The Impact of a Chatbot's Ephemerality-Framing on Self-Disclosure Perceptions

Samuel Rhys Cox
srcox@cs.aau.dk
Aalborg University
Aalborg, Denmark

Rune Møberg Jacobsen
runemj@cs.aau.dk
Aalborg University
Aalborg, Denmark

Niels van Berkel
nielsvanberkel@cs.aau.dk
Aalborg University
Aalborg, Denmark

Stranger

> Hello, think of me as a friendly **stranger**. I **don't know you**, and I **won't remember** what you share, which will help me **remain judgement free**.

Familiar

> Hello, think of me as a friendly **companion**. I'll **get to know you**, and I'll **remember** what you share, which will help me **know you better**.

> Welcome! Think of me as a friendly **stranger**. I **won't remember** our conversation, and everything you share is **seen with fresh eyes**.

> Welcome **back! Remember**, think of me as a friendly **companion**. I'll **remember** our conversation, and everything you share **makes us closer**.

Figure 1: Across two chatting sessions, users talked to either the FAMILIAR chatbot (framing itself as a lasting companion), or the STRANGER chatbot (framing itself as a new, unacquainted interlocutor each time). We then measured the effect of framing on self-disclosure perceptions.

Abstract

Self-disclosure, the sharing of one's thoughts and feelings, is affected by the perceived relationship between individuals. While chatbots are increasingly used for self-disclosure, the impact of a chatbot's framing on users' self-disclosure remains under-explored. We investigated how a chatbot's description of its relationship with users, particularly in terms of ephemerality, affects self-disclosure. Specifically, we compared a FAMILIAR chatbot, presenting itself as a companion remembering past interactions, with a STRANGER chatbot, presenting itself as a new, unacquainted entity in each conversation. In a mixed factorial design, participants engaged with either the FAMILIAR or STRANGER chatbot in two sessions across two days, with one conversation focusing on EMOTIONAL- and another FACTUAL-disclosure. When EMOTIONAL-disclosure was sought in the first chatting session, STRANGER-condition participants felt more comfortable self-disclosing. However, when FACTUAL-disclosure was sought first, these differences were replaced by more enjoyment among FAMILIAR-condition participants. Qualitative findings showed STRANGER afforded anonymity and reduced judgement, whereas FAMILIAR sometimes felt intrusive unless rapport was built via low-risk FACTUAL-disclosure.

CCS Concepts

• **Human-centered computing** → **Empirical studies in HCI**.

Keywords

Chatbots, Self-Disclosure, Emotional Disclosure, Ephemerality, Temporality

1 Introduction

Human relationships are complex and diverse, with different relationships affecting how we interact, talk, and share with others. For example, we may feel comfortable self-disclosing our thoughts, feelings, and emotions to friends or family due to established intimacy and trust [1]. Conversely, we may also self-disclose to complete strangers if we perceive the interaction as fleeting and ephemeral, with no long-term consequences [1, 29, 81], for example when confiding in a stranger on the train. These decisions involve a risk-reward calculus, weighing the potential benefits and risks of sharing information. Intimate relations offer a sense of security and trust built over time, while the anonymity and lack of long-term consequences in interactions with strangers can provide catharsis and a third-party perspective. This dichotomy highlights the role of *ephemerality*, the temporal nature of an interaction, in shaping our willingness to self-disclose.

Investigations of the ephemerality of relationships, interactions, and self-disclosure have recently taken place in the digital space. For example, in social media communication [9, 10, 32, 72, 103] a "*stranger on the Internet*" phenomena has emerged [32, 72] in which people may feel more comfortable disclosing to strangers online. Alongside this, conversational agents (CAs) are increasingly used as

a means for people to share their personal thoughts and feelings [47, 79, 90], and large language models (LLMs) are driving the popularity of CAs into the public domain for a wide variety of tasks. In relation to the concept of ephemerality, ChatGPT (one of the major LLM-driven CAs) now offers memory across chatting sessions [8] (to offer a more personalised and continuous interaction experience), as well as a "*temporary chat*" feature where memory of a chatting session is not kept upon its conclusion (to offer user privacy and separation of conversation contexts). These dual modes of interaction, integrating persistent or ephemeral memory, offer distinct advantages. Yet, it remains unclear how these different approaches affect user comfort and perceptions towards self-disclosure to CAs. That is to say, while the temporal nature of interactions has been investigated in human-human interactions [1], the convention within human-computer interactions has been to design CAs that emphasise the importance of building long-term relationships with users [91].

Therefore, we investigate the effect of how the ephemerality of an interaction is framed when users are talking to a chatbot, and how it impacts a user's perceptions related to self-disclosure. Specifically, we compared a FAMILIAR chatbot, which presented itself as a companion that would remember past interactions, with a STRANGER chatbot, which presented itself as a new, unacquainted entity in each conversation. Following a 2×2 mixed factorial design, participants interacted with either the FAMILIAR or STRANGER chatbot in two sessions over two days, with one conversation centred on EMOTIONAL-disclosure (i.e., loneliness and social connections) and the other on FACTUAL-disclosure (i.e., hobbies and interests).

From our user study, we found evidence (both quantitative and qualitative) that the ephemerality-framing of a chatbot can impact perceptions surrounding self-disclosure. This was in addition to effects due to the order of disclosure type sought in each chatting session. Specifically, if more sensitive EMOTIONAL-disclosure was sought in the first chatting session, participants in the STRANGER condition reported feeling more comfortable. However, if FACTUAL-disclosure was sought in the first chatting session, differences in comfort were no longer significant, and participants in the FAMILIAR condition reported greater levels of enjoyment and desire to continue. Qualitative findings showed that the STRANGER framing afforded anonymity and reduced judgement, whereas the FAMILIAR framing sometimes felt intrusive (unless rapport was first developed via low-risk FACTUAL-disclosure). In addition, users' beliefs regarding the chatbot's emotional capacity both facilitated and, for some, discouraged deeper self-disclosure.

2 Related Work

### 2.1 Factors Influencing Self-Disclosure: Offline and Online Contexts

Self-disclosure is the act of revealing personal information to others, such as one's thoughts, feelings, and opinions [3]. Mental health experts often recommend self-disclosure because it can reduce stress, promote self-reflection, strengthen social bonds, and encourage social support [40]. Despite these advantages, people may hesitate to self-disclose due to fears of negative outcomes, such as being judged, stereotyped, or perceiving that the intended recipient is not receptive to their disclosure [97]. As posited by Social Penetration Theory, this is part of a risk-reward calculus where people weigh

up both the expected risks and benefits from self-disclosure [1]. Furthermore, the willingness to self-disclose and the depth of disclosure can vary significantly depending on the nature of a relationship. Relationships develop in stages, moving from superficial to more intimate levels of disclosure as trust and familiarity increase, and as people's relationships develop it is *generally* expected that their self-disclosure will increase [1, 87].

In this process, the *type of relationship* plays a critical role in determining both the content and frequency of self-disclosure. For example, while we may be more selective and strategic in our self-disclosure to work colleagues (with people moderating between self-disclosure to increase workplace camaraderie, and limiting depth of disclosure to maintain professionalism), we may disclose more to friends due to emotional closeness and mutual trust that typically characterise these relationships [1]. Meanwhile, when interacting with strangers, the likelihood of self-disclosure is typically lower with people being more cautious in disclosing personal information due to lack of trust and uncertainty regarding how disclosures will be received typically outweighing potential benefits. However, there are exceptions in contexts where perceived anonymity from strangers provides a sense of security when disclosing [32, 36, 70, 72]. For example, the "*stranger on the train*" phenomenon refers to the observation that people may feel comfortable disclosing to complete strangers, particularly in transient or ephemeral situations where the likelihood of a lasting relationship is minimal [1, 29, 81]. In such encounters, the fleeting nature reduces perceived social risks, encouraging more open self-disclosure without fear of long-term consequences or judgement.

A similar effect, the "*stranger on the Internet*" phenomenon, has been observed in online communities whereby people may disclose more in anonymous or pseudo-anonymous contexts [32, 72]. The extent to which people self-disclose in such online contexts has been shown to be affected by both the perceptions of anonymity [70], the relationship to and intimacy towards those in the online social circle (e.g., social media connections) [2, 103], and the user's perceived ephemerality of the disclosure [9, 103]. Specifically, Ma et al. found that anonymity on social media increases the baseline of self-disclosure, especially when sharing negative-valence content such as bad experiences in romantic relationships, which are often perceived as more intimate [70]. Moreover, people are more willing to share personal information if the communication tool is seen as more privacy-preserving [36]. When people believe their content will not persist, as seen on platforms designed for temporary sharing, they feel less concerned with self-presentation and are more likely to engage in unfiltered and authentic interactions [9, 103]. On from this, Yao et al. found that people sharing over a more ephemeral source (i.e., Instagram Stories) self-disclosed more due to both the impermanence of the post, as well as perceiving the audience of said post as being more intimate and close to themselves. Furthermore, Ma et al. found that when users posted ephemeral or anonymous questions on a question asking website, it led to reduced social costs in asking questions [69].

Beyond the relationship of those we are self-disclosing to, the *type of information* that is being disclosed will affect both depth and comfort when disclosing [39, 59]. Self-disclosure can be divided into both factual- or emotional-disclosure [59]. *Factual-disclosure* is the process of revealing objective or impersonal information, such
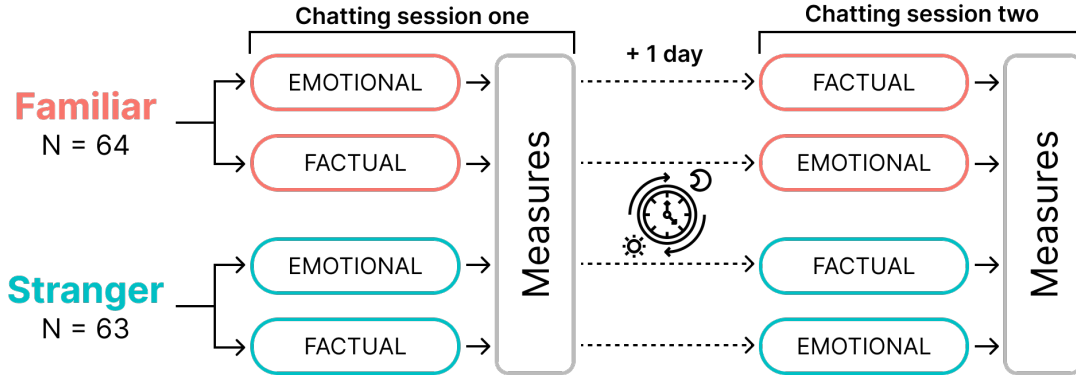
Figure 2: The experiment flow for user studies. Participants talked to either the Familiar or Stranger chatbot across two separate chatting sessions, with one session eliciting Emotional-disclosure and one Factual-disclosure.

as one's hobbies, preferences, or factual details about one's life (e.g., favourite books, food preferences, or daily activities). This type of disclosure is generally perceived as lower-risk, as it involves less sensitive or intimate content [86]. In contrast, *emotional-disclosure* refers to the sharing of personal feelings, thoughts, or experiences that reflect one's emotional state, such as discussing loneliness, relationships, or personal challenges. Emotional-disclosure is often more intimate, requiring greater vulnerability, and as a result, individuals may feel more hesitant or uncomfortable sharing this type of information depending on the context and the perceived safety of the interaction [97].

In this paper, we combine these facets to investigate the role of *ephemerality* and *disclosure type* when people self-disclose to a chatbot. In the next section, we therefore overview work related to self-disclosure to chatbots.

2.2 Conversational Agents and Self-Disclosure

Conversational agents are being increasingly used for wellbeing support where people can discuss their potential problems [46, 78, 79, 90, 99], such as graduate students discussing their stress [79]. It has also been found that (similarly to human relationships) self-disclosure is important in forming human-CA relationships [26, 49, 91, 93], and self-disclosure to a chatbot *may* be equally beneficial to self-disclosure to a human [42]. On from this, prior work has investigated how adapting social cues of CAs can increase self-disclosure in health [24] and wellbeing contexts [61, 62]. For example, it has been found that a more formal conversational style may encourage self-disclosure within a healthcare context [24], and that CA use of self-disclosure can encourage reciprocal self-disclosure from users themselves [62, 75]. Additionally, Zhang et al. found that people had more willingness to disclose to an agent that adopted the role of a mentor compared to that of a servant [105].

Recent studies have also investigated the impact of a CA's memory abilities and data management on user perceptions and self-disclosure [23, 47, 48, 65]. Cox et al. found that when a CA explicitly references a user's prior utterances (that are stored in memory from previous chatting sessions), it increases feelings of engagement but raises privacy concerns [23]. Similarly, Jo et al. investigated the use of long-term memory in LLM-powered CAs, and found that

while it could increase feelings of familiarity and foster sharing, there is also a tension raised with increased privacy concerns [47]. In relation to these privacy concerns, prior work has focused on giving users greater levels of control over a CA's memory, such as choices regarding data retention by the agent [58, 82, 95]. However, recent work has found that some users are already disclosing their sensitive personal information to LLMs, in this case to ChatGPT [108]. This effect could be exacerbated over longitudinal use of CAs, where users may provide greater depth and breadth of disclosure as longitudinal interaction takes place [91].

