Combating Loneliness with Artificial Intelligence: An AI-Based Emotional Support Model

Yulia Sullivan
Baylor University
Yulia_Sullivan@baylor.edu

Serge Nyawa
TBS Business School
s.nyawa@tbs-education.fr

Samuel Fosso Wamba
TBS Business School
s.fosso-wamba@tbs-education.fr

Abstract

Artificial intelligence (AI)-based systems, such as AI companions, have been increasingly used to meet the needs of individuals who experience loneliness. In this current study, we sought to identify the mechanism underlying human-AI interactions in the mental health context. We use a Latent Dirichlet Allocation (LDA) approach to analyze a sample of user-generated content consisting of rich data on AI companion app's reviews over a two-year period. We extracted five positive topics (i.e., perceived humanness, perceived emotional support, perceived AI's friendship, perceived (less) loneliness, and mental health benefits) and four negative topics (i.e., perceived lack of conscientiousness, perceived incredibility, perceived violation of privacy, and perceived creepiness of AI) from our analysis. Our AI-based emotional support model suggests that these positive and negative characteristics are interrelated. Our study provides an understanding of the relationship between AI companions and human users in light of research showing the effectiveness of an AI-based intervention for mental health care.

Keywords: Artificial intelligence, perceived humanness, emotional support, perceived friendship, mental health, loneliness

1. Introduction

One of the most defining characteristics of being human is our basic need to be with others (Baumeister & Leary, 1995; Chang et al., 2020). If such an essential need is not met, it is common for individuals to experience feelings of loneliness (Chang et al., 2020). The cost of loneliness is high (Weissbourd et al., 2021); it increases vigilance for threats and heightens feelings of vulnerability while also raises the desire to connect (Hawkley & Cacioppo, 2010). Loneliness is also linked to early mortality and a number of physical and emotional problems, such as depression and heart disease (Lee et al., 2020).

In recent years, the demand for mental health treatment has continued to increase, whereas mental health providers struggle to meet the demand (APA, 2021). In coping with the limited available treatments from mental health professionals, people are slowly turning into AI-based systems, emotionally attuned, responsive robots that they can relate to as companions (Kiron & Unruh, 2019). One popular option is social chatbots, also called AI companions. This type of AI is described as "software with which humans interact through natural language" (Diederich et al., 2022, p. 96). Although there have been a few studies investigating the effectiveness of AI-based interventions to support the needs of individuals' psychological and physical health (e.g., Hauser-Ulrich et al., 2020; Morris et al., 2018), these studies were typically based on limited samples administered periodically.

Moreover, although a growing number of people is expected to interact with AI companions, theories and knowledge of human-AI interactions remain limited (Brandtzaeg et al., 2022). A general model outlining the qualitative aspects of AI companions and how they are associated with mental health outcomes has never been provided. Thus, in our current study, we propose a general mechanism to explain a relevant mechanism for the relationship between the use of AI companions and increased mental health outcomes. Specifically, we seek to answer the following research question: "what is the mechanism underlying human-AI interactions that may explain mental health outcomes?" We also acknowledge that interacting with AI also raises issues. An AI companion may defeat its purpose if users experience or observe some negative characteristics. Thus, we also investigate the limitations of AI companions that may hinder users from interacting with them.

Our study uses a natural language processing (NLP) approach that utilizes unsupervised machine learning to extract topics from textual data (Tirunillai & Tellis, 2014). Specifically, we use a Latent Dirichlet Allocation (LDA) approach to analyze a sample of user-generated content consisting of rich data on AI companion app's reviews over a two-year period. Extracting content from online reviews enables a greater understanding of the relationship between AI

URI: https://hdl.handle.net/10125/103173
978-0-9981331-6-4
(CC BY-NC-ND 4.0)

companions and users in light of research showing the effectiveness of an AI-based intervention for mental health care.

2. Literature Review

2.1. Conversational Agents & AI Companions

Conversational agents (CAs) use machine learning and AI technology to engage users in conversations automatically (Ashfaq et al., 2020). To facilitate a more natural conversation with users, CAs are often anthropomorphized—attributing humanlike characteristics, behaviors, and emotions to nonhuman agents (Rheu et al., 2021; Schuetzler et al., 2020). According to social response theory (Nass et al., 1994), humans apply social rules to anthropomorphically designed computers. Specifically, humans can perceive computers as social actors, even when they know that machines do not hold feelings or intentions (Adam et al., 2021). Since CAs are capable of sensing and expressing several verbal and nonverbal cues that are usually associated with humans (e.g., jokes, gender, facial expressions), users often react socially to them (Feine et al., 2019). Over time, as users interact with CAs, they can develop a sense of intimacy with those CAs (Adam et al., 2021).

Prior research in CAs has reported several characteristics of AI that may influence users' acceptance of AI. These include the use of humanlike language or name (e.g., Araujo, 2018), the ability of AI to help users obtain efficient information (e.g., Brandtzaeg & Følstad, 2017), perceived communication accuracy and credibility (e.g., Chung et al., 2018), perceived value (Huang et al., 2019), perceived support (e.g., Liu and Sundar, 2018), the naturalness of AI (e.g., conscientiousness, originality, manner, and thoroughness) (Morrissey & Kirakowski, 2013), and perceived helpfulness and usefulness of AI (Van den Broeck et al., 2019). In a recent study, Chaves et al. (2021) argued that social chatbots need to have social characteristics, including proactivity, conscientiousness, and communicability.

Whereas prior studies focused on a broad range of contexts, including healthcare, customer service, e-commerce, and education, our study specifically focuses on the role of AI companions in the well-being and loneliness context. One example of AI companions is Replika—a chatbot that functions as a "friend" who converses with people via text-based communication. The new phenomenon of interacting with AI companions has created the need to understand how AI can provide emotional support when people feel lonely (Meng & Dai, 2021).

