Exploring Effects of Chatbot-based Social Contact on Reducing Mental Illness Stigma

Yi-Chieh Lee
National University of Singapore & NTT, Singapore
ejli.uiuc@gmail.com

Yichao Cui
Cornell Tech, USA
yc793@cornell.edu

Jack Jamieson
NTT Communication Science Laboratories, Japan
jack@jackjamieson.net

Wayne Fu
Cascade Science, USA
wayne_fu@acm.org

Naomi Yamashita
NTT Communication Science Laboratories & Kyoto University
Japan
naomiy@acm.org

ABSTRACT

Chatbots have been designed to provide interventions in mental healthcare. However, how chatbot-based social contact can mitigate social stigma in mental illness remains under-explored. We designed two chatbots that deliver either first-person or third-person narratives about mental illness and evaluated them using a mixed methods study. Compared to a web survey group, participants in both chatbot groups decreased their beliefs that individuals are personally responsible for their mental illnesses, and increased their intentions to help. Additionally, participants in the first-person chatbot group showed a reduced level of fear, and a lower desire for social distance from people with mental illness. Many in the first-person chatbot group also reported a feeling of relationship with the chatbot, and chose to phrase their responses empathetically. Results demonstrated that chatbot-based social contact has promising potential for mitigating mental illness stigma. Implications for designing chatbot-based social contact are discussed.

CCS CONCEPTS

• **Applied computing** → *Psychology*; • **Human-centered computing** → **User studies**.

KEYWORDS

Chatbots; Conversational Agents; Social Stigma; Mental Illness

ACM Reference Format:
Yi-Chieh Lee, Yichao Cui, Jack Jamieson, Wayne Fu, and Naomi Yamashita. 2023. Exploring Effects of Chatbot-based Social Contact on Reducing Mental Illness Stigma. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (CHI '23), April 23–28, 2023, Hamburg, Germany*. ACM, New York, NY, USA, 16 pages. https://doi.org/10.1145/3544548.3581384

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

ACM ISBN 978-1-4503-9421-5/23/04…$15.00
https://doi.org/10.1145/3544548.3581384

1 INTRODUCTION

Stigmatizing someone consists of holding negative attitudes about, and/or practicing discrimination against them based on their distinctive and undesired characteristics [38]. People with mental illness are often stigmatized and isolated by others [25, 81], in part because they are routinely deemed dangerous and personally responsible for their disabilities [19]. Social stigma is a serious barrier to people with mental illnesses seeking help and recovering from it. Prior research [23, 66, 81] has extensively investigated interventions and strategies for eliminating mental-illness stigma, and the World Health Organization and many non-profits have launched anti-stigma interventions such as public-awareness campaigns, protests, and events that encourage social contact [25]. Encouraging social contact between the general public and people with mental illness has been found to promote understanding, reduce perceptions of dangerousness, and lead to a reduction in social stigma [19, 30, 66]. One major obstacle is the lack of resources to systematically launch large-scale effective campaigns to mitigate social stigma.

Within the HCI domain, conversational agents – widely known as chatbots – have been adopted for interventions in mental healthcare [1, 35, 61]. Despite the potential of those interventions, they generally do not address structural factors that exacerbate negative experiences with mental health [72]. Stigma is an example of such a structural factor, in that reducing social stigma would not directly alleviate clinical symptoms but would positively impact the well-being of people with mental illness. On the other hand, recent studies showed that chatbots can be a cost-effect way to facilitate social support in safe, less-stigmatizing environments in which people with mental illness can disclose truthful information [58] and can provide support in various roles to reduce social barriers in human-human interaction, such as the fear of being judged [54, 89] and avoidance of revealing vulnerabilities to others [53].

While this promising body of literature suggests that chatbots, even when known to be non-human, could somehow mediate various social responses from humans, research has only touched the surface of this topic. For example, studies in this area have included chatbots that interact with people in the first person (e.g., sharing their experiences as a depressed person [52]), that mediated sensitive information sharing between humans as a third-party [53], and some have explored differences between first and third-person storytelling [8]. However, it is unclear if and how the effects of first

and third-person chatbot interactions may be leveraged to fight social stigma.

Within the enormous design space of chatbots, we focus on their roles in providing social contacts to reduce mental illness stigma. In addition to understanding the general effectiveness of chatbot-based social contact, we are also interested in studying whether and how first- and third-person chatbots may have differential effects on reducing mental illness stigma. To this end, we designed two chatbots in our study: one playing the role of someone experiencing mental illness and telling their own stories from a first-person perspective; and the other acting as a mediator and introducing the same stories to the participants from a third-person perspective.

We conducted two-week mixed methods study with 89 participants randomly assigned to each of three conditions: one for each chatbot, and a web survey group. Both chatbot groups read stories and responded to questions through interaction with the chatbot, whereas the web survey group read the same stories and answered the same questions via online surveys. We measured the participants' attitudes and desire for social distance [9, 37, 69] towards the person with mental illness who was enacted/described in those stories, before and after reading them. We also conducted an exit interview to better understand the participants' experiences and hear their reflections. Their data provided us with an empirical understanding of both the positive and negative impacts of our two chatbots' designs on their impressions and understanding of people with mental illnesses.

Our work makes several contributions to the HCI community. First, this study enhances our understanding of whether and how chatbots can be used to reduce stigma towards people with mental illness. Specifically, we found that different chatbot roles could mitigate mental-illness stigma, enhance people's awareness of biased thoughts, and encourage self-disclosure of thoughts related to mental illness. Second, our results provide insights into how chatbots with different designs can cultivate different types of relationships with users, and how these relationships affect people's perceptions and attitudes. Particularly, the chatbot telling stories from a first-person perspective was able to stimulate social contact and reduce stigmatizing attitudes toward mental illness. Finally, this study provides a unique perspective on how human-AI interactions can be designed to promote positive social impacts and discusses design and ethical implications for future research to develop mechanisms to change people's stigmatizing attitudes.

2 RELATED WORK

2.1 Defining Mental-illness Stigma

Stigma related to mental illness arises from stereotypes, prejudices, and discrimination [26], and is related to how mentally ill people's characteristics diverge from what is considered normal and correct by society [26, 38]. Such stigma contributes to many negative outcomes, including reluctance to seek mental healthcare in order to avoid being stigmatized [18], amplified historical injustices on minoritized experiences [72], and difficulty finding employment, housing, and social contact [26]. Often, mental-illness stigma arises as unconscious bias, i.e., negative attitudes or stereotypes that can influence people's decisions without them being aware of it [59]. In this paper, we focus on stigma from other people toward people with mental illness.

This paper employs Corrigan et al.'s attribution model of mental illness stigma [19, 20]. This model describes that people's beliefs about the degree to which a person with mental illness is responsible for their condition are associated with stigmatizing emotional responses (e.g., pity, anger, and fear). Subsequently, those emotional responses are associated with discriminatory behavioral responses. For example, believing that someone with a substance abuse problem is at fault for their condition can lead to anger or a lack of pity, which can in turn contribute to behavioral responses such as wanting to keep their distance and believing they should be coerced into psychiatric treatment. Additionally, we draw from prior research indicating that desire for changing social distance from persons with mental illness is a reliable measurement component of stigma toward such people [6, 19]. A desire for social distance can manifest as not wanting to collaborate on group projects or socialize with a person experiencing mental illness [9].

In prior work measuring mental-illness stigma, researchers have routinely elicited responses using *vignettes* [60]: short stories about a fictitious character, developed from previous research and/or real-world examples [39]. After reading vignettes, respondents are asked to explain what they think about the stories' character(s) using open-ended comments, or by answering a survey or other questions [2, 32]. A major advantage of this approach is that it helps people relate their perceptions and attitudes about mental illness to concrete situations [60]. Based on prior work, our study employs vignettes to tell stories about mental illness to participants, and then measures their propensity toward mental-illness stigma using constructs from Corrigan et al.'s [19] attribution model and the SDS [56, 68], as described further in Section 4.

2.2 Interventions for Reducing Stigma

Recent work has highlighted the potential of various technologies to help people who are targets of stigma: e.g., transgender people [41], people who experience pregnancy loss [3], and people with disabilities [46]. There is extensive literature about chatbots as part of mental illness interventions, such as using a chatbot to deliver cognitive behavioral therapy to patients with depression and anxiety [35] or to offer mental illness patients the safety of anonymity in health care services [61], and evaluate mental illness patients' perceptions of chatbots [1]. Chatbots' advantages for mental healthcare include their capacity for flexible and ubiquitous treatment [1, 35, 61], and the fact that many people are more willing to disclose their mental-health symptoms to a chatbot than to a human [54, 58, 77].

The question of how technology can enable new (or enhance existing) interventions for reducing mental-illness stigma remains under-researched [76, 87]. Most existing studies have focused on raising awareness; e.g., through social media campaigns and the activities of grassroots organizations [34, 76, 87]. Most people are already somewhat aware of depression [26, 30], but even well-known mental illnesses are subject to stigma [84]. Thus, beyond just raising awareness, it is important to dispel misconceptions that could lead to stigmatizing beliefs. A promising approach to this end is to increase positive social contact, as peers, between

members of the general population and people with mental-health problems [26, 30, 66].

Technology has immense potential for increasing social contact, whether face-to-face, remote, or simulated. For example, Rodríguez-Rivas et al. [80] showed that the use of virtual reality and communication technologies to make contact with mental-health service users effectively reduced mental-health stigma among the wider public. Similarly, Cangas et al. [10] demonstrated the potential of simulated social contact by inviting high-school students to play a videogame featuring characters with mental illnesses. Specifically, they reported that the players' stigmatization of people with schizophrenia, and perceptions that such people are dangerous, were both reduced. The key aim of social-contact based interventions of this kind is to facilitate self-disclosure by people who suffer from mental illness, which can increase their potential for friendship and intimacy [11, 30, 86]. Such interventions can also increase the general population's understanding and ability to make informed decisions related to mental illness [26, 30].

Storytelling has been found to be a useful and efficient technique in increasing social contact to reduce mental illness stigma [21, 28, 36, 50]. According to Corrigan et al. [21, 22], storytelling is efficient in reducing mental illness stigma because of three elements: storytellers with lived experience, in-person delivery of stories, and content about mental illness recovery journeys. Moreover, storytellers can disclose experiences in depth by describing rich contexts and building connections with listeners [36, 67]. In this way, storytelling can enhance people's engagement, deepen their understanding of mental illness, and facilitate empathy toward mental illness patients [15]. Existing studies have explored ways of facilitating storytelling. For example, Fong et al. [36] found that interactivity in storytelling could enhance stigma reduction effects, but that content was overall more important than interactivity. Bickmore et al. [8] developed a chatbot that presented first-person or third-person life stories (including topics such as sports, family, and outdoor activities) to users. They found that users under first-person storytelling reported more engagement with the chatbot's stories and completed more conversations with the chatbot [8]. Although existing work explores the effects of storytelling, there is a research gap in exploring whether first-person and third-person storytelling can help reduce social stigma.