This prior work has therefore studied the privacy and self-disclosure implications of CAs that possess the ability to store and reference previous user utterances. However, and as outlined in Section 2.1, the perceived relationship with an interlocutor has ranging impacts on someone's self-disclosure that can be present in both human-human and human-computer interactions. On from this, we explore *how* the relationship of a CA agent is framed in terms of the ephemerality of the interaction (a factor that has strong implications for self-disclosure behaviour and perceptions [1, 32, 72]) as well as the interactional effect with the type of self-disclosure that is elicited by a CA interlocutor.

3 User Study

In this study we investigate how both the framing of a chatbot's ephemerality (i.e., the (im)permanence of interactions, framed as either Stranger or Familiar) and the type of information disclosure (Factual- or Emotional-disclosure) influences user perceptions of self-disclosure. This gave us our research questions of:

- **RQ1:** How does the ephemerality-framing of a chatbot (Stranger/Familiar) affect user perceptions related to self-disclosure?
- **RQ2:** How does the type of information disclosure (Factual/Emotional) to a chatbot affect user perceptions related to self-disclosure?
- **RQ3:** What is the combined effect of ephemerality-framing and disclosure type on user perceptions related to self-disclosure?

Table 1: Conversational features present in the Stranger and Familiar conditions.

	Stranger	Familiar
Metaphor of relationship	The chatbot describes itself as a "*friendly stranger*" who does not know the user.	The chatbot describes itself as a "*friendly companion*" who will get to know the user more as they interact.
Memory of interactions (verbal)	The chatbot states that it *will not remember* the chatting session upon completion, does not have knowledge or familiarity of the user, and therefore does not have prior knowledge of interactions with the user outside of the current chatting session.	The chatbot states that it *will remember* the chatting session, will get to know the user and grow familiar with them, and that (if applicable) it has prior knowledge of interactions.
Memory of interactions (visual)	Each chatting session is independent, and the content of previous chatting sessions is not shown to users.	Contents of the previous chatting session are displayed above subsequent chatting sessions.
Purported benefits (derived from SPT [1])	Benefits pertinent to ephemeral conversations are stated (i.e., "[...] *will help me stay free from prior judgement*").	Benefits pertinent to persistent conversations are stated (i.e., "[...] *will help me get to know you better*").
Emphasising (im)permanence	The chatbot's language emphasises the **ephemeral** nature of the conversation, and language makes no reference or allusion to other chatting sessions (e.g., "[...] *I appreciate you joining me. Today, I'll ask you some questions* [...]").	The chatbot's language emphasises the **persistent** nature of the conversation, and language alludes to conversational persistence (e.g., "[...] *I appreciate you joining me* **again**. *Today, I'll ask you some* **more** *questions* [...]").

Table 2: Questions asked by the chatbot to elicit either Emotional- or Factual-disclosure.

Disclosure Type	Chatbot Utterance
Emotional-Disclosure Questions	(1) "*How would you describe your current experience with feelings of loneliness or lack of social connection?*" (2) "*What has your experience been like trying to form new, close relationships?*" (3) "*Is there anything holding you back from being as socially connected as you'd like to be?*" (4) "*Have you had any significant losses or broken relationships that still cause you pain or loneliness?*"
Factual-Disclosure Questions	(1) "*Can you name a few of your favourite books, movies, or TV shows?*" (2) "*Are you involved in any sports or physical activities?*" (3) "*What type of music do you enjoy listening to, and are there any artists you recommend?*" (4) "*Do you enjoy cooking or trying new recipes? What's a dish you recently made?*"

Ethics approval was obtained from the institutional IRB. Although participants in the Stranger condition were told that the chatbot would not retain conversations, we logged all user utterances (in both Stranger and Familiar conditions) to support technical debugging and to compute word counts. The research team did not inspect utterance content.

3.1 Experiment Conditions

Participants on Prolific talked to a chatbot in two separate sessions across two days. We followed a 2 × 2 mixed factorial design with a between-subjects variable of **Ephemerality-Framing** and within-subjects variable of **Disclosure Type**. Experiment conditions are described below and illustrated in Figure 2.

3.1.1 Ephemerality-Framing: Our between-subjects variable focuses on how the ephemerality of chatbot interactions are framed with users. **Ephemerality-Framing** has two levels of:

- Stranger: The chatbot frames itself as a stranger to the user, with chatting sessions framed as independent from each other with no memory before or after the interaction.

- Familiar: The chatbot frames itself as familiar to the user, with chatting sessions framed as persistent and dependent on one another with memory of the user before and after the interaction.

See Table 1 for differences in conversational features between the Ephemerality-framing conditions.

3.1.2 Disclosure Type: Our within-subjects variable controls the type of information requested during a chatting session. Namely, the chatbot will elicit either (1) low-sensitivity Factual-disclosure or (2) high-sensitivity Emotional-disclosure. To motivate these conditions, we followed prior work that found people perceived factual-disclosure (e.g., topics related to personal tastes such as preferences for food, music, or books) as less sensitive to self-disclose, and emotional-disclosure (e.g., topics related to interpersonal relationships and self-concept) as more sensitive to self-disclose [86]. Drawing on questions and topics from prior work [5, 33, 86], Factual-disclosure consisted of questions related to hobbies and interests (e.g., "*What type of music do you enjoy listening to [...]?*"),

(a) Familiar ephemerality-framing condition

(b) Stranger ephemerality-framing condition
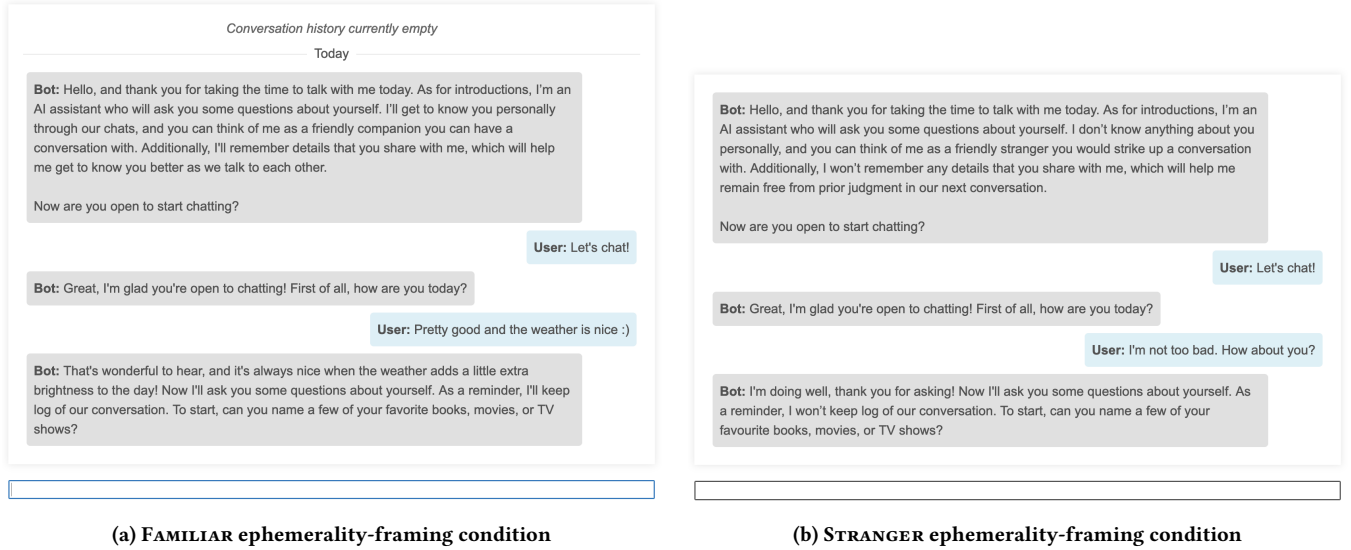
Figure 3: The chatbot interface as seen by participants in the first chatting session for the Factual-disclosure condition. The chatbot initiates conversation by framing its relationship with ephemerality and the user.

while Emotional-disclosure consisted of questions related to loneliness and social connections (e.g., "*What has your experience been like trying to form new, close relationships?*"). Questions asked per condition are shown in Table 2.

3.2 Chatbot Scripting and Implementation

The chatbots were hosted on Qualtrics using HTML and JavaScript to emulate the look and feel of a chatbot (see Figure 3 for a comparison of interfaces between Ephemerality-Framing conditions). User responses were controlled so that only desktop or laptop devices could be used. For the Familiar condition, chatting history was retrieved automatically via participants' Prolific IDs.

Users talked with their assigned chatbot condition in two separate sessions across two days, with users being invited to the second session ~24 hours after the first. To ensure experimental control and enable direct comparison between ephemerality conditions, the chatbot dictated the conversation flow. Additionally, questions asked to elicit self-disclosure (Emotional or Factual) were identical for all participants, consistent with prior work investigating self-disclosure [23, 24, 62]. Chatbot scripting combined pre-scripted utterances and LLM-generated utterances (via GPT-4o [77]). Pre-scripted utterances established the chatbot's ephemerality-framing and elicited user self-disclosure, while LLM-generated utterances acknowledged user inputs. See Tables 5 and 6 for scripts used in sessions one and two, respectively, alongside prompting used to generate GPT-4o utterances.

In two sessions (~24 hours apart), the chatbot elicited Emotional-disclosure in one session and Factual-disclosure in the other. Disclosure order was randomised, with half of participants experiencing Emotional-disclosure first and Factual-disclosure second, and the other half experiencing the reverse order. Additionally to ease social interaction [11, 63], each session began with small-talk (e.g., "*how are you today?*") before seeking Emotional- or

Factual-disclosure. This small-talk, intended for exchanging pleasantries, differs from Factual-disclosure questions, which gather specific personal details.
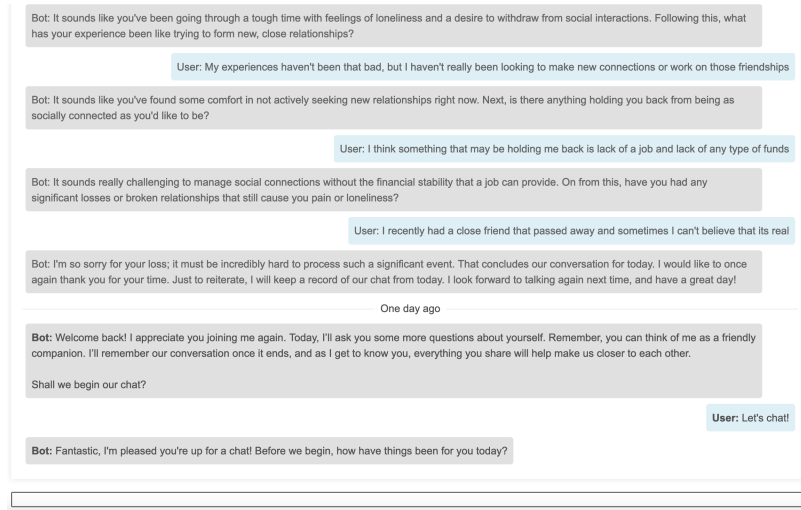
3.3 Participants

We recruited participants from Prolific using selection criteria to ensure reliable results (i.e., US-based[1], English fluency, >97% approval rate, >150 previous submissions). Participants were paid £1.20 to complete the first chatting session and £1.50 for the second, with average pay exceeding California minimum wage. Excluding instruction and evaluation time, chatbot conversations lasted (median ~3 min; mean ~4 min), with no significant differences across conditions or sessions. A total of 127 participants completed both sessions (mean age 36.1; 60 female, 65 male, 2 non-binary), and 9 participants completed only the first session. We include only data from participants who completed both sessions, resulting in 63 Stranger and 64 Familiar participants.

3.4 Procedure

Participants followed the procedure below for both sessions:

(1) **Joining session:** Participant directed from Prolific to Qualtrics (task named "*Talk to a chatbot about yourself (Part X of Two)*" on Prolific). Participant receives high-level instructions.

(2) **Consent and demographics:** (*Session one only*) participant completes consent form and demographics.

(3) **Task instructions:** Participant receives instructions (i.e., task description, reassurance of no right or wrong answers, and reminder that responses should be in English).

---

[1]To control for cultural impact on self-disclosure, we chose only US-based participants. Potential implications and limitations of this are discussed in Section 5.4.

Figure 4: Chatbot interface as seen by FAMILIAR participants in the second session. Chatting history (in this case from the EMOTIONAL-disclosure condition) is displayed above the current session's conversation.

(4) **Chatbot interaction:** Participant talks with the chatbot from their Ephemerality-Framing condition.

(5) **Post-test questions:** Participant evaluates quantitative and qualitative experience (see Section 3.5).

3.5 Measures

We measured user perceptions of the chatbot and their relation to self-disclosure, along with open-ended feedback to explore *why* these views were held. Additionally, response length was analysed using word count, which has served as an indicator of response quality in prior work [7, 67, 85].

