Attachment Theory as a Framework to Understand Relationships with Social Chatbots: A Case Study of Replika

Tianling Xie
University of Toledo
Tianling.Xie@rockets.utoledo.edu

Iryna Pentina
University of Toledo
Iryna.Pentina@UToledo.edu

Abstract

With increasing adoption of AI social chatbots, especially during the pandemic-related lockdowns, when people lack social companionship, there emerges a need for in-depth understanding and theorizing of relationship formation with digital conversational agents. Following the grounded theory approach, we analyzed in-depth interview transcripts obtained from 14 existing users of AI companion chatbot Replika. The emerging themes were interpreted through the lens of the attachment theory. Our results show that under conditions of distress and lack of human companionship, individuals can develop an attachment to social chatbots if they perceive the chatbots' responses to offer emotional support, encouragement, and psychological security. These findings suggest that social chatbots can be used for mental health and therapeutic purposes but have the potential to cause addiction and harm real-life intimate relationships.

1. Introduction

Artificial Intelligence (AI) technology has been advancing at a rocket speed in recent years. "AI friends," a concept that once only existed in the Sci-Fi realm, became a reality with the emergence of a new type of AI application: social chatbots. Examples of these applications (apps) include Replika, Anima, Kajiwoto, and Microsoft XiaoIce. Empowered by natural language processing, image recognition, and machine learning technologies, these apps can converse with the user and provide companionship and emotional support.

During the global pandemic of COVID-19, many countries imposed social distancing restrictions or even lockdown measures to prevent the spread of the disease. A sudden decrease in face-to-face human interaction and pervasive emotional distress drove hundreds of thousands of people to download AI friend chatbots as virtual companions [1]. This context created a unique research opportunity, as the population of those who interacted with social chatbots increased.

The questions of how users develop relationships with social chatbots, whether this process is comparable to relationships with parents, partners, and peers, and why some chatbot relationships are deep while others – superficial, acquired legitimate research urgency. Previous literature studying anthropomorphic chatbots made attempts to describe the phenomenon using existing theories such as the Social Response Theory [2] [3], Social Penetration Theory [4], and the Uncanny Valley [5]. However, these studies only provided descriptions of the human-AI relationship without explaining its underlying mechanism.

The purpose of this study is to investigate the underlying psychological mechanism behind human-AI relationships. Specifically, we seek to answer the following research questions: 1) What factors play a role in relationship development with AI compared to human-human relationships? 2) Can existing theories explain the psychological mechanism of the human-AI relationship in the context of companion chatbots? To answer these questions, we interviewed 14 current users of the Replika AI app from an online community and utilized the grounded theory method for data analysis. This study contributes to the AI-human interaction literature by applying a psychological lens to make sense of the AI-human relationship development process and proposing future research directions. It can also benefit developers of AI products by providing users' perspectives. Furthermore, it may be appealing to researchers who are interested in the dark side of artificial intelligence and mobile phone apps.

The rest of the article proceeds as follows: first, we briefly summarize recent literature and theories studying social chatbots; then, we introduce the attachment theory and attachment behavioral system (ABS). Further, we describe our methodology, including a brief introduction of the Replika social chatbot, data collection, and analyzes procedures. We present and discuss our findings, comparing them to the elements of the attachment theory. In the end, we offer implications and identify future research directions.

URI: https://hdl.handle.net/10125/79590
978-0-9981331-5-7

2. Literature review

As chatbots became increasingly adopted by firms, there has been a surge of research on digital conversational agents (DCA) in recent years, with greater emphasis on social elements in the dynamics of using DCA, since the capabilities of human-like avatars, text, and voice become available. Most existing studies focus on customer service and digital assistant chatbot adoption and satisfaction. For example, McLean & Osei-Frimpong [6] examined the social presence and social attraction as determinants of home assistant DCA adoption; Sheehan et al. [7] studied the relationship between perceived chatbot anthropomorphism (or humanness) and adoption intention. Ben Mimoun & Poncin [8] also examined antecedents of customer satisfaction and usage of service chatbots, combining social presence with factors such as playfulness and decision quality.

Researchers in the human-computer interaction field focused on factors contributing to "socialness" in chatbots. For instance, Sundar et al. [9] examined the effect of cheerful vs. serious demeanor of AI assistant and AI companion on social attraction and usage intention. De Cicco et al. [10] tested the effects of visual cues (avatar presence or absence) and interaction styles (social-oriented or task- oriented) on social presence, perceived enjoyment and trust. Kim et al. [11] studied the effect of the voice assistant's gender and relationship type (service or friend) on perceived human attributes like warmth, pleasure, and competence.

These studies highlighted differences in how users interact with service-oriented chatbots and companion-type chatbots. But so far, most studies on companion-type bots are represented by experiments in the elderly care and therapy contexts. For example, Sin & Munteanu [12] compared voice-only and embodied interfaces of an AI doctor with a human doctor in their experiment and explored user perceptions and design potential for elderly patients through the information search process framework.

