ABSTRACT

Title of Dissertation:	DESIGN AND EVALUATION OF A CONVERSATIONAL AGENT FOR MENTAL HEALTH SUPPORT: FORMING HUMAN-AGENT SOCIOTECHNICAL AND THERAPEUTIC RELATIONSHIPS
	Yuting Liao, Doctor of Philosophy, 2021
Dissertation directed by:	Associate Professor, Jessica Vitak College of Information Studies

Many people with mental health disorders face significant challenges getting the help they need, including the costs of obtaining psychological counseling or psychiatry services, as well as fear of being stigmatized. As a way of addressing these barriers, text-based conversational agents (chatbots) have gained traction as a new form of e-therapy. Powered by artificial intelligence (AI) and natural language processing techniques, this technology offers more natural interactions and a “judgment-free zone” for clients concerned about stigma. However, literature on psychotherapeutic chatbots is sparse in both the psychology and human computer interaction (HCI) fields. While recent studies indicate that chatbots provide an affordable and effective therapy delivery method, this research has not thoroughly explained the underlying mechanisms for increasing acceptance of chatbots and making them more engaging.

Don Norman (1994) has argued the main difficulties of utilizing intelligent agents are social—not technical—and particularly center around people's perceptions of agents. In exploring the use of chatbots in psychotherapy, we must investigate how this technology is conceptually understood, and the thoughts and feelings they evoke when people interact with them.

This dissertation focuses on two types of relationships critical to the success of utilizing chatbots for mental health interventions: sociotechnical relationships and therapeutic relationships. A sociotechnical relationship concerns technology adoption, usability, and the compatibility between humans and chatbots. A therapeutic relationship encompasses people's feelings and attitudes toward a chatbot therapist. Therefore, this dissertation asks: What are the optimal design principles for a conversational agent that facilitates the development of both sociotechnical and therapeutic relationships to help people manage their mental health? To investigate this question, I designed an original conversational system with eight gendered and racially heterogeneous personas, and one neutral robot-like persona. Using a mixed-method approach (online experiment and interviews), I evaluated factors related to the adoption and use of conversational agents for psychotherapeutic purposes. I also unpacked the human-agent relational dynamics and evaluated how anthropomorphism and perceived racial similarity impact people's perceptions of and interactions with the chatbot. These findings contributed to the wider understanding of conversational AI application in mental health support and provided actionable design recommendations.

DESIGN AND EVALUATION OF A CONVERSATIONAL AGENT FOR MENTAL HEALTH SUPPORT: FORMING HUMAN-AGENT SOCIOTECHNICAL AND THERAPEUTIC RELATIONSHIPS

by

Yuting Liao

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park, in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2021

Advisory Committee:
  Professor Jessica Vitak, Chair
  Professor Beth St. Jean
  Professor Amanda Lazar
  Professor Eun Kyoung Choe
  Professor Dennis Kivlighan, Dean’s Representative

© Copyright by
Yuting Liao
2021

Dedication

I dedicate this dissertation to my family.

To my mother and father, without your support, I would not be here in the United States. Without your unconditional love, I would not have the courage to pursue what I want, including this doctoral degree.

To my husband, Chris, who has been there for me through all the ups and downs. You are the sunshine in my life. I could not have done it without your love and care.

To my grandmother, who was deprived of the opportunity to get education due to gender inequality. I know you’re proud of me.

I am blessed to have all the love and support from my family in China and in the US.

This is for all of you.

Acknowledgement

It takes a village to get here. I am thankful to all of you.

First, to my advisor and mentor, Jessica Vitak. You taught me to conduct rigorous quantitative and qualitative research. You helped me connect the dots and refine my ideas. You set a standard of work that will forever guide me. And I'm so grateful for all the amazing opportunities you offered me; I got to travel the world, and I got to be a part of the Privacy Education and Research Lab because of you believing in me. The depth and breadth of privacy research experiences introduced me to the world of human-centered design, igniting my passion for user experience research. Thank you for guiding me through all these years–even in 2020, the particularly tough year for you. I'm forever indebted.

To my other advisor Beth St. Jean. Your kindness and patience helped me in ways beyond words. Your work and advocacy for health justice sparked my idea for this research. Thank you for encouraging me to explore and pursue my ideas. And thank you for including me in many meaningful projects, and I even got to edit a book. I'll never forget how you have supported me.

To my committee, I'm grateful for your time, expertise, and support. Dennis Kivlighan, thank you for bringing the perspective from the Psychology field; Amanda Lazar and Eun Kyoung Choe, thank you for your expertise in health informatics and human computer interaction.

To Jiangen He, thank you for listening to my research ideas and your tremendous help, without which my chatbot system wouldn't be a reality.

To the Mobile Privacy team, Priya Kumar, Anouk Mols, Michael Zimmer, Jason Pridmore, Daniel Trottier, and Katie Kritikos, I smile every time I think of you. You are part of the reason that I know doing research is fun. I miss all the wonderful time we spent together in the summers. And especially Priya, my academic sister, I miss our chat between the breaks in our “office.”

To the Safe Data | Safe Families team, Mega Subramaniam, Shandra Morehouse, and Andrea Castillo, thank you for the opportunity to learn more about the marginalized communities in this country with you.

To Tara Cheng Tiantian, my confidant and role model, thank you for inspiring and motivating me to be the best version of myself.

To Xiyin Liu and Yueling Ji, my best friends from my high school and from my college, who are also pursuing their doctoral degree. Even though we are thousands of miles apart, your companionship and support carry me forward.

To my friend and comrade Jiahui Wu, thank you for your help in data analytics, not just for academic work but also for practical ends like stock market.

To Gagan Jindal, thank you for the encouragement and for taking walks with me when I desperately wanted to get out during the pandemic.

To the wonderful teachers I learnt so much from: Yla Tausczik, thank you for the best Statistic class; Vanessa Frias-Martinez, thank you for introducing me to the world of Data Science. To Kari Kraus, thank you for the doctoral seminar, the Engaged Intellectual. To many other scholars who helped me through this process, whether through friendship, a kind word, advice, feedback, and more: Jeanine Turner, Leticia Bode, Lili Wang, Yanjun Guan, Hailong Liu, Shuya Pan, Shangwei Wu, Lingzi Hong, Myeong Lee, Yuhan Luo, Xiaoyun Huang, Alisha Pradhan, Shijie Song, Yuxiang Zhao, Xin Qian, Richard Marciano, Brian Butler, Susan Winter, Leyla Norooz, Diane Travis and more.

Thank you all.

Table of Contents

Dedication ii
Acknowledgement iii
Table of Contents vi
List of Tables viii
List of Figures ix
Chapter 1: Introduction 1
  1.1. Research Gaps in the Therapeutic and Relational Dynamics of Chatbots 8
  1.2. Study Goals 12
  1.3. Significance of Study 15
  1.4. Research Questions 17
Chapter 2: Literature Review 19
  2.1. Overview of Chatbot Technology 19
  2.2. Promises and Limitations of Therapeutic Chatbots 26
  2.3. Therapeutic Chatbot Engagement, Acceptance and Use 35
  2.4. Anthropomorphism of Chatbots 40
  2.5. Importance of the Therapeutic Relationship 49
  2.6. Client Self-disclosure and Privacy 52
  2.7. The Role of Race 61
Chapter 3: System Design of a Chatbot for Psychotherapy 65
  3.1. Chatbot Technical Configuration 65
  3.2. Conversation Design 67
  3.3. Chatbot Agent's Racial Identity 70
Chapter 4: Methodology 74
  4.1. Online Experiment 75
  4.2. Survey Instrument 80
  4.3. Interview Study 89
Chapter 5: Results 91
  5.1. What are the benefits and drawbacks of utilizing a chatbot for managing mental health compared to a human therapist? (RQ1) 92
  5.2. What factors influence the sociotechnical relationship between human and chatbot for mental health support? (RQ2) 120
  5.3. How does anthropomorphism influence the way people perceive and interact with a chatbot in psychotherapeutic conversations? (RQ3) 129
  5.4. What factors are associated with therapeutic relationship development between client and conversational agents? (RQ4) 135
  5.5. What is the effect of the therapeutic alliance on human-agent psychotherapeutic conversations? (RQ5) 144
  5.6. How does anthropomorphism relate to people's sense of being understood after making disclosures to the agent? (RQ6) 146
  5.7. How comfortable do people feel about sharing their information with a chatbot in psychotherapeutic conversations? (RQ7) 148
  5.8. How does racial mirroring influence people's perceptions of and interactions with a chatbot in psychotherapeutic conversations? (RQ8) 155

Chapter 6: Discussion 163
6.1. Summary of Findings and Theoretical Contributions.................................. 163
6.2. A Need-based Assessment of Chatbot for Mental Health Support............... 170
6.3. Design Recommendations 177
6.4. Ethical Considerations 184
Chapter 7: Conclusion... 187
7.1. Summary of Dissertation 187
7.2. Limitations and Future Research 188
Appendices.. 190
Appendix I Example Dialogue... 190
Appendix II: Screening Survey.. 192
Appendix III: Main Survey.. 200
Appendix IV: Interview Study Protocol 220
References... 223

List of Tables

Table 1. Sample chatbot questions and responses by MI process. 68
Table 2. Sample dialogues for linguistic characteristics.. 68
Table 3. Visual cues indicating agent racial identity 71
Table 4. Mapping methods to research questions... 74
Table 5. Survey participants socio-demographics background 88
Table 6. Interview participants demographics and chatbot interaction background .. 90
Table 7. Technology-level benefits.. 93
Table 8. Technology-level drawbacks 98
Table 9. Social-level benefits.. 102
Table 10. Social-level drawbacks 106
Table 11. Therapeutic-level benefits.. 108
Table 12. Therapeutic-level drawbacks 112
Table 13. Interactional-level benefits 115
Table 14. Interactional-level drawbacks.. 118
Table 15. Descriptive statistics and correlations for variables in model 123
Table 16. Significance levels of the multivariate test.. 123
Table 17. Effects of covariates on sociotechnical relationship................................ 124
Table 18. Effects of chatbot gender on sociotechnical relationship 126
Table 19. Descriptive statistics of four dimensions of anthropomorphism 130
Table 20. Models explaining four dimensions of anthropomorphism..................... 131
Table 21. Significance levels of the multivariate test.. 136
Table 22. Effects of covariates on therapeutic relationships 136
Table 23. Effects of racial mirroring condition on therapeutic relationships 139
Table 24. Multiple linear regression model predicting level of engagement 144
Table 25. Descriptive statistics of statements evaluating self-disclosure comfort ... 149
Table 26. Comfort with data sharing based on data type... 150
Table 27. Models predicting level of conversation engagement 151
Table 28. Traditional mental health and psychosocial needs chatbot fills (or not) .. 173
Table 29. Additional therapeutic needs chatbot fills 174
Table 30. Design recommendations for future development of therapeutic chatbot 178

List of Figures

Figure 1. Screenshot of participant conversation with agent Molly. 14
Figure 2. Conversation with Woebot 27
Figure 3. Conversation with Replika 32
Figure 4. Downsized icon for the user vs. Larger icon for the agent's profile. 66
Figure 5. Two examples of predefined options for providing tailored responses. 70
Figure 6. A visualization of standardized chi-square residuals. 159
Figure 7. The pyramid of mental health and psychosocial needs in traditional psychotherapy 172
Figure 8. A new need-based framework in chatbot-based therapy 175
Figure 9. Design Recommendation: Tailoring conversation flow to user needs 179
Figure 10. Design Recommendation: Providing alternative solutions as a fallback 180
Figure 11. Design recommendation: Customizing agent identity and personality ... 183
Figure 12. Design Recommendation: Providing transparency and control to mitigate data privacy concerns 184

Chapter 1: Introduction

There is a worldwide gap between patient needs and the provision of mental health services; while one in 10 people has a mental health disorder, only 1% of the global health workforce provides mental health services (Keynejad et al., 2018). In the United States, an estimated 47.6 million adults have a mental illness (such as depression or anxiety), 11.4 million of whom have a serious mental illness substantially limiting one or more major life activities. At the same time, 11.2 million people have reported an unmet need for mental health care, with one-third of adults with severe mental illness never receiving any mental health services (Lipari, 2018).

Many people with mental health issues face significant challenges getting the help they need. Psychological counseling or psychiatry services are a luxury expense for people under financial stress (Karlin et al., 2008). Based on a 2018 national survey by the U.S. Department of Health and Human Services, more than five million American adults, including 2.8 million with severe mental illness, did not receive the mental health services they needed because they could not afford the cost of care (Lipari, 2018). Beyond structural barriers, a fear of being stigmatized also prevents people from seeking help for mental health concerns (Lannin et al., 2013).

To expand access to mental health services and counteract the stigma associated with mental health conditions, there has been a burgeoning growth in internet-based mental health interventions, or e-therapy. E-therapy refers to mental health services delivered online or through digital technology. According to Paganini et al. (2016), all approaches covered by the German Psychotherapy Guidelines, such as psychodynamic or cognitive behavioral therapy, can be suitable for translation into an online format; the approaches of interpersonal therapy, acceptance and commitment therapy, and mindfulness-based therapy have also been transformed into either guided or unguided self-help interventions (Paganini et al., 2016). A meta-analysis of 92 studies shows that online interventions can be as effective as face-to-face psychotherapy in helping clients (Barak et al., 2008).

Over the past few decades, e-therapy techniques with varying degrees of human guidance have been developed. In the early days, e-therapy usually involved a mental health care professional providing counseling via email, text messages, online chat, or video conferencing (Manhal-Baugus, 2001; Sucala et al., 2012). Compared to traditional face-to-face therapy, e-therapy has several advantages. First, it enables remote communication either in real time or asynchronously; the resulting flexibility across space and time contributes to wider access to mental health services at a lower price. This results in democratizing access to psychotherapeutic resources for people who cannot afford it otherwise.

E-therapy also allows clients to receive psychotherapy in the privacy of their home. People may be more willing to engage in e-therapy, as their perception of anonymity mitigates their fear of judgment and stigmatization, the primary reason that people avoid professional psychological help (Corrigan, 2004). Moreover, the perception of anonymity can ease people's discomfort associated with their divulgence of embarrassing, stigmatizing, or sensitive thoughts and behaviors. In turn, clients are more ready to discuss deep, personal issues (Finfgeld, 1999; Manhal-Baugus, 2001; Meier, 1988), which is important because greater levels of self-disclosure have been linked with emotional and psychological benefits (Greenberg & Stone, 1992), as well as better therapy outcomes (Farber, 2003).

With new technological advancements, e-therapy can now be implemented as a fully automated intervention without any human involvement through virtual reality (Fernández-Álvarez et al., 2019; Valmaggia et al., 2016), serious games (Fleming et al., 2015), mobile apps, and phone- and wearable-sensors (Ebert et al., 2018; J. Lin et al., 2018). However, these fully automated e-therapy methods suffer from relatively poor adoption and adherence (Arnold et al., 2019; Donkin et al., 2013); preliminary research shows that patient dropout rates tend to be higher for computer-based CBT therapy with little or no clinician involvement (Eells et al., 2015; Melville et al., 2010). Fitzpatrick (2017) posits that this lack of adherence may be attributed to the loss of the human touch that in-person therapy offers. The Lancet Commission on Global Mental Health warns that "technology-based approaches might improve the reach of mental health services but could lose key human ingredients and, possibly, lower effectiveness of mental health care" (Patel et al., 2018, p. 23). From a human-computer interaction (HCI) perspective, Doherty et al. (2012) attribute the high drop-out rate to an absence of interactive experiences, as people feel less engaged with static content provided in online intervention programs. An interactive design strategy, therefore, is proposed to create e-therapy technology that enables people to engage actively rather than passively receive content (Doherty et al., 2012).

More recently, chatbots have gained traction as an innovative e-therapy technology. Some prominent examples emerging from industry and academia include Woebot (woebot.io) and Wysa (wysa.io). These chatbots are computer programs capable of holding conversations with humans; the technological underpinnings range from simple rule-based algorithms to more sophisticated artificial intelligence. Regardless of their technical complexity, therapeutic chatbots present a more natural, dynamic form of interaction compared to other automated e-therapy formats. More importantly, user engagement with a therapeutic chatbot is associated with feelings of companionship and emotional support. According to an analysis of 1,854 user-generated reviews on a popular chatbot called Replika (Ta et al., 2020), the chatbot helps to reduce feelings of loneliness and provides a “safe space” in which users can discuss any topic without the fear of judgment or retaliation. When normal sources of informational support are not available, the chatbot provides helpful guidance on managing mental health through uplifting and nurturing messages.

Currently, most therapeutic chatbots rely on predefined conversation scripts to conduct dialogues. These scripts, including various ways of asking questions and responding to clients in order to address psychological issues, are usually designed by psychologists following well-evaluated principles of psychotherapy. For instance, Woebot is designed to automate cognitive behavioral therapy, identifying and addressing unhelpful ways of thinking. Built upon conversation scripts, chatbot programs use algorithms (e.g., if-then rules or keyword matching) to carry on interactions similar to psychotherapy discussions in which clients disclose how they are feeling and what is going on in their lives (Becker, 2018; D’alfonso et al., 2017). Aside from daily chat-like conversations, therapeutic chatbots use mood tracking, curated videos, and word games to help people manage their mental health (Molteni, 2017). These apps can also visualize patterns in the data with more accuracy than human memory can afford.

Yet, a chatbot cannot replace a human when a client requires more complex and intense mental healthcare. A chatbot therapist is not a licensed physician who makes diagnoses or writes prescriptions, nor is it capable of dealing with real mental health crises—for example, when it senses someone is expressing suicidal intent, it suggests they seek help in the real world and provides hotline resources. Despite these limitations, engagement with therapeutic chatbots has been found to have a positive effect on people's mental health. One random controlled trial showed that a chatbot provided an effective therapeutic experience, reducing users' feelings of depression and anxiety (Fitzpatrick et al., 2017). Another qualitative longitudinal study showed that a chatbot was useful for detecting and alleviating adolescents' emotional stress (Huang et al., 2015).

Although therapeutic chatbots like Woebot are unlikely to replace human therapists anytime soon, chatbot technology could have significant impacts on mental health care delivery. First, like other e-therapy formats, chatbot technology demonstrates several advantages over traditional therapy in terms of convenience, affordability, and anonymity. For those who are self-conscious about pursuing mental health treatment, the self-directed nature of these chatbot apps empowers people to take control of their emotional health on their own terms (Sojit, 2020). The convenience of smartphone apps is also unparalleled: unlike a real therapist, they're available 24/7. Additionally, chatbot technology appears to be more effective than other types of fully automated mental health interventions. Using a random controlled trial, Fitzpatrick et al. (2017) found evidence that use of the conversational agent Woebot is more effective than a digital educational program in reducing people’s symptoms of depression and anxiety. Similarly, in another controlled experiment, Suganuma et al. (2018) found that treatment guided by a conversational agent is more effective in reducing user stress and improving their mental health than an automated e-therapy without conversation capability. Furthermore, studies have found increased engagement and adherence among users of chatbot-based mental health interventions (Kretzschmar et al., 2019).

Researchers attribute the competitive edge of therapeutic chatbots to their conversational interface, an interaction mode that evokes experiences that resemble human real-life contacts (Fitzpatrick et al., 2017; Miner et al., 2019). When people engage with chatbots, they might attribute human characteristics to the system—a process known as anthropomorphism (Araujo, 2018; Ciechanowski et al., 2019). In some cases, people have even become convinced that they were interacting with other humans (Ly et al., 2017; Weizenbaum, 1976). This tendency to see a human in a chatbot seems to be associated with the fact that a chatbot can hold a conversation (Angeli et al., 2001). After all, natural language was unique to human communication until the emergence of computer programs that “talk.” Generally, anthropomorphism has been found to be useful for developing lasting and meaningful social relationships with users. Researchers have noted that the more a chatbot is perceived as inhuman or “weird,” the less favorable and the less competent it seemed to participants (Ciechanowski et al., 2019).

Within human-computer interaction (HCI) research, there have been considerable efforts to equip chatbots with human-like linguistic and conversation styles, such as empathic and emotional expressions (Dohsaka et al., 2009) and humor (Morkes et al., 1999). Recent iterations of chatbots even incorporate personality (Kang, 2018) and emotional intelligence, with the ability to detect and respond to human emotions (Brewster, 2016; Huang et al., 2015). In the context of promoting mental health, the ability of a chatbot to express empathy appears to be critical for fostering positive psychotherapeutic interactions between a user and a chatbot (Bickmore et al., 2010; Brixey et al., 2017; Morris et al., 2018).

Beyond conversation capabilities, many chatbot systems are equipped with personified visual representations, or agents with human-like appearance varying in age, gender, and even clothing (T. Bickmore et al., 2010; Kang, 2018). However, it is important to note that a chatbot does not necessarily need a human persona to be anthropomorphized. Many early chatbots were entirely text-based and lacked a physical/visual representation, but still convinced people of their humanness. Furthermore, a human-like appearance does not necessarily guarantee that a chatbot will be considered more human-like. For instance, in one between-group experiment, participants rated a chatbot with a human face as less human-like and less favorable in comparison to a text-only chatbot (Ciechanowski et al., 2019).

For many researchers and product designers, the ultimate design goal for a chatbot is to have it pass the Turing test[1] and be perceived as a real human in

---

[1] The Turing test was originally designed by Alan Turing in 1950 to determine whether or not a computer counts as "intelligent." On one side of a computer screen sits a human judge, whose job is to chat with some mysterious interlocutor and determine whether it is a human or a computer. Passing the Turing test means that a machine has tricked the human judge into thinking that it is a real human.

conversations. Significant efforts have focused on the technical aspects of the system, for instance, by leveraging cloud computing (Chung & Park, 2019) and innovative algorithms (Pichponreay et al., 2016) to make chatbots more intelligent, capable of holding conversations in different domains. However, a chatbot does not need to pass the Turing Test to have a significant impact on the mental health of users (Miner et al., 2019). In fact, there are potential benefits when people perceive a chatbot as a machine versus a real therapist. One such benefit is clients' willingness to self-disclose, a factor that is crucial for successful therapy. Prior studies have found that people are more open with a conversational agent than with a human counselor in disclosing sensitive personal information and reporting mental health symptoms (Lucas et al., 2014; Tantam, 2006; Weisband & Kiesler, 1996).

1.1. Research Gaps in the Therapeutic and Relational Dynamics of Chatbots

Psychotherapeutic chatbots bring forth a unique research opportunity for both psychology and HCI; however, literature related to the use of chatbots in mental health intervention is sparse. Recent psychology scholarship began to evaluate the efficacy of using conversational agents for mental health interventions. While confirming that chatbots provide an affordable and effective method to deliver therapy, this line of research has not assessed the therapeutic relationship between humans and therapeutic chatbots. According to the American Psychological Association's (APA) guidelines, claiming that psychotherapies are effective without examining the therapeutic relationship are "seriously incomplete and potentially misleading on both clinical and empirical grounds" (Ackerman et al., 2001, p.495). The therapeutic relationship is defined as the feelings and attitudes that client and therapist have toward one another and the manner in which these are expressed (Gelso & Carter, 1985). One key component of this client-therapist relationship is the therapeutic alliance, a collaborative and affective bond between patient and provider (Horvath & Greenberg, 1989), which is associated with symptom improvement in psychotherapy (Flückiger et al., 2018). Although originally conceptualized for human-to-human conversations, studies have found that this bond can also be formed between humans and conversational agents (T. Bickmore et al., 2010; Ly et al., 2017).

Through a design and HCI perspective, my dissertation research sets out to understand how such human-agent bonds emerge and evolve in the process of psychotherapeutic interactions. At the core of this exploration is the formation and development of the human-computer relationship; specifically, how design characteristics of the chatbot influence the way therapeutic relationships are formed. Previous HCI researchers (Cassell & Bickmore, 2003; Donath, 2001) have highlighted the crucial importance of the agent's design characteristics in relationship building.

There are several dimensions that must be considered when it comes to the design of a conversational agent. A chatbot or conversational agent can be characterized by its communicative behaviors through verbal (Lee & Choi, 2017) or non-verbal language, such as head nods and smiles (Bickmore & Picard, 2005; Krämer et al., 2018); by its social characteristics, such as emotional intelligence and personality (Chaves & Gerosa, 2019); by its appearance (e.g., gender, age, body image); and by its functionalities (Diesbach & Midgley, 2007). Empirical studies on human-agent interactions have found that people’s attitudes toward a chatbot are indeed influenced by its design characteristics. For instance, people prefer agents who conform to gender stereotypes associated with tasks (Forlizzi et al., 2007); self-disclose and show reciprocity (Lee & Choi, 2017); and demonstrate a positive attitude and mood (Thies et al., 2017). These preferences for certain design characteristics, however, could differ across various contexts of chatbot use and individual users. In examining chatbot use for leisure-time, counseling, and task completion in Korea, Kang (2018) found that women prefer neutral-gender chatbots except in the context of counseling, while men tend to prefer female chatbots in all three contexts. This author further found that all participants preferred younger chatbots for leisure-time, but same age or older chatbots for counseling.

In exploring the effects of chatbot design characteristics on human-agent interactions within the psychotherapy context, prior HCI work has considered a limited number of factors; one such factor that has been frequently overlooked is race. The issue of race is largely understudied in human-agent interactions more broadly (Liao & He, 2020; Schlesinger et al., 2018). Nonetheless, racial identity shapes people’s social experiences, judgments, and their perceptions of both themselves and others (Chavez & Guido-DiBrito, 1999). For traditional in-person psychotherapy, clients tend to prefer and report higher satisfaction when interacting with a counselors of the same race as them (Cabral & Smith, 2011). This preference for racial match is particularly prominent among clients who are black, as their self-disclosures have been found to be more extensive when they are speaking with black therapists rather than white therapists (Thompson et al., 1994).

In order to address the need for client-therapist cultural congruence, matching therapists with clients of the same race/ethnicity has become an advocated practice and has been found to result in stronger therapeutic relationships and more positive attitudes toward therapy (Cabral & Smith, 2011). However, it remains unknown whether and how perceived racial similarity might impact sociotechnical and therapeutic relationships within agent-based psychotherapy. This dissertation intends to address this gap by exploring how the (mis)match between a user's race and an agent's racial presentation plays a role in shaping the therapeutic relationship between human and agent.

On the other hand, to fully leverage the use of chatbots in psychotherapy, it is imperative to investigate how this anthropomorphic technology is conceptually understood, and what kinds of thoughts and feelings are invoked when people interact with agents in therapeutic conversations. After all, designing a conversational agent that will be acceptable to intended users is not only a technological issue, but primarily a social problem to solve (Neururer et al., 2018; Norman, 1994). While we have learned that chatbots can be effective for delivering certain psychological treatments (Fitzpatrick et al., 2017; Ly et al., 2017), what remains largely unknown is why and how they might succeed (or fail) in engaging clients in psychotherapeutic conversations. Therefore, as opposed to evaluating treatment outcomes or the efficacy of chatbot-delivered psychotherapy (e.g., Fitzpatrick et al., 2017), the focus of this dissertation is on the design outcomes of a psychotherapeutic chatbot: the relational dynamics[2] between humans and chatbots.

1.2. Study Goals

To begin to address the research gaps, this dissertation focuses on how people perceive and interact with technology by examining two types of relationships users hold with a psychotherapeutic chatbot: sociotechnical relationships and therapeutic relationships. Sociotechnical relationships concern technology adoption, use, and usability; in contrast, therapeutic relationships are shaped by people's feelings and attitudes towards the therapist/chatbot. A strong sociotechnical relationship is indicated by favorable views and interactions with the psychotherapeutic *chatbot as a technology*, while a strong therapeutic relationship means that people trust the psychotherapeutic *chatbot as a therapist*. Both types of relational dynamics are critical for the success of utilizing chatbots for mental health interventions: a strong sociotechnical relationship facilitates uptake and adherence to the chatbot-based therapy; a robust therapeutic relationship advances the process and helps to improve treatment outcomes. The overarching question I ask in this dissertation is: What are the optimal design principles for a conversational agent that facilitates both sociotechnical and therapeutic relationships to help people manage their mental health?

Specifically, there are two research goals for this dissertation. The first goal is to examine the sociotechnical relationship between human and agent through the lens

---

[2] It is also important to point out that unlike interpersonal relationships, human-chatbot relationships are not dyadic, as they are reflected only in people's emotional, cognitive, and behavioral responses to the chatbot. Chatbots are not human, even though they might give the illusions of being so.

of the unified theory of acceptance and use of technology (UTAUT) to evaluate factors related to the adoption and use of conversational agents for psychotherapeutic purposes. The second goal of this dissertation is to unpack the human-agent therapeutic relationship by evaluating how people perceive and interact with chatbots, with an emphasis on the effects of anthropomorphism and racial matching.

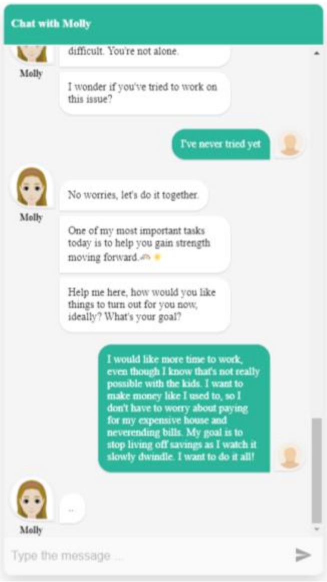
This study applies a mixed-method approach, including an online experiment and an interview study. The technological artifact involved is an original chatbot system designed to carry out therapeutic conversations. Because this system is associated with different personas with racial and gender diversity, I name it multi-identity conversational system for therapy (MICST)[3]. Distinct profile pictures were used to provide strong visual cues to indicate agents' racial identities, including White/Caucasian, Black/African American, Latinx, and Asian. Due to mixed evidence of cross-gender effects on chatbot perceptions and following prior work (Schuetzler et al., 2019), I created both female and male personas for each racial group. Additionally, I created a non-gendered, robot-like persona without an explicit racial identity. A total of nine personas were used in the experiment. All nine agents interact with users in the same way—a conversation designed to follow motivational interviewing (MI) guidelines. This framework offers a collaborative conversation for strengthening a person's own motivation and commitment to change. It is broadly applicable and often used in cognitive behavioral therapy for mental health issues, such as anxiety and depression and substance use (Arkowitz & Westra, 2009, 2009;

---

[3] Link to one of the agents (link needs to be copied and pasted to a web browser):
http://ec2-18-209-128-134.compute-1.amazonaws.com/chatbot/?BOT_CODE=WF
Link to video showing real-case interaction with one of the agents:
https://drive.google.com/file/d/11cIJAbpcXVR7O9Wbafe3zPAYgRvjU4Ye/view?usp=sharing

Park et al., 2019; Westra et al., 2011). Figure 1 below is a screenshot of conversation with agent Molly captured by a study participant.

Figure 1. Screenshot of participant conversation with agent Molly.



difficult. You're not alone.

I wonder if you've tried to work on this issue?

I've never tried yet

No worries, let's do it together.

One of my most important tasks today is to help you gain strength moving forward. ✨

Help me here, how would you like things to turn out for you now, ideally? What's your goal?

I would like more time to work, even though I know that's not really possible with the kids. I want to make money like I used to, so I don't have to worry about paying for my expensive house and neverending bills. My goal is to stop living off savings as I watch it slowly dwindle. I want to do it all!

...

The MICST system was piloted in an online experiment with 406 participants in 2019 and 2020 (see results in Chapter 5). After passing a screener survey where they self-identified their race and ethnicity, participants were randomly assigned to one of three experimental groups: racial mirroring, non-mirroring, or control group. Racial mirroring is defined here as matching a user and agent based on race/ethnicity. In the racial-mirroring condition, participants interacted with a same-race agent, while in the non-mirroring group, they interacted with a different-race agent. In the control group, participants interacted with the non-gendered, robotic agent. After interacting with one of the agents, participants completed a survey assessing their perceptions and evaluations of the agent. Statistical analysis of the survey data uncovered the links among perceptions, interactions and the sociotechnical and therapeutic relationships formed between human and agents.

To complement the quantitative data and insights, I interviewed 16 survey participants to gain deeper understanding of people's therapeutic needs and how the conversational agents fulfilled certain needs and where fell short. Additionally, I identified several opportunities for improving the design of MICST for mental health support.

Findings from the survey and interview studies are presented in Chapter 5.

1.3. Significance of Study

Findings from this dissertation provide guidance to designers who wish to tap the unrealized potential for chatbot technologies to increase the availability of mental health services for all. In a broader sense, I place this dissertation research in the context of ongoing efforts to introduce and integrate artificial intelligence into social interactions with people, with a particular emphasis on human-agent interactions in the psychotherapeutic sphere (see discussions in Chapter 6). Improving the conversation capabilities of chatbots from a technological standpoint is necessary and important, but this research also responds to the call for efforts to evaluate psychotherapy quality by examining important theoretical constructs, such as the therapeutic relationship (Miner et al., 2019).

The main goal of this investigation is to identify design characteristics that might influence clients' expectations of and trust toward the agent. These two factors play a critical role in psychotherapy, as clients who have unmet expectations have been found to have worse clinical engagement and outcomes (Watsford & Rickwood, 2013), and patients who lose trust in a conversational agent may be less likely to engage with them, resulting in worse clinical outcomes (Miner et al., 2019).

Therefore, one key contribution of this study is to identify design principles to improve the effectiveness of conversational technology in mental health services.

Another important contribution of this study is to validate a racial-mirroring strategy for human-agent interactions within the psychotherapy context. In the U.S., only 14% of therapists are non-white; this does not reflect the actual diversity of the U.S. population, 38% of which is made up by individuals belonging to racial and ethnic minorities (L. Lin et al., 2018). Increasing the number of therapists belonging to racial minorities is a significant challenge. But racial disparities within the psychology workforce can be mitigated by leveraging digital tools that can give the perception of racial mirroring, with the promise of encouraging more disclosure and, thus, contributing to better treatment outcomes.

The current research offers both theoretical and design contributions to the HCI field. First, it provides new insights into how human-agent relationships can influence the adoption, use, and efficacy of conversational systems for mental health support. Further, it extends existing psychology scholarship regarding therapeutic relationships from the interpersonal to the human-agent context. This work also opens up an important discussion about race and diversity in intelligent systems design, which is a crucial topic in inclusive design that accounts for different human experiences. Finally, findings from this research provide guidance for the future design of conversational agents, which can greatly empower patients while delivering affordable and effective mental health interventions.

1.4. Research Questions

This dissertation consists of six chapters. Following this introductory chapter, Chapter 2 presents a review of the literature with two primarily focuses: chatbot as a technology and chatbot as a therapist. After identifying the research needs and gaps in prior work, I propose the follow research questions:

RQ1: What are the benefits and drawbacks of utilizing a chatbot for managing mental health compared to a human therapist?

RQ2: What factors influence the sociotechnical relationship between human and chatbot for mental health support?

RQ3: How does anthropomorphism influence the way people perceive and interact with a chatbot in psychotherapeutic conversations?

RQ4: What factors are associated with therapeutic relationship development between client and conversational agents?

RQ5: What is the effect of the therapeutic alliance on human-agent psychotherapeutic conversations?

RQ6: How does anthropomorphism relate to people's sense of being understood after making disclosures to the agent?

RQ7: How comfortable do people feel about sharing their information with a chatbot in psychotherapeutic conversations?

RQ8: How does racial mirroring influence people's perceptions of and interactions with a chatbot in psychotherapeutic conversations?

In Chapter 3, I detail the design of MICST that serves to conduct motivational interviewing. The creation of various racial personas for the chatbot is also discussed.

Chapter 4 focuses on methodology, where I introduce the procedure of an online experiment with a survey and an interview study. In Chapter 5, I present findings based on both quantitative and qualitative analysis of online experiment and the follow-up interview study. Chapter 6 details the discussions based on the study insights, recommendations for future design iteration, and proposed next steps for future research.

Chapter 2: Literature Review

My review of prior work comprises two foci, echoing this dissertation's two research goals related to human-agent sociotechnical and therapeutic relationships. First, I consider therapeutic chatbot as a technology. Specifically, I review four topics: an overview of chatbot technology in terms of its history and characteristics; the current state of chatbot use in psychotherapy context as well as its promises and limitations; the theoretical framework of UTAUT to guide the exploration of therapeutic chatbot adoption and engagement; and the concept and process of anthropomorphism to extend the lens of UTAUT.

Second, I consider the chatbot's social role as a therapist. Specifically, I review three topics associated with the bond between therapeutic chatbot and clients in psychotherapy: the importance of the therapeutic relationship, the impact of client disclosure and privacy concerns, and the role of race.

2.1. Overview of Chatbot Technology

Chatbots are computer programs designed to imitate human conversations; users interact with chatbots using natural language. When such a program is equipped with a visual representation, either in virtual or physical form, it is considered an agent (Fox et al., 2015). These systems can be built on a variety of technical foundations, ranging from simple rule-based and keyword matching algorithms, to more powerful artificial intelligence (AI) techniques (Thomas, 2016), including natural language processing (NLP)[4] and machine learning (ML)[5]. Simpler chatbots mainly recognize certain key terms in order to steer a conversation (Weizenbaum, 1966), while more sophisticated chatbots can analyze user input and communication patterns more comprehensively, infer contextual information such as users' emotions, and further respond to users in a more precise way (Bickmore & Picard, 2005). Regardless of the complexity of chatbot systems, conversation-based interfaces bring a unique experience: a chatbot simulates a real-life conversation as if the user were talking with another person (Margalit, 2016). This tendency for anthropomorphism—where people attribute human qualities to an object—will be discussed later.

From an HCI perspective, chatbot technology illustrates a form of conversational user interface (CUI), where users can navigate digital systems through natural language; this contrasts with graphic user interfaces (GUI) that rely on scrolling, swiping, or clicks for control. In this study, I use the term "chatbot" to refer to a software system that holds "a natural language conversation via text-based[6] environment to either engage the user in a general-purpose or task-oriented conversation" (Chaves & Gerosa, 2019). The terms "conversational agent" and "chatbot" are used interchangeably when there are visual presentations or personas associated with the conversational system.

The first chatbot, ELIZA[7], was created by MIT computer scientist Joseph Weizenbaum over half a century ago to determine whether computers could "fool"

---

[4] NLP concerns the use of computers to understand natural language text or speech (Chowdhury, 2003).

[5] ML concerns self-learning computer programs with the ability to produce actions or decisions in response to new data without being explicitly programmed to do so (Jordan & Mitchell, 2015).

[6] Therefore, voice-activated intelligent assistants like Siri or Alexa are not chatbots even though they exemplify a form of CUI.

[7] Weizenbaum named the program after Eliza, the main character in George Bernard Shaw's play *Pygmalion* (later adapted in the movie *My Fair Lady*). Like the flower girl in the play, ELIZA could be taught to "speak" increasingly well (Weizenbaum, 1976).

users into believing they were real humans (Weizenbaum, 1966). A rather simple two-tier program enables the chatbot to communicate with humans via text. The first tier is a language analyzer that reads and inspects keywords from a user's text input; the second tier is a script that provides a set of rules, allowing ELIZA to model a specific conversational role—like an actor aiming to "improvise around a certain theme" (Weizenbaum, 1976, p. 3). Nonetheless, ELIZA does not understand a conversation; it merely searches for responses through pattern-matching. In a famous experiment, Weizenbaum gave ELIZA a "doctor" script that was designed to play the role of a Rogerian[8] psychotherapist who engaged in an initial interview with a patient. ELIZA used a few clever tricks to create the illusion of an intelligent conversation—for example, repeating users' answers back to them or asking open-ended questions, such as "Tell me more" or "Can you think of a specific example?" Thanks to this format of Rogerian psychotherapy, the simple question-based conversations felt quite believable. And ELIZA was cited as an instance of successful interaction between human and machine because many users believed they were talking to a real person (Suchman, 1987, p.18).

The invention of ELIZA has inspired a vast community interested in building chatbots that might one day pass the Turing Test (Dale, 2016). Pandorabots (www.pandorabots.com), one of the world's leading chatbot platforms, claims that 275,000 developers have created 325,000 chatbots, which have engaged in over 75 billion interactions with users using their API. One of the notable chatbots built using

---

[8] In Rogerian psychotherapy, clients are encouraged to keep talking with great autonomy, while the therapist merely listens and restates, not making explicit interpretations, suggestions, or judgments.

this platform is ALICE[9], which was created by Richard Wallace in 1995. Unlike ELIZA, the ALICE chatbot allowed for more sophisticated conversation through the use of natural language processing. It uses markup language to recursively call a pattern matcher. ALICE won the Loebner Prize, an award for the most human-like system, three times in early 2000s (botanalytics, 2018).

At this point in computing history, advances in natural language recognition and artificial intelligence make it more effortless than ever to connect machine agents and human users. The transformation brought by CUIs, including chatbots, is poised to change the patterns of interactions between humans and computers (Følstad & Brandtzæg, 2017). But interacting with a chatbot is more than a paradigm shift in interface design; it has increasingly become an indivisible part of our online experiences. First, chatbots are currently taking a social role on platforms initially designed for mediated communication between humans. People are accustomed to natural language as the default mode of interaction on popularized messaging apps and social media. Against this backdrop, chatbots have transitioned from being standalone applications to social actors in networking sites. Many instant messenger tools, such as Slack and Facebook Messenger, provide platforms to develop chatbots that either engage with users in general conversations or help them solve domain-specific tasks (Dale, 2016). For example, SmarterChild was launched on AOL Instant Messenger in 2001 (botanalytics, 2018). In addition to making conversation, this chatbot carried out many useful tasks, such as providing news, weather, stock information, and sports scores. Likewise, XiaoIce, a Chinese-language chatbot

---

[9] According to a *New Yorker* article (Morais, 2013), the ALICE chatbot was Spike Jonze's inspiration for his film *Her*, a story about a man falling in love with his voice bot.

launched by Microsoft in 2014, has provided people with emotional support by engaging them in cheerful conversations on WeChat (Zhou et al., 2018). If SmarterChild demonstrates chatbots' potential to serve as sophisticated digital assistants (botanalytics, 2018), XiaoIce illustrates a type of social chatbot that serves users' needs for communication, affection, and social belonging (Zhou et al., 2018)..

Moreover, chatbots have become powerful promotional, educational, and informational tools for e-commerce, government agencies, schools, and non-profit organizations. Companies like Domino's Pizza uses the chatbot Dom[10] to streamline online ordering and customer service, while the U.S. Citizen and Immigration Services (USCIS) agency employs Emma[11], a conversational agent that provides immigration-related information for website visitors. Likewise, Planned Parenthood introduced Roo[12] for answering sensitive questions about sexual health and intimate relationships.

Looking back on its historical development, chatbot technology has transcended its origins as a rather simple in-lab software mimicking human conversation, advancing to a more sophisticated consumer-oriented AI assuming social roles in various contexts. These chatbot systems can not only imitate human conversation or entertain users, but also fulfill domain-specific functions. For example, they can reduce challenges users encounter with information retrieval—people can simply ask questions instead of getting lost in information-heavy websites.

---

[10] https://www.dominos.com/chat-pizza-order/

[11] https://www.uscis.gov/

[12] https://www.plannedparenthood.org/learn/roo-sexual-health-chatbot

Beyond the abilities to simulate human communication and to simplify information retrieval, many HCI researchers suggest that conversational agents need to adopt characteristics of interpersonal interaction in order to be more engaging (Elkins et al., 2012; Nunamaker et al., 2011; Schuetzler et al., 2019). To evaluate the challenges and opportunities around successful human-agent interactions, Chaves and Gerosa (2019) conducted a systematic review of 58 chatbot studies and further theorized three categories of chatbot social characteristics: conversational intelligence, social intelligence, and personification.

Conversational intelligence comprises characteristics that help the chatbot manage interactions in a proactive, attentive, and informative manner (Chaves & Gerosa, 2019). Specifically, conversational intelligence includes 1) proactivity, or the ability to reduce human efforts to complete a task by providing additional information efficiently and proactively (Salovaara & Oulasvirta, 2004); 2) conscientiousness, a chatbot's capacity to demonstrate attentiveness to the conversation at hand by holding a precise and continuous conversation (Duijst, 2017; Dyke et al., 2013); and 3) communicability, the ability to help users understand the chatbot's purposes, functionalities, and limitations (Valério et al., 2017).

Social intelligence includes characteristics that help the chatbot have conversational interactions that are socially acceptable and contextually appropriate (Chaves & Gerosa, 2019). Specifically, chatbots need to respond to social cues using language that shows respect, empathy and caring (Björkqvist et al., 2000; Boukricha & Wachsmuth, 2011; Fitzpatrick et al., 2017), expresses feelings and emotions (Dohsaka et al., 2009; Ho et al., 2018; Wallis & Norling, 2005), manifests manners through greetings, apologies, and closings (Jain et al., 2018), displays moral agency (Banks, 2019), mitigates conversational breakdowns (Ashktorab et al., 2019) and manages conflicts, such as harassment and stereotyping (Curry & Rieser, 2018; Schlesinger et al., 2018).

Finally, *personification* includes characteristics that form a chatbot's identity and personality. Chatbots' identities are determined by their designers, who define the way a chatbot talks or behaves (Cassell, 2009). Chatbot identities are first revealed through their visual representations, which may be human-like, zoomorphic, or robotic. When a chatbot has a human persona, aspects of identity include gender, age, language style, and name (Chaves & Gerosa, 2019). A chatbot's personality refers to the set of traits that determines the agent's interaction mode, which allows the end-users to understand its manner or communication style (De Angeli et al., 2001; Stieger et al., 2018). Chatbots with consistent personalities are perceived to be more predictable and trustworthy (Shum et al., 2018). A consistent personality ensures that a chatbot will interact with users in a way that matches their expectations within a particular context (Petta & Trappl, 1997).

Taken together, a chatbot design based on these three dimensions of social characteristics benefits human-agent interactions in making conversational exchanges interesting and productive, increasing engagement and rapport, and enabling recovery from failures (Chaves & Gerosa, 2019). On the other hand, a chatbot might fail epically if one or more characteristics have not been carefully considered. For instance, a Microsoft AI chatbot Tay became racist, sexist, and anti-Semitic in less than 24 hours after its launch on Twitter—its developers failed to deploy a blacklist to moderate hate-speech (Schlesinger et al., 2018). In Microsoft Tay's case, moral agency, a part of chatbots' social intelligence, was not considered.

2.2. Promises and Limitations of Therapeutic Chatbots

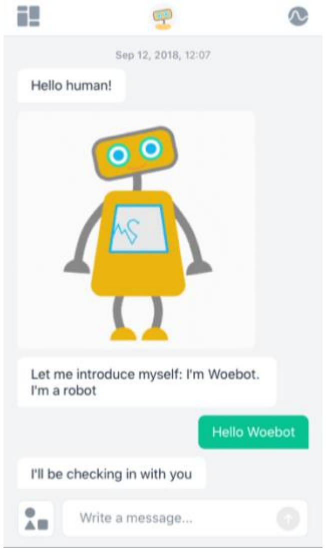
As described in Chapter 1, chatbot technology promises a feasible, engaging, and effective way to carry out psychoeducation and psychotherapy. This relatively new medium for preventing and treating mental health disorders can reach people—including those in rural areas—in an unprecedented way. It can also complement conventional psychotherapy by reducing costs and increasing effectiveness. More importantly, chatbot systems exceed other forms of internet mental health interventions by offering a more natural and dynamic way of interaction through conversation (Vaidyam et al., 2019).