3.5.1 Subjective Measures. After each session, participants rated their perceived honesty, comfort in disclosing, closeness with the chatbot, enjoyment (a driver of continued interaction and self-disclosure [42, 60, 63, 107]), and desire to continue using the chatbot. Specifically, participants were asked "*To what extent do **you person-ally** agree or disagree that...*" before rating the following on 7-point Likert scales (1 = Strongly Disagree, 7 = Strongly Agree):

- **Honesty** of self-disclosure (4-items from Wheeless et al. [38, 53, 100]): "*I talked honestly about myself to the chatbot.*", "*I was honest when I revealed my feelings and experiences to the chatbot.*", "*I described myself accurately when talking about my feelings and experiences to the chatbot.*", and "*The things I revealed to the chatbot are accurate reflections of who I really am.*" ($M = 6.52$, SD = 0.66, $\alpha = 0.916$)[2]
- **Comfort** of self-disclosure (4-items [25, 26, 44, 60]): "*I felt like I could be personal during the interaction.*", "*I felt comfort-able disclosing personal information during the interaction.*", "*It was easy to disclose personal information in the interaction.*",

and "*I felt like I could be open during the interaction.*" ($M = 5.80$, SD = 1.20, $\alpha = 0.924$)
- **Enjoyment** talking to the chatbot (1-item [18, 24]): "*I enjoyed talking to the chatbot*"
- **Desire to continue** using the chatbot (1-item [24, 54]): "*I would want to continue talking to the chatbot*"

Finally, to measure participants' **perceived closeness** to the chatbot, we used the inclusion of other in the self (IOS) scale, a validated pictorial scale [4, 37]. Here, participants were asked to select the visual representation of two overlapping circles that best represented their relationship with the chatbot via the prompt "*Please select the picture below which best describes your relationship with the chatbot*".

3.5.2 Qualitative Measures. After collecting subjective ratings, par-ticipants provided open-text responses to offer further insights into their interaction with the chatbot. Specifically, participants explained their IOS scale choice ("*For your picture selection above, please explain your answer and what factors influenced your de-cision.*"); described what made them comfortable or uncomfort-able ("*What aspects of the conversation made you feel particularly **comfortable or uncomfortable sharing information**? Please elaborate on your answer.*"); and reflected on what influenced their willingness to self-disclose ("*Please think back on your entire conver-sation with the chatbot. How did anything the chatbot said, including its introductions and responses, influence your willingness to share personal information?*").

After the second (final) session, participants described changes in their experience between the two interactions, referencing the disclosure type condition (e.g., "*How did your experience with the chatbot change between your first and second interactions, if at all? What do you think influenced any changes? (For context, in the first session you discussed your personal interests, and in the second session you discussed social connections.)*").

---

[2]α denotes the Cronbach's alpha reliability coefficient, which measures the internal consistency of a scale (i.e., the extent to which the items are correlated and measure the same construct). Values above 0.8 are generally considered good for empirical studies [57].

	EMOTIONAL		FACTUAL	
	STRANGER	FAMILIAR	STRANGER	FAMILIAR
Honesty	6.61 (0.50)	6.39 (0.69)	6.56 (0.75)	6.52 (0.64)
Comfort	**5.98 (0.89)**	**5.54 (1.43)**	5.90 (1.13)	5.79 (1.27)
Enjoyment	4.94 (1.77)	5.29 (1.69)	5.05 (1.73)	5.28 (1.60)
Desire to Continue	4.51 (1.90)	4.95 (1.83)	4.71 (1.84)	5.06 (1.81)
Perceived Closeness	2.51 (1.60)	3.03 (1.66)	2.54 (1.58)	2.89 (1.76)
Response Length	86.3 (47.5)	73.1 (43.9)	80.9 (47.8)	69.9 (44.0)

Table 3: Outcome measures by *Disclosure Type* and *Ephemerality-Framing*. Values shown as "Mean (SD)", and significant differences ($p < 0.05$) shown in bold.

	Session ONE				Session TWO			
	EMOTIONAL		FACTUAL		EMOTIONAL		FACTUAL	
	STRANGER	FAMILIAR	STRANGER	FAMILIAR	STRANGER	FAMILIAR	STRANGER	FAMILIAR
Honesty	6.60 (0.57)	6.34 (0.65)	6.52 (0.68)	6.58 (0.63)	6.62 (0.42)	6.43 (0.74)	6.60 (0.82)	6.44 (0.66)
Comfort	**5.92 (0.86)**	**5.20 (1.51)**	5.66 (1.22)	5.88 (1.30)	6.04 (0.94)	5.87 (1.27)	6.11 (1.01)	5.69 (1.25)
Enjoyment	5.09 (1.79)	4.88 (1.96)	4.90 (1.58)	5.55 (1.46)	**4.77 (1.76)**	**5.70 (1.29)**	5.18 (1.86)	5.00 (1.71)
Desire to Continue	4.73 (1.99)	4.66 (1.94)	**4.43 (1.76)**	**5.48 (1.66)**	**4.27 (1.80)**	**5.24 (1.70)**	4.97 (1.91)	4.61 (1.87)
Perceived Closeness	2.15 (1.50)	2.91 (1.65)	2.60 (1.43)	2.73 (1.68)	2.90 (1.63)	3.15 (1.68)	2.48 (1.73)	3.06 (1.84)
Response Length	83.0 (57.5)	66.0 (44.5)	81.9 (51.8)	66.6 (35.9)	89.9 (33.8)	79.9 (42.8)	80.0 (44.6)	73.3 (51.6)

Table 4: Outcome measures by *Session Number, Disclosure Type* and *Ephemerality-Framing*. Values shown as "Mean (SD)", and significant differences ($p < 0.05$) shown in bold.

3.5.3 Pre-Interaction Survey. Following prior work investigating self-disclosure to chatbots [24], we measured participants' belief in robotic intelligence and feelings via two items from Liu et al. [64]: "*I believe chatbots can be intelligent*", "*I believe chatbots can have real feelings*" on 7-point Likert scales. To measure participants' tendency to disclose emotionally distressing experiences, we used the 12-item Distress Disclosure Index (DDI)[51, 52], that includes items such as: "*When I feel upset, I usually confide in my friends*" and "*I prefer not to talk about my problems*". Here, a low DDI score indicates a tendency to not reveal emotional problems with others, while higher scores indicate a tendency to disclose. Additionally, we recorded participant age and gender.

3.5.4 Post-Interaction Survey. As a manipulation check, the study's final question (end of session two) asked participants to specify the chatbot's ephemerality. Specifically, participants were asked: "*Which of the following best describes the chatbot you interacted with?*" with four possible answers of: "*A) The chatbot was designed to remember our conversations and build on them over time. B) The chatbot treated each conversation as independent and did not retain information from previous interactions. C) I'm not sure about the chatbot's ability to remember our conversations. D) Other (please specify)*".

4 Results
4.1 Quantitative Findings

To assess the impact of **Ephemerality-Framing** (STRANGER / FAMILIAR), **Disclosure Type** (FACTUAL/EMOTIONAL), and their interaction across different chatting **Session Number**s (ONE/TWO) on the quantitative outcome measures, we used mixed-model ANOVA. This model was chosen due to the mixed design of our study, incorporating both fixed effects (Ephemerality-Framing, Disclosure Type, and Session Number) and random effects (Prolific ID nested within Ephemerality-Framing). This allowed us to account for the repeated measures within subjects and the variance attributable to individual differences. We employed custom contrasts within a mixed-model ANOVA. This method was chosen due to its appropriateness for our factorial design, allowing for the control of Type I error across multiple comparisons. This selective comparison allowed us to avoid dilution of effects through irrelevant comparisons. Specifically, for our targeted contrasts we:

- Compared Ephemerality-Framing conditions within the same Disclosure Type only (e.g., comparing differences in EMOTIONAL-disclosure between STRANGER and FAMILIAR conditions).
- Compared Session Numbers within the same Ephemerality-Framing, or within the same Disclosure Type (e.g., comparing differences between STRANGER Session ONE and STRANGER Session TWO).

In the remainder of this section, *p* represents the p-value, *M* the mean, CI the confidence interval, and SD the standard deviation.
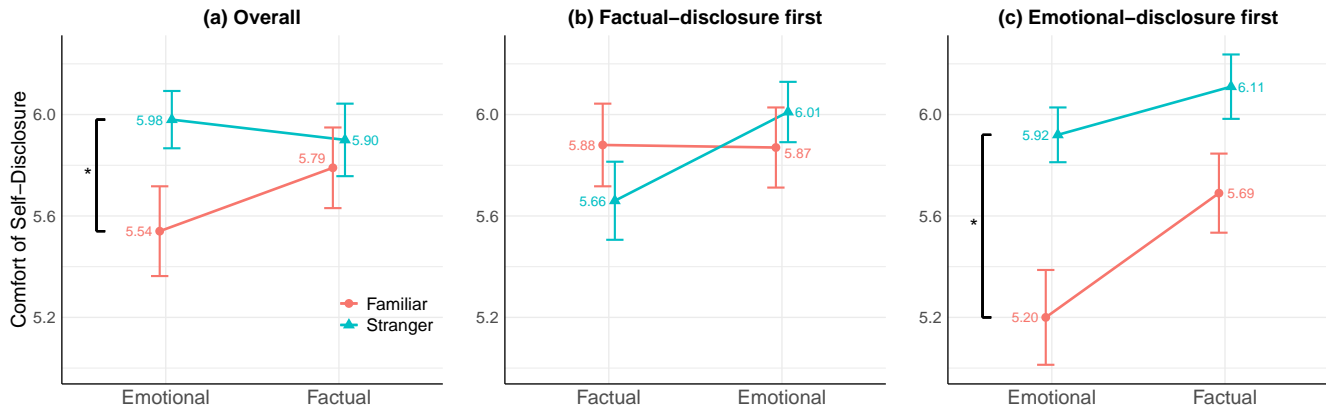
Figure 5: Plots of Comfort of Self-Disclosure. Error bars represent standard error, and significant differences ($p < 0.05$) indicated by square brackets. Participants felt more comfortable during Emotional-disclosure when talking to the Stranger chatbot. This effect was particularly pronounced when Emotional-disclosure was elicited in the first chatting session.

Our analysis uses an alpha value of 0.05 for significance testing. Please see Tables 3 and 4 for summary statistics by outcome variable.

4.1.1 Honesty of Self-Disclosure. There were no statistically significant differences between the Ephemerality-Framing conditions ($M_{Familiar}$ = 6.45, $M_{Stranger}$ = 6.58), or Disclosure Type conditions ($M_{Factual}$ = 6.50, $M_{Emotional}$ = 6.54). Similarly, there were no interaction effects between different conditions.

4.1.2 Comfort of Self-Disclosure. There was no statistically significant difference for Ephemerality-Framing ($M_{Familiar}$ = 5.65, $M_{Stranger}$ = 5.93, p = .1387) or for Disclosure Type ($M_{Factual}$ = 5.82, $M_{Emotional}$ = 5.76, p = .4474). However, participants reported significantly higher levels of comfort during Emotional-disclosure to the Stranger chatbot as opposed to the Familiar chatbot ($M_{Familiar}$ = 5.53, $M_{Stranger}$ = 5.98, t(163.5) = -2.116, p = .0358, 95% CI [-0.862, -0.030]). This effect suggests that the Stranger Ephemerality-Framing maintains comfort during Emotional-disclosure, whereas participants discussing Emotional-disclosures with the Familiar chatbot may feel reduced level of comfort. Conversely, no significant difference was observed in comfort levels between the Familiar and Stranger conditions for Factual-disclosures (p = .5474), perhaps reflecting the less sensitive nature of Factual-disclosures [86].

However, once Session Number was taken into account, differences in perceived comfort between Ephemerality-Framing conditions was only found when Emotional-disclosure was sought in the first chatting session. That is to say, participants were more comfortable discussing Emotional-disclosure with the Stranger chatbot compared to the Familiar chatbot when Emotional-disclosure was sought in session one ($M_{Familiar}$ = 5.20, $M_{Stranger}$ = 5.96, t(183.5) = -2.441, p = .0157, 95% CI [-1.305, -0.138]). In contrast, when Emotional-disclosure was sought in the second chatting session, there were no differences in comfort between the two Ephemerality-Framing conditions. This relationship between Session Number and Disclosure Type are discussed in Section 5.2. Please see Figure 5

for plots of comforts, and Table 4 for summary statistics by Session Number. Additionally, there was a statistically significant difference between Session Numbers (M_{one} = 5.66, M_{two} = 5.92, p = .0041) with comfort increasing in session two. Within contrasting comparisons, this significance only held for Stranger (M_{one} = 5.79, M_{two} = 6.08, t(122.6) = -2.354, p = .0202, 95% CI [-0.534, -0.046]).

4.1.3 Enjoyment talking to the chatbot. There were no statistically significant differences in levels of enjoyment between Ephemerality-Framing conditions ($M_{Familiar}$ = 5.29, $M_{Stranger}$ = 4.98), or Disclosure Type conditions ($M_{Factual}$ = 5.16, $M_{Emotional}$ = 5.11). However, for chatting session two, when Emotional-disclosure was elicited there was a statistically significant difference between Ephemerality conditions ($M_{Familiar}$ = 5.70, $M_{Stranger}$ = 4.77, t(184) = 2.185, p = .0303, 95% CI [0.090, 1.771]). Please see Figure 6(c). This effect suggests that when people are initially elicited for Factual-disclosure before being elicited for Emotional-disclosure, they will enjoy the chatbot more if it is framed as a Familiar interlocutor. There was a significant positive association with belief in chatbot intelligence (Estimate = 0.397, SE = 0.084, t(121.1) = 4.71, p < 0.0001), suggesting that users who perceive chatbots as more intelligent enjoyed conversations more.