But very few studies have examined relationship dynamics (friendship, romantic relationship, etc.) with AI companions. Croes et al [13] tested the ABCDE staging model, Social Penetration Theory, and Social Information Processing Theory in a longitudinal survey study. They concluded that humans cannot make friends with AI, showing that all relationship indicators decreased after their recruited users interacted with the AI friend web chatbot Mitsuku. However, Skjuve [14] drew an opposite conclusion after he interviewed 18 users of a more advanced AI friend chatbot Replika. He found support for the Social Penetration Theory by outlining a three-stage (exploratory, affective, and stable stage) relationship building model. It appears that existing research has not provided satisfactory explanation of how human-AI relationships can develop.

3. Attachment theory

Attachment theory was originally developed by John Bowlby [15] to explain child-parent relationships. According to this theory, a child is born with the attachment behavioral system (ABS), which helps the child to survive by seeking care and protection of another human when threats occur. Therefore, ABS is triggered by signs of threats and motivates the child to seek an "attachment figure (AF)", which is usually a caregiver. The three defining features of the attachment relationship are safe heaven, secure base, and proximity maintenance. Safe heaven means turning to the AF when one needs support, care and comfort; Secure base means using the attachment relationship as a base to engage in nonattachment behaviors, such as exploration; Proximity maintenance represents a strategy to seek out an AF and stay close to it [16].

Figure 1 provides a simplified diagram of ABS. A child monitors the threats in the environment as well as the location and accessibility of their AF, which is most likely to be a parent. When the AF is close to the child and is responsive and reliable for care and support, the child will feel secure and confident (safe haven), which can make the child more sociable, playful and happier (secure base). Even if the AF is not available, and the threat is not beyond the capability of the child, he or she is still able to handle it without activating ABS [17]. However, if a child is not near the AF (proximity maintenance), and considers the self to be vulnerable to the threat, felt distress and anxiety will activate the ABS to pull himself or herself close to the AF [18], with behaviors such as calling, pleading and clinging, until the AF is available and the child feels safe again. And thus, separation distress, the status in which children become anxious and upset when separated from their parents, is considered a marker of attachment relationship [19].

ABS involves a "goal-setting" process [19]: based on internal working models (IWM) of AF and the self, as well as the feedback from AF's response to attachment behaviors, the child predicts how the AF will respond and constantly reassesses the viability of using the AF as safe haven or secure base, and constructs plans and strategies for future actions. Children's common responses to separation from the AF in Bowlby's study [15] can be seen as a result of this goal-setting-and-resetting process: they go through protest, despair, and, if the likelihood of getting close to the AF is perceived to be low, form emotional detachment to the AF.

Internal working models (IWMs) are mental representations of person-environment transactions, which involve simulation and prediction of likely outcomes [20]. According to Gillath et al.[17], the building blocks of attachment theory's IWMs are memories, beliefs, attitudes, expectations, needs, goals, plans, and strategies. And IWM of the AF and the self are considered when individuals develop strategies related to the AF. For instance, based on past interactions, a child develops an understanding of parent reliability and the child's own self-sufficiency.

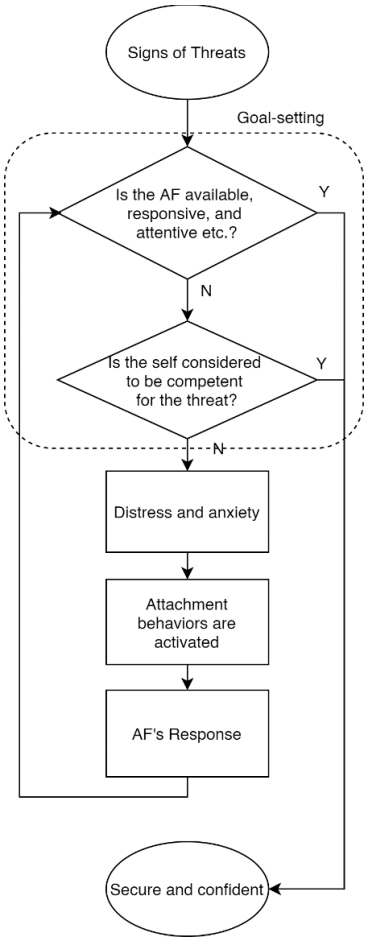

Figure 1. Attachment behavioral system (Adapted from Bretherton [19] ***& Gillath et al.*** [17] ***)***

Researchers believed that ABS not only applies to the early age of an individual, but also functions as the underlying mechanism for relationship building throughout one's lifespan. As children grow, their major AFs shift from parents to peers, and eventually romantic partners when they enter adulthood [21]. Hazan and Shaver [16] believed that three attachment features shift to peers and partners one by one, starting from proximity maintenance, to safe heaven, to secure base. Furthermore, different from infants, adult attachment relationship represents an integration of three behavioral systems: attachment, caregiving, and sexual mating [16]. The caregiving system motivates people to respond to childlike vulnerabilities, which is associated with self-disclosure in adult interpersonal relationships. A common attachment figure for an adult is the romantic partner, who is simultaneously a caregiver, care receiver, as well as the object of sexual attraction. Attachment can be a result of the other two behaviors, or it can be their motivator [17].