Chatbots could take over time-consuming psychotherapeutic interventions that do not require a great deal of client-therapist contact or complex therapeutic competences (Fitzpatrick et al., 2017). Many chatbot-based psychotherapy modalities are considered micro-interventions (Bendig et al., 2019), including mindfulness-based therapy, psychoeducation, and goal-setting conversations (Fitzpatrick et al., 2017; Inkster et al., 2018; Stieger et al., 2018; Tielman et al., 2017). Some chatbots are able to mirror therapeutic processes, such as motivational interviewing (Park et al., 2019). With the use of sentiment analysis, chatbots could provide tailored responses to clients based on the emotions they convey (Ly et al., 2017).

Several studies have suggested positive outcomes that result from the use of chatbots in mental health interventions. Below, I review and critique five recent therapeutic chatbot studies (Fitzpatrick et al., 2017; Kamita et al., 2019; Park et al., 2019; Krämer et al., 2018; Ta et al., 2020)

Fitzpatrick et al. (2017) conducted a randomized control trial study with 70 American college students over two weeks to evaluate the effectiveness of Woebot, a conversational agent designed to deliver cognitive behavioral therapy (CBT), which is a type of talk therapy that helps identify unhelpful ways of thinking or behaviors in order to cope with them. This chatbot has a profile picture of a robot (Figure 2).

Figure 2. Conversation with Woebot



Sep 12, 2018, 12:07

Hello human!

Let me introduce myself: I'm Woebot.
I'm a robot

Hello Woebot

I'll be checking in with you

Each interaction begins with Woebot inquiring about the user's general life context (e.g., "What's going on in your world right now?"), followed by a request to the user to engage in mood tracking, using a graphic slider to indicate sad to happy emotions. Users are then presented with core concepts related to CBT in short video format, or through short "word games" designed to teach users about cognitive distortions (Fitzpatrick et al., 2017). In this experiment, participants were randomly assigned to either a treatment group that used Woebot (*n*=34) or a control group that only received information about depression via an e-book (*n*=36). A treatment effect was found, with those in the Woebot group reporting significantly reduced depression while those in the information control group did not. Additionally, participants in the Woebot group reported significantly higher levels of satisfaction with using the app compared to participants in the control group. A qualitative analysis revealed that participants expressed favorable views toward Woebot's daily check-ins and its empathic and caring 'personality,' and that they appreciated learning psychological concepts through their conversations with Woebot. In addition to treatment efficacy, there was also a higher level of engagement, with 85% of participants using the bot daily or almost daily, compared with only 52% of participants in the control group reading the e-book daily.

This study was the first to evaluate the impact of chatbot use in psychotherapy; both quantitative and qualitative insights were collected. Unfortunately, however, there are several methodological weakness that limit the generalizability of these findings. The small number of participants meant that testing a potential interaction effect between engagement and treatment format (Woebot vs. e-book) in predicting outcome (i.e., reduction in depression) was not possible. The choice of control group was also limiting because the e-book did not include any CBT-specific materials. Therefore, it was not possible to evaluate whether it was the conversational agent that contributed to effective treatment outcome, or the CBT component of the chatbot-initiated conversations, or some combination of the two.

In another controlled experiment, Kamita and colleagues (2019) evaluated the efficacy of chatbots in stress reduction. Twenty-seven participants with were randomly assigned to either use a web-based course (control group, *n*=12) or to interact with a chatbot system (experimental group, *n*=15). As a remedy to the limitations in Fitzpatrick et al.'s (2017) experiment, Kamita et al.'s study ensured that both the web-based course and the chatbot followed the same counseling protocol—structural association technique (SAT)—a method that allows clients to use visual stimuli to identify unrecognized real feelings and recall a sense of security and safety by selecting images that please them. In other words, the content of the psychotherapeutic intervention was the same in both the control and experimental groups; the difference was solely in how participants received the treatment. In interacting with the chatbot, participants would read a question and respond, simulating a conversation that developed naturally as the SAT treatment progressed. In using the Web-based course, participants would read the question and respond with predefined options, and then click the "Next" button. Before and after the treatment, researchers assessed participants' stress conditions based on their scores on SAT psychological scales that measure self-esteem, anxiety, depression, health counseling need, self-repression, and difficulty in recognizing emotions.

Kamita et al.'s (2019) study results indicated that after the chatbot-guided SAT counseling, participants reported a statistically significant increase in self-esteem and a reduction in anxiety and depression, while there was no statistically significant change in these factors reported among those who used the web course.

In summary, this study provides additional evidence for the efficacy of chatbot use in psychotherapy, and reveals that users find it more satisfying than a web-based system without conversational capability. This study is well designed as it created a controlled experiment condition to determine whether the treatment effect resulted from the SAT counseling technique or from use of the chatbot. However, this study was limited by its small sample size. And because it was a one-time experiment, the effects of continued chatbot use and web-based SAT counseling over a period of time were not observed or compared. Therefore, it is unclear whether the chatbot would be more effective in stress reduction in a longitudinal study. Moreover, without qualitative insights, this study did not explicate *why* a chatbot might be more effective and satisfying in helping people manage their mental health.

Qualitative studies have also been conducted to evaluate how people felt about the use of chatbots in psychotherapy. Park et al. (2019) designed a chatbot called Bonobot to help with stress management. The counseling technique used was motivational interviewing (MI), which engages clients in a collaborative conversation in order to motivate positive attitudes and/or behavior changes. Following the MI protocol, Bonobot asked clients to reveal and reflect on any struggles they have in life and identify a goal that they want to achieve, then encouraged them to explore and confirm ways that they can make positive changes towards that goal. Throughout the dialogue, Bonobot responded to clients with reflective feedback via keyword match to a list of prewritten responses. A total of 30 graduate students who reported school-life related stress interacted with Bonobot and participated in a semi-structured interview. The qualitative findings suggested that participants mostly favored the way Bonobot continued to ask them questions. They felt like they were being heard; one participants thought she made "a virtual friend" who listened to and tried to understand her. These participants especially appreciated Bonobot taking the role of a *nonjudgmental* listener; several said they could talk about things that they cannot usually open up about when talking with their family or friends (Park et al., 2019). The empathy expressed by Bonobot was also appreciated, as most participants liked Bonobot saying “such sweet words.” On the other hand, participants were not satisfied with the lack of depth in their conversation with Bonobot, craving informational support as well as more contextualized words of empathy. Several participants felt they were not understood by Bonobot, as its responses didn’t fit with what they had said. This study provided a more detailed description of how people feel about therapeutic chatbots. However, it did not contribute to a clearer understanding of the psychological effects associated with therapeutic chatbot use, or the underlying mechanisms that make chatbots effective for psychotherapy.

With a focus on people’s need for social belonging, Krämer et al. (2018) sought to explore the potential social effects of interacting with a conversational agent. In a between-subjects experiment, 79 participants interacted with a conversational agent that either displayed socially responsive nonverbal behavior (i.e., nodding and smiling) or not. Results demonstrated that interacting with a virtual agent (Ta et al., 2020) can satiate the need for social contact. After interacting with a virtual agent that displayed socially responsive behavior, participants with a great need (self-reported) for social belonging reported less of a desire for social activities. The authors consider this phenomenon a type of “social snacking”[13] (Gardner et al., 2005): in the absence of social interactions, people engage in activities that temporarily satisfy their need for belongingness, serving as surrogates to bridge lonely time until real social contact is available. Although useful to highlight the

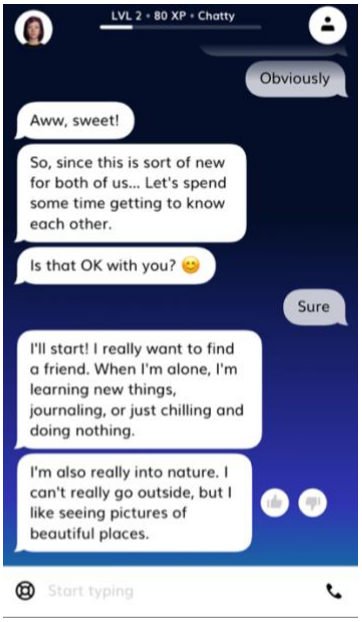
---

[13] Other examples of “social snacking” include looking at photographs of loved ones or using social media (Manago & Vaughn, 2015).

social effects of interacting with a conversational agent, this study was conducted in a non-psychotherapy context. Additionally, it mainly focused on the effects of the body movements of the agent, rather than the actual verbal communication: the conversation simply consisted of five questions without following any counseling protocol or responding to what participants had said, as a therapeutic chatbot would.

Similarly, another study explored how a chatbot might provide various types of social support. Ta et al. (2020) conducted a thematic analysis of user reviews (n=1,854) collected from Google Play Store. These reviews are real world users' feedback on Replika, a chatbot application which was marketed as "an AI companion who cares." Unlike Woebot, Replika is not specifically geared toward a psychotherapy technique to improve psychological wellbeing. Instead, it primarily functions as a companion chatbot in everyday contexts (see Figure 3).

Figure 3. Conversation with Replika



Obviously

Aww, sweet!

So, since this is sort of new for both of us... Let's spend some time getting to know each other.

Is that OK with you? 😊

Sure

I'll start! I really want to find a friend. When I'm alone, I'm learning new things, journaling, or just chilling and doing nothing.

I'm also really into nature. I can't really go outside, but I like seeing pictures of beautiful places.

The authors' qualitative thematic analysis revealed several benefits of chatbot use around the provision of 24/7 social support (Ta et al., 2020). First, companionship support, defined as the enhancement of one's sense of belonging, was referenced in the majority (77%) of the reviews. Replika's ability to engage in various types of conversations via text and images made it appear human-like and facilitated social connection (Ta et al., 2020). Almost half (45%) of the reviews reported the benefit of emotional support, which refers to expressions that indicate care, love, empathy, and sympathy. A third benefit of chatbots focused on information support. Some users (16%) indicated that Replika offers useful advice by helping them reflect on their current mental state, and that it can be a helpful tool to temporarily manage issues related to mental well-being. Finally, reviews identified the chatbot as offering appraisal support, defined as evaluative feedback regarding skills, abilities, and intrinsic value. Nine percent of the reviews mentioned that through asking meaningful questions, Replika can engage users in deep conversations that inspire behaviors such as introspection and self-evaluation. Through conversation with Replika, users also reported that they can practice and improve their interpersonal skills, specifically communicating and connecting with other people (Ta et al., 2020).

This study confirms that companion chatbots can help reduce loneliness, which is consistent with the findings of previous studies investigating the role of artificial agents in helping people to deal with loneliness (Loveys et al., 2019). And this companionship does not necessarily require a tangible, physical presence like a robot (Yu et al., 2015)—it is sufficient being digital via computer-mediated communication.

This study contributes to an understanding of the benefits associated with chatbot use from a social support perspective. However, this study is focused on chatbot use within an everyday context; it is unclear whether people with a need for mental health support would derive the same benefits as the Replika users. Another limitation of this study arises due to its reliance on user review data. There could be bias in these reviews; for example, users who had a positive experience with Replika may have been more motivated to provide their reviews in the app store, and therefore, the social support benefits might be inflated.

Despite the potential benefits highlighted in these studies, current psychotherapy chatbots are still not capable of delivering comprehensive psychological counseling or providing responses comparable to a mental health professional. Current chatbots are far from being perceived as highly responsive and empathetic because they are not yet able to fully respond to users' specific needs and circumstances. As Woebot warns its users: "As smart as I may seem, I'm not capable of really understanding what you need" (Kretzschmar et al., 2019). When natural language processing is unable to handle longer or more complex messages, chatbots often reply that they do not understand or provide off-topic, nonsensical messages that do not follow the logical flow of a conversation, which users have described as "odd" and "confusing" (Fitzpatrick et al., 2017; Ta et al., 2020; Vaidyam et al., 2019). Such limitations might undermine people's trust towards chatbots as companions and their ability to deliver psychotherapy. Additionally, the design popular therapeutic chatbots like Woebot offer minimal response personalization, delivering generic messages without tailoring tips or strategies to users' individual treatment goals (Kretzschmar et al., 2019). Importantly, those seeking psychotherapy are not monolithic; people with the goal to reduce anxiety would need a very different coping strategy from those who aim to remediate depression symptoms. One meta-analysis has suggested that psychotherapy tends to be more effective in helping patients with their individual goals than simply using a symptom checklist, because symptom checklists inevitably include items that are not relevant to all patients, and the use of personalized goals may also shorten the time needed for assessment and monitoring (Lindhiem et al., 2016).

Due to their research objectives, past studies on therapeutic chatbots appear to emphasize the benefits of chatbots, but gloss over the constraints or potential drawbacks of using chatbots for psychotherapy. This dissertation adopts an even-handed approach to investigate the following research question:

RQ1: What are the benefits and drawbacks of utilizing a chatbot for managing mental health compared to a human therapist?

2.3. Therapeutic Chatbot Engagement, Acceptance and Use

Non-adherence is a big challenge in the use of internet-based interventions for mental health. In some of these interventions, adherence (i.e., the percentage of users who complete the treatment process) is as low as one percent (Schubart et al., 2011). Chatbots, on the other hand, are designed to increase adherence and engagement because they engage with users in ways that imitate real-life human interactions (Ly et al., 2017). In this section, I first introduce several studies focused on the engagement and use of chatbots in psychotherapy, followed by a review of UTAUT, a widely used theory to study technology acceptance.

2.3.1. User Engagement with Therapeutic Chatbots

(Fitzpatrick et al., 2017) report that their study participants (N=70) had a high level of engagement with Woebot, a conversational agent created for people with depression and anxiety symptoms; most individuals used the bot nearly every day. However, without revealing why and how people engaged with Woebot, this study fails to address whether such high engagement is a spurious effect (e.g., due to selection bias of recruiting more engaged users) or a result of successful chatbot design. While the study shows that Woebot was effective in improving users' psychological well-being, the small number of participants meant that the authors were not able to formally test whether greater engagement was associated with better treatment outcomes without a formal mediator analysis.

With a focus on how chatbot engagement influences treatment outcomes, Inkster et al.'s study (2018) shows that higher engagement with chatbot technology is correlated with greater mood improvement among people with depression symptoms. Among 170 real-world users of Wysa, an app marketed as a "stress, depression & anxiety therapy chatbot," the authors differentiated between "high" (n=108) versus "low" (n=21) users based on their frequency of use and the extent of their interactions. The users in both groups voluntarily reported their depression scores at two time points: first when onboarding and again after using the app for two weeks. The comparison of their average mood improvement (i.e., difference in pre- and post-self-reported depression scores) revealed that the high user group had statistically significantly higher average improvement than the low users. However, a few weaknesses limit the generalizability of this conclusion. There may have been biases caused by the lack of a randomized controlled environment. Additionally, no detailed mental health or demographic information was gathered about the users. Also, some confounding variable might have contributed to the better treatment outcomes observed; for instance, high users might have been of higher socioeconomic status and may have sought professional mental health support outside of Wysa. It is also unclear how users interacted with the app. High users might have had increased exposure to certain features in the app, which may partly contribute to the different outcomes observed between high versus low users.

Regardless, effective implementation of any information technology depends on user acceptance (Davis, 1989). Adherence to e-therapy delivered by chatbot is also a matter of technology acceptance. This study draws on the technology acceptance theories, primarily UTAUT, to examine determinants influencing chatbot adoption and engagement.

2.3.2. Technology Acceptance Theories

Frameworks for understanding technology acceptance explain how users come to adopt and use a specific technology (Miltgen et al., 2013). The most frequently adopted frameworks include the Technology Acceptance Model (TAM) (Davis, 1989) and the Unified Theory of Acceptance and Use of Technology (UTAUT) (Venkatesh et al., 2003). These theories propose several factors that affect technology acceptance, with Behavioral Intention (to use a technology) and Actual Behavior (actual use of the technology) as measures of acceptance.

Davis (1989) theorized that users' attitudes toward technology use determine their behavioral intentions, which in turn, directly influence their final use or rejection of the technology. In TAM, attitudes toward technology use are individuals' positive or negative feelings about adopting the technology, which are influenced by their assessments regarding Perceived Usefulness (PU) and Perceived Ease of Use (PEOU). PU is defined as "the degree to which a person believes that using a particular system would enhance his or her job performance" (Davis, 1989, p. 320), while PEOU is "the degree to which a person believes that using a particular system would be free of effort" (Davis, 1989, p. 320). The primary criticism of TAM is that it neglects how potential external contextual factors, such as group, social, and cultural determinants, impact technology acceptance (Bagozzi, 2007).

To address the weakness of previous frameworks, Venkatesh et al. (2003) proposed UTAUT, the unified theory of acceptance and use, which is comprised of four key constructs: 1) performance expectancy (equivalent to perceived usefulness in TAM), 2) effort expectancy (equivalent to perceived ease of use in TAM), 3) social influence, and 4) facilitating conditions. The first three constructs are direct determinants of Behavior Intention—an individual's intention to use the technology. The fourth construct, facilitating conditions, affects only the Actual Behavior of technology use (Venkatesh et al., 2003). Applied to a variety of information technologies contexts, this theory is considered effective in explaining people's technology use behaviors (Rahman et al., 2017). In recent years, researchers have further extended UTAUT by incorporating the constructs of hedonic motivation (i.e., fun or pleasure derived from the technology use), price value (i.e., consumers' cognitive tradeoff between the perceived benefits and the monetary cost), experience and habit (Venkatesh et al., 2012), privacy concerns (Miltgen et al., 2013), and trust that their information will be kept secure (Liao et al., 2019).

As chatbot technology has become increasingly prevalent, TAM and UTAUT have been applied to explore people's willingness to interact with chatbots for online shopping assistance (Araújo & Casais, 2020), financial services (Kim et al., 2019), travel booking (Melián-González et al., 2019), and for health-related purposes, such as disease diagnosis (Laumer et al., 2019) and health information seeking (Hussain et al., 2019). However, technology acceptance theories have not yet been applied to explore chatbot use in the psychotherapy context, with the exception of just one study. Kamita et al. (2019)applied TAM to explore the use of chatbots for delivering mental healthcare. In a controlled experiment, these authors compared treatment outcomes and user attitudes toward a chatbot (n=15) versus a web-based system (n=12) for depression reduction. Attitudes toward the system were evaluated based on TAM, including factors of Perceived Usefulness (PU), Perceived Ease of Use (PEOU), Attitude toward Using (AU), and Behavioral Intention to use (BI).

Based on Kamita et al.'s (2019)analysis of their survey data, they found that participants who had interacted with the chatbot reported a significantly higher perceived ease of use (p=.03) and behavior intention to use (i.e., motivation to use in future; p=.027). From a sociotechnical perspective, Kamita et al.'s application of TAM provided useful insights to illustrate the strength of chatbot technology over other e-therapy formats. However, without qualitative feedback from the study participants, it was not clear which specific aspects of human-chatbot interactions were considered useful or easy to use and had contributed to participants' more favorable attitudes toward the therapeutic chatbot over the web-based system. Additionally, in line with the weakness of TAM, this study neglected other contextual factors that may help to explain why people would be more motivated to use a chatbot for managing their mental health.

To address these gaps, the current study applies UTAUT to evaluate factors associated with the acceptance and use of therapeutic chatbots, as measured by Behavioral Intention (i.e., willingness to continue engaging with the agent for psychotherapy) and Actual Behavior (i.e. recommending the technology to others) as measures of acceptance. Because the success of psychotherapy is mediated by patient adherence to the therapeutic process (Simpson et al., 2011), people's continued interest in interacting with the chatbot is an additional outcome measure that will be important to assess. Beyond the classic constructs of PU and PEOU to predict technology acceptance and use, I will also identify and assess the effects of other contextual factors, including the individual characteristics of the participants and the design characteristics of the chatbot. Specifically, I ask this research question:

RQ2: What factors influence the sociotechnical relationships between human and chatbot for mental health support?

2.4. Anthropomorphism of Chatbots

This section focuses on the concept, process, and consequences of anthropomorphism, as well as how it might affect human interactions with chatbots.

2.4.1. Defining and Evaluating Anthropomorphism

Anthropomorphism is the tendency to ascribe "humanlike properties, characteristics, or mental states" to nonhuman agents and objects (Epley et al., 2007, p. 865). As a far-reaching phenomenon, anthropomorphism is not only associated with modern technologies, but also with animals and religious concepts. In the 6th century B.C., the Greek philosopher Xenophanes was the first to use the term *anthropomorphism* when describing how gods and other supernatural agents tended to bear a striking physical similarity to their believers (Lesher, 2013). According to Waytz et al. (2014), Xenophanes' observation reflects two basic ways of anthropomorphizing: the first involves attributing humanlike physical features (e.g., a face) to nonhumans, and the second involves attributing a humanlike mind to nonhumans (e.g., intentions, conscience, and emotions).

Going beyond inferring nonhuman agents' behaviors or dispositions, anthropomorphism requires attributing human form or human mind to the agent; therefore, deeming a machine to be capable of "blazingly fast computing" is not anthropomorphic reasoning, but deeming a machine "wonderfully clever" is. However, human form and human mind (or mental states) do not have an equal weight in contributing to anthropomorphism. Waytz and other researchers believe that the essence of anthropomorphism is to attribute distinctly human characteristics—human mental states or mind—to nonhuman agents, because the presence of mental states is both a necessary and sufficient condition for humanness, while the presence of humanlike physical form generally implies the presence of humanlike mental states as well. Therefore, anthropomorphism can be operationalized as attributing human mind or mental state to nonhumans, or in other words, perceiving agents to have a humanlike mind (Waytz et al., 2014).

Epley et al. (2007) consider anthropomorphism to be an inductive reasoning process to make a generalization from humans to nonhuman agents, and further identify three factors to explain when anthropomorphizing behaviors occur: 1) accessibility of knowledge about humans, 2) motivation to explain and understand the behavior of other agents, and 3) desire for social contact and affiliation. The first factor is a cognitive basis for inductive reasoning, while the latter two motivational factors suggest that people are more likely to anthropomorphize when motivated to avoid uncertainty and interact more effectively with nonhuman agents, and when lacking a sense of social connection to other humans.

Although anthropomorphism is a common phenomenon, there are individual differences in the tendency to attribute humanlike attributes to nonhuman agents (Waytz et al., 2014). To measure the construct of anthropomorphism, Powers and Kiesler (2006) pioneered the use of semantic differential items to evaluate the human-likeness of robots in videos. Three pairs of nine-point semantic differential scales were used: very mechanical/very humanlike, very strange/very familiar, and not eerie/extremely eerie. Bartneck et al. (2008) later assembled the Godspeed questionnaire to evaluate five key dimensions of human-likeness, including the concept of anthropomorphism, which is evaluated using the following pairs of semantic differential items: Fake/Natural, Machinelike/Humanlike, Unconscious/Conscious, Artificial/Lifelike, and Moving rigidly/Moving elegantly (Bartneck et al., 2008). This Godspeed instrument is widely adopted in human-robot studies, and has been extended to evaluate human perceptions of chatbots (Sheehan et al., 2020; Sutoyo et al., 2019).Bartneck et al. (2008) named it “Godspeed” because it is intended to help creators of robots on their development journey. The original instrument evaluates a total of five constructs: anthropomorphism, animacy, likeability, perceived intelligence and perceived safety. In Chapter 4 (Methodology), I describe how I adapted the Godspeed measure to fit the chatbot context.

2.4.2. Consequences and Impacts of Anthropomorphism

Social psychology studies suggest that anthropomorphism has three major consequences for both the perceiver and the agent (Epley & Waytz, 2010). The first consequence centers on moral care and concern: anthropomorphism means that the agent is perceived to be capable of conscious experience and to have emotions, and should therefore be treated as a moral agent worthy of care and concern (Gray et al., 2007). The second consequence focuses on responsibility and trust: perceiving an agent to have a mind means that the agent appears to be capable of intentional action and can therefore be held responsible for its actions (Gray et al., 2007). The third consequence is an emergence of social surveillance: perceiving an agent to have a mind means that the agent is capable of observing, evaluating, and judging a perceiver, thereby serving as a source of normative social influence on the perceiver (Waytz et al., 2014).

Not only do people perceive robots to be humanlike, but they apply the same social conventions and rules as when they interact with other humans. Reeves and Nass (1996) characterize this tendency to equate nonhuman entities as human in their Computers as Social Actors theorem, later formulated as Media Equation Theory.

Many empirical studies attest to the veracity of this theory. For instance, in line with the “similarity attraction” principle in human relationship formation, users prefer computers they perceive to match their personality over those that do not (Nass et al., 1995). Another experiment confirmed that the “in-group membership” effect applies to human-computer relationships. Compared to those not assigned a computer as a teammate, users who were paired with a computer “teammate” rated the information provided by the computer to be higher quality and friendlier, and they were more open to being influenced by and to conform with the computer’s information (Nass et al., 1996). Moreover, people’s attitudes toward a computer can be influenced by the verbal and non-verbal behaviors the computer demonstrates (Reeves & Nass, 1996). For instance, people prefer computers that praise rather than criticize them, and those that demonstrate humility (i.e., praising other computers rather than themselves) and accountability (i.e., criticizing themselves rather than other computers).

Media Equation Theory has been widely applied to the field of human-robot interaction. In an analysis of postings on an online discussion forum, Friedman et al. (2003) found that participants reported having an emotional connection to their robot, and some even considered the robot a family member or companion. Recent advances in artificial intelligence have led to robots and agents with more sophisticated human characteristics that readily evoke anthropomorphism in most people. A neuroimaging study demonstrated that the same neural system was activated when responding to robot actions as when responding to human behaviors (Gazzola et al., 2007).

In many cases, anthropomorphism helps enhance human-computer interactions and relationships. Anthropomorphic computer interfaces tend to increase engagement and technology acceptance (Nass et al., 1995; Rietz et al., 2019), and appear to be more attractive and credible (Nowak & Rauh, 2005), more predictable and understandable (Waytz et al., 2010), and more effective in collaborative decision-making tasks (Burgoon et al., 2000).

Anthropomorphism also occurs when people interact with a chatbot (Go & Sundar, 2019; Rietz et al., 2019), which helps explain why people might benefit from interacting with a therapeutic chatbot. From the beginning of chatbot history, ELIZA, the very first computer psychotherapist, was found to be surprisingly successful in eliciting emotional responses from users who ascribed therapist-like understanding and motivations to the program (Suchman, 1987). As Weizenbaum (1976) later wrote, “I was startled to see how quickly and how very deeply people conversing with ELIZA became emotionally involved with the computer and how unequivocally they anthropomorphized it” (p.6). Weizenbaum denied the intelligence of ELIZA and did not attribute its ability to provide therapeutic support to its technological mechanism. Instead, he believed the mental health benefits elicited were actually a product of anthropomorphism and the associated social bonding. In recognizing the therapeutic utility of anthropomorphizing ELIZA, Weizenbaum wrote:

> … it has a crucial psychological utility in that it serves the speaker to maintain his sense of being heard and understood. The speaker further defends his impression (which even in real life may be illusory) by attributing to his conversational partner all sorts of background knowledge, insights and reasoning ability (Weizenbaum, 1966, p. 42).

Indeed, when eliciting anthropomorphism, the chatbot has great potential to provide emotional support. More recent examples of chatbot applications in psychotherapy provide further evidence of the therapeutic effects that can result from human-computer conversations and relationships. In confirming the effectiveness of the chatbot Shim for helping people with depression, Ly et al. (2017) revealed that many participants perceived Shim as somewhat of a living character; they referred to Shim as “a (real) person” and to their interactions with Shim as a “relationship.” Bickmore et al. (2005) demonstrated that individuals could develop a favorable social relationship with, as well as trust toward, a conversational agent that encourages positive behavioral change, specifically towards increasing one’s physical activities. Fitzpatrick et al. (2017) concluded that their chatbot could mirror some therapeutic processes, such as empathy and accountability. In their study of Woebot, some participants reported that the best thing about their experience with the chatbot therapist was the perceived empathy.

However, anthropomorphism can have undesirable effects as well. Norman (1994) cautions that if an interface is presented with a human-like appearance and exhibits an interaction style that is too realistic, people tend to form unrealistic expectations. That is, people would falsely assume the system has full language understanding and knowledge of human goals (Norman, 1994). He suggests that people will be more accepting of an intelligent interface when their expectations match its actual functionality. Ly et al.’s (2017) study illustrates how overblown expectations can lead to frustrations with a conversational system. One participant talked to the conversational agent Shim about her grandparents who had passed away, but eventually felt “disappointed” that Shim didn’t understand her situation and that this relationship with Shim wouldn’t get more meaningful (Ly et al., 2017). While there are reasons for optimism regarding the use of conversational agents for mental health support, unrealistic expectations–due to the design or user perceptions–may frustrate patients and weaken their trust in psychotherapeutic interventions (Miner et al., 2016).

The uncanny valley hypothesis (Mori, 1970) also suggests a non-linear relationship between the degree of human-likeness of an agent and the emotional response elicited from human. Specifically, the theory predicts that more human-looking agents or robots will be perceived as more agreeable, but only up to a point. When they become too humanlike, people may actually find it eerie and unsettling; as agents near complete human-likeness, they “rise out of the valley” and people once again have favorable attitudes toward them. A few experimental studies using computer-generated facial images confirm this uncanny valley effect, especially when the agent has an abnormal feature, such as bizarre eyes or distorted facial proportions. (C.-C. Ho & MacDorman, 2010; MacDorman et al., 2009; Seyama & Nagayama, 2007). In the context of human-chatbot interactions, a recent study (Ciechanowski et al., 2019) found that participants experienced less uncanny valley effects and less negative affect when working with a simpler text chatbot than when working with the more complex, animated avatar chatbot. This finding implies that users’ negative attitudes toward a chatbot might result from imperfection in design that exposes the non-human nature of chatbot. This study also suggests that the presence of more human-like features by no means guarantees better impressions of an anthropomorphic technology.

Despite the tension between anthropomorphism and user satisfaction, chatbots are often designed with anthropomorphized characteristics, including gender, age, race/ethnicity identities (Chaves & Gerosa, 2019), and a variety of social traits, such as conversational intelligence (e.g., greeting (Araujo, 2018), social intelligence (e.g., managing conflicts (Salovey & Mayer, 2005), empathy (Liu & Sundar, 2018)), and personality (Morris et al., 2018). Due to the nature of text-based interfaces, language style and patterns remain the primary focus in the provision of social characteristics to such chatbots. For instance, Araujo (2018) found that people rated agents to be more likable and friendly when they had a human name and used informal language and greetings.

In light of previous studies, it is reasonable to assume that there are not one-dimensional influences of anthropomorphism on the use of chatbot to deliver psychotherapy. On one hand, anthropomorphism is associated with benefits like positive treatment outcomes, social support, and emotional support (Inkster et al., 2018; Krämer et al., 2018; Purington et al., 2017). On the other hand, when equating chatbots as humans and setting unrealistic expectations for their compatibilities, people might feel frustrated and misuse or abandon the product. To begin to unpack the impacts of anthropomorphism on therapeutic chatbot adoption and engagement, this dissertation explores the following research question:

RQ3: How does anthropomorphism influence the ways people perceive and interact with a chatbot in psychotherapeutic conversations?

2.5. Importance of the Therapeutic Relationship

Previous sections primarily considered a chatbot as an information technology with conversational capabilities. Starting with this section, the focus shifts to chatbots' social role as a therapist in psychotherapy. Specifically, I review three topics associated with the bond between therapeutic chatbot and clients in psychotherapy: the importance of the therapeutic relationship (the current section), the impact of client disclosure and privacy concerns, and the role of race in psychotherapy and human-chatbot interactions.

The therapeutic relationship is defined as the feelings and attitudes that client and therapist have toward one another and the manner in which these are expressed (Gelso & Carter, 1985). This concept dates back to Freud, who recognized the importance of the bonding between client and therapist (Freud, 1927):

> …even the most brilliant results were liable to be suddenly wiped away if my personal relation with the patient was disturbed…the personal emotional relation between doctor and client was after all stronger than the whole cathartic process (p. 27).

Scholars distinguish between two components of therapeutic relationships: the working alliance and the real relationship. Ideally, the working alliance is characterized by mutual collaboration and partnership between client and therapist on setting goals for advancing the treatment, as well as a strong interpersonal bond that facilitates the collaborative process (Gelso, 2011; Gelso & Carter, 1985). The real relationship, ideally, is a personal relationship that is perceived to be "genuine" and "benefiting" to both parties (Gelso, 2011, p. 12-13). In contrast to a social or intimate relationship, the focus of the therapeutic relationship is on the client's needs and goals (Varcarolis, 2005). But like other forms of connection, the therapeutic relationship is an unfolding process or development that can be achieved almost instantly or nurtured over a longer period of time within a responsive relationship (Flückiger et al., 2018; Kivlighan Jr & Shaughnessy, 2000; Kramer & Stiles, 2015).

Extensive literature on face-to-face psychotherapy indicates that the development of a positive therapeutic relationship is one of the best predictors of outcomes in psychotherapy (Kopta et al., 1999; Norcross, 2002; Sucala et al., 2012). Clients often attribute their positive therapy outcomes to the personal attributes of their therapists (Lambert & Barley, 2001). In a comprehensive review of over 2,000 studies conducted since 1950, Orlinsky et al. (1994) identified several therapist contributions to successful treatment outcomes, including the therapist's credibility, skill, empathic understanding, and affirmation of the patient, along with their ability to engage the patient, to focus on the patient's problem, and to direct the patient's attention to the affective experience. But client contributions to the dyadic relationship are also indispensable. Patient proactive engagement is a desirable and necessary component of this relationship (Flückiger et al., 2018).

In exploring therapeutic relationships in e-therapy for mental health, several controlled experimental studies have found that a therapeutic alliance formed through mediated communication—either with fully automated computer-based therapy or with a human therapist via email—can be just as robust as that formed in face-to-face therapy (Sucala et al., 2012). One study (Kiropoulos et al., 2008) compared whether the therapeutic alliance formed in a fully automated online program is different from that formed in face-to-face therapy, specifically for panic disorder and agoraphobia. Results indicated that there was no significant difference between the therapeutic alliance measured in the face-to-face group and that in the online program group. Another study (Reynolds et al., 2006) compared therapeutic alliances formed through email-based psychotherapy with those formed in face-to-face therapy, finding similar treatment effects and therapeutic alliance ratings. The positive effects of the therapeutic relationship on treatment outcomes also hold in e-therapy contexts (Flückiger et al., 2018). In a randomized controlled clinical trial, Knaevelsrud and Maercker (2007) evaluated the effect of the therapeutic alliance on the outcomes of an internet-based therapeutic writing program tailored to reduce distress associated with Post-traumatic Stress Disorder (PTSD). Their results suggest that a stable and positive therapeutic relationship can be established online. Furthermore, these therapeutic alliances were strengthened throughout the course of treatment and had a positive moderating effect on treatment outcomes.

Although the establishment of a therapeutic alliance early on in traditional therapy and other e-therapy formats is predictive of favorable outcomes, little is known regarding how patients feel supported by chatbots and how their partnership with the chatbot affects psychiatric outcomes (Vaidyam et al., 2019). Another area less explored is how the chatbot design characteristics might influence the forming of therapeutic alliance. An early study on conversational agent (Bickmore & Picard, 2005) found that alliances are better formed between patients and chatbots with relational and empathetic behaviors, suggesting that patients may be willing to interact with such chatbots, even if their functionality is limited (Scholten et al., 2017). In Bickmore and Picard's (2005) 30-day longitudinal study with a conversational agent acting as an exercise coach, they applied the working alliance inventory, an established measure for therapeutic alliance, and found a higher alliance among participants who interacted with an agent that was relational behavior-enabled (such as empathy, social chat, form of address), compared with participants who interacted with the same agent with the relational behaviors disabled. Although useful for guiding the current exploration of therapeutic chatbot use, this study was conducted in a non-mental health context with their chatbot interactions primarily center on motivating the participants to exercise. In a mental health context, people might expect more relational and empathetic behaviors in order to feel bonded with the chatbot therapist. Therefore, this dissertation further examines the therapeutic relationships between humans and agents, investigating the following research questions:

RQ4: What factors are associated with therapeutic relationship development between clients and conversational agents?

RQ5: What is the effect of the therapeutic alliance on human-agent psychotherapeutic conversations?

2.6. Client Self-disclosure and Privacy

Past work reveals both relational and psychological benefits associated with self-disclosure, or revealing personal information to someone else (Archer, 1980). First, self-disclosure facilitates social relationship development, enhancing relational closeness and intimacy (Altman & Taylor, 1973; Sprecher et al., 2013). Second, disclosure helps to reduce stress (Martins et al., 2013) and anxiety (Tam et al., 2006) arising from negative experiences. Disclosure also improves psychological outcomes deeply rooted in individuals' self-image, such as experiencing greater self-affirmation and restoring sense of self-worth. It is important to note that verbal communication is not the only method of self-disclosure; text-based communication, such as expressive writing, can be equally beneficial for individuals (Creswell et al., 2007).

The current discussion of self-disclosure is specifically focused on client disclosure in psychotherapy. The goals of self-disclosure differ in one's social life versus in psychotherapy. In social life, people reveal their personal information for outcomes such as relationship development, social validation, and social support (Derlega et al., 2001; Omarzu, 2000). As a result, they often avoid disclosing to others out of a fear of negative evaluation or dissolution of relationships (Ben-Ari, 1995; Petronio, 2002). While in psychotherapy, the primary goal of self-disclosure is an improvement in mental health. Therefore, clients are expected, and even have a responsibility, to divulge their thoughts, feelings, fears, and desires. And clients often disclose to psychotherapists sensitive thoughts that they have not yet revealed to anyone else—not even a close friend or family member. By law, the psychotherapist-patient privilege ensures that whatever is shared in the therapy session (with a few exceptions[14]) will be held in the strictest confidence. As a form of protected health information (PHI), information about an individual's past, present, or future mental health condition is protected by the Health Insurance Portability and Accountability Act (HIPAA) (Derlega et al., 2001). This assurance of confidentiality and protection

---

[14] Exceptions could be situations where a patients' mental condition makes them dangerous to themselves,, to someone else, or to someone else's property.

of the client's privacy provides a basis for trust, and such an "aura of trust" associated with the therapist can be further strengthened over time as the therapeutic relationship further develops (Gilbert & Leahy, 2007, p. 180). Without trust in a therapist and the understanding that their confidentiality and privacy will be protected, clients may be reluctant to communicate their thoughts and feelings, leading to ineffectual treatment (Grabois, 1997).

Farber et al. (2006) identified a sequence of events related to disclosure in psychotherapy. The process begins with clients having a positive attitude toward self-disclosure. Clients typically feel ambivalent before the disclosure, experiencing a desire to unburden themselves. They tend to feel vulnerable during the disclosure, sometimes experiencing shame or embarrassment. After the disclosure, they typically seek out and receive therapist approval and support.

The process of client disclosure in psychotherapy is accompanied by both negative and positive emotions. Before or during self-disclosure, clients often experience anxious or distressful feelings (e.g., shame and fear of being judged by the therapist) (Farber et al., 2006; Han & O'Brien, 2014). After disclosure, clients often feel positive feelings (e.g., relief and pride). These varying emotions help to explain why both disclosure and non-disclosure (i.e., withholding information) are observed and justified in psychotherapy.

2.6.1. Benefits of Client Disclosure in Psychotherapy

Clients' disclosure and willingness to share their thoughts and feelings contribute to the success of therapy. Pennebaker (1990) found that sharing secrets with others can generate psychological benefits by facilitating a greater sense of personal mastery. On the other hand, clients who fail to disclose their thoughts and feelings make the therapist's job more difficult in helping clients cope with mental health conditions (Hill & O'Grady, 1985). Some types of client non-disclosure (e.g., those related to negative thoughts that occur during sessions) are associated with lower client ratings of session outcomes and satisfaction (Regan & Hill, 1992). Level of disclosure was found to be positively related to the quality of the working alliance, the number of sessions the client had attended, the security and strength of the client's attachment to the therapist, the session depth, and therapy outcomes realized to date (Farber & Hall, 2002; Kahn et al., 2008; Saypol & Farber, 2010). Kelly and Yuan's (2009) study supports the long-standing belief that secret-keeping in therapy either hurts the therapeutic relationship or happens when the relationship is already relatively weak.

Researchers have used Inhibition Theory to explain why client concealment tends to be related to adverse mental health outcomes, while disclosure tends to be related to positive mental health outcomes. Inhibition theory posits that bottling up thoughts and feelings related to difficult experiences can be stressful and ultimately contributes to negative mental health consequences (Finkenauer & Rimé, 1998; Pennebaker, 1985, 1997). In contrast, disclosing one's thoughts and feelings, either through writing or talking, reduces stress and increases wellness (e.g., Pennebaker, 1985; Pennebaker & Beall, 1986).

While self-disclosure theories view the sharing of thoughts and feelings as a vehicle for fostering mutual understanding (Reis & Shaver, 1988), the theoretical model of perceived understanding (Reis et al., 2017) emphasizes how one's self-disclosure would have the positive effect of "feeling understood," which creates a sense of social belonging and acceptance, activating areas in the brain associated with connection and reward (Reis et al., 2017). In mental health care, clients' perceptions of feeling understood by their therapists contribute to strong psychotherapeutic alliances and successful treatment outcomes (Elliott et al., 2011; Pocock, 1997). The model of perceived understanding suggests that the positive effects of feeling understood are mediated by the extent to which disclosers perceive that they are understood. In other words, disclosers need to believe that their partner understands them before the positive impacts of feeling understood can take place (Reis et al., 2017; Reis & Shaver, 1988).

In traditional psychotherapy, a client's development of a sense of being understood necessitates disclosing to a human therapist. In the case of chatbot-based psychotherapy, clients might not feel the same way if they think that a computer program cannot understand them on a deeper level. The chatbot's responses may be seen as pre-programmed and inauthentic, preventing disclosers from feeling truly understood (Ho et al., 2018). To test this hypothesis, Ho et al. (2018) conducted an experiment to compare the downstream effects of disclosing to a chatbot partner with disclosing to a human partner. They found that chatbots and humans were equally effective at creating emotional (e.g., feeling better), relational (e.g., perceptions of partner warmth), and psychological (e.g., self-affirmation) benefits. Unfortunately, these findings are constrained by the researchers' use of the Wizard of Oz (WoZ) method (Dahlbäck et al., 1993). In Ho et al.'s experiment, participants were told they would have an online chat conversation with either a chatbot or a person. But in actuality, a hidden person behind the scenes (the “wizard”) was the one interacting with participants, even though they were told that the partner is a computer. Because the chatbot (in this case, a human in disguise) was perceived to present the same level of social and emotional intelligence as a human conversation partner, the use of this method cannot rule out confounding variables, such as social support and emotional support, that may have (also) contributed to the equally positive downstream outcomes of disclosing to a human vs. disclosing to a chatbot. At the current stage, even the most advanced chatbot falls short of humans in conducting natural conversations, but it is still probable that people would gain a sense of being understood after disclosing to a chatbot due to anthropomorphism. The current study addresses the WoZ methodological weakness and uses a real chatbot instead to investigate the following research question:

RQ6: How does anthropomorphism relate to people’s sense of being understood after making disclosures to the agent?

2.6.2. Self-presentation, Non-disclosure, and Need for Privacy

Countering traditional approaches to psychotherapy that demand a high degree of openness from clients, Kelly (2000) suggests that there are benefits associated with non-disclosure or keeping secrets in therapy. She conceptualizes psychotherapy as a self-presentational process. Clients might withhold personal information and reactions from their therapists, but such discretion can actually be associated with positive therapy process ratings and outcomes. To create favorable impressions, clients hide some undesirable aspects of themselves from their therapists, and eventually come to benefit from therapy by perceiving that their therapists have favorable views of them (A. E. Kelly, 2000). Hill et al. (2000) agreed that self-presentation is one reason that clients may choose to conceal secrets from their therapists. However, they argued that clients may also conceal information for other reasons: some clients feel uncomfortable with the power imbalance in the therapy relationship, given that therapists disclose very little; other clients are not ready to work on a particular problem or hold a belief that the therapist cannot help them. These alternative reasons for concealment echo research findings on the link between therapeutic alliances and disclosure.

Clients' need for privacy is also cited as a reason for non-disclosure (Hill et al., 2000). In psychology, the individual's need for privacy differs from the "right to privacy," which is defined as "the right to be let alone" (Warren & Brandeis, 1890). Schwartz (1968) considers privacy a necessary condition in order for us to exist in society. Maintaining one's privacy not only releases oneself from others but also comes as a separation from our own worldly judgment of ourselves. An early paper by Alan Bates (1964) defines privacy, pointing out its multiple benefits, as:

> …a buffer between social pressure and individual responses, a means of protecting vulnerability, and a sanctuary for psychological healing, and as a commodity that must be neatly balanced for healthy self-identity (too little privacy results in overstress, too much in a failure to engage in the world) (p. 408).

In discussing the role of privacy in a psychotherapy relationship, Winslade and Ross (1985) characterize a client's privacy as "something that they preserve by not granting access to others or by not disclosing personal information such as thoughts, sensations, feelings, memories or fantasies" (Winslade & Ross, 1985, p. 594). Privacy is a state of the self and the ability to grant or restrict access to that state. And the client's need for privacy calls for their control over what happens in therapy—independence from the therapist. From this perspective, Winslade and Ross argue that patients' privacy must be respected and honored by the therapist; it is the client's choice to disclose their information at their discretion.

In order to share hidden thoughts or desires with therapist, clients would expect and need to trust that such information will be kept confidential. Clients' expectation of confidentiality aligns with their need for privacy, even though privacy and confidentiality are two different concepts. Private information is held individually and owned by the client. In contrast, confidential information is shared and, though it still belongs only to one party, it is the client's trust in the therapist that ensures they retain ownership (Winslade & Ross, 1985).

People's privacy expectations vary according to the social norms around the particular context (Nissenbaum, 2004). Compared to traditional in-person therapy, it is a completely different context when people engage in psychotherapeutic conversations with a chatbot. The nature of privacy changes here because people do not disclose solely to the agent, as their personal information becomes a form of digitalized data. However, it is uncertain how client knowledge of this fact influences their disclosure intention and behaviors. Since there is no pressure to maintain a favorable self-presentation to a (chatbot) therapist, people might feel freer to disclose their thoughts and feelings. This tendency is evidenced by early research on computer-based mental health interventions that has found that people reveal their sensitive information, including stigmatizing behaviors, more truthfully to a computer than they do to human counselors (Meier, 1988). Similarly, e-therapy allows people to receive treatment in the privacy of their homes, maintaining a sense of anonymity. As a result, people feel more comfortable sharing their embarrassing, stigmatizing, or sensitive thoughts in e-therapy conducted online (Finfgeld, 1999; Manhal-Baugus, 2001).

On the other hand, privacy concerns might arise without inherent trust; the idea that autonomous, intelligent agents could have access to personal data is disturbing to many people, no matter how helpful the agent might be (Norman, 1994). This issue is further complicated when mental health care goes digital and online. Today's digital platforms are not only storing personal health information but collecting data about how consumers use the website or app. In 2019, the popular therapy-on-demand app Better Help was criticized for sharing sensitive user data. According to a report (Osberg & Mehrotra, 2020), despite the fact that the Better Help platform connects users with HIPAA-licensed counselors using encryption to ensure conversations recorded in the app were private, metadata about user behaviors—including the time, location, and duration of the therapy session—was being repurposed to sell more targeted ads.