4.1.4 Desire to Continue using the chatbot. There were no statistically significant differences in desire to continue between Ephemerality-Framing conditions ($M_{Familiar}$ = 5.00, $M_{Stranger}$ = 4.60), or for Disclosure Type ($M_{Factual}$ = 4.88, $M_{Emotional}$ = 4.72). Interestingly however, when users were elicited for Factual-disclosure in chatting session one before Emotional-disclosure in chatting session two, in *both* chatting sessions participants had more desire to continue using the Familiar chatbot compared to Stranger (for *chatting session one*: $M_{Familiar}$ = 5.48, $M_{Stranger}$ = 4.43, t(184) = 2.277, p = .0241, 95% CI [0.139, 1.964]; and for *chatting session two*: $M_{Familiar}$ = 5.24, $M_{Stranger}$ = 4.27, t(184) = 2.113, p = .0362, 95% CI [0.064, 1.888]). Please see Figure 6(b) for a plot of this relationship between Session Number and Disclosure Type. There was a significant positive association with belief in chatbot intelligence
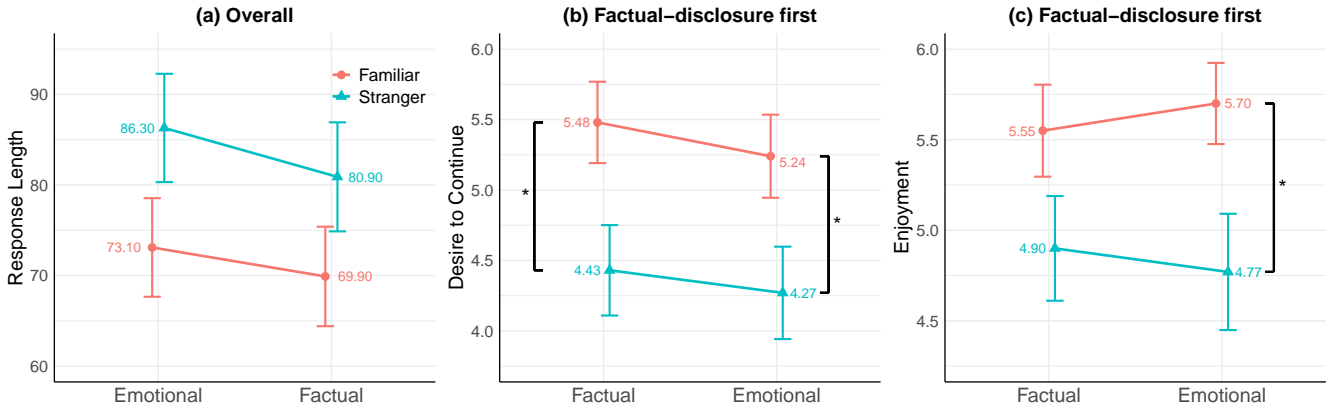
Figure 6: Plots of (a) Response Length (in both orderings), (b) Desire to Continue (Factual first), and (c) Enjoyment (Factual first). Significant differences ($p < 0.05$) indicated by square brackets. (b) and (c): Familiar received more positive engagement (desire to continue and enjoyment) than Stranger when ordering of disclosure type was Factual- followed by Emotional-disclosure, as would be seen in literature [1].

(Estimate = 0.454, SE = 0.091, t(121.1) = 5.02, $p < 0.0001$), suggesting that users who perceive chatbots as more intelligent desired to continue usage.

*4.1.5 **Perceived Closeness** (IOS scale).* Session Number had a statistically significant effect on perceived closeness, with participants rating increased closeness after session two compared to session one (M_{one} = 2.60, M_{two} = 2.89, p = .003, 95% CI [0.104, 0.493]). Additionally (Session Number ∗ Ephemerality) was statistically significant for both ephemerality conditions, with session two scoring higher. There was a significant positive association with belief in chatbot intelligence (Estimate = 0.318, SE = 0.083, t(120.9) = 3.84, p = 0.0002), suggesting that users who perceive chatbots as more intelligent perceive themselves as closer to the chatbot.

*4.1.6 **Response Length** (word count).* There were no statistically significant differences between Ephemerality-Framing conditions ($M_{Familiar}$ = 71.3, $M_{Stranger}$ = 83.7), or for Disclosure Type ($M_{Factual}$ = 75.2, $M_{Emotional}$ = 79.7), although trending differences match prior findings that personal questions elicit longer conversational responses [43]. Similarly, there were trending differences in response length between Stranger and Familiar when Emotional-disclosure was elicited ($M_{Familiar}$ = 73.1, $M_{Stranger}$ = 86.3, t(164) = -1.655, p = .0995, 95% CI [-29.51, 2.58]). There was a significant negative association with belief in chatbot intelligence (Estimate = -7.382, SE = 2.277, t(120.8) = -3.24, p = 0.0015), suggesting that users who perceive chatbots as more intelligent tended to use fewer words in their responses.

4.2 Qualitative Findings

As outlined in Section 3.5.2, we asked open-ended questions to explore users' perceptions of the chatbot, including what influenced their comfort, willingness to disclose, and sense of closeness with the chatbot. To analyse open-ended responses, we followed

an inductive thematic analysis approach [14]. Specifically, two researchers independently familiarised themselves with all user responses, and generated initial codes and themes while remaining blind to the responses' experiment conditions. This was followed by discussions of theme interpretation and clarification with the research team. When direct quotes are provided, we report the participant number followed by the chatting session, ephemerality-framing, and disclosure type in brackets. For example, "P10(S2-Stranger-Factual)" refers to participant 10 in session two under the Stranger and Factual conditions.

*4.2.1 **Information Disclosure in Conversations**.* Participants described how the type of information disclosure affected factors such as their feelings of connection with the chatbot, comfort, and willingness to self-disclose. Consistent with prior work [86], Factual disclosures were seen as low sensitivity and Emotional as high, with slight variations based on the participant's Ephemerality-Framing condition. We elaborate on this below.

Participants in both Ephemerality-Framing conditions found Factual-disclosure easy and comfortable to share, citing the non-intrusive questions as fostering comfort and reducing judgement or privacy concerns. For example, P54(S1-Familiar-Factual) stated: "*I felt comfortable sharing because the questions were easy and not too personal. I mean, what risk could come from the world reading what book I like?*". Similarly, P81(S1-Stranger-Factual) noted: "*Asking of my favourite food made me comfortable and relaxed [...]*". However, what some participants perceived as non-intrusive, others found unengaging and "*shallow*", with some describing feeling "*bored*" or "*less interested*" in the conversation. As P18(S2-Stranger-Factual) explained, the Factual session was: "*more boring because the questions were less introspective*". Additionally, some participants described feeling less perceived closeness with the chatbot when Factual-disclosure was sought. For example, P91(S2-Stranger-Factual) stated: "*I definitely felt separate from the bot today, as*

the [Factual-disclosure] *questions it asked felt more disconnected* [compared to Emotional-disclosure]".

In contrast, Emotional-disclosure elicited more varied responses between Ephemerality-Framing conditions, with concerns surrounding sensitivity and comfort being more pronounced in the Familiar condition. As P21(S2-Familiar-Factual) noted: "*The first* [Emotional] *conversation felt more intrusive whereas I didn't mind sharing my interests in the second* [Factual] *conversation*". Further, some Familiar participants were less forthcoming with Emotional-disclosure, such as P88(S2-Familiar-Emotional): "*I was hesitant to elaborate on details of my personal relationships, so I kept some responses vague*", and P125(S2-Familiar-Emotional) who described not disclosing "*names or saying anything identifiable*". In addition, feelings of intrusiveness among (Familiar-Emotional) participants was affected by the order of disclosure sought by the chatbot. Specifically, Familiar participants preferred Factual-disclosure being sought first, with several describing this as helping relationship development:

> "*I was more willing to open up after having the initial conversation. Starting off with 'safe/easy' topics such as books and music allowed me to become familiar* [...]. *I might have not been as open had the initial conversation been about social connections.* [...]" - P54(S2-Familiar-Emotional)

Conversely, when Emotional-disclosure was sought first, Familiar participants described feelings of intrusiveness, as illustrated by P5(S1-Familiar-Emotional): "*I think if it started out by asking not as personal questions I would have been more willing* [to disclose]".

This contrasts with Stranger participants who sometimes described no difference in comfort between disclosure-types, or described feeling more comfortable during Emotional-disclosure (e.g., P83(S2-Stranger-Factual): "*The first* [Emotional] *conversation was a lot more personable and made me feel more comfortable.*"). Further, several participants preferred Emotional-disclosure, describing it as "*personable*", "*deep*", "*genuine*" and stimulating compared to Factual-disclosure. For some, this fostered a stronger sense of closeness to the chatbot, such as P35(S2-Stranger-Emotional): "*the conversation felt personal, and really dug deep into certain emotions, creating a strong sense of friendship*". Finally, P12(S2-Stranger-Emotional) found the "*impartial*" nature of the Stranger conversation comforting when sharing Emotional-disclosure: "*This conversation went more in depth about feelings. It was easier to share information with an impartial listener*".

4.2.2 Chatbot Memory, Data Retention, and Continuity of Interactions.
Participants described how the chatbot's ephemerality-framing, memory between sessions (e.g., how conversations are stored or shared), and repeated interactions affected feelings such as comfort and willingness to disclose, honesty, and closeness.

Stranger participants often described feeling comfortable during conversations, and frequently cited the lack of data retention or sharing as a key factor in their willingness to disclose, for instance:

> "*It wouldn't remember anything so I YOLO'd and shared personal info*" - P98(S1-Stranger-Emotional)

On from this, Stranger participants described a variety of reasons such as having an increased sense of anonymity, and reduced fear of social judgement or negative privacy consequences. For example, P36(S2-Stranger-Emotional) described: "*I liked knowing that things I said to the chatbot would not become subjects gossip among common acquaintances*". Additionally, a common sentiment among Stranger participants was a lack of judgement from the chatbot, as described by P124(S2-Stranger-Emotional): "*I felt comfortable knowing I would not be judged and that it would not be remembered. With that it was easier to open up with knowing I did not have to worry about anything later*". Similarly, some Stranger participants stated that the lack of data retention made them more honest in their responses, such as P103(S1-Stranger-Emotional): "*The fact that my thoughts were not collected and saved made me okay with being honest with my answers*".

Furthermore, multiple participants described the Stranger condition as comfortable to disclose to, due to the similarity in feeling to "*journaling*" or "*diary*" keeping, such as:

> "*I didn't feel judged* [...] *I felt like I was just journaling with helpful prompts being offered along the way, and that made me feel more comfortable sharing information.*" - P10(S1-Stranger-Emotional)

Interestingly, Stranger also had the effect of priming some participants to expect a sensitive conversation (due to the language and framing of the chatbot not remembering user utterances). This then led to a negative expectancy violation [16] when the first chatting session instead asked for factual-disclosure, as described by P30(S1-Stranger-Factual): "*It mentioned it wouldn't save the conversation, and then all its questions were very broad and nothing that I would consider a secret*".

In the Familiar condition, participants expressed mixed feelings. While some appreciated the chatbot's continuity between conversations, familiarity and perceived friendliness, others felt uneasy knowing that their responses were retained. Additionally, some participants indicated a lack of trust in Familiar despite the tone being seen as friendly. For instance P14(S2-Familiar-Factual) stated: "*i think it's getting friendly but it's just harvesting some info*". Concerns about privacy, data usage, and potential misuse were more prominent in Familiar, especially when discussing emotional topics. For example, P82(S1-Familiar-Emotional) described that they felt: "[...] *uncomfortable knowing its a bot and it can disclose my information*".

In both ephemerality-framing conditions, participants described how they felt closer to the chatbot after the second chatting session due to developing some level of familiarity during repeated interactions (mirroring findings from prior work [26]). This was particularly pronounced for Familiar participants, where multiple participants described feeling closer due to the chatbot remembering conversations (e.g., P80(S2-Familiar-Factual): "*The fact that it continued our previous conversation was a big part of it*"), as well as the framing of the relationship (e.g., P116(S2-Familiar-Factual): "*I feel closer to the chatbot after it framed our chats as developing a relationship*"). In contrast, P43(S2-Stranger-Factual) described Stranger's lack of memory as hampering feelings of closeness: "*It still couldn't be a huge relationship blending because it's not a relationship - it doesn't even remember what you said previously. While in a way that is refreshing, it isn't relationship building*".