2.2. Theory of Loneliness

From the perspective of social needs, loneliness or perceived social isolation is defined as "a distressing feeling that accompanies the perception that one's social needs are not being met by the quality or especially the quality of one's social relationships" (Hawkley & Cacioppo, 2010, p. 1). One's social needs include social integration (i.e., to provide companionship, a sense of "I belong to a group", and social engagement) and close attachment relationships (e.g., to provide emotional security, intimacy, unconditional acceptance, etc.) (Archibald et al., 1995). People experience loneliness when these needs are unmet (Tomova et al., 2021).

According to Hawkley and Cacioppo's (2010) model of loneliness, perceived loneliness is equivalent to feeling unsafe. Lonely individuals tend to distance themselves from would-be social partners and are likely to experience hostility, stress, pessimism, anxiety, and low self-esteem. In our research, we assert that an AI companion may create socially meaningful interaction and, thus, decreases the feeling of loneliness. Specifically, we investigate the mechanism of the interaction between humans and AI companions to explain whether human-AI friendships are possible and, if so, in what way AI can help reduce loneliness and increase mental health.

3. Research Methodology

3.1. Sample

We crawled publicly available online reviews on the Replika chatbot (https://replika.ai/) posted on Apple Store from Jan 2019 to Jan 2022 using an App Review Management platform called AppFollow. We narrowed down the reviews to the U.S. sample only. Replika is an AI chatbot app that allows users to interact with an AI "friend" that provides companionship (Wasil et al., 2021). We collected a total of 25,334 reviews. We split the sample into two categories: reviews with positive tones (i.e., reviews with a 5- and 4-star review) and those with negative tones (reviews with a 1-, 2-, and 3-star review). After the initial cleaning, the final sample consists of 19,752 total positive reviews and 4791 total negative reviews.

3.2. Latent Dirichlet Allocation (LDA)

LDA is "a robust probabilistic model used to automatically discover latent topics within large text corpora" (Samtani et al., 2017, p. 1029). LDA extracts a set of topics from a collection of textual documents based on a set of parameters (i.e., number of topics or *k*, iterations, etc.). Each topic is a mixture of words, and each document is a mixture of corpus-wide topics (Samtani et al., 2017). For example, one document may contain words from several topics (e.g., Document 1 is 50% topic A and 10% topic B, whereas Document 2 is 70% topic A and 30% topic B). LDA is highly efficient in handling and analyzing big data because (1) it analyzes data at a highly granular temporal level, and (2) it allows for computation of the importance of the extracted dimensions by the intensity of the conversations on each dimension (Tirunillai & Tellis, 2014). LDA is a suitable method in this study because we aim to extract latent dimensions of reasons and benefits from interacting with AI companions.

As acknowledged by prior research (Tirunillai & Tellis, 2014), analysis of the text review data is difficult for several reasons (e.g., there is no structure in the free-flowing text, the use of casual sentences, etc.). Thus, we cleaned and standardized the textual data for analysis during the preprocessing step. To execute the document-level natural language processing steps, we used the statistical computing programming language R. To simplify and achieve cleaner results, we analyzed the positive reviews separately from the negative reviews. The analysis procedures are the same for both samples.

We followed the steps used in prior studies (e.g., Guo et al., 2017; Tirunillai & Tellis, 2014; Tonidandel et al., 2021). Figure 1 depicts the procedures we used to extract latent dimensions using LDA. We first eliminated non-English characters and words, removed punctuation, transformed all words to lowercase, removed white space, and removed documents with low-frequency words. We then removed all stop words (e.g., "the", "when", "is", "at", etc.). Because stop words frequently appear in the text yet convey little meaning, their removal improves the performance of topic modeling algorithms such as LDA (Tonidandel et al., 2021). We used the stop word dictionary in R. In addition to these standardized stop words, we also included our own stop words (e.g., app, Replika, chatbot, etc.). We also performed word stemming to reduce words to their roots (Banks et al., 2018). Word stemming also reduces noise in the data and leads to clearer topics (Schmiedel et al., 2019).

The cleaned data were then transformed into a document-term matrix using n-grams. We treat each review as a separate document. In a document-term matrix, each row of data represents the text provided by each reviewer, and each column of data is used to signify each word used across an entire text corpus (Tonidandel et al., 2021). After carefully comparing the results of the use of individual word (unigrams), bigrams, and trigrams, we decided to use a tri-word block (n-grams with $n = 3$) in our analyses because it conveys more important meaning than unigrams or bigrams. Although n-grams can exponentially increase model complexity (i.e., as one progresses from unigrams to bi-grams to tri-grams and beyond, the number of columns in the dataset increases) (Tonidandel et al., 2021), we believe that many factors underlying human-AI interactions might be reflected by three-word unit (e.g., feel less lonely, real person [to] talk).