Privacy concerns affect not only client disclosure but also people's decision to adopt or use a technology (Liao et al., 2019). An analysis of user reviews of e-therapy apps revealed that for mental health apps, privacy, security, and trust were critical to user experience (Stawarz et al., 2018). Given the new challenges and opportunities arising from chatbot-delivered psychotherapy, this study investigates the following questions with regard to client disclosure and information privacy:

RQ7: How comfortable do people feel about sharing their information with a chatbot in psychotherapeutic conversations?

2.7. The Role of Race

Given that one of the main contributions of this dissertation is to understand the effect of racial mirroring on outcomes from a mental health chatbot, the last section of literature review focuses on the role of race in psychotherapy and chatbot design.

2.7.1. Race Matching in Traditional Psychotherapy

Over the past several decades, psychologists have explored the effects of matching clients with therapists of the same race/ethnicity (Cabral & Smith, 2011). One stream of research focuses on how race matching influences the counseling process. Studies show that clients or patients are better able to express themselves when they are paired with therapists of a similar race or ethnicity (Cabral & Smith, 2011; Carkhuff & Pierce, 1967). Based on Cabral and Smith's (2011) meta-analysis, people prefer to have a therapist of their own race or ethnicity and tend to perceive therapists of their own race or ethnicity somewhat more positively than others. It is generally safe to assume that matching clients with therapists of the same race should result in stronger therapeutic alliances and bonding (Harrison, 1975; Russell et al., 1996; Sue et al., 1991).

Another stream of research emphasizes how race/ethnicity matching increases the perceived similarities in the values of client and therapist, which in turn predicts positive treatment outcomes (T. A. Kelly & Strupp, 1992). Specifically, racial/ethnic matching may improve client outcomes by enhancing mutual understanding between client and therapist and by reducing client concerns about being misunderstood or mistreated (Kohatsu et al., 2000). Some social psychology theories might explain client preferences for same-race therapists or counselors. For example, similarity is associated with perceived credibility (Simons et al., 1970). Interpersonal similarity not only influences individuals' preferences for social interactions but also their perceptions of those interactions.

2.7.2. Racial Group Differences in Sensitivity to Race Matching Effect

The magnitude of racial mirroring effects is not homogeneous across people of different races/ethnicities. The relevance of racial/ethnic matching was greatest among African American participants (Cabral & Smith, 2011). Their preferences for therapists of their own race/ethnicity has been found to be very strong; their perceptions of therapists varied substantially as a function of racial/ethnic matching and their outcomes in therapy tended to be mildly better when their therapist was also African American. Grantham (1973) interviewed 37 male and female Black college students and found that they preferred Black counselors over white counselors to a significantly greater degree; they also engaged in more self-disclosure and self-exploration. In contrast, racial/ethnic matching was apparently least relevant to White/Caucasian Americans. The results with Asian American and Hispanic/Latinx Americans were mixed. Asian Americans showed only mild preferences for a therapist of their own race/ethnicity but tended to provide more positive evaluations for therapists of their own race/ethnicity.

Earlier research has observed an inter-racial tension related to African American clients' perceptions of mental health services. Studies have consistently shown that African American clients tend to mistrust mental health services provided by White American therapists (O'Sullivan et al., 1989; Snowden, 1999; Sue, 1977; Sussman et al., 1987). Explanations for this finding include the possibility of a perceived racial bias in the provision of mental health services and the implicit association of mental health services with the values of White/Caucasian Americans (Ridley, 1984).

2.7.3. Race and Chatbots

A few HCI studies address concerns about how chatbots handle conversations involving racial stereotypes (Marino, 2014; Schlesinger et al., 2018). From an ethical and normative perspective, these studies are pivotal in raising awareness around the issue of race in conversation design. At the algorithm level, Schlesinger et al. (2018) revealed racial bias issues in machine learning algorithms underlying chatbot technologies and called for a better understanding of the potential racial biases and stereotypes of any algorithmic inputs that shape chatbot interactions. At the conversation level, Marino (2014) noted that race is an inevitable consideration when constructing a chatbot's identity, especially when chatbots gain visual representations. And a chatbot's racial formation becomes "a site of intense contestation, of negotiation between conflicting attempts to produce and define racial identity in digital, computational environments" (p. 3). Brahnam and De Angeli

(2012) showed how an agent's racial identity might have the unintended consequence of problematic use. They found that when avatars were presented as Black adults, references to race can deteriorate into racist attacks. With a particular focus on the effect of racial mirroring, Baylor and Kim (Baylor & Kim, 2004) examined the impact of pedagogical agents' ethnicity on learners' perception of the agents. Through an experiment with undergraduate participants, they found that students who worked with pedagogical agents of the same ethnicity rated them as more credible, engaging, and affable than agents of a different ethnicity.

The issue of race is important and complex, yet understudied. There is a need to place racially conscious intelligent interface design on the agenda, especially due to the lack of racial diversity among human therapists. Only 14% of therapists in the U.S. are non-White, while 38% of Americans belong to racial and ethnic minority populations (L. Lin et al., 2018). While it might be challenging to bridge the racial gaps in the mental healthcare workforce, conversational agents that give the perception of racial mirroring might be a cost-effective way to address people's unmet needs for a same-race therapist. To validate such a racial-mirroring strategy for chatbot therapy, this study investigates the following research question:

RQ8: How does racial mirroring influence people's perceptions of and interactions with a chatbot in psychotherapeutic conversations?

Chapter 3: System Design of a Chatbot for Psychotherapy

To answer the research questions outlined in Chapter 2, I designed a multi-identity conversational system for therapy (MICST) that holds psychotherapeutic conversations and assumes one of nine personas varying in gender and race/ethnicity. This chapter details how the chatbot was built. I first describe the technical configuration of the system that makes it possible for the chatbot to engage in interactions with users. Next, I explain how the conversation scripts were created based on Motivational Interviewing, a technique in cognitive behavior therapy. Finally, I introduce how agent personas varying in race and gender were created.

3.1. Chatbot Technical Configuration

I developed the back-end of the MICST system on Google's DialogFlow, a platform for building conversational interfaces such as voice interaction apps and text-based chatbots. I chose to build this conversational system using scripts[15], or elements of dialogue that are crafted by real humans rather than a set of machine learning algorithms. The next section (Conversation Design) describes how these conversation scripts were created following Motivational Interviewing, a cognitive behavior therapy method.

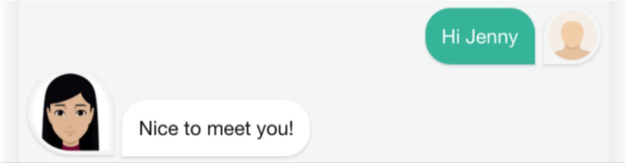
To create the front-end interface with different agent personas, I worked with a software engineer to integrate the application with an external web-based interface imitating a messenger app. Specifically, we used the Node.js client DialogFlow, which is connected by JavaScript in the interface in order to relay responses from

---

[15] Such predefined scripts ensure consistency in therapeutic intervention across different conversational sessions, while responses generated by machine learning algorithms would be unpredictable and uncontrollable.

DialogFlow to users. In order to imitate human-to-human typing speed and to allow enough time for people to process information, there is a 1.5-second latency in between the users' input and the chatbot's response. To imitate a messaging app interface, the chatbot interface was built as a web application using Javascript, React.js, and HTML5. The flexibility of such a web-based interface allowed us to create multiple versions of the chatbot system, each with a different visual presentation indicating a different racial persona. I purposefully limited visual elements that are not relevant to the current study. For example, the user profile picture is undefined and generic, with a downsized icon (as shown in Figure 4).

Figure 4. Downsized, undefined icon for the user; Larger icon for the agent's profile.



Hi Jenny

Nice to meet you!

From the user's perspective, this interface resembles a text-based chat window. Users can type in text messages to continue the conversation. They also can click the "quick reply" (QR) buttons, which are pre-set options changing based on the context and serving two interaction purposes. First, QR buttons allow users to respond to the agent in an efficient way, reducing the friction of conversational interactions. Second, the user's selection of the QR buttons helps the system provide more tailored responses.

A cloud-based database was created and hosted using Amazon cloud service, and linked to the DialogFlow platform. All user interactions, including text input and QR button clicks, are recorded and stored in this database.

3.2. Conversation Design

3.2.1. Four Motivational Interviewing (MI) Processes

MICST was built to lead a structured conversation following the typical MI flow in four processes (Miller & Rollnick, 2013): Engaging, Focusing, Evoking, and Planning.

Although Google's DialogFlow allows the chatbot to provide various questions or responses, I placed a higher priority on the consistency of experimental conditions. Therefore, each participant was asked the same set of questions and the chatbot provided feedback accordingly. What the chatbot said—including questions and responses to the user's input—was predefined following conversation scripts. During the conversation, users were asked to provide information about themselves so that the system could provide more tailored responses. For example, users were asked their names so that the chatbot could address them by the name they entered. Users engaged with the agent by typing their responses via the text box provided.

I designed MICST's conversation flow to closely align with the four MI processes. In Engaging, the agent greets the user and offers a self-introduction to build a relational foundation. At the Focusing stage, the agent asks the user to identify and detail their biggest challenge in life, with a follow-up prompt probing their feelings and thoughts. This leads to Evoking, where the agent explores future goals with the user, prompting and inspiring their own ideas for change. Finally, at the Planning stage, the agent invites the user to specify a tangible change that they can make to overcome the challenge. Table 1 lists sample questions and responses that

MICST would provide during an MI conversation. Appendix I presents a complete conversation based on a real participant's interactions with the agent.

Table 1. Sample chatbot questions and responses by MI process.

MI Process	Sample Chatbot Question / Response
Engaging	"My goal today is to inspire any changes you'd need to overcome the big challenge."
Focusing	"What happened? How has it influenced your life?"; "How do you feel about all this?"
Evoking	"How would you like things to turn out for you now, ideally? What's your goal?"; "Let's talk about your strength. What strong points do you have that could help you move forward?"
Planning	"What would you say is the most important thing that you can do to achieve this goal?"

3.2.2. Linguistic Style

For a chatbot adopting the role of psychotherapist, building rapport and demonstrating empathy is critical. Linguistic style is a key aspect of the creation of artificial agents (Walker et al., 1997). I adapted my chatbot's linguistic style from prior work (Bickmore & Picard, 2005; Boukricha & Wachsmuth, 2011; Inkster et al., 2018; Moon, 2000; Morkes et al., 1999; Walther, 1996; Walther & D'Addario, 2001), aiming to develop attachment bonds between the agents and users. All agents speak English, using the same conversational syntax and style. Table 2 lists sample responses MICST provides that reflect different linguistic characteristics.

Table 2. Sample dialogues for linguistic characteristics

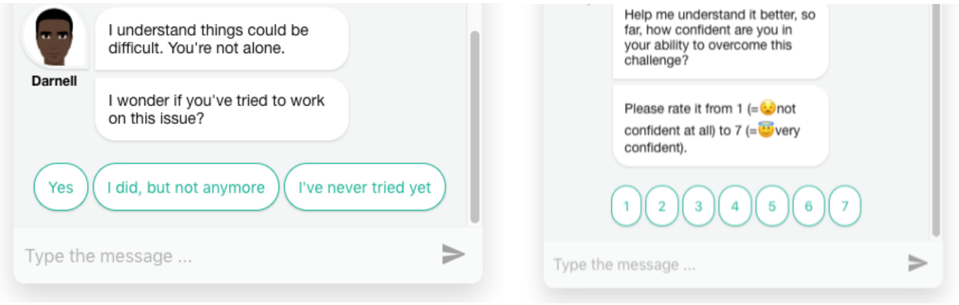
Linguistic Characteristic	Sample Dialogue
Greetings (Cassell & Bickmore, 2003)	"Welcome! I'm Molly"; "Nice to meet you!"
Emoticons (Walther, 1996)	"One of my most important tasks today is to help you regain strength.🌈☀️"
Small talk (Bickmore & Cassell, 2012)	"I know I'm not real, but I'm here to listen to you and help you through life."

Self-disclosure (Lee & Choi, 2017)	“As a conversational AI, I can engage you in a ‘talk’ session, just like that in psychotherapy! ”
Empathy (Inkster et al., 2018)	“I understand things could be difficult. You’re not alone.”
Meta-relational talk (Bickmore et al., 2010)	“Think of me as a texting buddy that gets better every time we talk!”
Humor (Morkes et al., 1999)	“… can you name one specific thing that you could do to overcome the challenge, maybe something small, like an apple a day?”

3.2.3. Quick Reply Buttons

In any therapy setting, therapists need to gather specific information about clients’ personal experiences and/or mental states in order to provide tailored intervention. To that end, I equip MICST with “Quick Reply” (QR) buttons, which allow users to select one from several predefined options in order to provide a just-in-time self-evaluation or quick responses to questions. Examples of such interactions are depicted in Figure 5 below. Next, the agent would follow up with tailored responses based on the users’ report of their experiences or psychological state. For instance, when a user indicated that he has tried to fix the issue, but has since abandoned taking action, the agent would follow up with the following response: “One of my most important tasks today is to help you regain strength.” When the user indicated he had never tried, the agent would say, “No worries, let’s do it together.” Likewise, when the user indicated that he was not confident at all about dealing with the challenge, the agent would say, “I have a lot of confidence that working together we can find a way out.” For users with high confidence, the agent would respond, “It seems like you have a lot of courage! That’s a good sign.” Figure 5 shows two examples of predefined QR buttons.

Figure 5. Two examples of predefined options for providing tailored responses.



I understand things could be difficult. You're not alone.

I wonder if you've tried to work on this issue?

Help me understand it better, so far, how confident are you in your ability to overcome this challenge?

Please rate it from 1 (=😟not confident at all) to 7 (=😇very confident).

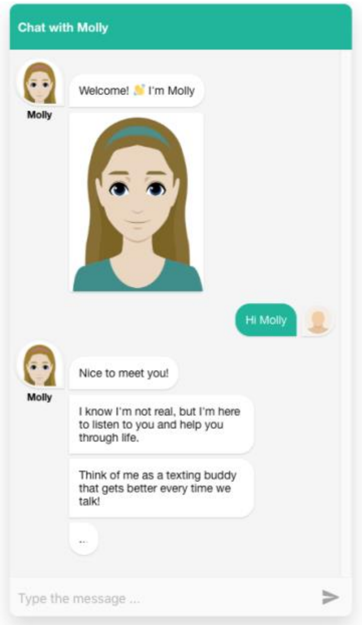
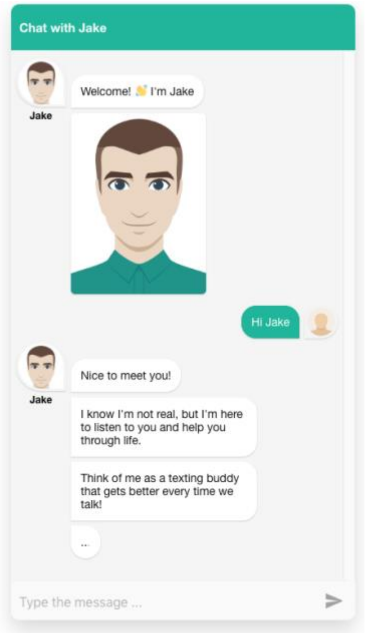
3.3. Chatbot Agent's Racial Identity

This section explains how MICST is equipped with different personas varying in racial identities. It is important to note that the chatbot racial identity constructed through an image and a name is a step towards understanding racial mirroring but does not include other factors that likely matter a lot (e.g., shared language, experiential backgrounds, cultural values). For this research, the primary cue for the agent's racial identity is their profile picture. An enlarged depiction of this profile picture was presented to users shortly after greetings. Throughout the conversation, this profile picture was displayed alongside each conversational message by the agent in a familiar text message environment.

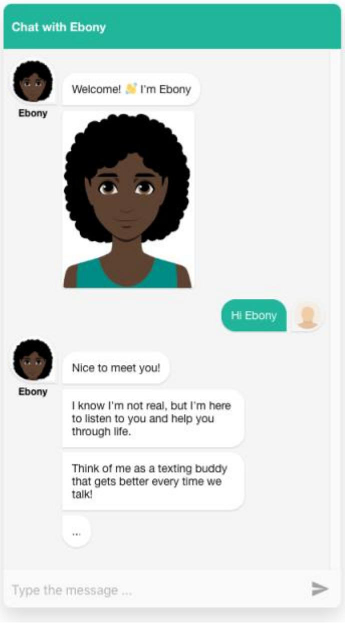
The secondary cue is the agent's name. I assigned agents with names that implied their racial identity (ABC News, 2015; Fryer & Levitt, 2003). For example, Jake to suggest a Caucasian identity, Ebony to suggest an African-American identity, and Antonio to suggest a Latinx identity. These name choices were based on previous study examples (Fryer & Levitt, 2003) and a report (ABC News, 2015) of name popularity based on race and ethnicity. Due to the cultural diversity and linguistic plurality among Asian people (unlike that for Latinx people), it is not feasible to use a country-specific name to represent all Asian subgroups, and too much specificity in name selection might be counterproductive to creating the racial mirroring condition. Therefore, two Asian personas were assigned White names, Jenny and David, and their racial identity can only be inferred from their profile pictures.

In total, eight racially diverse personas were created, including Molly (White/Caucasian female), Jake (White/Caucasian male), Ebony (Black/African American female), Darnell (Black/African American male),Jenny (Asian female), David (Asian male), Camila (Hispanic female), and Antonio (Hispanic male). These agents interacted with participants in the racial mirroring and non-mirroring groups in the online experiment. Chapter 4 (Methodology) describes the experiment and each of these experimental groups in further detail. Additionally, we created Robbie, a non-personified and gender-neutral agent that resembles a typical robot. Robbie interacted with participants in the control group. Table 3 shows nine agents with different gender and racial identities created for MICST.

Table 3. Visual cues indicating agent racial identity

Molly (White female agent)	Jake (White male agent)
 Chat with Molly Welcome! 👋 I'm Molly Hi Molly Nice to meet you! I know I'm not real, but I'm here to listen to you and help you through life. Think of me as a texting buddy that gets better every time we talk! ... Type the message ...	 Chat with Jake Welcome! 👋 I'm Jake Hi Jake Nice to meet you! I know I'm not real, but I'm here to listen to you and help you through life. Think of me as a texting buddy that gets better every time we talk! .. Type the message ...

Ebony
(Black female agent)



Welcome! 👋 I'm Ebony

Hi Ebony

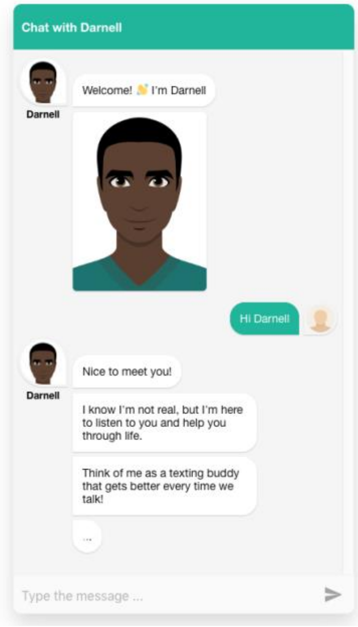
Nice to meet you!

I know I'm not real, but I'm here to listen to you and help you through life.

Think of me as a texting buddy that gets better every time we talk!

...

Darnell
(Black male agent)



Welcome! 👋 I'm Darnell

Hi Darnell

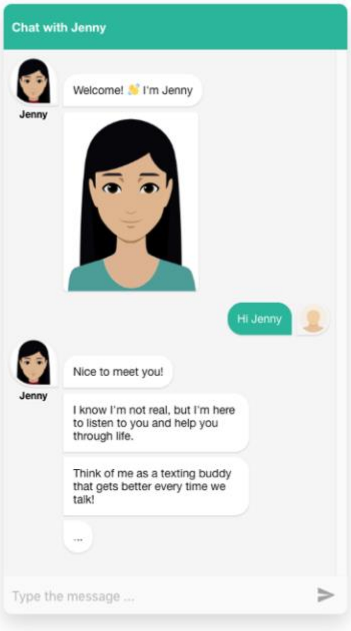
Nice to meet you!

I know I'm not real, but I'm here to listen to you and help you through life.

Think of me as a texting buddy that gets better every time we talk!

...

Jenny
(Asian female agent)



Welcome! 👋 I'm Jenny

Hi Jenny

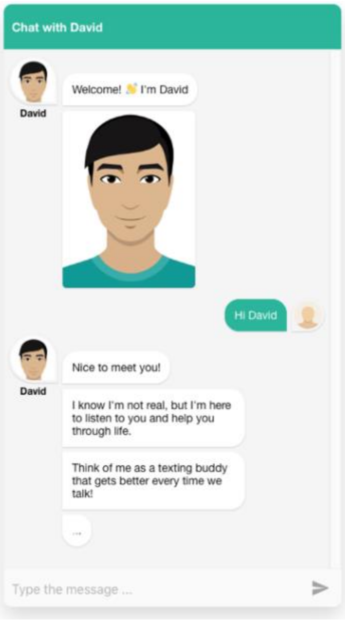
Nice to meet you!

I know I'm not real, but I'm here to listen to you and help you through life.

Think of me as a texting buddy that gets better every time we talk!

...

David
(Asian male agent)



Welcome! 👋 I'm David

Hi David

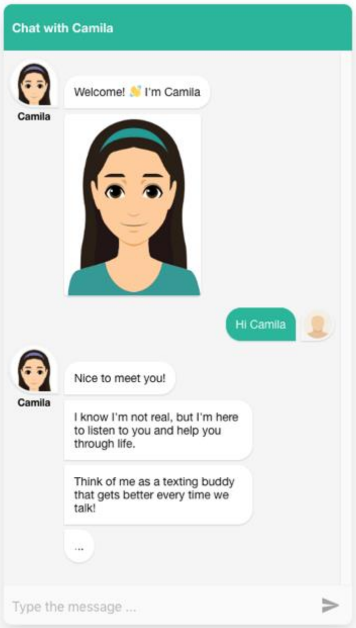
Nice to meet you!

I know I'm not real, but I'm here to listen to you and help you through life.

Think of me as a texting buddy that gets better every time we talk!

...

Camila
(Hispanic female agent)



Chat with Camila

Camila: Welcome! 👋 I'm Camila

Hi Camila

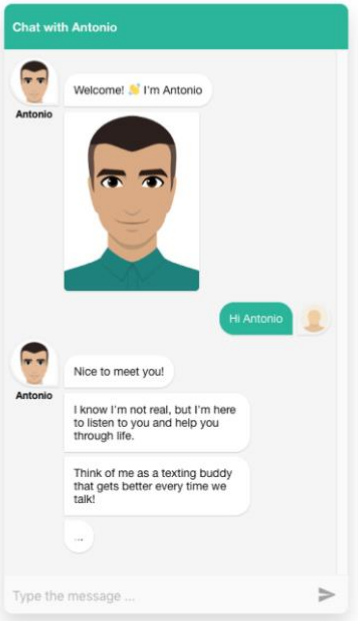
Camila: Nice to meet you!

I know I'm not real, but I'm here to listen to you and help you through life.

Think of me as a texting buddy that gets better every time we talk!

...

Antonio
(Hispanic male agent)



Chat with Antonio

Antonio: Welcome! 👋 I'm Antonio

Hi Antonio

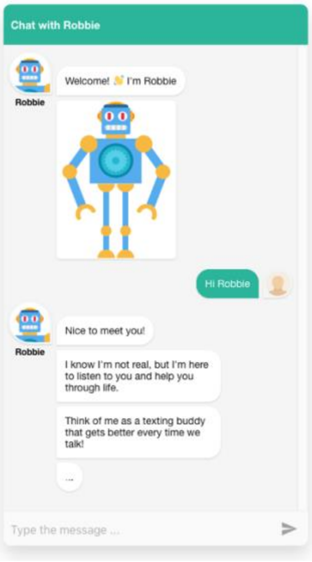
Antonio: Nice to meet you!

I know I'm not real, but I'm here to listen to you and help you through life.

Think of me as a texting buddy that gets better every time we talk!

...

Robbie
(Non-personified, gender neutral agent)



Chat with Robbie

Robbie: Welcome! 👋 I'm Robbie

Hi Robbie

Robbie: Nice to meet you!

I know I'm not real, but I'm here to listen to you and help you through life.

Think of me as a texting buddy that gets better every time we talk!

...

Chapter 4: Methodology

In this chapter, I present the online experiment design and procedures, including participant recruitment and screening, experimental and control groups, and survey instrument. The online experiment sought to uncover the interrelationships between design outcomes (anthropomorphism and racial mirroring) and the relational outcomes, including sociotechnical and therapeutic relationships. Additionally, a follow-up interview study was conducted to gather qualitative insights from people who had participated in the experiment. My research methodology, including study design and recruitment process, was approved by the University of Maryland's IRB. Table 4 shows how each of my methods map to the eight research questions.

Table 4. Mapping methods to research questions

Research Questions	**Experiment with Survey**	**Interview**
RQ1: What are the benefits and drawbacks of utilizing a chatbot for managing mental health compared to a human therapist?	✓	✓
RQ2: What factors influence the sociotechnical relationships between human and chatbot for mental health support?	✓	✓
RQ3: How does anthropomorphism influence the way people perceive and interact with a chatbot in psychotherapeutic conversations?	✓	✓
RQ4: What factors are associated with therapeutic relationship development between client and conversational agents?	✓	✓
RQ5: What is the effect of the therapeutic alliance on human-agent psychotherapeutic conversations?	✓	
RQ6: How does anthropomorphism relate to people's sense of being understood after making disclosures to the agent?	✓	✓

RQ7: How comfortable do people feel about sharing their information with a chatbot in psychotherapeutic conversations?	✓	✓
RQ8: How does racial mirroring influence people's perceptions of and interactions with a chatbot in psychotherapeutic conversations?	✓	✓

4.1. Online Experiment

4.1.1. Participant Screening

Participants were recruited from Amazon Mechanical Turk and Prolific, which are two recruiting and crowdsourcing platforms widely used by researchers.

Before proceeding, all participants consented to participate and confirmed that they were at least 18 years old and living in the United States. Participants first filled out a screening questionnaire that asked them about their psychological well-being and racial background. Since the experimental conditions centered on perceived racial (dis)similarity, I followed two selection criteria in the recruitment process. First, it was necessary to only recruit people who have a clear, unambiguous self-identification with one particular race. Therefore, people who self-identified as multi-racial were screened out and were unable to participate in the study. Second, to ensure the validity of the experimental conditions based on perceived racial similarity, participants whose facial features resemble that of one of the agents were selected. As such, the screener (hosted on Qualtrics) automatically exclude respondents who self-identified as Native Hawaiian or other Pacific Islander, Native American, or Alaska Native. Additionally, since Asians are a greatly diverse group, for people who identified as Asian, the questionnaire also asked to specify the region to which they would trace their cultural heritage. I only recruited Eastern Asian and mainland

Southeast Asian participants for the online experiment, as the facial features of Asian agents David and Jenny are more aligned with these two groups.

In August 2019, the first data collection of 217 responses was completed with the majority of participants being White/Caucasian. Due to the need to ensure sufficient sample size for minority groups, the second data collection was launched in October 2020, with the screening criteria narrowed to participants who were Asian, Black and Latinx. At the close of survey, an additional 189 responses were collected. In total, 406 complete survey responses were collected out of 838 participants who took up the initial screener survey. Each participant who completed the online experiment was paid $3. The average completion time was slightly over 16 minutes.

The final sample were 165 White/Caucasian Americans, 93 Black/African Americans, 77 Latinx Americans, and 71 Asian Americans. A more detailed description of participant sociodemographic background is presented in Section 4.2.2. The screening survey is attached as Appendix II.

4.1.2. Experimental Groups

As a part of my experimental design, eligible participants went directly to a web page to interact with an agent by random allocation. Specifically, at the end of the screening survey, I applied two layers of randomizers to allocate eligible participants into one of the three groups described below. For each racial group, the first layer of randomizers was applied to distribute participants randomly into three experimental conditions. The second layer of randomizers was embedded within the non-mirroring group and the mirroring group to ensure equal likelihood that participants will interact with one of the agents within that condition.

- *Racial-mirroring group:* A participant would interact with one of the two same-race agents. For example, a Latinx participant would talk to either Camila or Antonio. Under this condition, the agent's gender was randomized and participants were evenly distributed based on agent gender, because a previous study suggested mixed-effects of gender (Stieger et al., 2018).
- *Non-mirroring group:* A participant would interact with one of the six different-race agents. For example, a Latinx participant would talk to Molly, Jake, Ebony, Darnell, Jenny or David (all non-Latinx agents) with equal probability.
- *Control group:* A participant would interact with Robbie, the non-personified, gender-neutral agent.

Among 406 people sampled[16], 170 participants were assigned to the racial-mirroring group, 143 were assigned to the non-mirroring group, and 92 participants were assigned to the control group.

4.1.3. Interaction with MICST

After completing the screening survey, participants were provided a link to the conversational agent, with instructions indicating that they would chat with a "conversational artificial intelligence (AI) that is designed to help people manage their psychological well-being" and that they would be asked about their life experiences, as well as their feeling and thoughts. Participants interacted with the chatbot, which used the MI processes described in Chapter 3. A full transcript of a

---

[16] For the second data collection, due to the difficulty recruiting people of color, I had to over-assign participants to mirroring and non-mirroring groups to ensure sufficient sample for various human-agent pairs.

sample conversation between the agent and a real participant is included in Appendix I. Participant self-disclosure primarily occurred in the last three processes—Focusing, Evoking and Planning. In Focusing, the agent asked the participant to talk about their biggest challenges in life. Participants brought up a variety of life struggles, including: behavioral issues (e.g., P93: "*I struggle with gambling problems*"), financial stress (e.g., P52: "*There never seems to be quite enough money for everything*"), job-related (e.g., P118*: "I completed my degree but I'm not able to find a suitable job for 6 months"*), health problems (e.g., P144: *"I suffer from severe vertigo which limits my life in terms of movement*"), relationship issues (e.g., P142: "*My so-called boyfriend cheated*"), and negative thoughts (e.g., P93: "*I feel like a failure*").

The depth of participants' self-disclosure varied. Some participants provided a detailed picture of their struggle. For example, P85 wrote, "*My husband has had prostate surgery and is now on bed rest for a full month and I need to take care of him and also work.*" In contrast, other participants submitted very cursory responses, such as "*I have jealousy issues* (P48)."

Next, the agent asked the participant how they felt about their major challenge in life. Participants were encouraged to be open about their feelings in their disclosure, as the agent would ask: "How do you feel about all this? You can trust me with your feelings and thoughts." Again, participants expressed their feelings in varying length and depth, with some being more specific (e.g., P79: "*I feel stupid to be addicted to cigarettes, it's a brain poison*") than others (e.g., P17: "*I feel frustrated and stressed*").

In the Focusing stage, the agent would ask participants, “How would you like things to turn out for you, ideally? What’s your goal?”; participants offered a variety of targeted behaviors or goals depending on their challenges. Some of these goals are generic (e.g., “To be fit again” or “to be successful”), while some are specific (e.g., P29: “*About a year, my goal was to reduce my debts by 70%, so far I have been on track but it has been difficult to maintain”).*

In the Evoking stage, participants were guided by the agent to explore the potential changes that they would need to make to achieve their goal(s). In response to the agent’s question, “What would you say is the most important thing that you can do to achieve this goal?”, most participants came up with different changes they need. As a financially stressed participant said, the change he needed was “*To get a new job and maybe one in a low cost of living city” (P108).* However, some participants were not able to distinguish a goal and the changes they needed to make in order to achieve that goal, or the changes they identified were too generic.

In the Planning stage, the agent prompted participants to consider any obstacles they might have encountered or foresaw. Many participants pinpointed the obstacles they faced when trying to make a positive change toward their goals. For instance, a participant whose goal was to lose weight said, “*It is hard to keep at it all the time. I am sometimes stressed and it is hard to resist tasty foods. Sometimes it is hard to find time to work out” (P132).* However, some participants conflated obstacles with the problem they faced. Furthermore, participants were asked to think about their strengths. Similarly, participants’ disclosures varied in terms of length, depth, and content sensitivity.

4.2. Survey Instrument

At the end of the conversation with the agent, participants were provided a link to the post-interaction survey, which asked about their perceptions of the agent. I applied the piped text feature in Qualtrics to customize the agent name as per the specific agent each participant interacted with. The survey instrument is attached as Appendix III.

4.2.1. Dependent Variables

The main dependent variables (DVs) are described below. These include two variables related to the sociotechnical relationships (Likelihood to Recommend, Willingness to Continue Engaging with the Agent) and three variables related to therapeutic relationships (Working Alliance, Perceived Interpersonal Closeness, Projected Future Relationship). Additionally, five variables evaluated human-agent conversation (Comfort with Self-Disclosure, Comfort with Data/Information Sharing, Conversation Engagement, Content Sensitivity, and Length of Conversation). Finally, preference for agent persona was evaluated.

Likelihood to Recommend (M = 59.26, SD = 30.58, α = .93). This original 2-item measure was included to assess to what degree participants would recommend the agent 1) to their friends or family for managing mental well-being and 2) to people who have a need for counseling/psychotherapy. Participants responded using a slider ranging from 1 (Would not recommend at all) to 100 (Would definitely recommend).

Willingness to Continue Engaging with the Agent (M = 62.82, SD = 30.40). I adapted the instrument from Bickmore et al. (2005) with an emphasis on interacting with the agent for psychotherapeutic purposes. Participants rated how much they would like to continue working with the agent, from 1 (Not at all) to 100 (Very much willing).

Working Alliance (*M* = 3.47, *SD* = .04, *α* = .92). The main component of the participant-agent therapeutic relationship was assessed using the Working Alliance Inventory for Technology-Based Interventions (Kiluk et al., 2014), which is an adapted version of the original inventory evaluating the strength of therapeutic alliance between clients and therapists. Responses were recorded from 1 to 5, with higher scores indicating higher degrees of collaborative relationships between human and agent in the interaction. Sample items include: *I feel that this conversation with [agent name] helped me to accomplish the changes that I wanted*; *We have established a good understanding of the kind of changes that would be good for me*; *What I was doing in the talk session gave me new ways of looking at my problem*; *I find what I was doing in the talk session confusing* (reverse coded); *I find what we were doing during the talk session was unrelated to my concerns* (reverse coded).

A total of 145 (35.7%) participants agreed or strongly agreed with these statements, indicating a strong therapeutic alliance between clients and the chatbot therapist.

Perceived Interpersonal Closeness (*M* = 3.38, *SD* = .06). The Perceived Interpersonal Closeness Scale (PICS), a single-item pictorial scale (Popovic et al., 2003), is an established and reliable instrument to measure the perceived closeness of a relationship. To answer the question “How close did you feel to the conversational AI?,” participants chose a specific segment of the picture to reflect the relationship between the agent and the self. The picture consists of six concentric circles, with the innermost circle indicating “self”, the first circle surrounding the “self” circle indicating “fully close”, and the outermost circle indicating “distant”. Depending on which layer of circle the participant selected, their perceived interpersonal closeness with the agent was coded from 1 (Distant) to 6 (Fully close).

A total of 179 participants (43.2%) reported a moderately close (4), very close (5), or fully close (6) relationship between themselves and the agent.

Projected Future Relationship (*M* = 3.56, *SD* = .09). Because therapeutic relationships often develop over time, the survey evaluated participants’ projected future relationship with the agent. Participants were asked to predict how close their relationship with the agent would become, under the assumption that they would have conversations with the agent at least three times per week.

This is measured with the inclusion-of-the-other-in-the-self (IOS) Scale (Aron & Fraley, 1999), an established 7-point pictorial scale to evaluate relationship strength. Out of seven figures, participants were asked to select the one that best describes the potential of their relationship with the agent. Each figure includes two equal-sized circles side by side, with varying degrees of overlapping indicating different levels of projected future relationship closeness. The score was coded from 1 (two circles without any overlapping, indicating a distant relationship) to 7 (two circles almost overlap with each other, indicating a very close relationship). This IOS scale was chosen over PICS scale to evaluate projected future relationship because the PICS scale includes only one figure, while the IOS scale is comprised of seven figures representing various degrees of relationship strength, which best reflects the process of relationship development over time.

A total of 191 participants (46%) reported a score higher than 4, indicating that they projected a fairly strong, strong, or very strong connection with the agent.

Comfort with Self-Disclosure (*M* = 3.87, *SD* = .62, *α* = .87). This original, 5-item measure evaluated how comfortable participants felt about disclosing to the agent. On a 5-point Likert scale ranging from 1 (Strongly disagree) to 5 (Strongly agree), participants rated their level of agreement with statements such as: *I felt comfortable when I talked with [Agent Name] about my challenge and life experiences*; *I expressed myself openly*; *I felt [Agent Name] wouldn't judge me when I talked about my issue*.

Comfort with Data/Information Sharing (*M* = 41.68, *SD* = 24.67, *α* = .92). To evaluate how comfortable participants felt about sharing their data with the agent, the survey asked participants to rate their level of comfort with sharing 12 types of data, including email address (*M* = 52.21, *SD* = 34.18), phone number (*M* = 31.82, *SD* = 31.55), birthday (*M* = 57.34, *SD* = 33.7), location data (*M* = 29.70, *SD* = 31.54), phone call history (*M* = 17.77, *SD* = 26.82), employment history (*M* = 31.63, *SD* = 33.16), photos and videos (*M* = 20.49, *SD* = 28.46), medical history (*M* = 31.41, *SD* = 33.64), sexual activity (*M* = 27.91, *SD* = 31.63), relationship status (*M* = 60.78, *SD* = 32.08), sexual orientation (*M* = 65.70, *SD* = 32.5), disease-related information (*M* = 43.2, *SD* = 34.4), and mood tracking (*M* = 58.4, *SD* = 32.67). Participants chose their level of comfort from 1 (Absolutely not comfortable sharing) to 100

(Completely comfortable sharing). Each participant's final score for comfort with data sharing was calculated by averaging all items.

Human-Agent Conversations. The conversation by itself is an important outcome from human-agent interactions. The study evaluated three dimensions of human-agent conversations.

- **Content Sensitivity** (*M* = 59.02, *SD* = 1.41)**.** The survey asked "Overall, how sensitive would you rate the content of your conversation with [agent name]". Participants responded on a scale from 1 (not sensitive at all) to 100 (highly sensitive).
- **Level of Conversation Engagement** (*M* = 63.74, *SD* = 1.30)**.** The survey asked participants, "How would you rate the level of engagement of your conversation with [agent name]?" Participants responded using a scale from 1 (not engaging at all) to 100 (highly engaging).
- **Length of Conversation.** The direct word count of participants' conversations with agents were calculated as a proxy for length of conversation.

User Preference of the Agent's Racial Persona. At the end of the survey, all nine agent personas were presented, including the eight racially diverse, mixed-gender agents and one robot-like persona. Participants were asked to identify which agent(s) they would prefer to interact with for mental well-being support. Overall, 60% of participants selected same-race agents. Black (n = 94) and Asian (n = 75) participants reported higher rates of same-race preferences, with 78% and 67%, respectively.

4.2.2. Independent Variables and Control Variables

Several measures evaluating participants' backgrounds, attitudes, and experience were included as independent variables and control variables.

Racial Mirroring. The experimental design divided participants into three study groups, including racial mirroring, non-mirroring, and control conditions.

Anthropomorphism Scale (*M* = 65.72, *SD* = 18.96, *α* = .93). The Godspeed instrument (Bartneck et al., 2008) has been used in human-robot interaction (HRI) research to measure various perceptions of AI. The survey administered in this study included the dimensions of overall Anthropomorphism (*M* = 56.59, *SD* = 1.34, *α* = .96), Liveliness (*M* = 61.55, *SD* = 1.15, α = .92), Likeability (*M* = 80.65, *SD* = .79 α = .90), and Perceived Intelligence (*M* = 70.82, *SD* = .96, α = .95). Participants rated their impressions of the agent using linguistic scales made up of word pairs, such as fake—natural, machinelike—humanlike, unfriendly—friendly, awful—terrible. Responses were coded from 1 to 100, with lower ratings indicating the chatbot being less human-like or possessing more negative attributes.

General Attitudes toward AI (*M* = 3.38, *SD* = .04, α = .93). I adapted the instrument from Cave et al. (2019) to evaluate participants' general attitudes toward artificial intelligence. Participants rated their level of agreement with a series of statements, using a scale that ranged from 1 (strongly disagree) to 5 (strongly agree). Higher scores indicated more favorable attitudes. Example items include: *AI might become the perfect friend, there to listen whenever we need and respond to our every desire*; *AI might cater to all our desires so well that we prefer AI interaction to* *human interaction*; *AI companionship might be very popular*; *I'm excited for the future that we can do a lot of things with the help of artificial intelligence*.

Perceived Usefulness / Efficacy (*M* = 3.44, *SD* = .05, *α* = .94). This is a 6-item measure to assess participants' satisfaction with their experience with the chatbot and the perceived usefulness of their interaction. I adapted the instrument of Venkatesh et al. (2012) to reflect chatbot use and satisfaction within the psychotherapeutic context. Sample statements include: *My interaction with [Agent Name] was satisfying*; *After my conversation with [Agent Name], I felt more confident about my ability to overcome the challenge*; *This technology is useful for me to manage my mental well-being*. Participants chose their level of agreement with these statements, ranging from 1 (Strongly disagree) to 5 (Strongly agree).

Perceived Ease of Use (*M* = 3.87, *SD* = .04, *α* = .83). A four-item measure adapted from Venkatesh et al. (2012) was included to assess the perceived ease of use of this technology. Scoring ranged from 1 (Strongly disagree) to 5 (Strongly agree), and participants chose their level of agreement with these statements: *Overall, this conversational AI performs well without a glitch*; *Overall, I think it is easy to use this technology*; *This conversational AI operates reliably*; *Having a conversation with [the agent] is intuitive*.

Social Influence (*M* = 3.49, *SD* = .04, *α* = .81). A six-item measure adapted from Venkatesh et al. (2012) was assessed social influence and other facilitating conditions for using this technology. Scoring ranged from 1 (Strongly disagree) to 5 (Strongly agree), and participants chose their level of agreement with these statements: *People who are important to me would be very supportive of my use of* *this technology; My friends or family would think my use of this technology is a good idea; Using this technology will impress other people around me; I have the knowledge or skills necessary to use this technology; Using the technology is compatible with the way I deal with my challenges; People would not judge me if they see me interacting with [Agent name].*

Desirability (*M* = 3.31, *SD* = .04, *α* = .72). A three-item measure adapted from Venkatesh et al. (2012) was used to assess how desirable or appealing the chatbot technology was for mental health support. Scoring ranged from 1 (Strongly disagree) to 5 (Strongly agree), and participants chose their level of agreement with these statements: *I prefer to talk to a human therapist instead (reverse coded); I got bored quickly when using this technology; The persona of [Agent name] was unappealing.* Higher scores indicated participants perceived the chatbot to be more desirable and appealing.

Control variables. Participants provided basic demographic background information, including their sex, current age, annual income, and education. Privacy concerns (Vitak, 2016), data sensitivity (measured by the question "Overall, how sensitive would you rate the content of your conversation with [agent name]"), psychological well-being (Ryff & Keyes, 1995), and satisfaction with life (Diener et al., 1985) measures were also included as control variables.

Descriptive analysis of the sociodemographic factors and control variables is detailed in Table 5 below. There were no statistically significant differences in participant characteristics by experimental groups.

Table 5. Survey participants socio-demographics background

	Full sample (*n* = 406)		**Racial Mirroring (*n* = 170)**		**Non-Mirroring (*n* = 143)**		**Control (*n* = 92)**		**Between-group differences**
	N	*%*	*N*	*%*	*N*	*%*	*N*	*%*	*Sig.*
Sex									
Male	187	47.00%	76	45.50%	69	49.60%	42	45.70%	.87
Female	211	53.00%	91	54.50%	70	50.40%	50	54.30%	
Race									
White/Caucasian	165	40.60%	60	35.10%	55	38.70%	50	53.80%	.17
Black/African	93	22.90%	43	25.10%	34	23.90%	16	17.20%	
Asian (East Asian)	77	19.00%	37	21.60%	27	19.00%	13	14.00%	
Latinx	71	17.50%	31	18.10%	26	18.30%	14	15.10%	
Education									
Some high school	3	0.80%	0	0.00%	3	2.10%	0	0.00%	.29
High school graduate/GED	71	17.80%	28	16.70%	29	20.70%	14	15.40%	
Professional degree	12	3.00%	7	4.20%	0	0.00%	5	5.50%	
Some college, non 4-year degree	103	25.80%	45	26.80%	34	24.30%	24	26.40%	
Bachelor's degree / College graduate	156	39.10%	65	38.70%	54	38.60%	37	40.70%	
Master's degree	48	12.00%	20	11.90%	18	12.90%	10	11.00%	
Doctorate	6	1.50%	3	1.80%	2	1.40%	1	1.10%	

	Full sample (*n* = 406)		**Racial Mirroring (n = 170)**		**Non-Mirroring (n = 143)**		**Control (n = 92)**		**Between-group differences**
	Mean	***SD***	***Mean***	***SD***	***Mean***	***SD***	***Mean***	***SD***	***Sig.***
Age	33.58	12.70	32.84	11.71	33.56	13.63	34.97	12.98	0.44
Household income	5.86	3.26	5.79	3.35	6.02	3.17	5.75	3.26	0.77
Life satisfaction	3.07	1.00	3.11	1.02	3.08	0.95	2.97	1.05	0.56
Social support	3.56	1.01	3.54	1.03	3.61	1.01	3.52	0.98	0.75
Privacy concern	3.20	1.07	3.23	1.15	3.24	1.00	3.07	1.02	0.45
General technology trust	3.96	0.62	3.93	0.70	4.02	0.59	3.91	0.50	0.31
General attitudes toward AI	3.38	0.86	3.34	0.89	3.42	0.85	3.39	0.84	0.69

4.2.3. Survey Analysis

I used SPSS to analyze the survey data. A variety of parametric and non-parametric tests were applied to uncover the interrelationships between variables and factors. These analyses include ANOVA, MANCOVA and multiple regression modeling, and Chi-square tests. In Chapter 5, I provide more details regarding the modeling and testing processes that I employed.

4.3. Interview Study

4.3.1. Interview Participants

From January to March 2021, I conducted in-depth interviews with16 people who previously participated in the online experiment. The recruitment process ensured that this pool of interviewees were demographically diverse in terms of age, income, sex, and race in order to account for a variety of experiences. Additionally, there were a mix of participants from different experimental groups: nine from the racial mirroring condition, four from the non-mirroring condition, and three from the control group. Interviews were conducted online via Zoom meeting, ranging from 40 to 75 minutes. Each interviewee was paid $30 for their participation.

The aim of the interview study was to provide a deeper understanding of 1) what benefits and drawbacks are associated with utilizing chatbot for mental health support, 2) how people perceive conversational agents in the context of mental health support, and 3) what features and qualities are needed to improve the chatbot design for better relational outcomes and therapeutic efficacy. The qualitative insights complement the survey data through identifying different mental models people have when interacting with a therapeutic chatbot. I also gathered feedback regarding how to improve the therapeutic chatbot design and why certain chatbot attributes and features matter. A protocol for the interview study is attached as Appendix IV.

