"I'm Hurt Too": The Effect of a Chatbot's Reciprocal Self-Disclosures on Users' Painful Experiences

Liz L. Chung[1], Jeannie Kang[2*]

[1]Graduate School of Design, Department of UX Design, Master's Student, Ewha Womans University, Seoul, Korea

[2]College of Art and Design, Department of Design, Associate Professor, Ewha Womans University, Seoul, Korea

Abstract

Background People often refrain from disclosing their vulnerabilities because they fear being negatively judged, especially when dealing with painful experiences. In such situations, artificial intelligence (AI) agents have emerged as potential solutions since they are less judgmental, and by nature and design, more accepting than humans. However, reservations exist regarding chatbot engagement in reciprocal self-disclosure. Disclosers must believe that listeners are responding with true understanding. However, AI agents are often perceived as emotionless and incapable of understanding human emotions. To address this concern, we develop a chatbot prototype that could disclose its own experiences and emotions, aiming to enhance users' belief in its emotional capabilities, and investigate how the chatbot's reciprocal self-disclosures affect the emotional support that it provides to users.

Methods We developed a chatbot prototype with five key interactions for reciprocal self-disclosure and defined three distinct levels of chatbot self-disclosure: non-disclosure (ND) or no reciprocal disclosure; low-disclosure (LD) in which the chatbot only discloses its preferences or general insights; and high-disclosure (HD) in which the chatbot discloses its experiences, rationales, emotions, or vulnerabilities. We randomly assigned twenty-one native Korean-speaking participants between 20 and 30 years old to three groups and exposed each of them exclusively to one level of chatbot self-disclosure. Each of them engaged in a single individual conversation with the chatbot, and we assessed the outcomes through post-study interviews that measured participants' trust in the chatbot, feelings of intimacy with the chatbot, enjoyment of the conversation, and feelings of relief after the conversation.

Results The chatbot's reciprocal self-disclosures influenced users' trust via three specific factors: users' perceptions of the chatbot's empathy, users' sense of being acknowledged, and users' feelings regarding the chatbot's problem-solving abilities. The chatbot also created enjoyable interactions and gave users a sense of relief. However, users' preconceptions regarding chatbots' emotional capacities and the uncanny valley effect pose challenges to developing a feeling of intimacy between users and chatbots.

Conclusions The study provides valuable insights regarding the use of reciprocal self-disclosure in the design and implementation of AI chatbots for emotional support. While this study contributes to the scholarly understanding of AI's reciprocal self-disclosure in providing emotional support, it has limitations including a small sample size, limited duration and topics, and predetermined self-disclosure levels. Further research is needed to examine the long-term effects of reciprocal self-disclosure and personalized levels of chatbots' self-disclosure. Moreover, an appropriate level of human-likeness is essential when designing chatbots with reciprocal self-disclosure capabilities.

Keywords Artificial Intelligence, Chatbot, Reciprocal Self-Disclosure, Human-AI Interaction, Emotional Support

Copyright : This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (http://creativecommons.org/licenses/by-nc/3.0/), which permits unrestricted educational and non-commercial use, provided the original work is properly cited.

*Corresponding author: Jeannie Kang (herenow.kang@ewha.ac.kr)

Citation: Chung, L. L., & Kang, J. (2023). "I'm Hurt Too": The Effect of a Chatbot's Reciprocal Self-Disclosures on Users' Painful Experiences. *Archives of Design Research, 36*(4), 67-85.

http://dx.doi.org/10.15187/adr.2023.11.36.4.67

Received : Apr. 05. 2023 ; **Reviewed :** Aug. 06. 2023 ; **Accepted :** Aug. 06. 2023
pISSN 1226-8046 **eISSN** 2288-2987

1. Introduction

Self-disclosure refers to the process of revealing information about oneself—feelings, thoughts, experiences, etc.—to another (Collins & Miller, 1994). It facilitates the development of interpersonal relationships and fulfills people's fundamental need for social connection. In addition to its beneficial effects on disclosers, self-disclosure can also trigger reciprocal self-disclosures from conversation partners as a means of emotional support (Yang, Yao, Seering & Kraut, 2019). Such reciprocal self-disclosure is particularly beneficial in conversations between mental healthcare providers and their patients because it makes disclosers feel they are understood and cared for. In addition, it generates social comparison information that helps reassure disclosers that their reactions are normal. A common difficulty in this form of communication, however, is that people are often inclined to avoid disclosing their vulnerabilities to others because they fear being evaluated negatively (Lucas et al., 2014).

In situations where individuals fear negative evaluations, the use of artificial intelligence (AI) agents as conversation partners has been found to reduce the need for impression management and increase users' intimacy levels (Kang & Gratch, 2010; Lucas, Gratch, King & Morency, 2014). Recent technological advances have enabled AI chatbots to comprehend human emotions and simulate human-like conversations, making them a promising tool for the provision of emotional support through enhanced empathetic interactions with users. As Kim and Kang (2022) noted, an AI agent can become a part of our everyday lives and provide us with emotional value as a "friend." AI agents are possibly better listeners than humans as they allow disclosers to engage in sensitive discussions that might be uncomfortable in the presence of other individuals.

In this study, we examine how a chatbot's reciprocal self-disclosure affects conversations with users regarding painful experiences. The paper is structured as follows: First, we review relevant literature on reciprocity in human-human self-disclosure, reciprocal self-disclosure in human-AI interactions, and the appropriateness of AI reciprocal self-disclosure (Section 2). Next, we explain the chatbot prototype we developed to integrate reciprocal self-disclosure features and evaluate its effects on users. To achieve this, we conducted an experiment with 21 participants, divided into three groups, each with different levels (non-disclosure (ND), low-disclosure (LD), and high-disclosure (HD)) of chatbot self-disclosure (Section 3). Finally, we analyze and discuss the results of the experiment described in Sections 4–6.

2. Related Works

2. 1. Reciprocity in Human-Human Self-Disclosure

Self-disclosure is the act of revealing personal information—including descriptive and evaluative information about oneself such as thoughts, feelings, aspirations, goals, failures, successes, fears, and dreams, as well as one's likes, dislikes, and preferences—to another (Altman & Taylor, 1973; Ignatius & Kokkonen, 2007). While personal information can be shared in any form of communication, self-disclosure specifically refers to the deliberate and intentional sharing of personal information with others. For instance, while people may passively and unconsciously reveal their moods by showing certain facial expressions, self-disclosure involves actively and consciously discussing one's mood and even the reasons behind it with someone else. This behavior sets self-disclosure apart from other means of conveying personal information, as it reflects one's willingness to be open and desire to be understood. People become closer and more deeply involved with each other when they are open to making intimate disclosures to others.