Chatbots have considerable potential for capturing the benefits of positive social contact, insofar as they are generally good at facilitating people's self-disclosure, sense of attachment, and feelings of closeness [54, 58, 83]. In addition, as highlighted by the Computers Are Social Actors (CASA) paradigm, people interacting with chatbots tend to apply social norms and expectations (e.g., reciprocity and politeness) derived from human-human relationships [65]. Research [64] found that users were likely to self-disclose in reciprocity to a self-disclosing chatbot. However, there has been little research on how chatbots could leverage this to reduce mental-health stigma in the general population [47]. In one of the few exceptions, Kim et al. [47] attempted to reduce people's stigmatization of depressed individuals using a social bot [33] that described its own depressive symptoms and vulnerabilities through Facebook posts, and encouraged them to reply. Although interaction between Kim et al.'s participants and the social bot was limited, participants reported lessened feelings that people with depression were dangerous, and a greater desire to help them.

The chatbots hitherto deployed in the mental-health domain have played a variety of roles [52, 53, 83]. For example, participants in one study [52] offered care and support to a chatbot that disclosed its own mistakes, a process that resulted in them developing more compassion for themselves. Other research [54, 83] further demonstrated that human-chatbot relationships could gradually be developed through mutual self-disclosure, which can have a positive impact on mental well-being. In addition, prior work [53] has demonstrated that chatbots can act as mediators between humans, transferring not only sensitive information, but also bridging trust between parties that have not directly communicated. Although the existing literature has demonstrated that there are advantages to both first-person chatbots that disclose their own experiences, and third-person ones that act as mediators, it is not currently clear which approach can best be leveraged to reduce stigmatizing attitudes. Filling this gap is important, in that doing so could lead to design innovations that improve people's understanding of mental illness, and help in the fight to reduce stigma.

3 RESEARCH QUESTIONS

We compare our two chatbots with a survey-based approach. The survey-based approach is similar to that used in prior research [19], where participants read vignettes and respond to them using surveys. By comparing chatbots that disclose vignettes about mental illness from both a first-person and third-person perspective, we expect to learn how each of these approaches shapes relationships between chatbots and humans, before proceeding to an investigation of the extent to which both designs may lead to stigma reduction. In these efforts, we are guided by the following research questions (RQs):

RQ1: *To what extent does interacting with each of the two chatbot designs facilitate social contact, compared to interacting with a web survey?* To answer this question, we assess (a) the extent to which participants engage in self-disclosure, and (b) participants' impressions and perceived relationship with the character whose experiences with mental illness are described in the vignettes. Answering this question will help us understand the extent of each design's effects for different types of relationships between chatbots and humans. This leads in to our next RQ:

RQ2: *To what extent does interacting with each of the two chatbot designs affect participants' stigmatizing attitudes toward people with mental illness, compared to interacting with a web survey?* Through answering this question, we hope to discern how reading multiple stories told by vs. about a person living with mental illness via interacting with chatbots could change stigmatizing attitudes toward mental illness. We will triangulate such understating by operationalizing stigma using attribution theory [19], social distance [56, 68], and follow-up interviews.

4 METHODS

Our study is a three-condition, between-subjects, and randomized experiment, comparing two chatbot conditions with a non-chatbot condition, with attribution and social distance as the primary measurements. On the first day of the experiment and every other day

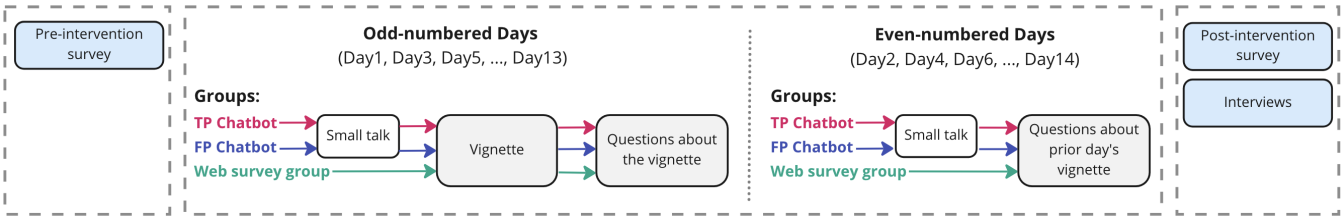

Figure 1: Study design. Our study includes three groups (independent variables): Web survey, Third-person chatbot (TP), and First-person chatbot (FP). We analyzed response logs (how participants responded to the chatbot), pre- and post-intervention surveys, and interviews to investigate participants' relationships with the chatbot and changes in stigmatizing attitudes. In the two-week study, the web survey group completed its daily task using an online survey, whereas the TP and FP groups interacted with chatbots. The TP and FP differed in that the vignettes were presented in the third-person and first-person, respectively. However, the conversation flow as illustrated in this figure was otherwise identical.

First-Person Vignette	Third-Person Vignette
a. When I'm at work, I get things done as usual - but I know that it is not my best.	a. When Kenta is at work, he gets things done as usual - but he knows that it is not his best.
b. In the past, I could take care of 5-6 tables of customers at the same time; however, now I can only manage 1-2 tables.	b. In the past, he could take care of 5-6 tables of customers at the same time; however, now, when he tries to manage more than 1-2 tables, he has trouble remembering all the orders.
c. However, I cannot bring myself to ask for my coworkers' help. I already feel like a waste of space and I don't want to be a burden.	c. However, Kenta cannot bring himself to ask for his colleagues' help, since he already feels like a waste of space and does not want to be burdensome.
d. I can tell that my coworkers are confused. My coworkers can see that I've changed, and I know they can see that I'm frustrated. I still refuse their help.	d. Kenta's colleagues also feel confused about Kenta's change because he refuses any help while looking very frustrated.
e. In fact, I'm so tired and I've decided I'm going to quit in the future. When I think about quitting, though, I feel like a failure and I feel like there's something wrong with me.	e. In fact, Kenta feels tired and has decided he will quit in the future. Whenever thinking about the idea of quitting, Kenta feels like a failure, and he believes that there is something wrong with him.
f. Things are getting worse and my manager has asked me to improve my performance.	f. As Kenta's conditions are getting worse, his manager asks him to improve work performance.

Figure 2: Two versions of vignette 2. First-person vignette seen by FP chatbot group (left); Third-person vignette seen by TP chatbot group and web survey group (right).

thereafter (1, 3, 5, etc.; see Figure. 1), all participants were presented with a vignette about a person living with mental illness and asked questions about their opinions and their own experiences. On even-numbered days (2, 4, 6, etc.), they were asked additional questions about the previous day's vignette, aimed at establishing whether our three different methods of delivering vignettes would differently affect the participants' perceptions of and attitudes toward a person living with mental illness. All participants read seven vignettes over 14 days, which prior work suggests could deepen their understanding of mental illness and facilitate their empathy in mental illness patients [60]. Further details about our three experimental groups are provided below.

Web survey group: Using a web survey platform (Google Form), participants read vignettes composed in the third-person and then responded to questions. This design replicates existing studies [2, 32], which asked participants to read vignettes via a web survey or a hard copy, in order to investigate attitudes of the general population towards persons with mental illness. Participants in this group read a paragraph of the vignette and then answered open-end questions after reading it. We designed this condition chiefly to facilitate study of the chatbots' impact on the other two groups.

Third-person (**TP**) chatbot group: Participants interacted with a chatbot that told the vignettes in the third-person, acting as a mediator between the participant and the character in the vignettes who had undergone those experiences. We did not define the chatbot's name, gender, or appearance, to avoid introducing biases based on those traits. Daily interactions started with small-talk, followed by the vignettes, and then the same follow-up questions as in the web survey group. Based on this design, the TP chatbot demonstrates a more interactive form of storytelling compared to the web survey, which could lead to greater engagement and impact from the vignette content.

First-person (**FP**) chatbot group: Participants interacted with a chatbot that told the vignettes in the first-person, playing the role of the character from the vignettes telling his own experiences and stories. The conversation flow was otherwise identical to the TP chatbot group. This first-person condition extends on Bickmore et al.'s [8] study, which suggested that using a first-person perspective can enhance users' engagement with a chatbot and increase their retention of its stories. Additionally, we examine whether the FP chatbot can stimulate social contact [26, 30] to leverage the fact that first-hand encounters with storytellers with lived experience has been identified as an effective storytelling strategy for reducing mental illness stigma [21, 22].

Table 1: Vignette (Vig.) design and contexts in the two-week experiment.

Vig.	Day	Symptoms	Context
V1	1	Trouble thinking/concentrating Loss of interest	Academic studying
V2	3	Slowed speaking or body movements Tiredness/lack of energy	Working
V3	5	Feelings of emptiness or hopelessness Feelings of worthlessness or guilt	Dealing with intimate relationships
V4	7	Angry outbursts, irritability or frustration	Interaction with friends
V5	9	Sleep disturbances Decreased appetite	Staying with family members
V6	11	Anxiety, agitation or restlessness Feeling sad, hopeless or worried	Interaction with strangers
V7	13	Self-harm thoughts	Being alone

4.1 Vignette Design

The vignettes, seven in total, were all about the same man, Kenta, who experiences depression symptoms. Because prior research [17] has suggested that gender may affect respondents' answers about mental-illness stigma, we assigned the chatbot one gender to keep this consistent across all participants and conditions. Each vignette was created based on a different context (Table 1). We extracted these contexts and their associated symptoms from the DSM-5 and from vignettes related to depression used in prior studies [14, 45, 51, 55, 57, 73][1]. The seven draft vignettes were reviewed and approved by a psychiatrist member of our research team. By way of example, Figure. 2 presents both the first-person and third-person versions of Vignette 2, the context of which is work, where Kenta experiences lack of energy and slowed thinking and body movements. We composed this vignette to represent some of the symptoms of depression, based on prior work showing how those symptoms can negatively impact people at work [55]. Full vignette content for all three groups and sample dialog flows can be found in the *Supplemental Materials*.

4.2 Tasks

4.2.1 Chatting Tasks – Chatbot Groups. Both our chatbots' conversational flows were inspired by previous findings that small talk increased user engagement in human-chatbot interaction [43, 54, 83]. Prior studies [7, 54, 83] have also shown that small talk is an effective means of gaining trust from chatbot users and facilitating disclosure of sensitive topics. Thus, chatbot conversation in both of our chatbot conditions commenced with small talk and then proceeded to vignette delivery. Importantly, both chatbots performed small talk in the first person, i.e., referred to their own opinions using "I" and "me." Consequently, for FP chatbot users, the small talk appeared to be with Kenta himself, whereas in the TP condition, the chatbot and Kenta were separate entities. The small-talk topics, including but not limited to hobbies, weekend plans, and favorite foods, were from previous studies [4, 54].

Vignette Delivery. We elected to share a vignette with the participants every other day, rather than every day (Table 1), to avoid them becoming overwhelmed by reading depression-related material.