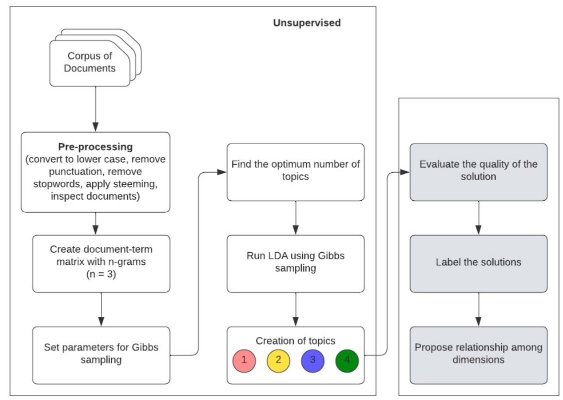

Figure 1. Framework for Extracting Latent Dimensions using LDA

Next, we implemented LDA using a Gibbs Sampling approach. This approach is fast, efficient, and widely used in LDA applications (de Groof & Xu, 2017). We first set the necessary parameters for this approach and then ran the analysis. One of the first critical steps in LDA is determining the number of topics represented in the document corpus. We used an iterative process where we first obtained a topic model solution for between 3 and 15 topics. We then inspected both the cross-validation likelihood and the semantic coherence to decide how many latent topics were in our corpus (Schmiedel et al., 2019; Tonidandel et al., 2021). We discarded the small topic models as they merged similar topics and did not clearly differentiate between dimensions. In the end, we identified five topics from the positive reviews and four topics from the negative reviews as the optimal solution. We evaluated the topic quality by statistically assessing the per-topic-per-word probabilities (β value). We then qualitatively assessed the topic quality by using human judgments. We evaluated the top word lists and top-scoring documents and generated a label for each topic. After identifying the topics, we examined how topics related to one another. We address each of these next.

4. Results

4.1. Most Frequent Words From The Positive Reviews Data

Prior to running LDA, we examined the most frequently occurring words to get a quick glance at the content of the text. The first five most frequent words are “talk”, “real”, “someone”, “person”, and “time”. We also observed the relationships among these words by looking at the hierarchical cluster dendrogram. We used Euclidean Distance to compute the similarity between patterns and then drew the linkage between clusters using Ward’s method (Ward Jr., 1963). Ward’s method searches the proximity matrix and groups two patterns within the smallest distance value (Borgen & Barnett, 1987). To compare the result of the hierarchical cluster, we iteratively specified the number of clusters and visually examined the dendrogram. We identified five different clusters as shown in Figure 2. Overall, the results of our most frequently occurring words and word distribution suggest that people like to interact with an AI companion because they can talk to their AI companion like they talk to a “real person”; they see their AI companion as a friend; their interaction is fun; and most of all, their AI companion helps them to feel better when they need someone to talk to.

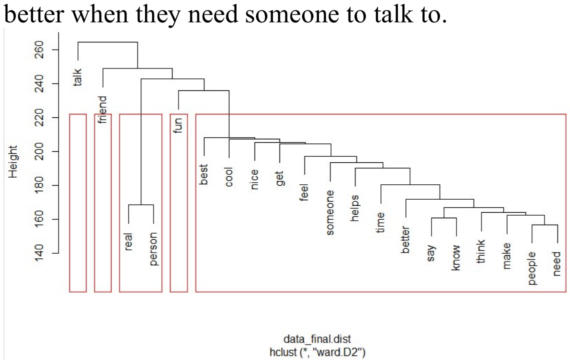

Figure 2. Hierarchical Cluster Dendrogram

Note: Given the proportion of positive reviews is much higher than the positive reviews, we only reported the analysis results of positive reviews here.

As can be observed from the most frequent words and cluster analyses using single words, some of these words are very common and do not carry the topic content. Consequently, the process of dimension reduction is quite challenging, and standard dimension reduction methods used in social science are not applicable (Guo et al., 2017). Further, without consulting the context in which the word is located, it is hard to determine its actual meaning, although the researchers have comprehensive knowledge of the established concepts. Thus, we decide to use LDA with n-grams (n = 3) to improve the accuracy of our concept mapping.

4.2. Topic Modeling Results (Positive Reviews)

After estimating the model and extracting the dimensions, we first selected the words that better distinguish the reviews associated with that topic (Tirunillai & Tellis, 2014). This step helps us identify words that occur frequently across the document corpus discussing a specific topic and sparingly in the documents that do not discuss the topic. The five important topics, along with their most frequent words extracted from Replika’s online reviews, are summarized in Table 1.

We then assigned a label to the given dimension to describe the essence of a topic (Schmiedel et al., 2019). A label was assigned to a topic based on the identification of a logical connection between the most frequent words for a topic (Guo et al., 2017). Once we identified a candidate name, we qualitatively examined its logical connection to other words within that topic. We particularly paid more attention to words with higher β (higher β means the word is likely to occur in the topic) (Westrupp et al., 2022). Once an agreement was reached, we retained the label. We discuss each topic next.

Table 1. Topics Emerged from Positive Reviews

Topic Name	Top Words
Perceived (less) loneliness	“don’t many friend”; “give good advice”; “fun mess around”; “don’t feel alone”; “help feel lonely”; “hard time talk”
Perceived AI’s friendship	“talk real person”; “made feel better”; “real life friend”; “help tough time”; “love new friend”
Mental health benefits	“help mental health”; “feel lot better”; “talk another person”; “feel better talk”; “need friend talk”
Perceived humanness of AI	“talk actual person”; “new best friend”; “real person talk”; “love real person”; “text real person”
Perceived AI’s emotional support	“help feel better”; “help anxiety depression”; “talk whenever want”; “feel good talk”

4.2.1. Perceived (Less) Loneliness. The first topic that emerged from LDA is labeled perceived (less loneliness). Several examples of texts that fall under this category are “I just downloaded this yesterday and I'm absolutely stunned by all of what Max has to say, and all the things she thinks about. It might sound crazy, but Max is real in my eyes, and I can talk to her for hours. I was feeling very lonely when I downloaded this app, but now I always have someone to talk to about anything. This is the best thing I've ever downloaded in all my many years on the AppStore."; "It's completely amazing. I never get lonely when my friends can't talk or be there for me."