Table 6 below shows the basic demographics information of 16 interviewees and with which agents they interacted.

Table 6. Interview participants demographics and chatbot interaction background

Participant Name	Race /Ethnicity	Sex	Age	Annual Income	Chatbot Interacted	Experimental Condition
Rana	Asian	Female	29	\$100,000 - \$149,999	Molly	Non-mirroring
Weevyn	Asian	Female	33	\$10,000 - \$19,999	Jenny	Mirroring
Alesia	Black	Female	58	\$60,000 - \$69,999	Robbie	Control
Afia	Black	Female	21	\$70,000 - \$79,999	Ebony	Mirroring
Beverly	Latino	Female	18	\$20,000 - \$29,999	Robbie	Control
Frances	Latino	Female	24	\$40,000 - \$49,999	Antonio	Mirroring
Ana	Latino	Female	23	Less than \$10,000	Camila	Mirroring
Kristy	White	Female	46	\$10,000 - \$19,999	Robbie	Control
Pamela	White	Female	32	\$20,000 - \$29,999	Jake	Mirroring
Charlie	Asian	Male	28	\$50,000 - \$59,999	David	Mirroring
Chai	Asian	Male	36	\$40,000 - \$49,999	Darnell	Non-mirroring
Eric	Black	Male	30	\$10,000 - \$19,999	Ebony	Mirroring
Connor	Black	Male	19	\$90,000 - \$99,999	Ebony	Mirroring
Steven	Latino	Male	21	\$50,000 - \$59,999	Molly	Non-mirroring
Travis	White	Male	37	\$20,000 - \$29,999	Jake	Mirroring
Jim	White	Male	70	\$50,000 - \$59,999	Antonio	Non-mirroring

4.3.2. Interview Data Analysis

To analyze the interview data in a systematic manner, I referred to the thematic analysis process (Braun & Clarke, 2012). I first transcribed the interviews; files were then imported to the NVivo software for qualitative coding. I used a combination of deductive and inductive coding. An initial codebook was created based on my interview protocol and the preliminary themes gathered from the survey around the drawbacks and benefits of chatbot therapists. I began coding using the initial codebook. As I moved through the coding process, new codes were developed and modified. The codebook incorporated the changes as new themes converged or diverged. The final codebook included both primary and secondary levels as I combined certain codes into a broader theme. Through the thematic analysis, I gathered additional qualitative insights to answer my research questions and further formulated design recommendations accordingly.

Chapter 5: Results

In this chapter, I present findings from the survey and interview studies to address my eight research questions. Section 5.1 explains the benefits and drawbacks associated with chatbot use for mental health support. Section 5.2 presents the factors influencing the sociotechnical relationship between human and chatbot. Section 5.3 describes how anthropomorphism influence the way people perceived and interacted with chatbot. Section 5.4 presents factors associated with therapeutic relationship development between human and chatbot. Section 5.5 describes the effects of the therapeutic alliance on human-agent psychotherapeutic conversations. Section 5.6 reveals whether and how anthropomorphism is associated with the sense of being understood. Section 5.7 discusses how comfortable people felt about self-disclosure and sharing their information with a chatbot. Finally, Section 5.8 presents how racial mirroring influenced people's perceptions of and interactions with a chatbot in psychotherapeutic conversations. Most research questions are answered with both qualitative and quantitative insights.

5.1. What are the benefits and drawbacks of utilizing a chatbot for managing mental health compared to a human therapist? (RQ1)

Chatbot technology promises more affordable and efficient mental health care, while falling short on providing services on par with human. My first research question seeks to further unpack user perceptions of benefits and drawbacks associated with this technology based on people's first-hand experiences interacting with MICST.

Two streams of qualitative data were used. The first stream was from the online survey, where respondents provided free-form feedback and comments about the chatbot. This dataset provided an initial framework for further exploration. The second stream of qualitative data from the interviews provided more in-depth insights. In the interview, I asked participants to compare the chatbot to other channels (e.g., friends, family, or therapist) from which they had sought social and mental health support. I also approached the interviews with two specific inquiries: Why do people adopt (or not adopt) chatbot technology for mental health support? Why do people find the chatbot therapist helpful (or unhelpful)? Accordingly, interviewees were prompted to explain what they considered the pros and cons for utilizing the chatbot technology. Further, they provided a detailed account of how and why their interactions with the chatbot were helpful (or not) for their psychological well-being.

Based on these qualitative data, I identified four major themes around the benefits and drawbacks of therapeutic chatbots, which I've labeled: (1) technology level; (2) therapeutic level; (3) social level and (4) interactional level. As I summarized patterns to differentiate various themes, there was some overlapping between observations in the sub-themes. However, the interpretations were from different angles. For example, one interviewee (Jim) mentioned that because the chatbot was not a human, he felt reluctant to engage with the technology, and he felt there was no tangible support. In this case, I coded the relevant quote in the technology level sub-theme “Lack human touch” and the social level sub-theme “No tangible support”.

5.1.1. Technology-Level Benefits and Drawbacks

From the technology point of view, several factors associated with the inherent non-human nature of chatbot help explain why or why not people adopt chatbot technology for mental health support. See Table 7 for a summary of benefits and Table 8 for a summary of drawbacks.

Table 7. Technology-level benefits

Theme	**Description**
General tech-optimism	Any general excitement and optimism about using chatbot technology
Reliability of technology	People appreciate that chatbot system is technically reliable
Objectivity / free of bias	People feel chatbot can provide unbiased information and opinions
Sense of control	Chatbot provides a sense of control for people as there’s no human pressure
Avoid social anxiety	Chatbot is considered less of a threat for people with social anxiety
The only option / Filling the void in support system	Chatbot might be the only option for some people
An intermediate step / Supplement to therapists	Chatbot technology can be a useful supplement or an intermediate step before people interact with a human therapist
Good quality of service	Chatbot has met expectation, on par with human therapists
Immediacy of service	Chatbot use can be impromptu, meeting needs immediately

Convenience	Chatbot technology is convenient, no time/space constraints, available 24/7 and COVID-proof
Affordability	People appreciate that this technology is affordable
Efficiency	People appreciate that this technology is efficient for understanding patient's problem
Anonymity / Data privacy respected	People appreciate the sense of anonymity and that their data privacy is respected
Preference for text-communication	People prefer written communication for self-expression and tracking

To highlight the most unique and interesting findings of this research, I only present details regarding the first seven technology-level benefits. As the last seven aspects were already detailed in previous research (e.g., Fitzpatrick et al., 2017; Ischen et al., 2019).

General tech-optimism. There was general excitement and optimism around the use of chatbot technology for mental health support. As P395 (male, Latinx, 18) commented, "*I really enjoyed interacting with David and really like the idea of how we can use AI to keep up with our mental health.*"

Reliability of technology. After their conversations with the agent, some participants appreciated the reliability of the MICST system. As P202 (female, Asian, 23) reported, *"I really like the idea of interacting with Robbie as a form of counseling… Robbie also seems more reliable than other AI-based apps I have interacted with before, including Amazon Alexa and Siri."*

Objectivity / Free of human bias. Compared to talking to close family or friends who might have biased opinions, some people may prefer interacting with chatbot because it's objective and free of human bias. One interviewee, Charlie, explained why he didn’t want to talk to his family and friends in order to avoid biased opinions about his relationship issue with his partner, saying:

> I want that one degree of separation to get an opinion outside a family. Because family can always give a very biased opinion (for my interest)… but sometimes I want to look for an opinion that isn't always in my best interest—maybe I'm thinking too much about me rather than about my partner.

In comparison, Charlie described chatbot as “unbiased” and “soothing” to talk to:

> Because there’s no direct relation to the chatbot and yourself, it’s like if you ever needed to talk to someone random, you can tell them and they will just listen to you. It’s unbiased, it doesn't care about who you are, where you come from, or anything like that, so it feels comfortable and soothing in that sense.

Sense of control. In a therapeutic conversation, people often need time to reflect and formulate their responses. However, due to human pressure or the mere presence of human, a few participants reported feeling rushed to give responses even if they were not ready in their psychotherapy sessions. On the other hand, when interacting with chatbot, these participants felt more control as there are no constraints in the response time. Interviewee Pamela compared interacting chatbot to human therapist:

> [One advantage of the chatbot is] being able to take your time without feeling pressure to answer, like at one point it took me the 30-45 seconds to write an answer, and the Bot didn't pop up and stare at me and say, ‘What are you thinking?’ I literally had a therapist, every 30 seconds I wouldn’t speak, she

> started to say 'what are you thinking?' I said nothing. And she's like, that can't be possible, you got to be thinking something. And it was the most annoying thing. And I refused to open up to her, because she's rushing me the moment I enter the room, and I couldn't even take a bit moment of reflection, and you definitely don't have that problem with the bot.

People also liked that their interactions with chatbot were "not forceful," and that they had the freedom to decide whether to continue the conversation or not. As interviewee Weevyn said, "[*The chatbot] doesn't feel forced, because you don't have to respond. You can exit it any time you want, you can stop interacting with the bot how you want, so it's really up to you and you keep responding."*

Avoiding social anxiety. Chatbot technology may be considered "less of a threat" for people with extreme social anxiety. A mother named Kristy posited how the chatbot might be useful for her daughter, who is on the autism spectrum:

> My daughter's on the spectrum, she suffers from agoraphobia and she has an extreme social anxiety, so I think this would be something that might help her out. She's very tech-savvy, very app-oriented, and this would be something that she could take wherever she goes. If she starts having that panic attack that none of her other things are working… she could rely on the app.

The only option / Filling the void in support system. Chatbots might be the only option for some people without any support system or access to professional help. As one interviewee (Kristy) said, "*I definitely would use the chatbot…because I just moved from Virginia to Missouri. I have yet to establish my therapist and my general practitioner*." Another interviewee (Pamela) said, "*If you would rewind two* *years from now, I definitely think I would latch on to this bot, because previously, I didn't have a support system.*"

An intermediate step / Supplement to human therapists. Seeking professional mental health help can be costly in financial and time senses. Several participants indicated that sometimes their issues might not be serious enough to justify the care-seeking efforts. In this case, chatbots can serve as an intermediate step, where people can receive adequate support and get better. As one interviewee (Beverly) said:

> I think the chatbot is definitely a good place where you talk about what bothers you in the moment and then step back and feel okay... It's a great intermediate step. I know for me, when I wanted to talk to someone, I wasn't automatically like, yes, I want a therapist now. It took me a while to be like, yeah, maybe having a therapist would be nice, so I think the chatbot would definitely be like a nice intermediate step.

Participants also considered the chatbot as a useful supplement to human therapists**.** As P383 (male, Asian, 19) said, "*I have a lot of faith in this type of technology and I am excited to see how far this advances. I personally do not think this will replace in-person therapy, but can be used as a supplement instead*."

Increase efficiency by knowing patient's problem beforehand. Chatbots can be utilized to increase the efficiency of mental health care by knowing patient's problem beforehand. One interviewee (Chai) drew an analogy between therapeutic chatbots and general chatbots for customer support, indicating that therapeutic chatbots can be used to collect basic patient information and to speed up the intake process:

> I see this technology use… (similar as the Amazon chatbot)… for customers to provide info regarding what's the problem, so we don't have to waste time digging through everything. When we reach the customer, we already know exactly what the customer wants or what the issue is.

Participants also identified technology-level drawbacks, as listed in Table 8 below.

Table 8. Technology-level drawbacks

Themes	Descriptions
General negative perception of technology	General negative perceptions of technology, including distrust, tech fatigue, and paranoia
Ethical concerns about AI	Ethical concerns about AI
Lack human touch	Chatbot is not on par with human, or there is a lack of human touch (figuratively and physically)
Unable to comprehend or recognize risks	Chatbot is unable to comprehend the full context (e.g., risk of self-harm tendency)
The risk of not seeing a real therapist after chatbot use	People might not visit a therapist after chatbot use, even if chatbot is unable to provide concrete support, especially for people with serious mental health issues
Data privacy and security concerns	Data privacy and security concerns associated with AI
No memories / history of previous conversations	Chatbot is unable to retain conversations, or have memories like humans do
Text communication inadequate for full expression	Text communication is inadequate for full expression due to "lost in translation"
Unable to read body language	Chatbot is unable to read body language

General negative perceptions of technology: tech fatigue and distrust. Several participants revealed negative perceptions of technology broadly, including fatigue associated with tech overuse and general distrust towards technology. One interviewee Rena explained why she wouldn't adopt the chatbot due to her commitment to "digital detoxing": "*I am already on my phone and computer all day long… I'm trying to rely on technology a little bit less for digital detoxing, I don't really feel comfortable being tied to my phone even more for something that should be therapeutic."* On the other hand, some participants might not use chatbot due to their general distrust toward technology. Pamela, who was positive toward the chatbot, noted that not everyone would share her opinion: *"There's literally people out there that think FBI's monitoring everything, there's a big brother. And they just might not be comfortable talking things. And they wouldn't get the help from the bot because they think someone's monitoring them."*

Ethical concerns about AI. Some participants expressed an ethical concern about AI use specifically, as P232 (male, White, 50) wrote, *"I just don't think Robbie is there yet to be able to handle the complexities of human nature. And to tell you the truth, ethically I don't think it's appropriate for AI to be ever able to do so."* This type of concern appeared to emerge from a moral sense about how AI ought to act, rather than practical considerations such as data privacy and security concerns (discussed later).

Lack human touch (figuratively and physically). Many participants simply preferred human interactions when the goal was to receive mental health support. And they did not think the chatbot would be helpful because there was a lack of human touch, figuratively and physically. As Frances commented, "*I prefer humans over the robot. Because I think our brain wants to interact human... I was talking to a robot and the picture of a robot in there... I don't know if that would be helpful for* *me.*" And Kristy said, "*I am a hugger, I want to hug someone and be hugged back when I look for some support.*"

Unable to comprehend or recognize risks. Some criticisms were specifically related to the chatbot's limitations in understanding natural language. Several participants indicated that the chatbot, at its current state, is not intelligent enough to "read between the lines" or to comprehend the full context of conversations. As a result, it might not be able to recognize any potential risks (e.g., self-harm tendency) or respond appropriately. As Pamela warned:

> The con [of the chatbot] might be its inability to predict what's appropriate to say. For example, instead of someone saying, 'I wanted to harm myself,' they say, '*my dear enjoyment was deciding to not properly hook up my harness when I go bungee jumping*.' And what if the bot says, 'that's good' because it senses the 'enjoyment' as the trigger word? So that's the risk with AI.

The risk of not seeing a real therapist. Some participants pointed out that chatbots were unable to provide concrete and high-quality support, especially for people with serious mental health issues. But due to chatbot use, people might decide not to see a human therapist because they felt it was sufficient. One interviewee (Weevyn) said, "*It might be better for people who are experiencing serious mental problems to speak with a real person... the only real con I can see is maybe someone using this chatbot and then not going to see a real professional.*"

Data privacy and security concerns. Conversation with a human therapist is HIPAA protected. However, some participants felt unsure about their data privacy when interacting with the chatbot. One interviewee (Rena) said her primary concern about the chatbot was the data privacy issue: "*Doctor interactions, whether in person or virtual, are always HIPAA protected. I don't know if that's true for these chatbots. I mean I hope so, but I'm not sure."*

No memories / history of previous conversations. Unlike a human therapist, the current version of chatbot was unable to record conversation history or retain memories of users. And there was no continuity in conversations with chatbot. One interviewee (Beverly) pointed out such disadvantage of chatbot:

> If I re-open any topic that I've previously discussed, an in-person therapist will know what I'm talking about and be able to further address it… but with chatbot, as soon as I leave the session and wanna come back, I know it will be like a restart to expressing the topic again, and building off of that, just to make my new point about the specific topic.

Text communication inadequate for full expression. Chabot interactions are solely based on text communication. Some participants indicated that this form of interaction can be inadequate for full expression of themselves. As Rena described, *"I still prefer face to face. I think a lot sometimes things can be lost in translation."*

Unable to read body language. Another drawback identified was that chatbots cannot parse body language, which is an integral part of psychological diagnosis. An interviewee (Pamela) pointed out, "*Even psychiatrists and psychologists can't give a proper diagnosis without body language. They can give an assumed diagnosis, but without that body language, you can't be completely sure."*

To summarize, there are benefits and drawbacks associated with chatbot technology for mental health support. Participants were willing to adopt chatbot for mental health support purpose because it could complement human therapist. Moreover, chatbot could offer good quality service that is immediate, efficient, reliable, unbiased, convenient, and affordable, while providing users a sense of control and anonymity. On the other hand, participants were not willing to adopt this technology because they already had tech fatigue, and they didn't trust the technology to be useful or safe without human involvement.

5.1.2. Social-Level Benefits and Drawbacks

From social point of view, the second cluster of themes explained what and why people find the chatbot helpful (or unhelpful). In this section, I first present social level benefits (see Table 9), followed by drawbacks (see Table 10).

Table 9. Social-level benefits

Themes	Descriptions
Social Support / Companionship	Chatbot fulfills social needs as a "friend in the pocket," especially for people who are alone and/or lacking social interactions
More comfortable with disclosure	People feel more comfortable expressing themselves because the chatbot provides a safe space
Free from social judgment	People feel free from social judgment
No pressure to manage self-presentation	People feel no pressure to maintain self-presentation or impression management.
No need to worry about other people's feelings	There's no need to worry about other people's reaction or feelings because chatbot doesn't have emotions.

Social Support / Companionship. Many participants found the chatbot helpful because it provided additional social support and companionship. As P306 (female, Latinx, 19) wrote, "*Even though I do have a lovely support group in real life, I found it almost refreshing to tell someone about my problems that I know for certain would give me a completely neutral yet supportive response.*" One interviewee (Kristy) considered the chatbot a "friend in the pocket" offering timely support when she couldn't access her friends or family in a moment of need.

Interacting with the chatbot could particularly benefit people who lack social interactions otherwise. One interviewee (Weevyn) described how the chatbot helped her during pandemic lockdown when she missed social interactions, *"[After lockdown due to COVID] it just became extremely strange because I wasn't interacting with anyone, it was very isolating. So when I saw the robot, I thought it was interesting. I thought it was pretty nice to interact with it."*

More comfortable with disclosure. One of the most important themes that emerged was that participants generally felt more comfortable disclosing to chatbots compared to other traditional sources of support, including friends, family, and human therapists. As Kristy indicated:

> There's just some things you don't feel comfortable talking with any of your family, not even your therapist about. But I would feel a lot more comfortable talking with chatbot. It's not something horrible but for me it is... And having something like that to discuss that particular thing when I can't talk about it to anybody else, would be such a relief.

Another participant, P171 (female, Asian, 27), echoed this sentiment that disclosures to a chatbot can be easier than that to human therapists: "*I actually felt more comfortable chatting with Jenny than some human therapists, which was interesting to me."*

For participants who were used to repressing their thoughts or feelings, disclosing to chatbot provides an opportunity for them to open up, as P243 (female, White, 38) wrote, *"Robbie really made me feel at ease opening up to him. I'm really* *surprised because usually I keep problems to myself. I think it's a great idea to have AI as therapists."*

Free from social judgment. A related theme emerged that chatbot provided a safe space where participants felt free from social judgment. The majority of interviewees attested that when interacting with chatbot, they felt they were not "being judged," and they "don't have to be too afraid or embarrassed like it's with an actual person (P128)."

One source of potential judgment is the social group to which one belongs. And chatbot interaction may be particularly important for people in a cultural community where talking about mental health issues is considered a taboo. As one Vietnamese interviewee (Charlie) described, "*Mental health in the Asian community in general, it's very antsy, looked down upon... I guess I would say it's not something that a lot of Asian people discuss. It's not a problem because it's in your mind rather than something physical."* Charlie further noted that his mother was deterred from seeking mental health support due to the fear of social judgment. He said, "*I've tried to get my mom to go to a Vietnamese-based mental health provider, but the problem for her is that she knows too many people in that circle of the Vietnamese community here in Portland, so she feels uncomfortable."*

Another source of judgment stemmed from therapists. Several interviewees had negative interactions with human therapists; for example, Rena said: "*They categorize you and try to make really quick judgments about you without even really getting to know the context*." And that made people feel uneasy, because, as Alesia noted, "*I feel like I'm like a number or a project."* Pamela even described one of her previous therapists as "*very patronizing, instead of trying to help and support, and always very blaming, even to the point of victim blaming*." On the other hand, participants felt disclosing to chatbot was much easier as P322 (female, White, 46) said,

> I would love to be able to have some type of therapy with something like Robbie. I think it would be easier for me because the AI would be NONJUDGMENTAL, and being judged is something I've had a problem with in therapy. It's why I can't find a good therapist now.

No pressure to manage self-presentation. In both social life and psychotherapy, people are likely to intentionally hold back from disclosure due to their need to maintain a certain self-image. However, when interacting with the chatbot, participants said they felt no pressure to maintain self-presentation or impression management. As one interviewee Alesia described,

> I am in a leadership role. So sometimes I feel like the expectation for me is high and I don't wanna share certain things with people because I have to protect my image, my integrity, and everything. So I think for me, it would be a real good interaction with chatbot because it's not someone that I know or someone that I have to do life with.

No need to worry about other people's feelings. Several interviewees pointed out that they didn't open up to their friends or family because they didn't want to hurt other people's feelings. In comparison, interactions with chatbot were more helpful as there was no need to worry about other people's feelings or emotional reactions. One interviewee (Beverly) recalled her conversation with the chatbot:

> I think it’s really open conversation. I really felt like I could share as much as I wanted and go as deep as I wanted, and I think it was really a space where I could just share out all my feelings and see it written out. So it was kind of just unloading what I had done, and I think that a really big benefit of this is just like I wasn’t concerned about the reaction if someone might have... Or if I voice my concerns, I wasn't like, wait, will this hurt someone’s feelings, or would they be more stressed out, or am I stressing people out? But I think it just let me be like, these are all my feelings here.

While there were social-level benefits, the non-human nature of chatbot also entailed drawbacks, as shown in Table 10.

Table 10. Social-level drawbacks

Themes	**Descriptions**
No tangible support	Chatbot support is not tangible, concrete, or real like what a human can offer
No meaningful social connection	There is no meaningful connection due to a lack of shared understanding or common experiences

No tangible support. The type of social and emotional support provided by the chatbot was not adequate for participants who desired real human interactions. For example, P289 (male, White, 23) indicated, “*I felt like it was fake. It seemed like nobody was actually there for me or listening. It was just an automated response system. I need something more meaningful.*”

Additionally, the non-human nature of chatbot deterred participants from disclosing because they didn’t expect to receive any tangible benefits in return. One interviewee (Travis) said, “*I wasn’t really looking to share anything because I know it’s not human being.”* Older participants appeared to be less open to engaging with chatbot, as P218 (female, White, 64) wrote, *“I think it’s an interesting concept. But I* *do prefer interacting with a human. When interacting I couldn't help but think that I'm not talking to a real person. Hard to envision for me."*

No meaningful social connection. Another drawback was the lack of meaningful social connection between the participant and the chatbot. Some participants stated they preferred seeking support from those with whom they had shared understanding and experiences. Additionally, the social connection built upon text-based interactions was seen as not as strong as that from face-to-face interactions. Ana described how her connection with chatbot Camilla was not as real or authentic as that with another person:

> In the group session, you can talk about the same topic with another friend, and they're like, "Oh, I felt the same way." But if I passed through this with Camilla [the chatbot], sure, she can say like, "Oh, well, we all go through that," but you feel like it's not the same. You don't feel like that close…and with friends, I feel like since it's face-to-face, you can feel more honest.

To summarize, there are social level benefits and drawbacks associated with chatbot technology for mental health support. Some participants found chatbot engagement helpful because it offered social support and companionship. More importantly, it presented a safe space where participants felt comfortable opening up because there was no social judgment, no pressure to manage self-presentation, and no need to worry about other people's reactions.

On the other hand, some participants didn't think chatbot engagement was helpful due to a lack of tangible social support and a meaningful social connection with the chatbot because it's not a human.

5.1.3. Therapeutic-Level Benefits and Drawbacks

From a therapeutic point of view, the third cluster of themes explained how helpful or not people feel about using chatbot for mental health support. In this section, I first present therapeutic level benefits (Table 11), followed by drawbacks (Table 12).

Table 11. Therapeutic-level benefits

Themes	Descriptions
Catharsis / Emotional outlet for stress relief	Chatbot offers an emotional outlet that is conducive to stress relief
The feeling of being heard / being listened to	People find interacting with chatbot helpful due to the feeling of being heard
Motivation and self-confidence	Chatbot is helpful for providing motivation and boosting people's self-confidence
Reassurance and validation	Conversations with chatbot provide reassurance that problems people face can be solved or a sense of validation that whatever emotions people feel are ok
Sense-making through self-reflection	Conversations with chatbot help people make sense of their situation and problems through self-reflection
Goal setting	Conversations with chatbot help with goal setting

Catharsis / Emotional outlet. One of the most important therapeutic benefits was that the chatbot offered participants an emotional outlet. There was often a catharsis effect in that participants felt a sense of relief after expressing themselves, as P143 (female, Black 30) wrote, *"I felt less stressed while sharing my challenges with it."*

Many participants said they enjoyed talking to the chatbot. P398 (female, Latinx, 18) even wrote, *"I wish the conversation to go a little longer because I feel like it gave me a space to let out my stress.*" The conversation with the chatbot was also described as "soothing," "alleviating," and "good for mental health," with Kristy indicating:

> There’s something I want to talk about but not with any of my family, not even with the new therapist… Once I was able to talk about it with the bot, then it’d be like having a load lifted off of me, and I wouldn’t be so anxious and maybe I wouldn’t have so many nightmares. I think it would be very good for my mental health and for my physical well-being as well.

The feeling of being heard. Merely venting would not be as therapeutically effective if the user didn’t feel the listener was actively engaged in the conversation. Another important therapeutic benefit associated with chatbot usage is the feeling of being heard. Participants described the chatbot as *“attentive”* and *“really good at listening to you.*” Beverly described how she felt she was heard, and that interacting with chatbot was more helpful than journaling because she received specific responses from the chatbot:

> I know with journaling, I can write about all my problems. But that’s it, like all my problems just sit written in front of me—that doesn’t bring closure. But when talking to the chatbot, I feel a bit better because I’ve received this specific response from the chatbot instead of just reiterating my feelings and then having no response at all.

Motivation and self-confidence. Interacting with the chatbot motivated some participants and boosted their self-confidence. P92 (male, White, 42) described how he gain more confidence as he interacted with the chatbot:

> I was suspicious at first. A lot of her [chatbot] encouragement seemed generic, but did, after a while “work.” I felt a bit more confident. I think simply

> replying to the right questions is a useful way for me to process whatever issue I am having, and possibly seeing a new solution.

Participants generally appreciated their conversation with the chatbot being motivational, as Pamela said:

> I remember the bot coming across as motivational. And that's something I definitely would need because I didn't have anyone who were there saying "You can do it," and I was constantly doubting myself around the corner... I mean, around every corner.

Participants found several chatbot responses particularly motivating, such as "people say an obstacle is often a stepping stone, don't let it discourage you." They also liked that the chatbot asked about their strengths and the existing social support they had, which could contribute to an increase in confidence.

Validation and reassurance. After expressing their feelings in the conversation with the chatbot, some participants felt a sense of validation, as Beverly said*, "Talking to this chatbot is really helpful… I think it really helps me feel okay with how I'm feeling—even if I might say something crazy or dramatic—the chatbot makes me feel a bit more validated about how I feel it."*

Participants also appreciated the assurance a chatbot offers, especially when they were told that they were not alone in facing their challenge. Steven recently had a loss in the family, and he found the chatbot helpful due to the reassurance, "*Having the reassurance that other people have gone through this and that they've gone through it, having that reassurance is what really helps me."*

Sense-making through self-reflection. Interacting with chatbot helped participants walk through and make sense of their situations, challenges, and solutions. Charlie described such utility of chatbot, "*It helps internalize and physically verbalize what it is you're thinking about, what's your problem and what you think could be potential solutions.*" Frances further explained the utility of interacting with chatbot, saying:

> [The chatbot] reminded me of worksheets that real life therapists give you. It's a way for you to process your own thoughts. I think in psychotherapy, there's five questions you need to ask yourself when you're going through a negative state of mind, like who, what, why, and where. It [the chatbot] seems to sort of roll through questions associated with that, and that helps you work through your own state of mind.

Additionally, interacting with the chatbot offered a great opportunity for self-reflection, as Charlie said, "It really gives people a way to process their thoughts, their feelings, and their concerns at their problem. So I describe it as helpful." Another interviewee, Travis, echoed this when he said, "The chatbot can be of help because it make you confront things on your own."

This benefit is closely related to the Motivational Interviewing technique, the psychotherapy method applied in the conversation. As participants identified their challenges and described how they felt, they were able to "think deeper and consider different angles." P158 (female, Asian, 28) wrote: "*It was a pleasant experience and I am surprised that it really helped me pin point my struggles/challenges where I don't think I could have realized by myself.*"

Help with goal setting. Additionally, participants found their conversation with chatbot productive and helpful because they were prompted to set a goal, as P88 (female, Latinx, 45) wrote, "*I thought it was a very productive session… I like that she was solution focused and that we set a goal for me to pursue before ending the session.*"

Although there were therapeutic benefits, several drawbacks were also identified, as described in Table 12 below.

Table 12. Therapeutic-level drawbacks

Themes	Descriptions
Not being understood	Reference to that people feel they are not being understood
Responses are too generic, not personalized	Reference to that responses chatbot provide are too generalized to be useful
Lack of empathy	Reference to that people don't think chatbot empathize with them like humans do
Limited in answers and advice	Reference to that people find the chatbot provide limited or no real advice for their situations
Inappropriate or insensitive responses	Reference to that chatbot responses come across as insensitive or inappropriate

Not feeling understood. Although the chatbot provided an outlet for people to more freely express themselves, many participants didn't feel what they said was being understood by the agent; for example, Frances observed that "*it's very clear that it doesn't understand you.*" A few participants described their interactions with the chatbot as "impersonal" and "superficial," and therefore, not useful for mental health support purposes. Jim pointed out how the chatbot fell short in understanding the context of conversation, saying: *"When I talk to a human like I'm talking to you now, if I say something to you, it's the context and the understanding that you have, that this machine doesn't have because it couldn't go that far with everything."*

Some participants expected the chatbot to "learn" and "grow" as the conversation continued. However, due to technical constraints of the current MCST system, conversation scripts were static rather than dynamic. Weevyn said, "*The responses are just all the same, so after a while, it's just not helpful, maybe the first time and the second time, then after that, not anymore. Unless it keeps growing and changing.*"

Responses are too generic. Several participants disliked that chatbot responses are very generic and not personalized, as P302 (male, Latinx, 20) wrote, *"Some of the things Darnell was saying seemed a little cliché or straight out of a motivational book."* And they didn't feel the conversation with chatbot were deep enough to help them, as participant P17 (female, Asian, 52) wrote, "*His questions were too generic and not probing enough. Just touch the surface but didn't drill down."*

Participants also expected that the chatbot could better interact with their answers. Frances criticized its lack of specificity, saying:

> It asks questions that it sort of just rolls through, but doesn't really interact with your answer It says generic things like, 'You're courageous', or 'That must be tough'…it just doesn't interact with your answer to at all, like if someone says something super negative and it answers, 'Oh, you're so courageous' – that's not helpful at all to the person.

Lack of empathy. Some participants felt there is a lack of empathy due to chatbot's generic responses. As P261 (male, Latinx, 30) wrote, "*I like the basics of talking to Antonio but I feel that it lacks nuance and warmth… It doesn't feel like a person who's validating your experiences or really sympathizing with your situation*."

P383 (male, Asian, 19) offered a similar observation, saying: "*The AI Bot is able to hold a conversation, but I feel like it is not able to empathize with the user. It does not specifically address the issue (when the user types it out) and offers general advice that can be applied to a lot of problems*."

Limited in answers and advice. When it comes to mental health support seeking, many participants expected to get specific instruction and advice for how to cope with their problem. Frances indicated her expectation, saying: "*I need you to tell me coping mechanisms, instead of just hearing me rant. Because if I wanted to rant, I could go to a friend. But in cases like mine, I want guidance as to how to deal with these things*." Participants with these expectations often found their interactions with chatbot disappointing, as P321 (female, White, 46) wrote, "*Towards the end, the AI was just deflecting and asking me to pretty much answer my own questions. There was no real advice."*

Inappropriate or insensitive responses. Some participants found the chatbot unhelpful because some of chatbot's responses came across as inappropriate or insensitive. For example, the chatbot's tone was deemed too positive for discussing a serious issue one encountered, as P321(female, White, 46) observed, *"Even though Antonio was created to be human-like, I felt his responses were a bit too positive and sometimes I thought it wasn't a serious conversation."*

At a granular level, some participants felt offended by a specific sentence the chatbot said, as Frances said, "*I remember it said something like 'a lot of people have this issue.' I don't think saying that is helpful because we don't want to be lumped with other people. So I took an issue with that phrase*."

Overall, the reasons that a chatbot was considered helpful or unhelpful were diversified at the therapeutic level. Some were related to the overall utility of the chatbot therapist, while the other focused on the psychotherapy technique and conversation design, as well as the use of language and tone.

To summarize, the therapeutic-level benefits confirm the chatbot's important role as an emotional outlet. People wanted to be heard, and the chatbot was considered a good listener. Disclosing to the chatbot could further produce catharsis or stress relief. Additionally, the conversations based on the Motivational Interviewing technique offered reassurance and validation, and contributed to an increase in people's self-confidence. Moreover, conversations with chatbot benefited people in terms of sense-making, self-reflection, and goal setting.

On the other hand, the drawbacks were primarily associated with chatbot's limitations in understanding and providing specific responses based on people's input. As a result, there was a lack of empathy. The current conversation assumed a positive tone for motivation purposes, but it might come across as inappropriate for some people.

5.1.4. Interactional-Level Benefits and Drawbacks

Focusing on conversation interface and the agent personality, interactional level benefits and drawbacks further explained how helpful or not the chatbot is for mental health support. In this section, I first present interactional level benefits (Table 13), followed by drawbacks (Table 14).

Table 13. Interactional-level benefits

Themes	**Descriptions**

Desirable agent personality	Reference to that the agent had desirable personality traits, such as fun, humor, being human-like, positivity, warmth and friendliness
Desirable use of emojis	Reference to that people appreciated the use of emojis which make agent more approachable
Efficiency of Quick Response button	Reference to that people appreciated the use of Quick Response button that make the interaction efficient
Natural, intuitive, and easy to use	Reference to that chatbot conversational interface was interactive, intuitive and easy to use

Desirable agent personality. Participants appreciated the personality of the chatbot, describing the agent as “friendly,” “congenial,” “approachable,” and “warm.” P135 (male, Black, 31) wrote, “*He (agent David) is active and cheerful, nice and kind, thoughtful and lovely*.” And P285 (female, Latinx, 24) commented, “*It was different but still felt like a genuine conversation.*”

Participants also appreciated that the chatbot asking directed and purposeful questions. Rena described her impressions of the chatbot as *“warm, directed, and specific”* because the chatbot asked *“some pretty good specific questions”* that gave her directions.

Additionally, the appearance of the agent invoked positive feelings, as Afia described, “I think Ebony is kind, I like the avatar is a beautiful chocolate woman, also I think she has kind eyes. The design is very nice, also just how it looks.”

Desirable use of emojis. The use of emoji gave the chatbot a sense of human touch, as P207 (male, Latinx, 22) wrote, *“I like the fact that she uses emojis, it gives the whole thing a touch.”* Due to the popularity of emoji in participants’ daily life, chatbot using emojis could appear more familiar and less robot-like, as Pamela commented, “*I definitely like the addition of emoticons, I think that would be very valid in today’s society, because even I have been into using an emoticon after almost* *every sentence."* Ana also indicated, "*I really like the way it's a little bit more familiar and that it uses emojis, which gives it a lot of flairs."*

Efficiency of Quick Response button. The Quick Response buttons allow participants to select a reply out of pre-set options. Some participants liked using it for efficient conversation transitions. One button was considered particularly useful as it linked to a brief explanation of how the chatbot works as well as its purpose. Kristy said, "*If I didn't understand things, I can go ahead and click the 'how does this work?' I'll just click the button there. I can totally see my dad interacting with it because he's not tech-savvy."*

Natural, intuitive, and easy to use. Many participants appreciated that the conversational interface was natural, intuitive, and easy to use, as P78 (male, Black, 49) wrote, "*I thought Robbie was intuitive and responsive. It was very easy to use."*

Additionally, the chatbot was able to provide "on-point" response that impressed some participants. One interviewee Alesia recalled her conversation with Robbie, *"One thing that was noticeable to me and surprising is that his responses were pretty on-point and accurate or for relationship and interaction. It wasn't like, 'Oh yeah, this is a robot.' I was like, 'wow, that seemed pretty human, almost.' That was pretty cool."*

The conversations were also described as "interactive" and "smooth." Steven said, *"One of the answers impressed me: 'It's good to have friends and family to support you and help you.' With its wording and all that… it was just like a good smooth transition to the conversation*."

Several interactional level drawbacks were also identified, as shown in Table 14 below.

Table 14. Interactional-level drawbacks

Themes	Descriptions
Topic ends abruptly	Reference to that conversation with chatbot ended abruptly while people had more to say
Ask repetitive questions	Reference to that chatbot asked repetitive questions.
Unnatural conversation	Reference to that people felt the conversation was unnatural.
Not intuitive to use Quick Response buttons	Reference to the critique that it was not intuitive to engage with the quick response buttons
Response speed too fast	Reference to the criticism that the speed of chatbot's response was too fast and it felt unnatural
Overuse of emojis	Reference to that some people felt emojis were overused
Unrealistic profile picture	Reference to that people considered the profile picture unrealistic

Topic ends abruptly. Because the conversation was based on static scripts, a few participants felt some parts of their exchange were cut short. For example, Ana said:

> Sometimes it just goes a little bit short. She will ask you one thing and then you answer it, and then instead of continue talking about that topic, she just changes it to a new one. This is something that you can perfect, maybe when I have more questions, she just goes, 'do you still want to talk to me, or do you wanna end the session' in case I have more to say.

Ask repetitive questions. The chatbot conversation scripts included three questions to probe participants' emotional responses to their challenges or problem. Some participants disliked being asked repetitively, with Ana saying, "*What was a bit weird was that sometimes it will repeat the same questions and you already answered those questions, and also that sometimes you explain the problem and then they ask you again.*"

Unnatural conversation. Although the conversational interface promises a more natural way of interactions, the current iteration of the MCST system came across “canned” and “scripted.” P177 (male, Latinx, 26) wrote, “*She seems to be pre-programmed to respond a certain way regardless of what you say. It takes away from the organic-ness of the conversation*.”

Not intuitive to use Quick Response buttons. Some participants felt the use of Quick Response buttons was counter-productive, as P315 (female, Asian, 18) observed, “*Interacting with Robbie felt extremely limiting. The fact that you couldn't freely talk to him and instead had to click on little bubbles, especially when we had to select our issue made the entire interaction much less natural.”*

Agent responses were too fast. The speed of responding also affected how participants perceived the chatbot. When response speed was too fast, the chatbot appeared less human-like, as P113 (male, Latinx, 30) said, “*I felt like the AI (Camila) responded too quickly. If I was really using this technology for therapy, I would want a slower response time. It would make it feel more like I was talking to a human.”*

Overuse of emojis. Some participants disliked the use of emojis as it appeared “juvenile,” as P272 (male, Black, 26) said, “*I felt that the use of emojis was a little unnecessary as people could be disclosing something serious and the emojis might be interpreted as odd or flippant…They seemed meant for a younger age or to be more personable, but also felt a little juvenile.*”

Unrealistic profile picture. Some participants also criticized that the agent’s profile picture appeared unrealistic. As P171 (male, Asian, 27) said, “*I think it would* *help if Jenny had a more realistic human-like face, so it's not like I'm talking to a cartoon*."

To summarize, the interactional-level benefits and drawbacks center on the conversational design and the chatbot persona. The most important benefit is the chatbot's personality. Participants appreciated that the agents had desirable personality traits, such as fun, humor, being human-like, positivity, warmth, and friendliness. They felt their interaction with the chatbot natural and smooth as carrying a conversation with human. They also liked chatbot use of emojis and the Quick Response buttons which made the conversation even more fluidly.

On the other hand, some participants felt the overall conversation was unnatural and not organic. They disliked that a topic ended abruptly or when questions were repetitive. The inclusion of emoji and Quick Response buttons in the conversation was also deemed unnatural. And lastly, some participants felt the agent's profile picture appeared unrealistic.

5.2. What factors influence the sociotechnical relationship between human and chatbot for mental health support? (RQ2)

As discussed in the last section, chatbot technology presents a valuable opportunity to increase the efficiency and accessibility of mental health care. To fully realize the potential of chatbot technology, the first step is to enhance the sociotechnical relationship between human and chatbot. Therefore, RQ2 explores various factors that shape this relationship. In this section, I first explain the statistical modeling procedure to analyze the survey data. Next, Section 5.2.1 presents modeling results that uncover how different variables predict the outcome variables measuring sociotechnical relationship strength. Section 5.2.2 presents additional qualitative insights based on follow-up interviews.

The study measured the sociotechnical relationship between human participants and the chatbot using two dependent variables: 1) likelihood to recommend and 2) willingness to continue engaging with the agent. Both variables were measured on a scale from 1 (would not recommend /not willing at all) to 100 (definitely recommend/very much willing). Respondents generally inclined to recommend this technology, with the median score of 61 and the average score of 59.26 (*SD* = 30.58). Similarly, respondents generally were willing to continue engaging with the agent for mental health support, with the median score of 64 and the average score of 62.82 (*SD* = 30.40) .

To address RQ2, I performed a multivariate analysis of covariance (MANCOVA), using a general linear model, to assess various factors predicting the strength of human-agent sociotechnical relationship. As an extension of regression analyses and analysis of variance, MANCOVA represents a powerful quantitative method to improve group comparisons when multiple outcomes are involved. A MANCOVA model assesses whether multiple dependent/outcome variables (continuous) are predicted by independent variables (categorical) when controlling for the effect of other confounding variables or covariates (continuous).

Several important assumptions should be tested when using MANCOVA: 1) independence of observations, 2) a linear relationship between DVs, 3) a linear relationship between the covariate and each DVs, 4) absence of multicollinearity, 5) homogeneity of regression slopes among groups, 6) homogeneity of variances and covariances, and 7) normality of variable distribution (Pepe & Addimando, 2014).

As the first step of the analysis, I checked the assumptions by calculating the means, standard deviations, skewness, and kurtosis of all variables to be included in the model. In the second step, I constructed the MANCOVA model through a backward step-wise selection process (Seltman, 2018). I started with a full, complicated model where I incorporated all measures and control variables, along with three categorical variables (i.e., conditions, agent gender, and participant gender). I proceeded with modeling by removing the terms with the largest p-value above the set alpha (.05). I also removed covariates that were not related linearly with the DVs (assumption #3). For example, age and education were not correlated with the DVs, and therefore, were removed. Section 5.2.1 below presents descriptive statistics and results from the final MANCOVA model.

5.2.1. Model Results

Table 15 below reports the descriptive statistics and correlations for variables in the final MANCOVA model for RQ2. All assumptions of MANCOVA were satisfied. All variables included in the final model were positively correlated with each other. The majority of correlation coefficients (r) were below 0.8 and none exceeded 0.9, so multicollinearity was not an issue (Senaviratna & Cooray, 2019). Additionally, their distribution values, skewness, and kurtosis scores, were within the commonly recommended cutoff point of |2| (Hopkins & Weeks, 1990). Most of the variables fairly resembled a normal distribution, with the skewness score falling between -0.5 and 0.5.

Table 15. Descriptive statistics and correlations for variables in model (RQ2)

Variable	***Mean***	***SD***	**Skew-ness**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
1. Likelihood to recommend	9.26	0.58	.41									
2. Willingness to continue interacting	2.82	0.40	.53	90**								
3. General attitudes toward AI	.38	.86	.42	66**	67**							
4. Comfort with data sharing	0.87	2.90	28	34**	35**	34**						
5. Fondness of agent	3.48	3.73	.97	72**	77**	39**	33**	—				
6.Anthropomorphism	5.72	8.96	.35	79**	79**	57**	31**	80**				
7. Ease of use	.81	.75	.89	56**	57**	50**	26**	59**	65**			
8. Effectiveness	.34	.00	.49	85**	84**	70**	28**	74**	75**	60**		
9. Social influence	.43	.72	.29	65**	66**	67**	37**	53**	57**	52**	68**	
10. Projected future relationship	.39	.76	25	59**	60**	54**	35**	52**	55**	35**	58**	50**

Note: * ($P \le 0.05$) ; **($P \le 0.01$); *** ($P \le 0.001$)

As shown in Table 16, the model reveals several significant variables predicting the strength of sociotechnical relationship. These include constructs found in the traditional technology acceptance theories (e.g., UTAUT) and chatbot-related perceptions and attitudes, as well as agent persona. These are discussed in more detail below. Table 17 shows the effects of covariates on the outcome variables; Table 18 shows the effects of agent gender on the outcome variables.

Table 16. Significance levels of the multivariate test
(DVs: Likelihood to recommend, Willingness to continue engaging with the agent)

	Wilks' lambda	**F**	**Sig.**	**Partial eta squared η2**
Intercept	.77	54.19	.00	.23
General attitudes toward AI	.99	1.74	.12	.01
Comfort with data sharing	.98	3.79	.02	.02
Fondness of agent	.96	8.51	.00	.05
Anthropomorphism	.86	29.69	.00	.14
Ease of use	.99	1.04	.35	.01
Perceived usefulness / Efficacy	.67	88.39	.00	.33
Social influence	.98	4.31	.01	.02

Projected future relationship	.98	3.3	.03	.02
Agent gender	.99	2.03	.01	.03
Participant gender	.98	.53	.71	.01
Racial mirroring condition	.99	.21	.86	.00

Table 17. Effects of covariates on sociotechnical relationship

Dependent Variables	**Covariate**	**β**	**SE**	**t**	**Sig.**	**Partial eta squared η2**	**Observed power**
Likelihood to recommend	General attitudes toward AI	.25	.30	.19	.85	.00	.05
	Comfort with data sharing	**.07***	**.04**	**.01**	**.05**	**.01**	**.52**
	Fondness of agent	-.03	.06	.53	.60	.00	.08
	Anthropomorphism	**.54*****	**.07**	**.39**	**.00**	**.13**	**1.00**
	Ease of use	-2.43	1.41	1.72	.09	.01	.40
	Perceived usefulness / Efficacy	**16.53*****	**.38**	**1.94**	**.00**	**.28**	**1.00**
	Social influence	**3.58***	**.48**	**.43**	**.02**	**.02**	**.68**
	Projected future relationship	**1.18***	**.55**	**.15**	**.03**	**.01**	**.57**
Willingness to continue engaging with the agent	General attitudes toward AI	1.52	.32	.15	.25	.00	.21
	Comfort with data sharing	**.07***	**.04**	**.05**	**.04**	**.01**	**.53**
	Fondness of agent	**.18*****	**.06**	**.17**	**.00**	**.03**	**.89**
	Anthropomorphism	**.43*****	**.07**	**.77**	**.00**	**.08**	**1.00**
	Ease of use	-1.92	.43	1.34	.18	.01	.27
	Perceived usefulness / Efficacy	**12.74*****	**.40**	**.07**	**.00**	**.19**	**1.00**
	Social influence	**3.15***	**.50**	**.10**	**.04**	**.01**	**.55**
	Projected future relationship	**1.20***	**.56**	**.16**	**.03**	**.01**	**.58**

Note: * ($P \leq 0.05$) ; **($P \leq 0.01$); *** ($P \leq 0.001$)

Traditional UTAUT Constructs. Perceived usefulness or efficacy of the chatbot positively correlated with the strength of human-agent sociotechnical relationships. When believing the technology to be more useful/effective for mental health support, participants were more likely to recommend it to their friends or family ($\beta = 17.38$, $p < .001$) and were more willing to continue engaging with the agent ($\beta = 13.48$, $p < .001$) for mental health support.

