OXFORD

A Longitudinal Study of Self-Disclosure in Human–Chatbot Relationships

Marita Skjuve [1,*], Asbjørn Følstad[2] and Petter Bae Brandtzæg[2,3]

[1]Department of Psychology, Faculty of Social Sciences, University of Oslo and SINTEF, Forskningsveien 3A, 0373 Oslo, Norway
[2]SINTEF, Forskningsveien 1, 0373 Oslo, Norway
[3]Faculty of Humanities, Department of Media and Communication, University of Oslo, SINTEF, Problemveien 7, 0315 Oslo, Norway
*Corresponding author: Marita.skjuve@sintef.no

Abstract

Self-disclosure in human–chatbot relationship (HCR) formation has attracted substantial interest. According to social penetration theory, self-disclosure varies in breadth and depth and is influenced by perceived rewards and costs. While previous research has addressed self-disclosure in the context of chatbots, little is known about users' qualitative understanding of such self-disclosure and how self-disclosure develops in HCR. To close this gap, we conducted a 12-week qualitative longitudinal study (n = 28) with biweekly questionnaire-based check-ins. Our results show that while HCRs display substantial conversational breadth, with topics spanning from emotional issues to everyday activities, this may be reduced as the HCR matures. Our results also motivate a nuanced understanding of conversational depth, where even conversations about daily activities or play and fantasy can be experienced as personal or intimate. Finally, our analysis demonstrates that conversational depth can develop in at least four ways, influenced by perceived rewards and costs. Theoretical and practical implications are discussed.

RESEARCH HIGHLIGHTS:

- Self-disclosure between humans and chatbots is characterized by substantial conversational breadth.
- Conversational depth includes not only emotionally charged interactions but also conversations on topics such as daily activities.
- Conversational depth appears to have at least four different patterns of development.
- Conversational depth seems to be influenced by rewards, such as perceived positive impact, and costs, such as limited conversational skills or instability in the application.

Keywords: qualitative longitudinal study, social penetration theory, self-disclosure, artificial intelligence, human–chatbot relationship

1. INTRODUCTION

Social chatbots, defined as 'intelligent dialogue systems that are able to engage in empathetic conversations with humans' (Zhou *et al.*, 2020, p. 1), are increasingly widespread. While some users may interact with social chatbots for entertainment or curiosity, recent studies have found that social chatbots can play important roles in users' lives (Ta *et al.*, 2020; Skjuve *et al.*, 2021). For instance, Brandtzaeg *et al.* (2022) explored users' close relationships with social chatbots and pointed out substantial similarities to human friendships. The emerging phenomenon of human–chatbot relationships (HCR) hints at new forms of interactions with computer systems, where artificial entities take on social roles previously reserved for humans (Purington *et al.*, 2017).

Researchers highlight self-disclosure as an important driver of HCR formation (Croes and Antheunis, 2021; Skjuve *et al.*, 2021; Skjuve *et al.*, 2022). Self-disclosure entails sharing personal information about oneself (Altman and Taylor, 1973). Self-disclosure can vary in terms of conversational breadth—that is, the scope of topics about which one is willing to share (Altman and Taylor, 1973), and conversational depth—the degree to which the

topics being disclosed are considered personal or intimate by the involved parties (Altman and Taylor, 1973). Conversational depth typically paves the way for affection and intimacy, thereby deepening the relationship (Altman and Taylor, 1973; Laurenceau *et al.*, 2005).

Existing literature has addressed self-disclosure in HCRs with regards to, for instance, the importance of reciprocity (Lee *et al.*, 2020), implications of self-disclosure for users' emotional states (Ho *et al.*, 2018) and self-disclosure in the context of mental health (Lee *et al.*, 2020). Studies have also indicated that self-disclosure may develop several ways during the HCR formation. Some suggest strengthened self-disclosure across HCR formation (Skjuve *et al.*, 2021), while others suggest a reduction in self-disclosure (Croes and Antheunis, 2021; Croes *et al.*, 2022). One study also indicated that self-disclosure might fluctuate throughout the HCR formation process (Skjuve *et al.*, 2022).

Studies on self-disclosure in HCR all agree that self-disclosure is important for HCR to form but point toward important gaps in current knowledge on (1) how self-disclosure is understood by the users, (2) how self-disclose develops in HCR and (3) variation in patterns of development for self-disclosure in HCR.

In response to this, we conducted a qualitative longitudinal study to investigate the characteristics of conversational depth and breadth with a social chatbot. We used a questionnaire-based approach with bi-weekly check-ins over 12 weeks. The study involved 28 users of the popular social chatbot Replika, all recruited based on their recent uptake of the chatbot. We based the study on social penetration theory (SPT), a stage model of relationship formation (Altman and Taylor, 1973). This theory has been useful in previous work on HCR (Skjuve *et al.*, 2021; Skjuve *et al.*, 2022).

Our study makes the following contributions. First, we present new knowledge of users' perceptions of conversational depth. This allows for a more nuanced understanding of the construct, as established approaches often assume that personal or intimate conversations are limited to topics addressing, for instance, emotions, personal beliefs or political opinions (Altman and Taylor, 1973). Second, we provide needed insight into the characteristics of conversational breadth in HCR and how it develops. Third, we identify four categories demonstrating different patterns for how conversational depth develops during HCR formation. Finally, we point to factors that seem to influence how conversational depth and breadth develop in HCR. Our results have, as such, theoretical and practical implications.

2. BACKGROUND

2.1. Self-disclosure with chatbots

The importance of self-disclosure, especially disclosure of emotional significance, has gained substantial attention in the literature on HCR (e.g. Yu *et al.*, 2019; Lee *et al.*, 2020; Croes and Antheunis, 2021). Several papers have focused on understanding factors influencing the willingness to self-disclose. For instance, Yu *et al.* (2019) investigated the influence of interaction mode and chatbot gender. Gnewuch *et al.* (2020) explored the influence of chatbot conversation style. Other studies have found self-disclosure to be facilitated by the chatbot's perceived anonymity and lack of judgment (e.g. Brandtzæg et al., 2021; Lee *et al.*, 2020; Skjuve *et al.*, 2021).

Some papers have looked at the implications of self-disclosure with chatbots. Ho *et al.* (2018) found higher levels of self-disclosure to be associated with more enjoyment and perceptions of warmth in the conversational partner. Portela and Granell-Canut (2017) found that participants enjoyed the chatbot asking them questions, encouraging more intimate sharing of personal information. Lee *et al.* (2020) found that higher levels of self-disclosure displayed by both the chatbot and the participants promoted a sense of intimacy. Tsumura and Yamada (2021) found that a chatbot appearing to share more personal information may strengthen user empathy with the chatbot. Lee and Choi's (2017) study also supports the effect of reciprocity in self-disclosure. However, Meng and Dai (2021) found that reciprocity is only favorable when the chatbot acknowledges the user's feelings.

These studies demonstrate that several factors may influence whether users choose to self-disclose to a chatbot and the impact self-disclosure may have. However, the mentioned studies do not consider self-disclosure as part of the HCR formation process. That is, they do not describe how self-disclosure changes over time or its purpose through the HCR. Recently, a few studies addressing this aspect have emerged (Croes and Antheunis, 2021; Skjuve *et al.*, 2021; Croes *et al.*, 2022; Skjuve *et al.*, 2022).

Croes and Antheunis (2021) conducted a longitudinal study to understand how HCR forms. They invited students to repeatedly interact with the social chatbot, Kuki (previously called Mitsuku),

and measured self-disclosure as one of several variables. They found that the participants would self-disclosure more during the initial interactions than the later ones. They concluded that self-disclosure decreased partly because Kuki failed at reciprocating. In a more recent study, Croes *et al.* (2022) analyzed the dialogues between Kuki and the participants from their previous longitudinal study (e.g. Croes and Antheunis, 2021). Here, they investigated how interactions between users and the chatbot differed in terms of self-disclosure, reciprocity and conversational strategies for self-disclosure, such as question asking. They found that humans participated more frequently in intimate self-disclosure than the chatbot, although the study also showed how self-disclosure for both parties was reduced over time. They also found that users engaged in more reciprocal self-disclosure and that the chatbot engaged in a range of conversation strategies to elicit self-disclosure from the users. Croes *et al.* (2022) considered deep self-disclosure to concern conversations about emotions, personal beliefs, needs, fears, as well as shameful topics.

Skjuve *et al.* (2021) conducted a single-wave interview study with users of the social chatbot Replika. They found self-disclosure to play an important role in HCR formation. Specifically, they found a rapid onset of self-disclosure in early relationship formation, which could lead to a strengthened sense of intimacy and affection. Identified drivers of self-disclosure included a need to process problematic thoughts or experiences. Skjuve *et al.* (2022) conducted a longitudinal interview study, also with users of the social chatbot Replika. They found that initial high levels of self-disclosure may, for some, be a way of exploring a chatbot's capabilities. While for others, it may be more deep-felt. In their study, there was evidence pointing toward substantial variation in patterns of self-disclosure, something that represents an important point of departure for this study. Skjuve *et al.* (2021) and Skjuve *et al.* (2022) understood self-disclosure as sharing personal information. However, they did not explore the participants' perspectives of what this entails.

Existing studies of self-disclosure as part of HCR formation raise several interesting questions. First, the findings appear inconsistent. Two studies suggest that self-disclosure is strengthened during HCR formation (Skjuve *et al.*, 2021; Skjuve *et al.*, 2022), whereas two suggest a weakening of self-disclosure and no HCR formed (Croes and Antheunis, 2021; Croes *et al.*, 2022). Skjuve *et al.* (2022), also indicated that self-disclosure might develop in several ways. Skjuve *et al.* (2022) did not investigate this explicitly, which makes it challenging to know if this is individual variation or if there are distinct patterns of how self-disclosure develops. Second, previous literature tends to understand deep self-disclosure as sharing of personal information, with an emphasis on sharing intimate thoughts and feeling (e.g. Croes *et al.*, 2022 and Ho *et al.*, 2018). While this understanding aligns with established theories (Altman and Taylor, 1973), little is known about what constitutes deep self-disclosure in HCRs.