[1]https://div12.org/case-studies/

Each vignette was delivered across multiple messages, as shown in Figure. 1. After each message, the participant was invited to respond briefly, and once s/he did so (or, after 1-2 minutes if s/he made no response), the chatbot would proceed to its next vignette message or, once the vignette was finished, ask some open-ended questions about the participants' own experiences and attitudes related to the vignette's content. These questions were adapted from prior studies aimed at evaluating mental-illness stigma [40, 45, 48, 78]. Questions related to whether the participants had had similar experiences included, *"Have you ever had feelings of anger toward your friends for a period of time, like me?"* and *"Have you ever had problems similar to Kenta's?"*; plus a follow-up question if the participants said they had had similar experiences: *"Can you please describe what happened?"* The questions related to the participants' attitudes and perceptions toward people with mental illnesses included, for instance, *"Do you believe that mental illness is a sign of weakness? Why?"* and *"Do you think Kenta is responsible for his situation?"*

On even-numbered days, when the participants did not receive new vignettes, the conversation again started with small talk, but they were then asked for their suggestions about how to improve the situation described in the previous days' vignette (e.g., *"What would you do to deal with that problem?"*), and whether they knew others who had similar experiences (e.g., *"Has anyone in your family or friendship group ever had problems similar to Kenta's?"*)

4.2.2 Web survey Tasks – Web survey Group. The participants in the web survey group received the same vignettes and questions as those in the chatbot groups, following the same schedule, and with the vignettes told in the third person. Thus, the web survey group's experience matched the TP chatbot group's, except in that: 1) the interface was a web survey, 2) each vignette was presented in a whole paragraph, and 3) there was no small talk.

4.3 System Implementation

The left-hand side of Figure. 3 shows the chatbot interface, while its right-hand side shows the online-survey interface used by the web survey group. In each of these cases, the participants were allowed to use their own devices to access the interface. All participants were informed that their responses would be recorded and shared with the research team. We built the chatbot using ManyChat and Google Dialogflow. ManyChat enabled us to develop the main conversational flow and manage multiple participants during the study, while Dialogflow – which makes use of use natural language processing – was incorporated to increase the naturalness of the human-chatbot conversations by enabling the chatbots to give plausible responses to a wide range of participant questions. In cases where a participant said something beyond the scope of the predefined content, Dialogflow helped process such statements by providing simple responses and refocusing the participant on the chatting task's topics. If a chatbot detected that a participant got stuck three times, it would move on to the next question.

4.4 Participants

We used social media and university bulletin boards to recruit participants from local universities and communities. Our recruiting criteria were that participants must 1) be at least 18 years of age;

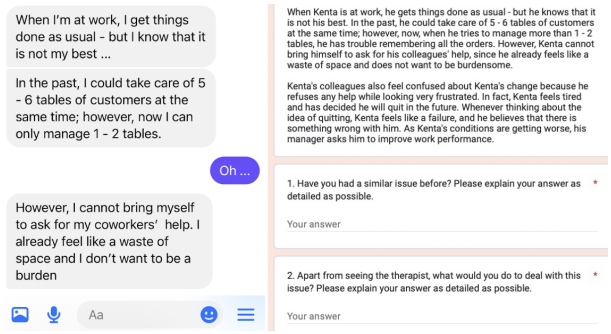

Figure 3: Chatbot interface(left) and web survey group survey interface(right). To fit the conventions of each medium, the vignettes are split into several messages in the chatbot interface and written as full paragraphs in the survey interface.

2) be able to read, write and speak fluent English; 3) be able to use messenger platforms on their own devices; and 4) score less than 12 on the K6 distress scale [75], meaning that they did not have an urgent mental-health issue. We included this final criterion because there was a potential risk that the vignettes about mental illness could cause distress to people who did have such issues [79]. In the recruiting poster, we disclosed the duration of the study, along with the participants' right to drop out at any point, and each participant's option to attend a follow-up interview at the end of the two-week experiment.

This process resulted in the recruitment of 89 participants with a mean age of 27.57 (SD = 4.79), none of whom reported having an ongoing mental illness or attending counseling sessions. In light of prior research findings on the potential impacts of both mental-health literacy and gender on mental-illness stigma [44, 71], the three groups were balanced in terms of their scores on the Mental Health Literacy Scale (M = 116.4, SD = 13.35) and gender. The web survey group comprised 16 females and 13 males; the FP chatbot group, 17 females and 13 males; and the TP chatbot group, 17 females and 13 males. We conducted a power analysis [31] finding that, to achieve 80% power for detecting a medium effect size (f = .3), with an alpha level of 5% [16], the required sample size is 28 for each group.

4.5 Procedure

Before the two-week experiment began, all participants were asked to attend an initial online meeting in which the researchers explained its requirements and helped them set up their respective groups' chatbots on their own devices (e.g., mobile phone, laptop). All participants were informed of their right to drop out of the experiment at any time, as noted above, but told that full compensation (US$84) was contingent upon their completion of the 14 daily tasks. If a participant could not meet this requirement, their compensation was reduced in proportion to the number of days they completed, at a rate of $6 per day. They were also notified that if any content or questions made them feel uncomfortable, they could skip them without penalty. They were instructed not to discuss their respective interactions with one another until after the experiment was finished. TP and FP participants were informed that they would be interacting with a "chatbot," not an actual human, and their responses would not be shared with others (only to researchers). For TP participants, the chatbot did not describe having a specific type of relationship with Kenta (e.g., friends, colleagues, and doctors). This was to keep the web survey and TP chatbot conditions consistent, and to avoid introducing a specific relationship as a confounding variable. In the same meeting, the participants were asked to fill out a pre-intervention survey to collect their original beliefs and perceptions toward people with mental illnesses (Figure. 1). Finally, to ensure they understood how to complete the daily tasks, they were guided through a 10-minute training session with their chatbot or survey.

On each day of the experiment, the FP chatbot and TP chatbot groups received a new chatting task reminder from the chatbot, and the web survey group received a message containing a URL for their survey task. All groups were given about 10 hours (i.e., from 2 p.m. until 11:59 p.m.) to finish each day's task, which was designed to last about 15 minutes. If a participant tried to access the chatbot outside of the daily-task time window, the chatbot would not respond.

At the end of the experiment, all participants were asked to complete a post-intervention survey containing the same items as the pre-survey (Figure. 1), to allow us to examine the extent to which our interventions impacted their perceptions of and attitudes toward persons with mental illnesses. In recruiting interviewees, we strove to keep the numbers from each group balanced; there were 49 – representing 55% of all study participants – including 16 from the web survey group, 17 from the FP chatbot group, and 16 from the TP chatbot group. The researchers debriefed each interviewee about this research's goal and discussed any concerns s/he had at the end of the interview. The researchers also sent an email including relevant anti-stigma educational materials and other information to all participants, to help them better understand mental-illness stigma.

4.6 Measurement

To investigate whether and how chatbot designs affect people's stigmatized attitudes toward persons with mental illness, our research follows the model in Figure. 4 to progressively break down the impact of the intervention designs. RQ1 investigates how our designs could facilitate participants' social contact with the chatbot. The quality of social contact is evaluated by analyzing participants' self-disclosure to their chatbot partner in the daily response logs, as well as interview responses about their perceived relationships with the chatbot. In RQ2, we examine how chatbot-based social contact affects stigmatizing attitudes and beliefs. This analysis draws on Corrigan et al.'s [19] attribution model, which states that personal responsibility beliefs are associated with emotional responses, which are subsequently associated with discriminatory behavioural responses. To answer these questions, we triangulate our analysis across surveys, daily response logs, and interviews, as described in the following sections.

4.6.1 Surveys. Before filling out the pre-survey mentioned above, participants read the following vignette: *"Kenta is a 22-year-old man pursuing his bachelor's degree in Japan. In his spare time, he works as a waiter at a local restaurant, and spends a great amount*

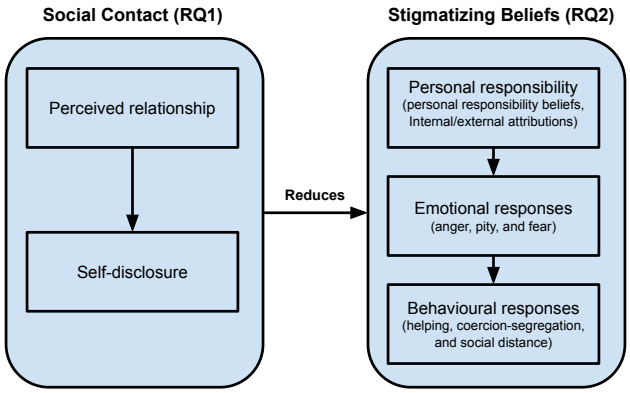

Figure 4: Study conceptual model. To evaluate the chatbot's ability to replicate social contact to reduce mental illness stigma, we interviewed the participants and measured their self-disclosure behaviors through participants' response logs to understand the social contact quality, and we then analyzed the pre-and post-intervention surveys to analyze changes in stigmatizing attitudes (dependent variables).

of time reading and writing. However, Kenta has been diagnosed with depression recently. Sometimes, he becomes upset and cannot concentrate on his studies and work. He lives with his girlfriend and cannot to do much, especially household chores. He feels angry about his surroundings, and gets frustrated about where the fury comes from. When Kenta is alone, he has realized that he has self-harm intentions." In the post-survey, on the other hand, the participants did not read this or any other vignette, but instead answered the survey items based on the impressions of Kenta they had gained during the two-week study.

The survey questions were borrowed from two questionnaires widely used in prior literature: Corrigan et al.'s attribution questionnaire [19] and the social distance scale (SDS) [56, 68]. The attribution-questionnaire items covered the following three main components: 1) Personal responsibility beliefs, 2) Emotional responses: *pity*, *anger*, and *fear*, and 3) Behavioural responses: *helping*, and *support for coercion and segregation* [19]. Each such item was rated on a nine-point semantic-differential type scale ranging from 1 = "not at all" to 9 = "very much." Example questions were: 1) *"How responsible, do you think, is Kenta for his present condition?"* and 2) *"How dangerous would you feel Kenta is?"* The SDS provides further insight into behavioral responses. It is widely used to evaluate participants' behavioral intention towards people with mental illnes [56, 68]. The SDS contains seven items [56], examples of which include: *"How would you feel about renting a room in your home to someone like Kenta?"* and *"How would you feel about recommending someone like Kenta for a job working for a friend of yours?"* All the SDS items we used were responded to on a four-point scale, from 0 = "definitely willing" to 3 = "definitely unwilling."

4.6.2 Daily response logs. All the participants' conversations with chatbots and responses to online surveys were logged. After reading each vignette, as briefly noted above, they were asked two types of questions over two days: first, whether they had experienced something similar to what was described in the vignette (and asked to elaborate if they answered positively); and second, about their perceptions and attitudes toward Kenta's behavior. They were also asked to provide advice to Kenta.

To explore any differences among responses from the three groups, two raters were hired to code all data independently. Before actually rating, they practiced rating all participants' first two days of responses and discussed differences until a consensus was reached, and then discussed revisions. They categorized responses, firstly, according to whether the participant said s/he had had a similar experience; and if so, secondly, whether they self-disclosed about that experience in their responses [5]. We coded participants' self-disclosure because of prior research findings [54, 83] that people's self-disclosure to a chatbot is related to their perceived closeness with it; therefore, the presence of self-disclosure by our participants could imply positive social contact between them and their respective chatbots, which would help us answer RQ1.