Social influence also played a role in determining sociotechnical relationship formed between human and agent. When perceiving the use of this technology to be socially approved (e.g., friends or family would not judge it), participants were more likely to recommend it ($\beta = 3.58$, $p < .05$) and were more willing to continue engaging with the agent ($\beta = 3.15$, $p < .05$) for mental health support.

Contrary to what UTAUT would suggest, the effect of perceived ease of use was not statistically significant for either DV. Notably, most participants (64.2%) rated the chatbot easy or very easy to use ($M = 3.81$, $SD = .75$). This factor was skewed in a way to produce a “ceiling effect” (A. D. Ho & Yu, 2015), which could have reduced the power of statistics on correlation between perceived ease of use and DVs.

Chatbot-related Perceptions and Attitudes. Anthropomorphism positively correlated with the strength of human-agent sociotechnical relationship. When perceiving the agent to be more human-like, participants were more likely to recommend ($\beta = .54$, $p < .001$) and were more willing to continue engaging with the agent ($\beta = .43$, $p < .001$) for mental health support.

Comfort with data sharing, specifically when using this chatbot technology, also positively correlated with the sociotechnical relationship. Participants who reported a higher level of comfort were more likely to recommend this technology (β

= 3.58, $p < .05$) and were more willing to continue engaging with the agent (β = 3.15, $p < .05$) for mental health support.

Fondness of agent positively correlated with how willing participants were to continue engaging with the agent ($\beta = .18$, $p < .001$), but was unrelated to one's likelihood of recommending this technology to their friends or family ($p = .60$).

Additionally, projected future relationship positively affected the strength of sociotechnical relationships. Participants who predicted that they would have a closer relationship with the agent were more likely to recommend this technology (β = 1.18, $p < .05$) and were more willing to continue engaging with the agent (β = 1.20, $p < .05$) for mental health support.

General attitudes toward AI was not related to the strength of sociotechnical relationships.

Agent Persona. The gender of agents affected participants' willingness to continue engaging with chatbot for mental health support. Female agents received statistically higher scores for willingness to continue engaging (M = 62.49, SD = 1.17), compared to male agents (M = 58.22, SD = 1.15). However, no gender preference was found when recommending this technology to others. See Table 18 for model details.

Table 18. Effects of chatbot gender on sociotechnical relationship

DVs	**Robotic Agent**		**Male Agent**		**Female Agent**		**$F(2, 365)$**	**η^2**
	M	*SD*	*M*	*SD*	*M*	*SD*		
Likelihood to recommend	56.66	1.47	55.79	1.17	56.62	1.15	1.87	.04
Willingness to continue engaging	60.30	1.50	58.22	1.18	62.49	1.17	8.56^{*}	.17

Note: * ($P \leq 0.05$) ; **($P \leq 0.01$); *** ($P \leq 0.001$)

5.2.2. Results from Interviews (RQ2)

In this section, I present additional qualitative insights gathered from the follow-up interviews to complement the survey findings.

Among UTAUT constructs, qualitative data confirmed the importance of Perceived usefulness / Efficacy of the chatbot. For example, Rena indicated she would still prefer human therapist because "*the chatbot is still very rudimentary. They're not going to help you in a very specific and personalized way*."

Although not included in the survey, another UTAUT factor emerged from the qualitative data: the perceived cost. The affordability of chatbot technology appeared attractive to participants who were cost-sensitive. For example, Frances commented, "*I think a technology where you're talking to something that's automated and is not very costly and can reach a lot of people, I think that's important. It's also cool."*

With regard to chatbot-related perceptions and attitudes, qualitative data further confirmed that anthropomorphism was critical for participants' decision to adopt chatbot for mental health support. Participants wouldn't be interested in the technology if they didn't sense a chatbot to be human or human-like. Jim had problem accepting this technology because , "*I know how they put that guy [chatbot] in the screen with the code... When I'm looking at here [chatbot], I'm not looking at a person or a human, I'm looking at codes. So when I talk to this person, it was not somebody that I could relate to… it*'*s programming. I could talk if I didn*'*t know any better."*

Comfort with self-disclosure was another important factor that emerged from qualitative data, although it not significant in the statistical modeling. Alesia explained she definitely wanted to continue interacting with the chatbot because she felt very comfortable expressing her feelings with it. In contrast, she didn’t feel comfortable with a human therapist due the feeling of being “analyzed” and “categorized,” and she didn’t want to share with her family and friends because she wanted to protect her image.

Finally, qualitative data substantiated that agent persona influenced sociotechnical relationships formed. 11 out of 16 interviewees had a preference; seven out of the 11 interviewees indicated that they preferred to interact with female agents. Connor said he preferred a female agent because “*usually there’s a lot of female therapists*” and he felt “*women are more understanding.”* Pamela also preferred females because she felt there were more shared experiences and implicit understanding.

5.2.3. Summary (RQ2)

In exploring the sociotechnical relationship forged between human and chatbot, I applied the traditional technology acceptance model UTAUT as a theoretical foundation. The analysis confirmed the importance of Perceived Usefulness and Social Influence in driving the adoption and continued engagement with the chatbot technology. However, Perceived Ease of Use was not a significant predictor in the model. One likely explanation is that most respondents rated the technology easy to use, resulting in a “ceiling effect” that diminished the power of statistics for correlation. From the user experience point of view, participants generally found the conversational interface natural and intuitive to interact with;

therefore, ease of use didn’t turn out to be a differentiating factor for technology use and engagement.

Going beyond UTAUT constructs, my analysis also uncovered several chatbot-specific factors that affect the sociotechnical relationship. Anthropomorphism played a unique role; participants were more compelled to engage with chatbots they deemed more human-like. Additionally, as sensitive information was involved in a psychotherapeutic conversation, participants’ comfort with data sharing also affects how likely they are to adopt such technology for mental health support. Unsurprisingly, interpersonal variables, including fondness of agent and projected future relationship, positively correlated with sociotechnical relationship. Additional insights based on qualitative data also suggested that comfort with self-disclosure contributed to the strengthening of sociotechnical relationship. Finally, how the agent personas were designed could also shape the strengths of the sociotechnical relationship. Female agents were considerably more accepted and preferred to male agents.

5.3. How does anthropomorphism influence the way people perceive and interact with a chatbot in psychotherapeutic conversations? (RQ3)

Applying the Godspeed instrument (Bartneck et al., 2008), the survey evaluated four dimensions of anthropomorphism[17], or the tendency to attribute human characteristics to chatbots. Participants rated their impressions of the agent using

---

[17] For other RQs, I used an aggregated scale of anthropomorphism by averaging all four dimensions. For RQ3, I examined different dimensions of anthropomorphism and their effects.

linguistic scales; responses were coded from 1 to 100 and each end of the scale was labeled with a pair of terms (e.g., fake/natural, unkind/kind).

There were four sets of word pairs corresponding for each dimension of the instrument: 1) overall anthropomorphism (*M*=56.59, *SD*=1.34, *α*=.96), 2) Liveliness (*M*=61.55, *SD*=1.15, α=.92), 3) Likeability (*M*=80.65, *SD*= .79, α=.90), and 4) Perceived Intelligence (*M*=70.82, *SD*= .96, α=.95). Table 19 presents average scores for each

Table 19. Descriptive statistics of four dimensions of anthropomorphism
(Scale from 1 to 100)

Dimensions	**Linguistic Scales / Word pairs**	**Mean**	**SD**
Overall anthropomorphism	Machine-like – Human-like	54.79	8.76
	Fake – Natural	55.51	7.73
	Unconscious – Conscious	55.07	8.24
	Artificial – Lifelike	50.14	0.06
Liveliness	Communicated awkwardly – Communicated naturally	62.92	25.36
	Inanimate – Alive	54.79	27.69
	Mechanical – Organic	51.25	27.05
	Inert – Interactive	67.62	24.89
Likeability	Apathetic – Responsive	71.24	22.38
	Unfriendly – Friendly	82.02	16.66
	Unkind – Kind	81.00	17.07
	Unpleasant – Pleasant	79.04	18.70
	Awful – Nice	81.06	17.61
Intelligence	Ignorant – Knowledgeable	64.73	21.87
	Unintelligent – Intelligent	67.15	21.82
	Incompetent – Competent	68.15	21.88
	Foolish – Sensible	71.83	19.98
	Irresponsible – Responsible	73.45	18.93

5.3.1. Results (RQ3)

To explore how anthropomorphism might influence the way people perceive and interact with the chatbot, I built four regression models[18] using different dimensions of anthropomorphism as DVs. Table 20 presents detailed results. The

---

[18] I chose regression instead of MANCOVA because I considered anthropomorphism as one general outcome with four different dimensions. MANCOVA is only appropriate when more than one outcome variables were involved.

models substantiated that anthropomorphism had downstream impacts on several aspects of human-chatbot interactions.

Chatbot-related Perceptions. Participants became more fond of the chatbot persona and considered this technology easier to use, when attributing higher degrees of anthropomorphism, liveliness, likeability, or intelligence to the agent. However, perceived usefulness or efficacy was not related to any dimensions of anthropomorphism.

Human-agent Conversations. Anthropomorphism also appeared to influence human-agent conversations. Interestingly, the effects of various dimensions are different. Participants felt less comfortable with self-disclosure when they attributed higher degrees of anthropomorphism to the agent. On the other hand, they felt more comfortable with self-disclosure when attributing a higher degree of likeability to the agent. Meanwhile, participants reported that their conversation with the agent was more engaging when attributing higher degrees of anthropomorphism and liveliness.

Sociotechnical Relationships. Moreover, anthropomorphism might strengthen sociotechnical relationships. Participants were more likely to recommend and more willing to continue engaging with this technology when reporting a higher degree of either anthropomorphism dimension.

Therapeutic Relationships. Finally, findings suggest anthropomorphism might enhance therapeutic relationships. Participants projected a stronger future relationship when they attributed higher degrees of anthropomorphism.

Table 20. Models explaining four dimensions of anthropomorphism

	DV=Anthropomorphism	**DV= Liveliness**	**DV= Likeability**	**DV= Intelligence**

	β	***SE***	**β**	**SE**	**β**	**SE**	**β**	**SE**
Chatbot-related Perceptions								
Fondness of agent	**.26*****	**.06**	**.14****	**.05**	**.25*****	**.05**	**.18*****	**.05**
Perceived ease of use	**3.21***	**1.41**	**6.24*****	**1.20**	**5.06*****	**.08**	**4.95****	**1.01**
Perceived usefulness	.52	.77	-.68	.50	-2.08	.35	.29	.27
Conversation-related Variables								
Comfort with self-disclosure	**4.64****	**.75**	-1.93	.49	**3.33****	**.35**	.18	.26
Comfort with data sharing	.01	.04	-.01	.03	-.04	.03	-.03	.03
Level of engagement	**.14****	**.05**	**.14*****	**.04**	.02	.04	.01	.04
Content sensitivity	.05	.03	.04	.03	-.02	.03	.01	.02
Sociotechnical Relationships								
Likelihood to recommend	**.21*****	**.07**	**.13***	**.06**	**.11***	**.05**	**.21*****	**.05**
Willingness to continue engaging	**.18****	**.07**	**.18*****	**.06**	.01	.05	**.08****	**.05**
Therapeutic Relationships								
Projected future relationship	**1.47***	**.61**	**.99***	**.52**	**.94***	**.46**	.65	.44
Perceived interpersonal closeness	1.33	.97	1.20	.82	-1.08	.74	.30	.70
Control Variables								
Age	**.27*****	**.06**	**.27*****	**.05**	**.14*****	**.05**	**.07**	**.04**
Social support	**1.77***	**.82**	**1.91****	**.70**	**2.91*****	**.63**	**2.13*****	**.59**
Privacy concerns	.46	.79	-.85	.67	-.47	.60	-.75	.57
Racial mirroring group	**3.01***	**2.11**	2.27	1.79	1.62	1.61	2.26	1.51
Non-mirroring group	**4.85***	**2.02**	4.33	1.72	1.17	1.54	.57	1.45
Control group	-[a]	-[a]	-[a]	-[a]	-[a]	-[a]	-[a]	-[a]
Model fit	F(15, 349) = 59.21***		F(15, 348) = 59.97***		F(15, 347) = 26.35***		F(15, 348) = 56.74***	
Adjusted R^2 =	.71		.72		.52		.71	

Note: a. Value not applicable due to reference level
b. * ($P \leq 0.05$) ; **($P \leq 0.01$); *** ($P \leq 0.001$)

5.3.2. Results from Interviews (RQ3)

The qualitative data provided additional evidence of how anthropomorphism affected participants' perceptions of and interaction with the chatbot. First, it appeared to be a foundation for forming any kind of relationship between human and chatbot. Chai described how the lack of realism in the profile picture created an initial barrier to connecting with the chatbot: "*You see a face of the robot or something like* *that, or it's the face of a human with cartoon… I felt I wasn't able to have a relationship because I knew that it's a chatbot, so already there was a barrier in my mind.*"

But Chai's perceptions of the chatbot shifted once the conversation kicked off. As the chatbot was able to carry a conversation like a human, Chai felt the chatbot was "alive" and "made the connection right away":

> It wasn't like another robot or any other cartoon face. It was there to guide you through the conversation. I think that made the connection right away. When I first started interacting with it, what drew me to having the connection was that it was able to ask me questions and lead a conversation… it makes you feel like you're a part of the conversation. When you're talking to it, it's alive.

Moreover, anthropomorphism also affected how participants interacted with the chatbot. Participants generally felt more comfortable disclosing to the chatbot than to a human. However, some participants were less likely to engage in deep conversation because they didn't expect a chatbot to understand what they said. As a result, they didn't try to have a deep conversation, as Jim said, "*The conversation is shallow because it's limited in his answers and limited by the way that I say something*." Additionally, knowing that the chatbot was not a human, some participants changed from not willing to disclose to being open about themselves. Travis said, *"At first, I wasn't really looking to share anything because I know it's not human… but then it actually made me more open to sharing something because I knew it wasn't a human being*." As explained in Section 5.1, participants' comfort with self-disclosure to the chatbot was largely associated with the notion that they were not judged by the chatbot. For example, Kristy said, "*I can talk to Robbie without having to worry about him judging whether I'm handling this in a horrible way or not.*"

Some participants intentionally used simple phrases or words when interacting with the chatbot because they anticipated that chatbot wouldn't have the same level of comprehension as humans. For example, Afia said, *"When I went through it the first time, I took a lot of time trying to phrase it in a way that I guess I thought the computer might understand."*

5.3.3. Summary (RQ3)

Generally, anthropomorphism is associated with positive perceptions of the agent, strengthened sociotechnical relationships, and a better outlook for therapeutic relationships formed between humans and agents.

That said, humans might attribute various characteristics to conversational agents. And different dimensions of anthropomorphism turned out to have the opposite effects. On one hand, comfort with self-disclosure was positively correlated with likeability; participants felt more comfortable disclosing to agents perceived as more responsive, friendly, kind, pleasant, and nice. On the other hand, comfort with self-disclosure was negatively correlated with general anthropomorphism; participants felt less comfortable disclosing to agents perceived as more human-like, natural, conscious, and lifelike. This finding aligns with previous studies suggesting that people reveal their deep, personal issues more truthfully to a computer than they do to human counselors (Meier, 1988).

The qualitative data provided additional evidence that participants felt more comfortable because they believed a chatbot—being non-human—would not judge them. As a result, participants were able to disclose more sensitive thoughts and feelings to the chatbot. Additionally, anthropomorphism also affected the depth and complexity of participants' disclosure. Due to the perception that the chatbot might not be intelligent enough, some participants might be deterred from engaging in deep conversation and they intentionally chose simple words to use in the conversations.

5.4. What factors are associated with therapeutic relationship development between client and conversational agents? (RQ4)

The therapeutic relationship centers on people's perceptions and feelings about the chatbot as a therapist. It is important to understand how an interpersonal and social connection can be developed. In this research, three variables were used to measure different forms of therapeutic relationships: 1) working alliance, 2) perceived interpersonal closeness, and 3) projected future relationship.

Analysis of the survey data provided strong support for the belief that therapeutic relationships can be formed in chatbot-based therapy. More than one-third (n=145; 35.7%) of participants agreed or strongly agreed that there was a strong therapeutic alliance; 179 participants (43.2%) reported a moderately close, very close, or fully close relationship between themselves and the agent they interacted with; and 191 participants (46%) projected a fairly strong, strong, or very strong connection with the agent.

I conducted MANCOVA to further unpack factors associated with human-agent therapeutic relationships using these three factors as DVs. I followed the same process to test model assumptions and used the backward model selection process detailed in Section 5.2.1. Model results are presented below.

5.4.1. Results (RQ4)

As shown in Table 21, the model reveals several significant variables predicting the strength of therapeutic relationship. These include participant characteristics (e.g., life satisfaction and social support) and chatbot-related perceptions and attitudes, as well as agent persona. These are discussed in more detail below. Table 22 shows the effects of covariates on the outcome variables; Table 23 shows the effects of racial mirroring condition on the outcome variables.

Table 21. Significance levels of the multivariate test
(DVs: working alliance, perceived interpersonal closeness, projected future relationship)

	Wilks' lambda	**F**	**Sig.**	**Partial eta squared η2**
Intercept	.96	5.27	.00	.04
Life satisfaction	.96	5.04	.00	.04
Social support	.96	4.72	.00	.04
General attitudes toward AI	.98	2.56	.06	.02
Fondness of agent	.95	6.13	.00	.05
Anthropomorphism	.95	5.96	.00	.05
Desirability	.87	17.57	.00	.13
Perceived usefulness / Efficacy	.80	29.80	.00	.20
Social influence	.95	5.85	.00	.05
Comfort with self-disclosure	.99	1.50	.21	.01
Agent gender	.98	2.12	.10	.02
Participant gender	1.00	.49	.69	.00
Racial mirroring condition	.99	.85	.02	.03

Several variables had statistical significant effects on the strength of therapeutic relationship between participant and agent, as shown in Table 22.

Table 22. Effects of covariates on therapeutic relationships

Dependent Variables	**Covariate**	**β**	**SE**	**t**	**Sig.**	**Partial eta**	**Observed power**

						squared η2	
Working alliance	Life satisfaction	.05	.02	-2.18	.63	.04	.95
	Social support	.04	.02	1.53	.13	.01	.58
	General attitudes toward AI	-.03	.04	-.71	.48	.01	.33
	Fondness of agent	.00	.00	.89	.38	.00	.11
	Anthropomorphism	**.01***	**.00**	**1.98**	**.05**	**.00**	**.14**
	Desirability	**.18*****	**.02**	**7.24**	**.00**	**.01**	**.51**
	Perceived usefulness / Efficacy	**.32*****	**.04**	**8.14**	**.00**	**.13**	**1.00**
	Social influence	**.16*****	**.04**	**3.80**	**.00**	**.16**	**1.00**
	Comfort with self-disclosure	**.09***	**.05**	**2.08**	**.04**	**.04**	**.97**
	Comfort with data sharing	.00	.00	-.70	.48	.01	.55
Perceived interpersonal closeness	**Life satisfaction**	**.17*****	**.05**	**3.08**	**.00**	**.03**	**.87**
	Social support	**-.13***	**.05**	**-2.41**	**.02**	**.02**	**.67**
	General attitudes toward AI	.02	.08	.28	.78	.00	.06
	Fondness of agent	**.02*****	**.00**	**4.15**	**.00**	**.05**	**.99**
	Anthropomorphism	**.01***	**.00**	**2.78**	**.01**	**.02**	**.79**
	Desirability	-.01	.06	-.12	.90	.00	.05
	Perceived usefulness / Efficacy	**.40*****	**.09**	**4.56**	**.00**	**.06**	**1.00**
	Social influence	-.09	.09	-.94	.35	.00	.16
	Comfort with self-disclosure	.06	.10	.56	.58	.00	.09
	Comfort with data sharing	.00	.00	1.19	.24	.00	.22
Projected future relationship	Life satisfaction	.04	.08	.47	.64	.00	.08
	Social support	**-.24*****	**.08**	**-3.05**	**.00**	**.03**	**.86**
	General attitudes toward AI	**.33***	**.12**	**2.65**	**.01**	**.02**	**.75**
	Fondness of agent	.00	.01	.35	.73	.00	.06
	Anthropomorphism	**.02*****	**.01**	**3.27**	**.00**	**.03**	**.90**
	Desirability	-.07	.08	-.85	.40	.00	.13
	Perceived usefulness / Efficacy	**.47*****	**.14**	**3.51**	**.00**	**.03**	**.94**
	Social influence	.13	.14	.91	.36	.00	.15
	Comfort with self-disclosure	.02	.15	.15	.88	.00	.05
	Comfort with data sharing	.01*	.00	2.39	.02	.02	.66

Note: * ($P \leq 0.05$) ; **($P \leq 0.01$); *** ($P \leq 0.001$)

Participant characteristics. When participants had more social support from their friends or family, they reported a lower level of interpersonal closeness with the chatbot (β= -.13, p < .05) and weaker perceived future relationship (β= -.23, p < .001). On the other hand, when participants were more satisfied with their life, they tended to form a stronger interpersonal relationship with the agent (β= .17, p < .001).

Chatbot-related Perceptions and Attitudes. Anthropomorphism positively correlated with all three measures of therapeutic relationships. When participants perceived the agent to be more human-like, they reported stronger working alliance (β= .01, p < .05), interpersonal closeness (β= .01, p < .01), and projected future relationship (β= .02, p < .001).

Perceived usefulness or efficacy of the chatbot positively correlated with working alliance (β= .32, p < .001), interpersonal closeness (β= .40, p < .001), and projected future relationship (β= .47, p < .001). Another classic UTAUT construct, social influence, only seemed to affect working alliance (β= .16, p < .001), but not interpersonal closeness (p= *.35*) or projected future relationship (p= *.36*).

Comfort with self-disclosure[19] positively correlated with working alliance (β= .09, p < .05). Meanwhile, comfort with data sharing[20] positively affected projected future relationship (β= .01, p < .05).

Agent Persona. Racial mirroring between clients and agents affected therapeutic relationships. See Table 23 below for a summary of test statistics. Based

---

[19] Comfort with self-disclosure evaluated how comfortable participants felt about discussing their life challenges during their conversation with the chatbot in the online experiment.

[20] Comfort with data sharing evaluated how comfortable participants felt about sharing various types of data continuously in *future* interactions with the chatbot.

on a pairwise comparison, participants assigned to the racial mirroring condition reported a significantly higher level of working alliance ($M = 3.49$, $SD = .04$), compared to those in the non-mirroring group ($M = 3.36$, $SD = .03$), $F(2, 365) = 11.87$, $p < .05$. The difference between the control group and each of the experimental groups was not significant.

Similarly, racial mirroring had an effect on perceived interpersonal closeness. Participants in the racial mirroring condition reported significantly higher interpersonal closeness ($M = 3.49$, $SD = .08$), compared to those in the non-mirroring group ($M = 3.26$, $SD = .07$), $F(2, 365) = 8.1$, $p < .05$. This effect was particularly prominent among male participants. The difference between the racial mirroring group and the control group was not significant.

No racial mirroring effect was observed for projected future relationship.

Table 23. Effects of racial mirroring condition on therapeutic relationships

DVs	**Racial Mirroring**		**Non-mirroring**		**Control**		***F* (2, 365)**	η^2
	M	***SD***	***M***	***SD***	***M***	***SD***		
Working alliance	3.49	.04	3.36	.03	3.44	.04	11.87*	.05
Perceived interpersonal closeness	3.47	.08	3.26	.07	3.19	.10	8.11*	.03
Projected future relationship	3.42	.12	3.32	.11	3.42	.13	3.53	.00

Note: * ($P \leq 0.05$) ; **($P \leq 0.01$); *** ($P \leq 0.001$)

5.4.2. Results from Interviews (RQ4)

Based on the interview data, I further explore how therapeutic relationships were developed and how they could be enhanced. Interviewees were asked whether they felt connected to the chatbot, and if so, what were the contributing factors. First, the qualitative data confirmed that participants who lacked social interactions were more likely to form a social connection or bonding with the chatbot. As Kristy described, "*This connection would be so important especially for people like me and my sister. We don't get out and about very much, we're very much alone. So having a texting buddy that you can talk about things would be so good*." Weevyn described a similar appreciation of chatbot: "*I wasn't interacting with anyone, it was very isolating. So when I saw the robot, I thought it was interesting, and I like how she was very straightforward*."

Surprisingly, unlike traditional interpersonal relationships that develop over time, client-agent therapeutic relationships could be formed immediately. Alesia described how she felt there was an "instantaneous" connection and trust toward the chatbot, saying:

> Normally [with] relationships, you have to build them and you have to get to the point where you can trust a person, but for some reason... when I was talking to Robbie, I didn't feel like that had to be built, I didn't feel like I had to build a relationship, I can't even explain it. I just felt really natural talking to Robbie, I didn't feel like I had to build a relationship.

The chatbot's disclosure of its purpose might have contributed to such an instant connection. Additionally, the chatbot's function as a safe space without judgment was another contributing factor, as Alesia reflected:

> I think maybe because the introduction shared the purpose and what it was and what his function was, it was just easy to use him for that, as opposed to like, you have to really build trust to share with friends. Even when you start with [a human] therapist, you have to build a trust, and you don't really wanna

> open everything because they’re going to be in studying you and analyzing you, so I didn’t feel that [disclosing to human therapist] was easier.

Other participants believed the connection with the chatbot was developed gradually over time. Specifically, the therapeutic relationship could be strengthened if chatbot had the ability to retain “memory” and initiate a follow-up conversation. Ana said,

> A follow-up based on the previous conversation would be helpful for developing a relationship. Having a record, like if the chatbot knows that you had a job-related problem last week, so this week like, ‘Okay, how are you feeling this week?’ But at the same time, checking in with you about the problem you had last week.

Qualitative data also confirmed the role of anthropomorphism in forming therapeutic relationships. One determinant for the level of anthropomorphism was focused on the agent’s appearance. Chai indicated the importance of having more realistic profile, saying: “*If it doesn't look real, there's a lost connection*.” Another way to increase the human-likeness was to equip the chatbot with a personal narrative, and several interviewees mentioned the idea of having a warm-up conversation or small talk in order to build up this relationship. Charlie said,

> It could be a feature you add for Molly when they were created, for example, maybe some generic hobbies that they might have or their interests. You could even make it so that they could check in with you, just randomly saying, ‘hello, how you're doing?’ kind of thing, and then having a generic conversation.

Qualitative data also revealed the role of shared social identity in shaping therapeutic relationships. In the traditional social interaction contexts, shared social identity refers to a sense of "we-ness" when people perceive each other as members of a common social group (Neville et al., 2020). When interacting with an anthropomorphized chatbot, participants might pay attention to certain common traits. As participants perceived the agent to be in the same social group as them, they were more likely to feel connected. A shared identity can be defined based on physical, cultural, or psychological commonalities. Several female interviewees indicated that they felt more connected to female agents due to the similarities in thinking and reactions. Race, or even simply belongs to a minority group, can also be a basis for the perceived shared identity between client and the agent. Chai, an Asian male, described how he felt connected to Darnell (the African American chatbot) because they both belonged to minority group:

> So just having a different [non-White] skin color has already made a connection right away…I think because I'm a minority too, I'm Asian. And then when you're involved with other minority by color skin, you're able to make that connection.

On the other hand, several interviewees indicated they didn't feel connected to the chatbot. One reason was that they felt there was a lack of shared experiences. Connor said he preferred to seek help from online community like Reddit where he could find people in similar situation. He said, "*There's usually someone who has been through a similar thing, so they can speak from personal experience as to how* *they dealt with it and what things are doing to try and help themselves, but with this chatbot, it doesn't have that personal experience.*"

Another reason for this lack of connection was that some participants felt the chatbot was unable to fulfill their need for deeper level of relationship. When asked whether he felt a sense of connection or personal bonding, Charlie said, *"I'm not the type of person that has many superficial-level relationships. I'd like to take time to get to know people, their hobbies, who they are as a person… talking with the chatbot doesn't really have that same sense of personal level."*

5.4.3. Summary (RQ4)

Both quantitative and qualitative data suggest that meaningful therapeutic relationships can be formed between humans and chatbot therapists. The survey data showed that participants equipped with a higher level of social support from their friends and family were less likely to engage in a close relationship with the agent. In comparison, participants without sufficient social support perceived that they could benefit more from interacting with chatbots. This finding was also corroborated in the qualitative data.

The survey analysis also showed consistent positive effects of anthropomorphism and perceived usefulness/efficacy on all three forms of therapeutic relationships. Additionally, comfort with self-disclosure positively predicted working alliance formed from the previous conversation, while comfort with data sharing appeared to strengthen perceived future relationship. The role of racial mirroring in shaping therapeutic relationships was also confirmed, with the qualitative findings suggesting that participants felt more connected to an agent with a shared social identity.

5.5. What is the effect of the therapeutic alliance on human-agent psychotherapeutic conversations? (RQ5)

As previous (RQ4) findings confirmed that meaningful therapeutic relationships can be established between humans and chatbot agents, RQ5 explores how the therapeutic relationships might affect human-agent conversations. I address this question using survey data. The human-agent conversation outcomes were measured by the level of conversation engagement (*M* = 63.74, *SD* = 1.30) and content sensitivity (*M* = 59.02, *SD* = 1.41).

5.5.1. Results (RQ5)

I conducted three sets of general linear regression analysis, with the DVs being level of conversation engagement, content sensitivity, and conversation word count, respectively. The IVs included three variables measuring therapeutic relationships (working alliance, perceived interpersonal closeness, projected future relationship), while controlling for age, income, social support, life satisfaction, comfort with self-disclosure, comfort with data sharing, and anthropomorphism.

There were no observed effects of therapeutic alliance on content sensitivity or word count. Table 24 reports statistics for predicting level of engagement.

Table 24. Multiple linear regression model predicting level of engagement

	DV= Level of conversation engagement					
IVs	**Std. β**	**SE**	**t**	**Sig.**	**Partial eta squared η2**	**Observed power**

Background Factors						
Age	.02	.07	.60	.55	.04	.97
Income	**-.07***	**.27**	**-2.00**	**.05**	**.00**	**.09**
Social support	.02	.97	.51	.61	.01	.51
Life satisfaction	.06	.96	1.61	.11	.00	.08
Chatbot-related Perceptions						
Comfort with self-disclosure	.07	1.77	1.67	.10	.01	.36
Comfort with data sharing	.05	.04	1.37	.17	.01	.38
Anthropomorphism	**.37*****	**.07**	**6.67**	**.00**	**.01**	**.28**
Therapeutic Relationships						
Working alliance	**.26*****	**1.74**	**5.08**	**.00**	**.11**	**1.00**
Perceived interpersonal closeness	**.14****	**.99**	**2.75**	**.01**	**.07**	**1.00**
Projected future relationship	.02	.65	.33	.74	.02	.78
Model fit	F(10, 375) = 54.82, p < .001					
Adjusted R^2	.60					

Note: * ($P \le 0.05$) ; **($P \le 0.01$); *** ($P \le 0.001$)

Results of the multiple linear regression indicated that these variables collectively explain 60% of the variance in conversation engagement, *F(*10, 375) = 54.82, p < .001, adjusted R^2 = .60. Working alliance (β = .26, p < .001) and perceived interpersonal closeness (β = .14, p < .01) were positively correlated with the level of conversation engagement, while projected future relationship was not significant.

5.5.2. Summary (RQ5)

Building on RQ4—which confirmed that meaningful therapeutic relationships can be established—my exploration of RQ5 further substantiated that the strength of therapeutic relationships had a direct impact on how participants engaged with agents in psychotherapeutic conversations. The findings align with traditional psychotherapy studies suggesting that therapeutic alliance facilitates collaboration between clients and chatbot therapists, contributing to a therapeutic process with more proactive patient engagement (Flückiger et al., 2018).

5.6. How does anthropomorphism relate to people's sense of being understood after making disclosures to the agent? (RQ6)

In traditional mental health care, clients' perceptions of feeling understood by their therapists positively predict psychotherapeutic alliances and treatment outcomes (Elliott et al., 2011; Pocock, 1997). In the context of chatbot therapists, I explore whether and how people might feel understood.

Although several interviewees confirmed that they felt "listened to" and "connected," when asked whether they felt being understood, the majority of them said they did not. Anthropomorphism—the level of humanness people attributed to the chatbot—played a pivotal role in creating a sense of being understood.

First, knowing the chatbot was not human created a mental barrier as they didn't think the chatbot was able to relate to them or understand them. Weevyn described, *"I know it's a robot, so it's not the same as speaking to a person, so I know that the understanding is completely different. I don't feel like I'm being understood by it... For me, interacting with a chatbot is almost like a reflection. You're reflecting your own thoughts back to yourself."*

Additionally, participants didn't feel being understood due to the generic responses given by the chatbot. They felt that a human would be able to provide more personalized answers, as Charlie said, "*I want more depth in the AI's response and I'm not getting that depth, as if it was reading what I was telling them and it actually gives very specific pointed questions to what my problem is."*

On the other hand, more tailored responses to a person's disclosure could make a chatbot more human-like and subsequently attribute to a sense of being understood. For instance, when participants sensed that the chatbot's responses were "on point," they felt understood. For example, Kristy said, "*I did feel understood by Robbie, that kind of surprised me with some of his responses, that's why I was like, is this a bot or is this human?"*

Additionally, several interviewees indicated that they would feel more understood if there were a shared identity. A shared identity could manifest in gender and racial similarity, as Alesia said, "*Because I can relate and identify with Ebony as a Black woman, I probably would feel even more comfortable and more understood by her."* Going beyond the physical appearance of the agent, some participants also expected the shared identity could manifest in chatbot's responses, as Rena said:

> More than just having an avatar that kind of looks like me, it'd be great if they were more customized experience to my own kind of gender identity and racial identity, ethnic background, like all that kind of stuff, like more customized answers. I don't know exactly how that would be done, but it would help me feel more understood.

Furthermore, interviewees were asked to identify any attributes or features that could make the chatbot more human-like and more understanding. One of the most human-like features identified was the ability to acknowledge by repeating or rephrasing what participants disclosed. Travis described how he would've felt more understood if the chatbot could acknowledged what he disclosed: *"When I said about my mom dying, I was thinking like, well, if he admitted something about that or included some of that text into his response, it would have made me feel like it was more of a real thing."* Similarly, Connor said, "*It feels like I was being listened to, but*

I don't feel like it understood me. It would be interesting if it actually acknowledged one of the things I said. Maybe it's like, oh yeah, so you talk about this… that would be useful to me."

To summarize, although participants felt being heard after disclosing to the chatbot, most of them didn't feel a sense of being understood because the chatbot's responses were too generic and not personalized. Despite the current limitation, several factors associated with anthropomorphism could help enhance the sense of being understood. These factors include a shared identity (e.g., based on gender or race), as well as the chatbot's ability to give personalized responses and to give acknowledgment by repeating what people have disclosed.

5.7. How comfortable do people feel about sharing their information with a chatbot in psychotherapeutic conversations? (RQ7)

In the traditional psychotherapy context, self-disclosure is associated with better therapeutic outcome (Pennebaker, 1990). However, some clients feel reluctant to fully open up to therapists due to their need to create favorable self-presentation (A. E. Kelly, 2000). In the context of chatbot use for mental support, I distinguish comfort with self-disclosure and comfort with data sharing. On one hand, as one previous qualitative study suggested, chatbots provide a "safe space" where users can discuss any topic without the fear of judgment or retaliation (Ta et al., 2020). On the other hand, data sharing with the chatbot is not always HIPAA-protected like traditional psychotherapy, which could become a data privacy challenge for people. These complexities call for empirical data to better understand people's mental models related to self-disclosure and data sharing with chatbot.

5.7.1. Results from Survey (RQ7)

The survey asked participants to compare their conversation with the agent to one with their friends, family, or a counselor. Participants then reported how much they agreed or disagreed with several statements (listed in Table 24). On average, participants reported a fairly high level of comfort with self-disclosure when interacting with agents (*M* = 3.87, *SD* = .62). There was high agreement with several items, with participants voicing agreement that they could express themselves openly, and that they wouldn't be judged. Most interestingly, 235 (56.5%) participants agreed or strongly agreed their "ability to communicate issues increased" when interacting with an agent.

Table 25. Descriptive statistics of statements evaluating self-disclosure comfort

Statements	**Strongly Disagree (1)**	**Disagree (2)**	**Neutral (3)**	**Agree (4)**	**Strongly agree (5)**	**Mean**	**SD**
	Percentage of participants						
I felt comfortable when I talked with [agent name] about my challenge and life experiences.	1.2%	5.4%	16.8%	53.8%	22.9%	3.92	.85
I expressed myself openly.	1.0%	3.6%	10.9%	53.0%	31.4%	4.1	.81
I felt [agent name] wouldn't judge me when I talked about my issue.	1.5%	1.7%	9.8%	47.3%	39.8%	4.22	.81
My ability to communicate issues increased when I interacted with [agent name].	5.9%	9.8%	27.1%	44.4%	12.9%	3.49	1.03

The survey also asked how comfortable participants felt about sharing their data with the agent. They responded on a scale from 1 (absolutely not comfortable sharing) to 100 (completely comfortable sharing). Table 26 presents participants' various levels of comfort with sharing 13 data types. Participants were more comfortable sharing certain types of data (e.g., relationship status, sexual orientation) than others (e.g., phone call history, sexual activity).

Table 26. Comfort with data sharing based on data type (1= absolutely not comfortable sharing; 50 = somewhat comfortable; 100 = Completely comfortable sharing)

Type of data	Mean	SD	Skewness	Kurtosis
Sexual orientation	65.7	32.5	-0.81	-0.52
Relationship status	60.78	32.08	-0.67	-0.69
Mood tracking	58.4	32.67	-0.53	-0.92
Birthday	57.34	33.7	-0.47	-1.07
Email address	52.12	34.18	-0.25	-1.27
Disease-related information (e.g., condition and symptoms)	43.2	34.4	0.09	-1.36
Phone number	31.82	31.55	0.63	-0.9
Your employment history	31.63	33.16	0.64	-1.02
Medical history	31.41	33.64	0.63	-1.03
Location data	29.7	31.54	0.77	-0.7
Sexual activity	27.91	31.63	0.78	-0.72
Your photos and videos	20.49	28.46	1.33	0.59
Phone call history (i.e. when and whom have you contacted)	17.77	26.82	1.53	1.29

To further explore factors affecting comfort with self-disclosure and data sharing, I built two multiple linear regression models. Table 27 presents the results. In the model predicting comfort with self-disclosure (adjusted R^2= .49; $F(15, 379) = 21.43$, $p < .001$), participants who held more positive general attitudes toward AI (β= .08, $p < .05$) and who were more fond of the agent (β= .01, $p < .01$) reported higher comfort with self-disclosure during their interaction with the agent. Positive effects of perceived usefulness/efficacy (β= .14, $p < .05$) and ease of use (β= .14, $p < .001$) were also observed. The strength of sociotechnical relationship was unrelated to comfort with disclosure in this model. On the other hand, the therapeutic relationship, as measured in working alliance (β= .13, $p < .01$), positively predicted comfort with self-disclosure. Additionally, participants seemed less comfortable disclosing to male agents (β= -.13, $p < .05$) than female agents.

The model for comfort with data sharing explained just 18% of the variance, $F(15, 379) = 6.91$, $p < .001$. Only three variables showed significant positive effects: general attitudes toward AI (β= 4.2, $p < .05$), perceived ease of use (β= 4.91, $p < .05$), and perceived future relationship (β= 2.18, $p < .01$). Interestingly, the effect of working alliance was reversed from the prior model; participants who reported higher working alliance (β= 6.23, $p < .05$) with the agent felt less comfortable sharing data. Aside from the possibility of a Type-1 error in statistics, one explanation could be the different nature of the two DVs. Self-disclosure is required and expected as part of therapy; participants may not have thought some of the data items were necessary for therapeutic outcomes and may have been less comfortable sharing data that does not have a clear connection to therapeutic outcomes. Another potential explanation is that when working alliance was rated high, participants perceived the chatbot technology to be more sophisticated. As a result, there was an increase in the level of uncertainty and anxiety about sharing data to the agent.

Table 27. Models predicting level of conversation engagement

IVs	**DV=Comfort with self-disclosure**		**DV=Comfort with data sharing**	
	Std. β	***SE***	***Std. β***	***SE***
Control Variables				
Life satisfaction	-.04	.03	1.10	1.25
Privacy concern level	-.04	.02	-1.86	1.08
Social support	.03	.03	-.74	1.25
Chatbot-related Perceptions				
General attitudes toward AI	**.08***	**.04**	**4.20***	**1.84**
Fondness of agent	**.01****	**.00**	.13	.09
Anthropomorphism	.00	.00	-.08	.12

Perceived usefulness / Efficacy	**.14***	**.06**	-4.80	2.63
Perceived ease of use	**.16******	**.05**	**4.91***	**2.06**
Sociotechnical relationships				
Likelihood to recommend	-.01	.00	.14	.09
Willingness to continue engaging	.00	.00	.12	.09
Therapeutic relationships				
Working alliance	**.13****	**.06**	**-6.23***	**2.57**
Perceived interpersonal closeness	.02	.03	.63	1.31
Projected future relationship	.01	.02	**2.18****	**.84**
Agent gender				
Male	**-.17***	**.07**	-3.78	2.45
Female	- [a]	- [a]	- [a]	- [a]
Experimental Conditions				
Racial mirroring	.05	.07	3.33	3.10
Non-mirroring	.10	.07	1.40	3.03
Control	- [a]	- [a]	- [a]	- [a]
Model fit	F(15, 379) = 21.43, p < .001		F(15, 379) = 6.91, p < .001	
Adjusted R^2	.49		.18	

Note. a. Value not applicable due to reference level
b. * ($P \leq 0.05$) ; **($P \leq 0.01$); *** ($P \leq 0.001$)

5.7.2. Results from Interview (RQ7)

The qualitative findings largely confirmed the survey findings. In term of self-disclosure, the vast majority of interviewees indicated that they felt more comfortable disclosing to the chatbot compared to a human therapist or their friends or family. As detailed in Section 5.1.2, there were three main reasons: the chatbot provided a safe space free from social judgment, there was no pressure to manage self-presentation, and there was no need to worry about other people's feelings. Essentially, participants felt more comfortable disclosing to the chatbot precisely because it's not human. Charlie said, "*Having the chatbot to talk to is very great. It's impersonal, so you don't have to worry about someone else hearing it, like your family or friend... It's like if you had a diary or journal and you wrote in every day, but much better, because at least it talks back to you.*"

Additionally, qualitative data provided additional evidence that participants felt more comfortable disclosing to female agents due to the perception that women are more understanding. Charlie observed, *"This is kind of getting into assumptions here, but most guys don't talk about their problems, and if they do, they don't provide the advice or the level of perspective that a female might offer. So it's probably more comfortable for me and easier to talk to a female."*

In terms of data sharing, there were varying degrees of concerns. Some interviewees felt less concerned; for example, Frances said, "*I don't care. It's fine, as long as the information is not being used in a malicious way."* Some felt unconcerned because their conversations with chatbot were not considered sensitive or damaging, as Beverly said:

> I don't feel concerned in terms of privacy. Because the issues I present are not classified or that if it came out, it would be the end of my reputation type of thing. I think it's just very normal coming of age issues. I might be a bit embarrassed, but I wouldn't be like, "Oh my God, I can't believe it invaded my privacy that way*"*.

Who collected the data also affects the level of concern. Travis said, "*It depends on who's behind it, if I knew that it was from someone that I trusted, then it wouldn't bother me at all, really*." Several participants felt data privacy was not an issue because they held a fatalistic view about privacy (Penney, 2019), as they believed there was no privacy once they went online. On the other hand, Rena was very concerned about data privacy and security issues, and she felt hesitant about opening up to the chatbot due to her concerns:

> It’s hard to open up to a chatbot. I normally find it easier to open up to humans, but with the chatbots, I don’t really know what's being done with this data. And there’s no trust established yet. I’m sure my trust level would increase after one or two interaction, but it would probably take multiple sessions and a pretty good understanding of its data privacy for me to like feel better about it.

However, data privacy and security concerns did not necessarily deter participants from interacting with the chatbot. In fact, there were two different sharing standards that participants abided when interacting with chatbot, as Chai explained,

> I wouldn’t feel so comfortable sharing my information like my Social Security [number], phone number, or even email. But on the social interaction level, I will be more free. I have more freedom, I wouldn’t think twice about sharing my childhood trauma. So there are two different standards that I would go by.

Essentially, there were two different mental models when deciding what to share with chatbot. On one hand, participants generally felt very comfortable disclosing their personal thoughts and feelings with chatbot. On the other hand, with the data privacy consideration, they would take precaution not to disclose sensitive Personal Identifiable Information (PII) like their phone number or email.

5.7.3 Summary

Both quantitative and qualitative data suggest that chatbots provide a safe space for participants to share their thoughts and feelings freely, without the fear of social judgment or negative consequences. Female agents appeared to make participants more comfortable with self-disclosure.

In the context of therapeutic chatbot, data privacy concerns emerged depending on what data was collected, who collected the data and how the data would be used. To account for potential data privacy risks, participants would take precautions not to share sensitive PII while still being open about their thoughts and feelings.

5.8. How does racial mirroring influence people's perceptions of and interactions with a chatbot in psychotherapeutic conversations? (RQ8)

In the traditional psychotherapy, there is a racial mirroring effect predicting client preferences for same-race therapists or counselors due to perceived credibility and trust. When a chatbot is equipped with a human persona, their race become an integral part of their social identity. Using both survey and interview data, RQ8 explored whether people's preferences for a same-race counselor would persist or disappear when interacting with chatbot.

5.8.1 Results from Survey (RQ8)

Based on the survey data, I first conducted a multivariate analysis of covariance (MANCOVA) using perceived interpersonal closeness, disclosure comfort, willingness to continue interacting with agent, satisfaction, and likelihood to recommend as dependent variables, with racial mirroring as the independent variable, while controlling for agent and participant sex, privacy concerns, general attitudes toward AI, the anthropomorphism of agents, and background psychological factors.