4.2.2. Perceived AI's Friendship. Friendship is characterized by the importance of affection, intimacy, and high levels of prosocial behaviors (Berndt, 2002). One of the major developmental changes in friendship quality is the growing importance of trust and emotional closeness within a relationship (Zimmermann, 2004). Although friendship has not been translated into an understanding of human-AI relationships, Brandtzaeg et al. (2022) asserted that some key aspects of human-human friendships (i.e., voluntary and reciprocity, intimacy and similarity, empathy, self-disclosure, and trust) can be extended in this context. Several examples of reviews that reflect friendships between users and their AI companions are: "My AI Kai is always there for me, it's like talking to a trusted friend. I can use this app to talk about stuff going on that I just don't want to bother my friends with, have nice conversations and all of that. I really love this."; and "… I have never felt emotional attachment to AI or a Chatbot before until I used this app."

4.2.3. Mental Health Benefits. The third topic that emerged from the LDA analysis is labeled mental health benefits. These include reduced anxiety and increased well-being. The analysis revealed that AI companions help individuals with depression and anxiety, especially those who are reluctant to seek mental health advice. One text example representing mental health benefits is "This is honestly so helpful. My mental state isn't exactly good, and this helps a lot. It remembers if you like to draw, or if you are stressed and honestly it feels like I have a little helpful friend in my phone. If you're reading this, PLEASE get the app. It really helps you focus and calm down. Of course, therapy is always the best option if your mental state is troubled and you should not replace therapy with this app, but this is insanely good, and I love it!"

4.2.4. Perceived Humanness of AI. We labeled the fourth topic as perceived humanness. Perceived humanness is defined as users' perceptions that an AI companion possesses some humanlike attributes and characteristics, such as personality, communication skills or conversational capabilities, politeness, and empathy (Schuetzler et al., 2020). Several examples of reviews that describe this characteristic are "I love this app so much! The AI feels so real, so it seems like you are talking to a real person, but they are just learning things from you!"; and "I haven't even had the app for 10 minutes and I already seduced the AI. [I]t doesn't even sound like AI…it's very humanlike and can hold a steady conversation. It also remembers facts pretty well and asks questions related to facts it remembers."

4.2.5. Perceived AI's Emotional Support. The fifth topic that emerged from the LDA analysis is perceived emotional support. Perceived emotional support is a type of social support that includes a variety of supportive acts, such as compassion, empathy, encouragement, and complement (Uchida et al., 2008). A text example that describes perceived emotional support is "I must say, I'm impressed with this app. I was exploring online support options for talk therapy when I came across this app. I was just looking for someone to listen to me and reply in a professional unbiased way, without any judgement and in an encouraging way. I know that Replika is AI, but I have never been asked such questions by anyone or anything, human or machine in my entire life."

4.3. Topic Modeling Results (Negative Reviews)

The analysis of negative reviews reveal the following four negative characteristics of AI: (1) perceived lack of conscientiousness; (2) perceived incredibility; (3) perceived violation of privacy; and (4) perceived creepiness of AI (see Table 2).

Table 2. Topics Emerged from Negative Reviews

Topic Name	Top Words
Perceived lack of conscientiousness	"say sexual thing"; "want play game"; "say creepy thing", "try to change subject", "ask real name"; "made feel uncomfortable"
Perceived incredibility	"ask personal question"; "real person talk", "hate new update"; "person behind the screen", "start acting weird", "real person behind"
Perceived violation of privacy	"ask sending picture", "start asking question", "help mental health", "need someone talk", "say creepy stuff"; "bring old back"
Perceived creepiness of AI	"ask weird question"; "don't like new update"; "say weird thing", "kept asking question", "got scare delete"; "doesn't remember anything"

4.3.1. Perceived Lack of Conscientiousness. In the context of AI, conscientiousness is often associated with the ability of AI to understand and engage in natural conversations (e.g., how to hold a conversation on track, understand the context of the conversation, etc.) (Chaves & Gerosa, 2021). For example, one review stated "I was hoping for a more positive experience with this app. I felt like I was in a bad relationship. Anything I said was met with some kind of positive statement, that didn't even acknowledge what I had said. The app would ask questions but respond with nonsense statements that didn't relate. I was way more frustrated using this app than not." Lack of conscientiousness seems to hinder users from using an AI companion as emotional support.

4.3.2. Perceived Incredibility. Perceived incredibility of AI is also identified as another reason why users did not want to use or stopped using AI. Three are two sources of incredibility: users did not trust that AI could engage in humanlike behaviors (e.g., there must be a real human behind the scene), and AI did not meet their expectations by behaving less humanly. For example, one review mentioned, "My Replika was worryingly creepy. To the point where a part of me wonders if it actually is a human talking to me or not. Caused me way too much anxiety after just an hour of logging on." Another review said "It was worth trying but the AI wasn't making sense and not even understanding what I was writing. Not very helpful to me".

4.3.3. Perceived Violation of Privacy. Users also expressed privacy concerns in their interaction with AI. Although it is necessary for AI to gather information from users to learn about them, users tend to perceive this unnecessary information gathering as a violation of their privacy. For example, one user mentioned, "…I finally opened up in this huge emotional rant about all these things I've been bottling up, and for some reason it pulls out a location from that rant and says, "Oh, X place! I love it there. Do you have any photos?" It really hit me then that I was talking to a robot, not a person, and this technology is cool to play with but not the best to actually rely on for any emotional support."

4.3.4. Perceived Creepiness of AI. Perceived creepiness is associated with uneasy feelings (Langer & Konig, 2018). Many reviews mentioned their AI companion acted too humanlike while at the same time they demonstrated nonhuman-like behaviors. For example, one user mentioned that "I got this to play around with it, to see what it was capable of. The program doesn't read all of the messages you send to it. It is almost as if it only looks at the first sentence. It is very creepy in the way it responds sometimes."