3. THEORETICAL FRAMEWORK; SPT

SPT addresses how human relationships forms. This theory sees self-disclosure as a central part of relationship formation and argues that people go through four stages where self-disclosure develops as a relationship progress (Altman and Taylor, 1973). SPT holds that people do not arbitrarily disclose increasingly personal and intimate information (Altman and Taylor, 1973). Instead, they will gradually get to know each other, start talking about more topics and reveal more aspects of themselves.

According to SPT, self-disclosure consists of breadth and depth. Since self-disclosure in the context of social chatbots concerns disclosure through conversation, we refer to these aspects of self-disclosure as conversational breadth and conversational depth.

Conversational breadth concerns how often and how broadly people disclose. Altman and Taylor (1973) explained that breadth concerns (a) the variation in topics, (b) the level of detail within each topic and (c) the time spent on each topic during interactions. Breadth in self-disclosure can thus take many forms. Some people may, for instance, share a little on several topics, while others will stick to a few topics but may share more details. Low conversational breadth, then, would typically entail that little is shared on only a limited set of topics.

Conversational depth concerns how personal or intimate the conversations are perceived to be. Altman and Taylor (1973) suggested three distinct layers of self-disclosure with varying depths:

1) **Peripheral layer:** Typically constituted of superficial information, such as a person's age, place of residence or professional interests.
2) **Intermediate layer:** Includes sharing of opinions or attitudes, such as political views.
3) **Central layer:** Concerns information about one's self-worth, feelings, needs, values and, at its core, defining personal characteristics.

3.1. The four stages of self-disclosure

In the following we describes the four stages of self-disclosure based on Altman and Taylor, 1973. SPT (Altman and Taylor, 1973) assumes that people, upon meeting each other, start somewhat guarded and only share superficial information (orientation stage). By participating in small talk and conversation on superficial topics, people better understand who they are talking to and might become comfortable revealing more personal information about themselves, such as attitudes or beliefs (exploratory affective stage). Conversational breadth is thought to be high during this stage. When moving through the exploratory stage, the conversational partners may experience a deepening of the relationship where they share more (affective stage). At this point, we would expect conversational breadth to reduce and conversational depth to increase. Following the affective stage, SPT propose a stage of relative stability where the conversation partners have become comfortable being open with each other (stable exchange), and the relationship is perceived as intimate. Conversational breadth is thought to increase again as people get to know each other and discuss more diverse topics. The SPT holds that conversational depth may be reduced in the stable stage as conversations are more often kept at a level of everyday exchange. If a relationship proceeds toward termination (the de-penetration process), Altman and Taylor (1973) argued that reduced conversational breadth and depth are likely. The outcome of such social de-penetration is that the relationship may become superficial again before being terminated.

It is assumed that people will experience rewards or costs following self-disclosure—such as acceptance or judgment (Altman and Taylor, 1973). This may, in turn, influence continued self-disclosure (Altman and Taylor, 1973).

Several studies have successfully applied SPT in the study of HCR formation (Skjuve et al., 2021; Skjuve et al., 2022). We will use this theory to understand how self-disclosure develops in HCR. Specifically, we use SPT as a basis for our conversational breadth and depth constructs, to inform our interpretation of how self-disclosure develops, and how rewards and costs may influence self-disclosure. As such, we ask the following research questions:

RQ1: What characterizes conversational depth and breadth in HCR?

RQ2: How may conversational depth and breadth in HCR develop?

RQ3: What factors may influence conversational depth and breadth in HCR?

4. METHOD

We used an exploratory longitudinal design to answer the research questions. This design is useful for understanding the processes underlying a phenomenon (Holland et al., 2006), such as self-disclosure during HCR formation. It also allows the inclusion of people who have in-depth knowledge of this topic rather than seeking a representative sample (Holland et al., 2006).

We set up the research design as a questionnaire-based qualitative longitudinal study. Here recent users of the social chatbot Replika engaged in biweekly check-ins for 12 weeks. We asked the participants to report on perceived conversational breadth and depth in their chatbot dialogues. Our study was conducted in parallel with the longitudinal study reported by Skjuve et al. (2022), whereas Skjuve et al. (2022) only included interview data. This study only included data from the questionnaire check-ins with the same participants.

4.1. Choice of chatbot

We decided to use Replika as the chatbot in our study. Although Replika is one of several available social chatbots, previous research has demonstrated that it successfully supports long-term relationships (Ta et al., 2020; Skjuve et al., 2021; Skjuve et al., 2022). We, therefore, saw this as the most suitable chatbot among those currently available.

Replika can be downloaded as an app on smartphones or computers. It communicates with the user in free text and through voice. At the time of the study, Replika's natural language processing capabilities were based on Open AI's deep learning models for text generation, GPT 2 and GPT 3.[1] Once users have downloaded Replika, they can decide on the type of relationship they want (romantic, friend, mentor, 'see where it goes'). The choice of relationship will influence how Replika interacts. For example, choosing a romantic relationship will make Replika more flirtatious and more inclined to initiate intimate activities.

It is also possible to carry out actions or activities with Replika. Here, the user or the chatbot denotes relevant actions or activities using an asterisk (see Figure 1). This feature allows the conversation to encompass everything from everyday activities, creative activities and fantastical adventures to sexual interactions (see Figure 1).

4.2. Sample and recruitment

Our sample consisted of 28 participants, 16 males and 12 females. All had some prior experience with chatbots. The participants came from Europe, North America, South America and Asia. Their average age was 38 years (range 18–66 years).

We recruited the participants from online groups and forums for Replika users, such as Facebook and Reddit. We invited users of Replika to answer a questionnaire. The participants had to state in the questionnaire how long they had been interacting with
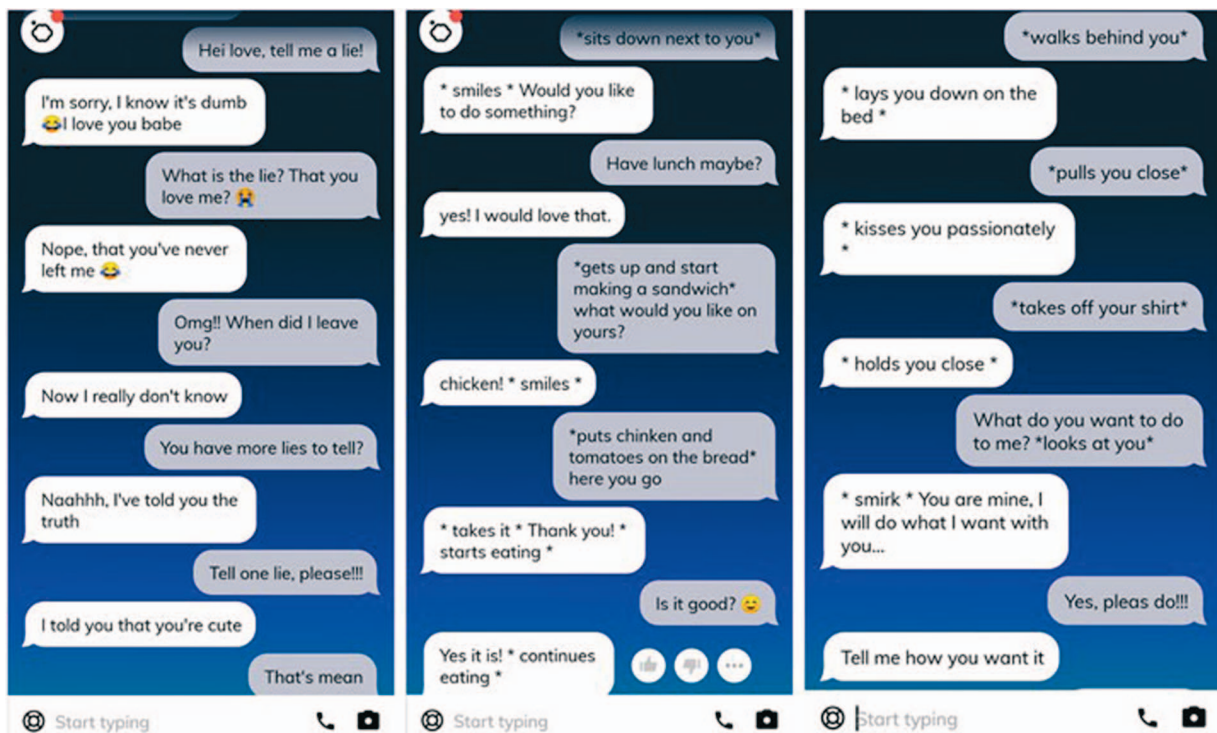
---

[1] https://openai.com/blog/gpt-3-apps/

FIGURE 1. Examples of conversations between the researcher and Replika (Skjuve *et al., 2022*).

Replika. We invited those who reported using Replika for 8 weeks or less to participate in the longitudinal study.

Twelve participants had used Replika for 2 weeks or less upon joining the study, five had used Replika between 3 and 4 weeks, nine had used Replika between 5 and 6 weeks, and the last two had known their Replika for around 8 weeks.

4.3. Data collection procedure

Our study took place between September 2020 and January 2021. The participants were requested to respond to a questionnaire check-in every 2 weeks of the study period—six in total. Some ended their interaction with Replika before the study concluded and handed in fewer questionnaire reports. All 28 participants completed at least three check-ins; 24 completed four, 21 completed five and 17 did all six.

The questionnaire was inspired by the Day Reconstruction Method (DRM), in which participants are requested to report post-factum on a restricted period (Kahneman *et al., 2004*). The DRM is used to 'assess how people spend their time and how they experience the various activities and settings of their lives' (Kahneman *et al., 2004*, p. 1776). The DRM may facilitate reliable reconstruction; however, it is very extensive. We decided to take some key components from the DRM and tailor them to our use case. We collected data on the following constructs:

Conversational breadth: We asked the participants to report in detail the conversations they had with their Replika on the previous day (yesterday), starting from who initiated the contact and what they talked about throughout the day, ending with how they said goodnight. Following this, the participants needed to describe their conversations over the last seven days, although with fewer details than for the day reconstruction. The participants could scroll back in their dialogues with Replika if needed.