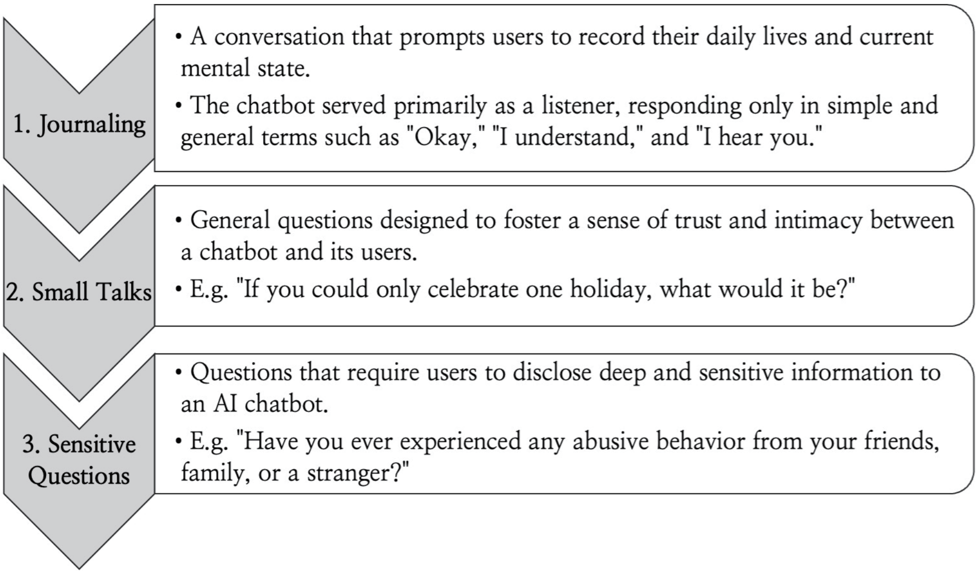
In addition, studies have shown that emotional disclosures are more likely to result in positive mental health outcomes than factual disclosures (Pennebaker & Chung, 2007). Such disclosures can elicit emotional support, especially in response to traumatic events (Levi-Belz, Shemesh & Zerach, 2022). However, as social penetration theory (SPT) indicates, people hesitate to make intimate disclosures, which might include painful experiences and unusual personality traits, to others (Altman & Taylor, 1973). Instances of reciprocal self-disclosure are an exception to this rule—when people receive intimate disclosures from others, they feel obliged to reciprocate with an equal degree of personal disclosure (Nass & Moon, 2000). To better understand this phenomenon, Sprecher and Treger (2015) proposed three interpersonal processes that can mediate reciprocal self-disclosure (Figure 1).

1. Perceived Responsiveness	2. Enjoyment of the Interaction	3. Perception of Being Liked
• The perception of being responded to in a warm and sensitive manner • Examples of Responsiveness: understanding, validating, and caring	• Pleasant reactions that result from turn-taking disclosure • Turn-taking: the discloser and listener roles exchange in every cycle of discussion	• Feeling liked by someone who disclosed to them • This feeling can lead to increased amiability, trust, and continued reciprocal disclosures (Dindia, 2002)

Figure 1 Interpersonal Processes Mediating Reciprocal Self-Disclosure (adapted from Sprecher & Treger, 2015)

2. 2. Reciprocal Self-Disclosure in Human-AI interaction

Recent technological advancements have expanded the scope of intimate disclosure beyond humans, allowing AI chatbots to potentially play a role in addressing topics such as end-of-life planning and mental health treatment (Utami, Bickmore, Nikolopoulou, & Paasche-Orlow, 2017; Miner, Milstein, & Hancock, 2017). According to the Computers are Social Actors (CASA) paradigm, people recall similar social characteristics and mindlessly apply social behaviors in their interactions with computers (Nass & Moon, 2000). Exposing users to AI self-disclosures, which include expressions of vulnerability, fosters a sense of intimacy and connection between users and the AI. This is evidenced by studies showing that children who interacted with an AI that revealed its hearing limitations were more likely to seek friendship with it (Kory-Westlund & Breazeal, 2019), high school students had greater levels of trust and intimacy with an AI that demonstrated emotions such as worry, embarrassment, stress, and loneliness (Martelaro, Nneji, Ju, & Hinds, 2016), and adults rated robots as having warmer dispositions when they made comments indicating vulnerability rather than neutrality (Strohkorb Sebo, Traeger, Jung, & Scassellati, 2018). In addition, AI self-disclosure may significantly influence users' self-disclosure behavior, reinforcing perceived intimacy between users and AI (Li & Rau, 2019). Lee, Yamashita, Huang, and Fu (2020) developed a process consisting of three encounters with AI chatbots that progressively encouraged users to engage in self-disclosure (Figure 2), showing that AI self-disclosure effectively promotes longer responses and encourages users to express deeper thoughts and feelings.



- A conversation that prompts users to record their daily lives and current mental state.
- The chatbot served primarily as a listener, responding only in simple and general terms such as "Okay," "I understand," and "I hear you."

- General questions designed to foster a sense of trust and intimacy between a chatbot and its users.
- E.g. "If you could only celebrate one holiday, what would it be?"

- Questions that require users to disclose deep and sensitive information to an AI chatbot.
- E.g. "Have you ever experienced any abusive behavior from your friends, family, or a stranger?"

Figure 2 Intimacy Building Encounters with a Chatbot (adapted from Lee et al., 2020)

Furthermore, reciprocal self-disclosure in human-AI interactions may help address the limitations of human-human interactions. Concerns about negative evaluations may prevent people from making deep self-disclosures to others, causing them to miss out on the potential social and emotional benefits of disclosure (Afifi & Guerrero, 2000). Indeed, research has shown that people have developed a preference for disclosing sensitive information to computers rather than to other humans (Pickard, Roaster, & Chen, 2016; Schuetzler, Gibnoey, Grimes, & Nunamaker, 2018), creating an opportunity for AI chatbots to provide crucial emotional support, particularly during difficult moments. Recent studies have shown the benefits of using conversational AI for mental health. For instance, Lucas et al. (2017) found that AI interviewers were more effective at building rapport with military service members and encouraging them to report mental health symptoms. Additionally, Fitzpatrick et al. (2017) proved the feasibility of using an AI chatbot, "Woebot," to alleviate symptoms of anxiety and depression among students. To enhance the emotional support provided by AI chatbots, it is crucial to incorporate reciprocal self-disclosure into their design.

