. 2002. Journaling about stressful events: Effects of cognitive processing and emotional expression. *Annals of Behavioral Medicine* 24, 3 (2002), 244–250.

[46] Hans Van der Heijden. 2003. Factors influencing the usage of websites: the case of a generic portal in The Netherlands. *Information & management* 40, 6 (2003), 541–549.

[47] Yi-Chia Wang, Moira Burke, and Robert Kraut. 2016. Modeling Self-Disclosure in Social Networking Sites. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing (CSCW '16)*. ACM, New York, NY, USA, 74–85. DOI: http://dx.doi.org/10.1145/2818048.2820010

[48] Joseph Weizenbaum and others. 1966. ELIZA—a computer program for the study of natural language communication between man and machine. *Commun. ACM* 9, 1 (1966), 36–45.

[49] Lawrence R Wheeless and Janis Grotz. 1977. The measurement of trust and its relationship to self-disclosure. *Human Communication Research* 3, 3 (1977), 250–257.

[50] Alex C. Williams, Harmanpreet Kaur, Gloria Mark, Anne Loomis Thompson, Shamsi T. Iqbal, and Jaime Teevan. 2018. Supporting Workplace Detachment and Reattachment with Conversational Intelligence. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18)*. ACM, New York, NY, USA, Article 88, 13 pages. DOI: http://dx.doi.org/10.1145/3173574.3173662