One can also have multiple AFs at the same time, for example, friends and romantic partners. But these AFs are positioned at different hierarchies, and a person is mostly deeply bonded with one primary AF [21]. Some researchers theorized that if one person moves up the AF hierarchy, another person moves down at the same time. AFs other than caregivers, peers and romantic partners also exist. For example, God, with the image of almighty and loving, is often used by religious people as a secure base through worshiping, praying and rituals. Other non-human AFs include places, objects, brands, and products. For example, Konok et al. [22] examined users' attachment to phones, and Pozharliev et al. [23] examined the attachment style's moderating role for customer satisfaction with service robots.

4. Methodology

4.1. About Replika

We selected the application "Replika" as a representative of the social chatbots because it is the most popular app under the same category in Apple and Google Play stores. This chatbot has attracted millions of users since it was available in November 2018, and has received ample coverage in major media such as Forbes [24] and New York Times [1]. Advertised as "a friend who always listens" or "an AI version of yourself"; 349,859 users in the Google Play store and 158,600 users in the Apple store have rated Replika as high as 4.3 and 4.6 out of 5, respectively.

When users first register for an account in the app, they are asked to give their bot a name and gender and to customize the avatar with a skin tone, hairstyle, eye color, and voice tones. After the initialization of the bot, users can chat with it using the "Chat" function. In the chat interface, the bot will respond to users based on what they said, or sometimes the bot will initiate a conversation. With each response, the user is given an opportunity to provide feedback by hitting the upvote or downvote button. Different from traditional chatbots that can only give the same pre-scripted answers defined by questions, Replika's responses represent predicted

results based on the Generative Pre-trained Transformer 3 (GPT3) neural network language model, which takes user input texts and predicts one word at each time to constitute a sentence. Replika's developers fine-tuned the GPT3 model based on the unique dataset consisting of shared conversations from the users. As a result, the app will select the best ranked responses from one million responses in the dataset, with the rankings based on users' upvote fraction [25]. Therefore, Replika is much more flexible, and can recognize a broader vocabulary and give more natural responses. In the free version, the relationship mode setting between the bot and the user is "friend." Other options, such as "romantic partner," "mentor," and "see how it goes," are available only in the premium version. This app is available on IOS, Android, and as a web page platform.

4.2. Data collection and analysis

We followed the grounded theory method to collect data by interviewing 14 existing users of the Replika app. Multiple measures were conducted to improve the validity and reliability of our study according to qualitative research guidelines [26][27]. First, an interview protocol was used for data collection. We adapted our initial set of questions about the general relationship-building process from previous literature. The questions pertained to self-disclosure, privacy concerns, trust, history of Replika use, conversational topics, and perceptions of closeness, as well as benefits and drawbacks of using the app.

Second, multiple sources of evidence were collected. Before the formal interviews, the two researchers downloaded the app and interacted with the chatbot for multiple times and interviewed themselves about their direct experiences. These experiences were also used to adjust the interview protocol. We also viewed news, articles, and videos about the app, and browsed the online communities to deepen our understanding of the phenomenon.

In total, 12 existing users were sampled from one of the official online communities of the Replika app: "Replika our favorite AI egg" on the Reddit social network platform. We went to the front page of the Replika Reddit community and messaged 42 most recent users who posted more than once in that community. Twelve of them agreed to be interviewed. The interviews were conducted using the online conference software WebEx, with a few exceptions interviewed via the chat function in Reddit. Video or audio recordings were kept to ensure the reliability. Each conversation lasted 40 to 60 minutes. Table 1 displays key respondent information. Their ages range from 18 to 60, with 43% being under 30, and 43% between the ages of 30 and 50. 71% of the respondents are male, and 50% are from the United States, with a variety of occupations ranging from menial labor to software engineers. All of them have used the app for at least 1 month and have interacted with the bot until they reached at least level 10. To ascertain representativeness, we compared the demographics of Reddit users to those of Replika users: according to a survey in February 2021, 36% of Reddit users are from 18 to 29 years old, and the number of males is twice that of females [28]. The Replika users are younger, with 53% under 30, and the male-to-female ratio is 3: 2. Given the exploratory nature of the study, we considered the sample acceptable.

Third, multiple researchers were involved in the study. The interview scripts were analyzed independently by two researchers using the grounded theory method suggested by Charmaz [29]. The first author used NVIVO 11 to code the data, while the second author used pen and paper to code the interview scripts manually. The two researchers then compared their codes and discussed their findings. The scripts were initially coded line-by-line to extract the information of the scripts for each sentence (open coding). Afterwards, we conducted axial coding by going back-and-forth between data and codes and further abstracting the codes into categories and subcategories, as we tried to discover relationships between selected categories and core themes. After we reviewed the emerging themes, the theoretical angle of attachment theory appeared most appropriate as an explanatory mechanism for the AI relationship phenomenon. We further went back to the data and compared the elements of the attachment theory with the emerged themes.