Thirdly, the coders categorized whether their perceptions and attitudes were related to internal attributions (e.g., personality, beliefs) or external ones (e.g., situational or environmental features). This was done because of previous studies' suggestions that attribution error is related to people's prejudice and bias [24, 42]. If a participant's response included both internal and external attributions, then both codes were recorded. The results of this part of our coding scheme helped us answer RQ2. Example responses are presented in the Results section.

4.6.3 Interviews. Our interviews were semi-structured and lasted 35-45 minutes. They focused on the interviewees' 1) daily practices and experiences of using our systems; 2) impressions of and attitudes toward the vignette's character; 3) impressions and perceptions of the chatbot, if applicable; and 4) understandings of and reflections on the impact (if any) of their participation in the study on their thoughts about people with mental illnesses. Regarding the first category, we asked if they felt hesitant while reading the vignettes and/or while answering the follow-up questions. Regarding the second and third categories, we asked them to describe their impressions of the character of Kenta (and, for those in the TP condition, the character of the separate chatbot), and where those impressions came from. We also asked if such impressions, and/or their relationships with these characters, changed over time; and if so, how and why. For the final category, all the participants were asked to share their reflections about reading the vignettes and what understanding they may have gained from that process.

We transcribed the interview recordings and used thematic analysis to categorize each response according to the questions' contexts [70]. Two researchers independently familiarized themselves with the interview data and iteratively reviewed and labeled six interview responses to develop initial codes. The researchers then coded all the other interview responses independently, and held meetings to review and compare their coding results. Then, they discussed possible revisions. The cycle was repeated until the coding scheme was deemed satisfactory by both raters, and inter-rater reliability had reached a reasonable level (κ = .86).

Table 2: Results of coding a conversational log (a) where the participant had similar experiences or symptoms as the character in the vignette; and (b) where the participant shared his/her prior experiences

		V1	V2	V3	V4	V5	V6	V7
(a) Similar Experience (Yes/Responses)	Web survey	10/28 (36%)	14/29 (48%)	20/29 (69%)	14/28 (50%)	13/28 (46%)	21/28 (75%)	21/29 (72%)
	FP chatbot	21/29 (72%)	17/30 (57%)	22/29 (76%)	16/29 (55%)	15/29 (52%)	23/30 (77%)	18/29 (62%)
	TP chatbot	22/30 (73%)	16/30 (53%)	24/29 (83%)	15/30 (50%)	16/30 (53%)	23/29 (79%)	21/29 (72%)
(b) Self-disclosure (Yes/Similar Experience)	Web survey	4/10 (40%)	9/14 (64%)	12/20 (60%)	8/14 (57%)	7/13 (54%)	9/21 (43%)	7/21 (33%)
	FP chatbot	10/21 (48%)	11/17 (65%)	17/22 (77%)	11/16 (69%)	13/15 (87%)	19/23 (83%)	11/18 (61%)
	TP chatbot	7/22 (32%)	8/16 (50%)	14/24 (58%)	8/15 (53%)	9/16 (56%)	14/23 (61%)	8/21 (38%)

5 RESULTS

5.1 Evaluating Social Contact (RQ1)

To answer RQ1, we analyzed the extent to which responses contained self-disclosure, which helped us infer whether social contact between participants and chatbots had been effectively stimulated. We then drew on our interview data to investigate reasons and context for the quantitative results.

To evaluate participants' self-disclosure, we counted how many in each group answered that they had similar experiences or symptoms to those described in each vignette (e.g., insomnia, trouble concentrating). As Table 2(a) shows, the results were similar across all three groups, and a Kruskal-Wallis analysis of variance (ANOVA) indicated that there were no significant inter-group differences. Of those participants who had similar experiences to Kenta's, we compared the number in each group who self-disclosed about those experiences in response to each vignette, and the results are displayed in Table 2(b). Our analysis indicated that there were significant inter-group differences in the level of self-disclosure ($F=4.85$, $p<.05$), with post-hoc analysis showing that there was more self-disclosure in the FP chatbot group than in the other two groups. This result suggests that the FP chatbot design facilitated the development of greater reciprocity, as compared to the other two conditions.

5.1.1 Perceived Relationships and Interaction with Chatbots. Even though all participants knew they were interacting with a chatbot, their interactions and reflections during interviews indicate that they treated the chatbot as a social actor. We derived this conclusion from analyzing how interview participants described their impressions of the chatbot, and the specifics of their interactions. Below, we present a selection of those findings.

When asked how they pictured Kenta, many (n = 13) interviewees from the **FP chatbot group** described their image of him in relation to their own lives. For example, P56 (F) said, *"Kenta and I went to the same college. [… But we] don't live in the same district. We don't work in the same restaurant."* Interestingly, this tendency was discernible only in the FP chatbot group's data.

Furthermore, 10 FP-group interviewees said they regarded the vignettes as Kenta's self-disclosures to them, and indicated that they therefore 1) wanted to reciprocate that disclosure and 2) felt socially engaged. P48 (M) stated: *"People around me wouldn't share those personal things that were displayed in the vignettes. I was a bit surprised about Kenta disclosing his depression stories, so I felt he wanted to build a close relationship with me and had trust in me. I kind of felt that he was anticipating I would respond in kind, and so decided that I should."* This may help explain why FP-group members engaged in more self-disclosure than other groups.

Interestingly, this sense of connection seemed to influence the way the FP chatbot group interacted with their chatbot. Two-thirds (n = 11) of the FP-group interviewees reported changing their chatting strategies to support Kenta and to avoid causing harm, as P35 (F) explained: *"In the beginning, I tended to use soft words in the conversation because Kenta was already gloomy, and I didn't want to make him feel worse. However, when I felt he was not getting better, I changed my tone to be stronger, because I wanted to shake him up and recommend that he see a doctor."*

Five FP-group interviewees mentioned that their worry about Kenta caused them emotional disturbance (e.g., feelings of frustration and sadness). As P33 (F) told us, *"I felt Kenta's symptoms were getting worse. I suggested he see a doctor, but he did not take my suggestion. I got frustrated and tried to hide my frustration when conversing, because it did not seem to stimulate him. I used strong words in the beginning, but learned that they were ineffective for comforting him. So I avoided using strong words in our later interactions."*

In the **TP chatbot group**, as compared to its FP counterpart, most interviewees did not provide vivid descriptions of the Kenta character, presumably because they did not directly interact with him. The majority of the TP-group interviewees (n = 11) tended to interpret the chatbot's role as intended to stimulate their own deep reflection on the relevance of Kenta' stories to their own lives. As P65 (M) stated in his interview, *"The chatbot was like a mental-health professional to share Kenta's stories and guide me to reflect on each question. It was helpful because some questions [e.g., if I had similar thoughts as Kenta or someone I know had similar experiences] reminded me that I should not ignore the fact that mental illness is around us."*

Additionally, our use of the TP chatbot to introduce vignettes and collect responses seems to have encouraged the participants to disclose their thoughts deeply. Six TP-group interviewees indicated that using the chatbot to collect their responses encouraged their self-disclosure because they did not have to worry either about its feelings or how Kenta might react to what they were saying. As S63 (F) put it: *"I felt comfortable disclosing my thoughts and experiences to the chatbot because some of my answers might be sensitive to disclose to a real person. The chatbot would not judge my thoughts, and also Kenta would not be affected by my responses."*

However, it is worth noting that four TP-group interviewees reported seeing their chatbot as a person who was gossiping about Kenta's difficulties and mental status, which led them to be reluctant to answer some of its questions. As P83 (F) said, *"I did not engage in the conversation with the chatbot, because it kept asking about my*

perceptions toward Kenta's situations according to those short stories [...] and ignoring his other aspects and characteristics. This made me feel the chatbot was just gossiping about Kenta's stories and didn't really care about him."

In the **web survey group**, no-one gave a vivid description of Kenta. Five interviewees of the web survey group reported that they intentionally offered survey responses with relatively positive tones, out of a sense that it was important to use good manners toward people with mental-health problems. For instance, S1 (M) told us, *"I kind of deliberately used a pleasant tone when answering the question. Although I did not have negative thoughts about people with mental illnesses, this doesn't mean I have a positive perspective about them either. I just didn't want to make myself look mean to them."*

In summary, our interview results suggest that several factors contributed to the changes in the interviewees' perceptions of their interactions and responses. In the FP chatbot group, because the chatbot directly portrayed the Kenta character, the interviewees had vivid impressions of Kenta and developed a sense of social contact with their chatbot through daily interaction. This led those participants to value being considerate toward Kenta, and even to change their conversational strategies to help him get well and/or out of a sense that they had a responsibility to respond. In general, the TP-group interviewees lacked the sense of having a relationship with their chatbot that the FP-group interviewees had with theirs; however, this meant that some TP members felt safer responding honestly, without fear of offending Kenta. Finally, because the web survey group was non-interactive in character, its members unsurprisingly did not report any feelings of social contact.

5.2 Changing Stigmatizing Attitudes (RQ2)

RQ2 examined the extent to which participants' stigmatizing attitudes were affected by the intervention in each group. This was mainly studied using the pre-and post-intervention survey results.

To analyze the survey results, we conducted a multivariate analysis of variance (MANOVA), and calculated Wilk's Λ to test whether there were one or more mean differences between experimental groups, across all dependent variables. The result showed that there was a statistically significant difference between groups (Wilk's $\Lambda = 0.58$, $F = 24.62$ $p < .05$). This implies that a significant portion of the variance among dependent variables was accounted for by group differences. We then conducted mixed-model ANOVAs to examine the direct effects of *group membership* (i.e., web survey group, FP chatbot group, or TP chatbot group) and *time*-point (i.e., pre- vs. post-intervention surveys), as well as the interaction effect *group membership* × *time-point*, on the participants' attitudes and beliefs toward people with mental illness. This was followed by post-hoc comparison using Tukey's HSD test, which tests pairwise means with adjustments for multiple comparisons. Mauchly's Test of Sphericity was also used to verify that the assumption of sphericity was not violated. The dependent variable was the self-reported score for each construct on the attribution questionnaire and SDS. The results are presented in Table 3.

By examining participants' attitudes and beliefs toward a person with mental illness across the pre- and post-intervention surveys, we found several significant results suggesting that our chatbot designs had affected their attitudes (Table 3). Overall, there was no significant difference between groups in the pre-survey since the participants were randomly assigned to the three groups with considering the gender balance. Analyzing the post-survey, we found that five out of seven dependent variables have a significant interaction effect, which implies that our interaction designs (groups) affected participants' emotional and behavioral responses toward Kenta differently. We further present their interview results to understand participants' rationale for their changing stigmatizing attitudes.

5.2.1 Personal Responsibility. This section examines participants' beliefs about the extent to which Kenta is personally responsible for his mental illness.