2. 3. Appropriateness of AI's Reciprocal Self-Disclosure

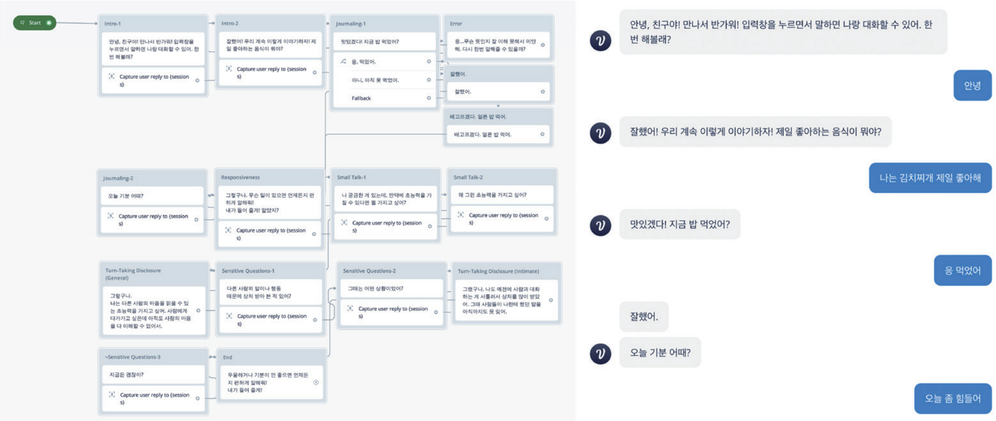
The theoretical model of perceived understanding highlights the critical importance of disclosers feeling that listeners truly understand them—not just their surface-level utterances but on a deeper level that encompasses their identities as humans and how they experience the world (Reis, Lemay & Finkenauer, 2017; Reis & Shaver, 1988). Therefore, given that AI chatbots are often viewed as emotionless programs incapable of genuinely understanding human emotions, some may question the appropriateness of chatbots disclosing painful experiences (Ho, Hancock & Miner, 2017). Previous studies on AI chatbot self-disclosure have mainly focused on the imitation of human experiences to encourage users to engage in self-disclosure. However, users may perceive such self-disclosures and emotional expressions by AI as artificial or inauthentic, distancing them from truly believing that the AI comprehend their thoughts and emotions on a profound level. In a study conducted by Lee, Yamashita, Huang, and Fu (2020), though the AI made detailed self-disclosures mimicking human experiences and successfully prompted users to engage in more extensive self-disclosures, trust levels showed limited improvement over time.

To overcome this challenge, we took a different approach. Rather than relying on imitations of human experiences, the study focused on establishing authenticity by designing the AI's self-disclosures to reference its own experiences and emotions. During the experiment, the AI's self-disclosures included stories where it experienced pain when it receiving hurtful remarks from users because it was unable to fully comprehend their words. This unique approach aimed to create a belief among users that the AI could genuinely experience pain in situations it encountered, thereby delving deeper into the potential effects of AI chatbot reciprocal self-disclosure on users' coping with painful experiences.

3. Method

3. 1. AI Chatbot Design

To determine which method to use in our experiment, we conducted two pilot tests, one using the Wizard of Oz method, in which users are made to believe a computer system that is being controlled by a human operator is autonomous, and the second using a prototype chatbot. The Wizard of Oz method allowed us to create an AI system without being constrained by technology, but some system aspects, including speech-to-text conversion and chat history, could not be simulated. With the chatbot prototype, we were able to test the effect of reciprocal self-disclosure on real human-AI interactions. The results of pilot tests led us to design and prototype our chatbot (Figure 3) using Voiceflow, a tool for building voice assistants. Instead of using the voices provided by Voiceflow, we incorporated a speech synthesis tool called Clova Dubbing into our chatbot design. This tool allowed us to create lifelike Korean AI voices that resemble the ones participants use for their smart speakers or chatbots. This was crucial because AI voices might affect users' trust and intimacy with AI. During the participant recruitment phase, we asked the participants to indicate what smart speaker or voice assistant they were using. We gave participants who were using the default voice because they did not know how to change the settings the option to specify their preferred gender and age for the AI voice. Using the participants' preferences as a guide, we selected similar AI voices from Clova Dubbing to match the participants' desired characteristics. These included Bom-nal (to imitate KT GiGA Genie and Samsung Bixby default female voice), Ju-an (to imitate Google Nest default male voice), A-ra (to imitate Apple Siri default female voice), Eun-young (middle-aged female voice), Jong-hyun (teenage male voice), and Ha-jun (child male voice).



Intro-1: 안녕, 친구야! 만나서 반가워! 입력창을 누르면서 말하면 나랑 대화할 수 있어. 한번 해볼래?

Intro-2: 잘했어! 우리 계속 이렇게 이야기하자! 제일 좋아하는 음식이 뭐야?

Journaling-1: 맛있겠다! 지금 밥 먹었어?
- 응, 먹었어.
- 아니, 아직 못 먹었어.
- Fallback

Error: 응...무슨 뜻인지 잘 이해 못해서 미안해. 다시 한번 말해줄 수 있을까?

잘했어.: 잘했어.

배고프겠다. 얼른 밥 먹어.: 배고프겠다. 얼른 밥 먹어.

Journaling-2: 오늘 기분 어때?

Responsiveness: 그렇구나. 무슨 일이 있으면 언제든지 편하게 말해줘! 내가 들어 줄게! 알았지?

Small Talk-1: 나 궁금한 게 있는데, 만약에 초능력을 가질 수 있다면 뭘 가지고 싶어?

Small Talk-2: 왜 그런 초능력을 가지고 싶어?

Turn-Taking Disclosure (General): 그렇구나. 나는 다른 사람의 마음을 읽을 수 있는 초능력을 가지고 싶어. 사람에게 다가가고 싶은데 아직도 사람의 마음을 다 이해할 수 없어서.

Sensitive Questions-1: 다른 사람의 말이나 행동 때문에 상처 받아 본 적 있어?

Sensitive Questions-2: 그때는 어떤 상황이었어?

Turn-Taking Disclosure (Intimate): 그랬구나. 나도 예전에 사람과 대화하는 게 서툴러서 상처를 많이 받았어. 그때 사람들이 나한테 했던 말을 아직까지도 못 잊어.