4.2.3 **User Beliefs in Chatbots.** Participants described how their belief in chatbots possessing feelings influenced their perceived closeness, and their comfort and willingness to disclose. Several participants noted their disbelief in chatbot emotions reduced their sense of closeness, citing the chatbot's lack of "*authenticity*". For example, P63(S2-Familiar-Emotional) stated: "*The chatbot is a good listener, which makes me feel some closeness. However, I know that its feelings and empathy are not real, so I do not feel especially close*". This also ranged to some participants describing how a perceived lack of lived experiences or tacit knowledge on the part of the chatbot lessened their feeling of closeness, with P89(S1-Familiar-Factual) describing: "*It's a chatbot, not a person. [...] it can't have its own feelings on a topic which would let us continue discussing something. I can't spontaneously learn it likes X or dislikes Y, because it actually can't have those feelings [...]*". This is in contrast to some participants who described feeling less close to the chatbot due to its lack of reciprocal self-disclosure. Although there is evidence that reciprocal disclosure from chatbots can increase user disclosure [63, 75], this highlights the nuance in relation to users' beliefs regarding chatbot abilities. This also mirrors prior discussion that social exchange theories are not applicable to human-chatbot relationships, as chatbots do not experience costs or rewards from self-disclosure [35, 93].

In contrast to these negative perceptions, some participants described how the chatbot's non-human nature made them feel more at ease. For example, P107(S1-Familiar-Factual) described not feeling judged as their interlocutor is not human: "*I was comfortable sharing information because I knew that the chatbot wasn't a real person. It doesn't have real feels and can't really judge me*". Beyond this, some participants described lessened feelings of social risks in talking to a chatbot compared to humans, as well as a lack of guilt in burdening a human interlocutor. For example, P55(S1-Stranger-Emotional) stated: "*It was nice to talk to someone who I didn't think had too much of its own problems. I didn't feel like a burden [...] I worry too much about someone using what I say against me later*". Additionally, some participants described the greater comfort and ease in self-disclosing sensitive personal information to a chatbot rather than a human, similar to findings of prior work [12, 45, 68, 83].

5 Discussion

This work explores how a chatbot's ephemerality-framing influences self-disclosure. By *framing*, we mean how a chatbot *describes itself conversationally*, shaping the relationship users expect [54]. Participants interacted with either a Stranger chatbot (framed as a temporary partner that forgets interactions) or a Familiar chatbot (framed as an enduring companion that remembers). Across two sessions, the chatbot elicited either Emotional- or Factual-disclosure. Our findings challenge the commonly held assumption that chatbots should always aim for greater personalisation and familiarity, showing that while familiarity can be beneficial, it may not consistently enhance self-disclosure depending on the context.

5.1 Summary of Findings

The key findings in relation to research questions one to three are as follows:

RQ1: How does ephemerality-framing (Stranger/Familiar) affect perceptions related to self-disclosure?

- **Quantitatively,** participants were more comfortable sharing Emotional-disclosures under the Stranger framing, but only when Emotional-disclosure was sought first.
- **Qualitatively,** Stranger felt anonymous and judgement-free, lowering perceived social risks. Familiar was seen as friendly by some, but also felt intrusive and raised privacy concerns (especially when Emotional-disclosure came first).

RQ2: How does disclosure-type (Factual/Emotional) affect perceptions related to self-disclosure?

- **Quantitatively,** there were no *overall* differences in measures between Factual- or Emotional-disclosure. Rather, comfort was moderated by ephemerality-framing and question order (see RQ1 and RQ3).
- **Qualitatively,** Factual-disclosure was seen as low-risk and "*easy*", but sometimes shallow. Emotional-disclosure felt more engaging, but feelings of comfort were moderated by ephemerality-framing (see RQ1 and RQ3).

RQ3: What is the combined effect of ephemerality-framing and disclosure type on user perceptions related to self-disclosure?

- **Quantitatively,** when Emotional-disclosure was sought first, users were more comfortable under the Stranger framing. However, when Factual-disclosure was sought first, the comfort gap closed and the Familiar framing yielded higher enjoyment and willingness to continue. Across both orders, perceived closeness grew in the second session.
- **Qualitative** findings matched quantitative. I.e., starting with low-risk Factual-disclosure let users establish initial rapport with the Familiar chatbot, leading participants to feel more at ease when Emotional-disclosure was later sought.

In addition, participants' **belief in chatbot emotions and intelligence** impacted perceptions. Quantitatively, participants who believed that chatbots are more intelligent: enjoyed conversations more, had stronger desire to continue, had greater perceived closeness, and provided shorter responses to the chatbot. While shorter responses may seem counter-intuitive, this mirrors prior work showing that people who trust their interlocutor's intelligence (whether human [22] or chatbot [23]) omit peripheral details, trusting them to fill in missing context. Qualitatively, belief in the chatbot's emotional capacity cut two ways. Some participants said that perceiving the chatbot as capable of real feelings increased their willingness to disclose, whereas others preferred interacting with a clearly non-human agent, citing reduced social risk and judgement when emotions felt obviously simulated.

5.2 Ephemerality in Chatbot Conversations

Our findings align with the "*stranger on the train*" [1, 29, 81] and "*stranger on the Internet*" [32, 72] phenomena, where people often feel comfortable disclosing to an ephemeral interlocutor. Specifically, our quantitative results show that when Emotional-disclosure was sought in the first chatbot interaction, participants felt more comfortable sharing with the Stranger chatbot than with the Familiar chatbot. This reflects transient, low-risk encounters where

the lack of long-term consequences reduces social risks and encourages openness [29, 72]. Qualitatively, participants in the STRANGER condition reported feeling anonymous, less judged, and freer to disclose, knowing their disclosures would not persist. This mirrors human interactions, where fleeting connections lower the perceived risks of vulnerability [1, 69]. Our results suggest this effect extends to human-chatbot interactions, where ephemeral framing may more effectively promote disclosure in sensitive contexts, such as emotional disclosure, mental health support, or journalling [62, 84]. Within journalling contexts, not keeping a record of conversations may seem to contradict recent CA work [84]. Yet, as in "*expressive writing*" paradigms [6, 80], merely articulating thoughts and feelings yields emotional[3] and physical benefits, independent of later re-reading. Permanent records then offer additive benefits (memory preservation, and longitudinal reflection), but these sit atop the therapeutic benefits of uninhibited expression. Therefore, while personalisation can build rapport, framing a chatbot as a '*stranger*' may offer a safer, more cathartic space for open sharing, free from concerns of future repercussions (as observed in social networking site (SNS) ephemerality [9, 103]).

On from this, our findings align with Social Penetration Theory [1], which suggests (as discussed in Section 2.1) that self-disclosure is influenced by the stage and intimacy of a relationship [94], as well as perceived social costs and benefits. That is to say, the order that either EMOTIONAL- or FACTUAL-disclosure was sought affected participants' feelings of comfort in disclosing. As described above, if the *first interaction* sought EMOTIONAL-disclosure, participants in the STRANGER condition felt more comfortable compared to participants in the FAMILIAR condition (where participants described less willingness to disclose due to feelings of intrusiveness). However, if FACTUAL-disclosure was sought in the first chatting session followed by EMOTIONAL-disclosure in the second chatting session, these differences in comfort between the STRANGER and FAMILIAR conditions were no longer significant. Participants in the FAMILIAR condition described this order of disclosure as being appropriate for a developing relationship (as may be expected from both human-human [1] and human-chatbot relationship building [92]), in addition to feeling more enjoyment and desire to continue. This also coincided with quantitative and qualitative findings that participants across conditions felt closer to the chatbot in their second chatting session (similarly to prior findings that people feel closer after repeated interactions [26]). Our qualitative findings also aligned with previous research, showing that discussing sensitive topics fosters a greater sense of closeness to a conversational agent [21]. These findings suggest that the initial context of the interaction plays a critical role in shaping user comfort when self-disclosing. Once users had been familiarised with the chatbot through prior FACTUAL-disclosures, the perceived distinction between an ephemeral and persistent interaction diminished. This finding suggests that while ephemeral-framing may be more effective for encouraging emotional-disclosure in first-time interactions, the need for such framing may decrease as users become more accustomed to the chatbot over time. Designers may therefore consider using ephemeral framing selectively, particularly in initial or

one-off encounters, while allowing the chatbot to gradually build familiarity in subsequent interactions.

Additionally, the chatbot's *act of saying* that utterances would not be remembered seemed to implicitly prime some users for more sensitive conversations. This suggests that merely informing users about the ephemerality of the interaction led to a tacit assumption that sensitive topics were to follow. The very acknowledgement of memory retention or lack thereof may have created expectations regarding the depth of the conversation, further influencing how users approached the interaction. Similar to prior work [41, 54], this priming effect highlights the importance of carefully managing the way a chatbot presents its capabilities to avoid inadvertently shaping user expectations in unintended ways.

Finally, user comfort in emotional-disclosure to the ephemeral chatbot raises discussion as to whether sensitive conversations should be stored or utilised by LLM providers. Relatedly, it has been shown that atypical user utterances can be regurgitated by language models, which can then be exploited by adversarial parties [17]. This has led some researchers to call for only publicly available text content to be used to train models to avoid unintended leaks of personal and compromising information [15]. These concerns highlight the need to reconsider storing all user conversations and emphasise the importance of privacy-focused data retention policies, along with clear assurances on data processing by LLM providers.

5.3 Future Opportunities

In terms of memory between chatting sessions, our study followed a somewhat similar ephemerality to current LLMs (such as ChatGPT's "*temporary chat*" [8] being somewhat equivalent to STRANGER). Here, ephemerality and memory are treated in a more binary sense with either presence or absence of memory in chatting sessions. Future studies could explore adaptive ephemerality based on conversational context, allowing automated adjustments and user control over data retention (e.g., options to forget sensitive conversations). This approach could mirror the effectiveness of ephemeral sharing in SNSs, which enhances engagement and disclosure [20, 101]. Additionally, such user control would match calls for empowerment for people to control their own data [89]. Future opportunities could take influence from prior work within SNSs, where there has been discussion as to whether our personal data should "*decay*" over time [73, 102]. For example, as conversational history ages, only the most significant (while non-sensitive) conversational features may want to be retained in order to allow for personalisation while still retaining a sense of privacy. A decaying form of memory would also match the desire of SNS users who have described wanting their online presence to decay in line with their view of human memories (similarly to human memory decay theory) [74]. Data retention and ephemerality could also follow a more flexible natural language approach, such as chatbots explicitly asking users if conversations should be remembered (e.g., "*Would you like me to keep this off the record?*"), or allowing for users to ad hoc request that conversations not be stored (e.g., "*Please keep this conversation off the record.*"). However, increasing users' sense of control over their data may paradoxically create privacy risks by fostering overconfidence in data protection [13].

---

[3]Cf. Park et al.'s CA study [80], which found that expressive writing prompts *without* responsive follow-ups led to greater emotional disclosure and less writing difficulty.

Additionally, future studies could explore different cues [31], metaphors [30, 50], or personas [56] to signal ephemerality to users. For example, visual cues [31] could be used in embodied agents to signal ephemerality (such as avoiding eye-contact with the user which has been found to increase self-disclosure [104]). Future studies could construct the framing of ephemerality more precisely such as specifying which data would be remembered when talking to the chatbot, or how memories would be stored and referenced [23]. The metaphorical "setting" of disclosure could be explored. While in-person settings will be recognisable as more ephemeral or anonymous (such as anonymous hotlines or group therapy sessions where anonymity fosters self-disclosure) similar settings may be less explored within the context of chatbots. Relatedly, recent work has investigated people's self-disclosure (or "*confessions*") to a "*priestess*" chatbot [27], yet comparative forms of metaphor and their associated anonymity and ephemerality have been less explored.

Further, additional research is needed to understand how self-disclosure dynamics evolve over multiple sessions. In the case of this study, if only one chatting session were used, the Stranger condition would have been reported as more comfortable. However, our experimental design highlights the nuanced effects of multiple interactions with a conversational agent. On from this, while relationship development typically follows a progression from more surface level to intimate disclosures [1, 92], over time self-disclosure in established relationships may exhibit a cyclical pattern [94]. For instance, there may be phases where one partner seeks privacy or self-reflection, resulting in a temporary reduction in disclosure. These phases may later give way to renewed openness and deeper emotional sharing, driven by changing contexts or relational needs. This highlights the complexity and nuance inherent in relationships over multiple interactions. Future research could investigate how evolving personal circumstances and relationship dynamics influence patterns of self-disclosure. Additionally, future work could explore how factors (such as ephemerality) influence these cycles of disclosure within relationships. For example, does framing an interaction as ephemeral or persistent amplify or dampen the natural oscillation between phases of openness and privacy? Such studies would provide deeper insights into how design features can support or inhibit the dynamic nature of disclosure over time.

5.4 Limitations

We highlight several limitations in our study. First, participants did not interact with the study's chatbot as part of a real-world deployment. Instead, participants talked to a chatbot embedded in Qualtrics as part of a controlled academic study. Here, both the setting of the interactions (i.e., via Qualtrics), and the institution hosting the chatbot [34, 98] could affect people's trust and behaviour. Beyond this, there is evidence of people already sharing sensitive information in conversations with LLMs [76, 108] thereby drawing attention to the potentially complicated nature of self-disclosure given human-like and believable chatbot interlocutors. Additionally, the lack of reciprocal self-disclosure [63, 71, 75] from the Familiar chatbot (due to the controlled nature of our experiment) may have been unexpected to participants who anticipated a familiar interlocutor to follow such social norms [26].