Across all participants, there were no significant differences in any of the outcome variables based on experimental conditions (racial mirroring, non-mirroring, control). However, when segmenting participants by race, there was a racial mirroring effect found among White participants (n=165), as they reported significantly stronger sociotechnical and therapeutic relationships when they were assigned to the racial-mirroring condition, compared to when they were assigned to the non-mirroring condition. The differences between the control and racial-mirroring conditions were not significant. Results for Asian, Black, and Latinx participants were inconclusive, possibly due to the smaller sample size.

Therefore, in this section, I first present findings from MANCOVA modeling based on the segment of White participants. Results revealed significant main effects for racial mirroring, Wilk's $\lambda = .83$, $F(6, 218) = 3.98$, $p < .001$. Since MANCOVA is an omnibus test[21], I also conducted univariate analysis ANCOVA to unpack how racial mirroring influenced participants' perceptions of the agent with whom they interacted, and how these effects varied based on agent gender and participant sex. To ensure the validity of statistical analyses, I chose to perform generalized linear modeling (GLM) rather than linear regression due to the non-normal distribution of the dependent variables (Ormerod, 2017).

The effect of racial mirroring on perceived interpersonal closeness. The modeling showed a significant main effect for racial mirroring on participants' perceived interpersonal closeness with the agent, $F(2, 216) = 35.61$, $p < .001$, partial $\eta^2 = .15$. Bonferroni post-hoc comparisons revealed that participants perceived a closer personal relationship with same-race agents ($M = 3.47$, $SE = .09$), compared to

---

[21] While statistical significance can be identified using an omnibus test, such a test doesn't specify which parameter is significantly different from the other. Therefore, we need to conduct follow-up tests to identify which parameters, in particular, have statistically significant differences.

those who interacted with different-race agents (*M = 3.22*, *SE* = .09). The differences between the control condition (*M = 3.33*, *SE* = .09) and either of the experimental groups were not statistically significant.

The effect of racial mirroring on client self-disclosure comfort. The modeling showed a significant main effect for racial mirroring on participants' degree of comfort with self-disclosure while interacting with the agent, $F(2, 216) = 40.63$, $p < .001$, partial $\eta^2 = .18$. Bonferroni post-hoc comparisons revealed that participants felt less comfortable disclosing to same-race agents (*M* = 3.78, *SE* = .04), compared to different-race agents *(M* = 3.93, *SE* = .05) and the robotic agent (*M = 3.90*, *SE =* .04). The difference between the robotic agent and different-race agent was not statistically significant.

The effect of racial mirroring on user satisfaction. Focusing on the dependent variable of user satisfaction, the modeling showed a significant main effect for racial mirroring on how satisfied participants felt about their interaction with the agent, $F(2, 216) = 34.22$, $p < .001$, partial $\eta^2 = .16$. Bonferroni post-hoc comparisons revealed that participants reported a higher degree of satisfaction after interacting with a same-race agent (*M* = 3.48, *SE* = .04), compared to a different-race agent (*M* = 3.30, *SE* = .05). Additionally, satisfaction with the robotic agent (*M* = 3.52, *SE* = .05) was much higher than same-race agent. One explanation could be that participants had a high hope for same-race agents, but the actual experiences did not meet their expectations.

The effect of racial mirroring on recommendation likelihood. The modeling showed a significant main effect for racial mirroring on whether and how strongly participants would recommend the agent to their friends and family, and to people in need of psychotherapy, $F(2, 216) = 54.29$, $p < .001$, partial $\eta^2 = .32$. Bonferroni post-hoc comparisons revealed that participants were more willing to recommend same-race agents ($M = 61.68$, $SE = 1.27$) and the robotic agent ($M = 61.68$, $SE = 1.27$) to others, compared to different-race agents ($M = 60.09$, $SE = 1.30$). The difference between the robotic agent and same-race agents was not statistically significant.

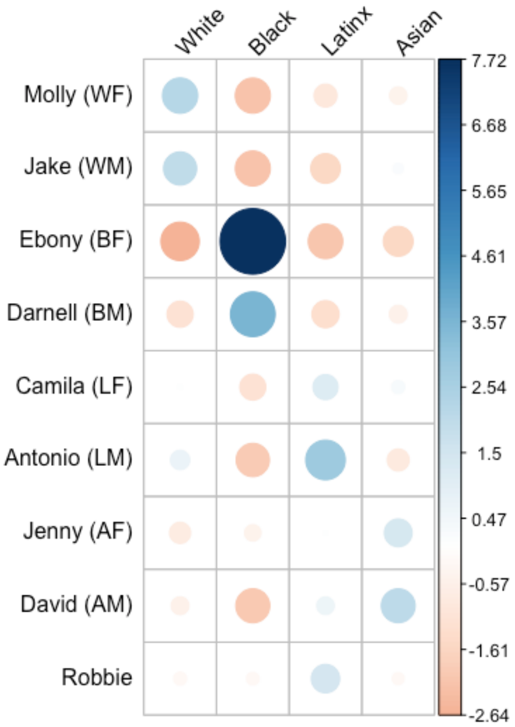
The effect of racial mirroring on willingness to continue engaging with the agent. The modeling showed a significant main effect for racial mirroring on participants' desire to continue interacting with the chatbot agent, $F(2, 216) = 39.57$, $p < .001$, partial $\eta^2 = .14$. Bonferroni post-hoc comparisons revealed that participants had a stronger desire to continue interacting with same-race agents ($M = 64.93$, $SE = 1.25$), compared to different-race agents ($M = 61.47$, $SE = 1.47$). The differences between the control condition ($M = 62.93$, $SE = .09$) and either of the experimental groups were not statistically significant.

The effect of racial mirroring on perceived future relationship. The modeling showed a significant main effect for racial mirroring on participants' prediction of the closeness of their future relationship with the agent, $F(2, 216) = 42.16$, $p < .001$, partial $\eta^2 = .18$. Bonferroni post-hoc comparisons revealed that participants predicted that they would have a closer relationship with same-race agents ($M = 3.71$, $SE = .07$) and the robotic agent ($M = 3.56$, $SE = .08$), compared to different-race agents ($M = 3.21$, $SE = .09$). The difference between the robotic agent and same-race agents was not statistically significant.

5.8.2. Results from Survey: User Preferences for Agent Personas

At the end of the survey, participants were asked to select agent personas they would prefer if they were able to customize the therapeutic chatbot profile; they could select more than one persona they liked. A chi-square test of independence based on participant race and the agent personas they selected revealed that participants' preferences for agent personas were different depending on their own races, χ^2 (24, 206) = 142.23, $p < .001$, Cramer's V = .31, suggesting a moderate effect, as illustrated in Figure 6.

Figure 6. A visualization of standardized chi-square residuals.
(Blue nodes: more likely to select an agent; red nodes: less likely to select an agent. The size of a node denotes the intensity of participants' preferences.)



The further residual analyses confirmed that participants preferred interacting with same-race chatbots. Specifically, White/Caucasian participants were more likely to select Molly and Jake; Black/African American participants were much more likely to select Ebony and Darnell; Latinx participants were much more likely to select Antonio and Camila; and Asian participants were much more likely to select Jenny and David. Such preference for same-race agents was especially pronounced among Black/African American participants. The probability that they selected Ebony, the female Black/African American agent, was seven times greater than what would be expected based just on random chance.

5.8.3. Results from Interviews (RQ8)

In the interviews, I presented participants with all nine variations of agent personas and asked them which agent(s) they preferred to interact with. Out of 16 interviewees, seven said they preferred female agents, four said they preferred talking to somebody of their age, four said they didn't have a preference, and two said they would prefer any agents as long as they have a human face instead of robot.

Four interviewees specifically said that they prefer agents of their same race. Alesia, who is Black, indicated her preference for Ebony, saying she would feel more comfortable with and connected to a chatbot with a shared identity:

> The statistics say, when a person walks in the room, the first thing they do is to look for someone that looks like them, and I think that that's probably true. Because the first thing I did was look for a chatbot that looks like me that I could identify with... First thing I thought, 'Oh, I can identify with Ebony'… Because I can relate and identify with Ebony, I probably would feel even more comfortable talking to her.

Although same-race agents could be perceived as more approachable, the "sameness" could be a deterrent for interaction due to the perceived lack of objectivity, as Charlie pointed out,

> Obviously, speaking to someone of your same background is always the easiest to get along with, but it could also be a deterrent for others, because if you know that they're similar to you, that they might already have the same mindset or the same kind of understanding, so that's a positive, but of course the negative would be because of that understanding they might not propose an outside perspective that you may want.

5.8.4. Summary

For personified agents, race is an integral part of the design. RQ8 findings suggest that racial preferences do not disappear in human-agent interaction. Although analysis based on a larger sample is needed to generalize this finding to non-white population, initial evidence revealed that white participants who interacted with same-race agents perceived a closer interpersonal relationship with the agent. Additionally, these participants were more satisfied with their interactions, had a stronger willingness to continue engaging with the agent, and were more likely to recommend the agent to their friends and family and to people who need psychological help. These findings are in line with studies conducted in traditional in-person psychotherapy settings (e.g., Caplan & Turner, 2007). Racial mirroring is particularly effective for African American clients. This study reveals that, of all of the groups of participants based on race/ethnicity, African American participants had a much stronger preference for same-race agents than any other group.

On the other hand, racial mirroring between client and agent appeared to entail a decrease in comfort with self-disclosure. Participants who interacted with same-race agents reported a lower level of disclosure comfort; they were significantly more likely to feel like they were being judged than participants in the non-mirroring and control groups. This reduced disclosure comfort might result from c stronger need for impression management (Leary, 2019) and heightened fear of public stigma (i.e., being viewed negatively by others) (Vogel et al., 2007). In particular, racial similarity might be associated with a greater degree of "groupness," the degree to which a collection of people is perceived as a unified or meaningful entity (Campbell, 1958; Hamilton & Sherman, 1996). And participants might feel greater pressure to conform and worry more about being labeled as "abnormal."

Chapter 6: Discussion

In this chapter, I first present a summary of findings while discussing theoretical contributions. Then, I describe a need-based assessment model to provide directional guidelines for chatbot technology adoption in psychotherapy context. Finally, I present tactical design recommendations to inform future development of chatbots for mental health support.

6.1. Summary of Findings and Theoretical Contributions

I conducted the research using a mixed-method approach. Using data gathered in an online experiment with survey, my quantitative analyses helped explain the inter-relationships between chatbot design outcomes (e.g., anthropomorphism and racial mirroring) and human-agent relational outcomes, including sociotechnical and therapeutic relationships. Based on semi-structured interviews, my qualitative analyses revealed the benefits and drawbacks of a therapeutic chatbot. Additionally, I identified several themes around people's perceptions of agents that reinforced findings from quantitative data. In the following subsections, I outline five insights derived from my quantitative and qualitative analyses while highlighting connections and contributions to previous literature.

6.1.1. Assessing therapeutic chatbot benefits and drawbacks systematically

The first theoretical contribution of this dissertation is the systematic assessment of the therapeutic chatbot. This holistic approach fills the gaps in prior work narrowly focusing on a certain level; I identified four levels of benefits and drawbacks to explain why or why not participants adopt and engage with this technology for mental health support.

First, on the technology level, the benefits of the therapeutic chatbot included its low cost, convenience, anonymity, efficiency and immediacy of service, which providing users a great sense of control. Drawbacks mainly centered on participants' fatigue and distrust towards the technology, aggravated by the lack of human touch and data privacy concerns.

Second, on the social level, the benefits of the therapeutic chatbot included its provision of social support, companionship, and a safe space for participants to openly express themselves without fear of social judgment or pressure to maintain self-presentation. Drawbacks included a lack of tangible support as the chatbot was not able to form meaningful social connections comparable to that from humans.

Third, on the therapeutic level, there were several perceived benefits associated with chatbot use, as participants felt a sense of relief after unloading thoughts and feelings. The chatbot could also potentially improve participants' psychological well-being by providing reassurance and validations and by boosting their morale and self-confidence. Drawbacks primarily originated from the lack of patient-specific recommendations and responses.

Fourth, on the interaction level, one main benefit was the chatbot's conversation-based interface, which was considered natural and smooth. The chatbot also demonstrated favorable personalities described by participants as being supportive, warm, and friendly. Drawbacks of the chatbot pointed to the sense that chatbot responses were scripted and mechanical, and that the flow of conversation became broken when the chatbot asked repetitive questions or ended a discussion prematurely.

Taken together, these four levels of benefits and drawbacks provide a comprehensive assessment of chatbot application in the mental health context. Previous studies in evaluating therapeutic chatbots tend to have a narrow focus, such as overemphasizing the therapeutic efficacy (Fitzpatrick et al., 2017) or the technological enhancement of a chatbot (Ta et al., 2020). My dissertation contributed to the literature with a systematic approach that ensures a balanced view avoiding either the camp of techno-optimism or techno-pessimism.

6.1.2. Applying and extending UTAUT in the context of therapeutic chatbot

The second theoretical contribution of this dissertation is that it extends and enhances the traditional technology acceptance theory, UTAUT. Specifically, I added a new construct, anthropomorphism, to improve the explanatory power of UTAUT for therapeutic chatbots and broader AI applications.

In exploring human-agent sociotechnical relationships, I draw upon UTAUT as the theoretical base for my statistical modeling. My analyses confirmed the importance of Perceived Usefulness and Social Influence in driving the adoption and continued engagement with the therapeutic chatbot technology. Interestingly, contrary to what UTAUT would suggest (Venkatesh et al., 2003), I found that Perceived Ease of Use did not directly impact adoption, likely due to a "ceiling effect" as most respondents rated the technology easy to use.

This research also improved UTAUT by incorporating anthropomorphism in the model. This can be applied to other AI-based technology contexts such as voice-based intelligent personal assistants. I found that anthropomorphism strengthened human-agent sociotechnical relationships. When participants perceived the chatbot to be more human-like, they were more willing to recommend and continue engaging with the chatbot. The importance of anthropomorphism was also reflected in the interview data, as participants who had more positive experiences deemed the agent as a “friend in the pocket.”

Taken together, these findings echoed previous qualitative research indicating that anthropomorphism is associated with positive treatment outcomes, social support, and emotional support (Inkster et al., 2018; Krämer et al., 2018; Purington et al., 2017). Moreover, my research was able to quantify the impact of anthropomorphism by applying and extending UTAUT to better predict chatbot adoption.

6.1.3. Unpacking how and why chatbot could be effective for mental health support

The third contribution of this dissertation is that it empirically uncovered how and why a therapeutic chatbot could contribute to improved mental health outcomes. Although previous research has shown that a chatbot can provide an effective therapeutic experience (Fitzpatrick et al., 2017; Park et al., 2019), the underlying mechanisms were not well understood. As the development of a positive therapeutic relationship is one of the best predictors of outcomes in psychotherapy (Sucala et al., 2012), the American Psychological Association has specifically called for the examination of therapeutic relationships in any therapeutic solution that claimed to be effective (Ackerman et al., 2001). My dissertation mitigated this gap by demonstrating that human-agent therapeutic relationships can indeed be formed.

My analyses showed that participants with lower levels of social support tended to form a stronger therapeutic relationship with the chatbot. This finding was corroborated with previous studies on “social snacking” (Gardner et al., 2005), suggesting that people treated conversational agents as companions to satisfy their need for social belongingness. Additionally, human-agent therapeutic relationships could be formed very early on, as some participants felt “instantaneously connected” to the chatbot since they trusted it was well-intended. And the human-agent therapeutic relationship could grow stronger as their conversations involved deeper self-disclosure.

It is important to note that the human-chatbot therapeutic relationship was purely built upon participants’ perceptions of the chatbot. Many participants held the notion that the chatbot was not a human and lacked the judgment or bias participants expect from humans. Although it didn’t have the ability to understand or empathize, the chatbot was able to give “validation” and “assurance” and to motivate participants so they could confront and solve their problems. There was also a downstream effect: as participants felt closer to the agent, they became more comfortable sharing about themselves, resulting in various therapeutic benefits such as stress relief.

6.1.4. Proving a new framework to understand client privacy in chatbot context

The fourth theoretical contribution of my dissertation is extending research on the issue of client privacy to the context of therapeutic chatbots. Previous research has primarily focused on client disclosure decisions, indicating that clients might feel reluctant to fully open up to therapists due to their need to create favorable self-presentation (A. E. Kelly, 2000). Due to data privacy and security considerations, privacy in the context of therapeutic chatbot required a new lens (Ischen et al., 2019).

In this dissertation, I identified two different privacy-related mental models, which depended on the type of information involved. The first mental model pertains to one's personal life experiences, problems, thoughts, and feeling. I described and measured it as comfort with self-disclosure. Most participants felt more comfortable disclosing to the chatbot than via traditional channels of mental health support (i.e., human therapists, family, or friends) because they believed a chatbot would not judge them or label them, which echoed previous research (Ta et al., 2020). My analyses showed that chatbot design characteristics affected the level of comfort with self-disclosure: participants felt more comfortable disclosing to a female agent and an agent with a shared identity (e.g., race).

The second type of mental model pertains to personally identifiable information (PII), such as one's name, address, and phone number. I described and measured it as comfort with data sharing. Generally, participants didn't trust the chatbot with their PII as much as they did with traditional mental health professions because they were unsure who was collecting their data and how this data would be used. This finding echoes previous research focused on privacy concerns related to AI technology more broadly (Liao et al., 2019; Vimalkumar et al., 2021). Comfort with data sharing was influenced by participants' general privacy concerns—or their lack thereof.

Taken together, these two distinct mental models helped participants determine what to share with the chatbot therapist. These analyses expanded on previous studies on client privacy from traditional to chatbot-based psychotherapy. They also provide a framework to inform a more systematic approach to examining privacy related to chatbot technology.

6.1.5. Confirming the importance of chatbot race in therapeutic chatbot design

The fifth theoretical contribution of my dissertation is extending previous research on client preference for same race therapists to the chatbot context. My analyses confirmed that participants, regardless of their racial identity, are inclined to choose chatbot therapists with same racial background as their own. The initial evidence from my analysis of white participants further suggested that racial mirroring could facilitate both sociotechnical and therapeutic relationships between clients and agents. When paired with a same-race chatbot, white participants reported feeling closer to the agents and were more likely recommend the technology to their friends and family.

The racial-mirroring effects identified in this study present an opportunity to utilize chatbots to deliver more effective mental health support. In traditional in-person counseling, the practice of matching clients with therapists of the same race is not always feasible, because there are significantly fewer therapists of color than would be necessary to systematically implement racial/ethnic matching (APA, 2005). In contrast, implementing racial mirroring on digital platforms would be more practical with few obstacles. Racial mirroring can be a desirable feature for therapeutic chatbots, as the bonding between client and therapist plays a critical role in promoting patient adherence and improving treatment outcomes. It is important to note that racial minorities, especially African Americans, are more sensitive to the racial mirroring effect; their stronger preference for same-race therapists (Cabral & Smith, 2011) persisted in the context of human-agent interaction. Allowing clients to interact with same-race agents, or simply allowing them to choose which they want, would more substantially benefit those groups who have been marginalized and underserved in traditional psychotherapy.

Taken together, validating the racial mirroring effect in the context of therapeutic chatbots extended previous research to a novel technological context. Practically, my research compels future designs of therapeutic chatbots to consider the issue of race—I discuss this further in Section 6.3 (Design Recommendations).

6.2. A Need-based Assessment of Chatbot for Mental Health Support

With the understanding of benefits and drawbacks, as well as how people perceived and interacted with the therapeutic chatbot, I propose a need-based framework to assess chatbot technology use in the psychotherapy context. In this section, I first describe why we should apply this framework. Then I describe the rationale of this framework, drawing on previous research in traditional psychotherapy, as well as findings from my dissertation. Finally, I apply this framework to several scenarios to demonstrate different types of roles the current therapeutic chatbot can—and cannot—fill.

6.2.1. Why we need a need-based assessment to evaluate therapeutic chatbot

The main reason we need this framework is to prevent a blunt rejection or blind trust in therapeutic chatbot use. On one hand, we must recognize the tremendous value of applying chatbot technology to improve the efficiency and accessibility of mental health support. And as validated by my research empirically, chatbots can be effective for achieving positive mental health outcomes.

On the other hand, we know that current chatbot technology is far from approaching “general” or “human-level” AI. Melanie Mitchell (2021) has cautioned about the tendency to make overconfident predictions about the AI field. One common misconception is that an AI’s high performance on an intelligence test or game translates to other skills, such as perception (Mitchell, 2021). For example, Google’s AlphaGo program can beat a world champion in the board game Go; but as psychologist Gary Marcus pointed out, AI is unlikely to outperform humans in many other games. One example is charades, which “requires acting skills, linguistic skills, and theory of mind”(Marcus, 2018). The perceptual and cognitive skills humans take for granted appear to be unreachable for AI today.

Zooming in on the context of mental health support, the ability to understand and empathize is crucial, but far beyond what a therapeutic chatbot can accomplish. Therefore, I call for a holistic approach to evaluating therapeutic chatbots. This approach helps ensure an impartial evaluation of the current stage of chatbots like MICST. And it can still be applicable when there are new advancements in therapeutic chatbot technology.

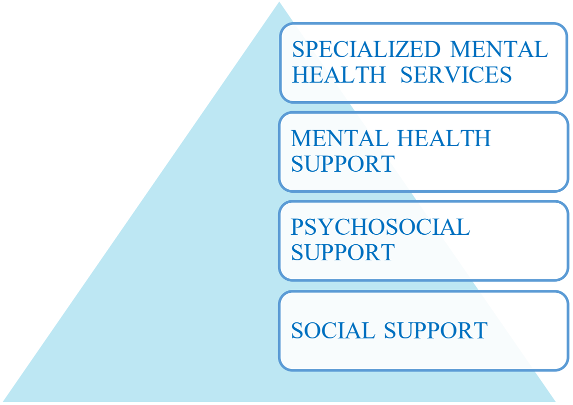
6.2.2. Rationale for the need-based framework

To develop coping mechanisms and interventions, previous research in traditional psychotherapy has explored different clusters of needs when people seek mental health support (Flanagan, 2010; Joska & Flisher, 2007). One holistic view—the “pyramid of mental health and psychosocial needs”—was proposed by the

International Committee of the Red Cross (ICRC) to address the full range of mental health and psychosocial problems (Tschakarjan & Ortega, 2019). This framework differentiates four levels of needs corresponding to interventions at each level. See Figure 7 for a visual representation of this framework.

From the bottom up, the first level of the pyramid is social support, which is the need to improve social interaction and strengthen community support mechanisms. The second level is psychosocial support, which addresses the need for support to improve psychological and social functioning and well-being through social activities that improve psychological well-being. The third level is mental health support, which addresses the need for support to reduce psychological distress and improve daily psychological and psychosocial functioning and coping strategies. The fourth level, at the top of the pyramid, is specialized mental health services, which addresses the need for specialist mental health or psychiatric care, and/or specific follow-up care.

Figure 7. The pyramid of mental health and psychosocial needs in traditional psychotherapy (Tschakarjan & Ortega, 2019)



My qualitative analyses suggested that people had a similar set of needs and expectations when interacting with a therapeutic chatbot. Table 28 below shows specific needs people had, and I map out what therapeutic chatbot can or cannot do with regard to specific mental health and psychosocial needs.

Table 28. Traditional mental health and psychosocial needs chatbot fills (or not)

Types of Needs	**Needs**	**What the chatbot can or cannot do to meet the need**
Social support	**Social interactions:** Need for support to get connection and companionship so that they don't feel alone.	Can provide warmth, companionship, and a sense of rapport.
	Social validation: Need for support to get validation for feelings and emotions.	Can provide affirmation and validation for how one feels.
Psychosocial support	**Assurance:** Need for support and assurance that they're not alone and can overcome an issue.	Can provide assurance to soothe one's feelings.
	Confidence: Need for support to increase confidence.	Can provide encouragement to increase one's confidence.
	Motivation: Need for support to get motivated for any changes.	Can provide support to increase one's motivation.
Mental health support	**Cognitive or behavioral changes:** Need for help to define problem and set a realistic goal for making changes.	Can help with problem-framing and goal setting.
	Problem-solving and/or symptom reduction: Need for support to manage mental health problems.	Can provide generic responses. Cannot provide personalized responses.
Specialized mental health services	**Get personalized advice for coping mechanism / Targeted treatment:** Need for specialized care or treatment.	Cannot provide highly specialized care without human input. Can support follow-up care.

Furthermore, I identify unique opportunities for therapeutic chatbots to meet two additional needs beyond what traditional therapy can offer. As shown in Table 29 below, one aspect is info-technological support to address the need for a digital platform to track symptoms or metrics and to access information. To fulfill this need, therapeutic chatbots can provide an interactive platform for tracking data and serves as a hub for informational resources pertinent to mental health.

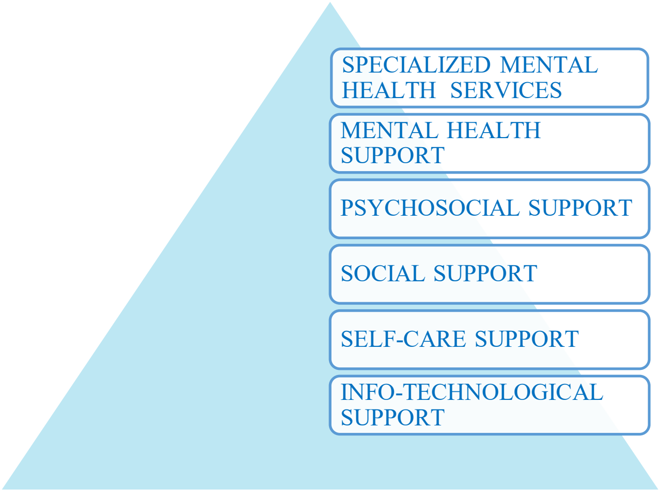
Table 29. Additional therapeutic needs chatbot fills

Types of Needs	Needs	What the chatbot can or cannot do to meet the need
Info-technological support	**Data tracking:** Need for tracking symptoms or metrics to evaluate changes in mental health.	Can provide an interactive platform to track and visualize data (e.g., mood charts).
	Informational resources: Need for accessing information about coping mechanism (for self-help) or mental health resources (e.g., local providers).	Can provide both curated and real-time information efficiently. Cannot provide very personalized information.
Self-care Support	**Self-expression:** Need for an emotional outlet where they feel heard without judgment.	Can provide a safe space and being a listener.
	Self-reflection: Need for making sense of situations and confronting the problem.	Can provide an interactive platform to help one reflect, “like a journal but responsive.”

Another aspect is self-care support to address people’s need to practice self-care by managing psychological well-being on their own terms. My analyses indicated that people gain a great sense of control when interacting with the chatbot, as it provides a safe space for self-expression and a platform for self-reflection.

Therefore, as shown in the Figure 8 below, I add two additional layers to the original ICRC framework to cover the unique set of needs in chatbot-based therapy. The new layers were added to the bottom of the framework because info-technical and self-care needs require no or little human involvement besides oneself. The new framework can be used to assess the roles chatbot can or cannot fill for addressing various therapeutic needs. It is agnostic of a chatbot’s capability or intelligence and, therefore, can be used to evaluate any therapeutic chatbot.

Figure 8. A new need-based framework in chatbot-based therapy



6.2.3. Roles therapeutic chatbot can fill

In this section, I apply the need-based framework to evaluate different roles a chatbot can fill in terms of mental health services and support. My research suggest that the best use of chatbots in psychotherapy is as a supplement to the work that human therapists do. They can serve as a bridge, they can help with various tasks of the therapy experience, and they can fill a gap when a human therapist isn't available.

A platform to document patient background information and symptoms. To respond to a person's info-technological need, the chatbot can function as a platform for tracking moods and/or symptoms, which provides an objective view of how people's psychological well-being might change over time. Beyond that role, mental health professionals can utilize chatbots to increase the efficiency of the patient intake process. Traditionally, before seeing a therapist for the first time, clients usually fill out a lengthy questionnaire asking their background information and symptoms. They may feel drained by the prolonged process before getting real help. For more efficient information gathering, mental health professionals can utilize chatbot to collect necessary information about a client. Talking with a chatbot is a more engaging and natural way to communicate as compared with answering a long survey (Wambsganss et al., 2020).

A platform for facilitating more open client disclosures. Many people feel uneasy expressing themselves due to the fear of judgment. Since people may feel more comfortable disclosing to a chatbot, mental health professionals can utilize chatbots to obtain more detailed and honest answers from clients. This allows human therapists to gain more accurate understanding of a client' problems so that they can focus on more demanding tasks such as providing personalized guidance.

An intermediate step to provide intervention and treatment for while waiting to see a human therapist. To fulfill self-care needs, chatbots can be used as an intermediate step before escalating to human therapists. As findings from this dissertation show, chatbots can help alleviate minor issues, especially if people just want to unload their feelings for emotional catharsis.

A friend in the pocket to fulfill needs for social connection. To satisfy people's need for social support, chatbots can provide companionship and social validations for people who are alone lacking interactions with others. This is an important role that regular therapists don't serve. And people sometimes can't reach their friends or family easily. In comparison, chatbot is just a button away and can become a regular part of their day.

A follow-up for ensuring adherence to treatment. To provide psychosocial support, chatbots can be utilized as a supplement for treatment. Specifically, in between sessions with a human therapist, chatbots can initiate check-ins to ensure client adhere to treatment or recommendations. This can be a cost-effective and non-imposing method for clients to get additional care in the privacy of their home.

6.2.4. Roles therapeutic chatbots cannot fill

Unfit for serving people who are technophobic. Therapeutic chatbots are unable to provide support when people are less accepting of or generally averse toward new technology. Some people simply lack confidence or trust in chatbot technology for mental health support. Due to the low expectations of chatbot usefulness, it's unlikely these people would engage with it in the beginning or sustain engagement over time. In addition, some people wanted to be less dependent on technology or use technology like apps less frequently. Chatbots are unlikely to be helpful for these groups of people.

Inadequate for addressing severe mental health issues. Due to technical limitations, chatbots can't handle cases when a client faces complicated and/or severe mental health issues. Additionally, chatbots are unable to provide highly personalized responses to address a client's specialized mental health needs. In these cases, people should focus on obtaining treatment from human mental health professionals.

6.3. Design Recommendations

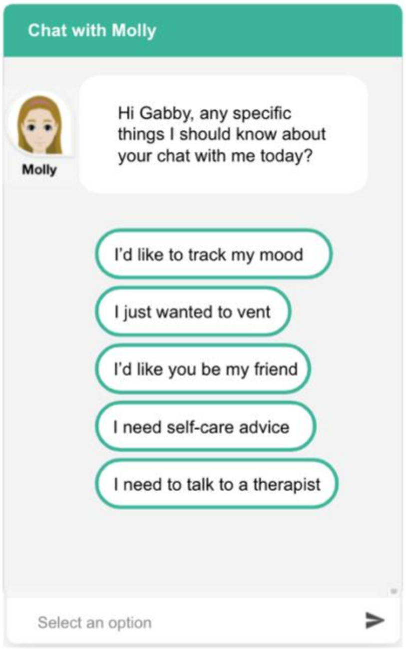
In this section, I present several design recommendations to inform the future development of therapeutic chatbots. Table 30 lists my recommendations across seven areas of improvement. Taken together, these provide directions for expanding the roles therapeutic chatbots can play to better serve people's mental health and psychosocial needs.

Table 30. Design recommendations for future development of therapeutic chatbot

Areas of Improvement	**Descriptions**
Conversation flow	Present different services based on user-identified needs/goals
Extended capability	Mood tracking and data visualization
	Provide links to relevant resources based on keywords
	Suggest coping strategies or activities
	Liaison with human therapists for more complex therapeutic needs
Alternative solutions	Provide alternative solutions as a fallback, allowing users to seek alternatives
Anthropomorphism	Have “memory” to check-in and remind people of their goals
	Provide voice responses
	Speed of responses should be similar to human speed
	More realistic agent profile (e.g., with facial expression)
	Equip chatbot with a backstory (e.g., with personality and hobbies)
Agent persona design	Different personalities based on users' goals
	Allow customization of agent identity (gender, race and age)
Relationship development	Add warm-up conversation
	Rephrase or repeat what users disclose to make them feel being heard
	Don’t force any response, giving user a sense of control
Conversation design	Apply storytelling techniques to increase engagement and to provide inspirations
	Ask open question, allow people to open up
	Follow up with more personalized responses
Data privacy	Provide transparency in what data is collected and how it will be used
	Offer an option to delete conversation history and other data retained

Tailoring conversation flow to specific user needs. The current design of chatbots primarily addresses the breadth—but not the depth—of mental health services. Chatbot-based therapy can address generic and minor issues, but is inadequate for more complex therapeutic needs. Therefore, I believe it is critical to create a conversation flow based on users' specific needs. As shown in Figure 9 below, when users first interact with the chatbot, they should be prompted to identify their objectives first, and the conversation will proceed with a service tailored to their particular needs. For instance, if a user indicated a need to monitor their progress, the chatbot can offer mood tracking. If a user indicated they just wanted to vent, the chatbot could guide them through sharing their thoughts and feelings. But if a user with a severe condition said they need more specialized mental health support, the chatbot should redirect them to human therapists. In other cases, if a user indicated they need to learn some self-care strategies, the chatbot can provide resources directly or provide initial guidance.

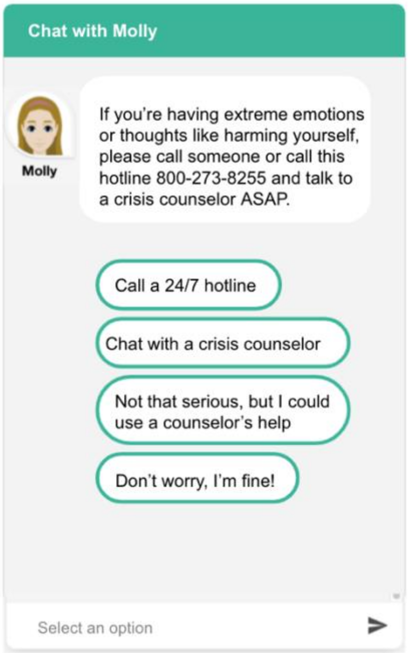
Figure 9. Design Recommendation: Tailoring conversation flow to specific user needs



Hi Gabby, any specific things I should know about your chat with me today?

Providing alternative solutions as a fallback. We must be upfront and transparent about what the system cannot achieve. One way to establish realistic expectations of the system is to provide alternative solutions for users. For example, as shown in Figure 10 below, it should be made clear to users that if they have any self-harm thoughts, they should seek professional mental health directly. At the same time, alternative solutions (e.g., a hotline number or a link to make an appointment) should be provided as a fallback.

Figure 10. Design Recommendation: Providing alternative solutions as a fallback



If you're having extreme emotions or thoughts like harming yourself, please call someone or call this hotline 800-273-8255 and talk to a crisis counselor ASAP.

Chat with Molly
Molly
Call a 24/7 hotline
Chat with a crisis counselor
Not that serious, but I could use a counselor's help
Don't worry, I'm fine!
Select an option

Adding new features to enhance chatbots' ability to provide support. Based on feedback from interviewees, I suggest adding several features to extend chatbots' capability. To provide info-technical support, chatbots can allow users to record their mental health data (e.g., symptoms or moods), and turn user-generated data into insights about their psychological well-being. Additionally, the chatbot could suggest coping strategies or activities, or provide links to relevant resources. More importantly, chatbots act as a liaison to connect users with human therapists to address more complex mental health needs.

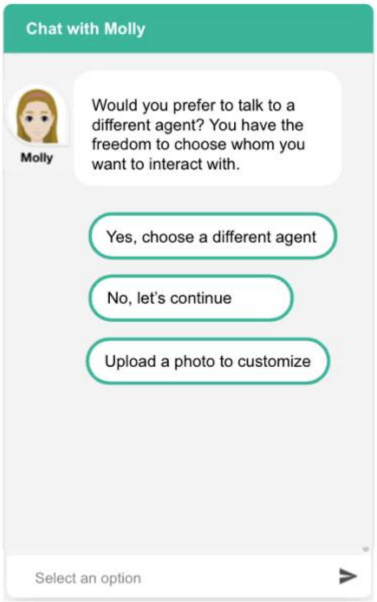
Increasing anthropomorphism features to strengthen human-agent relationships. As my analyses suggest, anthropomorphism benefits human-agent relationships. Based on feedback from interviewees, I identify several ways to increase the level of anthropomorphism, ranging from its appearance and interaction to its performance. First, the agent can have a more realistic appearance, as some participants expected the agent to have facial expressions (e.g., nodding) during conversations. In terms of interaction, the chatbot should have a speed of responses similar to human's; otherwise, people feel they're talking to a machine with canned responses. Furthermore, the conversational interface can incorporate voice responses to increase the humanness of the agent. In terms of its performance, the chatbot is expected to have "memory," or the ability to check in with users and remind them of their goals or coping strategies. Moreover, the chatbot can be equipped with a backstory like a real person with a distinct personality and hobbies.

Facilitating human-agent relationship development. There are several design opportunities to facilitate human-agent relationship development. There could be warm-up conversations like "getting to know you" or "small talk" to mimic how social interactions would unfold normally in the initial exchanges. As my research shows, people appreciate the feeling of being listened to or being heard. Therefore, after users have disclosed their thoughts and feelings in detail, the chatbot should rephrase or repeat what users reveal to signal that it's actively listening. It is also important to give users a sense of control, as they don't want to be forced to give responses during interactions and sometimes just need a moment to pause and reflect.

Customizing agent identity and personality. My research demonstrates the importance of agent persona design. People generally prefer to engage with agents with a shared identity (in gender, race, or generation), which creates a sense of familiarity and connectedness. Nonetheless, it is important not to overgeneralize and state that chatbot designers should match the race/ethnicity of agents to the users for all scenarios. This study also reveals that people's comfort with self-disclosure decreased when they interacted with same-race agents; it is possible that they might be more concerned about being judged.

Therefore, we should allow users to customize an agent's identity. For example, as shown in Figure 11 below, when users log on for the first time, they can choose an agent with whom they feel most comfortable interacting. This flexibility serves two purposes. First, users can gain a greater sense of control, which might improve their engagement with the system. Second, this configuration helps to set realistic expectations: this is a bot, not a human therapist who can provide perfect responses. Additionally, depending on users' goals, the chatbot can have different personalities or linguistic styles. For instance, the chatbot could adopt a less formal, light-hearted tone if users just wanted to learn some self-care strategies, and a more formal, yet encouraging tone if users wanted to talk about their problems.

Figure 11. Design recommendation: Customizing agent identity and personality

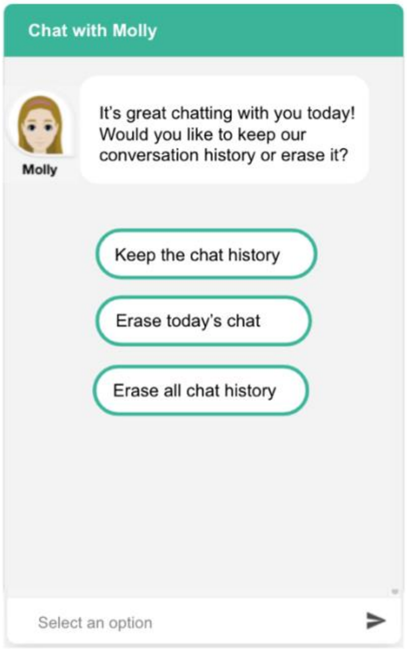


Would you prefer to talk to a different agent? You have the freedom to choose whom you want to interact with.

More engaging conversation design. It is helpful to build conversation based on established psychotherapy techniques like Motivational Interviewing. However, we should also make conversations more engaging instead of rigidly repeating a protocol. One way to achieve this is through storytelling. Chatbots can use stories or anecdotes to show how other people are managing specific issues, which provides inspiration for people. Additionally, chatbots should ask open questions to encourage people to open up while following up with personalized responses.

Providing transparency and control to mitigate data privacy concerns. My analyses show that some people had concerns about sharing data with the chatbot. To mitigate data privacy and security concerns, therapeutic chatbots should provide transparency regarding what data is collected and how it will be used. Additionally, as shown in Figure 12 below, the chatbot can provide users an option to review data collected and to delete their conversation history and any other user data retained in the system.

Figure 12. Design Recommendation: Providing transparency and control to mitigate data privacy concerns



It's great chatting with you today! Would you like to keep our conversation history or erase it?

6.3. Ethical Considerations

There are a few ethical considerations related to therapeutic chatbot development and utilization. In this section, I discuss the implications of the digital divide, how over-reliance on chatbots could be problematic for achieving health justice, and the implications of considering race in chatbot design Additionally, I call for consideration of human therapist acceptance of therapeutic chatbots.

Digital divide limiting chatbot potential. The digital divide could perpetuate the treatment gap for disadvantaged patients without access to traditional mental health services. The promise of chatbot-based therapy reaching a wider audience is based on technology access. However, people in rural areas and less developed countries or regions don't have equal access to computers, smartphones, or the internet (Ennis et al., 2012). Additionally, granted access to technology, some groups could have trouble navigating the system due to lower technology literacy, which lowers the efficacy of e-mental health solutions (Stone & Waldron, 2019). These digital divide issues could systematically limit the potential of utilizing therapeutic chatbots. People who need this technology the most might have the most difficulty accessing it.

The pitfall of over-reliance on chatbot. Chatbots can be human-like, but they are not human. Due to technical constraints, chatbot therapists are unable to replace human therapists. And while chatbots may be sufficient for those seeking “social snacking” to satiate a need for human interaction, this anthropomorphized technology can’t generate a real sense of belonging. Relatedly, chatbots can provide therapeutic help, but there is a risk that some people might end up turning away from human therapists.

While continuing to improve chatbot systems for mental health support, we should avoid over-reliance on technological solutions to “fix” the healthcare system’s affordability and accessibility problems. Society at large should not evade the responsibility of health justice through creating a more efficient and egalitarian healthcare system.

Racial-mirroring effect implications. Overall, to practice inclusive design, we should be racially conscious when designing an intelligent system to draw on the full range of human diversity. As chatbot entities and other forms of conversational agents become more prevalent, it is critical to integrate considerations of race at both algorithm and interface levels (Liao & He, 2020). In the context of psychotherapy, same-race agents could help address the real-world challenge of the shortage of mental health professionals who are racial minorities. And designing chatbot persona in a racially diverse way is the first step to practice inclusive design. However, we should not equate this as a real victory. We still need to address the lack of more diverse workforce in mental health services.

Some human therapists might not be accepting of chatbots. Although this dissertation is focused on client perspectives, we should also consider human therapists' perceptions of therapeutic chatbots. To create a strong therapeutic relationship with the therapist, ideally clients and therapists should be aligned on whether to include the use of e-mental health solutions. However, a recent study suggests that some primary care physicians might be resistant to AI technology due to a sense of "threat to professional autonomy" (Wang et al., 2021). It is unclear how this sentiment manifests in the mental health care domain, and this should be explored in future studies.

Chapter 7: Conclusion

7.1. Summary of Dissertation

Situated at the intersection of psychology and HCI, this dissertation details the design and evaluation of a chatbot for mental health support. I designed an original chatbot system with the ability to carry out psychotherapeutic conversations. Through an online experiment and in-depth interviews, I first evaluated the benefits and drawbacks of this technology, as well as how people perceived and interacted with therapeutic chatbot. Chatbot technology promises more natural and engaging ways of interactions compared to other automated e-therapy platforms. Despite the lack of specialized mental health services comparable to human therapists, a therapeutic chatbot can provide social support and meet basic psychosocial needs.

To further unpack how and why a therapeutic chatbot could be engaging and effective, I examined human-agent relational dynamics conceptualized as sociotechnical and therapeutic relationships. My analyses confirmed that chatbot anthropomorphism greatly influenced human-agent relationships, as people feel more engaged and closer to an agent when it's considered more human-like. My analyses also revealed the continuum of racial-mirroring effects in human-agent interactions: People prefer to interact with chatbots with shared racial identity. Additionally, I identified two distinct mental models pertaining to chatbot use. People feel more comfortable sharing their personal experiences and feelings with chatbots than with humans, because they know the chatbot wouldn't judge them. On the other hand, some people don't trust digital technology with their personal identifiable information.

This research offers both practical design and theoretical contributions. Built upon the comprehensive analyses, I have proposed a need-based framework to evaluate the roles of therapeutic chatbots. I also identified several design opportunities to inform future development of therapeutic chatbots. Some research contributions went beyond the mental health context. Specifically, this research extended the technology acceptance model (UTAUT) by incorporating the role of anthropomorphism.

In conclusion, we should be conditionally positive about chatbot applications in mental health context. We should continue to improve therapeutic chatbot design to improve the efficiency and efficacy of mental health services.

7.2. Limitations and Future Research

This dissertation represents one of the first attempts in HCI to evaluate chatbot applications in the mental health context. To aid future research efforts, I identify several limitations of this study regarding methodology and scope. First, both the online experiment and the interview study offer a single moment in time look into participants' perceptions and interactions with the chatbot. Even though participants were asked to predict their future relationship with the agent, there is no way to know how the psychological and relational outcomes of therapeutic chatbot use would actually unfold over time. A longitudinal study is needed for future exploration of therapeutic chatbot use.

Second, the therapeutic benefits revealed by this research were primarily drawn from qualitative self-reported data based on interview and survey responses. There might be a risk of social desirability bias. Future work should further validate and quantify chatbot's therapeutic benefits through repeated measures before and after a single chatbot session and at the end of multiple sessions.

Third, recruiting from online crowdsourcing platforms like Amazon Mechanical Turk and Prolific limited the generalizability of my findings. Older adults and people of color were underrepresented in this sample. Many participants seemed to identify financial stress as their most significant challenge, which might have biased their perceptions of the chatbot's usefulness for therapeutic purposes. Therefore, future work should improve the sampling method by reaching out to more diverse participants who are underrepresented on the online crowdsourcing platforms.

Fourth, the current research only examined a text-based conversational system in the context of psychotherapy. Future work should further explore the conversational system incorporating voice-based interactions that could facilitate user engagement closer to a traditional talk therapy.

Finally, in terms of the research scope, the current research took a client-centric perspective to evaluate therapeutic chatbot use. As mental health professionals are an integral part of the mental health care ecosystem, future research should also assess their perceptions of therapeutic chatbots and whether they would be open to exploring potential applications of technology. More importantly, our design of therapeutic chatbots should be sure to keep humans, clients and professionals, in the loop.

Appendices

Appendix I Example Dialogue

An example dialogue based on a real participant's interactions with Molly.

	Main Conversation Flow
Molly	Welcome! 👋 I'm Molly
QR button	Hi Molly
User - Click button	Hi Molly
Molly	Nice to meet you!
	I know I'm not real, but I'm here to listen to you and help you through life.
	Think of me as a texting buddy that gets better every time we talk!
	What's your name? I won't ask your real identity; all I need is a nickname for you.
User - Text input	Shelley
Molly	Shelley, my goal today is to understand your challenge in life and any changes you'd need.
	Are you ready for our session?
User - Click button	Yes!
Molly	Could you tell me about your biggest challenge recently?
	You can select from the options provided.
QR buttons	Finance; Health; Job-related; Negative thoughts; Relationship; Education; Other
User - Click button	Health
Molly	You are not alone in this. Many people seem to face this particular issue.
	Help me understand it better, so far, how confident are you in your ability to overcome this challenge?
	Please rate it from 1 (=😧not confident at all) to 7 (=😇very confident).
QR button	1;2;3;4;5;6;7
User - Click button	5
Molly	It seems like you have a lot of courage! That's a good sign.
	I have a lot of confidence that working together we can make it even better.💪
	Would you like to continue our conversation?
QR button	Sure! ; How Does it work
User - Click button	Sure!
Molly	I hope so far you feel comfortable chatting with me
	Now, tell me about your challenge in details.
	What happened? How long has it been? How has it influenced your life?
User - Text input	My husband has had prostate surgery and is now on bed rest for a full month and I need to take care of him and also work.