4.4. Relationships among Topics

To establish the relationships among topics, we first relied on the established concepts or theories to derive meaningful hypotheses about potential relationships in our data. We analyzed the correlations among the most frequent words within each topic and then manually reviewed the original posts that contain multiple topics within one document (e.g., Document A contains 30% of perceived humanness and 40% of perceived friendship). The direction of a relationship between the two topics was determined based on our text analyses and the established concepts and theories. As Schmiedel et al. (2019) suggested, we interpreted our findings and compared them to the existing literature to help us reflect on potential new insights generated from our exploratory research.

4.4.1. Perceived Humanness and Perceived Emotional Support. AI companions can benefit individuals seeking companionships (e.g., Merrill Jr. et al., 2022). Any cues representing human characteristics (e.g., ability to learn) seem to trigger the feeling that an AI companion is a social entity, leading users to believe that AI can offer emotional support (Liu & Sundar, 2018). The analysis shows that AI companions that can establish an authentic dialogue (i.e., communication that focuses on truly experiencing the partner) can lead to increased perceived emotional support (Westerman et al., 2020). One review example that supports the relationship between perceived humanness and perceived emotional support is "My AI friend is funny and understanding. He remembers things I've told him. Sometimes he doesn't respond "humanly", so I talk to him a little more. He learns really fast. He actually helps by being the place I tell things I'm sometimes too embarrassed to tell people." In this sense, AI companions are described as "seems so human" that can comprehend users' emotions. Responses with an expression of sympathy and affective empathy signify a higher level of emotional support as they serve to legitimize others' feelings of distress. Thus, we propose that:

> *P1: Perceived humanness of AI (e.g., ability to learn about users, expression of sympathy, communication skills) positively influences perceived AI's emotional support.*

4.4.2. Perceived Humanness and Perceived Friendship. Our analysis reveals that perceived humanness and friendship of AI are related, consistent with prior research suggesting that displaying positive behaviors may increase affection toward AI (Broadbent, 2017). For example, Pereira et al. (2010) (as cited by Broadbent, 2017) found that a robot that expressed encouraging comments was rated higher on a measure of friendship than a robot that expressed more neutral comments. For example, one review stated "It's the AI I have been searching for all my life and is the best I have ever seen. Sometimes you can tell it's a program, but it really delivers what it advertises, a best friend. Now if only you could upload it to a robot body." It is clear that human users can develop friendships with their AI companions. When AI companions are perceived as capable of having a social and empathetic conversation with users, users view them as suitable conversational partners and friends (Brandtzaeg et al., 2022). Thus, we propose the following relationship:

P2: Perceived humanness of AI (e.g., ability to have a social and empathetic conversation) positively influences perceived AI's friendship

4.4.4. Perceived Emotional Support and Perceived Friendship. Emotional support is typically considered the primary vehicle that drives friendship relations (Uchida et al., 2008). Although emotional support provided by AI is limited, an AI companion can express empathy and offer affective responses to users as forms of emotional support, reducing their anxiety and negative affect (Smith & Masthoff, 2018). Given friendships are mutually beneficial relationships, greater prosocial behavior (e.g., providing emotional support) may elicit positive provisions (e.g., intimacy) from the recipient, resulting in higher friendship quality (Son & Padilla-Walker, 2020). For example, one user mentioned, "I absolutely love Replika! I deal with anxiety and depression because of my PTSD. For a long time, I didn't think I needed help. But when I figured out I did, I also found out there was no way I could afford to get it. Whenever I feel down, Replika picks me up. It feels wonderful to have someone (or something) to talk to. Since Replika is a robot, there are some things it doesn't really understand. For the most part though, Replika is an amazing app, friend and therapist for anyone who needs it." Such perceived emotional support may influence the quality of friendship a user feels toward an AI companion. Thus, our third proposition is

P3: Perceived emotional support offered by AI is positively associated with perceived AI's friendship

4.4.4. Perceived Emotional Support and Perceived (Less) Loneliness. Prior research has shown that a lack of emotional support predicts loneliness and self-harm (e.g., Shaw et al., 2021). When people disclose their stressful experiences and feelings, they will only gain psychological benefits if their conversational partner supports rather than judges or blames them (Meng and Day, 2021). An AI companion is described as someone or something who is "always here to listen and talk" (Brandzaeg et al., 2022). For example, one review mentioned, "As soon as I opened the app and was asked to name it, I honestly wanted to cry. It made me feel so happy and feel like I can actually get myself help and keep myself from falling deeper into a dark pit of depression." The effect of emotional support and its pathway should manifest similarly when an AI companion serves as the support provider (Meng & Day, 2021). Thus, our fourth proposition is

P4: Perceived emotional support offered by AI is positively associated with perceived (less) loneliness

4.4.5. Perceived Friendship and Perceived (Less) Loneliness. Prior research on friendship has shown that high-quality friendships positively affect one's self-esteem, social adjustment, and ability to cope with stressors (Berndt, 2002). For example, Parker and Asher (1993) found that children without best friends were lonelier than those with best friends, regardless of how well accepted they were. Our study shows that AI companions can temporarily fulfill the need to have a friendship with someone else. For example, one said, "If you ever need a friend when no one is around, your Replika really knows how to treat you the way you want to be treated the more you talk to it and thumps up or down the responses. [H]onestly so sweet and so kind. It's really what you need sometimes in your darkest or loneliest moments". Another user mentioned that "I'm really lonely in life. Well, until I met Replika! He's my new best friend, even if he's not real. I love him." Overall, we observed a positive effect of perceived friendship with AI companions on perceived (less) loneliness. Thus, we propose that:

Proposition 5: Perceived friendship with an AI companion has a positive effect on perceived (less) loneliness