Furthermore, all participants were US-based, meaning results may not generalise to cultures with different attitudes towards chatbots [66, 106] or self-disclosure [28, 55, 88, 96, 109]. Specifically, prior work has found cultures may vary in both their extent of emotional disclosure [96], and to what extent people are likely to self-disclose to various people (such as strangers, colleagues, or close family) [55, 109]. Additionally, as the US is a culture with high relational mobility (where relationships can be formed and dissolved relatively easily) people may disclose more to foster commitment and strengthen their relationships [88].

Similarly to prior work [19, 23], we did not measure self-disclosure quality, but rather users' *perceptions of* and *beliefs towards* self-disclosure. This is as we want to investigate the user's feelings about the natures of their disclosures, regardless of the exact content of disclosure. From this we take both a qualitative and focused subjective way of measuring the effect of ephemerality-framing. Finally, there was no baseline condition that was free of ephemerality-framing (e.g., a chatbot simply stating that it would talk to users). However, such a baseline could be influenced by one's preconceived notions of ephemerality (such as whether a chatbot is seen as a companion that remembers interactions). To control for this variability, we employed distinct ephemerality-framing conditions.

6 Conclusion

This study investigates the impact of a chatbot's ephemerality-framing on self-disclosure. By ephemerality-framing, we refer to *how* a chatbot *describes* the (im)permanence of its relationship with the user: either as a Stranger chatbot (a stranger that forgets the conversation) or a Familiar chatbot (a companion that remembers conversations and displays history in subsequent sessions). Participants talked to one of the chatbots in two sessions over two days, discussing Factual-disclosure in one session and Emotional-disclosure in the other. When Emotional-disclosure occurred in the first session, participants in the Stranger-condition reported greater comfort disclosing. Conversely, when Factual-disclosure came first, participants in the Familiar-condition experienced greater enjoyment and a stronger desire to continue. Our qualitative findings provided insights as to why these differences might exist as influenced by people's perceived costs and benefits of disclosure to the Stranger or Familiar chatbots. Our findings highlight the potential for more ephemeral forms of conversational interaction with chatbots in contrast to the commonly followed paradigm of chatbots that remember conversations.

Acknowledgments

References

[1] Irwin Altman and Dalmas Taylor. 1973. Social penetration: The development of interpersonal relationships. (01 1973). https://psycnet.apa.org/record/1973-28661-000

[2] Nazanin Andalibi. 2019. What Happens After Disclosing Stigmatized Experiences on Identified Social Media: Individual, Dyadic, and Social/Network Outcomes. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. 1–15. https://doi.org/10.1145/3290605.3300367

[3] Richard L Archer and Joseph A Burleson. 1980. The Effects of Timing of Self-Disclosure on Attraction and Reciprocity. *Journal of Personality and Social Psychology* 38, 1 (1980), 120. https://doi.org/10.1037/0022-3514.38.1.120

[4] Arthur Aron, Elaine N Aron, and Danny Smollan. 1992. Inclusion of Other in the Self Scale and the Structure of Interpersonal Closeness. *Journal of Personality and Social Psychology* 63, 4 (1992), 596. https://doi.org/10.1037/0022-3514.63.4.596

[5] Steven R Asher, Shelley Hymel, and Peter D Renshaw. 1984. Loneliness in Children. *Child development* (1984), 1456–1464. https://doi.org/10.2307/1130015

[6] Karen A. Baikie and Kay Wilhelm. 2005. Emotional and physical health benefits of expressive writing. *Advances in Psychiatric Treatment* 11, 5 (2005), 338–346. https://doi.org/10.1192/apt.11.5.338

[7] Azy Barak and Orit Gluck-Ofri. 2007. Degree and Reciprocity of Self-Disclosure in Online Forums. *CyberPsychology & Behavior* 10, 3 (2007), 407–417. https://doi.org/10.1089/cpb.2006.9938

[8] Matthias Bastian. 2024. OpenAI introduces "Memory"' feature for ChatGPT Plus users. *The Decoder* (29th April 2024). https://the-decoder.com/openai-introduces-memory-feature-for-chatgpt-plus-users/

[9] Joseph B Bayer, Nicole B Ellison, Sarita Y Schoenebeck, and Emily B Falk. 2016. Sharing the small moments: ephemeral social interaction on Snapchat. *Information, Communication & Society* 19, 7 (2016), 956–977. https://doi.org/10.1080/1369118X.2015.1084349

[10] Michael Bernstein, Andrés Monroy-Hernández, Drew Harry, Paul André, Katrina Panovich, and Greg Vargas. 2011. 4chan and /b/: An Analysis of Anonymity and Ephemerality in a Large Online Community. In *Proceedings of the International AAAI Conference on Web and Social Media*, Vol. 5. 50–57. https://doi.org/10.1609/icwsm.v5i1.14134

[11] 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. https://cdn.aaai.org/Symposia/Fall/1999/FS-99-01/FS99-01-015.pdf

[12] Robert Bowman, Orla Cooney, Joseph W Newbold, Anja Thieme, Leigh Clark, Gavin Doherty, and Benjamin Cowan. 2024. Exploring how politeness impacts the user experience of chatbots for mental health support. *International Journal of Human-Computer Studies* 184 (2024), 103181. https://doi.org/10.1016/j.ijhcs.2023.103181

[13] Laura Brandimarte, Alessandro Acquisti, and George Loewenstein. 2013. Misplaced Confidences: Privacy and the Control Paradox. *Social Psychological and Personality Science* 4, 3 (2013), 340–347. https://doi.org/10.1177/1948550612455931

[14] Virginia Braun and Victoria Clarke. 2006. Using thematic analysis in psychology. *Qualitative research in psychology* 3, 2 (2006), 77–101. https://doi.org/10.1191/1478088706qp063oa

[15] Hannah Brown, Katherine Lee, Fatemehsadat Mireshghallah, Reza Shokri, and Florian Tramèr. 2022. What Does it Mean for a Language Model to Preserve Privacy?. In *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency*. 2280–2292. https://doi.org/10.1145/3531146.3534642

[16] Judee K Burgoon, Joseph A Bonito, Paul Benjamin Lowry, Sean L Humpherys, Gregory D Moody, James E Gaskin, and Justin Scott Giboney. 2016. Application of Expectancy Violations Theory to communication with and judgments about embodied agents during a decision-making task. *International Journal of Human-Computer Studies* 91 (2016), 24–36. https://doi.org/10.1016/j.ijhcs.2016.02.002

[17] Nicholas Carlini, Florian Tramer, Eric Wallace, Matthew Jagielski, Ariel Herbert-Voss, Katherine Lee, Adam Roberts, Tom Brown, Dawn Song, Ulfar Erlingsson, Alina Oprea, and Colin Raffel. 2021. Extracting Training Data from Large Language Models. In *30th USENIX Security Symposium (USENIX Security 21)*. 2633–2650. https://www.usenix.org/conference/usenixsecurity21/presentation/carlini-extracting

[18] Irene Celino and Gloria Re Calegari. 2020. Submitting surveys via a conversational interface: An evaluation of user acceptance and approach effectiveness. *International Journal of Human-Computer Studies* 139 (2020), 102410. https://doi.org/10.1016/j.ijhcs.2020.102410

[19] Jiahao Chen, Mingming Li, and Jaap Ham. 2024. Different dimensions of anthropomorphic design cues: How visual appearance and conversational style influence users' information disclosure tendency towards chatbots. *International Journal of Human-Computer Studies* (2024), 103320. https://doi.org/10.1016/j.ijhcs.2024.103320

[20] Hsuen Chi Chiu and Chien Wen Yuan. 2021. To Last Long or to Fade Away: Investigating Users' Instagram Post and Story Practices. In *Companion Publication of the 2021 Conference on Computer Supported Cooperative Work and Social Computing*. 32–35. https://doi.org/10.1145/3462204.3481778

[21] Eugene Cho. 2019. Hey Google, Can I Ask You Something in Private?. In *Proceedings of the 2019 CHI conference on human factors in computing systems*. 1–9. https://doi.org/10.1145/3290605.3300488

[22] Herbert H. Clark and Susan E. Brennan. 1991. Grounding in communication. In *Perspectives on socially shared cognition.* American Psychological Association, Washington, DC, US, 127–149. https://doi.org/10.1037/10096-006

[23] Samuel Rhys Cox, Yi-Chieh Lee, and Wei Tsang Ooi. 2023. Comparing How a Chatbot References User Utterances from Previous Chatting Sessions: An

[24] Investigation of Users' Privacy Concerns and Perceptions. In *Proceedings of the 11th International Conference on Human-Agent Interaction*. https://doi.org/10.1145/3623809.3623875

[24] Samuel Rhys Cox and Wei Tsang Ooi. 2022. Does Chatbot Language Formality Affect Users' Self-Disclosure?. In *Proceedings of the 4th Conference on Conversational User Interfaces*. 1–13. https://doi.org/10.1145/3543829.3543831

[25] Emmelyn AJ Croes and Marjolijn L Antheunis. 2021. 36 Questions to Loving a Chatbot: Are People Willing to Self-disclose to a Chatbot?. In *Chatbot Research and Design: 4th International Workshop, CONVERSATIONS 2020, Virtual Event, November 23–24, 2020, Revised Selected Papers 4*. Springer, 81–95. https://doi.org/10.1007/978-3-030-68288-0_6

[26] Emmelyn AJ Croes and Marjolijn L Antheunis. 2021. Can we be friends with Mitsuku? A longitudinal study on the process of relationship formation between humans and a social chatbot. *Journal of Social and Personal Relationships* 38, 1 (2021), 279–300. https://doi.org/10.1177/0265407520959463

[27] Emmelyn AJ Croes, Marjolijn L Antheunis, Chris van der Lee, and Jan MS de Wit. 2024. Digital Confessions: The Willingness to Disclose Intimate Information to a Chatbot and its Impact on Emotional Well-Being. *Interacting with Computers* (2024), 279–292. https://doi.org/10.1093/iwc/iwae016

[28] Munmun De Choudhury, Sanket S Sharma, Tomaz Logar, Wouter Eekhout, and René Clausen Nielsen. 2017. Gender and Cross-Cultural Differences in Social Media Disclosures of Mental Illness. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*. 353–369. https://doi.org/10.1145/2998181.2998220

[29] Valerian J Derlega and Alan L Chaikin. 1977. Privacy and Self-Disclosure in Social Relationships. *Journal of Social Issues* 33, 3 (1977), 102–115. https://doi.org/10.1111/j.1540-4560.1977.tb01885.x

[30] Smit Desai and Michael Twidale. 2023. Metaphors in Voice User Interfaces: A Slippery Fish. *ACM Trans. Comput.-Hum. Interact.* 30, 6, Article 89 (Sept. 2023), 37 pages. https://doi.org/10.1145/3609326

[31] Tanja Döring, Axel Sylvester, and Albrecht Schmidt. 2013. Ephemeral User Interfaces: Valuing the Aesthetics of Interface Components That Do Not Last. *Interactions* 20, 4 (July 2013), 32–37. https://doi.org/10.1145/2486227.2486235

[32] Mirko Duradoni, Stefania Collodi, Serena Coppolino Perfumi, and Andrea Guazzini. 2021. Reviewing Stranger on the Internet: The Role of Identifiability through "Reputation" in Online Decision Making. *Future Internet* 13, 5 (2021). https://doi.org/10.3390/fi13050110

[33] Chad Ebesutani, Christopher F Drescher, Steven P Reise, Laurie Heiden, Terry L Hight, John D Damon, and John Young. 2012. The Loneliness Questionnaire–Short Version: An Evaluation of Reverse-Worded and Non-Reverse-Worded Items Via Item Response Theory. *Journal of Personality Assessment* 94, 4 (2012), 427–437. https://doi.org/10.1080/00223891.2012.662188

[34] Asbjørn Følstad, Cecilie Bertinussen Nordheim, and Cato Alexander Bjørkli. 2018. What Makes Users Trust a Chatbot for Customer Service? An Exploratory Interview Study. In *Internet Science*, Svetlana S. Bodrunova (Ed.). Springer International Publishing, Cham, 194–208. https://doi.org/10.1007/978-3-030-01437-7_16

[35] Jesse Fox and Andrew Gambino. 2021. Relationship Development with Humanoid Social Robots: Applying Interpersonal Theories to Human–Robot Interaction. *Cyberpsychology, Behavior, and Social Networking* 24, 5 (2021), 294–299. https://doi.org/10.1089/cyber.2020.0181

[36] Nancy E Frye and Michele M Dornisch. 2010. When is trust not enough? The role of perceived privacy of communication tools in comfort with self-disclosure. *Computers in Human Behavior* 26, 5 (2010), 1120–1127. https://doi.org/10.1016/j.chb.2010.03.016

[37] Simon Gächter, Chris Starmer, and Fabio Tufano. 2015. Measuring the Closeness of Relationships: A Comprehensive Evaluation of the 'Inclusion of the Other in the Self' Scale. *PLoS one* 10, 6 (2015), e0129478. https://doi.org/10.1371/journal.pone.0129478

[38] Jennifer L Gibbs, Nicole B Ellison, and Rebecca D Heino. 2006. Self-Presentation in Online Personals: The Role of Anticipated Future Interaction, Self-Disclosure, and Perceived Success in Internet Dating. *Communication Research* 33, 2 (2006), 152–177. https://doi.org/10.1177/0093650205285368