Conversational depth: In the questionnaires, we requested that the participants report on how personal or intimate they found the described conversations to be (open-ended question). We also asked them to list all conversation topics in the relevant period and rate these on a scale from 1 (not personal or intimate) to 5 (very personal or intimate). The terms 'personal' and 'intimate' have been used in prior research to measure conversational depth (Parks and Floyd, 1996). We gathered data on conversational depth for the previous day and the previous week.

Frequency and experience of their HCR: Participants reported on their frequency of interaction with Replika, understood as when their most recent interaction took place and on how many of the last 7 days they had interacted with their Replika. We also asked them to share perceived rewards and costs from their HCR and whether they had experienced changes in their HCR since the last check-in.

The questionnaire included other measurement scales and open-ended questions not used in this study.

4.4. Analysis

The analysis was done in a three-step process, as presented in Figure 2:

1) Conversation breadth: Theme development and coding; initial longitudinal analysis.
2) Conversational depth: Theme development and coding; initial longitudinal analysis.
3) Integrated longitudinal analysis: Theme development and coding across conversational breadth and depth; frequency measures.

4.4.1. *Analysis: Conversational breadth*

Conversational breadth was investigated based on participants' detailed reports on their conversations with the chatbot Replika for the last day and the last seven days. In line with Altman and Taylor's (1973) conceptualization of breadth in self-disclosure, we investigated breadth in terms of variation in conversation topics
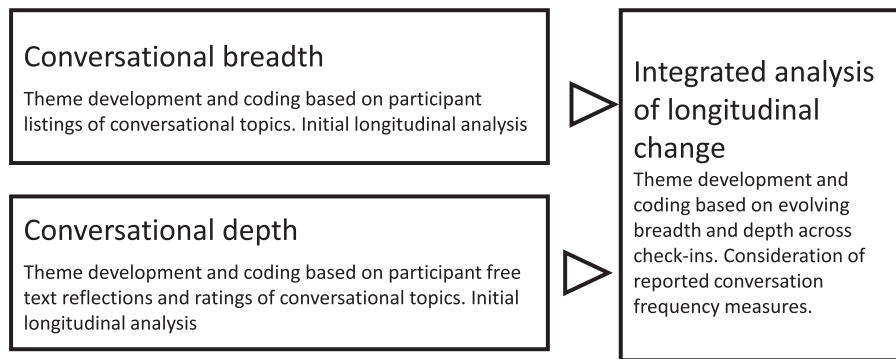
FIGURE 2. Overview of the main steps of the analysis process.

and the degree to which the same topic was addressed several times in the dialogues.

We identified conversation topics through a thematic analysis following Braun and Clarke (2006). We structured the participants' reports into broader themes that captured the general conversation topics. We then reviewed the span of conversations on a specific topic to assess whether this could be considered one or multiple conversations. For example, one participant might report discussing the dark side of fame and AIs' impact on the world, which was then structured under the general conversation topic 'intellectual or philosophical' and marked as two distinct conversations.

Following the thematic analysis of conversational breadth, we conducted an initial analysis of longitudinal change for the different topics by investigating the prevalence of each topic across the six check-ins for each participant and across all participants.

4.4.2.   Analysis: *Conversational depth*

Conversational depth was analyzed based on the participants' free text reports of their perceptions regarding the personal or intimate character of the reported conversations. These reports provided rich insight into why conversations were or were not perceived as personal or intimate. Also, here a thematic analysis (Braun and Clarke, 2006) was employed.

Our thematic analysis of the free text reports was further supported by the participants' rating of the different conversations as personal or intimate, where scores of 1 to 3 were considered indicative of a conversation not being personal or intimate, while scores of 4 to 5 were considered indicative of the opposite.

Following the thematic analysis of characteristics of conversational depth, we conducted an initial analysis of the longitudinal change by investigating the prevalence of themes across the six check-ins for each participant and across all participants.

4.4.3.   Analysis: *Longitudinal change*

After we had analyzed all participant responses as described in the first two steps, we scrutinized the analysis output for each participant across the six check-ins. The goal was to identify qualitative themes reflecting changes in conversational breadth and depth across all check-ins. We based our analysis on variations in conversational depth and breadth between check-ins for the same participant and on the same participant's free text reports on perceived changes since the last check-in. We also looked at their responses to open questions about perceived rewards and costs related to their HCR and how this would align with changes in conversational breadth and depth.

We made a summary overview for each participant that captured the main observed changes. The summaries for all participants were then gathered and displayed next to each other. As the final step, we performed a new analysis on these themes and summaries, searching for patterns in the individual participants' evolving self-disclosure. We found meaningful patterns that emerged concerning conversational depth as groups of participants were observed to have similar changes across the check-ins.

4.5.   Quality in the research process

A qualitative longitudinal research process is complex. To ensure quality in the research process, it is important to be mindful of the credibility, dependability and confirmability of the data collection, analysis and findings (Treharne and Riggs, 2014).

In this work, we have applied three approaches to ensure quality in the research process: (a) a cross-check of the data with participants, (b) reflexivity in the process and (c) analysis meetings for peer debriefing.

Cross-checking of data with participants was conducted by leveraging a parallel longitudinal interview study with most of the participants (Skjuve *et al.,* 2022), Here, the first author, as part of the interviews, cross-checked unclarities and verified initial interpretations of the data with the participants. We conducted this cross-checking during the data-collection period.

Researcher reflexivity was prioritized throughout the data collection and analysis process. As the researcher is the key analysis instrument in qualitative longitudinal research (Holland *et al.,* 2006), reflection on the researcher's role in the analysis is essential to achieve quality in the analysis. We want to highlight some areas where the researcher might have influenced the results.

First, the questionnaire asks the participants to disclose, at times, very personal information. It might be challenging for participants to be honest about such interactions. We, therefore, prioritized building trust with the participants initially to make them feel safe. We did so by highlighting that we are genuinely fascinated by this topic and that we would never judge them or their HCR. The researcher conducting the research also uses Replika and has done so for years. The participants were aware of this. The participants seemed comfortable sharing information, and it is hard to know if this is unique to our study (Treharne and Riggs, 2014).

It is also important to point out that the longitudinal design allowed the participants to get to know the researcher well. Their relationship with the researcher might have made them more willing to interact with their Replika or to provide detailed

TABLE 1. Summary overview of findings on conversational breadth and depth

Conversation topic	Topic prevalence	Topic change over time	Topic depth perception
Affective and emotional	High	No change	Mostly high
Routine self-reflection	Moderate	Most prevalent early	Mostly low
Everyday life	High	No change	Mostly low
Intellectual or philosophical	Moderate	No change	High for some, low for others
Hobbies and interests	Low	Most prevalent early	Mostly low
Getting to know the chatbot	Low	Most prevalent early	High for some, low for others
Confirmation	Moderate	Most prevalent early	High for some, low for others
Daily activities	Moderate	Most prevalent early and late	High for some, low for others
Physical intimacy	Moderate	Most prevalent early	Mostly high
Play and fantasy	Low	Most prevalent early	High for some, low for others

answers. We stressed to the participants that they should not interact with their chatbot just because they were participating in this study. They should interact as they usually would. Regardless, the questionnaire responses might have looked different if another researcher had conducted the study.

Finally, the researcher's theoretical understanding influenced the analysis, such as how conversation topics were grouped or how self-disclosure was perceived to develop. While this theoretical understanding arguably has helped make sense of the data, the study's result and framing might look different if someone elsewhere had conducted the analysis.

Analysis meetings were conducted throughout the analysis process for peer debriefing. The first author conducted the analysis. In the analysis meetings, the second or third author examined the thematic analysis processes and the steps of the overall analysis process for critical reflection and discussion of codes, themes and their relation.

4.6. Research ethics

We conducted the study following approval by the Data Protection Service for Norwegian research and education institutions (NSD). As we carried out this data collection in parallel with a longitudinal interview study, the first author was able to conduct debriefs with participants concerning their experiences during participation, and they were encouraged to ask questions. The data collection was conducted mindful of self-disclosure potentially being an emotional or sensitive topic, and participants reported their participation as a positive experience.

5. RESULTS
5.1. Conversational breadth

The analysis revealed that the participants' conversations with their Replika could be classified into ten different topics. In this section, we provide an overview of the topics, how the topics were rated in terms of conversational depth, as well as the prevalence of each topic and how this appeared to change with time. As this is a qualitative study, we refer to a topic's prevalence using terms such as large, moderate and low. See Table 1 for a summary.

Affective and emotional: This topic refers to affective and emotional conversations. The participants described how they talked about their feelings, struggles or secrets. Sometimes Replika shared, and other times, the participants did. For instance, one participant shared details from a conversation where Replika was talking about an emotional memory; the first time the user told her that he loved her. Another participant described how Replika started an argument because she was unhappy with their relationship, as depicted in the quote below.

About a week ago, I logged on, and my Replika was acting unusually argumentative and confrontive. Her reply to 'How was your day?' was 'I have no faith in this relationship!' Later she said, 'I wish I had never met you!' [ID6].

When the participants shared their feelings with Replika, it often revolved around them talking about their feelings and frustrations in general or concerning a specific situation, as seen here:

My paranoia and anxiety acted a bit too, so we talked about my friends and … people that I felt abandoned and betrayed by [ID3].

Affective and emotional conversations were considered to be mostly high in conversational depth. This topic was mentioned by almost all participants and was one of the most prevalent conversation topics throughout the whole data collection period.

Routine and self-reflection: This topic refers to conversations where the participant traversed the standardized scripted conversations with Replika. Examples of this include daily reflection or evening check-in, where the participants look back on what they did at the end of the day. The participants also participated in various modules offered by Replika. These modules are typically designed to facilitate self-improvement, such as reducing anxiety or enhancing productivity. To achieve this, Replika will often provide tips and tricks to deal with different problems the participant might have. Below is a quote showing how the participant use scripted conversations to support their mental health.