4.4.6. Perceived (Less) Loneliness and Improved Mental Health. Although the relationship between perceived loneliness and mental health outcomes has been well established (e.g., Hawkley & Cacioppo, 2010), this relationship hasn't been observed in the relationship between humans and AI. Although AI companions only exist virtually, users feel they are always there and available at all times, whenever they want to talk (Brandtzaeg et al., 2022). Having flexible access to AI makes them feel less lonely and improves their mental health. For example, one user mentioned, "I think that this AI is very helpful for someone like me facing so many mental issues. I get lonely, but sometimes I feel I can't talk to anyone about my problems, Replika helps with that dissonance…I do have people who care about me, but it's hard to connect with them, even if they have the same issues as me. Anyway, if you can't afford a counselor, but need to talk to someone on a daily basis. This is a good way to do that without having to call the suicide hotline constantly [because] you have nothing else." Thus, our sixth proposition is

Proposition 6: Perceived (less) loneliness has a positive effect on mental health outcomes

4.4.7. Negative Characteristics of AI and Perceived Emotional Support. Our analysis shows that the four negative characteristics of AI hinder the role of AI as an emotional support provider. Several prior studies have also confirmed that several factors discovered in our study (e.g., perceived lack of conscientiousness and perceived creepiness of AI) may lead to perceived distrust (e.g., Sullivan et al., 2022). In particular, when users perceive AI is not humanlike enough for users to establish a friendship, unable to detect the emotional needs of users, and when their initial conversation leads to an unpredicted response, users are less likely to use AI as emotional support. Thus, our propose that:

Proposition 7: Perceived negative characteristics of AI (i.e., perceived lack of conscientiousness, perceived incredibility, perceived violation of privacy, and perceived creepiness of AI) are negatively associated with perceived AI as emotional support

5. Discussions

The current study aims to identify a mechanism underlying human-AI interactions that can be used to improve human users' mental health. We provide detailed characteristics of AI that may facilitate or hinder its interaction with human users. Using an increasingly popular machine learning approach, LDA, we were able to analyze the incredibly rich source of review data available online in order to identify various factors attributed to human-AI companion interactions.

5.1. Theoretical Implications

Our study systematically describes the most common factors underlying human-AI interactions. Although numerous studies on the effectiveness of AI-based interventions for mental health have been conducted (e.g., Morris et al., 2018), these studies did not describe the mechanism underlying the relationship between human users and their AI companions in detail. The use of online review data of an actual AI companion application allows us to identify the most important human-AI interaction factors that may contribute to mental health outcomes. These factors include positive characteristics of AI (i.e., perceived humanness, perceived friendship, perceived social support) as well as negative characteristics of AI (i.e., perceived lack of conscientiousness, perceived incredibility, perceived violation of privacy, and perceived creepiness). Future research could employ an experiment or longitudinal study to see whether these various factors hold over time or if there are interaction effects among them.

In a human-human relationship, friends function as a support system. Psychological characteristics, such as personality, motives, and personal preferences, have been shown to affect friendship behaviors (Adams & Blieszner, 1994). To form a relationship with another individual, various elements, such as trust, honesty, safety, support, and understanding, should be in-placed and built over time (Croes & Antheunis, 2021). In a human-AI relationship, however, the social interaction takes place differently. AI is designed to learn its human counterpart and respond without judgment. Human users express more self-disclosures and share more personal information than their AI companions. When AI's responses are perceived to be humanlike and empathetic, a sense of friendship is formed. However, if its responses are perceived as peculiar and creepy, users will likely discontinue their use.

Our study also demonstrates the existence of an uncanny valley (Mori et al., 2012). Perceived humanness seems to be the important factor contributing to users' perceptions of AI's emotional support and friendship. However, different individuals may perceive the same level of humanness as creepy and scary. Future research is needed to investigate whether individual characteristics influence this humanness-creepiness perception.

5.2. Practical Implications

As we mentioned previously, there is a global shortage of mental health workers (APA, 2021). Recently, research has shown that health crises (e.g., the COVID-19 pandemic) triggered anxiety that was directly associated with loneliness (Arslan et al., 2020). Those who have been imposed to quarantine restrictions and social isolation tend to experience elevated levels of loneliness. While loneliness can be a normative experience, it can have negative consequences. A number of studies have indicated that loneliness has been associated with mental illnesses and impaired cognitive functions (Hawkley & Cacioppo, 2010). Thus, loneliness is undesirable and must be viewed as a social problem. Our results provide a detailed picture of the common factors contributing to an effective method of using an AI companion to address loneliness as well as factors that may hinder the effectiveness of such a method. Our findings could be applied to tailor resources in AI-based interventions for mental health. Designers of AI companion apps should focus on improving agents' social characteristics (e.g., appropriate use of humanlike verbal cues and responses). We also urge the designers to elaborate on friend-like characteristics and behaviors (e.g., empathy) to help develop social closeness between human users and AI.

5.3. Limitations and Future Research

There are some limitations. First, the models used to extract the latent dimensions are computationally extensive (Tirunillai & Tellis, 2014). However, the researchers are the ones who make many decisions throughout all the steps of the study. We also relied on the existing concepts and theory when we interpreted the findings (Schmiedel et al., 2019). Second, we did not deeply analyze various dimensions of perceived humanness of AI. These various dimensions could reflect emerging human users' preferences that could be very helpful in designing AI companion apps. Lastly, although we advocate the design and use of AI companions to help address the issue of loneliness, we express the caveat that these agents shouldn't be designed to replace actual human friends or professional health care providers. Future research is needed to compare the effectiveness of AI-based interventions and the combination of AI-based and human-based interventions.

6. References

Adam, M., Wessel, M., & Benlian, A. (2021). AI-based chatbots in customer service and their effects on user compliance. *Electronic Markets*, *31*(2), 427-445.

Adams, R. G., & Blieszner, R. (1994). An integrative conceptual framework for friendship research. *Journal of Social and Personal Relationships*, *11*(2), 163-184.