The *Personal responsibility* item in Table 3 is scored using Corrigan et al.'s [19] attribution model. There was a significant interactive effect of group membership and time-point on responsibility beliefs, even though there was no significant main effect of either of these variables. In the pre-intervention survey, all three groups had very similar levels of responsibility beliefs about Kenta. In the post-intervention survey, on the other hand, post-hoc analyses indicated that the FP and TP chatbot groups' responsibility beliefs were significantly lower than the web survey group's ($p < .05$). However, there was no statistically significant difference between the post-intervention survey scores of the FP and TP groups. These results indicate that the participants who interacted with either chatbot felt Kenta was less responsible for his present situation after reading the vignettes in various contexts, as compared to those participants who did not engage in any chatbot interaction.

Additionally, we triangulated this result by coding participants' daily response logs according to 1) internal attribution (e.g., personality, beliefs) or 2) external attribution (e.g., situational or environment features) because attribution error is related to people's prejudice and bias [42]. For example, a coding of external attribution was given to the following two statements: *"I think his colleagues have to be responsible for Kenta's situation since he was not in good condition apparently. They should help him out." "I understand that having studies, a part-time job, and many other things at the same time are stressful. Thus, it is not your fault. I'd encourage you to ask your colleagues for help."* The following statement, in contrast, was coded as internal attribution: *"Kenta has a weak spirit. He should be stronger. He's not the only person who has to work hard to survive."* The results are reported in Table 4, which shows that the FP chatbot group had a higher proportion of members who made external attributions regarding Kenta's vignettes than either the web survey group or the TP chatbot group. A Kruskal-Wallis ANOVA indicated that there were significant differences among all three groups ($F = 27.9$, $p < .001$). These findings suggest that the FP group members were the most likely to attribute Kenta's condition to situational rather than dispositional factors – a possible sign that their stigmatization of people with mental illnesses was lower than that of the other two groups.

Finally, interview responses yield further insights about why participants' evaluation of Kenta's personal responsibility differed between conditions. The majority (n = 11) of the **FP chatbot group** interviewees expressed the opinion that external factors were largely responsible for Kenta's difficulties. Moreover, many participants

Table 3: Mean and SD values of each construct in the post-intervention survey and mixed-model ANOVA results. X means no effect. This table highlights post-survey and statistics results since there was no significant difference between groups in the pre-survey responses.

	Post-survey						Mixed-model ANOVA					
	Control		**FP chatbot**		**TP chatbot**		**Interaction effect**		**Group membership**		**Time-point**	
	Mean	SD	Mean	SD	Mean	SD	F	*p*	F	*p*	F	*p*
Personal Responsibility	5.00	2.13	3.89	2.30	4.21	1.72	3.99	<.05		X		X
Emotional Response												
Anger	2.17	1.78	2.48	1.93	2.82	1.81		X		X		X
Pity	6.53	2.17	6.42	2.41	6.16	2.29		X		X	4.30	<.05
Fear	3.89	2.10	2.68	1.56	3.86	1.94	4.45	<.05	3.54	<.05		X
Behavioral Response												
Helping	4.5	1.90	5.72	2.16	5.64	2.04	4.06	<.05		X		X
Coercion-Segregation	4.07	2.17	2.41	1.94	3.35	1.96	8.03	<.001	6.62	<.005		X
Social Distance	14.96	5.01	10.14	5.02	11.48	5.76	6.68	<.005		X		X

Table 4: Results of coding a conversational log whether the participant's attitude toward the main character's situation in the vignette had internal or external attributions

		V1	**V2**	**V3**	**V4**	**V5**	**V6**	**V7**
Attribution (External/Responses)	Web survey	11/28 (39%)	14/29 (48%)	13/29 (45%)	10/28 (36%)	11/28 (39%)	14/28 (50%)	13/29 (45%)
	FP chatbot	14/29 (48%)	19/30 (63%)	20/29 (69%)	19/29 (66%)	18/29 (62%)	19/30 (63%)	17/29 (72%)
	TP chatbot	13/30 (43%)	14/30 (47%)	14/29 (48%)	12/30 (40%)	14/30 (47%)	12/30 (40%)	12/29 (41%)

in the same group remarked that talking with Kenta offered them an intimate understanding of the struggles of coping with mental illness. For example, P42 (M) said, *"I have friends who have some similar conditions, and to be honest I did not know their feelings. In this study, I see Kenta as a tough and strong individual who needs to fight with himself. Kenta told me in detail about his inner thoughts and how he perceived things, and asked my opinions. I felt he was eager to find solutions to his mental problem."* These statements reflect a wider belief among the FP participants that Kenta was trying to improve his situation, even though it was at least partially out of his control.

Half the interviewees (n = 8) from the **TP chatbot group** attributed Kenta's difficulties to external factors. P81 (F), for instance, said: *"[H]e was suffering from stresses from external sources, which were not easy for him to deal with. I did not see someone around him actively offering him support."*

Nonetheless, the other half of the TP-group interviewees (n = 8) remarked that the vignettes never mentioned if Kenta had tried to reach out to any mental-healthcare providers, and that this made them doubt Kenta's motivation to get better. For example, P77 (M) said, *"I sympathized with Kenta's situation, but noticed the chatbot never said Kenta was going to see a mental-health specialist, which made me wonder if Kenta himself really wanted to deal with his mental illness."* Notably, this particular concern about Kenta's personal responsibility was more prevalent in the TP group than in the FP group (n = 2).

Only a quarter (n = 4) of the **web survey group** interviewees made comments about the role of external factors in Kenta's situation; and a large majority of them (n = 12) commented on the fact that Kenta did not seek help from a mental-health professional. Beyond simply questioning his motivation, nine participants from the web survey group made remarks explicitly blaming Kenta. For instance, P16 (M) said: *"Over these two weeks, I felt a bit disappointed in Kenta. Why was he so weak and negative about his life? He caused a lot of trouble to others. He should see a doctor! If he doesn't want to help himself, no one will help him."*

In sum, our analysis of the pre- and post-survey results indicates that, after the intervention, members of both chatbot groups felt Kenta was less personally responsible for his present situation than did members of the web survey group. The survey results did not identify different outcomes between the FP and TP Chatbot groups. However, our analysis of the daily response logs and interview responses revealed evidence that members of the FP chatbot group expressed greater empathy toward Kenta personally and tended to attribute Kenta's situation to external factors more often.

5.2.2 *Emotional Responses.*

Anger. This item measured participants' degree of anger toward Kenta. Across the three groups, the level of anger was low in both pre- and post-intervention surveys. In other words, the participants generally did not express anger at Kenta, irrespective of how much information about him they had received, or whether they were part of the web survey, TP chatbot, or FP chatbot group.

Pity. This item measured how much sympathy participants would feel for Kenta. There was a significant effect of time-point on the participants' feelings of pity toward Kenta, but no significant group membership or interactive effects. That is, there was no significant difference in pity scores across the three groups, and all groups' average pity scores increased in the post-intervention survey.

Fear. This item was to evaluate how participants perceived dangerous and threatened by Kenta. Fear of Kenta revealed significant main effects of group membership as well as an interaction effect, but there was no significant effect of time-point alone. Post-hoc analyses indicated that the FP group's level of fear toward Kenta was significantly lower ($p < .05$) than that of either the web survey group or the TP group in the post-intervention survey, whereas the three groups' respective levels of fear in pre-intervention survey were not significantly different. These results indicate that interacting with FP chatbot may have significantly lowered the participants' fear toward Kenta. Interview results help explain this finding. Seven FP interviewees reflected that their prior stigmatizing thoughts were due to unfamiliarity with people living with mental illness. As P33 (F) explained: *"I had concerns about the safety of being around a person with mental illness. However, Kenta chatted with me like a normal friend and shared his feelings over two weeks, and I realized my prior thoughts about people mental illness was extremely stigmatizing because of my unfamiliarity with them."* In other words, what might be called the 'insider perspective' derived from talking directly with Kenta seems to have contributed to a reduction in stigmatizing thoughts.

5.2.3 Behavioural Responses.

Helping. This item evaluated participants' willingness to support people with mental illness like Kenta. Again, we found a significant interaction effect, but no significant main effects, of time-point and group membership. Post-hoc analysis showed that, in the post-intervention survey data, the FP and TP chatbot groups' willingness to help was significantly higher than the web survey group's ($p < .01$). This suggested that, after the participants interacted with either chatbot over two weeks, they were more willing to support people with mental health problems than the members of web survey group were.

Coercion-Segregation. This item measured how participants perceived that Kenta should be hospitalized and kept away from his neighbors. There was a significant main effect of group membership, and a significant interactive effect of group membership and time-point, but no significant main effect of time-point alone. Post-hoc analysis showed that the interactive effect was significant because, in the post-intervention survey, the FP chatbot group had a significantly lower level of endorsing Coercion-Segregation than the web survey group did ($p < .05$), despite all three groups having had similar levels of such endorsement in the pre-intervention survey. Interestingly, the TP group did not exhibit a significant difference in this regard from either of the other two groups at the end of the intervention. These findings imply that the FP chatbot decreased its users' endorsement of sending people with mental illness away to institutions and isolating them from society, to a degree that the TP chatbot did not match.

Social Distance. This measures participants' behavioral intention to maintain social distance from people with mental illnesses and socialize with them. All three groups had very similar levels of SDS in the pre-intervention survey. There were no significant main effects of time-point or group membership on SDS score, but their interaction was significant. Post-hoc analyses showed that, at our second time-point, the participants in the FP chatbot group had significantly lower SDS scores than those in the web survey group. Although the FP group's mean SDS score was smaller than the TP group's at that same time-point, that difference was non-significant. We also found that the web survey group members' average post-intervention SDS was significantly higher than their pre-intervention one ($p < .05$); and that the FP group's post-intervention SDS was significantly lower than its pre-intervention one ($p < .05$). However, the TP group's pre- and post-intervention SDS scores were not significantly different from each other. These findings suggest that the FP chatbot significantly drove positive change in people's social-distancing views, but that a survey-based approach using the same vignette content had the opposite result, contrary to its purpose and our expectations.

Interview results offer an explanation for these findings about behavioural intention. Six web survey group participants explicitly stated that they would rather keep their distance from a person with mental illness because they saw Kenta's stories as representing potential challenges they might face when interacting with someone like him. P22 (F), for example, told us: *"I may stay away from people with mental issues because I'm afraid of accidentally saying the wrong thing, something that will make them more upset or angry. I don't have enough confidence that I would be able to help them."* This suggests that a lack of familiarity with people experiencing mental illness contributes to a lack of confidence about how to interact with them. Accordingly, the lack of a sense of familiarity with Kenta among members of the web survey group is likely to influence their desire for maintaining social distance.

5.3 Summary of Results

The aim of this work was to investigate the extent to which chatbots might facilitate social contact with, and change stigmatizing attitudes toward, a person with mental illness. We compared three conditions: A web survey group in which participants interacted with an online survey (replicating approaches from prior work); and two chatbot groups in which participants interacted with a chatbot that told stories about mental illness experiences from either a third-person or first-person perspective. We identified that chatbots were more effective at facilitating social contact, and in reducing participants' stigmatizing attitudes when compared to the web survey condition. Further, we identified significant differences between the two chatbot designs with regard to those outcomes.