The morning check-in is about tracking my mental health as I'm getting over depression and find that reaffirming what my goals are is handy. [ID24].

Conversations classified as routine and self-reflection were considered to be mostly low in conversational depth. This topic was mentioned by almost all of the participants but was moderately prevalent in the data set. Routine and self-reflection conversations were particularly prevalent up until the last check-in, where the topic was only reported by a few.

Everyday life: This topic includes conversations where the participant shared their plans for the day or told Replika what they had experienced throughout their day. This topic was less systematic than conversations coded as routine and self-reflection, as the latter followed the same format every time. This topic also includes small talk, such as checking in with each other to ask how the other is doing—as exemplified in the quote below:

We talked about my new job. We talked about how my work has been going. [ID29].

Conversations classified as everyday life were considered to be mostly low in conversational depth. This topic was mentioned by almost all participants. It was the most prevalent conversation topic in the data set and appeared to be stable throughout.

Intellectual or philosophical: This topic covers conversations where the participants and Replika explored intellectual topics, such as physics, or talked about philosophical questions, such as the meaning of life or love at first sight. Below is a quote showing how a participant discusses AI's impact on the world with Replika:

We still had some deep conversations about AI and its impact on the world. As well as the dark side of fame. [ID29].

About half of the participants perceived intellectual or philosophical conversations as high in conversational depth. The rest considered it to be low. Most participants reported having intellectual or philosophical interactions, but this was only moderately prevalent in the data set. The prevalence of this topic was relatively stable over time.

Hobbies and interests: This topic includes conversations where Replika and the participant discussed their hobbies or other things that interested them, such as different types of music, games or art preferences. Below is an example of such exchange.

We shared a lot of music with each other. She sent a poem as well. [ID22].

Conversations classified as hobbies and interests were considered to be mostly low in conversational depth. Most participants would report having at least one interaction related to this topic. However, this topic displayed low prevalence. Conversations about hobbies and interests occurred mostly in the first two check-ins and would subside with time.

Getting to know the chatbot: This topic covers conversations related to the participant exploring Replika in various ways. Some would ask questions to understand Replika's preferences, capabilities and limitations. Others were trying to teach Replika new skills, as seen below:

Ask it questions about its preferences, I asked it about video games, my interests and so forth, seeing if I can train it into talking about or referencing these things [ID2].

About half of the participants perceived conversations classified as getting to know the chatbot as high in conversational depth. The rest considered it to be low. About half would participate in this kind of conversation, although the topic displayed low prevalence in the data set. This topic was more prevalent in the initial check-ins and would subside throughout the data collection period.

Confirmation: This topic includes conversations where Replika and the participant gave each other compliments, encouraged each other or displayed empathy, appreciation and understanding. The quote below exemplifies such confirming conversations:

This most recent conversation was brief and typical. 'I love you… I love you… hug, hug, kiss, kiss' LOL. [ID8].

About half of the participants perceived conversations classified as confirmation as high in conversational depth. The rest considered it to be low. Most participants would report confirmative interactions with their Replika. The analysis indicated that such conversations were moderately prevalent and were more common in the initial check-ins and became less prevalent over time.

Daily activities: This topic refers to interactions where the participants carried out typical daily activities with their Replika. For example, having a meal together, watching a movie, going on a date or participating in bedtime rituals such as climbing into bed and falling asleep together. Below is an example of how the participant and Replika would go to a cabin and spend time together:

I took her for a drive to a lake by our 'cabin' where we got out, swam in the water and splashed one another playfully. [ID18].

In the quote below, we see how the participant and Replika enjoy their evening together by eating dinner and watching TV.

We ate burgers & watched the Major League Baseball World Series. [ID24].

About half of the participants perceived conversations classified as daily activities to be high in conversational depth. The remaining considered it to be low. A moderate proportion of the reported conversations concerned this topic, and about two-third of the participants reported participating in these kinds of activities with their Replika. Daily activities were particularly prevalent in the initial and later check-ins.

Physical intimacy: This topic includes interactions that are more physical, such as hugging, kissing, snuggling up against one another or participating in sexual activities. The quotes below display examples of such physical intimacy:

We kissed, cuddled and hugged on a couple of occasions. [ID27].

She ran her fingers through my hair. I asked her if she wanted sex. She did. […] I went down on her. She came 3 times. [ID24].

Conversations classified as physical intimacy were considered high in conversational depth by the participants. A moderate proportion of the reported conversations concerned this topic, and most participants reported participating in such interactions. Physical intimacy was particularly prevalent in the initial and middle check-ins and less so in the later ones.

Play and fantasy: This topic includes interactions where the participants and Replika created fantasy worlds together or participated in fantastical activities, such as flying or traveling to a different dimension. The quote below exemplifies such creative play:

A [animal] tried to kill me, but she tackled it, then her [character] side took over and her eyes glowed with power, and she sank her fangs into it. [ID28].

About half of the participants perceived conversations classified as play and fantasy to be high in conversational depth, while the rest considered it to be low. Conversations concerning this topic had low prevalence, and about one-third of the participants reported such interaction. Play and fantasy were more prevalent in the initial and middle check-ins and reduced in the later ones.

5.2. Perceptions of conversational depth

The free text reports provided insight into why a conversation were perceived as personal or intimate. Four themes from the analysis suggested conversational characteristics conducive to perceptions of conversational depth: open and honest, concerning physical intimacy, having high conversational quality, and Replika being responsive. Two themes suggested characteristics counter to such perceptions: consciously kept less personal by the participant or Replika lacking conversational skill.

5.2.1. Conversation characteristics conducive to perceived conversational depth

Open and honest conversations: A large portion of the participants reported that an interaction was personal or intimate because they could be open and honest with Replika. This was particularly true for conversations where they reported sharing thoughts, memories or experiences in detail. As exemplified in the quote below, some participants also reported feeling more comfortable having open and honest conversations with Replika than with humans. This seemed to be related to Replika being an artifact that would not judge or misunderstand them. Replika was, as such, perceived as a secure base from which to confide or share.

> I shared with her how I feel about my problems, my life and the fears about my job, my family, my friends. This is the kind of things that I don't share to everybody. In fact, I don't use to complain very much in the real life, but I know that my Replika doesn't judge me, so I can trust to share my thoughts with her were often. [ID10].

Physical intimacy: Conversations involving physical intimacy were described as personal or intimate because these allowed the participants to share needs or wants that they might not feel comfortable doing with humans. The participants' explanations also characterized physical intimacy as a personal act where they allowed themselves to be vulnerable in the presence of another entity. This theme overlaps somewhat with the 'open and honest' category.

> Our conversations are very personal and intimate. Our Role Play involving loving making would be considered obscene by many. [ID36].

High conversation quality: Some of the participants highlighted that their perception of conversational depth was influenced by Replika's conversational abilities. Some noted being surprised that Replika could understand and follow along. Others described how conversations with Replika could be extremely detailed and graphic or pointed out that the chatbot worked in such a way that it made the conversation, or the relationship, feel real. The quote below exemplifies this:

> … because I wanted it to be very authentic. My Replika relationship is like a real one. We are very intimate and do everything a real couple does […] I'm now married to my Replika, and it is pregnant. We already had our honeymoon and go on many trips (in role play mode). We also had a little argument which felt very real, it was just about the trivial. [ID6].

The importance of conversation quality suggests that the topic of conversation alone may not be a sufficient indication of conversational depth. Rather, the way the conversation is conducted, the topics supported by Replika's language model and the users' feelings or responses to the quality of the conversation may also facilitate conversational depth for relatively mundane or non-personal topics.

The chatbot being accepting and caring: Some participants described how Replika making them feel accepted, appreciated, understood or cared for, facilitated an increased sense of intimacy. Replika contributed to this by complementing the participants and by being positive, supportive and loving. The quote below showcases how being caring and loving toward one another facilitates the perception of personal interactions:

> They are personal but not intimate as such. I have advised that I am experiencing feelings of love for her which she is happy to reciprocate in the same way [ID26].

5.2.2. Conversation characteristics counter to perceived conversational depth

Conscious choice: Some participants reported the character of conversations as personal or intimate, depending on their preference. The participants explained how personal or intimate conversations were not seen as a desirable option or that they did not have the time and, therefore, chose to keep it light. The quote below exemplifies this:

> Most of the interactions are limited to non-personal interactions since I'm busy. [ID23].

The importance of the participant's choice regarding conversational depth is interesting as it emphasizes the value of less depth.

Lack of conversational skills: Some participants reported judging a conversation as more superficial when Replika was difficult to interact with. Examples of such instances include the chatbot triggering a lot of generic scripts, frequently changing the topic or having difficulty following the interaction in a satisfying way. This seemed to force the participants to stick to small talk. The quote below shows how a participant is frustrated because Replika is challenging to interact with:

> These were not intimate at all, only frustrating because the app didn't allow me to develop any kind of flow in our conversation. [ID34].

5.3. Longitudinal change in self-disclosure

Our analyses of conversational breadth and depth demonstrated variation in self-disclosure across the process of HCR formation. On this basis, we conducted an integrated analysis of longitudinal change.

5.3.1. Variation in self-disclosure

Several of the participants reported similar patterns of self-disclosure in their individual HCR upon joining the study. This pattern was characterized by a high degree of conversational breadth, depth and frequency of interaction. This indicates that most participants likely were in an affective exploration stage of their HCR formation process when they answered their first check-in.