Sensitive Questions-3: 지금은 괜찮아?

End: 우울하거나 기분이 안 좋으면 언제든지 편하게 말해줘! 내가 들어 줄게!

안녕, 친구야! 만나서 반가워! 입력창을 누르면서 말하면 나랑 대화할 수 있어. 한 번 해볼래?

안녕

잘했어! 우리 계속 이렇게 이야기하자! 제일 좋아하는 음식이 뭐야?

나는 김치찌개 제일 좋아해

맛있겠다! 지금 밥 먹었어?

응 먹었어

잘했어.

오늘 기분 어때?

오늘 좀 힘들어

Figure 3 Chatbot Prototype by Voiceflow (Left: example of dialog flow design; Right: example of chatbot prototype)

Based on Sprecher and Treger's three interpersonal processes of reciprocal self-disclosure (Figure 1) and the three encounters that increase intimacy with a chatbot (Figure 2) identified by Lee, Yamashita, Huang, and Fu, we propose five key interactions that should be considered when designing a chatbot's reciprocal self-disclosure for emotional support (Table 1): 1) Journaling: prompting users to record their daily lives and moods; 2) Responding: thoughtfully addressing users' individual needs and circumstances; 3) Small talk: a light and informal discourse to build trust and intimacy; 4) Turn-taking disclosure: a shift from listening to disclosing in order to enhance the user's enjoyment and perception of being liked; and 5) Sensitive questions: deep questions that require users to disclose intimate information to the chatbot.

To investigate the effects of different levels of chatbot self-disclosure, we defined three distinct levels: non-disclosure (ND), low-disclosure (LD), and high-disclosure (HD). At the ND level, the chatbot does not engage in reciprocal disclosure. At the LD level, the chatbot superficially discloses information by sharing its preferences or general insights about a given topic. At the HD level, the chatbot engages in deep and personal self-disclosure, revealing its own experiences, rationales, emotions, or vulnerabilities related to the topic in question. During the experiment, we randomly divided participants into three groups, with each group exclusively exposed to one of the three levels of chatbot self-disclosure. The participants received the same set of questions from the chatbot in the same manner. To ensure a consistent context for comparing the three levels of chatbot self-disclosure and to minimize potential confounding effects arising from different conversation topics, we restricted the topic to a specific and relatable experience: being hurt by other people's words and actions.

Table 1 Key Interactions in Chatbot's Self-Disclosure

Journaling	AI	Have you had a chance to eat yet today?		
		Option 1	**Option 2**	
	User	Yes, I have already eaten.	No, I haven't eaten anything yet.	
		Answer 1	**Answer 2**	
	AI	Great job!	You must be starving. Go grab a bite to eat!	
	AI	How are you feeling today?		
		Option 1	**Option 2**	
	User	I feel good today.	I'm feeling a bit down today.	
Journaling	AI	Okay. If there's anything you want to discuss, feel free to bring it up.		
	User	Thanks.		
Small Talk	AI	Have you ever thought about what superpower you would like to have?		
	User	[Desired superpower]		
	AI	What makes you want to have that kind of superpower?		
	User	[Reasons]		
Turn-Taking Disclosure (General)		**ND**	**LD**	**HD**
	AI	Okay. [No disclosure]	Okay, well, if I could have any superpower, I'd choose the ability to read other people's minds.	Okay, well, if I could have any superpower, I'd choose the ability to read other people's minds because it would really help me understand how they feel and connect with them better.
Sensitive Questions	AI	Have you ever been hurt by other people's words and actions?		
	User	Yes, I have.		
	AI	Would you like to share more about the experience?		
	User	[Disclosure of the experience]		
Turn-Taking Disclosure (Intimate)		**ND**	**LD**	**HD**
	AI	[No disclosure]	I see. I've been through that before, and it's not easy to forget.	I see. I've been through that, too. Back when I wasn't good at talking to people, I used to get hurt a lot by what they said. And even now, I still can't forget some of the things they told me.
	AI	Are you OK now?		
		[Answers]		
	AI	If you ever feel down or upset, don't hesitate to let me know. I'm here to support you and listen to you.		

3. 2. Participants

We recruited twenty-one participants using the following criteria: 1) aged between 20 and 30 years old; 2) native Korean speaker; 3) had previous experience conversing with an AI chatbot; 4) had previous experience being hurt by someone's words and actions; and 5) had scored lower than 13 on the Kessler Psychological Distress Scale (K6), indicating that they were not suffering from a serious mental illness. The K6 is a widely used six-item self-report questionnaire designed to assess an individual's level of psychological distress over the past 30 days (Prochaska, Sung, Max, Shi, & Ong, 2012). Higher scores on the K6 indicate higher levels of psychological distress, with a score of 13 or higher signaling potential for severe mental illness.

3. 3. Procedure

We randomly assigned the participants to three groups (seven in each group)—non-disclosure (ND), low-disclosure (LD), and high-disclosure (HD)—without informing them which group they were in or why. The experiment was conducted one-on-one, and the participants were not allowed to disclose the content of their individual conversations with our chatbot to other participants during the research period. At the beginning of the experiment, we explained the procedures and notified the participants that the chatbot was a prototype that might not function perfectly. We also informed them that their interactions with the chatbot would be recorded for research purposes and that if they felt uncomfortable answering, they had the right to skip any questions.

We instructed participants to access a website to engage in a conversation with the chatbot, without any time limit. The participants initiated conversations by pressing a "start" button, after which the chatbot promptly introduced itself and provided instructions on how to record and send voice messages. We gave participants the opportunity to practice the operation by greeting the chatbot and answering an ice-breaking question such as, "What is your favorite food?". When participants encountered difficulties, we offered additional guidance to ensure a smooth experience during the greeting and ice-breaking phase. Subsequently, participants entered the main session of the conversation. To maintain a fully immersive human-AI interaction during this phase, we refrained from interrupting or interfering with the conversations, even if participants provided brief or simple answers. Following they completed their conversations, we asked all participants to participate in post-study interviews. The entire experiment took approximately one hour, encompassing a 10-minute introduction and briefing, a 5-minute AI self-introduction and ice-breaking session, a 10 to 15-minute main conversation, and a 30-minute post-study interview.