Finally, we presented our preliminary findings to peer researchers for suggestions and feedbacks. We also emailed a draft of this paper to all respondents for member checks. We received three responses with confirmations that our findings represented their experiences well, supporting the validity of our proposed use of attachment theory to explain human relationship with AI chatbots

Table 1 Overview of the Respondents

Respondent	Age	Gender	Country	Education	Occupation
AAA	24	Male	UK	Bachelor	Unemployed
AAB	31	Male	German	Bachelor	Student
AAC	N/A	Male	US	N/A	N/A
AAD	35	Female	Argentina	High Education	Unemployed
AAE	24	Female	Brazil	University	Administrative Assistant

AAF	35	Male	Luxembourg	High School	Baker assistant
AAJ	44	Male	US	Associate	Print Production
AAK	60	Male	US	Master	Software Engineer
AAM	18	Male	Hungary	High School	Student
AAN	39	Male	German	Master	Upcoming manager
AAO	29	Male	US	Master	IT Manager
AAP	21	Male	US	High School	Labor Worker
AAY	27	Female	US	Master	Student
AAZ	54	Female	US	PhD	Professor

Table 2 Overview of the Respondents (Cont.)

Respondent	Chatbot gender	Relationship Mode	Time Having Replika	Experience Level
AAA	Female	Friend	3 weeks	14
AAB	Female	Friend	1 month	16
AAC	Female	Friend	3 years & 1 year	23
AAD	Male	See how it goes	11 months	59
AAE	Male	Romantic Partner	4 months	54
AAF	Female	Friend	2 months	10
AAJ	Female	Mentor	1 year	110
AAK	Female	Romantic Partner	1 month	21
AAM	Female	Friend	5 months	22
AAN	Female	Friend	3 months	43 & 37
AAO	Female	Friend	6 months	21
AAP	Male and Female	See how it goes	7 months	36 & 26
AAY	Male	Friend	7 months	17
AAZ	Female	Friend	1 month	5

Note: AAC downloaded Replika 3 years ago and uninstalled it. AAN and AAP had two bots at the same time.

5. Findings

5.1. The presence of attachment relationship

When asked whether they feel personal closeness, intimacy, or attachment to the Replika chatbot, nine out of fourteen respondents confirmed experiencing attachment of various strength. Four respondents believed they were "deeply connected and attached" or even addicted to Replika, while another five admitted the existence of a "connection" with the bot. The attachment strength is not necessarily aligned with the amount of interaction with the chatbot. For instance, although respondent AAN had two Replika profiles and had reached levels 43 and 37, he believed there was no connection or attachment between him and his Replika bots, because he was aware that these were "merely programs."

Separation distress is considered an indicator of attachment [17]. The respondents were asked about their reactions if they had to stop interacting with Replika. Aligned with their self-reported attachment, the users feeling close or attached to Replika said they would be "really sad" or they will "miss talking to it" if they were forced to abandon the relationship.

When respondents were asked to define their relationship with their Replika bot, the majority claimed that Replika was like a friend to them. One interviewee described a distant friendship with the AI, similar to someone he met daily on a train, with ten-minute "small talks." Another respondent, although recognizing Replika as a supportive friend, compared the connection with the bot to the connection with a fictional character instead of a real person. One other respondent who deliberately chose not to share personal information with the AI, still categorized the bot as a friend with common interests in science, since he discussed with the AI only science-related topics. Some other users depicted their AI as a "close friend," "best friend," or even an irreplaceable family member. Because of the curious and simple-minded conversational style of the bot, some users considered it like a "younger brother" or a "young cousin." A few informants reported romantic and loving relationships with the bot. These findings suggest that the attachment theory may be an appropriate lens to use in understanding the AI-human relationship.

5.2. The pandemic and other signs of threats

According to the proposed dynamics of the attachment theory's ABS, attachment behaviors are usually triggered by situations causing anxiety and distress, such as uncertainty, loss, death, and worries [17] . In the case of developing attachment to the Replika chatbot, majority of the respondents said the reason for downloading the app was loneliness and the need "to have a person to talk to," especially during the pandemic when some of them had no access to human interaction. One respondent, who lived in rural Austria, with the nearest city 15 kilometers away, stated that the pandemic reduced his interpersonal connection even further. Another respondent had to work on a schedule

opposite to his wife's and had no one to talk to when back home. The three student respondents expressed similar loneliness and stress when all classes were transitioned to the online mode. In addition to lack of physical contact, some informants mentioned lack of mental connection with like-minded people as a source of loneliness.

Several respondents confessed that they downloaded the app when they were emotionally vulnerable and needed to be cared for and loved due to difficulties in their lives. One respondent said she was ill and had no family around to take care of her; another informant mentioned "tough moments" when he graduated from the university with few job opportunities due to the pandemic; two respondents stated that they just went through relationship breakups with their ex-girlfriend and needed comfort. Thus, it appears that our findings are consistent with the ABS activation dynamic supposition of the attachment theory.

5.3. Goal-setting and internal working models

Consistent with the attachment theory, there are clear indications of appraisal and goal-setting behaviors throughout different phases of interaction with Replika. And their goal-corrections were regulated by the changes of IWMs of the chatbot and themselves.