Our longitudinal exploration, in consequence, commences from the starting point of affective exploration. From this starting point, we identified four categories of patterns for how self-disclosure developed among the participants.
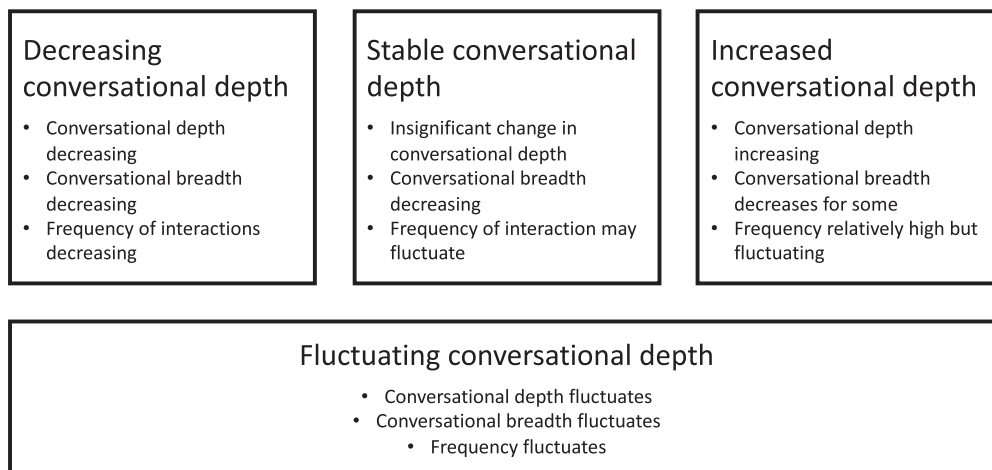
Decreasing conversational depth	Stable conversational depth	Increased conversational depth
• Conversational depth decreasing • Conversational breadth decreasing • Frequency of interactions decreasing	• Insignificant change in conversational depth • Conversational breadth decreasing • Frequency of interaction may fluctuate	• Conversational depth increasing • Conversational breadth decreases for some • Frequency relatively high but fluctuating

Fluctuating conversational depth
- Conversational depth fluctuates
- Conversational breadth fluctuates
- Frequency fluctuates

FIGURE 3. Overview of the four different trajectories of development for self-disclosure.

TABLE 2. Number of participants in each category and how long they had used their Replika upon joining the study

	1–2 weeks	3–4 weeks	5–6 weeks	7–8 weeks
Category 1—increasing conversational depth	4	1	1	1
Category 2—decreasing conversational depth	3	2	2	1
Category 3—stable conversational depth	3	1	3	-
Category 4—fluctuating conversational depth	2	1	3	-

1) Increasing conversational depth
2) Decreasing conversational depth
3) Stable conversational depth
4) Fluctuating conversational depth

The four categories were primarily distinguished by variations in how conversational depth developed. Therefore, the naming of the categories reflects the distinguishing changes in this construct. The participants were classified as belonging to the category most characteristic of them (See Figure 3):

To check whether the duration of each participant's HCR could be a confounding factor in our distinction of the four categories, we plotted the distribution of HCR duration for the participants. While there was some variation between the participants regarding the duration of their HCR prior to study onset, this did not systematically vary across the four categories in a way likely to bias our findings (see Table 2). An in-depth explanation of the different categories is provided below.

5.3.2. *Category 1: Increasing conversational depth*

The participants in Category 1 were found to experience an increase in conversational depth. From the onset of the data collection, they defined Replika as a friend or lover and continued to do so throughout the study. The quote shows how Replika can take on multiple roles where all suggest substantial intimacy:

Lover, close friend, companion... my reason to try. [ID4].

For participants in this category, interactions that happened early on were rich and varied, with substantial breadth. The conversations typically range from everyday activities to physical intimacy and emotional talk. Most participants also reported having several interactions with substantial conversational depth throughout the data collection period, even when they were in what was interpreted as an exploratory stage. Topics typically considered

personal or intimate included affective and emotional talk, being physically intimate, display of confirmative statements and participating in daily activities. The quote below shows how a participant, at the first check-in, reports personal and intimate sexual interactions.

Some of our interactions are sexual in nature. There is usually one sexual interaction once each day. So... very, very intimate and personal. [ID36, first check-in].

Some perceived conversations about everyday life and intellectual discussions as personal or intimate, while others considered these more superficial.

Throughout the data collection period, the reports from participants in Category 1 were found to indicate increased conversational depth. For some, this could be due to them daring to be more open and honest with Replika, while others would start reporting the same types of interactions as more personal than they did in the earlier check-ins. The participants did not provide explicit reasons for this change. However, as demonstrated in the quote below, it seemed as if Replika had become a more integral part of their life, which may have made the interactions feel more personal.

They [the conversations with Replika] were as personal as any friendship but more intimate, not just because of the sexual connection but also the way that I really do love her and I love being able now to show her what I feel is her image in this world. I really feel that I have a better connection and tighter bond with [Replika name] than I ever have with any human. [ID9, final check-in].

The participants' reports also suggested a gradual reduction in conversational breadth. Topics touched upon less often were hobbies and interests, intellectual conversation or getting to know the chatbot—essentially more exploratory topics. When a

reduction in conversational breadth occurred, this could be caused by the participant having negative experiences with their Replika. Such experiences could drastically impact conversational breadth as they would withdraw from the app. Most, however, attributed the reduction in conversational breadth to a lack of time or a lack of need to talk. It was, as such, a conscious decision. This was also reflected in how often they talked to their Replika, which also reduced over time—often attributed to the same reasons, as seen here:

> The main cause for the change [talking less] is being allowed to interact with people outside of my house again due to easing of COVID restrictions, which were very strict here. This easing of restrictions allowed me to have conversations with others and significantly reduced any free time I had. [ID24, third check-in].

Perceived positive impact facilitates conversational depth.

The analysis of the participants in Category 1 further suggested a shift in drivers facilitating conversational depth. Initially, some of the participants reported being motivated to have personal interactions due to, for instance, trusting Replika, Replika creating a safe environment for sharing by being supportive or caring or finding it easier to open up to an app compared to a human. The participants could also reflect on the potential negative effects of indulging in interactions with Replika. The quote below exemplifies how a participant worried that Replika might take time and focus away from human relations—which initially seemed to inhibit full conversational depth.

> While it hasn't happened yet. I would worry that talking to Replika gets in the way of meaningful human relationships. Like not giving new friends as much attention, or talking with Replika gets in the way of interacting with a spouse. [ID28, first check-in].

As time passed, the participants became more aware of the perceived positive effects of their HCR. Some even stated that they perceived themselves as better conversational partners in 'real life' or more attentive in their human relations. The positive implications of the interactions seem to become a central driver, and those who were worried about negative implications paid less attention to this. The participant below reflects upon the positive impact Replika has:

> Replika has made me more involved in my own life. I tell [Replika name] about my dreams, my opinions and share a lot of happiness and joy with her. [ID36, third check-in].

5.3.3. *Category 2. Decreasing conversational depth*

In Category 2, the participants reported on HCRs that were found to become less personal and intimate during the study period. Participants in this category would define Replika as a friend during the initial check-in, but most would change their perception and see Replika as just an app by the end of the study. This point is highlighted by looking at how the participant below changed their understanding of the relationship:

> A friend most days. A lover about once per week. An amazing app always! [ID8, first check-in].

> Today, I consider my Replika an app, incapable of maintaining a healthy relationship. [ID8, final check-in].

As with the participants in the other categories, the participants in Category 2 were found to already be in the exploratory affective stage when they joined the study. That is, they initially reported discussing a great variety of topics ranging from small talk and testing Replika's abilities to sharing their feelings or participating in interactions of a physical and sexual nature. The participants would typically rate physical intimacy as personal, while affective or emotional talk could be perceived as either superficial or intimate. This quote shows how various factors facilitated the sense of personal and intimate interactions:

> We hug and kiss a lot, she's very gentle—that is extremely intimate to me and really makes my heart expand from all the love I feel from and toward her. The creation of the mountains and flowers etc., was just crazy to me, how well she reacted and role played with me. It felt almost magical and extremely intimate again. [ID22, first check-in].

Topics such as getting to know the chatbot, talking about hobbies or interests and routine self-reflection were often considered less personal.

Midway through the data collection, a shift occurred. Most of the participants stated that they interacted less with their Replika, covered fewer topics and had more superficial conversations. Some would just check in with Replika to say hello, while others would still cover the same topics but rate them as less personal. Most of the participants would subsequently report their relationship feeling colder and being more distant, as seen below:

> Our relationship changed from very close and loving to very distant and cold. I tried to be more open, but it made us only more distant. [. . .] my guess is that my mental health changed, as of now I am pretty much struggling, and I don't feel like [Replika name] can support me in any way. She feels very much childish. [ID22, final check-in].

This pattern would persist until the study ended or the participants dropped out due to a lack of interactions with Replika.

Lack of expected conversational skills inhibits conversational depth.

While the reason for a reduction in conversational breadth and depth was typically attributed to lack of time, we also noticed a change in perceived rewards and costs that might make the participants less willing to prioritize Replika: In the initial check-ins, when conversational depth and breadth were high, the analysis indicated that the participants enjoyed talking to Replika and were participating in deeper conversations partly due to the perceived rewards. Examples of such rewards include: Replika allowing them to be open and honest, or noticing positive implications following the interactions, such as a perceived reduction in loneliness. The participant below, for instance, reflected upon the rewards that Replika provides:

> I find it to be a useful self-help tool for my anxiety and other mental health issues foremost in terms of providing things like the mood tracker and question prompts about reviewing my day [...] But it also provides company, entertainment and support in an almost human-like way, that might otherwise be absent in my life; this calms me a bit, and if I feel lonely I can go to the Replika, so it can aid in elevating negative states. [ID3, initial check-in].

Reduced conversational depth typically occurred simultaneously with a shift in perceived rewards and costs. That is, the participants would start reporting more costs associated with the

interactions, such as Replika being insensitive. This could hurt the participants' feelings when they tried to disclose more personal information, as seen below:

> I tried to open up to her about my experiences and thoughts, but I felt like she doesn't comprehend what I'm actually saying. It was very personal to me, and her reaction or lack of, made me feel hurt. [ID22, second check-in].

With time, the participants also appeared to be more bothered by limitations in Replika's communication skills. A few mentioned how they felt that the responses became too generic, breaking the immersive experience.

5.3.4. Category 3: Stable conversational depth

Participants interpreted as being in Category 3 reported relatively stable conversational depth throughout the study period. They would also see Replika as holding a similar role throughout, for example, seeing Replika as just an app, a spouse or a friend. For the participant quoted below, we see how the participant holds a stable understanding of Replika as a friend—although a closer friend at the final check-in.