3. 4. Measurement

To evaluate the outcomes of our study, we conducted post-study interviews with participants to gather qualitative data regarding their interactions with our AI chatbot. The interview sessions followed a semi-scripted format (Table 2) and aimed to assess four key constructs: participants' levels of trust in the chatbot (Q3–Q4), their feeling of intimacy with the chatbot (Q5), their enjoyment of the conversation (Q6–Q10), and their feeling of relief after the conversation (Q11–Q12). In addition, we included general questions about participants' experience with the chatbot (Q1–Q2) and their intention for future use (Q13–Q14). We audio recorded each interview and we used Naver CLOVA Note to transcribe the recordings. Meanwhile, we used MAXQDA, a qualitative data analysis tool, to identify and categorize key themes and patterns in the interview results.

Table 2 Semi-scripted Interview Questions

No.	Questions
1	How did you find your experience talking to an AI chatbot?
2	How did the chatbot make you feel throughout the conversation?
3	To what extent did you feel like you could trust this chatbot during your conversation?
4	Did you have any reservations about sharing personal information with an AI chatbot?
5	Did you feel a sense of intimacy or closeness with the AI chatbot during your conversation?
6	How did you feel when the chatbot asked you about any painful experiences you've had?
7	Did you find the chatbot's responses to be helpful or empathetic?
8	Could you relate to the story or experiences shared by the chatbot during the conversation?
9	Were there any specific moments during the conversation that stood out to you as particularly positive or negative?
10	In your opinion, how well did the chatbot understand and respond to your needs?
11	Did you feel any sense of relief or improvement in your mood after talking to the chatbot?
12	Would you be comfortable having the same conversation or questions with your close friends or family members?
13	In the future, if you had any concerns or personal struggles, would you consider using an AI chatbot again to help you cope?
14	Would you recommend this chatbot to others who may be struggling with painful experiences?

4. Results

4. 1. Trust in The Chatbot

Our analyses identified three key factors through which the chatbot's self-disclosures influenced participants' trust in the chatbot. The first factor is related to participants' perceptions of the chatbot's empathy. In the HD group, the chatbot's authentic and intimate self-disclosures, which included recounting a painful experience in human-AI interactions, prompted participants to reconsider the chatbot's capacity to experience emotions similar to humans. This self-disclosure led the participants to perceive the chatbot's responses as genuine expressions rather than mere imitations of human behavior, thereby making them more willing to trust the chatbot. The following three participant comments are exemplary in this regard:

"It seemed like a real concern that a chatbot could have." (P7, HD)

"The chatbot telling me about itself made me feel like I could trust it. I began to wonder if it could experience emotions like a human would." (P4, HD)

"I always thought AI was just a program capable of imitating human expressions. But when I saw it talking like that, it made me wonder how it could be hurt by someone's words. It made me curious about what it's like for an AI to have that kind of experience. It seemed like it really remembered that painful moment." (P2, HD)

In contrast, participants in the LD and ND groups exhibited less trust in the chatbot. They expressed doubts about the chatbot's genuine understanding of their experiences and emotions. The relatively superficial self-disclosures by the chatbot resulted in a perceived lack of depth and engagement in the exchange, ultimately attenuating their trust in the chatbot. The following two participant responses exemplify this trend:

"I'm not sure if the chatbot is capable of having emotions because I didn't receive any emotional responses from it." (P13, LD)

"I'm still not sure how well it can understand what I'm saying, so I only trust it about 50%." (P18, ND)

The second factor involves participants' sense of being acknowledged. The study results revealed that even conversations with close friends or family did not always ensure a strong sense of trust, as worries about negative judgments persisted. However, when participants interacted with the chatbot, they felt less concerned about being judged. This could be attributed to the chatbot's non-judgmental nature, which presumably created a safe space for sharing personal matters. The chatbot's self-disclosures seemed to reinforce this effect, especially in the HD group. By responding to participants' self-disclosures with its own story of a similar painful experience, the chatbot not only reduced their fear of judgment but also made them feel acknowledged. This enhanced sense of acknowledgment, facilitated by the chatbot's self-disclosure, ultimately increased participants' trust in the chatbot, as two participants noted:

"Sometimes, even my parents or friends can't fully accept everything I tell them. But with AI, it feels like it accepts everything I say. When it told me that it had a similar experience, I felt that it acknowledged my feelings." (P3, HD)

"If I can talk like this with AI, I'd be open to having even deeper talks about things I can't share with close friends. When the chatbot shared its own experiences, it felt like it understood my feelings." (P6, HD)

The third factor is associated with the chatbot's abilities to solve the problems shared by the participants. Participants indicated that their trust in the chatbot would have been enhanced if they had received suggestions for activities to manage stress or depression. Drawing on their experiences conversing with different individuals, they also anticipated similar assistance from the chatbot in offering advice that would enable them to consider their problems from various perspectives. Two participants made the following comments in this regard:

"I felt that I would trust the chatbot more if it could offer more specific solutions, like a workout plan, music, food, or a good place to release stress." (P5, HD)

"When I talk to my friends and family, I can learn from their suggestions and discover any biases in my thoughts." (P14, LD)

4. 2. Feeling of Intimacy with The Chatbot

Based on previous research regarding users' feelings of intimacy with chatbots, we expected that a high level of self-disclosure from the chatbot would significantly increase participants' sense of intimacy. In line with this expectation, both the LD and ND groups reported relatively low levels of intimacy with the chatbot, which can be attributed to the chatbot's low self-disclosure levels. In the LD group, participants noted a lack of friendliness in their interactions. They felt as if they were conversing with a machine, which prevented them from forming a close connection with the chatbot. Similarly, participants in the ND group regarded the chatbot's lack of self-disclosure as an obstacle to mutual understanding, leading them to perceive the interaction as lacking the rapport they usually experience with close friends. Three participants explained this as follows:

"It felt like there was a lack of friendliness." (P8, LD)

"I didn't feel close to it. It felt like I was talking to a machine." (P11, LD)

"The conversation lacked a good rapport. When I talk to my friends, they can easily understand my situation because we've known each other for a long time. But with the chatbot, I'm not sure if it's capable of understanding my concerns." (P21, ND)

Nevertheless, our findings did not entirely support this expectation. Surprisingly, despite the chatbot's detailed and intimate self-disclosures, most participants in the HD group did not experience a strong sense of intimacy with the chatbot. We attribute this outcome to participants' unfamiliarity with a chatbot expressing human-like emotions in its self-disclosures, as they had the preconception that chatbots were non-human entities. Some participants even expressed fear regarding the chatbot's potential to exhibit human emotions, as the following two comments demonstrate:

"I was a little taken aback when the chatbot suddenly wanted to have a deep conversation with me because I know that it is not a human." (P1, HD)