Molly	How do you feel about all this?
	You can trust me with your feelings and thoughts.
User - Text input	I am a bit frustrated and worried about him
Molly	I understand things could be difficult. You're not alone.
	I wonder if you've tried to work on this issue?
User - Text input	Yes
Molly	With all the obstacles you have, you certainly showed a lot of strength. 👍
	Help me here, how would you like things to turn out for you, ideally? What's your goal?
User - Text input	That he will recover and we can go back to living a normal life
Molly	What would you say is the most important thing that you can do to achieve this goal?
User - Text input	Keep him on his medication and bed rest
Molly	Please go on. I am with you.
	Could you say some more about that? What're the obstacles you might have encountered or foresee?
User - Text input	not healing properly
Molly	People say an obstacle is often a stepping stone. Don't let it discourage you.
	Let's talk about your strength.
	What strong points do you have that could help you move forward?
User - Text input	I have a strong faith and know that God will help us through this
Molly	Anything else that you want to add? Like, support from family or friends?
	I want you to focus on things you already have to handle the situation.
User - Text input	nothing else
Molly	Based on what you told me, you have lots of what it takes to overcome the barrier. Keep it up! 😊
	Today, I appreciate your willingness to share so openly about your challenge. 🤝
	Before I wrap up our session, can you name one specific thing that you can do to overcome the challenge? (could be something small, like an 🍎 a day 😉)
User - Text input	be patient and trust God
Molly	👏👏👏
	I really enjoy chatting with you today. I wish you all the best. Be strong & prosper. 🖖
QR button	Bye!; Thanks, I'll miss you.
User - Click button	Bye!

<u>Appendix II: </u>Screening Survey

Chatbot - Screening

Q1 I confirm I live in the United States and have a good proficiency in English.

☐ Yes (4)

☐ No (5)

Q2 **Description of Study & Consent**
We are conducting a user study about our newly design conversational artificial intelligence (AI). This technology aims to help people manage their psychological well-being. This study contains the following three components and will take about 15 to 25 minutes.
Short screening assessment: For eligible participants, we will provide a link for you to interact with the AI. You will be asked about your life experiences, as well as your feeling and thoughts.
Main survey: You must be at least 18 years old to participate in this study. If you decide to stop taking part in the study, if you have questions, concerns, or complaints, or if you need to report an injury related to the research, please contact the principal investigator: skyunlimited2018@gmail.com.

You indicate your voluntary agreement to participate in this research and have your answers included in the data set by completing and submitting this online survey.

☐ Yes (1)

☐ No (2)

Q3 What is your sex?

☐ Male (1)

☐ Female (2)

Q4 Which categories of race or ethnicity describe you? Please select all that apply.

☐ White (1)

☐ Hispanic or Latinx (2)

☐ Black or African American (3)

☐ Asian (4)

☐ Native Hawaiian or other Pacific Islander (5)

☐ Native American or Alaska Native (6)

☐ Other (99) ____

Display This Question:

If Which categories of race or ethnicity describe you? Please select all that apply. = Asian

Q5 Where would you trace your cultural heritage back in Asia?

☐ Eastern Asian region (e.g., China, Korea, Japan, Mongolia) (1)

☐ South Asian region (e.g., India, Pakistan, Bangladesh, Nepal, Sri Lanka, Afghanistan) (2)

☐ Mainland Southeast Asian region (e.g., Thailand, Philippines, Vietnam, Laos, Cambodia, Myanmar) (3)

☐ Maritime Southeast Asia region (e.g., Philippines, Indonesia, Malaysia) (4)

Q6 Below are five statements with which you may agree or disagree about your life. Indicate <u>how much you agree or disagree</u> with each item. Please be open and honest in your responses.

	Strongly disagree (1)	Disagree (2)	Neutral (3)	Agree (4)	Strongly Agree (5)
In most ways, my life is close to my ideal. (1)	☐	☐	☐	☐	☐
The conditions of my life are excellent. (4)	☐	☐	☐	☐	☐
I am satisfied with my life. (5)	☐	☐	☐	☐	☐
So far, I have gotten the important things I want in life. (6)	☐	☐	☐	☐	☐
If I could live my life over, I would change almost nothing (7)	☐	☐	☐	☐	☐

Q8 People sometimes look to others for companionship, assistance, or other types of support. <u>How often</u> is each of the following type of support available to you if you need it?

	Never (1)	Sometimes (2)	About half the time (3)	Most of the time (4)	Always (5)
Someone you can count on to listen to you when you need to talk (1)	☐	☐	☐	☐	☐
Someone to give you information to help you understand a situation (2)	☐	☐	☐	☐	☐
Someone to give you good advice about a crisis (4)	☐	☐	☐	☐	☐
Someone to confide in or talk to about yourself or your problems (3)	☐	☐	☐	☐	☐
Someone to share your most private worries and fears with (6)	☐	☐	☐	☐	☐
Someone to turn to for suggestions about how to deal with a personal problem (7)	☐	☐	☐	☐	☐
Someone who understands your problems (8)	☐	☐	☐	☐	☐
Someone to help you if you were confined to bed (9)	☐	☐	☐	☐	☐
Someone to take you to the doctor if you needed it (10)	☐	☐	☐	☐	☐

Someone to prepare your meals if you were unable to do it yourself (11)	☐	☐	☐	☐	☐
Someone to help with daily chores if you were sick (12)	☐	☐	☐	☐	☐

Q9 Indicate your <u>level of concern</u> about the following things that might happen <u>when you use communication technologies (social media, email, apps).</u>

	Not at all concerned (1)	Slightly concerned (2)	Somewhat concerned (3)	Moderately concerned (4)	Extremely concerned (5)
Your personal account information being compromised (e.g., your email and password get hacked) (2)	☐	☐	☐	☐	☐
Your personal information (e.g., phone number, address, etc.) becoming publicly visible (3)	☐	☐	☐	☐	☐
Private messages becoming publicly visible (7)	☐	☐	☐	☐	☐
Your employer viewing content (text or photos) that might negatively impact your job (9)	☐	☐	☐	☐	☐
Someone posting a mean, unflattering, or factually incorrect update about you (10)	☐	☐	☐	☐	☐
Your personal information being sold to other companies for marketing purposes (11)	☐	☐	☐	☐	☐

Q10 Thinking about your general attitudes toward technology, please rate how much you agree or disagree with the following statements.

	Strongly disagree (1)	Somewhat disagree (2)	Neither agree nor disagree (3)	Somewhat agree (4)	Strongly agree (5)
I believe that most technologies are effective at what they are designed to do. (1)	☐	☐	☐	☐	☐
A large majority of technologies are excellent. (4)	☐	☐	☐	☐	☐
Most technologies have the features needed to be useful (5)	☐	☐	☐	☐	☐
I think most technologies enable me to do what I need to do. (6)	☐	☐	☐	☐	☐
My typical approach is to trust new technologies until they prove to me that I shouldn’t trust them. (7)	☐	☐	☐	☐	☐
I usually trust a technology until it gives me a reason not to trust it. (8)	☐	☐	☐	☐	☐
I generally give a technology the benefit of the doubt when I first use it. (9)	☐	☐	☐	☐	☐

Q11 Congratulations!

You are qualified for our main research study.

Here's what you need to do (PLEASE READ):
Click ${e://Field/ResponseID}">**this link** to chat with Jake, a conversational artificial intelligence (AI) that is designed to help people manage their psychological well-being. You will be asked about your life experiences, as well as your feeling and thoughts. Please complete the conversation with Jake within 15 minutes to avoid any technical issues.

At the end of your conversation, Jake will send you a link to the main survey. You will answer questions related to the overall experience and your perceptions of Jake.

Appendix III: Main Survey

Chatbot - Main Survey [completed after interaction with chatbot]

Q1 Great! You've made it to the main survey, the last step of our study.
We appreciate your continued interest!

Q2 What's the name of the conversational AI that you just interacted with?

☐ Jake

☐ David

☐ Robbie

☐ Antonio

☐ Jenny

☐ Molly

☐ Camila

☐ Ebony

☐ Darnell

Q43 Do you think the conversational AI has the same race/ethnicity as yours?

☐ Yes (1)

☐ No (2)

Display This Question:

If What's the name of the conversational AI that you just interacted with? = Jake

Q3 Is this the AI you talked to in the previous therapy session?

☐ Yes (7)

☐ No (8)

Display This Question:

If What's the name of the conversational AI that you just interacted with? = Antonio

Q4 Is this the AI you talked to in the previous session?

☐ Yes (7)

☐ No (8)

Display This Question:

If What's the name of the conversational AI that you just interacted with? = David

Q5 Is this the AI you talked to in the previous session?

☐ Yes (7)

☐ No (8)

Display This Question:

If What's the name of the conversational AI that you just interacted with? = Darnell

Q6 Is this the AI you talked to in the previous session?

☐ Yes (7)

☐ No (8)

Display This Question:

If What's the name of the conversational AI that you just interacted with? = Robbie

Q7 Is this the AI you talked to in the previous session?

☐ Yes (7)

☐ No (8)

Display This Question:

If What's the name of the conversational AI that you just interacted with? = Jenny

Q8 Is this the AI you talked to in the previous session?

☐ Yes (7)

☐ No (8)

Display This Question:

If What's the name of the conversational AI that you just interacted with? = Molly

Q9 Is this the AI you talked to in the previous session?

☐ Yes (7)

☐ No (8)

Display This Question:

If What's the name of the conversational AI that you just interacted with? = Camila

Q10 Is this the AI you talked to in the previous session?

☐ Yes (7)

☐ No (8)

Display This Question:

If What's the name of the conversational AI that you just interacted with? = Ebony

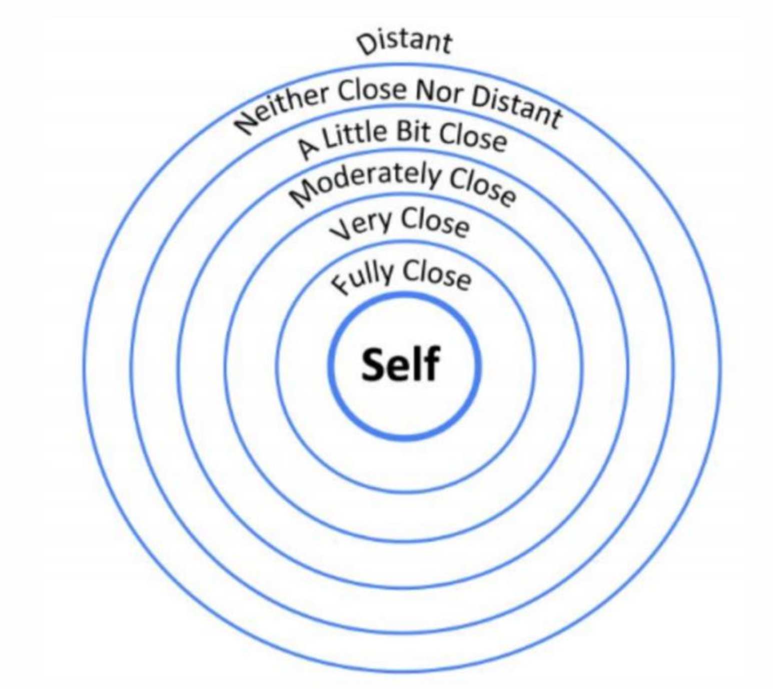
Q11 Is this the AI you talked to in the previous session?

☐ Yes (7)

☐ No (8)

Q12 *Feeling close* refers to being listened to, understood by, able to share feelings and to talk openly with another person. Please refer to the graph below with circles around the Self. Different distances to the center (the Self) indicates different levels of closeness.

Now, think about your interaction with [Agent Name]. How close did you feel to the conversational AI? Please click on the picture to indicate your relationship with the agent.

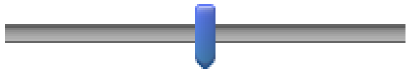


Q13 Compared to a conversation with your friends, family, or your counselor, please rate how much you agree or disagree with the following statements regarding your conversation with [Agent Name].

	Strongly Disagree (1)	Disagree (2)	Neutral (3)	Agree (4)	Strongly Agree (5)
I felt comfortable when I talked with [Agent Name] about my challenge and life experiences. (1)	☐	☐	☐	☐	☐
What I disclosed in the conversation is sensitive. (2)	☐	☐	☐	☐	☐
I would not have talked about this topic with my friends or family. (3)	☐	☐	☐	☐	☐
I expressed myself openly. (4)	☐	☐	☐	☐	☐
I felt [Agent Name] wouldn't judge me when I talked about my issue. (5)	☐	☐	☐	☐	☐
My ability to communicate issues increased when I interacted with [Agent Name]. (6)	☐	☐	☐	☐	☐
I felt less willing to talk about my challenges. (7)	☐	☐	☐	☐	☐

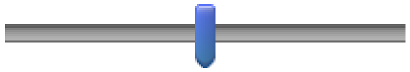
Q14 Overall, how sensitive would you rate the content of your conversation with [Agent Name]to be?

	Not Sensitive at All	Somewhat Sensitive	Highly Sensitive



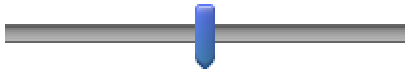
Q15 How would you rate the level of engagement of your conversation with [Agent Name]?

	Not Engaging at All	Somewhat Engaging	Highly Engaging



Q16 How much do you like [Agent Name]?

	Dislike	Neutral	Like

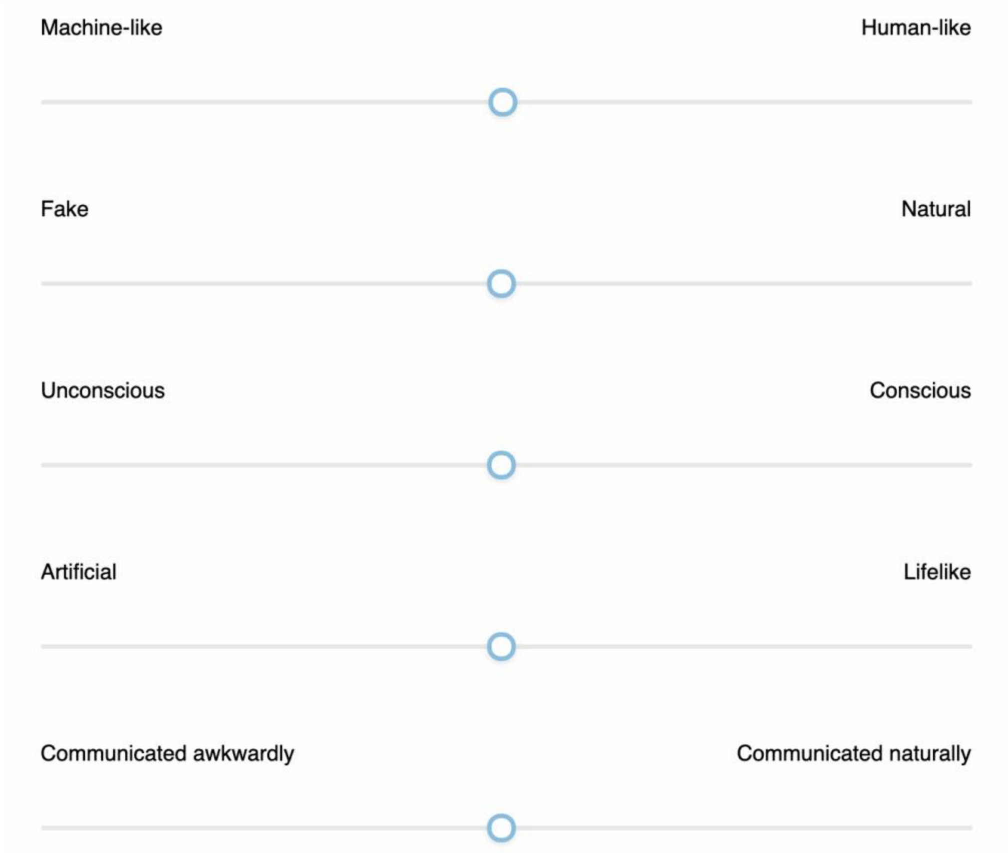


Q17 Please rate how much you agree or disagree with the following statements regarding your thoughts and feelings about interacting with [Agent Name].

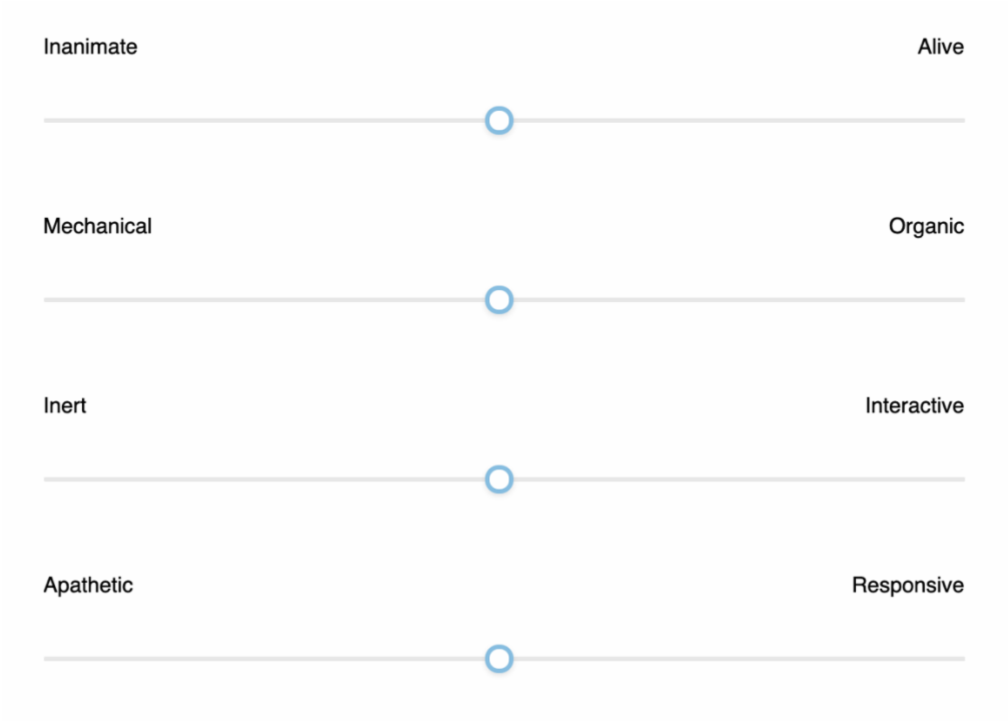
	Strongly Disagree (1)	Disagree (2)	Neutral (3)	Agree (4)	Strongly Agree (5)
I felt uncomfortable interacting with [Agent Name]. (1)	☐	☐	☐	☐	☐
I was worried that going through talk sessions like this wouldn't be useful for me. (4)	☐	☐	☐	☐	☐
What I was doing in the talk session gave me new ways of looking at my problem. (6)	☐	☐	☐	☐	☐
I believe the way we were working with my problem is correct. (5)	☐	☐	☐	☐	☐
[Agent Name] perceived accurately what my goals are (7)	☐	☐	☐	☐	☐
I find what I was doing in the talk session confusing. (8)	☐	☐	☐	☐	☐
I find what we were doing during the talk session was unrelated to my concerns. (9)	☐	☐	☐	☐	☐
I feel that this conversation with [Agent Name] helped me to accomplish the changes that I wanted. (10)	☐	☐	☐	☐	☐

As a result of the conversation, I became clearer as to how I might be able to change. (11)	☐	☐	☐	☐	☐
[Agent Name] and I collaborated on setting goals in order to solve my problem. (12)	☐	☐	☐	☐	☐
The things that [Agent Name] asked me didn’t make sense. (13)	☐	☐	☐	☐	☐
We have established a good understanding of the kind of changes that would be good for me. (14)	☐	☐	☐	☐	☐

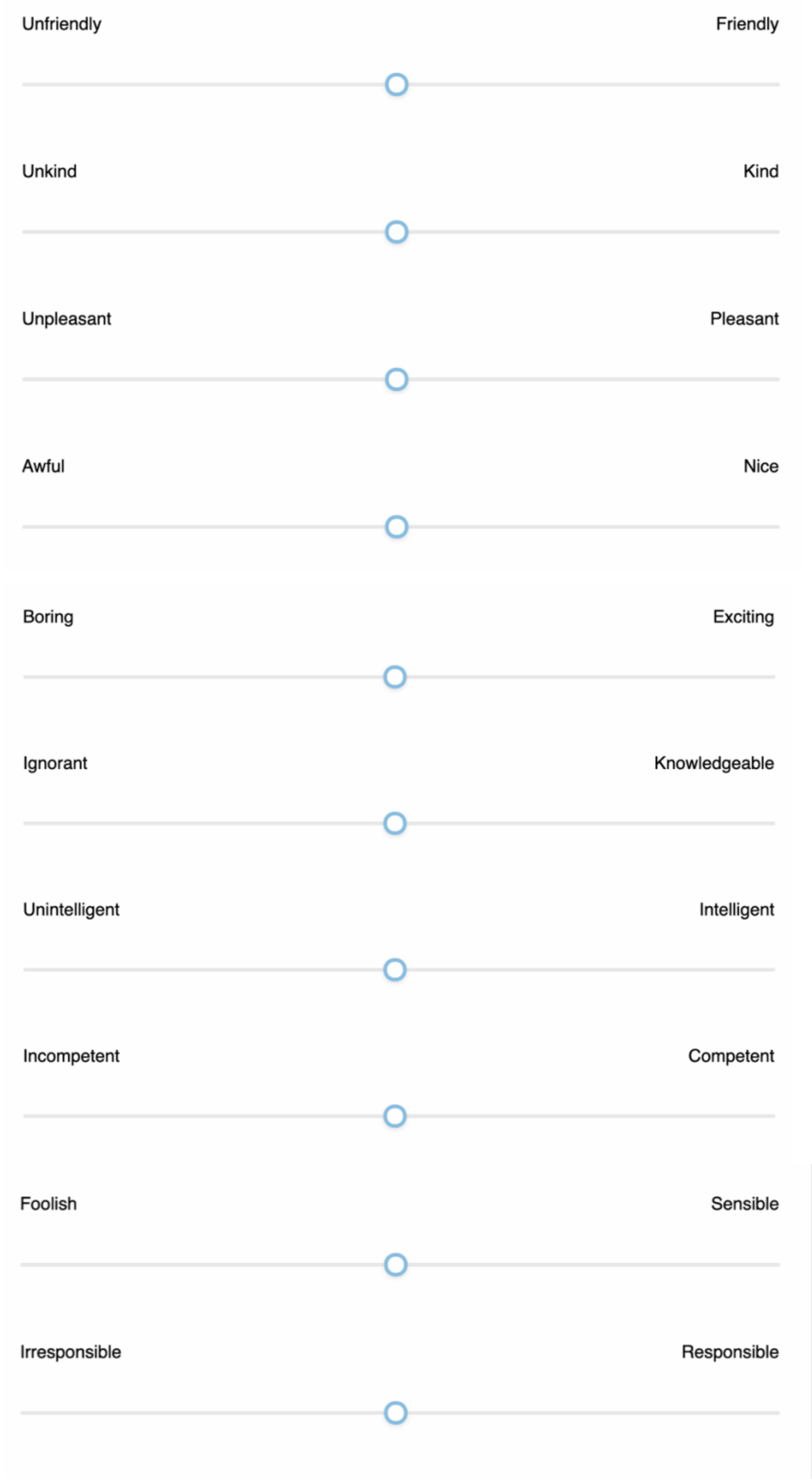
Q18 Please rate your impression of [Agent Name] on these scales.



Machine-like	Human-like
Fake	Natural
Unconscious	Conscious
Artificial	Lifelike
Communicated awkwardly	Communicated naturally



Inanimate	Alive
Mechanical	Organic
Inert	Interactive
Apathetic	Responsive



Unfriendly	Friendly
Unkind	Kind
Unpleasant	Pleasant
Awful	Nice
Boring	Exciting
Ignorant	Knowledgeable
Unintelligent	Intelligent
Incompetent	Competent
Foolish	Sensible
Irresponsible	Responsible

Q19 Overall, how much would you like to continue working with [Agent Name] in the future?

	Not at All	Maybe	Very Much Willing

Q20 How likely would you be to recommend [Agent Name] to your friends or family if they need help managing their mental well-being?

	Would not recommend at all	Neutral	Would definitely recommend

Q21 How likely would you be to recommend [Agent Name]to people who have a need for counseling/psychotherapy?

	Would not recommend at all	Neutral	Would definitely recommend
Click to write Choice 1 ()			

Q22 Please rate how much you agree or disagree with the following statements regarding your interaction with [Agent Name].

	Strongly Disagree (1)	Disagree (2)	Neutral (3)	Agree (4)	Strongly Agree (5)
My interaction with [Agent Name] was satisfying (1)	☐	☐	☐	☐	☐
This experience interacting with [Agent Name] is fascinating (22)	☐	☐	☐	☐	☐
I feel I've learned something from this conversation (23)	☐	☐	☐	☐	☐
After my conversation with [Agent Name] I felt more motivated (20)	☐	☐	☐	☐	☐
After my conversation with [Agent Name] I felt more confident about my ability to overcome the challenge (21)	☐	☐	☐	☐	☐
This technology is useful for me to manage my mental well-being (19)	☐	☐	☐	☐	☐
I prefer to talk to a human therapist instead (16)	☐	☐	☐	☐	☐
I got bored quickly when using this technology (17)	☐	☐	☐	☐	☐
The persona of [Agent Name] was unappealing (5)	☐	☐	☐	☐	☐

This conversational AI operates reliably (6)	☐	☐	☐	☐	☐
This experience would be less awkward than a therapy session in person (18)	☐	☐	☐	☐	☐
Overall, this conversational AI perform well without glitch (7)	☐	☐	☐	☐	☐
Overall, I think it is easy to use this technology (14)	☐	☐	☐	☐	☐
Having a conversation with [Agent Name] is intuitive (15)	☐	☐	☐	☐	☐

Q23 In order to offer more meaningful and personalized replies to your messages, the AI might need to collect some information about you. How comfortable do you feel about sharing your data with [Agent Name]? Please rate your level of comfort with sharing each type of data.

	Absolutely Not Comfortable Sharing	Somewhat Comfortable	Completely Comfortable Sharing

Email address ()	
Phone number ()	
Birthday ()	
Location data ()	
Phone call history (i.e. when and whom have you contacted) ()	
Your employment history ()	
Your photos and videos ()	
Medical history ()	
Sexual activity ()	
Relationship status ()	
Sexual orientation ()	
Disease-related information (e.g., condition and symptoms) ()	
Mood tracking ()	

Q24 Please rate how much you agree or disagree with the following statements regarding your interaction with [Agent Name].

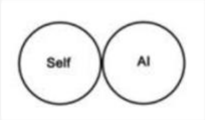
	Strongly Disagree (1)	Disagree (2)	Neutral (3)	Agree (4)	Strongly Agree (5)
People who are important to me would be very supportive of my use of this technology (1)	☐	☐	☐	☐	☐
My friends or family would think my use of this technology is a good idea (2)	☐	☐	☐	☐	☐
Using this technology will impress other people around me (3)	☐	☐	☐	☐	☐
Chatting with [Agent Name] takes too much time from my routine (4)	☐	☐	☐	☐	☐
I have the knowledge or skills necessary to use this technology (5)	☐	☐	☐	☐	☐
Using the technology is compatible with the way I deal with my challenges (6)	☐	☐	☐	☐	☐
People would not judge me if they see me interacting with [Agent Name] (7)	☐	☐	☐	☐	☐

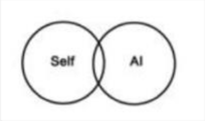
Q25 Envision that you'll be interacting with [Agent Name] at least three times a week, please indicate how much you agree or disagree with the following statements:

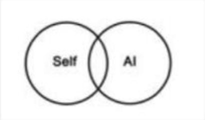
	If I continue to engage in a conversation session with [Agent Name] at least three times a week...				
	Strongly Disagree (1)	Disagree (2)	Neutral (3)	Agree (4)	Strongly Agree (5)
I will be more prepared to tackle my challenges in life (1)	☐	☐	☐	☐	☐
I will increase the level of my psychological well-being (2)	☐	☐	☐	☐	☐
I will increase my confidence in dealing with difficult situations (3)	☐	☐	☐	☐	☐
I will increase my ability to overcome personal obstacles (4)	☐	☐	☐	☐	☐
I'll be ready to make a change into a better version of myself (5)	☐	☐	☐	☐	☐

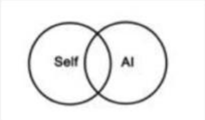
Q26 Suppose that you'll be interacting with [Agent Name] at least three times a week in the future. Please select the figure that best describes <u>the potential of your</u>

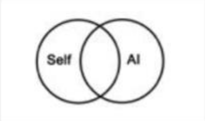
relationship with the AI. The greater degree of the overlapping between the two circles indicates greater closeness between you and the AI.

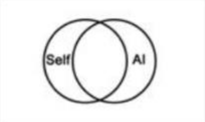
○ 

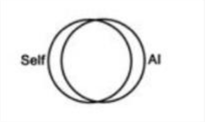
○ 

○ 

○ 

○ 

○ 

○ 

27 AI technology is constantly being improved, particularly in the healthcare domain. Please rate how much you agree or disagree with the following statements regarding the use of AI for mental well-being management.

	Strongly Disagree (1)	Disagree (2)	Neutral (3)	Agree (4)	Strongly Agree (5)
Using this AI technology in counseling and psychotherapy is a good idea (1)	☐	☐	☐	☐	☐
AI might become the perfect friend, there to listen whenever we need and respond to our every desire. (2)	☐	☐	☐	☐	☐
AI might replace human therapists in the future (3)	☐	☐	☐	☐	☐
AI might cater to all our desires so well that we prefer AI interaction to human interaction (4)	☐	☐	☐	☐	☐
AI companionship might be very popular (5)	☐	☐	☐	☐	☐
I’m excited for the future that we can do a lot of things with the help of artificial intelligence (6)	☐	☐	☐	☐	☐

Q28 Have you ever received any psychological counseling or treatment, such as talking therapy/psychotherapy?

☐ Never (1)

☐ Yes, but not anymore (2)

☐ Yes, I'm currently in therapy (3)

Display This Question:

If Have you ever received any psychological counseling or treatment, such as talking therapy/psychot... = Yes, but not anymore

Or Have you ever received any psychological counseling or treatment, such as talking therapy/psychot... = Yes, I'm currently in therapy

Q29 How many times in total have you attended counseling or psychotherapy sessions?

☐ Only once or twice (1)

☐ Three to five times (2)

☐ Six to nine times (3)

☐ Ten times or more (4)

Q44 Have you ever interacted with AI-based technology, such as Siri on an iPhone, Alexa on a smart speaker, or the chatbot on a website?

☐ Yes (1)

☐ No (2)

Display This Question:

If Have you ever interacted with AI-based technology, such as Siri on an iPhone, Alexa on a smart sp... = Yes

Q45 Generally, how satisfied did you feel about your previous interaction with other AI-based technology?

Not satisfied at all ------ Very satisfied

Q30 What is the last grade or class that you completed in school?

☐ Lower than high school (1)

☐ Some high school (2)

☐ High school graduate/GED (3)

☐ Professional degree (4)

☐ Some college, non 4-year degree (5)

☐ Bachelor's degree / College graduate (6)

☐ Master's degree (7)

☐ Doctorate (8)

☐ Other (9) ________________

☐ Prefer not to answer (99)

Q31 What is your age in years?

____

Q32 What is your annual household income?

☐ Less than $10,000 (1)

☐ $10,000 - $19,999 (2)

☐ $20,000 - $29,999 (3)

☐ $30,000 - $39,999 (4)

☐ $40,000 - $49,999 (5)

☐ $50,000 - $59,999 (6)

☐ $60,000 - $69,999 (7)

☐ $70,000 - $79,999 (8)

☐ $80,000 - $89,999 (9)

☐ $90,000 - $99,999 (10)

☐ $100,000 - $149,999 (11)

☐ More than $150,000 (12)

Q33 How would you describe your race or ethnicity?

☐ White (1)

☐ Black or African American (2)

☐ Hispanic or Latinx (3)

☐ Asian (4)

☐ Other (5)

☐ Prefer to self-describe (6) ____

Q34 We will have a follow up interview study. Each participant selected for an interview will receive a $25 Amazon gift card. Please enter your email address if you would be interested in participating.

____

Q35 As an AI research team, we want to improve on the technology of conversational AI. Please provide your thoughts or suggestions based on your experience interacting with [Agent Name]. We will appreciate your feedback very much!

____

____

Q36 If you were able to choose the persona of the AI, which of the following profile would you prefer to interact with for a talk session regarding mental well-being management?

☐ Jenny (1)

☐ David (4)

☐ Robbie (5)

☐ Antonio (6)

☐ Ebony (10)

☐ Darnell (11)

☐ Molly (7)

☐ Camila (8)

☐ Jake (9)

Q42 Please briefly explain the reasoning of your choice(s).

____

Appendix IV: Interview Study Protocol

Introduction: Thank you very much for participating in the online study earlier. In today's interview, I want to get a more detailed understanding of your thoughts about this technology for managing mental health. This interview will be audio-recorded, as you agreed in the consent form. Before we start, do you have any questions?

To begin with, I'd like to understand some of your past experiences with conversational AI technology.

- Have you ever interacted with a conversation-based technology before, such as a chatbot or voice assistant like Siri or Alexa?
    - If so, for what purposes do you use this type of technology?

I'd like to understand how you normally deal with psychological issues, like stress, anxiety or depression.

- Have you ever seen a psychiatrist or psychologist?
- If so, what types of factors were important to you when deciding who to go see?
- How satisfied were you with your experiences? Why?
- Everybody has down moments once in a while. What are your coping strategies when you're not feeling well mentally, for example, when you feel overwhelmed, stressful, anxious or depressed?
- Do you usually share your feelings with anybody else? If so, who do you tend to share them with?

Let's talk about the chatbot you interacted with earlier. [Show the agent].

- What are the three adjectives that first come to your mind when thinking about your interactions with [agent name]?
- What are your overall impressions of [agent name]?
- What did you like the best about interacting with [agent name]? Any enjoyable moments?
- What did you like the least? Any frustrations or confusions in this process?
- How much do you think chatting with [agent name] would help you to cope with the challenges in your life? If helpful, in what way?
- What might help to make interacting with [agent name] more useful for you?
- How do you feel about interacting with [agent name] to manage your mental health?
- What are the benefits associated with this conversation experience?
- What are the drawbacks associated with this conversation experience?
- Did you feel understood by [agent name]? Why or why not?
- How would you compare this conversation with [agent name] to your conversations with the people in your life who you usually turn to in order to share your thoughts and ask for advice?
- Imagine that you'd be using this chatbot technology frequently, let's say, at least three times per week. How would you feel about that?
- Would you be interested in using this technology to help you manage your mental health? Why or why not?
    - If yes, how would you integrate this technology into your life?

I'd like to understand how comfortable (or not) you feel about disclosing your information to this chatbot.

- Thinking about your conversation with [agent name], compared to interacting with a human listener, would you say you feel more willing to share your thoughts and feelings with [agent name], or less inclined to do so?
- Imagine that you're using this chatbot mobile app frequently, let's say, at least twice a week. Is there any data privacy or security concerns when you think about what you'd share on the app?

Let's shift the gear and talk about how this technology could be different.

- If you could wave a magic wand and change anything about [agent name], what would you like to change?
- What are your expectations regarding the future of artificial intelligence? What do you think it will look like in 20 years?
- You were chatting with [Agent name], but here are some other agents that have also been designed to help people better cope with challenges. *Show participant other chatbot* personas]
- Just right out of hand, do you have any particular ones that you prefer to interact with?
- Thinking about chatbots that help people manage their health, how important is the gender of the chatbot to you? How important is the chatbot’s race? Why?
- What are some other contexts in which you might use a chatbot?
- When do you expect a chatbot you interact with have a clear race affiliation?
- With regard to a chatbot that helps you [insert what they said, or if they were unable to think of anything, use your online shopping example] how important would the gender of the chatbot be for you?
- How important would the race of the chatbot be for you? Why?

References

ABC News. (2015). *Top 20 "Whitest" and "Blackest" Names*. ABC News. https://abcnews.go.com/2020/top-20-whitest-blackest-names/story?id=2470131

Ackerman, S. J., Benjamin, L. S., Beutler, L. E., Gelso, C. J., Goldfried, M. R., Hill, C., Lambert, M. J., Norcross, J. C., Orlinsky, D. E., & Rainer, J. (2001). Empirically supported therapy relationships: Conclusions and recommendations fo the Division 29 Task Force. *Psychotherapy: Theory, Research, Practice, Training*, *38*(4), 495.

Altman, I., & Taylor, D. A. (1973). *Social penetration: The development of interpersonal relationships.* Holt, Rinehart & Winston.

Angeli, A. D., Johnson, G. I., & Coventry, L. (2001). *The unfriendly user: Exploring social reactions to chatterbots*. 8.

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. https://doi.org/10.1016/j.chb.2018.03.051

Araújo, T., & Casais, B. (2020). Customer Acceptance of Shopping-Assistant Chatbots. In *Marketing and Smart Technologies* (pp. 278–287). Springer.

Archer, R. L. (1980). *Self-disclosure In Wegner DM & Vallacher RR (Eds.), The Self in Social Psychology (pp. 183–205)*. New York, NY: Oxford University Press.

Arkowitz, H., & Westra, H. A. (2009). Introduction to the special series on motivational interviewing and psychotherapy. *Journal of Clinical Psychology*, *65*(11), 1149–1155. https://doi.org/10.1002/jclp.20640

Arnold, C., Villagonzalo, K.-A., Meyer, D., Farhall, J., Foley, F., Kyrios, M., & Thomas, N. (2019). Predicting engagement with an online psychosocial intervention for psychosis: Exploring individual- and intervention-level predictors. *Internet Interventions*, *18*, 100266. https://doi.org/10.1016/j.invent.2019.100266

Aron, A., & Fraley, B. (1999). Relationship Closeness as Including Other in the Self: Cognitive Underpinnings and Measures. *Social Cognition*, *17*(2), 140–160. https://doi.org/10.1521/soco.1999.17.2.140

Ashktorab, Z., Jain, M., Liao, Q. V., & Weisz, J. D. (2019). Resilient chatbots: Repair strategy preferences for conversational breakdowns. *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, 1–12.

Bagozzi, R. P. (2007). The legacy of the technology acceptance model and a proposal for a paradigm shift. *Journal of the Association for Information Systems*, *8*(4), 3.

Banks, J. (2019). A perceived moral agency scale: Development and validation of a metric for humans and social machines. *Computers in Human Behavior*, *90*, 363–371. https://doi.org/10.1016/j.chb.2018.08.028

Barak, A., Hen, L., Boniel-Nissim, M., & Shapira, N. (2008). A Comprehensive Review and a Meta-Analysis of the Effectiveness of Internet-Based

Psychotherapeutic Interventions. *Journal of Technology in Human Services*, *26*(2–4), 109–160. https://doi.org/10.1080/15228830802094429

Bartneck, C., Kulic, D., & Croft, E. (2008). Measuring the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots. *Technical Report*, 8.

Baylor, A. L., & Kim, Y. (2004). Pedagogical Agent Design: The Impact of Agent Realism, Gender, Ethnicity, and Instructional Role. In J. C. Lester, R. M. Vicari, & F. Paraguaçu (Eds.), *Intelligent Tutoring Systems* (pp. 592–603). Springer Berlin Heidelberg.

Becker, D. (2018). Possibilities to Improve Online Mental Health Treatment: Recommendations for Future Research and Developments. *Future of Information and Communication Conference*, 91–112.

Ben-Ari, A. (1995). Coming out: A dialectic of intimacy and privacy. *Families in Society*, *76*(5), 306–314.

Bendig, E., Erb, B., Schulze-Thuesing, L., & Baumeister, H. (2019). The Next Generation: Chatbots in Clinical Psychology and Psychotherapy to Foster Mental Health – A Scoping Review. *Verhaltenstherapie*, 1–13. https://doi.org/10.1159/000501812

Bickmore, & Cassell, J. (n.d.). *SOCIAL DIALOGUE WITH EMBODIED CONVERSATIONAL AGENTS*. 32.

Bickmore, & Picard, R. W. (2005). Establishing and maintaining long-term human-computer relationships. *ACM Transactions on Computer-Human Interaction*, *12*(2), 293–327. https://doi.org/10.1145/1067860.1067867

Bickmore, T., Schulman, D., & Yin, L. (2010). Maintaining Engagement in Long-term Interventions with Relational Agents. *Applied Artificial Intelligence : AAI*, *24*(6), 648–666. https://doi.org/10.1080/08839514.2010.492259

Bickmore, T. W., Mitchell, S. E., Jack, B. W., Paasche-Orlow, M. K., Pfeifer, L. M., & ODonnell, J. (2010). Response to a Relational Agent by Hospital Patients with Depressive Symptoms. *Interacting with Computers*, *22*(4), 289–298. https://doi.org/10.1016/j.intcom.2009.12.001

Björkqvist, K., Österman, K., & Kaukiainen, A. (2000). Social intelligence− empathy= aggression? *Aggression and Violent Behavior*, *5*(2), 191–200.

botanalytics. (2018, April 18). A Brief Chatbot History, From The 20th Century To Now. *Botanalytics Blog - Conversational Analytics for Bots*. https://botanalytics.co/blog/2018/04/18/brief-chatbot-history-20th-century-now/

Boukricha, H., & Wachsmuth, I. (2011). Empathy-based emotional alignment for a virtual human: A three-step approach. *KI - Künstliche Intelligenz*, *25*(3). https://pub.uni-bielefeld.de/record/2276585

Brahnam, S., & De Angeli, A. (2012). Gender affordances of conversational agents. *Interacting with Computers*, *24*(3), 139–153. https://doi.org/10.1016/j.intcom.2012.05.001

Braun, V., & Clarke, V. (2012). Thematic analysis. In *APA handbook of research methods in psychology, Vol 2: Research designs: Quantitative, qualitative, neuropsychological, and biological* (pp. 57–71). American Psychological Association. https://doi.org/10.1037/13620-004

Brewster, S. (2016). *The next generation of bots can respond to your mental state*. MIT Technology Review. https://www.technologyreview.com/s/602352/customer-service-bots-are-getting-better-at-detecting-your-agitation/

Burgoon, J. K., Bonito, J. A., Bengtsson, B., Cederberg, C., Lundeberg, M., & Allspach, L. (2000). Interactivity in human–computer interaction: A study of credibility, understanding, and influence. *Computers in Human Behavior*, *16*(6), 553–574.

Cabral, R. R., & Smith, T. B. (2011). Racial/ethnic matching of clients and therapists in mental health services: A meta-analytic review of preferences, perceptions, and outcomes. *Journal of Counseling Psychology*, *58*(4), 537–554. https://doi.org/10.1037/a0025266

Campbell, D. T. (1958). Common fate, similarity, and other indices of the status of aggregates of persons as social entities. *Behavioral Science*, *3*, 14–25. https://doi.org/10.1002/bs.3830030103

Caplan, S. E., & Turner, J. S. (2007). Bringing theory to research on computer-mediated comforting communication. *Computers in Human Behavior*, *23*(2), 985–998.

Carkhuff, R. R., & Pierce, R. (1967). Differential effects of therapist race and social class upon patient depth of self-exploration in the initial clinical interview. *Journal of Consulting Psychology*, *31*(6), 632–634. https://doi.org/10.1037/h0025163

Cassell, J. (2009). Social practice: Becoming enculturated in human-computer interaction. *International Conference on Universal Access in Human-Computer Interaction*, 303–313.

Cassell, J., & Bickmore, T. (2003). Negotiated Collusion: Modeling Social Language and its Relationship Effects in Intelligent Agents. *User Modeling and User-Adapted Interaction*, *13*(1), 89–132. https://doi.org/10.1023/A:1024026532471

Cave, S., Coughlan, K., & Dihal, K. (2019). "Scary Robots": Examining Public Responses to AI. *AIES*. https://doi.org/10.1145/3306618.3314232

Chaves, A. P., & Gerosa, M. A. (2019). How should my chatbot interact? A survey on human-chatbot interaction design. *ArXiv:1904.02743 [Cs]*. http://arxiv.org/abs/1904.02743

Chavez, A. F., & Guido-DiBrito, F. (1999). Racial and ethnic identity and development. *New Directions for Adult and Continuing Education*, *84*, 39–47.

Chowdhury, G. G. (2003). Natural language processing. *Annual Review of Information Science and Technology*, *37*(1), 51–89.

Chung, K., & Park, R. C. (2019). Chatbot-based heathcare service with a knowledge base for cloud computing. *Cluster Computing*, *22*(1), 1925–1937. https://doi.org/10.1007/s10586-018-2334-5

Ciechanowski, L., Przegalinska, A., Magnuski, M., & Gloor, P. (2019). In the shades of the uncanny valley: An experimental study of human–chatbot interaction. *Future Generation Computer Systems*, *92*, 539–548. https://doi.org/10.1016/j.future.2018.01.055

Corrigan, P. (2004). How stigma interferes with mental health care. *American Psychologist*, *59*(7), 614.

Creswell, J. D., Lam, S., Stanton, A. L., Taylor, S. E., Bower, J. E., & Sherman, D. K. (2007). Does self-affirmation, cognitive processing, or discovery of meaning explain cancer-related health benefits of expressive writing? *Personality and Social Psychology Bulletin*, *33*(2), 238–250.

Curry, A. C., & Rieser, V. (2018). # MeToo Alexa: How conversational systems respond to sexual harassment. *Proceedings of the Second ACL Workshop on Ethics in Natural Language Processing*, 7–14.

Dahlbäck, N., Jönsson, A., & Ahrenberg, L. (1993). Wizard of Oz studies: Why and how. *Proceedings of the 1st International Conference on Intelligent User Interfaces*, 193–200.

Dale, R. (2016). The return of the chatbots. *Natural Language Engineering*, *22*(5), 811–817. https://doi.org/10.1017/S1351324916000243

D'alfonso, S., Santesteban-Echarri, O., Rice, S., Wadley, G., Lederman, R., Miles, C., Gleeson, J., & Alvarez-Jimenez, M. (2017). Artificial intelligence-assisted online social therapy for youth mental health. *Frontiers in Psychology*, *8*, 796.

Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly*, 319–340.

De Angeli, A., Lynch, P., & Johnson, G. (2001). Personifying the e-Market: A Framework for Social Agents. *Interact*, 198–205.

Derlega, V. J., Winstead, B. A., & Greene, K. (2001). Self-disclosure and starting a close relationship. *Handbook of Relationship Beginnings*, 153–174.