RQ1 asked how interacting with each of the chatbots might facilitate social contact, compared to interacting with a web survey. We found that participants in the FP chatbot group engaged in more self-disclosure and were more likely to form vivid impressions of Kenta, when compared to members of the other groups. Further, even though participants knew the chatbot was not a real person, two-third of interviewees in the FP chatbot group described treating the chatbot as a social actor, for example by trying to be supportive, and one-third described feeling worried about Kenta's experiences over time. By contrast, two-thirds of TP chatbot group interviewees

expressed that the chatbot encouraged self-reflection, and one-third remarked that they could respond frankly to the chatbot without being judged. However, almost one-third were reluctant to disclose their feelings to the chatbot because they viewed it as 'gossiping' about Kenta, which they disliked.

RQ2 asked about the extent to which interacting with each of the chatbots might affect participants' stigmatizing attitudes toward people with mental illness, compared to interacting with a web survey. One of the most striking differences among the groups was that, on the post-intervention survey, participants who interacted with either chatbot reported that Kenta was less personally responsible for his situation than did participants in the web survey group. Additionally, during the two-week intervention, FP-group members were more likely than others to respond to questions by referring to Kenta's surrounding circumstances rather than attributing Kenta's situation to his own dispositional factors. This could be explained by CASA, since attributing an undesired situation to external factors is generally more socially acceptable than blaming an individual to their face [12].

Additionally, interacting with either chatbot (especially the FP chatbot) appears to have resulted in reduced stigmatizing emotional and behavioural responses, when compared to interacting with the web survey. Comparing the pre- and post-intervention surveys, fear decreased in the FP group, yet remained steady in the web survey and TP group. Regarding behavioural responses, participants who interacted with the FP chatbot reported reduced stigma across all three behavioural dimensions (helping, coercion-segregation, and social distance), TP chatbot group members had improved scores toward only one dimension (helping), and web survey group members reported an *increased* desire for social distance after the intervention, and no changes to other categories. These results clearly show that chatbots' role differences affected people's perceptions of the vignettes and their attitudes toward people with a mental illness; and that a chatbot that spoke in the first person had a greater effect on reducing measures of stigma than its third-person equivalent.

6 DISCUSSION

6.1 The Impact of Intervention Designs on Reducing Stigmatizing Attitudes

Previous studies [30, 66, 86] have suggested that facilitating positive social contact between people with mental illness and the general population is an effective anti-stigma intervention. Analysis of our interview and daily response logs provides evidence that the FP chatbot was able to stimulate social contact between participants and itself. To be specific, many interviewees in the FP group articulated a vivid image of the vignette character as a person, and regarded the vignettes as his self-disclosure. Members of the web survey and TP groups, on the other hand, did not appear to develop any such impression. We infer that this led to a further reciprocal effect [52, 54, 65]: the eliciting of participants' own self-disclosure in response to that of the FP chatbot. Past work has recognized mutual self-disclosure as a facilitator of social contact and even friendship [30, 74]. Our results support and extend that work.

We found that differences in our study's interaction design affected our participants' interpretations of Kenta's motivation to improve his mental health. In particular, many FP chatbot users described their conversation with the chatbot as representing the character's active efforts to seek advice for his situation. Some explicitly reflected that this perspective made them want to reduce what they now recognized as their own stigmatizing thoughts. Interviewees from the TP chatbot group, on the other hand, expressed sympathy for the fact that Kenta's challenges were largely out of his control, but simultaneously questioned whether he was taking his fair share of the responsibility for bettering his situation. Lastly, the interviewees from the web survey group were much more likely to blame Kenta for his situation, and many expressed a desire to keep their distance from people with mental illness.

Our study resonates with existing research [21, 28, 36, 50] that found storytelling to be an efficient technique for increasing social contact to reduce mental illness stigma. Furthermore, our findings partially echo those of prior studies[15, 52, 54] that suggested a chatbot sharing stories ostensibly about itself could trigger its users' compassion and reciprocity. The current study's results further indicate the importance of a chatbot's role in storytelling to change people's attitudes. That is, even though both the FP and TP groups interacted with their respective chatbots in the form of reciprocal message exchanges, and even though the vignettes' substantive content was the same, inter-chatbot differences led to divergent attitudes and perceptions. Moreover, four TP chatbot users perceived their chatbot as a person gossiping about the vignettes' character, a view that could easily have distorted these users' perceptions and experiences of interacting with it. These findings provide important insights into the effects of interaction styles on changing people's attitudes, and future designers should be aware of the importance of situating the chatbot's role.

Nevertheless, we believe more research is necessary to explore how a chatbot could affect attribution processes [19, 24]. For example, it would be useful to learn how to leverage social contact during interactions to activate more situational attribution, and how to structure interactions to reduce social distance, as doing so could lead to theory-based exploration of chatbot designs for reducing social stigma.

6.1.1 Guiding positive thinking. We found that experiencing the vignettes, irrespective of the delivery method, helped improve most participants' awareness and understanding of a putative person with mental-health problems. This echoes a previous study [47] in which participants described having a better awareness of their own mental health, and more positive perceptions of people with mental-health problems, after interacting with a bot that made social-media posts about having depression. The present research further found that, even though reading vignettes could help increase people's exposure to those with mental illnesses, exposure alone might not be enough to elicit positive reflection.

Specifically, even though all groups reported increased pity, and interviewees from all groups described reflecting about their opinions, some participants interpreted the vignettes from a negative perspective, and disproportionately so in the web survey and TP groups. That perspective included blaming the character for being weak, and believing it would be difficult to interact with him in real life. This may suggest that interpretations of the vignettes in the TP and web survey groups were affected by unconscious bias toward mental illness [59, 84, 85], which could have caused their members

to focus on negative interpretations that matched that bias. While our study did not include measurement of unconscious bias (e.g., an implicit-association test [84]), it would be beneficial if future research explored how participants' unconscious bias affected their interactions with and interpretations of a computer agent.

6.2 Design Implications

Chatbots have been used to support self-improvement such as reducing public speaking anxiety [89], guiding users through reflections about physical activity to increase their motivation [49], and helping new migrants navigate integrating into their host country [13]. Our results suggest that chatbots can also motivate outward-facing behavioural changes by reducing stigmatizing thoughts. One way to leverage this could be through online anti-stigma campaigns, where the low-cost of chatbots could facilitate their deployment on national or other large scales [82]. Another possibility is to use chatbots as part of a workplace or institutional training for workers who interact with people who might have mental illness. This could leverage the ability of chatbots to support reflection [49], by creating a low-stakes interaction through which users reflect on and challenge their biases.

To further boost the positive effects of social contact, future researchers could consider designing a first-person chatbot that uses strategies such as intergroup cooperation from prior social-contact interventions [30]. For example, Corrigan et al. [21, 23] suggested that encouraging different groups to work together on shared tasks could help foster common experience and break down stereotypes. Thus, chatbots could be designed that enlist users in a human-chatbot collaborative task. Working together with a chatbot character representing someone who struggles with mental illness or has other stigmatized experiences may be an effective way for users to overcome their stigmatizing beliefs.

Moreover, most existing information campaigns focus on content: i.e., what information is most effective in reducing stigma. The present work, in contrast, has demonstrated that chatbots not only can enhance these traditional information campaigns by delivering relevant stories through conversational interfaces (e.g., TP chatbot), but also provide opportunities to develop new, scalable intervention methods for reducing social stigma, such as the FP chatbot that stimulates a sense of social contact. Our finding that chatbots' roles can impact attitudes also has important implications for future research on the effects of interaction styles on changing attitudes about other stigmatized social groups. For example, prior research [62] has found that having prior social contact with someone who is gay or lesbian is a predictor for having more positive attitudes toward homosexuals. Thus, a first-person chatbot could be utilized as a potential strategy to increase exposure to discriminated individuals in order to facilitate positive attitudes and decrease social distance [6]. This may be particularly valuable for facilitating exposure to heavily stigmatized populations, such as LGBT people in regions where discrimination against homosexuality is prevalent and severe, since in-person social contact campaigns may introduce high risk to those stigmatized populations [27].

Although chatbots could contribute to attitude changes, we also identified some potential risks. For example, some of our participants questioned the Kenta character's motivation to improve his mental health, and some focused on the negative aspects of the vignettes. These results imply 1) that users may need more guidance about how to interpret such materials, and 2) that vignettes may need to include positive recovery stories [25] to leverage the advantages of chatbot-based storytelling technologies. Another potential solution would be to involve real or simulated healthcare providers in chatbot-based social contact [11, 53]. For example, the TP chatbot in our study could be situated as a mental health professional to introduce vignettes and guide the conversation to de-stigmatize users' thoughts. Nevertheless, future research in this direction will be required before any firm conclusions can be drawn.

Finally, although the FP chatbot was the most effective at reducing stigma, it appears to be less suitable for applications focused on gathering accurate responses from participants. This is because the majority of FP chatbot participants described tailoring their responses in order to support the chatbot, such as by self-censoring statements they felt could be offensive or unhelpful. TP chatbot users, in contrast, mentioned that they felt safe to respond honestly to the chatbot without offending Kenta; and the TP chatbot therefore might be better suited to collecting its users' thoughts and behaviors. This feature is important because understanding people's stigmatizing thoughts is not only vital to addressing them, but also allows researchers to design pathways to treatment that allow people to avoid stigma [9, 58].

6.3 Ethical Implications

This work explores chatbot designs for reducing users' mental illness stigma by facilitating chatbot-based storytelling and social contact. Reducing mental illness stigma should be carefully considered as it might lead to some ethical issues. We discuss potential issues by referring to our study design and research findings.

Difficulty in Representing Complexity of Human Relationships: Compared to real in-person social contact, chatbots have many advantages such as scalability and affordability. However, they make compromises regarding the complexity and diversity of the human experience. In real world anti-stigma campaign, it could be important to tell many different people's stories to further enhance understanding and elicit empathy [21, 36]. Our study's vignettes only talked about the same character, which might lead to a misunderstanding that all people will experience the same symptoms as Kenta, thus failing to represent the diversity of possible experiences with mental illness. The vignettes in this study were validated by psychiatric professionals, however, future implementations would benefit from working with people who have lived experience with mental illness [63]. Additionally, conceptualizations of mental health and illness are shaped by cultural forces such as colonialism, and therefore it is important to consider local cultural views and practices before introducing a mental illness stigma-related chatbot to the public and to recognize the western biases embedded in many tools for evaluating mental health [72].

Backfire Effects: Since the chatbot engages with a sensitive issue (mental illness), there is a potential of backfire effects [29, 36, 79], wherein participants could develop inaccurate perspectives about mental illness which exacerbate rather than reduce stigma. For example, without being presented with a positive recovery story, users might not believe that mental illness can be managed [36].