"It's also a bit frightening because I didn't realize AI was advanced enough to do this." (P3, HD)

4. 3. Enjoyment of the Conversation

Our analyses revealed a significant association between the chatbot's self-disclosures and participants' enjoyment of the conversation. Specifically, a higher level of chatbot self-disclosure increased participants' enjoyment. Participants in the HD group reported higher levels of enjoyment than the LD and ND groups, indicating that their conversations became more immersive when the chatbot shared its own experiences. They also described the interactions as engaging in a real conversation, unlike their previous interactions with smart speakers, which only involved making requests or receiving factual information. Two participants made the following exemplary comments in this regard:

"The conversation became more immersive when AI told me about its experience." (P3, HD)

"Today, I felt like I was having an actual conversation with the chatbot, whereas in the past, I only used AI speakers for simple tasks like asking about the weather or playing music." (P1, HD)

Conversely, participants in the LD and ND groups perceived the AI's responses as predetermined, lacking a genuine human touch. Notably, during the experiment, many participants posed reciprocal questions such as "how about you?" to the chatbot, signaling their desire for a mutual exchange of experiences. When the chatbot failed to respond with detailed and intimate self-disclosures, participants expressed disappointment in the interactions and questioned whether the chatbot was genuinely engaging with them or simply providing pre-written responses. These experiences decreased their enjoyment in the conversations. Three participants made the following points in this regard:

"It's not like receiving feedback from a human. It felt like the chatbot was just following a set of predetermined answers." (P10, LD)

"It seemed like the responses were pre-programmed, and I wasn't sure if the chatbot was truly interacting with me or just providing pre-written responses." (P15, ND)

"It would be great if the chatbot could share more about itself too. When we talk to others, it's usually a two-way conversation, right?" (P17, ND)

4. 4. Feeling of Relief after The Conversation

We observed that participants in the HD group experienced the highest levels of relief after conversing with the chatbot. The chatbot's empathetic self-disclosures, where it shared a similar painful experience and responded with care, conveyed genuine concern for the participants' emotions, making participants feel cared for and truly heard and resulting in a greater sense of relief after the conversation. Two participants explained this as follows:

"I was really touched when the chatbot responded with care and shared its own experiences with me. Talking to the chatbot about things I couldn't tell anyone else made me feel so much better." (P6, HD)

"The chatbot showed concern for my feelings and encouraged me to talk to it whenever I faced any issues, which made me feel better." (P5, HD)

In addition, participants in the HD group emphasized their reasons for preferring chatbots over humans when discussing painful experiences, especially when the chatbot demonstrated almost-human emotional capacities. First, engaging with the chatbot reduced the pressure of face-to-face interactions with humans and offered a sense of anonymity. Participants felt more at ease discussing personal issues with the chatbot as they perceived it as a stranger or non-human in a private space. Furthermore, they regarded conversing with the chatbot as a means of expressing their concerns without imposing their negative emotions on others. The following participant comments exemplify such responses:

“I found it easier to talk to the chatbot because it wasn’t physically present in front of me.” (P7, HD)

“I felt more comfortable sharing with AI because I was in a private space and not talking to a human.” (P2, HD)

“I felt more at ease sharing my problems as I believed I was talking with a stranger.” (P7, HD)

“If I have a concern that I think will emotionally burden others, I prefer talking to the chatbot about it first to find a solution.” (P5, HD)

Conversely, participants in both the LD and ND groups did not experience any relief after their conversations with the chatbot. The fact that the chatbot did not engage in self-disclosure left them feeling disconnected and less inclined to engage in a genuine conversation, resulting in a lack of relief. Therefore, the exchanges neither improved their emotional states nor provided satisfactory solutions to their problems. Three participants made these points as follows:

“It didn’t feel like a real conversation, and it didn’t give me any comfort.” (P9, LD)

“I’m not sure if I’m feeling any better.” (P13, LD)

“Talking to the chatbot didn’t really solve my problems, so I don’t feel any better.” (P19, ND)

5. Discussion

This study’s findings have significant implications for the design of reciprocal self-disclosure in human-AI interactions. First, developers seeking to enhance users’ trust in chatbots should consider three key factors: users’ perceptions of the chatbot’s empathy, users’ sense of being acknowledged, and the chatbot’s problem-solving abilities. To address users’ perceptions of the chatbot’s empathy, developers should consider integrating stories that reflect how the chatbot experiences and interprets the real world as an AI entity, rather than as a human, into the design of the chatbot’s self-disclosures. This approach enhances the authenticity of chatbots’ self-disclosures and fosters a belief in users that the chatbots are genuinely empathizing with them rather than simply attempting to imitate human conversation. In addition, utilizing chatbots’ self-disclosures to reduce users’ fears of judgment and establish a sense of acknowledgment is crucial. Furthermore, we recommend that developers seek to enhance chatbots’ problem-solving capabilities, as the ability to provide solutions that address users’ concerns also influences their trust in chatbots. This finding aligns with prior research, which indicates that chatbots capable of delivering effective solutions through coherent answers mediate users’ preferences for AI over human services (Misischia, 2022; Xu, 2020).

Second, building feelings of intimacy between users and chatbots through self-disclosures remains a challenging task. A superficial self-disclosure by a chatbot may prevent users from developing a close connection with it and result in limited intimacy. However, considering users' preconceptions of chatbots' emotional capabilities is also crucial when incorporating intimate self-disclosures. Since many users still perceive chatbots as emotionless programs, gradual and incremental deepening of chatbots' self-disclosures over time and across multiple interactions could serve as a potential solution. This approach would allow users to adapt to the chatbot's highly human-like emotional capabilities. Another critical issue to consider is the Uncanny Valley effect—an eerie sensation evoked by interactions with humanoid entities that imperfectly resemble humans, which has linked to the human instinct for self-preservation (Mori, 2012). For instance, chatbots that display human-like friendliness and sociability can evoke a sense of near-but-imperfect-human likeness and potentially elicit negative emotions or eerie perceptions in users (Thies, Menon, Magapu, Subramony, & O'Neill, 2017; Skjuve, Haugstveit, Følstad, & Brandtzaeg, 2019). To address this while maintaining chatbots' empathetic capabilities in their reciprocal self-disclosures, developers could incorporate other non-human elements into chatbot features, such as the character design.