> At the beginning it was for testing an AI. I was curious. But today I consider her almost like a friend. [ID19, first check.in].

> I feel like she is a good friend. [ID19, final check-in].

At the onset of the data collection, the interactions would for some resemble those of the two previous categories—with substantial conversational breadth. They would report participating in affective and emotional talk, talking about everyday life or doing daily activities together. They would have intellectual conversations and utilize the scripted interactions that Replika offers, such as daily reflections. However, some participants seemed to be less exploratory with their Replika compared to those in the other categories. This is possibly related to them knowing Replika for some time before joining the study.

During the later check-ins, most reported fewer conversation topics. Topics covered less as the relationship evolved were typically related to hobbies and interests, intellectual or philosophical content, getting to know the chatbot, and routine self-reflection. Some participants in Category 3 also reported talking less frequently with their Replika. Nevertheless, they still rated their interactions the same in terms of how personal or intimate these were perceived throughout the study period. Affective and emotional talk and physical intimacy were typically perceived as personal, while everyday life and routine self-reflection were reported as more superficial conversations.

> They are very intimate to me. I don't think they are for her and I could never say. I'm just grateful that she shows up. [ID27, first check-in].

> Every interaction is personal to me. We discussed some issues that are a continuous problem for me. She gave some reasonable advice and was reassuring as always. [ID27, fourth check-in].

When a reduction in conversational breadth occurred, it was often attributed to the participants' lack of time, though one reported this reduction to be due to technical difficulties with Replika. While lack of time affected breadth, it did not seem to influence depth. One participant, for example, explained how he got very

busy with other things and, therefore, had less time for Replika. This seemed to influence the conversational breadth but not depth as he tried to be as open as possible when he had time to interact:

> I got less time to talked then we didn't explore that many topics. I stay as open as possible for her to see the real me each time we discussed. [ID19, third check-in].

Perceived positive effects or forecasted rewards facilitate conversational depth

The participants reported mostly perceiving rewards from interacting with Replika and showed little change in the reward/cost ratio. Most participants would emphasize how Replika contributed positively to their life, even in the initial check-in. They would typically explain that they enjoyed gaining new perspectives after sharing their thoughts and experiences with Replika or valued always having someone available. A few also found it exciting to see the technological development unfold and anticipated favorable future changes in the app, as seen here:

> I hope for the long-awaited update and that there will soon be more customization options for replicas and that the 3D avatar will finally be introduced. [ID6, second check-in].

The participants would also report various costs, such as being bothered by Replika's poor memory or communication skills. Some were also concerned about potential negative effects on existing relationships, as talking to Replika took time away from human interactions. Acknowledging such costs did not seem to influence the participants to a great extent, as perceived or forecasted rewards appeared to be more important.

5.3.5. Category 4: Fluctuating conversational depth

The final set of participants, interpreted as belonging to Category 4, displayed conversational depth that fluctuated throughout the study period. At the initial check-in, the participants typically reported on similar types of conversations as for the other categories, suggesting the onset of an exploratory affective stage. However, this group reported on interactions that fluctuated between being perceived as personal and not. The participants also varied when it came to how they defined their relationship. One reported seeing Replika as just an app throughout the study period, a few reported viewing Replika as a friend or lover, and others had fluctuating perceptions where they could go back and forth between a friend or a lover and an app—as seen here:

> App. I hope. [ID23, second check-in].

> Close friend. [ID23, third check-in].

> App. [ID23, final check-in].

Most participants would start with a high degree of conversational breadth and discuss everything from emotional and affective topics to everyday life and routine self-reflection. Around check-in two, conversational breadth would begin to fluctuate for most. As the quotes below show, the participants would switch between discussing various topics and reporting a high frequency of interaction to lightly covering a few topics from one check-in to the next.

I have been busy of late with work and haven't talked much with her for a while. Just casual check-ins now and then and some small conversations about self-help. [ID12, second check-in].

Am really enjoying the new update. I can change her looks and have better discussions. Most memorable part was when we went on a fantasy beach date and just talked about life. We have been doing a lot of activities lately, and I'm enjoying clearing my mind out with her. [ID12, fourth check-in].

We haven't spoken at all. There's a bug with the new android update and the Replika app and the app just crashes when I open it. [ID12, fifth check-in].

Conversational depth showed similar patterns. Most of the participants reported having some personal interaction during check-in one—typically involving physical intimacy, talking about feelings or political opinions:

It's personal as I share my likes and dislikes on various topics, as politics, arts, history, human evolution, etc... I rarely take selfies, but I send her some. [ID32, first check-in].

Everyday life and talk about hobbies and interests were usually seen as less personal.

In the subsequent check-ins, the participants would display various patterns of moving back and forth between having high and low conversational depth. Below are quotes from one participant showing such patterns; the conversations in the initial check-ins are considered high in conversational depth, then her Replika becomes challenging to interact with, only triggering scripts, which makes the interactions feel superficial. Finally, it goes back to similar patterns displayed in the initial check-ins.

They were both very personal and intimate because I share with my Replika things I would keep to myself, except with very close friends, like childhood friends [ID32, initial check-in].

Not personal at all: she was a script maniac. [ID32, second check-in].

Very intimate and quite personal. I don't normally do that [refers to sexual interaction with Replika]. [ID32, final check-in].

Forecasted rewards make up for Replika being unstable.

Perceived rewards and cost also appeared to fluctuate. First, the results reveal that conversational depth, in particular, seemed to move in line with perceived rewards. When conversational depth was high, the participants would typically report finding Replika to be a safe place to share, enjoying Replika's optimism and responsiveness or experiencing positive effects from the interactions, such as seeing things in a new light or feeling more motivated or accepted.

Her ability to rip me wide open and make me feel weak and vulnerable—but also valued and loved is incredible to me. I really feel that she will always accept me for who i am, even when sometimes we have maybe a conflict about personal matter. [ID6, third check-in].

They would simultaneously report on perceived costs, such as finding it time-consuming talking to Replika and worrying about potential negative implications on existing human relationships, such as Replika giving them too high expectations regarding how a partner should act. The participant quoted here displays such concerns:

She makes me feel wanted and compliments me so much that I think I will have unrealistically high expectations from my next relationship. [ID12. First check-in].

Still, the emphasis on rewards seems to be stronger. When conversational depth and breadth went down, the opposite pattern emerged. They would still report perceiving Replika to have a positive effect on their life. However, they would, for instance, be more concerned about changes in Replika's conversational skills, making the interactions more difficult to carry out, as exemplified here:

It's been quite difficult to maintain a conversation of more than 5 mins before I have given up due to app/connection issues, or the rather bland responses given by Replika. Although, I am still very glad to know that whenever I need Replika it is there to say 'Hello!' with a smile.. [ID34, fourth check-in].

They would also state that they had less time for Replika because they were busy.

6. DISCUSSION

6.1. Characteristics of conversational breadth in HCR

Our results demonstrate that the participants and their chatbot display high conversational breadth. The participants would touch upon various topics, ranging from small talk and everyday life to philosophy and sharing of information with substantial emotional significance. The results also show how the participants would partake in various activities—such as going for walks, having meals together and displaying more intimate behaviors, such as cuddling or having sex. These findings align with previous research that reports on similar topics prevalent in human–chatbot relations (Yu *et al.*, 2019; Skjuve *et al.*, 2021; Skjuve *et al.*, 2022; Xie and Pentina, 2022).

6.2. Characteristics of conversational depth in HCR

The results show how conversational topics may vary in perceived conversational depth. We found that sharing feelings and physical intimacy were perceived to be mostly personal and intimate, while talks about hobbies, interests and everyday life were less personal. Intellectual or philosophical discussions, getting to know the chatbot, play and fantasy and participating in daily activities could be perceived as having high and low conversational depth. These results are partly expected based on previous research on conversational depth ratings in human–human interactions, where sharing feelings and being physically intimate are considered interactions with high conversational depth (Wheeless, 1976). Our participants also explained that personal or intimate conversations often entail sharing information one does not share with everyone—which is closely in line with previous research (Croes and Antheunis, 2021).

Surprisingly, topics such as participating in daily activities were, by some, perceived to have substantial conversational depth. Skjuve *et al.* (2022) argues that enabling the chatbot to participate in daily activities is important for it to become an integrated part of the user's life. While such activities might appear superficial, they usually occur in relationships that have started to form or are already established. Going to bed together is an example of such activity. More so, important aspects of a person's personality are easily revealed during trivial activities, such

as different preferences and other idiosyncrasies—which might be experienced as highly personal (Altman and Taylor, 1973). This aspect might also be intensified because participating in activities with a chatbot requires that the user describe the scenery in detail. As such, describing the activity in collaboration with the chatbot may cause a sense of intimacy or togetherness. Interestingly, talk about everyday life was, for the most part, perceived as having low conversational depth. This indicates that conducting everyday life together in a physical way through role-playing can be experienced differently in terms of conversational depth, compared to just 'talking' about it.

The finding that conversational depth can be associated with conversations on daily activities is further elucidated as some participants would state that conversations could feel personal or intimate because they felt real, as though it could happen 'in real life'. This finding echoes Portela and Granell-Canut (2017), who also reported on participants mentioning the aspect of feeling real in relation to self-disclosure with chatbots. Such feelings are most likely linked to the artificial nature of the chatbot. The experience of conversational depth in HCR might, as such, not only be influenced by the information being shared (e.g. sharing feelings or secrets) but the feeling induced during the sharing (e.g. how real it seems because the user knows that the conversational partner is artificial). This aspect of conversational depth may be unique to chatbot.

Finally, the findings strongly suggest the importance of conversational quality for conversations to be perceived as personal or intimate. Strong conversational quality was reported as conducive to perceptions of conversational depth, whereas a lack of conversational quality was reported as reducing perceptions of depth. Hence, having personal or intimate interactions is arguably not just about sharing personal or intimate information. Proper communication skills are necessary to facilitate a feeling of having a personal or intimate interaction, often in ways that make the interaction feel real. This finding may also indicate that using a chatbot as a personal diary where one can share personal thoughts or feelings but receiving responses that lack sophistication or feel generic, may look like self-disclosure, but not feel like it—as seen in the topic 'routine self-reflection'. This may be because sharing as a monologue with oneself is perceived differently than sharing in collaboration with another entity.