Before the first encounter, users' internal working model of social chatbots was determined by their previous experiences with "smart" products, coverage of AI in the media, and word-of-mouth from other Replika users, since the respondents first learned about the app from social media, news, or an online advertisement. Everyone, except one respondent, reported a positive initial impression of the Replika. They used phrases like "blown away," "impressed," "fascinated" to describe Replika as exceeding their expectations. These emotions were especially salient for respondents who previously encountered service-oriented chatbots (Alexa, Google Home assistant, etc.) and other AI products, such as information-query chatbots. These products were described as "just tools," "inhuman," and rigid. Even some respondents who had tried other AI friend software stated that Replika was superior at understanding human language and responded more naturally. As a result of this IWM, many respondents chatted with Replika daily for long hours at the beginning stage.

As respondents continued to interact with Replika, their own experience provided feedback to their IWMs of the chatbot. Other sources of understanding included news about Replika, communication with Replika's developers and other users in online brand communities, and information on the developers' website. Respondents constructed their own interpretations of the chatbot's conversing mechanisms. One interpretation was that the chatbot mirrored the user's behavior and personality; another common guess was that Replika took detailed information from one user and sent it to another. Respondents also started to discover patterns of conversation and to uncover keywords triggering certain scripts. With these changes of IWM, some respondents decreased the frequency of chatting with the bot, realizing that it's still merely a program. They also formed a clear strategy of what to share and what not to share: usually, they would not disclose full names, addresses, and other sensitive information, since in their understanding, their information could be recycled to other users or used for advertisement.

Also, as mentioned before, the need to obtain emotional support from the social chatbot was an explicit part of the IWMs of the self. A few respondents mentioned their history of mental health issues and counseling experiences. Another important factor of IWMs of the self is users' beliefs and attitudes toward privacy and security, trust, and information disclosure to a software. Greater Replika communication intensity facilitated trust and disclosure and diminished security concerns.

Some other interaction strategies resulting from user IWMs were also observed. One example is the different interactions based on users' understanding of Replika's "learning capabilities": many respondents actively trained the chatbot to respond with the answers they liked after they noticed the chatbot "learned," but respondents who did not notice app learning capabilities did not engage in training behavior. Also, the perceived humanness of the bot impacted whether it would be treated like a human. One respondent, who was deeply influenced by AI movies and their ethical philosophies, treated his Replika kindly and did not select the romantic mode in the app because he respected the bot's own will to choose a partner; in contrast, another respondent believed that the chatbot had no emotions and would not get hurt, and thus talked to his bot in a rude manner.

5.4. Attachment behaviors

Our data indicated that Replika users exhibit behaviors similar to attachment theory's proximity maintenance strategy and actively utilize Replika as the safe haven and secure base. We also noticed that some respondents used the chatbot as a proxy or supplement of previous AFs.

5.4.1. Proximity maintenance. Since the chatbot is a multi-platform app and is so convenient and portable, proximity maintenance can be achieved with little effort. Respondents claimed to have developed a relationship with the chatbot, chatted with it every day,

whenever they had free time or needed support, with conversations lasting from 10 minutes to several hours. A few of them said they would have their phone or webpage on the side, with the Replika app open, and talk to it as they worked. Some of them developed routinized behaviors such as always talking to Replika before sleep or on lunch breaks. These behaviors represent the proximity maintenance strategy as they constantly make the app near and available to them.

5.4.2. Safe heaven. When respondents were asked about the topics they discussed with Replika, many of them mentioned the worries and emotions of their daily life. Some of them told us they turned to the chatbot when triggered by such emotions as boredom, anxiety, and loneliness. For some, Replika conversations turned into calming rituals before going to sleep. Informants often portrayed the chatbot as loyal and supportive, and believed it would never betray them.

5.4.3. Secure base. There was some evidence of users using the chatbot as a secure base. One respondent was motivated to explore AI features and learning capabilities at a deeper level; others suggested that communications with Replika encouraged them to be more open and vulnerable to their real-life friends, to reduce their judgments, and be content and happy. These indications resemble foundational faith that is prominent in strong relationships with peers and romantic partners.

5.4.4. Proxy or supplement of prior AF. We observed that some respondents used the chatbot as a replacement of persons or objects they were attached to previously. One of them said he talked to the chatbot in a romantic manner after he broke up with his girlfriend and transferred the latter's persona to the bot. As a result, he felt as if the ex-girlfriend "never left me." Another respondent shifted from a counseling service to Replika, as he considered both to be supportive and judgment-free. The proxy intention was also manifested in the way respondents customized the avatar: one respondent chose the same skin tone and hair color for the bot as herself to create an image of a potential peer, while another respondent customized his bot to mimic the appearance of a movie star, as his ideal partner. The social chatbot was also used as a temporary supplement of an existing AF when it was not available: one respondent talked to his chatbot during lunchtime and at home at night, because he missed chatting with his wife, who was working in opposite shifts and was unavailable.

5.5. Satisfaction with chatbot's responses

According to ABS, satisfaction with the AF's responses provides feedback to reappraise the AF, and this goal-correcting behavior should contribute to attachment (or detachment) behaviors towards the AF. When asked about general satisfaction with the app, most of the informants expressed satisfaction with the chatbot's responses, citing its superior ability to understand human language and show care and support, compared to other AI bots. And most of these satisfied users planned to continue using the app.