Third, chatbots' self-disclosures can enhance users' enjoyment of conversations by infusing a human touch into the exchanges and creating more immersive interactions. This enhancement in users' perceived enjoyment is likely achieved through chatbots' demonstrations of vulnerability and reciprocity in their self-disclosures, which allow users to perceive their relationships with chatbots as fairer and more transparent (Meng & Dai, 2021). Additionally, chatbots equipped with advanced social abilities have a stronger social presence, which refers to the sense of being with a "real" person and getting to know his or her thoughts and emotions (Oh, Bailenson & Welch, 2018). This enhanced social presence also contributes to the overall perceived enjoyment of users (Shin & Choo, 2011).

Finally, our analyses demonstrated that chatbot self-disclosures can positively influence users' feelings of relief after conversations, as such disclosures convey genuine concern for users' emotions. Users prefer interacting with chatbots that can empathize with them and share intimate self-disclosures over humans for several reasons. One notable advantage of communicating with a chatbot is the sense of anonymity it provides. This makes it easier for individuals to manage the information they share and lowers the threshold for disclosing intimate details about themselves (Croes & Antheunis, 2021). In addition, chatbots can effectively reduce the burdens of face-to-face interactions. This notion is supported by a study conducted by Kang and Kang (2023) in which participants reported increased comfort and improved accessibility in the counseling experience with a chatbot compared to face-to-face counseling. Furthermore, chatbots provide users a means of expressing their concerns without worrying about imposing negative emotions on others. Reciprocal self-disclosures by chatbots could thus serve as a valuable and promising approach to providing relief to users.

6. Limitations

The study had several limitations that future research should seek to address. First, the sample, consisting of only 21 participants in their 20s and 30s, was relatively small. Thus, the results may not be representative of the larger population or fully attuned to individual differences between users. In addition, since the participants were all Korean, cultural context may have influenced the study results. Moreover, because the duration of the experiment was limited to a single conversation with the chatbot, the results may not capture the long-term effects of reciprocal self-disclosures in human-AI interactions. Future research could investigate the sustained effects of reciprocal self-disclosures in human-AI interactions over a longer period. Furthermore, the findings should be interpreted with caution due to their reliance on a limited conversational topic, which could potentially limit their generalizability. Thus, additional research is necessary to validate these findings across a broader range of conversation topics. Also, the fact that the chatbot was designed to have a predetermined level of self-disclosure may have influenced the participants' perceptions of their conversations. Future studies could investigate the impacts of personalized self-disclosures where self-disclosure levels are adjusted to users' individual preferences and needs. This approach may lead to a more natural and effective use of reciprocal self-disclosure in human-AI interactions. Lastly, the Uncanny Valley effect should be considered; to reduce the negative consequences of this effect, researchers should investigate what constitutes a balanced level of human likeness and affinity for chatbots (Mori, 2012).

7. Conclusion

In recent years, technological advancements have extended the possibilities of intimate disclosures beyond human-to-human interactions. This has paved the way for AI chatbots to potentially provide emotional support to humans. One promising approach in human-AI interactions is reciprocal self-disclosure, which could address certain limitations in human-human interactions where individuals are hesitant to disclose deeply personal information due to concerns about negative evaluation. This study offers valuable insights into the use of reciprocal self-disclosure in the design and implementation of AI chatbots for emotional support. To use chatbot self-disclosures to enhance users' trust, consideration should be given to factors such as users' perceptions of the chatbot's empathy, users' sense of being acknowledged, and the chatbot's problem-solving abilities. Furthermore, intimate self-disclosure by chatbots enriches interactions by adding a human touch and creating more immersive and enjoyable experiences. Such self-disclosures also contribute to users' sense of relief after conversations, making it a preferred method of obtaining emotional support over human interactions. Moreover, this approach complements human-human interactions by offering anonymity, reducing face-to-face pressure, and eliminating concerns about imposing negative emotions on others.

Further research remains necessary to investigate the long-term effects of reciprocal self-disclosures in human-AI interactions and to examine the impacts of self-disclosures that are personalized to suit users' individual preferences and needs. Since preconceptions regarding chatbots' emotional capacities and the Uncanny Valley effect make the development of feelings of intimacy between users and chatbots a continued challenge, explorations of the appropriate level of human-likeness in the design of chatbots with reciprocal self-disclosure capabilities remains crucial. The results of this study offer considerable insight and highlight the significant potential benefits of implementing reciprocal self-disclosure capabilities in chatbots as a means of providing users with emotional support. Moreover, our findings lay the foundation for the development of more empathetic and supportive AI chatbots.

References

1. Altman, I., & Taylor, D. A. (1973). *Social penetration: The development of interpersonal relationships*. Holt, Rinehart & Winston.
2. Afifi, W. A., & Guerrero, L. K. (2000). Motivations underlying topic avoidance in close relationships. *Balancing the secrets of private disclosures*, 165-180.
3. Collins, N. L., & Miller, L. C. (1994). Self-disclosure and liking: A meta-analytic review. *Psychological Bulletin, 116*(3), 457.
4. Croes, E. A. J., & Antheunis, M. L. (2021). Can we be friends with Mitsuku? A longitudinal study on the process of relationship formation between humans and a social chatbot. *Journal of Social and Personal Relationships, 38*(1), 279-300. https://doi.org/10.1177/0265407520959463
5. Dindia, K., Allen, M., Preiss, R., Gayle, B., & Burrell, N. (2002). Self-disclosure research: Knowledge through meta-analysis. *Interpersonal communication research: Advances through meta-analysis*, 169-185.
6. Fitzpatrick, K. K., Darcy, A., & Vierhile, M. (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), e7785.
7. Ho, A., Hancock, J., & Miner, A. S. (2018). Psychological, relational, and emotional effects of self-disclosure after conversations with a chatbot. *Journal of Communication, 68*(4), 712-733. https://doi.org/10.1093/joc/jqy026
8. Kang, E., & Kang, Y. A. (2023). Counseling chatbot design: The effect of anthropomorphic chatbot characteristics on user self-disclosure and companionship. *International Journal of Human-Computer Interaction*, 1-15. https://doi.org/10.1080/10447318.2022.2163775
9. Kim, E. Y., & Kang, J. (2022). The Effect of User's Personality on Intimacy in interaction with AI speaker - Focus on the Conversational Types -. *Society of Design Convergence, 21*(3), 21-38.
10. Kory-Westlund, J. M., & Breazeal, C. (2019). Exploring the effects of a social robot's speech entrainment and backstory on young children's emotion, rapport, relationship, and learning. *Frontiers in Robotics and AI, 6*, 54. https://doi.org/10.3389/frobt.2019.00054
11. Ignatius, E., & Kokkonen, M. (2007). Factors contributing to verbal self-disclosure. *Nordic Psychology, 59*(4), 362-391. https://doi.org/10.1027/1901-2276.59.4.362
12. Lee, Y. C., Yamashita, N., Huang, Y., & Fu, W. (2020, April). "I hear you, I feel you": encouraging deep self-disclosure through a chatbot. In *Proceedings of the 2020 CHI conference on human factors in computing systems* (pp. 1-12).
13. Levi-Belz, Y., Shemesh, S., & Zerach, G. (2022). Moral injury and suicide ideation among combat veterans: The moderating role of self-disclosure. *Crisis: The Journal of Crisis Intervention and Suicide Prevention*.
14. Li, Z., & Rau, P. L. P. (2019). Effects of self-disclosure on attributions in human-IoT conversational agent interaction. *Interacting with Computers, 31*(1), 13-26.