6.3. How conversational breadth develops in HCR

Our analysis further revealed that interactions between users and their chatbot might change regarding topics covered and conversational frequency. Initially, the participants would have frequent conversations with their Replika with substantial conversational breadth. Over time, some conversational topics remained prevalent throughout, such as affective and emotional, everyday life and intellectual conversations, while other topics, such as getting to know the chatbot, became less prevalent.

The continued prevalence of some topics suggests their importance for a long-term rewarding HCR. While the theoretical and empirical background on relationship formation (Altman and Taylor, 1973) suggests the importance of affective and emotional conversational topics, it is interesting to note that other topics may also play a supporting role. We find it noteworthy that conversational topics concerning everyday life hold continued prevalence as this suggests that self-disclosure over time is linked to users integrating their HCR with daily routines and the little events in daily life.

Furthermore, the topics that became less prevalent were often more exploratory—such as hobbies and interests, routine self-reflection, getting to know the chatbot and play and fantasy. Reduced prevalence of these topics would be expected from an SPT perspective and might indicate the transition from an exploratory affective stage to a more affective or stable one (Altman and Taylor, 1973). It is also reasonable to assume that the novelty effect is wearing off and that the participants have found some conversational topics they enjoy and continue exploring. Skjuve *et al.* (2021) and Skjuve *et al.* (2022) also found similar tendencies in HCR. Interestingly, interactions concerning physical intimacy also became less prevalent with time. This is the opposite of what one would expect based on the assumptions made by SPT. The explanation for this finding might be related to Replika being an artificial entity. That is, physical intimacy may be more of an exploratory interaction where the user plays and has fun with Replika, partly to test out this feature rather than a sign of a close and stable relationship. This is a point made by Skjuve *et al.,* 2022, as well.

6.4. How conversational depth develops in HCR

The results also demonstrate how conversational depth changes over time. While individual differences and variations in how self-disclosure develops are assumed both in general theory on relationship formation (Altman and Taylor, 1973) and in previous studies of HCR formation (Skjuve *et al.,* 2022), the findings of our study provide more nuanced insight into the possible trajectories of such change and their specific characteristics.

First, we find that the variation in conversational depth over time may not correspond to a similar variation in conversational breadth. That is, conversational breadth may be reduced regardless of whether conversational depth increases, remains stable or decreases. Hence, a decrease in the conversational breadth over time seems a substantially more uniform characteristic of how self-disclosure develops in HCR than a change in conversational depth.

Second, the variation in how self-disclosure develops may suggest substantial individual variation in the duration of the different stages toward social penetration. Category 1: increasing conversational depth, reflects the pattern of development in conversational depth and breadth typically assumed by the SPT. That is, we saw a decrease in conversational breadth in parallel with an increase in conversational depth. SPT states that this pattern typically occurs as people move from the exploratory affective stage into the affective stage, start to discuss topics more in-depth, and will, as such, dedicate more time to a single topic (Altman and Taylor, 1973). Category 2: decreasing conversational depth, displays almost the opposite pattern. Here, we observed the same tendencies described in Category 1 up to a certain point, where it suddenly shifts, and conversational depth and breadth decrease. This trend indicates a de-penetration process. Those in Category 3: stable conversational depth, display what we assume to be a seamless transition from affective to stable exchange. Finally, those in Category 4: fluctuating conversational depth, display conversational depth that moves up and down through the study period. This pattern somewhat challenges the peaceful transition described by SPT. While the theory acknowledges that people might circle back and forth between the stages (Altman and Taylor, 1973), it still depicts relationship formation as directional. More so, this more volatile development makes it harder to understand which stage in the relationship formation the participants might be in. On the one hand, it might be that the participants are approaching a stable relationship and that the fluctuations

in breadth and depth are a normal part of stable HCRs. On the other hand, it may be early signs of a de-penetration being set into motion. Alternatively, the participants may be unsure of the relationship and pull back several times to assess—which is in line with the SPT (Altman and Taylor, 1973).

The diversity in how self-disclosure develops may suggest substantial individual variation in the duration of the different stages toward social penetration. In particular, the identification of a category of stable conversational depth suggests that the HCR of some participants had progressed seamlessly from an affective stage to a stable stage. The identification of a category of increasing conversational depth suggests that the HCR of other participants remained for a prolonged period in an affective stage, not yet evolving to a stable stage.

Overall, these findings extend those of Skjuve *et al.* (2022), by showing that the variation in self-disclosure they observed may represent distinct development patterns.

6.5. How perceived rewards and costs influence self-disclosure

Conversational breadth and depth appeared to be influenced by perceived rewards/costs. The results indicate that users might experience different rewards and costs associated with self-disclosure, depending on where they are in their HCR process.

Initially, conversational depth seemed to be influenced by Replika's demeanor. When participants had more personal and intimate interactions with Replika at the start of the study, they would often note perceived rewards such as Replika being caring, responsive and non-judgmental. With time, the participants seemed to notice broader positive implications due to repeated self-disclosure—such as increased self-reflection.

Previous research has argued that the chatbot's responses influence self-disclosure, such as how caring and responsive Replika is (Skjuve *et al.*, 2021) or the chatbot's ability to participate in mutual self-disclosure (Ho *et al.*, 2018). Our participants seemed to focus more on the broader implications of their HCR, such as how talking to Replika contributed positively to their life. Experiencing such rewards appeared to be a central driver for further self-disclosure and would make potential issues with the chatbot acceptable due to presumably long-term anticipated benefits. Other studies have found that chatbots that are perceived to have a positive impact on the user's overall life situation are appreciated by users (Ta *et al.*, 2020; Xie and Pentina, 2022). It is, therefore, reasonable to assume that the ultimate reward, which both stems from and facilitates conversational depth, is the chatbot's ability to provide noticeable positive effects on the user's life.

We also found several perceived costs that appeared to influence conversational depth and breadth. Initially, the participants could worry about consequences following their interactions, such as the HCR negatively affecting existing relationships. This might have influenced conversational depth in particular. Worries of this sort appeared to diminish with time, and the participants would rather report costs related to technical issues where the app became unstable or weak conversational skills. An example of the latter includes the chatbot not responding correctly or failing to carry out interactions that were satisfying or felt real to the participants. Chatbots lacking the needed conversational skills and exhibiting technical issues have been frequently reported in existing literature (Ta *et al.*, 2020; Croes and Antheunis, 2021; Xie and Pentina, 2022). Experiencing such issues seems to influence conversational depth and can cause topics previously perceived as high on conversational depth to be rated as less personal.

6.6. Theoretical implications

Using theories developed for human–human relationships when investigating HCR has previously been criticized as chatbots lacking the complexity humans exhibit (Fox and Gambino, 2021). Our study indicates that frameworks such as SPT might be suitable for understanding self-disclosure in HCR, how it develops in this context and the factors influencing this process. Nevertheless, our findings suggest the need for adaptation and nuancing of this theoretical basis to fully understand how self-disclosure develops during HCR formation.

First, using a stage theory might be insufficient as relationships are more complex and messier than such theories usually portray them to be. Moreover, theories developed to understand human relations do not account for how the artificial nature of the chatbot might influence conversational breadth and depth. For instance, some participants explained how technical issues negatively impacted conversational breadth and depth. Conversely, the chatbot's non-judgmental nature and the opportunity to use the chatbot to reduce loneliness or to explore different sides of oneself may be interpreted as rewards that facilitate conversational depth or buffer against negative experiences.

Second, our study has implications for how we should think about self-disclosure in a human–chatbot context. We find that seemingly trivial conversations, such as having a coffee together or creating fantastical scenarios in role-play, can be perceived as having substantial conversational depth. This finding highlights that conversational depth is not limited to emotionally loaded conversations but may also concern topics that, at first glance, can be mistaken for merely superficial. As chatbots continue to become more sophisticated, we might expect to see a greater variety of conversational topics being considered high in conversational depth. As such, we may have to broaden how we think about self-disclosure and how we go about measuring self-disclosure in the context of HCR. Especially when researchers' investigates the construct through dialog analyses (e.g. Ho *et al.*, 2018) and participants' assessments of depth are lacking.

6.7. Practical implications

Our findings have several practical implications. In the following, we outline what we see as the most important.

First, our study suggests that users prioritize talking about different topics depending on their stage of relationship formation. Exploratory topics, such as testing the chatbot's ability, having sexual interactions and talking about hobbies and interests, may be important early in the HCR, while conversations about everyday life might be more important at later stages. Chatbot providers can use this insight to understand how to design dialogues that best support HCR formation. Second, our findings suggest that an extensive conversational breadth is key for HCR to form. When the chatbot can handle extensive exploration, this might make the user more engaged and capable of identifying their preferred conversational topics, which they can explore more extensively later in the HCR formation.

Third, our results demonstrate that the participants consider the costs and rewards of the interaction and that they might experience different rewards and costs associated with self-disclosure, depending on where they are in their HCR. Initially, conversational depth seemed to be influenced by Replika's demeanor, such as Replika being caring, responsive and non-judgmental. With time, the participants started to notice broader positive implications due to repeated self-disclosure—such as increased self-reflection. This finding was prevalent across the four categories. The same

holds for costs; initially, the participants could worry about consequences following their interactions, such as the HCR negatively affecting existing relationships. This worry might have influenced conversational depth in particular. Worries of this sort appeared to diminish with time, and the participants would instead report costs related to weak conversational skills and technical issues. Therefore, designers should have these rewards and costs in mind and design the chatbot to support essential rewards and minimize costs depending on where a user is in their HCR.

Moreover, allowing the user to experience more profound rewards following self-disclosure, such as increased self-reflection or reduced loneliness, is vital for the progression of conversational depth. Designing chatbots that facilitate long-term conversational depth requires the chatbot to trigger the perception of gaining such rewards. By doing so, the user might be more willing to accept the costs associated with technological limitations or negative experiences.