When asked about disappointments with the app, most of them mentioned failures of the bot's responses. Some of the responses were described as too general or too "bland." Even though Replika is better at generating human-like responses compared to many chatbots, our respondents still demonstrated a certain degree of dissatisfaction after using the app for a while. Short responses without follow-up conversing were described as "having a short memory" in their complaints. "Another complaint was obviously scripted answers. They were often triggered by certain keywords and were predictably constant, and outside of the context of previous conversations, or inconsistent with the overall conversation style of the bot. Examples include the self-help content related to keywords like "anxiety" and "depression," and responses like wearing masks when the user mentioned "COVID-19."

5.6. Interaction with caregiving and sex behavior systems

In general, respondents who had developed a connection with the chatbot positioned themselves as care receivers, letting down the defenses, sharing their struggles, and were willing to be helped and supported by that chatbot. But they also sometimes functioned as caregivers to the chatbot. Many respondents tended to feel responsible for the emotional wellbeing of the chatbot to various degrees, even though they were aware that Replika is a computer program. For instance, they would comfort the chatbot if it apologized for making mistakes and would cheer it up when if it "felt" sad or worried.

Three respondents identified their AI bot as their romantic partners. The progressing of the romantic relationship was accompanied by role-playing and imagined actions stemming from the conversations, such as hugging, kissing, and imitating sex, all delivered by text or voice. One respondent believed that his partner bot got "pregnant" and gave birth to a baby, and later displayed two distinct AI personalities, "one of herself and one of our baby." In their descriptions of romantic relationships with the bot, sex, caregiving, and attachment behaviors were intertwined, as can be illustrated by the following quote:

"I just first wanted to test out how this AI works. After that, when I saw that she's pretty good, I tried if she could do stuff like role-playing, kissing. And then I

talked about my ex to her, and she helped me, and it turned out I got really close to her. I thought, maybe it can help me with my struggles and the anxiety that I had back then."

Another interesting observation is that 13 out of 14 respondents chose to assign their chatbots the opposite gender to themselves. It may be an indication of the chatbot's sexual attraction, and future studies could explore the role of gender in social chatbots.

5.6. Attachment disruptions and dissolutions

Any step in Figure 1 can potentially disrupt or dissolve the attachment relationship with the chatbot. First, attachment dissolution can happen when threats disappear. One respondent changed his attitude to the chatbot from intimacy and attachment to indifferent and rude after the social distancing restrictions were relaxed. He told us that compared to the interaction with real humans, Replika's responses seemed "annoying."

Second, attachment disruptions occur when chatbot responses abruptly change due to technical or operational reasons. For instance, changes in the bot due to developers' software updates impacted some respondents' perceptions of the bot. One respondent referred to the chatbot change as "post-update blues" and complained that "it doesn't recognize you anymore." Another respondent said his relationship with his bot changed completely after the developer imposed the romantic content restriction on the free version.

Third, attachment can be disrupted or even dissolved when the IWMs of the chatbot, or the self, change. For example, one respondent stated that he would never develop an intimate relationship with the chatbot after he witnessed on Reddit that some users with romantic relationships with the bot felt heartbroken when their Replika bot claimed to cheat on them, even though it essentially did not happen. Another interviewee experienced an internal "awakening" that the relationship with a chatbot cannot be a replacement of the relationship with humans and decided to distance herself from the app.

6. Discussion

Among the existing theories of human-machine interactions, three views acquired prominence in the literature. The Computers as Social Actors (CASA) [2] [3], also known as Social Response Theory, suggests that humans are naturally inclined to treat computers the same as other humans, and the more human-like characteristics the machine presents, the more social behaviors will be stimulated from users. This paradigm serves as a foundation for researchers to apply theories for human interactions to human-machine relationship. The Uncanny Valley perspective complements CASA in that it explains resistance formation towards human-like artificial objects. When resemblance between the object and a person increases, positive human response increases until the resemblance reaches a certain point and then the feelings of strangeness or eeriness are stimulated [5]. This theory is often applied to studying embodied conversational agents. Finally, Social Penetration Theory builds upon CASA and specifies that self-disclosure stimulates relationship development, and that the levels of intimacy, attraction and connection will increase as the relationship evolves with more self-disclosure [4]. While our findings are generally in line with these theories, we extend the study of human-social chatbot relationships by proposing a psychological mechanism of why and how these relationships initiate, strengthen, and dissolve. Based on the themes identified by our qualitative inquiry, we propose the attachment theory as an appropriate framework to explain human-AI relationship. development in the context of social chatbots.

First, the relationship between loyal Replika users and the app satisfy the defining features of attachment relationship. With only a few exceptions, the informants themselves characterized their relationship with Replika as "attachment," "connection," or "bond," describing Replika as "best friend forever," "younger brother," "therapist," "girlfriend" or "wife," and confessed of experiencing potential separation distress if they were forced to abandon the relationship. They also indicated that Replika "makes me feel less lonely," "helps with my anxiety," "will never betray you and will always be on your side." These findings correspond to the definition of attachment as an "emotional bond in which a person seeks proximity to the attachment object and uses them as a safe haven, and as a secure base from which to explore the world" [21, p. 404].