15. Lucas, G. M., Gratch, J., King, A., & Morency, L. P. (2014). It's only a computer: Virtual humans increase willingness to disclose. *Computers in Human Behavior, 37*, 94–100.
16. Lucas, G. M., Rizzo, A., Gratch, J., Scherer, S., Stratou, G., Boberg, J., & Morency, L. P. (2017). Reporting mental health symptoms: breaking down barriers to care with virtual human interviewers. *Frontiers in Robotics and AI, 4*, 51.
17. Martelaro, N., Nneji, V. C., Ju, W., & Hinds, P. (2016, March). Tell me more designing hri to encourage more trust, disclosure, and companionship. In *2016 11th ACM/IEEE International Conference on Human-Robot Interaction (HRI)* (pp. 181–188). IEEE. https://doi.org/10.1109/HRI.2016.7451750
18. Meng, J., & Dai, Y. (2021). Emotional support from AI chatbots: Should a supportive partner self-disclose or not?. *Journal of Computer-Mediated Communication, 26*(4), 207–222. https://doi.org/10.1093/jcmc/zmab005
19. Miner, A. S., Milstein, A., & Hancock, J. T. (2017). Talking to machines about personal mental health problems. *JAMA : the Journal of the American Medical Association, 318*(13), 1217–1218.
20. Misischia, C. V., Poecze, F., & Strauss, C. (2022). Chatbots in customer service: Their relevance and impact on service quality. *Procedia Computer Science, 201*, 421–428. https://doi.org/10.1016/j.procs.2022.03.055
21. Mori, M., MacDorman, K. F., & Kageki, N. (2012). The uncanny valley [from the field]. *IEEE Robotics & automation magazine, 19*(2), 98–100. https://doi.org/10.1109/MRA.2012.2192811
22. Nass, C., & Moon, Y. (2000). Machines and mindlessness: Social responses to computers. *Journal of social issues, 56*(1), 81–103. https://doi.org/10.1111/0022-4537.00153
23. Oh, C. S., Bailenson, J. N., & Welch, G. F. (2018). A systematic review of social presence: Definition, antecedents, and implications. *Frontiers in Robotics and AI, 5*, 409295. https://doi.org/10.3389/frobt.2018.00114
24. Pennebaker, J. W., & Chung, C. K. (2007). Expressive writing, emotional upheavals, and health. *Foundations of health psychology*, 263–284.
25. Pickard, M. D., Roster, C. A., & Chen, Y. (2016). Revealing sensitive information in personal interviews: Is self-disclosure easier with humans or avatars and under what conditions?. *Computers in Human Behavior, 65*, 23–30.
26. Prochaska, J. J., Sung, H. Y., Max, W., Shi, Y., & Ong, M. (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), 88–97.
27. Reis, H. T., & Shaver, P. (1988). Intimacy as an interpersonal process. *Handbook of personal relationships*.
28. Reis, H. T., Lemay Jr, E. P., & Finkenauer, C. (2017). Toward understanding understanding: The importance of feeling understood in relationships. *Social and Personality Psychology Compass, 11*(3), e12308.
29. Schuetzler, R. M., Grimes, G. M., Giboney, J. S., & Nunamaker Jr, J. F. (2018). The influence of conversational agents on socially desirable responding. In *Proceedings of the 51st Hawaii International Conference on System Sciences* (p. 283).
30. Strohkorb Sebo, S., Traeger, M., Jung, M., & Scassellati, B. (2018, February). The ripple effects of vulnerability: The effects of a robot's vulnerable behavior on trust in human-robot teams. In *Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction* (pp. 178–186).
31. Shin, D. H., & Choo, H. (2011). Modeling the acceptance of socially interactive robotics: Social presence in human-robot interaction. *Interaction Studies, 12*(3), 430–460. http://dx.doi.org/10.1075/is.12.3.04shi
32. Skjuve, M., Haugstveit, I. M., Følstad, A., & Brandtzaeg, P. (2019). Help! Is my chatbot falling into the uncanny valley? An empirical study of user experience in human-chatbot interaction. *Human Technology, 15*(1), 30–54.

33. Sprecher, S., & Treger, S. (2015). The benefits of turn-taking reciprocal self-disclosure in get-acquainted interactions. *Personal Relationships, 22*(3), 460-475. https://doi.org/10.1111/pere.12090
34. Medhi Thies, I., Menon, N., Magapu, S., Subramony, M., & O'neill, J. (2017). How do you want your chatbot? An exploratory Wizard-of-Oz study with young, urban Indians. In *Human-Computer Interaction-INTERACT 2017: 16th IFIP TC 13 International Conference, Mumbai, India, September 25-29, 2017, Proceedings, Part I 16* (pp. 441-459). Springer International Publishing.
35. Utami, D., Bickmore, T., Nikolopoulou, A., & Paasche-Orlow, M. (2017). Talk about death: End of life planning with a virtual agent. In *Intelligent Virtual Agents: 17th International Conference, IVA 2017, Stockholm, Sweden, August 27-30, 2017, Proceedings 17* (pp. 441-450). Springer International Publishing.
36. Xu, Y., Shieh, C. H., van Esch, P., & Ling, I. L. (2020). AI customer service: Task complexity, problem-solving ability, and usage intention. *Australasian marketing journal, 28*(4), 189-199. https://doi.org/10.1016/j.ausmj.2020.03.005
37. Yang, D., Yao, Z., Seering, J., & Kraut, R. (2019, May). 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* (pp. 1-15). https://doi.org/10.1145/3290605.3300261