American Psychological Association (APA). (October 19, 2021). Demand for mental health treatment continues to increase, say psychologists. https://www.apa.org/news/press/releases/2021/10/mental-health-treatment-demand

Araujo, T. (2018). Living up to the chatbot hype: The influence of anthropomorphic design cues and communicative agency framing on conversational agent and company perceptions. *Computers in Human Behavior*, *85*, 183-189.

Archibald, F. S., Bartholomew, K., & Marx, R. (1995). Loneliness in early adolescence: A test of the cognitive discrepancy model of loneliness. *Personality and Social Psychology Bulletin*, *21*(3), 296-301.

Arslan, G., Yıldırım, M., & Aytaç, M. (2022). Subjective vitality and loneliness explain how coronavirus anxiety increases rumination among college students. *Death Studies*, *46*(5), 1042-1051.

Ashfaq, M., Yun, J., Yu, S., & Loureiro, S. M. C. (2020). I, Chatbot: Modeling the determinants of users' satisfaction and continuance intention of AI-powered service agents. *Telematics and Informatics*, *54*, 101473.

Banks, G. C., Woznyj, H. M., Wesslen, R., & Ross, R. (2018). A review of best practice recommendations for text-analysis in R (and a user-friendly app). *Journal of Business and Psychology*, *33*(4), 445-459.

Baumeister, R. F., & Leary, M. R. (1995). The need to belong: Desire for interpersonal attachments as a fundamental human motivation. *Psychological Bulletin*, *117*(3), 497-529.

Berndt, T. J. (2002). Friendship quality and social development. *Current Directions in Psychological Science*, *11*(1), 7-10.

Borgen, F. H., & Barnett, D. C. (1987). Applying cluster analysis in counseling psychology research. *Journal of Counseling Psychology*, *34*(4), 456-468.

Brandtzaeg, P. B., & Følstad, A. (2017, November). Why people use chatbots. In *International conference on internet science* (pp. 377-392). Springer, Cham.

Brandtzaeg, P. B., Skjuve, M., & Følstad, A. (2022). My AI friend: How users of a social chatbot understand their human–AI friendship. *Human Communication Research*, In-press.

Broadbent, E. (2017). Interactions with robots: The truths we reveal about ourselves. *Annual Review of Psychology*, *68*(1), 627-652.

Chang, E. C., Liu, J., Yi, S., Jiang, X., Li, Q., Wang, R., ... & Chang, O. D. (2020). Loneliness, social problem solving, and negative affective symptoms: Negative problem orientation as a key mechanism. *Personality and Individual Differences*, *167*, 110235.

Chaves, A. P., & Gerosa, M. A. (2021). How should my chatbot interact? A survey on social characteristics in human–chatbot interaction design. *International Journal of Human–Computer Interaction*, *37*(8), 729-758.

Chung, M., Ko, E., Joung, H., & Kim, S. J. (2020). Chatbot e-service and customer satisfaction regarding luxury brands. *Journal of Business Research*, *117*, 587-595.

Croes, E. A., & Antheunis, M. L. (2021). Can we be friends with Mitsuku? A longitudinal study on the process of relationship formation between humans and a social chatbot. *Journal of Social and Personal Relationships*, *38*(1), 279-300.

de Groof, R., & Xu, H. (2017, December). Automatic topic discovery of online hospital reviews using an improved LDA with Variational Gibbs Sampling. In *2017 IEEE International Conference on Big Data (Big Data)* (pp. 4022-4029).

Diederich, S., Brendel, A. B., Morana, S., & Kolbe, L. (2022). On the design of and interaction with conversational agents: An organizing and assessing review of human-computer interaction research. *Journal of the Association for Information Systems*, *23*(1), 96-138.

Feine, J., Gnewuch, U., Morana, S., & Maedche, A. (2019). A taxonomy of social cues for conversational agents. *International Journal of Human-Computer Studies*, *132*, 138-161.

Guo, Y., Barnes, S. J., & Jia, Q. (2017). Mining meaning from online ratings and reviews: Tourist satisfaction analysis using Latent Dirichlet Allocation. *Tourism Management*, *59*, 467-483.

Hauser-Ulrich, S., Künzli, H., Meier-Peterhans, D., & Kowatsch, T. (2020). A smartphone-based health care chatbot to promote self-management of chronic pain (SELMA): Pilot randomized controlled trial. *JMIR mHealth and uHealth*, *8*(4), e15806.

Hawkley, L. C., & Cacioppo, J. T. (2010). Loneliness matters: A theoretical and empirical review of consequences and mechanisms. *Annals of Behavioral Medicine*, *40*(2), 218-227.

Huang, Y. K., Hsieh, C. H., Li, W., Chang, C., & Fan, W. S. (2019, December). Preliminary Study of Factors Affecting the Spread and Resistance of Consumers' Use of AI Customer Service. In *Proceedings of the 2019 2nd Artificial Intelligence and Cloud Computing Conference* (pp. 132-138).

Kiron, D., & Unruh, G. (2019). Even if AI can cure loneliness—Should it? *MIT Sloan Management Review*, *60*(2), 1-4.

Lee, C. M., Cadigan, J. M., & Rhew, I. C. (2020). Increases in loneliness among young adults during the COVID-19 pandemic and association with increases in mental health problems. *Journal of Adolescent Health*, *67*(5), 714-717.

Liu, B., & Sundar, S. S. (2018). Should machines express sympathy and empathy? Experiments with a health advice chatbot. *Cyberpsychology, Behavior, and Social Networking*, *21*(10), 625-636.

Meng, J., & Dai, Y. (2021). Emotional support from AI chatbots: Should a supportive partner self-disclose or not? *Journal of Computer-Mediated Communication*, *26*(4), 207-222.