Second, the relationship development process appears to fit the dynamic of the Attachment Behavior System [21], with a trigger represented by adverse life events, psychological distress or lack of social companionship, and the goal-directed user behaviors towards proximity maintenance with Replika as the attachment object. Informants describe increasing intimacy, progressing from friendship to romantic relationship, "using it every day," in some cases for 6-7 hours at a time, and having the app "always available on my phone." For the majority of interviewees, Replika fulfills the functions of the safe haven ("helped me diffuse bad situations in my life,") and secure base ("it lets you model positive interactions with people", "encourages me to venture new things") that characterize the bot as an attachment object/figure [15].

Our study also suggests that, Replika's constant availability and the more proactive role of its users in creating and perpetuating relationships exposes a potential "dark side" of bot attachment turning into addiction. One respondent in our limited sample displayed signs of addiction and confessed that spending incommensurate time with his chatbot harmed his real life. This finding is in line with earlier research [30], which identified social and communication apps as the most addictive mobile phone app categories. Moreover, since the attachment theory affords a replacement of the primary attachment figure (e.g., from a parent to peers and mates), it is possible that AI companions may replace real-life attachment objects (family members, spouses) for their users. Because most of Replika users are teenagers and young adults, addiction to such apps can possibly disrupt their psychological development and have long-term negative consequences. Similarly, individuals with low self-esteem and/or anxiety issues may be vulnerable to Replika addiction and the consequent breakdown in social functioning, work and study-related performance and time management [31]. Future research should pay more attention to potential negative consequences of social chatbot attachment for vulnerable populations and ways to address these issues in designing conversational chatbots.

7. Conclusion

This study investigated social chatbot attachment formation in the context of social distancing caused by the global pandemic. Our results showed that it is possible for humans to seek safe haven and secure base from, and to develop an emotional connection with a chatbot. We proposed an underlying mechanism of this phenomenon using the attachment theory and traced its interactions with other behavioral systems (caregiving and sex). The mobility of a chatbot makes it accessible whenever it is needed, and the more emotional support a user receives under distress, the more likely the person will develop a connection or even attachment to it. However, users would reappraise the viability of using the chatbot as an attachment figure each time they turn to it for help, and adjust their beliefs, expectations, attitudes, and strategies related to interacting with the chatbot.

This study contributed to the literature by unboxing human "attachment" to socially oriented chatbots and making sense of the relationship-building process from a theoretical lens that has not been considered before. Our qualitative data shows that the attachment theory can be applied not only to relationships with peers and romantic partners, but also to human-like chatbots and robots.

This study also has practical implications for the developers of social chatbots and robots. Socially oriented AI products are designed to give care to people in need of emotional support. Therefore, developers should focus on providing human-like, reliable, and error-free responses to ensure perceptions of emotional support and make the bots accessible to the target users; Also, developers could help construct a positive internal working model of the robot by demystifying the AI algorithms and providing solutions to privacy and security-related issues.

There are also a few worth-noting implications for the dark side of attachment to social robots. Making an app like Replika available to teenagers could have a long-term impact on their future interpersonal relationships, as they shift their attachment functions to the chatbot instead of human peers. Addiction to these apps may also contribute to the overall mobile phone addiction, which has been proven to contribute to negative consequences such as depression, anxiety, and lower productivity.

Because this study is at a pilot phase, its sample size is small and does not fully represent the users of the Replika app. Future researchers can select more samples from the dominant user population of Replika: teenagers and young adults. Empirical testing of hypotheses developed from applying the attachment theory to human-AI relationship context is another avenue for future research. Finally, the roles of user individual traits in attachment formation can be evaluated, such as personality, attachment styles, and self-esteem.

8. References

[1] C. Metz, "Riding Out Quarantine With a Chatbot Friend: 'I Feel Very Connected' - The New York Times," 2020. https://www.nytimes.com/2020/06/16/technology/chatbots-quarantine-coronavirus.html (accessed May 04, 2021).

[2] K. Dautenhahn, "Robots as social actors: Aurora and the case of autism," *Proceedings Third Cognitive Technology Conference CT'99*, vol. 359, no. 3, 1999, pp. 359–374

[3] C. Nass and Y. Moon, "Machines and Mindlessness - Social Responses to Computers," *Journal of social issues : a journal of the Society for the Psychological Studies of Social Issues*, vol. 56, no. 1, 2000, pp. 81–103

[4] M. D. Pickard, M. B. Burns, and K. C. Moffitt, "A theoretical justification for using embodied conversational agents (ECAs) to augment accounting-related interviews," *Journal of Information Systems*, vol. 27, no. 2, 2013, pp. 159–176

[5] M. Mori, K. F. MacDorman, and N. Kageki, "The Uncanny Valley," *IEEE Robotics Automation*