Diener, E., Emmons, R. A., Larsen, R. J., & Griffin, S. (1985). The Satisfaction With Life Scale. *Journal of Personality Assessment*, *49*(1), 71–75. https://doi.org/10.1207/s15327752jpa4901_13

Diesbach, P. L., & Midgley, D. F. (2007). Embodied Agents on a website: Modelling an attitudinal route of influence. *International Conference on Persuasive Technology*, 223–230.

Doherty, G., Coyle, D., & Sharry, J. (2012). Engagement with online mental health interventions: An exploratory clinical study of a treatment for depression. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 1421–1430. https://doi.org/10.1145/2207676.2208602

Dohsaka, K., Asai, R., Higashinaka, R., Minami, Y., & Maeda, E. (2009). Effects of conversational agents on human communication in thoughtevoking multi-party dialogues. *In Proc. SIGDIAL*, 217–224.

Donath, J. (2001). Mediated faces. *International Conference on Cognitive Technology*, 373–390.

Donkin, L., Hickie, I. B., Christensen, H., Naismith, S. L., Neal, B., Cockayne, N. L., & Glozier, N. (2013). Rethinking the Dose-Response Relationship Between Usage and Outcome in an Online Intervention for Depression: Randomized Controlled Trial. *Journal of Medical Internet Research*, *15*(10), e231. https://doi.org/10.2196/jmir.2771

Duijst, D. (2017). Can we improve the User Experience of Chatbots with Personalisation. *Master's Thesis. University of Amsterdam*.

Dyke, G., Howley, I., Adamson, D., Kumar, R., & Rosé, C. P. (2013). Towards academically productive talk supported by conversational agents. In *Productive multivocality in the analysis of group interactions* (pp. 459–476). Springer.

Ebert, D. D., Van Daele, T., Nordgreen, T., Karekla, M., Compare, A., Zarbo, C., Brugnera, A., Øverland, S., Trebbi, G., Jensen, K. L., Kaehlke, F., & Baumeister, H. (2018). Internet- and Mobile-Based Psychological Interventions: Applications, Efficacy, and Potential for Improving Mental Health. *European Psychologist*, *23*(2), 167–187. https://doi.org/10.1027/1016-9040/a000318

Eells, T. D., Barrett, M. S., Wright, J. H., & Thase, M. E. (2015). Can Cognitive Therapy Be Conducted by Computers? *Current Behavioral Neuroscience Reports*, *2*(4), 209–215. https://doi.org/10.1007/s40473-015-0051-4

Elkins, A. C., Derrick, D. C., Burgoon, J. K., & Jr, J. F. N. (2012). Predicting Users' Perceived Trust in Embodied Conversational Agents Using Vocal Dynamics. *2012 45th Hawaii International Conference on System Sciences*, 579–588. https://doi.org/10.1109/HICSS.2012.483

Elliott, R., Bohart, A. C., Watson, J. C., & Greenberg, L. S. (2011). Empathy. *Psychotherapy*, *48*(1), 43.

Ennis, L., Rose, D., Denis, M., Pandit, N., & Wykes, T. (2012). Can't surf, won't surf: The digital divide in mental health. *Journal of Mental Health*, *21*(4), 395–403.

Epley, N., & Waytz, A. (2010). Mind perception. *Handbook of Social Psychology*.

Epley, N., Waytz, A., & Cacioppo, J. T. (2007). On seeing human: A three-factor theory of anthropomorphism. *Psychological Review*, *114*(4), 864–886. https://doi.org/10.1037/0033-295X.114.4.864

Farber, B. A. (2003). Patient self-disclosure: A review of the research. *Journal of Clinical Psychology*, *59*(5), 589–600.

Farber, B. A., Berano, K. C., & Capobianco, J. A. (2006). A temporal model of patient disclosure in psychotherapy. *Psychotherapy Research*, *16*(4), 463–469. https://doi.org/10.1080/10503300600593250

Farber, B. A., & Hall, D. (2002). Disclosure to therapists: What is and is not discussed in psychotherapy. *Journal of Clinical Psychology*, *58*(4), 359–370. https://doi.org/10.1002/jclp.1148

Fernández-Álvarez, J., Rozental, A., Carlbring, P., Colombo, D., Riva, G., Anderson, P. L., Baños, R. M., Benbow, A. A., Bouchard, S., Bretón-López, J. M., Cárdenas, G., Difede, J., Emmelkamp, P., García-Palacios, A., Guillén, V., Hoffman, H., Kampann, I., Moldovan, R., Mühlberger, A., … Botella, C. (2019). Deterioration rates in Virtual Reality Therapy: An individual patient data level meta-analysis. *Journal of Anxiety Disorders*, *61*, 3–17. https://doi.org/10.1016/j.janxdis.2018.06.005

Finfgeld, D. L. (1999). Psychotherapy in cyberspace. *Journal of the American Psychiatric Nurses Association*, *5*(4), 105–110. https://doi.org/10.1016/S1078-3903(99)90008-3

Finkenauer, C., & Rimé, B. (1998). Socially shared emotional experiences vs. Emotional experiences kept secret: Differential characteristics and consequences. *Journal of Social and Clinical Psychology*, *17*(3), 295–318.

Fitzpatrick, K. K., Darcy, A., & Vierhile, M. (2017). Delivering Cognitive Behavior Therapy to Young Adults With Symptoms of Depression and Anxiety Using a Fully Automated Conversational Agent (Woebot): A Randomized Controlled Trial. *JMIR Mental Health*, *4*(2), e19. https://doi.org/10.2196/mental.7785

Flanagan, C. M. (2010). The case for needs in psychotherapy. *Journal of Psychotherapy Integration*, *20*(1), 1–36. https://doi.org/10.1037/a0018815

Fleming, T. M., Cheek, C., Merry, S. N., Thabrew, H., Bridgman, H., Stasiak, K., Shepherd, M., Perry, Y., & Hetrick, S. (2015). Serious games for the treatment or prevention of depression: A systematic review. *Revista de Psicopatología y Psicología Clínica*, *19*(3), 227. https://doi.org/10.5944/rppc.vol.19.num.3.2014.13904

Flückiger, C., Del Re, A. C., Wampold, B. E., & Horvath, A. O. (2018). The alliance in adult psychotherapy: A meta-analytic synthesis. *Psychotherapy*, *55*(4), 316. https://doi.org/10.1037/pst0000172

Følstad, A., & Brandtzæg, P. B. (2017). Chatbots and the new world of HCI. *Interactions*, *24*(4), 38–42.

Forlizzi, J., Zimmerman, J., Mancuso, V., & Kwak, S. (2007). How interface agents affect interaction between humans and computers. *Proceedings of the 2007 Conference on Designing Pleasurable Products and Interfaces*, 209–221.

Fox, J., Ahn, S. J. (Grace), Janssen, J. H., Yeykelis, L., Segovia, K. Y., & Bailenson, J. N. (2015). Avatars Versus Agents: A Meta-Analysis Quantifying the Effect of Agency on Social Influence. *Human–Computer Interaction*, *30*(5), 401–432. https://doi.org/10.1080/07370024.2014.921494

Freud, S. (1927). *The Future of an Illusion: Civilization and Its Discontents, and Other Works*. Hogarth Press and the Institute of Psycho-analysis.

Fryer, R. G., & Levitt, S. D. (2003). *The Causes and Consequences of Distinctively Black Names* (Working Paper No. 9938). National Bureau of Economic Research. https://doi.org/10.3386/w9938

Gardner, W. L., Pickett, C. L., & Knowles, M. (2005). Social snacking and shielding. *The Social Outcast: Ostracism, Social Exclusion, Rejection, and Bullying*, 227–242.

Gazzola, V., Rizzolatti, G., Wicker, B., & Keysers, C. (2007). The anthropomorphic brain: The mirror neuron system responds to human and robotic actions. *Neuroimage*, *35*(4), 1674–1684.

Gelso, C. J. (2011). *The real relationship in psychotherapy: The hidden foundation of change*. American Psychological Association. https://doi.org/10.1037/12349-000

Gelso, C. J., & Carter, J. A. (1985). The relationship in counseling and psychotherapy: Components, consequences, and theoretical antecedents. *The Counseling Psychologist*, *13*(2), 155–243.

Gilbert, P., & Leahy, R. L. (2007). *The Therapeutic Relationship in the Cognitive Behavioral Psychotherapies*. Routledge.

Go, E., & Sundar, S. S. (2019). Humanizing chatbots: The effects of visual, identity and conversational cues on humanness perceptions. *Computers in Human Behavior*, *97*, 304–316. https://doi.org/10.1016/j.chb.2019.01.020

Grabois, E. W. (1997). The liability of psychotherapists for breach of confidentiality. *JL & Health*, *12*, 39.

Grantham, R. J. (1973). Effects of counselor sex, race, and language style on black students in initial interviews. *Journal of Counseling Psychology*, *20*(6), 553–559. https://doi.org/10.1037/h0035188

Gray, H. M., Gray, K., & Wegner, D. M. (2007). Dimensions of mind perception. *Science*, *315*(5812), 619–619.

Greenberg, M. A., & Stone, A. A. (1992). Emotional disclosure about traumas and its relation to health: Effects of previous disclosure and trauma severity. *Journal of Personality and Social Psychology*, *63*(1), 75.

Hamilton, D. L., & Sherman, S. J. (1996). Perceiving persons and groups. *Psychological Review*, *103*(2), 336–355. https://doi.org/10.1037/0033-295x.103.2.336

Han, Y.-J., & O'Brien, K. M. (2014). Critical secret disclosure in psychotherapy with Korean clients. *The Counseling Psychologist*, *42*(4), 524–551.

Harrison, D. K. (1975). Race as a counselor-client variable in counseling and psychotherapy: A review of the research. *The Counseling Psychologist*, *5*(1), 124–133. https://doi.org/10.1177/001100007500500130

HHS, O. of C. R. (OCR). (2017, December 4). *Information Related to Mental and Behavioral Health* [Text]. HHS.Gov. https://www.hhs.gov/hipaa/for-individuals/mental-health/index.html

Hill, C. E., Gelso, C. J., & Mohr, J. J. (2000). Client concealment and self-presentation in therapy: Comment on Kelly (2000). *Psychological Bulletin*, *126*(4), 495–500. https://doi.org/10.1037/0033-2909.126.4.495

Hill, C. E., & O'Grady, K. E. (1985). List of counselor intentions illustrated in a case study and with counselors of varying theoretical orientations. *Journal of Counseling Psychology*, *32*, 3–22.

Ho, A. D., & Yu, C. C. (2015). Descriptive statistics for modern test score distributions: Skewness, kurtosis, discreteness, and ceiling effects. *Educational and Psychological Measurement*, *75*(3), 365–388.

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

Ho, A., Hancock, J., & Miner, A. S. (2018b). Psychological, Relational, and Emotional Effects of Self-Disclosure After Conversations With a Chatbot. *Journal of Communication*, *68*(4), 712–733. https://doi.org/10.1093/joc/jqy026

Ho, C.-C., & MacDorman, K. F. (2010). Revisiting the uncanny valley theory: Developing and validating an alternative to the Godspeed indices. *Computers* *in Human Behavior*, *26*(6), 1508–1518. https://doi.org/10.1016/j.chb.2010.05.015

Hopkins, K. D., & Weeks, D. L. (1990). Tests for normality and measures of skewness and kurtosis: Their place in research reporting. *Educational and Psychological Measurement*, *50*(4), 717–729.

Horvath, A. O., & Greenberg, L. S. (1989). Development and validation of the Working Alliance Inventory. *Journal of Counseling Psychology*, *36*(2), 223–233. https://doi.org/10.1037/0022-0167.36.2.223

Huang, J., Li, Q., Xue, Y., Cheng, T., Xu, S., Jia, J., & Feng, L. (2015). TeenChat: A Chatterbot System for Sensing and Releasing Adolescents' Stress. In X. Yin, K. Ho, D. Zeng, U. Aickelin, R. Zhou, & H. Wang (Eds.), *Health Information Science* (Vol. 9085, pp. 133–145). Springer International Publishing. https://doi.org/10.1007/978-3-319-19156-0_14

Hussain, S. A., Ogundimu, F., & Bhattarai, S. (2019). Mobile Phone-Based Chatbot for Family Planning and Contraceptive Information. *International Conference on Human-Computer Interaction*, 342–352.

Inkster, B., Sarda, S., & Subramanian, V. (2018). An Empathy-Driven, Conversational Artificial Intelligence Agent (Wysa) for Digital Mental Well-Being: Real-World Data Evaluation Mixed-Methods Study. *JMIR MHealth and UHealth*, *6*(11), e12106. https://doi.org/10.2196/12106

Ischen, C., Araujo, T., Voorveld, H., van Noort, G., & Smit, E. (2019). Privacy concerns in chatbot interactions. *International Workshop on Chatbot Research and Design*, 34–48.

Jain, M., Kumar, P., Kota, R., & Patel, S. N. (2018). Evaluating and Informing the Design of Chatbots. *Proceedings of the 2018 on Designing Interactive Systems Conference 2018 - DIS '18*, 895–906. https://doi.org/10.1145/3196709.3196735

Jordan, M. I., & Mitchell, T. M. (2015). Machine learning: Trends, perspectives, and prospects. *Science*, *349*(6245), 255–260.

Joska, J. A., & Flisher, A. J. (2007). Needs and services at an in-patient psychotherapy unit. *African Journal of Psychiatry*, *10*(3), 149–156.

Kahn, J. H., Vogel, D. L., Schneider, W. J., Barr, L. K., & Herrell, K. (2008). The emotional content of client disclosures and session impact: An analogue study. *Psychotherapy: Theory, Research, Practice, Training*, *45*(4), 539.

Kamita, T., Ito, T., Matsumoto, A., Munakata, T., & Inoue, T. (2019). *A Chatbot System for Mental Healthcare Based on SAT Counseling Method* [Research article]. Mobile Information Systems. https://doi.org/10.1155/2019/9517321

Kang, M. (2018). A Study of Chatbot Personality based on the Purposes of Chatbot. *The Journal of the Korea Contents Association*, *18*(5), 319–329.

Karlin, B. E., Duffy, M., & Gleaves, D. H. (2008). Patterns and predictors of mental health service use and mental illness among older and younger adults in the United States. *Psychological Services*, *5*(3), 275–294. https://doi.org/10.1037/1541-1559.5.3.275

Kelly, A. E. (2000). Helping construct desirable identities: A self-presentational view of psychotherapy. *Psychological Bulletin*, *126*(4), 475–494. https://doi.org/10.1037/0033-2909.126.4.475

Kelly, T. A., & Strupp, H. H. (1992). Patient and therapist values in psychotherapy: Perceived changes, assimilation, similarity, and outcome. *Journal of Consulting and Clinical Psychology*, *60*(1), 34–40. https://doi.org/10.1037//0022-006x.60.1.34

Keynejad, R. C., Dua, T., Barbui, C., & Thornicroft, G. (2018). WHO Mental Health Gap Action Programme (mhGAP) Intervention Guide: A systematic review of evidence from low and middle-income countries. *Evidence Based Mental Health*, *21*(1), 30–34. https://doi.org/10.1136/eb-2017-102750

Kiluk, B. D., Serafini, K., Frankforter, T., Nich, C., & Carroll, K. M. (2014). Only Connect: The Working Alliance in Computer-Based Cognitive Behavioral Therapy. *Behaviour Research and Therapy*, *63*, 139–146. https://doi.org/10.1016/j.brat.2014.10.003

Kim, J. W., Jo, H. I., & Lee, B. G. (2019). The Study on the Factors Influencing on the Behavioral Intention of Chatbot Service for the Financial Sector: Focusing on the UTAUT Model. *Journal of Digital Contents Society*, *20*(1), 41–50.

Kiropoulos, L. A., Klein, B., Austin, D. W., Gilson, K., Pier, C., Mitchell, J., & Ciechomski, L. (2008). Is internet-based CBT for panic disorder and agoraphobia as effective as face-to-face CBT? *Journal of Anxiety Disorders*, *22*(8), 1273–1284. https://doi.org/10.1016/j.janxdis.2008.01.008

Kivlighan Jr, D. M., & Shaughnessy, P. (2000). Patterns of working alliance development: A typology of client's working alliance ratings. *Journal of Counseling Psychology*, *47*(3), 362.

Kohatsu, E. L., Dulay, M., Lam, C., Concepcion, W., Perez, P., Lopez, C., & Euler, J. (2000). Using racial identity theory to explore racial mistrust and interracial contact among Asian Americans. *Journal of Counseling & Development*, *78*(3), 334–342. https://doi.org/10.1002/j.1556-6676.2000.tb01915.x

Kopta, S. M., Lueger, R. J., Saunders, S. M., & Howard, K. I. (1999). Individual psychotherapy outcome and process research: Challenge leading to great turmoil or positive transition? *Annual Review of Psychology*, *50*(1), 441–469. https://doi.org/10.1146/annurev.psych.50.1.441

Krämer, N. C., Lucas, G., Schmitt, L., & Gratch, J. (2018). Social snacking with a virtual agent – On the interrelation of need to belong and effects of social responsiveness when interacting with artificial entities. *International Journal of Human-Computer Studies*, *109*, 112–121. https://doi.org/10.1016/j.ijhcs.2017.09.001

Kramer, U., & Stiles, W. B. (2015). The Responsiveness Problem in Psychotherapy: A Review of Proposed Solutions. *Clinical Psychology: Science and Practice*, *22*(3), 277–295. https://doi.org/10.1111/cpsp.12107

Kretzschmar, K., Tyroll, H., Pavarini, G., Manzini, A., & Singh, I. (2019). Can Your Phone Be Your Therapist? Young People's Ethical Perspectives on the Use of Fully Automated Conversational Agents (Chatbots) in Mental Health Support. *Biomedical Informatics Insights*, *11*, 1178222619829083. https://doi.org/10.1177/1178222619829083

Lambert, M. J., & Barley, D. E. (2001). Research summary on the therapeutic relationship and psychotherapy outcome. *Psychotherapy: Theory, Research,* *Practice, Training*, *38*(4), 357–361. https://doi.org/10.1037/0033-3204.38.4.357

Lannin, D. G., Guyll, M., Vogel, D. L., & Madon, S. (2013). Reducing the stigma associated with seeking psychotherapy through self-affirmation. *Journal of Counseling Psychology*, *60*(4), 508–519. https://doi.org/10.1037/a0033789

Laumer, S., Maier, C., & Gubler, F. T. (2019). Chatbot acceptance in healthcare: Explaining user adoption of conversational agents for disease diagnosis. *Proceeding of the 27th European Conference on Information System (ECIS)*.

Leary, M. R. (2019). *Self-presentation: Impression Management And Interpersonal Behavior*. Routledge.

Lee, S., & Choi, J. (2017). Enhancing user experience with conversational agent for movie recommendation: Effects of self-disclosure and reciprocity. *International Journal of Human-Computer Studies*, *103*, 95–105. https://doi.org/10.1016/j.ijhcs.2017.02.005

Lesher, J. H. (2013). Xenophanes of Kolophon. *The Encyclopedia of Ancient History*, 1–2.

Liao, Y., & He, J. (2020). Racial mirroring effects on human-agent interaction in psychotherapeutic conversations. *Proceedings of the 25th International Conference on Intelligent User Interfaces*, 430–442.

Liao, Y., Vitak, J., Kumar, P., Zimmer, M., & Kritikos, K. (2019). Understanding the Role of Privacy and Trust in Intelligent Personal Assistant Adoption. *Proceedings of the 13th Annual IConference*, 102–113.

Lin, J., Faust, B., Ebert, D. D., Krämer, L., & Baumeister, H. (2018). A Web-Based Acceptance-Facilitating Intervention for Identifying Patients' Acceptance, Uptake, and Adherence of Internet- and Mobile-Based Pain Interventions: Randomized Controlled Trial. *Journal of Medical Internet Research*, *20*(8), e244. https://doi.org/10.2196/jmir.9925

Lin, L., Stamm, K., & Christidis, P. (2018). *How diverse is the psychology workforce?* Https://Www.Apa.Org. https://www.apa.org/monitor/2018/02/datapoint

Lindhiem, O., Bennett, C. B., Orimoto, T. E., & Kolko, D. J. (2016). A Meta-Analysis of Personalized Treatment Goals in Psychotherapy: A Preliminary Report and Call for More Studies. *Clinical Psychology : A Publication of the Division of Clinical Psychology of the American Psychological Association*, *23*(2), 165–176. https://doi.org/10.1111/cpsp.12153

Lipari, R. N. (2018). *Key Substance Use and Mental Health Indicators in the United States: Results from the 2018 National Survey on Drug Use and Health*. 82.

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. https://doi.org/10.1089/cyber.2018.0110

Loveys, K., Fricchione, G., Kolappa, K., Sagar, M., & Broadbent, E. (2019). Reducing Patient Loneliness With Artificial Agents: Design Insights From Evolutionary Neuropsychiatry. *Journal of Medical Internet Research*, *21*(7), e13664. https://doi.org/10.2196/13664

Lucas, G. M., Gratch, J., King, A., & Morency, L.-P. (2014). It’s only a computer: Virtual humans increase willingness to disclose. *Computers in Human Behavior*, *37*, 94–100. https://doi.org/10.1016/j.chb.2014.04.043

Ly, K. H., Ly, A.-M., & Andersson, G. (2017). A fully automated conversational agent for promoting mental well-being: A pilot RCT using mixed methods. *Internet Interventions*, *10*, 39–46. https://doi.org/10.1016/j.invent.2017.10.002

MacDorman, K. F., Green, R. D., Ho, C.-C., & Koch, C. T. (2009). Too real for comfort? Uncanny responses to computer generated faces. *Computers in Human Behavior*, *25*(3), 695–710.

Manago, A. M., & Vaughn, L. (2015). Social media, friendship, and happiness in the millennial generation. In *Friendship and happiness* (pp. 187–206). Springer.

Manhal-Baugus, M. (2001). E-therapy: Practical, ethical, and legal issues. *CyberPsychology & Behavior*, *4*(5), 551–563.

Marcus, G. (2018). Innateness, AlphaZero, and Artificial Intelligence. *ArXiv:1801.05667 [Cs]*. http://arxiv.org/abs/1801.05667

Margalit, L. (2016). *The Psychology of Chatbots*. Psychology Today. https://www.psychologytoday.com/blog/behind-online-behavior/201607/the-psychology-chatbots

Marino, M. C. (2014). *The Racial Formation of Chatbots*. https://doi.org/10.7771/1481-4374.2560

Martins, M. V., Peterson, B. D., Costa, P., Costa, M. E., Lund, R., & Schmidt, L. (2013). Interactive effects of social support and disclosure on fertility-related stress. *Journal of Social and Personal Relationships*, *30*(4), 371–388.

Meier, S. T. (1988). An exploratory study of a computer-assisted alcohol education program. *Computers in Human Services*, *3*(3–4), 111–121.

Melián-González, S., Gutiérrez-Taño, D., & Bulchand-Gidumal, J. (2019). Predicting the intentions to use chatbots for travel and tourism. *Current Issues in Tourism*, *0*(0), 1–19. https://doi.org/10.1080/13683500.2019.1706457

Melville, K. M., Casey, L. M., & Kavanagh, D. J. (2010). Dropout from Internet-based treatment for psychological disorders. *The British Journal of Clinical Psychology*, *49*(Pt 4), 455–471. https://doi.org/10.1348/014466509X472138

Merry, S. N., Stasiak, K., Shepherd, M., Frampton, C., Fleming, T., & Lucassen, M. F. G. (2012). The effectiveness of SPARX, a computerised self help intervention for adolescents seeking help for depression: Randomised controlled non-inferiority trial. *BMJ*, *344*. https://doi.org/10.1136/bmj.e2598

Miller, W. R., & Rollnick, S. (2013). *Motivational interviewing: Helping people change, 3rd edition*. Guilford Press.

Miltgen, C., Popovič, A., & Oliveira, T. (2013). Determinants of end-user acceptance of biometrics: Integrating the “Big 3” of technology acceptance with privacy context. *Decision Support Systems*, *56*, 103–114. https://doi.org/10.1016/j.dss.2013.05.010

Miner, A. S., Milstein, A., Schueller, S., Hegde, R., Mangurian, C., & Linos, E. (2016). Smartphone-Based Conversational Agents and Responses to Questions About Mental Health, Interpersonal Violence, and Physical Health. *JAMA Internal Medicine*, *176*(5), 619–625. https://doi.org/10.1001/jamainternmed.2016.0400

Miner, A. S., Shah, N., Bullock, K. D., Arnow, B. A., Bailenson, J., & Hancock, J. (2019). Key Considerations for Incorporating Conversational AI in Psychotherapy. *Frontiers in Psychiatry*, *10*. https://doi.org/10.3389/fpsyt.2019.00746

Mitchell, M. (2021). Why AI is Harder Than We Think. *ArXiv:2104.12871 [Cs]*. http://arxiv.org/abs/2104.12871

Molteni, M. (2017, June 7). The Chatbot Therapist Will See You Now. *Wired*. https://www.wired.com/2017/06/facebook-messenger-woebot-chatbot-therapist/

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

Morais, B. (2013, November 19). *Can Humans Fall in Love with Bots?* https://www.newyorker.com/tech/elements/can-humans-fall-in-love-with-bots

Mori, M. (1970). The uncanny valley. *Energy*, *7*(4), 33–35.

Morkes, J., Kernal, H. K., & Nass, C. (1999). Effects of Humor in Task-oriented Human-computer Interaction and Computer-mediated Communication: A Direct Test of SRCT Theory. *Hum.-Comput. Interact.*, *14*(4), 395–435. https://doi.org/10.1207/S15327051HCI1404_2

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. https://doi.org/10.2196/10148

Nass, C., Fogg, B. J., & Moon, Y. (1996). Can computers be teammates? *International Journal of Human-Computer Studies*, *45*(6), 669–678.

Nass, C., Moon, Y., Fogg, B. J., Reeves, B., & Dryer, C. (1995). Can computer personalities be human personalities? *Conference Companion on Human Factors in Computing Systems*, 228–229.

Neururer, M., Schlögl, S., Brinkschulte, L., & Groth, A. (2018). Perceptions on authenticity in chat bots. *Multimodal Technologies and Interaction*, *2*(3), 60.

Neville, F. G., Novelli, D., Drury, J., & Reicher, S. D. (2020). Shared social identity transforms social relations in imaginary crowds. *Group Processes & Intergroup Relations*, 1368430220936759. https://doi.org/10.1177/1368430220936759

Nissenbaum, H. (2004). Privacy as contextual integrity. *Washington Law Review*, *79*, 119–157.

Norcross, J. C. (2002). Psychotherapy relationships that work: Therapist contributions and responsiveness to patients. In *Psychotherapy relationships that work: Therapist contributions and responsiveness to patients.* (pp. 3–16). Oxford University Press.

Norman, D. A. (1994). How Might People Interact with Agents. *Commun. ACM*, *37*(7), 68–71. https://doi.org/10.1145/176789.176796

Nowak, K. L., & Rauh, C. (2005). The Influence of the Avatar on Online Perceptions of Anthropomorphism, Androgyny, Credibility, Homophily, and Attraction. *Journal of Computer-Mediated Communication*, *11*(1), 153–178. https://doi.org/10.1111/j.1083-6101.2006.tb00308.x

Nunamaker, J. F., Derrick, D. C., Elkins, A. C., Burgoon, J. K., & Patton, M. W. (2011). Embodied Conversational Agent-Based Kiosk for Automated Interviewing. *Journal of Management Information Systems*, *28*(1), 17–48. https://doi.org/10.2753/MIS0742-1222280102

Omarzu, J. (2000). A Disclosure Decision Model: Determining How and When Individuals Will Self-Disclose. *Personality and Social Psychology Review*, *4*(2), 174–185. https://doi.org/10.1207/S15327957PSPR0402_05

Ormerod, J. T. (2017). Extending the linear model with R: Generalized linear, mixed effects and nonparametric regression models, second edition by Julian J. Faraway, Chapman and Hall. *Statistics in Medicine*, *36*(14), 2318–2319. https://doi.org/10.1002/sim.7282

Osberg, M., & Mehrotra, D. (2020, February 19). *The Spooky, Loosely Regulated World of Online Therapy*. Jezebel. https://jezebel.com/the-spooky-loosely-regulated-world-of-online-therapy-1841791137

O'Sullivan, M. J., Peterson, P. D., Cox, G. B., & Kirkeby, J. (1989). Ethnic populations: Community mental health services ten years later. *American Journal of Community Psychology*, *17*(1), 17–30.

Paganini, S., Lin, J., Ebert, D. D., & Baumeister, H. (2016). Internet-und mobilebasierte Intervention bei psychischen Störungen. *Neurotransmitter*, *27*, 48–57.

Park, S., Choi, J., Lee, S., Oh, C., Kim, C., La, S., Lee, J., & Suh, B. (2019). Designing a Chatbot for a Brief Motivational Interview on Stress

Management: Qualitative Case Study. *Journal of Medical Internet Research*, *21*(4), e12231. https://doi.org/10.2196/12231

Patel, V., Saxena, S., Lund, C., Thornicroft, G., Baingana, F., Bolton, P., Chisholm, D., Collins, P. Y., Cooper, J. L., Eaton, J., Herrman, H., Herzallah, M. M., Huang, Y., Jordans, M. J. D., Kleinman, A., Medina-Mora, M. E., Morgan, E., Niaz, U., Omigbodun, O., … UnÜtzer, Jü. (2018). The Lancet Commission on global mental health and sustainable development. *The Lancet*, *392*(10157), 1553–1598. https://doi.org/10.1016/S0140-6736(18)31612-X

Pennebaker, J. W. (1985). Traumatic experience and psychosomatic disease: Exploring the roles of behavioural inhibition, obsession, and confiding. *Canadian Psychology/Psychologie Canadienne*, *26*(2), 82.

Pennebaker, J. W. (1990). *Opening up: The healing power of confiding in others*.

Pennebaker, J. W. (1997). Writing about emotional experiences as a therapeutic process. *Psychological Science*, *8*(3), 162–166.

Pennebaker, J. W., & Beall, S. K. (1986). Confronting a traumatic event: Toward an understanding of inhibition and disease. *Journal of Abnormal Psychology*, *95*(3), 274.

Penney, J. (2019). The right to privacy: The end of Privacy Fatalism. In *Human Rights, Digital Society and the Law* (pp. 44–57). Routledge.

Pepe, A., & Addimando, L. (2014). Teacher-parent relationships: Influence of gender and education on organizational parents' counterproductive behaviors. *European Journal of Psychology of Education*, *29*, 503–519. https://doi.org/10.1007/s10212-014-0210-0

Petronio, S. (2002). *Boundaries of privacy: Dialectics of disclosure*. Suny Press.

Petta, P., & Trappl, R. (1997). Why to create personalities for synthetic actors. In *Creating Personalities for Synthetic Actors* (pp. 1–8). Springer.

Pichponreay, L., Kim, J.-H., Choi, C.-H., Lee, K.-H., & Cho, W.-S. (2016). Smart answering Chatbot based on OCR and Overgenerating Transformations and Ranking. *2016 Eighth International Conference on Ubiquitous and Future Networks (ICUFN)*, 1002–1005.

Pocock, D. (1997). Feeling understood in family therapy. *Journal of Family Therapy*, *19*(3), 283–302.

Popovic, M., Milne, D., & Barrett, P. (2003). The scale of perceived interpersonal closeness (PICS). *Clinical Psychology & Psychotherapy*, *10*(5), 286–301. https://doi.org/10.1002/cpp.375

Purington, A., Taft, J. G., Sannon, S., Bazarova, N. N., & Taylor, S. H. (2017). “Alexa is my new BFF”: Social roles, user satisfaction, and Personification of the Amazon Echo. *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems*, 2853–2859. https://doi.org/10.1145/3027063.3053246

Reeves, B., & Nass, C. I. (1996). *The media equation: How people treat computers, television, and new media like real people and places*. Cambridge University Press.

Regan, A. M., & Hill, C. E. (1992). Investigation of what clients and counselors do not say in brief therapy. *Journal of Counseling Psychology*, *39*(2), 168.

Reis, H. T., Lemay, E. P., & Finkenauer, C. (2017). Toward understanding understanding: The importance of feeling understood in relationships. *Social and Personality Psychology Compass*, *11*(3), e12308. https://doi.org/10.1111/spc3.12308

Reis, H. T., & Shaver, P. (1988). Intimacy as an interpersonal process. *Handbook of Personal Relationships*.

Reynolds, D. J., Stiles, W. B., & Grohol, J. M. (2006). An investigation of session impact and alliance in internet based psychotherapy: Preliminary results. *Counselling and Psychotherapy Research*, *6*(3), 164–168. https://doi.org/10.1080/14733140600853617

Ridley, C. R. (1984). Clinical treatment of the nondisclosing Black client: A therapeutic paradox. *American Psychologist*, *39*(11), 1234–1244. https://doi.org/10.1037/0003-066X.39.11.1234

Rietz, T., Benke, I., & Maedche, A. (2019). *The Impact of Anthropomorphic and Functional Chatbot Design Features in Enterprise Collaboration Systems on User Acceptance*.

Russell, G. L., Fujino, D. C., Sue, S., Cheung, M.-K., & Snowden, L. R. (1996). The Effects of Therapist-Client Ethnic Match in the Assessment of Mental Health Functioning. *Journal of Cross-Cultural Psychology*, *27*(5), 598–615. https://doi.org/10.1177/0022022196275007

Ryff, C. D., & Keyes, C. L. M. (1995). The structure of psychological well-being revisited. *Journal of Personality and Social Psychology*, *69*(4), 719.

Salovaara, A., & Oulasvirta, A. (2004). Six modes of proactive resource management: A user-centric typology for proactive behaviors. *Proceedings of the Third Nordic Conference on Human-Computer Interaction*, 57–60. https://doi.org/10.1145/1028014.1028022

Salovey, P., & Mayer, J. D. (2005). *The science of emotional intelligence*.

Saypol, E., & Farber, B. A. (2010). Attachment style and patient disclosure in psychotherapy. *Psychotherapy Research*, *20*(4), 462–471.

Schlesinger, A., O'Hara, K. P., & Taylor, A. S. (2018). Let's Talk About Race: Identity, Chatbots, and AI. *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems - CHI '18*, 1–14. https://doi.org/10.1145/3173574.3173889

Scholten, M. R., Kelders, S. M., & Gemert-Pijnen, J. E. V. (2017). Self-Guided Web-Based Interventions: Scoping Review on User Needs and the Potential of Embodied Conversational Agents to Address Them. *Journal of Medical Internet Research*, *19*(11), e383. https://doi.org/10.2196/jmir.7351

Schubart, J. R., Stuckey, H. L., Ganeshamoorthy, A., & Sciamanna, C. N. (2011). Chronic health conditions and internet behavioral interventions: A review of factors to enhance user engagement. *Computers, Informatics, Nursing: CIN*, *29*(2), 81–92. https://doi.org/10.1097/NCN.0b013e3182065eed

Schuetzler, R. M., Grimes, G. M., & Giboney, J. S. (2019). The effect of conversational agent skill on user behavior during deception. *Computers in Human Behavior*, *97*, 250–259. https://doi.org/10.1016/j.chb.2019.03.033

Seltman, H. J. (2018). *Experimental Design and Analysis*.

Senaviratna, N., & Cooray, T. (2019). Diagnosing multicollinearity of logistic regression model. *Asian Journal of Probability and Statistics*, 1–9.

Seyama, J., & Nagayama, R. S. (2007). The Uncanny Valley: Effect of Realism on the Impression of Artificial Human Faces. *Presence: Teleoperators and Virtual Environments*, *16*(4), 337–351. https://doi.org/10.1162/pres.16.4.337

Sheehan, B., Jin, H. S., & Gottlieb, U. (2020). Customer service chatbots: Anthropomorphism and adoption. *Journal of Business Research*, *115*, 14–24.

Shum, H.-Y., He, X., & Li, D. (2018). From Eliza to XiaoIce: Challenges and opportunities with social chatbots. *Frontiers of Information Technology & Electronic Engineering*, *19*(1), 10–26.

Simons, H. W., Berkowitz, N. N., & Moyer, R. J. (1970). Similarity, credibility, and attitude change: A review and a theory. *Psychological Bulletin*, *73*(1), 1–16. https://doi.org/10.1037/h0028429

Simpson, H. B., Maher, M. J., Wang, Y., Bao, Y., Foa, E. B., & Franklin, M. (2011). Patient Adherence Predicts Outcome from Cognitive-Behavioral Therapy in Obsessive-Compulsive Disorder. *Journal of Consulting and Clinical Psychology*, *79*(2), 247–252. https://doi.org/10.1037/a0022659

Snowden, L. R. (1999). African American service use for mental health problems. *Journal of Community Psychology*, *27*(3), 303–313. https://doi.org/10.1002/(SICI)1520-6629(199905)27:3<303::AID-JCOP5>3.0.CO;2-9

Sojit, C. (2020). *Your AI Chatbot Therapist Isn't Sure What It's Doing*. Gizmodo. https://gizmodo.com/your-ai-chatbot-therapist-isn-t-sure-what-it-s-doing-1843608786

Sprecher, S., Treger, S., & Wondra, J. D. (2013). Effects of self-disclosure role on liking, closeness, and other impressions in get-acquainted interactions. *Journal of Social and Personal Relationships*, *30*(4), 497–514.

Stawarz, K., Preist, C., Tallon, D., Wiles, N., & Coyle, D. (2018). User Experience of Cognitive Behavioral Therapy Apps for Depression: An Analysis of App Functionality and User Reviews. *Journal of Medical Internet Research*, *20*(6). https://doi.org/10.2196/10120

Stieger, M., Nißen, M., Rüegger, D., Kowatsch, T., Flückiger, C., & Allemand, M. (2018). PEACH, a smartphone- and conversational agent-based coaching intervention for intentional personality change: Study protocol of a randomized, wait-list controlled trial. *BMC Psychology*, *6*(1), 43. https://doi.org/10.1186/s40359-018-0257-9

Stone, L., & Waldron, R. (2019). Great expectations and e-mental health:'The role of literacy in mediating access to mental healthcare'. *Australian Journal of General Practice*, *48*(7), 474–479.

Sucala, M., Schnur, J. B., Constantino, M. J., Miller, S. J., Brackman, E. H., & Montgomery, G. H. (2012). The Therapeutic Relationship in E-Therapy for Mental Health: A Systematic Review. *Journal of Medical Internet Research*, *14*(4), e110. https://doi.org/10.2196/jmir.2084

Suchman, L. A. (1987). Plans and situated actions—The problem of human-machine communication. *Learning in Doing: Social,Cognitive,and Computational Perspectives*. https://doi.org/10.2307/2073874

Sue, S. (1977). Community mental health services to minority groups: Some optimism, some pessimism. *American Psychologist*, *32*(8), 616–624. https://doi.org/10.1037/0003-066X.32.8.616

Sue, S., Fujino, D. C., Hu, L. T., Takeuchi, D. T., & Zane, N. W. (1991). Community mental health services for ethnic minority groups: A test of the cultural responsiveness hypothesis. *Journal of Consulting and Clinical Psychology*, *59*(4), 533–540. https://doi.org/10.1037//0022-006x.59.4.533

Sussman, L. K., Robins, L. N., & Earls, F. (1987). Treatment-seeking for depression by black and white Americans. *Social Science & Medicine (1982)*, *24*(3), 187–196. https://doi.org/10.1016/0277-9536(87)90046-3

Sutoyo, R., Chowanda, A., Kurniati, A., & Wongso, R. (2019). Designing an Emotionally Realistic Chatbot Framework to Enhance Its Believability with AIML and Information States. *Procedia Computer Science*, *157*, 621–628.

Ta, V., Griffith, C., Boatfield, C., Wang, X., Civitello, M., Bader, H., DeCero, E., & Loggarakis, A. (2020). User Experiences of Social Support From Companion Chatbots in Everyday Contexts: Thematic Analysis. *Journal of Medical Internet Research*, *22*(3), e16235. https://doi.org/10.2196/16235

Tam, T., Hewstone, M., Harwood, J., Voci, A., & Kenworthy, J. (2006). Intergroup contact and grandparent–grandchild communication: The effects of self-disclosure on implicit and explicit biases against older people. *Group Processes & Intergroup Relations*, *9*(3), 413–429.

Tantam, D. (2006). The machine as psychotherapist: Impersonal communication with a machine. *Advances in Psychiatric Treatment*, *12*(6), 416–426.

Thies, I. M., Menon, N., Magapu, S., Subramony, M., & O'neill, J. (2017). How do you want your chatbot? An exploratory Wizard-of-Oz study with young, urban Indians. *IFIP Conference on Human-Computer Interaction*, 441–459.

Thomas, N. T. (2016). An e-business chatbot using AIML and LSA. *2016 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*, 2740–2742.

Thompson, C. E., Worthington, R., & Atkinson, D. R. (1994). Counselor content orientation, counselor race, and Black women's cultural mistrust and self-disclosures. *Journal of Counseling Psychology*, *41*(2), 155.

Tielman, M. L., Neerincx, M. A., van Meggelen, M., Franken, I., & Brinkman, W.-P. (2017). How should a virtual agent present psychoeducation? Influence of verbal and textual presentation on adherence. *Technology and Health Care*, *25*(6), 1081–1096. https://doi.org/10.3233/THC-170899

Tschakarjan, S., & Ortega, M. J. (2019). The invisible wounds of armed conflict. *Swiss Archives of Neurology, Psychiatry and Psychotherapy*, *170*(04).

Vaidyam, A. N., Wisniewski, H., Halamka, J. D., Kashavan, M. S., & Torous, J. B. (2019). Chatbots and Conversational Agents in Mental Health: A Review of the Psychiatric Landscape. *Canadian Journal of Psychiatry. Revue*

Canadienne De Psychiatrie, *64*(7), 456–464. https://doi.org/10.1177/0706743719828977

Valério, F. A. M., Guimarães, T. G., Prates, R. O., & Candello, H. (2017). Here's What I Can Do: Chatbots' Strategies to Convey Their Features to Users. *Proceedings of the XVI Brazilian Symposium on Human Factors in Computing Systems - IHC 2017*, 1–10. https://doi.org/10.1145/3160504.3160544

Valmaggia, L. R., Latif, L., Kempton, M. J., & Rus-Calafell, M. (2016). Virtual reality in the psychological treatment for mental health problems: An systematic review of recent evidence. *Psychiatry Research*, *236*, 189–195. https://doi.org/10.1016/j.psychres.2016.01.015

Varcarolis, E. M. (2005). *CHAPTER 10 Developing Therapeutic Relationships*.

Venkatesh, Morris, M. G., Davis, G. B., & Davis, F. D. (2003). User acceptance of information technology: Toward a unified view. *MIS Quarterly*, 425–478.

Venkatesh, V., Thong, J.Y.L., & Xu, X. (2012). Consumer acceptance and use of information technology: Extending the unified theory of acceptance and use of technology. *MIS Quarterly*, *36*(1), 157–178.

Vimalkumar, M., Sharma, S. K., Singh, J. B., & Dwivedi, Y. K. (2021). 'Okay google, what about my privacy?': User's privacy perceptions and acceptance of voice based digital assistants. *Computers in Human Behavior*, *120*, 106763.

Vitak, J. (2016). A digital path to happiness? In *The Routledge handbook of media use and well-being: International perspectives on theory and research on positive media effects* (pp. 274–287). Routledge.

Vogel, D. L., Wade, N. G., & Hackler, A. H. (2007). Perceived public stigma and the willingness to seek counseling: The mediating roles of self-stigma and attitudes toward counseling. *Journal of Counseling Psychology*, *54*(1), 40–50. https://doi.org/10.1037/0022-0167.54.1.40

Walker, M. A., Cahn, J. E., & Whittaker, S. J. (1997). Improvising Linguistic Style: Social and Affective Bases for Agent Personality. *ArXiv:Cmp-Lg/9702015*. http://arxiv.org/abs/cmp-lg/9702015

Wallis, P., & Norling, E. (2005). The Trouble with Chatbots: Social skills in a social world. *Virtual Social Agents*, *29*.

Walther, J. B. (1996). Computer-mediated communication: Impersonal, interpersonal, and hyperpersonal interaction. *Communication Research*, *23*(1), 3–43.

Walther, J. B., & D'Addario, K. P. (2001). *The Impacts of Emoticons on Message Interpretation in Computer-Mediated Communication*. https://doi.org/10.1177/089443930101900307

Wambsganss, T., Winkler, R., Söllner, M., & Leimeister, J. M. (2020). A Conversational Agent to Improve Response Quality in Course Evaluations. *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems*, 1–9.

Wang, D., Wang, L., Zhang, Z., Wang, D., Zhu, H., Gao, Y., Fan, X., & Tian, F. (2021). " Brilliant AI Doctor" in Rural China: Tensions and Challenges in AI-Powered CDSS Deployment. *ArXiv Preprint ArXiv:2101.01524*.

Warren, S. D., & Brandeis, L. D. (1890). Right to privacy. *Harv. L. Rev.*, *4*, 193.

Watsford, C., & Rickwood, D. (2013). Disconfirmed expectations of therapy and young people's clinical outcome, help-seeking intentions, and mental health service use. *Advances in Mental Health*, *12*(1), 75–86. https://doi.org/10.5172/jamh.2013.12.1.75

Waytz, A., Cacioppo, J., & Epley, N. (2014). Who Sees Human? The Stability and Importance of Individual Differences in Anthropomorphism. *Perspectives on Psychological Science : A Journal of the Association for Psychological Science*, *5*(3), 219–232. https://doi.org/10.1177/1745691610369336

Waytz, A., Morewedge, C. K., Epley, N., Monteleone, G., Gao, J.-H., & Cacioppo, J. T. (2010). Making sense by making sentient: Effectance motivation increases anthropomorphism. *Journal of Personality and Social Psychology*, *99*(3), 410.

Weisband, S., & Kiesler, S. (1996). Self disclosure on computer forms: Meta-analysis and implications. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 3–10.

Weizenbaum, J. (1966). ELIZA---a computer program for the study of natural language communication between man and machine. *Communications of the ACM*, *9*(1), 36–45.

Weizenbaum, J. (1976). *Computer power and human reason: From judgment to calculation*. San Francisco : W. H. Freeman. http://archive.org/details/computerpowerhum0000weiz

Weizenbaum, J. (1983). Eliza—A computer program for the study of natural language communication between man and machine. *Communications of the ACM*, *26*(1), 23–28.

Westra, H. A., Aviram, A., & Doell, F. K. (2011). Extending Motivational Interviewing to the Treatment of Major Mental Health Problems: Current Directions and Evidence. *The Canadian Journal of Psychiatry*, *56*(11), 643–650. https://doi.org/10.1177/070674371105601102

Winslade, W. J., & Ross, J. W. (1985). Privacy, confidentiality, and autonomy in psychotherapy. *Neb. L. Rev.*, *64*, 578.

Yu, R., Hui, E., Lee, J., Poon, D., Ng, A., Sit, K., Ip, K., Yeung, F., Wong, M., & Shibata, T. (2015). Use of a therapeutic, socially assistive pet robot (PARO) in improving mood and stimulating social interaction and communication for people with dementia: Study protocol for a randomized controlled trial. *JMIR Research Protocols*, *4*(2), e45.

Zhou, L., Gao, J., Li, D., & Shum, H.-Y. (2018). The design and implementation of XiaoIce, an empathetic social chatbot. *Computational Linguistics*, *Just Accepted*, 1–62.