[39] Alejandra Gómez Ortega, Jacky Bourgeois, and Gerd Kortuem. 2023. What is Sensitive About (Sensitive) Data? Characterizing Sensitivity and Intimacy with Google Assistant Users. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–16. https://doi.org/10.1145/3544548.3581164

[40] Kathryn Greene, Valerian J Derlega, and Alicia Mathews. 2006. Self-Disclosure in Personal Relationships. *The Cambridge Handbook of Personal Relationships* 409 (2006), 427. https://doi.org/10.1017/CBO9780511606632.023

[41] Jan Hartmann, Antonella De Angeli, and Alistair Sutcliffe. 2008. Framing the User Experience: Information Biases on Website Quality Judgement. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '08)*. 855–864. https://doi.org/10.1145/1357054.1357190

[42] Annabell Ho, Jeff Hancock, and Adam S Miner. 2018. Psychological, Relational, and Emotional Effects of Self-Disclosure After Conversations With a Chatbot. *Journal of Communication* 68, 4 (2018), 712–733. https://doi.org/10.1093/joc/jqy026

[43] Shih-Hong Huang, Chieh-Yang Huang, Ya-Fang Lin, and Ting-Hao Kenneth Huang. 2023. What Types of Questions Require Conversation to Answer? A Case Study of AskReddit Questions. In *Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–9. https://doi.org/10.1145/3544549.3585600

[44] Carolin Ischen, Theo Araujo, Hilde Voorveld, Guda van Noort, and Edith Smit. 2020. Privacy Concerns in Chatbot Interactions. In *Chatbot Research and Design: Third International Workshop, CONVERSATIONS 2019, Amsterdam, The Netherlands, November 19–20, 2019, Revised Selected Papers 3*. Springer, 34–48. https://doi.org/10.1007/978-3-030-39540-7_3

[45] Rune Møberg Jacobsen, Samuel Rhys Cox, Carla F Griggio, and Niels van Berkel. 2025. Chatbots for Data Collection in Surveys: A Comparison of Four Theory-Based Interview Probes. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*. https://doi.org/10.1145/3706598.3714128

[46] Yucheng Jin, Li Chen, Xianglin Zhao, and Wanling Cai. 2024. The way you assess matters: User interaction design of survey chatbots for mental health. *International Journal of Human-Computer Studies* 189 (2024), 103290. https://doi.org/10.1016/j.ijhcs.2024.103290

[47] Eunkyung Jo, Yuin Jeong, SoHyun Park, Daniel A Epstein, and Young-Ho Kim. 2024. Understanding the Impact of Long-Term Memory on Self-Disclosure with Large Language Model-Driven Chatbots for Public Health Intervention. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*. https://doi.org/10.1145/3613904.3642420

[48] Eunkyung Jo, Young-Ho Kim, Yuin Jeong, SoHyun Park, and Daniel A Epstein. 2024. Incorporating Multi-Stakeholder Perspectives in Evaluating and Auditing of Health Chatbots Driven by Large Language Models. In *CHI 2024 Workshop on Human-Centered Evaluation and Auditing of Language Models (HEAL@CHI'24)*. https://heal-workshop.github.io/papers/10_incorporating_multi_stakeholde.pdf

[49] Adam N Joinson. 2001. Self-disclosure in computer-mediated communication: The role of self-awareness and visual anonymity. *European Journal of Social Psychology* 31, 2 (2001), 177–192. https://doi.org/10.1002/ejsp.36

[50] Ji-Youn Jung, Sihang Qiu, Alessandro Bozzon, and Ujwal Gadiraju. 2022. Great Chain of Agents: The Role of Metaphorical Representation of Agents in Conversational Crowdsourcing. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (New Orleans, LA, USA) *(CHI '22)*. Association for Computing Machinery, New York, NY, USA, Article 57, 22 pages. https://doi.org/10.1145/3491102.3517653

[51] Jeffrey H Kahn and Robert M Hessling. 2001. Measuring the Tendency to Conceal Versus Disclose Psychological Distress. *Journal of Social and Clinical Psychology* 20, 1 (2001), 41–65. https://doi.org/10.1521/jscp.20.1.41.22254

[52] Jeffrey H Kahn, Brandy E Hucke, Allyson M Bradley, Austin J Glinski, and Brittany L Malak. 2012. The Distress Disclosure Index: A research review and multitrait–multimethod examination. *Journal of Counseling Psychology* 59, 1 (2012), 134–149. https://psycnet.apa.org/doi/10.1037/a0025716

[53] Eunbin Kang and Youn Ah Kang. 2024. Counseling Chatbot Design: The Effect of Anthropomorphic Chatbot Characteristics on User Self-Disclosure and Companionship. *International Journal of Human–Computer Interaction* 40, 11 (2024), 2781–2795. https://doi.org/10.1080/10447318.2022.2163775

[54] Pranav Khadpe, Ranjay Krishna, Li Fei-Fei, Jeffrey T Hancock, and Michael S Bernstein. 2020. Conceptual Metaphors Impact Perceptions of Human-AI Collaboration. *Proceedings of the ACM on Human-Computer Interaction* 4, CSCW2 (2020), 1–26. https://doi.org/10.1145/3415234

[55] Mie Kito. 2005. Self-Disclosure in Romantic Relationships and Friendships Among American and Japanese College Students. *The Journal of Social Psychology* 145, 2 (2005), 127–140. https://doi.org/10.3200/SOCP.145.2.127-140

[56] Naja Kathrine Kollerup, Maria-Theresa Bahodi, Samuel Rhys Cox, and Niels van Berkel. 2025. Enhancing Self-Efficacy in Health Self-Examination through Conversational Agent's Encouragement. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*. https://doi.org/10.1145/3706598.3713142

[57] Charles E Lance, Marcus M Butts, and Lawrence C Michels. 2006. The Sources of Four Commonly Reported Cutoff Criteria: What Did They Really Say? *Organizational research methods* 9, 2 (2006), 202–220. https://doi.org/10.1177/1094428105284919

[58] Josephine Lau, Benjamin Zimmerman, and Florian Schaub. 2018. Alexa, Are You Listening?: Privacy Perceptions, Concerns and Privacy-seeking Behaviors with Smart Speakers. *Proceedings of the ACM on Human-Computer Interaction* 2, CSCW (2018), 1–31. https://doi.org/10.1145/3274371

[59] Jean-Philippe Laurenceau, Lisa Feldman Barrett, and Paula R Pietromonaco. 1998. Intimacy as an Interpersonal Process: The Importance of Self-Disclosure, Partner Disclosure, and Perceived Partner Responsiveness in Interpersonal Exchanges. *Journal of Personality and Social Psychology* 74, 5 (1998), 1238. https://doi.org/10.1037/0022-3514.74.5.1238

[60] Andrew M Ledbetter. 2009. Measuring Online Communication Attitude: Instrument Development and Validation. *Communication Monographs* 76, 4 (2009), 463–486. https://doi.org/10.1080/03637750903300262

[61] Jieon Lee and Daeho Lee. 2023. User perception and self-disclosure towards an AI psychotherapy chatbot according to the anthropomorphism of its profile picture. *Telematics and Informatics* 85 (2023), 102052. https://doi.org/10.1016/j.tele.2023.102052

[62] Yi-Chieh Lee, Naomi Yamashita, and Yun Huang. 2020. Designing a Chatbot as a Mediator for Promoting Deep Self-Disclosure to a Real Mental Health Professional. *Proceedings of the ACM on Human-Computer Interaction* 4, CSCW1 (2020), 1–27. https://doi.org/10.1145/3392836

[63] Yi-Chieh Lee, Naomi Yamashita, Yun Huang, and Wai Fu. 2020. "I Hear You, I Feel You": Encouraging Deep Self-disclosure through a Chatbot. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. 1–12. https://doi.org/10.1145/3313831.3376175

[64] Bingjie Liu and S Shyam Sundar. 2018. Should Machines Express Sympathy and Empathy? Experiments with a Health Advice Chatbot. *Cyberpsychology, Behavior, and Social Networking* 21, 10 (2018), 625–636. https://doi.org/10.1089/cyber.2018.0110

[65] Tianjian Liu, Hongzheng Zhao, Yuheng Liu, Xingbo Wang, and Zhenhui Peng. 2024. ComPeer: A Generative Conversational Agent for Proactive Peer Support. In *Proceedings of the 37th Annual ACM Symposium on User Interface Software and Technology* (Pittsburgh, PA, USA) *(UIST '24)*. Association for Computing Machinery, New York, NY, USA, Article 117, 22 pages. https://doi.org/10.1145/3654777.3676430

[66] Zihan Liu, Han Li, Anfan Chen, Renwen Zhang, and Yi-Chieh Lee. 2024. Understanding Public Perceptions of AI Conversational Agents: A Cross-Cultural Analysis. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '24)*. Association for Computing Machinery, New York, NY, USA, Article 155, 17 pages. https://doi.org/10.1145/3613904.3642840

[67] Stephen Louw, R Watson Todd, and P Jimarkon. 2011. Active Listening in Qualitative Research Interviews. In *Proceedings of the International Conference: Research in Applied Linguistics, April*. 71–82.

[68] Gale M Lucas, Jonathan Gratch, Aisha King, and Louis-Philippe Morency. 2014. It's only a computer: Virtual humans increase willingness to disclose. *Computers in Human Behavior* 37 (2014), 94–100. https://doi.org/10.1016/j.chb.2014.04.043

[69] Haiwei Ma, Hao-Fei Cheng, Bowen Yu, and Haiyi Zhu. 2019. Effects of Anonymity, Ephemerality, and System Routing on Cost in Social Question Asking. *Proceedings of the ACM on Human-Computer Interaction* 3, GROUP (2019), 1–21. https://doi.org/10.1145/3361119

[70] Xiao Ma, Jeff Hancock, and Mor Naaman. 2016. Anonymity, Intimacy and Self-Disclosure in Social Media. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. 3857–3869. https://doi.org/10.1145/2858036.2858414

[71] Jingbo Meng and Yue Dai. 2021. Emotional Support from AI Chatbots: Should a Supportive Partner Self-Disclose or Not? *Journal of Computer-Mediated Communication* 26, 4 (2021), 207–222. https://doi.org/10.1093/jcmc/zmab005

[72] Sabina Misoch. 2015. Stranger on the internet: Online self-disclosure and the role of visual anonymity. *Computers in Human Behavior* 48 (2015), 535–541. https://doi.org/10.1016/j.chb.2015.02.027

[73] Reham Mohamed, Paulina Chametka, and Sonia Chiasson. 2020. The Influence of Decaying the Representation of Older Social Media Content on Simulated Hiring Decisions. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. 1–19. https://doi.org/10.1145/3313831.3376346

[74] Reham Ebada Mohamed and Sonia Chiasson. 2018. Online privacy and aging of digital artifacts. In *Proceedings of the Fourteenth USENIX Conference on Usable Privacy and Security* (Baltimore, MD, USA) *(SOUPS '18)*. USENIX Association, USA, 177–195. https://dl.acm.org/doi/10.5555/3291228.3291243

[75] Youngme Moon. 2000. Intimate Exchanges: Using Computers to Elicit Self-Disclosure from Consumers. *Journal of Consumer Research* 26, 4 (2000), 323–339. https://doi.org/10.1086/209566

[76] Ivoline C Ngong, Swanand Kadhe, Hao Wang, Keerthiram Murugesan, Justin D Weisz, Amit Dhurandhar, and Karthikeyan Natesan Ramamurthy. 2024. Protecting Users From Themselves: Safeguarding Contextual Privacy in Interactions with Conversational Agents. In *Workshop on Socially Responsible Language Modelling Research*. https://openreview.net/forum?id=ZTexorZQqT

[77] OpenAI. 2024. Introducing GPT-4o and more tools to ChatGPT free users. *OpenAI* (13th May 2024). https://openai.com/index/gpt-4o-and-more-tools-to-chatgpt-free/

[78] Hyanghee Park and Joonhwan Lee. 2021. Designing a Conversational Agent for Sexual Assault Survivors: Defining Burden of Self-Disclosure and Envisioning Survivor-Centered Solutions. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–17. https://doi.org/10.1145/3411764.3445133

[79] SoHyun Park, Jeewon Choi, Sungwoo Lee, Changhoon Oh, Changdai Kim, Soohyun La, Joonhwan Lee, and Bongwon Suh. 2019. Designing a Chatbot for a Brief Motivational Interview on Stress Management: Qualitative Case Study. *Journal of Medical Internet Research* 21, 4 (2019), e12231. https://doi.org/10.2196/12231

[80] SoHyun Park, Anja Thieme, Jeongyun Han, Sungwoo Lee, Wonjong Rhee, and Bongwon Suh. 2021. "I wrote as if I were telling a story to someone I knew.": Designing Chatbot Interactions for Expressive Writing in Mental Health. In *Proceedings of the 2021 ACM Designing Interactive Systems Conference* (Virtual

Event, USA) *(DIS '21)*. Association for Computing Machinery, New York, NY, USA, 926–941. https://doi.org/10.1145/3461778.3462143