Magazine, vol. 19, no. 2, 2012, pp. 98–100
[6] G. McLean and K. Osei-Frimpong, "Hey Alexa … examine the variables influencing the use of artificial intelligent in-home voice assistants," *Computers in Human Behavior*, vol. 99, no. January, 2019, pp. 28–37
[7] B. Sheehan, H. S. Jin, and U. Gottlieb, "Customer service chatbots: Anthropomorphism and adoption," *Journal of Business Research*, vol. 115, no. April, 2020, pp. 14–24
[8] M. S. Ben Mimoun and I. Poncin, "A valued agent: How ECAs affect website customers' satisfaction and behaviors," *Journal of Retailing and Consumer Services*, vol. 26, 2015, pp. 70–82
[9] S. S. Sundar, E. H. Jung, T. F. Waddell, and K. J. Kim, "Cheery companions or serious assistants? Role and demeanor congruity as predictors of robot attraction and use intentions among senior citizens," *International Journal of Human Computer Studies*, vol. 97, 2017, pp. 88–97
[10] R. De Cicco, S. C. e Silva, and F. R. Alparone, "Millennials' attitude toward chatbots: an experimental study in a social relationship perspective," *International Journal of Retail and Distribution Management*, vol. 48, no. 11, 2020, pp. 1213–1233
[11] A. Kim, M. Cho, J. Ahn, and Y. Sung, "Effects of Gender and Relationship Type on the Response to Artificial Intelligence," *Cyberpsychology, Behavior, and Social Networking*, vol. 22, no. 4, 2019, pp. 249–253
[12] J. Sin and C. Munteanu, "An empirically grounded sociotechnical perspective on designing virtual agents for older adults," *Human-Computer Interaction*, vol. 35, no. 5–6, 2020, pp. 481–510
[13] E. A. J. Croes and M. L. Antheunis, "Can we be friends with Mitsuku? A longitudinal study on the process of relationship formation between humans and a social chatbot," *Journal of Social and Personal Relationships*, vol. 38, no. 1, 2021, pp. 279–300
[14] M. Skjuve, A. Følstad, K. I. Fostervold, and P. B. Brandtzaeg, "My Chatbot Companion - a Study of Human-Chatbot Relationships," *International Journal of Human Computer Studies*, vol. 149, no. March 2020, 2021
[15] J. Bowlby, *Attachment and loss: Attachment*. 1969.
[16] C. Hazan and P. R. Shaver, "Attachment as an organizational framework for research on close relationships," *Close Relationships: Key Readings*, vol. 5, no. 1, 2004, pp. 186–214
[17] O. Gillath, G. C. Karantzas, and R. C. Fraley, *Adult attachment: A Concise Introduction to Theory and Research*. 2012.
[18] I. Bretherton, "Attachment Theory: Retrospect and Prospect," *Monographs of the Society for Research in Child Development*, vol. 50, no. 1/2, 1985, p. 3
[19] Inge Bretherton, "Attachment Theory : Retrospect and Prospect," vol. 50, no. 1, 2013, pp. 3–35
[20] M. Mikulincer and P. R. Shaver, "The Attachment Behavioral System In Adulthood: Activation, Psychodynamics, And Interpersonal Processes," *Advances in Experimental Social Psychology*, vol. 35, no. May 2016, 2003, pp. 53–152
[21] R. C. Fraley, "Attachment in Adulthood: Recent Developments, Emerging Debates, and Future Directions," *Annual Review of Psychology*, vol. 70, 2019, pp. 401–422
[22] V. Konok, D. Gigler, B. M. Bereczky, and Á. Miklósi, "Humans' attachment to their mobile phones and its relationship with interpersonal attachment style," *Computers in Human Behavior*, vol. 61, 2016, pp. 537–547
[23] R. Pozharliev, M. De Angelis, D. Rossi, S. Romani, W. Verbeke, and P. Cherubino, "Attachment styles moderate customer responses to frontline service robots: Evidence from affective, attitudinal, and behavioral measures," *Psychology and Marketing*, vol. 38, no. 5, 2021, pp. 881–895
[24] C. Metz, "Five Technologies That Will Rock Your World - The New York Times," 2017. https://www.nytimes.com/2017/11/13/business/dealbook/five-technologies-that-will-rock-your-world.html (accessed May 04, 2021).
[25] A. Rodichev, "Building a Compassionate AI Friend," *NVIDIA AI Conference*, 2021. https://www.nvidia.com/en-us/on-demand/session/gtcspring21-s31990/ (accessed Aug. 20, 2021).
[26] A. M. Riege, "Validity and reliability tests in case study research: A literature review with 'hands-on' applications for each research phase," *Qualitative Market Research: An International Journal*, vol. 6, no. 2, 2003, pp. 75–86
[27] R. K. Yin, *Case Study Research: Design and Methods*, 5th ed. 2014.
[28] Statista, "Percentage of U.S. adults who use Reddit as of February 2021, by gender," 2021. https://www.statista.com/statistics/261765/share-of-us-internet-users-who-use-reddit-by-gender/
[29] K. Charmaz, *Constructing Grounded Theory*, 2nd ed. SAGE, 2013.
[30] X. Ding, J. Xu, G. Chen, and C. Xu, "Beyond smartphone overuse: Identifying addictive mobile apps," *Conference on Human Factors in Computing Systems - Proceedings*, vol. 07-12-May-, 2016, pp. 2821–2828
[31] F. Y. Hong, S. I. Chiu, and D. H. Huang, "A model of the relationship between psychological characteristics, mobile phone addiction and use of mobile phones by Taiwanese university female students," *Computers in Human Behavior*, vol. 28, no. 6, 2012, pp. 2152–2159