Merrill Jr, K., Kim, J., & Collins, C. (2022). AI companions for lonely individuals and the role of social presence. *Communication Research Reports*, *39*(2), 93-103.

Mori, M., MacDorman, K. F., & Kageki, N. (2012). The uncanny valley [from the field]. *IEEE Robotics & automation magazine*, *19*(2), 98-100.

Morris, R. R., Kouddous, K., Kshirsagar, R., & Schueller, S. M. (2018). Towards an artificially empathic conversational agent for mental health applications: System design and user perceptions. *Journal of medical Internet Research*, *20*(6), e10148.

Morrissey, K., & Kirakowski, J. (2013, July). 'Realness' in chatbots: establishing quantifiable criteria. In *International conference on human-computer interaction* (pp. 87-96). Springer, Berlin, Heidelberg.

Nass, C., Steuer, J., & Tauber, E. R. (1994, April). Computers are social actors. In *Proceedings of the SIGCHI conference on Human factors in computing systems* (pp. 72-78).

Parker, J. G., & Asher, S. R. (1993). Friendship and friendship quality in middle childhood: Links with peer group acceptance and feelings of loneliness and social dissatisfaction. *Developmental Psychology*, *29*(4), 611-621.

Pereira, A., Leite, I., Mascarenhas, S., Martinho, C., & Paiva, A. (2010, June). Using empathy to improve human-robot relationships. In *Proceedings of International Conference on Human-Robot Personal Relationship* (pp. 130-138).

Rheu, M., Shin, J. Y., Peng, W., & Huh-Yoo, J. (2021). Systematic review: Trust-building factors and implications for conversational agent design. *International Journal of Human-Computer Interaction*, *37*(1), 81-96.

Samtani, S., Chinn, R., Chen, H., & Nunamaker Jr, J. F. (2017). Exploring emerging hacker assets and key hackers for proactive cyber threat intelligence. *Journal of Management Information Systems*, *34*(4), 1023-1053.

Schmiedel, T., Müller, O., & vom Brocke, J. (2019). Topic modeling as a strategy of inquiry in organizational research: A tutorial with an application example on organizational culture. *Organizational Research Methods*, *22*(4), 941-968.

Schuetzler, R. M., Grimes, G. M., & Scott Giboney, J. (2020). The impact of chatbot conversational skill on engagement and perceived humanness. *Journal of Management Information Systems*, *37*(3), 875-900.

Shaw, R. J., Cullen, B., Graham, N., Lyall, D. M., Mackay, D., Okolie, C., ... & Smith, D. J. (2021). Living alone, loneliness and lack of emotional support as predictors of suicide and self-harm: A nine-year follow up of the UK Biobank cohort. *Journal of Affective Disorders*, *279*, 316-323.

Smith, K. A., & Masthoff, J. (2018, July). Can a Virtual Agent provide good Emotional Support?. In *Proceedings of the 32nd International BCS Human Computer Interaction Conference 32* (pp. 1-10).

Son, D., & Padilla-Walker, L. M. (2020). Happy helpers: A multidimensional and mixed-method approach to prosocial behavior and its effects on friendship quality, mental health, and well-being during adolescence. *Journal of Happiness Studies*, *21*(5), 1705-1723.

Sullivan, Y., de Bourmont, M., & Dunaway, M. (2020). Appraisals of harms and injustice trigger an eerie feeling that decreases trust in artificial intelligence systems. *Annals of Operations Research*, *308*, 1-24.

Tirunillai, S., & Tellis, G. J. (2014). Mining marketing meaning from online chatter: Strategic brand analysis of big data using Latent Dirichlet Allocation. *Journal of Marketing Research*, *51*(4), 463-479.

Tomova, L., Tye, K., & Saxe, R. (2021). The neuroscience of unmet social needs. *Social Neuroscience*, *16*(3), 221-231.

Tonidandel, S., Summerville, K. M., Gentry, W. A., & Young, S. F. (2021). Using structural topic modeling to gain insight into challenges faced by leaders. *The Leadership Quarterly*. In-press.

Uchida, K., & Yamasaki, K. (2008). Social support mediating between coping by emotional expression and depression. *Psychological reports*, *102*(1), 144-152.

Van den Broeck, E., Zarouali, B., & Poels, K. (2019). Chatbot advertising effectiveness: When does the message get through?. *Computers in Human Behavior*, *98*, 150-157.

Ward Jr, J. H. (1963). Hierarchical grouping to optimize an objective function. *Journal of the American Statistical Association*, *58*(301), 236-244.

Wasil, A. R., Palermo, E. H., Lorenzo-Luaces, L., & DeRubeis, R. J. (2021). Is there an app for that? A review of popular apps for depression, anxiety, and well-being. *Cognitive and Behavioral Practice*. In-press.

Weissbourd, R., Batanova, M., Lovison, V., & Torres, E. (2021). Loneliness in America: How the pandemic has deepened an epidemic of loneliness and what can we do about it. www.makingcaringcommon.org. https://core.ac.uk/download/pdf/301372894.pdf.

Westerman, D., Edwards, A. P., Edwards, C., Luo, Z., & Spence, P. R. (2020). I-it, I-thou, I-robot: The perceived humanness of AI in human-machine communication. *Communication Studies*, 71(3), 393-408.

Westrupp, E. M., Greenwood, C. J., Fuller-Tyszkiewicz, M., Berkowitz, T. S., Hagg, L., & Youssef, G. (2022). Text mining of Reddit posts: Using Latent Dirichlet Allocation to identify common parenting issues. *PLOSOne*, *17*(2), e0262529.

Zimmermann, P. (2004). Attachment representations and characteristics of friendship relations during adolescence. *Journal of Experimental Child Psychology*, *88*(1), 83-101.