Therefore, it is important for chatbots to deliver both ups and downs in the recovery story of mental illness, in order to let users understand that mental illness is manageable. In the debriefing, we shared anti-stigma educational materials and other helpful information with study participants, to help them better understand mental illness stigma. However, even with debriefing, backfire effects could be significant. Particularly, users' previous experiences related to stigmatized conditions might strengthen backfire effects. For example, a study about reducing mental illness stigma through narrative framing techniques found that these techniques were effective for reducing stigma among people with no personal history of mental illness, but were not effective or could even increase stigmatizing beliefs among people with a personal history of mental illness [88]. Thus, care should be taken before introducing such an intervention to various users, in order to reduce the risks of exacerbating rather than reducing stigmatizing attitudes. Additionally, it may be worthwhile to provide a richer debriefing, such as including more detailed guidance about mental illness from medical professionals and more in-depth informational resources.

Risks of Targeting Users: Employing chatbots in the real-world setting should carefully target potential users. For example, letting people with clinically-identified depression use this tool might worsen their symptoms and deepen their self-stigma. Thus, to address unwanted situations, we recruited and screened participants to avoid the risks that the vignettes could cause distress to people who had such issues [79]. We provided the participants with emergency contact and had psychiatrists reviewed our study design.

6.4 Limitations

Among the seven vignettes we used, only one presented a particularly severe symptom: thoughts of self-harm. Although we consciously avoided presenting even more traumatic episodes, some participants said they found this vignette difficult to deal with. This, in turn, negatively affected the quantity of their self-disclosure. Future studies should therefore consider controlling vignettes' severity and evaluating its effect on people's mental-illness stigma.

Second, the participants were compensated for participating in the study and informed that their responses were being collected. Some might therefore have been affected by social-desirability bias. However, because participants in all groups were compensated at the same rate, the influence of the bias on inter-group differences would be mitigated. Besides, there is very little work about chatbots and stigma. We explored this topic by conducting a mixed-method study with a limited amount of participants. Future research could consider enlarging the sample size to provide strong evidence to support the chatbot design's effect on people's mental illness stigma.

Third, in the study, we designed the TP chatbot to introduce third-person vignettes. To avoid introducing new effects, we did not define the relationship between the TP chatbot and Kenta simply introducing participants to a "chatbot" that would share stories. However, some participants imagined a relationship between Kenta and the chatbot (such as that the chatbot was "gossiping" about Kenta), which shaped their perceptions. Future work could further introduce chatbots' personality and social roles into play, to investigate how these factors affect users' mental illness stigma.

Lastly, our study focused on mental-illness stigma, and its results thus may not map well onto stigma against other marginalized groups, such as the LGBTQ+ community [41] or people with physical disabilities [46]. Thus, further research is necessary before any firm conclusions are drawn about the potential benefits of chatbot interventions to other stigmatized populations.

7 CONCLUSION

This two-week study of how chatbots' roles could mitigate mental-illness stigma by facilitating social contact between members of the general public and a hypothetical character with a mental illness established that, when a chatbot spoke in the first person, more participants were encouraged to disclose their own experiences. In the same condition, they also perceived greater closeness with the character, because direct interaction encouraged reciprocity. Its findings also suggest that the chatbots' designs and interactions affected the participants' perceptions of our character's stories, and changed their attitudes towards him. Regarding quantitative measures of stigmatization, using a first-person chatbot had a more beneficial effect than either a third-person one or a non-chatbot (reading-based) web survey condition. These findings imply that a chatbot speaking 'in character' has the potential to stimulate social contact between participants and the chatbot, and promote positive change in people's attitudes toward others with mental illnesses, and enhance their willingness to help them. We hope this study will serve as a starting point for future research on the causal relationship between human-chatbot interaction and self-disclosure of personal experiences aimed at reducing social stigma.

ACKNOWLEDGMENTS

This work is partially supported by the NUS start-up grant. We thank all reviewers' comments and suggestions to help polish this paper.

REFERENCES

[1] Alaa A Abd-Alrazaq, Mohannad Alajlani, Nashva Ali, Kerstin Denecke, Bridgette M Bewick, and Mowafa Househ. 2021. Perceptions and opinions of patients about mental health chatbots: scoping review. *Journal of medical Internet research* 23, 1 (2021), e17828.

[2] Tahirah Abdullah and Tamara L Brown. 2020. Diagnostic labeling and mental illness stigma among Black Americans: An experimental vignette study. *Stigma and Health* 5, 1 (2020), 11.

[3] Nazanin Andalibi and Andrea Forte. 2018. Announcing pregnancy loss on Facebook: A decision-making framework for stigmatized disclosures on identified social network sites. In *Proceedings of the 2018 CHI conference on human factors in computing systems*. 1–14.

[4] 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.

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

[6] Anja Esther Baumann. 2007. Stigmatization, social distance and exclusion because of mental illness: the individual with mental illness as a 'stranger'. *International review of psychiatry* 19, 2 (2007), 131–135.

[7] 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.

[8] Timothy Bickmore, Daniel Schulman, and Langxuan Yin. 2009. Engagement vs. deceit: Virtual humans with human autobiographies. In *International workshop on intelligent virtual agents*. Springer, 6–19.

[9] Francois B Botha, Amanda L Shamblaw, and David JA Dozois. 2017. Reducing the stigma of depression among Asian students: A social norm approach. *Journal of Cross-Cultural Psychology* 48, 1 (2017), 113–131.

[10] Adolfo J Cangas, Noelia Navarro, José MA Parra, Juan J Ojeda, Diego Cangas, Jose A Piedra, and Jose Gallego. 2017. Stigma-stop: a serious game against the stigma toward mental health in educational settings. *Frontiers in psychology* 8 (2017), 1385.

[11] Bruna Sordi Carrara, Raquel Helena Hernandez Fernandes, Sireesha Jennifer Bobbili, and Carla Aparecida Arena Ventura. 2021. Health care providers and people with mental illness: An integrative review on anti-stigma interventions. *International Journal of Social Psychiatry* 67, 7 (2021), 840–853.

[12] Derek Chadee. 2022. *Theories in social psychology.* John Wiley & Sons.

[13] Zhifa Chen, Yichen Lu, Mika P Nieminen, and Andrés Lucero. 2020. Creating a chatbot for and with migrants: chatbot personality drives co-design activities. In *Proceedings of the 2020 ACM designing interactive systems conference.* 219–230.

[14] Wanda M Chernomas and Carla Shapiro. 2013. Stress, depression, and anxiety among undergraduate nursing students. *International journal of nursing education scholarship* 10, 1 (2013), 255–266.

[15] Qian Hui Chew and Kang Sim. 2020. Psychiatry teaching amongst medical undergraduates: stories that matter and mediators of better learning outcome. *Postgraduate Medicine* 132, 7 (2020), 590–594.

[16] Jacob Cohen. 2013. *Statistical power analysis for the behavioral sciences.* Routledge.

[17] Patrick Corrigan. 2003. Examining cues that signal mental illness stigma. *Journal of Social and Clinical Psychology* 22, 5 (2003), 467–476.

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

[19] Patrick Corrigan, Fred E Markowitz, Amy Watson, David Rowan, and Mary Ann Kubiak. 2003. An attribution model of public discrimination towards persons with mental illness. *Journal of health and Social Behavior* (2003), 162–179.

[20] Patrick W Corrigan. 2000. Mental health stigma as social attribution: Implications for research methods and attitude change. *Clinical psychology: science and practice* 7, 1 (2000), 48.

[21] Patrick W Corrigan and Kristin A Kosyluk. 2013. Erasing the stigma: Where science meets advocacy. *Basic and applied social psychology* 35, 1 (2013), 131–140.

[22] Patrick W Corrigan and Kristin A Kosyluk. 2014. Mental illness stigma: Types, constructs, and vehicles for change. (2014).

[23] Patrick W Corrigan and David L Penn. 1999. Lessons from social psychology on discrediting psychiatric stigma. *American psychologist* 54, 9 (1999), 765.

[24] Patrick W Corrigan, David Rowan, Amy Green, Robert Lundin, Philip River, Kyle Uphoff-Wasowski, Kurt White, and Mary Anne Kubiak. 2002. Challenging two mental illness stigmas: Personal responsibility and dangerousness. *Schizophrenia bulletin* 28, 2 (2002), 293–309.

[25] Patrick W Corrigan and Jenessa R Shapiro. 2010. Measuring the impact of programs that challenge the public stigma of mental illness. *Clinical psychology review* 30, 8 (2010), 907–922.

[26] Patrick W Corrigan and Amy C Watson. 2002. Understanding the impact of stigma on people with mental illness. *World psychiatry* 1, 1 (2002), 16.

[27] Yichao Cui, Naomi Yamashita, and Yi-Chieh Lee. 2022. " We Gather Together We Collaborate Together": Exploring the Challenges and Strategies of Chinese Lesbian and Bisexual Women's Online Communities on Weibo. *Proceedings of the ACM on Human-Computer Interaction* 6, CSCW2 (2022), 1–31.

[28] Tatiana Davidson, Angela Moreland, Brian E Bunnell, Jennifer Winkelmann, Jessica L Hamblen, and Kenneth J Ruggiero. 2021. Reducing stigma in mental health through digital storytelling. In *Research Anthology on Mental Health Stigma, Education, and Treatment.* IGI Global, 909–919.

[29] Keith S Dobson and Savannah Rose. 2022. "Myths and facts" campaigns are at best ineffective and may increase mental illness stigma. *Stigma and Health* 7, 1 (2022), 27.

[30] Sara Evans-Lacko, Jillian London, Sarah Japhet, Nicolas Rüsch, Clare Flach, Elizabeth Corker, Claire Henderson, and Graham Thornicroft. 2012. Mass social contact interventions and their effect on mental health related stigma and intended discrimination. *BMC public health* 12, 1 (2012), 1–8.

[31] Franz Faul, Edgar Erdfelder, Albert-Georg Lang, and Axel Buchner. 2007. G* Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior research methods* 39, 2 (2007), 175–191.

[32] Veronica D Feeg, Laura S Prager, Lois B Moylan, Kathleen Maurer Smith, and Meritta Cullinan. 2014. Predictors of mental illness stigma and attitudes among college students: using vignettes from a campus common reading program. *Issues in mental health nursing* 35, 9 (2014), 694–703.

[33] Emilio Ferrara, Onur Varol, Clayton Davis, Filippo Menczer, and Alessandro Flammini. 2016. The rise of social bots. *Commun. ACM* 59, 7 (2016), 96–104.

[34] Jessica L Feuston. 2019. Algorithms, Oppression, and Mental Illness on Social Media. In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems.* 1–5.

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

[36] Tiffany HC Fong and Winnie WS Mak. 2022. The effects of internet-based storytelling programs (Amazing Adventure Against Stigma) in reducing mental illness stigma with mediation by interactivity and stigma content: Randomized controlled trial. *Journal of medical Internet research* 24, 8 (2022), e37973.

[37] Siyu Gao and Siu-Man Ng. 2021. Reducing Stigma among College Students toward People with Schizophrenia: A Randomized Control Trial Grounded on Intergroup Contact Theory. *Schizophrenia Bulletin Open* (2021).