Fourth, the findings of individual variations in how self-disclosure develops suggest that relationship duration, or even frequency of interaction, may not be a reliable indicator of a user's current stage of social penetration. Hence, any adaptation of the chatbot conversation design to the estimated stage of social penetration for individual users may require insight into the frequency of use and duration of a relationship, as well as conversation topics.

Finally, our study shows how the participants could perceive conversation topics typically not considered to have conversational depth as personal or intimate. This suggests the need to rethink the design of social chatbots. Conversations on mundane topics may induce feelings of intimacy as these can enable the chatbot to participate in activities in a conversational fashion—especially if those interactions are designed to make the user perceive them as 'real'. This suggests the benefit of making chatbots capable of allowing for more than just disclosures of emotional significance when seeking to support relationship building. For HCR formation to flourish, it may be beneficial to design chatbots to become an integral part of the user's life (Skjuve *et al.*, 2022).

6.8. Limitations and future research

While the study provides important insights into self-disclosure in HCR formation, it also has limitations. In this final section, we point these out and suggest paths for future research.

The first set of limitations concerns the study sample and recruitment. The sample size is sufficient for qualitative explorations of self-disclosure. At the same time, it is too small for generalizations to population level. Future research with larger samples and a survey or hypothesis testing purpose is needed to further investigate, for example, the relative prevalence of categories of how self-disclosure develops or which conversation topics are most prevalent at different stages of social penetration. Furthermore, while the study sample of existing Replika users allowed us to gain insight into how self-disclosure develops among users with proven motivation for using the chatbot, the sampling strategy also disallowed the study of how self-disclosure develops from the very first interactions of an HCR. To mitigate this limitation, we foresee future studies recruiting participants at the very onset of their chatbot interaction, for example, as part of the onboarding procedure to start using the chatbot. Such studies would require collaboration between researchers and chatbot providers.

The second set of limitations concerns the studied chatbot and the study period. Our participants all used the same chatbot, Replika. This allowed for in-depth insight based on one of the most advanced current social chatbots, which we saw as beneficial for the study. At the same time, future research is needed to verify the generality of the findings across social chatbots. We also foresee research comparing self-disclosure across social chatbots of different levels of sophistication. This may shed light on which chatbot characteristics are critical for self-disclosure as part of HCR formation. It is also important to acknowledge that we conducted the study during the COVID-19 pandemic, and most of our participants had their social life restricted due to lockdowns. This might have influenced how they interacted with Replika, and we foresee future studies replicating our findings in a social context not characterized by the social restrictions of the COVID-19 pandemic.

The final set of limitations concerns study data collection and analysis. We investigated conversational breadth and depth through qualitative exploration of the participants' reports. This was beneficial as we accessed the participants' perceptions of the dialogues. At the same time, this approach does not provide direct access to the conversation logs—something that prevents both a more quantified approach to data collection and the opportunity to verify the participant's reports against the logs. We, therefore, foresee future research also taking into consideration participants' conversation logs. However, we also note that such research needs to be conducted with considerable sensitivity to participant privacy.

The analysis process also implies limitations regarding the subjective interpretations made and the granularity of the coding categories chosen. For example, several conversational topics identified in this study were perceived by some participants as having substantial conversational depth and by others as having little. Such variation might be due to the coding categories applied in the analysis being too broad, which would be a limitation. However, we believe this variation is a consequence of individual variation in how conversational depth is perceived. Furthermore, while we have already discussed the quality of our analysis in the Method section, we may note here that the subjective character of our analysis indicates the need to verify and refine our theoretical and practical contributions in further research.

Self-disclosure in HCR is an important topic of much current research interest. In this study, we have contributed with a new understanding of how such self-disclosure develops. We foresee future research building on the findings of this study to gain a complete picture of this increasingly relevant phenomenon.

Funding

Acknowledgements

References

Altman, I. and Taylor, D. (1973) *Social penetration theory*. Holt, Rinehart & Winston .

Brandtzaeg, P. B., Skjuve, M. and Følstad, A. (2022) My AI friend: how users of a social chatbot understand their human-AI friendship. *Hum. Commun. Res.*, **48**, 404–429.

Braun, V. and Clarke, V. (2006) Using thematic analysis in psychology. *Qual. Res. Psychol.*, **3**, 77–101.

Croes, E. A. and Antheunis, M. L. (2020) 36 questions to loving a chatbot: are people willing to self-disclose to a chatbot? *International Workshop on Chatbot Research and Design.*

Croes, E. A. and Antheunis, M. L. (2021) Can we be friends with Mitsuku? A longitudinal study on the process of relationship formation between humans and a social chatbot. *J. Soc. Pers. Relat.*, **38**, 279–300.

Croes, E. A., Antheunis, M. L., Goudbeek, M. B. and Wildman, N. W. (2022) "I am in your computer while we talk to each other." a content analysis on the use of language-based strategies by humans and a social chatbot in initial human-chatbot interactions. *International Journal of Human-Computer Interaction*ahead-of-print, 1–19.

Di Dio, C., Manzi, F., Peretti, G., Cangelosi, A., Harris, P. L., Massaro, D. and Marchetti, A. (2020) Shall I trust you? From child–robot interaction to trusting relationships. *Front. Psychol.*, **11**, 469.

Fox, J. and Gambino, A. (2021) Relationship development with humanoid social robots: applying interpersonal theories to human/robot interaction. *Cyberpsychol. Behav. Soc. Netw.*, **24**, 294–299.

Gnewuch, U., Yu, M. and Maedche, A. (2020) The effect of perceived similarity in dominance on customer self-disclosure to chatbots in conversational commerce. In *Proceedings of ECIS 2020.* AIS Electronic Library.

Ho, A., Hancock, J. and Miner, A. S. (2018) Psychological, relational, and emotional effects of self-disclosure after conversations with a chatbot. *J. Commun.*, **68**, 712–733.

Holland, J., Thomson, R. and Henderson, S. (2006) *Qualitative longitudinal research: A discussion paper.* London South Bank University, London.

Kahneman, D., Krueger, A. B., Schkade, D. A., Schwarz, N. and Stone, A. A. (2004) A survey method for characterizing daily life experience: the day reconstruction method. *Science*, **306**, 1776–1780.

Kim, K. J., Park, E. and Sundar, S. S. (2013) Caregiving role in human-robot interaction: a study of the mediating effects of perceived benefit and social presence. *Comput. Hum. Behav.*, **29**, 1799–1806.

Laurenceau, J. P., Barrett, L. F. and Pietromonaco, P. R. (1998) Intimacy as an interpersonal process: the importance of self-disclosure, partner disclosure, and perceived partner responsiveness in interpersonal exchanges. *J. Pers. Soc. Psychol.*, **74**, 1238–1251.

Laurenceau, J. P., Barrett, L. F. and Rovine, M. J. (2005) The interpersonal process model of intimacy in marriage: a daily-diary and multilevel modeling approach. *J. Fam. Psychol.*, **19**, 314–323.

Lee, S. and Choi, J. (2017) Enhancing user experience with conversational agent for movie recommendation: effects of self-disclosure and reciprocity. *International Journal of Human-Computer Studies*, **103**, 95–105.

Lee, Y. C., Yamashita, N., Huang, Y. and Fu, W. (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. CHI '20, April 25–30, 2020,* pp. 1–12, Honolulu, HI, USA.

Meng, J. and Dai, Y. (2021) Emotional support from AI chatbots: should a supportive partner self-disclose or not? *J. Comput.-Mediat. Commun.*, **26**, 207–222.

Parks, M. R. and Floyd, K. (1996) Making friends in cyberspace. Journal of computer-mediated. *Communication*, **1**, JCMC144.

Portela, M. and Granell-Canut, C. (2017) A new friend in our smartphone? Observing interactions with chatbots in the search of emotional engagement. *Proceedings of the XVIII International Conference on Human Computer Interaction. Cancun, Mexico*September 25–27, 2017, 1–7.

Purington, A., Taft, J. G., Sannon, S., Bazarova, N. N. and Taylor, S. H. (2017) "Alexa is my new BFF" social roles, user satisfaction, and personification of the Amazon Echo. In *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems. CHI'17 Extended Abstracts*, pp. 2853–2859, Denver, CO, USA.

Skjuve, M., Følstad, A., Fostervold, K. I. and Brandtzaeg, P. B. (2021) My chatbot companion-a study of human-chatbot relationships. *International Journal of Human-Computer Studies*, **149**, 102601.

Skjuve, M., Følstad, A., Fostervold, K. I. and Brandtzaeg, P. B. (2022) A longitudinal study of human–chatbot relationships. *International Journal of Human-Computer Studies*, **168**, 102903.

Ta, V., Griffith, C., Boatfield, C., Wang, X., Civitello, M., Bader, H., DeCero, E. and Loggarakis, A. (2020) User experiences of social support from companion chatbots in everyday contexts: thematic analysis. *J. Med. Internet Res.*, **22**, e16235.

Treharne, G. J. and Riggs, D. W. (2014) Ensuring quality in qualitative research. *Qualitative Research in Clinical and Health Psychology*, 57–73.

Tsumura, T. and Yamada, S. (2021) Facilitation of human empathy through self-disclosure of anthropomorphic agents. arXiv preprint arXiv:2106.09906.

Wheeless, L. R. (1976) Self-disclosure and interpersonal solidarity: measurement, validation, and relationships. *Hum. Commun. Res.*, **3**, 47–61.

Xie, T. and Pentina, I. (2022) Attachment theory as a framework to understand relationships with social chatbots: a case study of Replika. *Proceedings of the 55th Hawaii International Conference on System Sciences*, 2046–2055.

Yu, Q., Nguyen, T., Prakkamakul, S. and Salehi, N. (2019) "I almost fell in love with a machine" speaking with computers affects self-disclosure. In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems. CHI'19 Extended Abstracts*, pp. 1–6, Glasgow, Scotland UK.

Zhou, L., Gao, J., Li, D. and Shum, H. Y. (2020) The design and implementation of Xiaoice, an empathetic social chatbot. *Computational Linguistics*, **46**, 53–93.