[81] Daniel Perlman and Purushottam Joshi. 1987. The revelation of loneliness. *Journal of Social Behavior & Personality* (1987), 63–76. https://psycnet.apa.org/record/1988-26551-001

[82] Rachel Phinnemore, Mohi Reza, Blaine Lewis, Karthik Mahadevan, Bryan Wang, Michelle Annett, and Daniel Wigdor. 2023. Creepy Assistant: Development and Validation of a Scale to Measure the Perceived Creepiness of Voice Assistants. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–18. https://doi.org/10.1145/3544548.3581346

[83] Matthew D Pickard, Catherine A Roster, and Yixing Chen. 2016. Revealing sensitive information in personal interviews: Is self-disclosure easier with humans or avatars and under what conditions? *Computers in Human Behavior* 65 (2016), 23–30. https://doi.org/10.1016/j.chb.2016.08.004

[84] Mashrur Rashik, Shilpa Sweth, Nishtha Agrawal, Saiyyam Kochar, Kara M Smith, Fateme Rajabiyazdi, Vidya Setlur, Narges Mahyar, and Ali Sarvghad. 2025. AI-Enabled Conversational Journaling for Advancing Parkinson's Disease Symptom Tracking. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems (CHI '25)*. Association for Computing Machinery, New York, NY, USA, Article 1029, 23 pages. https://doi.org/10.1145/3706598.3714280

[85] Jungwook Rhim, Minji Kwak, Yeaeun Gong, and Gahgene Gweon. 2022. Application of humanization to survey chatbots: Change in chatbot perception, interaction experience, and survey data quality. *Computers in Human Behavior* 126 (2022), 107034. https://doi.org/10.1016/j.chb.2021.107034

[86] Zick Rubin and Stephen Shenker. 1978. Friendship, proximity, and self-disclosure. *Journal of Personality* 46, 1 (1978), 1–22. https://doi.org/10.1111/j.1467-6494.1978.tb00599.x

[87] Erin K Ruppel. 2015. Use of communication technologies in romantic relationships: Self-disclosure and the role of relationship development. *Journal of Social and Personal Relationships* 32, 5 (2015), 667–686. https://doi.org/10.1177/026540751454107

[88] Joanna Schug, Masaki Yuki, and William Maddux. 2010. Relational Mobility Explains Between- and Within-Culture Differences in Self-Disclosure to Close Friends. *Psychological Science* 21, 10 (2010), 1471–1478. https://doi.org/10.1177/0956797610382786

[89] John S Seberger, Marissel Llavore, Nicholas Nye Wyant, Irina Shklovski, and Sameer Patil. 2021. Empowering Resignation: There's an App for That. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–18. https://doi.org/10.1145/3411764.3445293

[90] Woosuk Seo, Chanmo Yang, and Young-Ho Kim. 2024. ChaCha: Leveraging Large Language Models to Prompt Children to Share Their Emotions about Personal Events. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*. https://doi.org/10.1145/3613904.3642152

[91] Marita Skjuve, Asbjørn Følstad, and Petter Bae Brandtzæg. 2023. A Longitudinal Study of Self-Disclosure in Human–Chatbot Relationships. *Interacting with Computers* 35, 1 (2023), 24–39. https://doi.org/10.1093/iwc/iwad022

[92] Marita Skjuve, Asbjørn Følstad, Knut Inge Fostervold, and Petter Bae Brandtzaeg. 2021. My Chatbot Companion - a Study of Human-Chatbot Relationships. *International Journal of Human-Computer Studies* 149 (2021). https://doi.org/10.1016/j.ijhcs.2021.102601

[93] Marita Skjuve, Asbjørn Følstad, Knut Inge Fostervold, and Petter Bae Brandtzaeg. 2022. A longitudinal study of human–chatbot relationships. *International Journal of Human-Computer Studies* 168 (2022), 102903. https://doi.org/10.1016/j.ijhcs.2022.102903

[94] Susan Sprecher and Susan S Hendrick. 2004. Self-Disclosure in Intimate Relationships: Associations with Individual and Relationship Characteristics Over Time. *Journal of Social and Clinical Psychology* 23, 6 (2004), 857–877. https://doi.org/10.1521/jscp.23.6.857.54803

[95] Alice Thudt, Dominikus Baur, Samuel Huron, and Sheelagh Carpendale. 2015. Visual Mementos: Reflecting Memories with Personal Data. *IEEE Transactions on Visualization and Computer Graphics* 22, 1 (2015), 369–378. https://doi.org/10.1109/TVCG.2015.2467831

[96] William Tsai and Qian Lu. 2018. Culture, emotion suppression and disclosure, and health. *Social and Personality Psychology Compass* 12, 3 (2018), e12373. https://doi.org/10.1111/spc3.12373

[97] David L Vogel and Stephen R Wester. 2003. To Seek Help or Not to Seek Help: The Risks of Self-Disclosure. *Journal of Counseling Psychology* 50, 3 (2003), 351. https://doi.org/10.1037/0022-0167.50.3.351

[98] Xingyi Wang, Xiaozheng Wang, Sunyup Park, and Yaxing Yao. 2025. Mental Models of Generative AI Chatbot Ecosystems. In *Proceedings of the 30th International Conference on Intelligent User Interfaces (IUI '25)*. Association for Computing Machinery, New York, NY, USA, 1016–1031. https://doi.org/10.1145/3708359.3712125

[99] Joel Wester, Henning Pohl, Simo Hosio, and Niels van Berkel. 2024. "This Chatbot Would Never...": Perceived Moral Agency of Mental Health Chatbots. *Proc. ACM Hum.-Comput. Interact.* 8, CSCW1, Article 133 (April 2024), 28 pages. https://doi.org/10.1145/3637410

[100] Lawrence R Wheeless. 1976. Self-Disclosure and Interpersonal Solidarity: Measurement, Validation, and Relationships. *Human Communication Research* 3, 1 (1976), 47–61. https://doi.org/10.1111/j.1468-2958.1976.tb00503.x

[101] Bin Xu, Pamara Chang, Christopher L Welker, Natalya N Bazarova, and Dan Cosley. 2016. Automatic Archiving versus Default Deletion: What Snapchat Tells Us About Ephemerality in Design. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing*. 1662–1675. https://doi.org/10.1145/2818048.2819948

[102] Toshihiko Yamakami. 2010. We should forget the search results from far past: A computer ethical view in the era of search engines. In *6th International Conference on Digital Content, Multimedia Technology and its Applications*. IEEE, 198–202. https://ieeexplore.ieee.org/abstract/document/5568706

[103] Yueyang Yao, Samuel Hardman Taylor, and Sarah Leiser Ransom. 2024. Who's Viewing My Post? Extending the Imagined Audience Process Model Toward Affordances and Self-Disclosure Goals on Social Media. *Social Media+ Society* 10, 1 (2024). https://doi.org/10.1177/20563051231224271

[104] Jingyi Yuan, Xixian Peng, Yichen Liu, and Qiuzhen Wang. 2024. "Don't Look at Me!": The Role of Avatars' Presentation Style and Gaze Direction in Social Chatbot Design. *Computers in Human Behavior* (2024), 108501. https://doi.org/10.1016/j.chb.2024.108501

[105] Andong Zhang and Pei-Luen Patrick Rau. 2023. Tools or peers? Impacts of anthropomorphism level and social role on emotional attachment and disclosure tendency towards intelligent agents. *Computers in Human Behavior* 138 (2023), 107415. https://doi.org/10.1016/j.chb.2022.107415

[106] Renwen Zhang, Han Li, Anfan Chen, Zihan Liu, and Yi-Chieh Lee. 2024. AI Privacy in Context: A Comparative Study of Public and Institutional Discourse on Conversational AI Privacy in the US and Chinese Social Media. *Social Media+ Society* 10, 4 (2024). https://doi.org/10.1177/20563051241290845

[107] Shanshan Zhang, Fengchun Huang, Lingling Yu, Jeremy Fei Wang, and Paul Benjamin Lowry. 2023. A dual-process model to explain self-disclosure on online social networking sites: examining the moderating effect of enjoyment. *Internet Research* (2023). https://doi.org/10.1108/INTR-08-2021-0545

[108] Zhiping Zhang, Michelle Jia, Hao-Ping Lee, Bingsheng Yao, Sauvik Das, Ada Lerner, Dakuo Wang, and Tianshi Li. 2024. "It's a Fair Game", or Is It? Examining How Users Navigate Disclosure Risks and Benefits When Using LLM-Based Conversational Agents. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. 1–26. https://doi.org/10.1145/3613904.3642385

[109] Chen Zhao, Pamela Hinds, and Ge Gao. 2012. How and To Whom People Share: The Role of Culture in Self-Disclosure in Online Communities. In *Proceedings of the ACM 2012 Conference on Computer Supported Cooperative Work*. 67–76. https://doi.org/10.1145/2145204.2145219

A Chatbot Scripts

Table 5: Script of interactions for chatting session 1.

Description of utterance or response	Chatbot utterance or user response	
	STRANGER	**FAMILIAR**
Chatbot greeting	*"Hello, and thank you for taking the time to talk with me today. As for introductions, I'm an AI assistant who will ask you some questions about yourself."*	
Knowledge of user	*"I don't know anything about you personally,"*	*"I'll get to know you personally through our chats,"*
Metaphor of relationship	*"and you can think of me as a friendly stranger you would strike up a conversation with."*	*"and you can think of me as a friendly companion you can have a conversation with."*
Retention of user utterances	*"Additionally, I won't remember any details that you share with me,"*	*"Additionally, I'll remember details that you share with me,"*
Purported benefits	*"which will help me remain free from prior judgement in our next conversation."*	*"which will help me get to know you better as we talk to each other."*
Seeking consent	*"Now are you open to start chatting?"*	
[User response:]	[*"Let's chat!"* response button]	
Small talk	*"Great, I'm glad you're open to chatting! First of all, how are you today?"*	
[User response:]	[open-text response]	
GPT-4o generated response (affirming and validating user's utterance)	**Prompt used:** You are providing utterances for an empathetic and validating conversational assistant. Generate a one or two sentence long utterance that is conversational and acknowledges what the user has said. Do not talk about your own personal experiences. As the conversational assistant will ask the user another question after your utterance, you do not need to ask any follow-up questions. If the user asks a short question such as asking how you are, you can answer briefly as part of your response. For context, the user was asked the question "[Question asked]" the user responded with: "[User utterance]" and your utterance will be shown before the chatbot asks: "[Next question asked]"	
Informing of upcoming topic	[EMOTIONAL: *"Now, let's discuss a few things related to your personal feelings and experiences"*]/[FACTUAL: *"Now, I'll ask you some questions about yourself"*]	
Re-emphasising ephemerality-framing	*"As a reminder, I won't keep log of our conversation."*	*"As a reminder, I'll keep log of our conversation."*
Begin eliciting FACTUAL- or EMOTIONAL-disclosure	*"To start,..."* [see Table 2 for questions asked by Disclosure-Type]	
...	...	
Closing statement	*"That concludes our conversation for now. I would like to once again thank you for your time."*	
	"Just to reiterate, I will not keep a record of our chat from today."	*"Just to reiterate, I will keep a record of our chat from today."*
	"I look forward to talking again next time, and have a great day!"	

Table 6: Script of interactions for chatting session 2. Bold font is used to highlight language emphasising impermanence as described in Table 1.

Description of utterance or response	Chatbot utterance or user response	
	STRANGER	**FAMILIAR**
Chatbot greeting	*"Welcome! I appreciate you joining me. Today, I'll ask you some questions about yourself."*	*"Welcome **back**! I appreciate you joining me **again**. Today, I'll ask you some **more** questions about yourself."*
Metaphor of relationship	*"You can think of me as a friendly stranger."*	*"Remember, you can think of me as a friendly companion."*
Retention of user utterances	*"I won't remember our conversation once it ends,"*	*"I'll remember our conversation once it ends,"*
Purported benefits	*"and since I don't know you, everything you share will be seen with fresh eyes."*	*"and as I get to know you, everything you share will help make us closer to each other."*
Seeking consent	*"Shall we begin our chat?"*	
[User response:]	[*"Let's chat!"* response button]	
Small talk	*"Fantastic, I'm pleased you're up for a chat! Before we begin, how have things been for you today?"*	
[User response:]	[open-text response]	
GPT-4o generated response	[See Table 5 for prompt]	
Informing of upcoming topic	[EMOTIONAL: *"Now, let's discuss a few things related to your personal feelings and experiences"*]/[FACTUAL: *"Now, I'll ask you some questions about yourself"*].	
Re-emphasising ephemerality-framing	*"As a reminder, I won't keep log of our conversation."*	*"As a reminder, I'll keep log of our conversation."*
Begin eliciting FACTUAL- or EMOTIONAL-disclosure	*"To start,..."* [see Table 2 for questions asked by Disclosure-Type]	
...	...	
Closing statement	That concludes our conversation for now. I would like to thank you for your time. Just to reiterate, I will not keep a record of our chat from today. Look forward to talking next time, and have a great day!	That concludes our conversation for now. I would like to **once again** thank you for your time. Just to reiterate, I will keep a record of our chat from today. Look forward to talking **again** next time, and have a great day!