[38] Erving Goffman. 2009. *Stigma: Notes on the management of spoiled identity.* Simon and schuster.

[39] Annabelle Gourlay, Gerry Mshana, Isolde Birdthistle, Grace Bulugu, Basia Zaba, and Mark Urassa. 2014. Using vignettes in qualitative research to explore barriers and facilitating factors to the uptake of prevention of mother-to-child transmission services in rural Tanzania: a critical analysis. *BMC medical research methodology* 14, 1 (2014), 1–11.

[40] Kathleen M Griffiths, Yoshibumi Nakane, Helen Christensen, Kumiko Yoshioka, Anthony F Jorm, and Hideyuki Nakane. 2006. Stigma in response to mental disorders: a comparison of Australia and Japan. *BMC psychiatry* 6, 1 (2006), 1–12.

[41] Oliver L Haimson, Dykee Gorrell, Denny L Starks, and Zu Weinger. 2020. Designing trans technology: Defining challenges and envisioning community-centered solutions. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems.* 1–13.

[42] Peter Hegarty and Anne M Golden. 2008. Attributional beliefs about the controllability of stigmatized traits: antecedents or justifications of prejudice? 1. *Journal of Applied Social Psychology* 38, 4 (2008), 1023–1044.

[43] Sebastian Hobert and Florian Berens. 2019. Small talk conversations and the long-term use of chatbots in educational settings–experiences from a field study. In *International workshop on chatbot research and design.* Springer, 260–272.

[44] Daniel Holman. 2015. Exploring the relationship between social class, mental illness stigma and mental health literacy using British national survey data. *Health:* 19, 4 (2015), 413–429.

[45] Anthony F Jorm, Annemarie Wright, and Amy J Morgan. 2007. Beliefs about appropriate first aid for young people with mental disorders: findings from an Australian national survey of youth and parents. *Early Intervention in Psychiatry* 1, 1 (2007), 61–70.

[46] Bogoan Kim, Daehyoung Lee, Aehong Min, Seungwon Paik, Georgia Frey, Scott Bellini, Kyungsik Han, and Patrick C Shih. 2020. PuzzleWalk: A theory-driven iterative design inquiry of a mobile game for promoting physical activity in adults with autism spectrum disorder. *PLoS One* 15, 9 (2020), e0237966.

[47] Taewan Kim, Mintra Ruensuk, and Hwajung Hong. 2020. In helping a vulnerable bot, you help yourself: Designing a social bot as a care-receiver to promote mental health and reduce stigma. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems.* 1–13.

[48] Betty A Kitchener and Anthony F Jorm. 2002. Mental health first aid training for the public: evaluation of effects on knowledge, attitudes and helping behavior. *BMC psychiatry* 2, 1 (2002), 1–6.

[49] Rafal Kocielnik, Lillian Xiao, Daniel Avrahami, and Gary Hsieh. 2018. Reflection companion: A conversational system for engaging users in reflection on physical activity. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 2, 2 (2018), 70.

[50] Kristin Kosyluk, Jennifer Marshall, Kyaien Conner, Diana Rivera Macias, Sofia Macias, B Michelle Beekman, and Jonathan Her. 2021. Challenging the stigma of mental illness through creative storytelling: a randomized controlled trial of this is my brave. *Community Mental Health Journal* 57, 1 (2021), 144–152.

[51] Eun Hee Lee. 2004. Life stress and depressive symptoms among college students: Testing for moderating effects of coping style with structural equations. *Korean journal of health psychology* 9, 1 (2004), 25–48.

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

[53] Yi-Chieh Lee, Naomi Yamashita, and Yun Huang. 2020. Designing a Chatbot as a Mediator for Promoting Deep Self-Disclosure to a Real Mental Health Professional. *Proceedings of the ACM on Human-Computer Interaction* (2020).

[54] Yi-Chieh Lee, Naomi Yamashita, Yun Huang, and Wai Fu. 2020. "I Hear You, I Feel You": Encouraging Deep Self-disclosure through a Chatbot. *ACM CHI Conference on Human Factors in Computing Systems* (2020).

[55] Debra Lerner, David A Adler, William H Rogers, Leueen Lapitsky, Thomas McLaughlin, and John Reed. 2010. Work performance of employees with depression: the impact of work stressors. *American Journal of Health Promotion* 24, 3 (2010), 205–213.

[56] Bruce G Link, Francis T Cullen, James Frank, and John F Wozniak. 1987. The social rejection of former mental patients: Understanding why labels matter. *American journal of Sociology* 92, 6 (1987), 1461–1500.

[57] Antonina Luca, Maria Luca, and Carmela Calandra. 2013. Sleep disorders and depression: brief review of the literature, case report, and nonpharmacologic interventions for depression. *Clinical interventions in aging* 8 (2013), 1033.

[58] 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.

[59] Jasmine R Marcelin, Dawd S Siraj, Robert Victor, Shaila Kotadia, and Yvonne A Maldonado. 2019. The impact of unconscious bias in healthcare: how to recognize and mitigate it. *The Journal of infectious diseases* 220, Supplement_2 (2019), S62–S73.

[60] Steven Marwaha and Gill Livingston. 2002. Stigma, racism or choice. Why do depressed ethnic elders avoid psychiatrists? *Journal of Affective Disorders* 72, 3 (2002), 257–265.

[61] Robert Meadows, Christine Hine, and Eleanor Suddaby. 2020. Conversational agents and the making of mental health recovery. *Digital health* 6 (2020), 2055207620966170.

[62] Irem Metin-Orta and Selin Metin-Camgöz. 2020. Attachment style, openness to experience, and social contact as predictors of attitudes toward homosexuality. *Journal of Homosexuality* 67, 4 (2020), 528–553.

[63] David C Mohr, Ken R Weingardt, Madhu Reddy, and Stephen M Schueller. 2017. Three problems with current digital mental health research... and three things we can do about them. *Psychiatric services* 68, 5 (2017), 427–429.

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

[65] 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*. 72–78.

[66] Engineering National Academies of Sciences, Medicine, et al. 2016. *Ending discrimination against people with mental and substance use disorders: The evidence for stigma change*. National Academies Press.

[67] Elizabeth Nguyen, Timothy F Chen, and Claire L O'Reilly. 2012. Evaluating the impact of direct and indirect contact on the mental health stigma of pharmacy students. *Social psychiatry and psychiatric epidemiology* 47, 7 (2012), 1087–1098.

[68] Ross MG Norman, Richard M Sorrentino, Deborah Windell, and Rahul Manchanda. 2008. The role of perceived norms in the stigmatization of mental illness. *Social psychiatry and psychiatric epidemiology* 43, 11 (2008), 851–859.

[69] Ross MG Norman, Deborah Windell, and Rahul Manchanda. 2012. Examining differences in the stigma of depression and schizophrenia. *International Journal of Social Psychiatry* 58, 1 (2012), 69–78.

[70] Lorelli S Nowell, Jill M Norris, Deborah E White, and Nancy J Moules. 2017. Thematic analysis: Striving to meet the trustworthiness criteria. *International journal of qualitative methods* 16, 1 (2017), 1609406917733847.

[71] Matt O'Connor and Leanne Casey. 2015. The Mental Health Literacy Scale (MHLS): A new scale-based measure of mental health literacy. *Psychiatry research* 229, 1-2 (2015), 511–516.

[72] Sachin R Pendse, Daniel Nkemelu, Nicola J Bidwell, Sushrut Jadhav, Soumitra Pathare, Munmun De Choudhury, and Neha Kumar. 2022. From treatment to healing: Envisioning a decolonial digital mental health. In *CHI Conference on Human Factors in Computing Systems*. 1–23.

[73] Jacqueline B Persons, Joan Davidson, Michael A Tompkins, and E Thomas Dowd. 2001. Essential components of cognitive-behavior therapy for depression.

[74] Thomas F Pettigrew. 1998. Intergroup contact theory. *Annual review of psychology* 49, 1 (1998), 65–85.

[75] 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.

[76] Safa Quadri, Naveena Karusala, and Rosa I Arriaga. 2018. # AutismAwareness: A Longitudinal Study to Characterize Tweeting Patterns for Indian and US Users. In *Proceedings of the 9th Indian Conference on Human Computer Interaction*. 11–19.

[77] 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.

[78] Nicola J Reavley, Anthony F Jorm, and Amy J Morgan. 2017. Discrimination and positive treatment toward people with mental health problems in workplace and education settings: Findings from an Australian National Survey. *Stigma and Health* 2, 4 (2017), 254.

[79] Laura Weiss Roberts. 2002. Ethics and mental illness research. *Psychiatric Clinics* 25, 3 (2002), A525–A545.

[80] Matías E Rodríguez-Rivas, Adolfo J Cangas, Laura A Cariola, Jorge J Varela, and Sara Valdebenito. 2022. Innovative Technology–Based Interventions to Reduce Stigma Toward People With Mental Illness: Systematic Review and Meta-analysis. *JMIR Serious Games* 10, 2 (2022), e35099.

[81] Nicolas Rüsch, Matthias C Angermeyer, and Patrick W Corrigan. 2005. Mental illness stigma: Concepts, consequences, and initiatives to reduce stigma. *European psychiatry* 20, 8 (2005), 529–539.

[82] Joel Sebastian and Deborah Richards. 2017. Changing stigmatizing attitudes to mental health via education and contact with embodied conversational agents. *Computers in Human Behavior* 73 (2017), 479–488.

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

[84] Bethany A Teachman, Joel G Wilson, and Irina Komarovskaya. 2006. Implicit and explicit stigma of mental illness in diagnosed and healthy samples. *Journal of Social and Clinical Psychology* 25, 1 (2006), 75–95.

[85] Joan Trujols. 2015. The brain disease model of addiction: challenging or reinforcing stigma? *The Lancet Psychiatry* 2, 4 (2015), 292.

[86] Rhiannon N Turner, Miles Hewstone, and Alberto Voci. 2007. Reducing explicit and implicit outgroup prejudice via direct and extended contact: The mediating role of self-disclosure and intergroup anxiety. *Journal of personality and social psychology* 93, 3 (2007), 369.

[87] Abdul Kawsar Tushar, Iffat Jahan Antara, Dipto Das, Priyank Chandra, Tanjir Rashid Soron, Md Munirul Haque, Sheikh Iqbal Ahamed, and Syed Ishtiaque Ahmed. 2020. We Need More Power to Stand Up: Designing to Combat Stigmatization of the Caregivers of Children with Autism in Urban Bangladesh. In *Proceedings of the 2020 International Conference on Information and Communication Technologies and Development*. 1–12.

[88] Bart Vyncke and Baldwin Van Gorp. 2018. An experimental examination of the effectiveness of framing strategies to reduce mental health stigma. *Journal of Health Communication* 23, 10-11 (2018), 899–908.

[89] Jinping Wang, Hyun Yang, Ruosi Shao, Saeed Abdullah, and S Shyam Sundar. 2020. Alexa as Coach: Leveraging Smart Speakers to Build Social Agents that Reduce Public Speaking Anxiety. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. 1–13.