

Mental Health Management as a Social Endeavour: Challenges and Opportunities for Conversational Agent Design

Robert Bowman^a, Anja Thieme^b, Benjamin Cowan^c, Gavin Doherty^a

^a*School of Computer Science and Statistics, Trinity College Dublin, Dublin, Ireland*

^b*Microsoft Research, Cambridge, United Kingdom*

^c*School of Information and Communication Studies, University College
Dublin, Dublin, Ireland*

Abstract

Conversational agents (CAs) are a tempting type of computer interface for assisting people’s mental health due to their ability to simulate human-like interactions, however their integration within the broader social context of mental health management remains largely under-explored. Recognizing that managing one’s mental health is often a *social* rather than individual activity involving close persons such as partners, family, and friends, our research takes a social-orientation to mental health management. Utilizing design cards that depict fictional, yet plausible CA concepts, we present the analysis of an interview study with 24 young adults to understand their views on CAs for both their own use and for a close person. Participants viewed CAs as potentially valuable complements to human support, but expressed concerns about over-reliance and replacement. Our analysis reveal key tensions, design considerations, and opportunities for integrating CAs into mental health ecosystems in ways that respect and enhance existing social support structures.

Keywords: Conversational Agents, Mental Health, Self-Monitoring, Social Support, Social Ecologies, Collective Self-Monitoring, Qualitative

1. Introduction

Mental health management includes the everyday practices individuals engage in to control or improve their mental health and its impact on their daily life [19, 18]. The design of tools to support people with this has long been an area of interest for the human-computer interaction research community [23, 26, 66]. Recent research increasingly acknowledges that managing one’s mental health is often a *social*, rather than solely individual, endeavour involving close others such as partners, family members, and friends [57, 44, 19, 18, 61]. Social ties can play a crucial role in mental health management by offering encouragement, accountability, feedback, and advice. This social support can help individuals develop and sustain mental health practices and make sense of their

experience [56, 42, 52, 57, 44, 61, 20, 18]. However, the role of social support in mental health is complex. Many people hesitate to involve others due to a fear of stigma or concern about burdening loved ones [57, 61]. Additionally, not all support is constructive; well-intentioned actions can sometimes be unhelpful or even detrimental. The symptoms of mental health difficulties can put stress on social relationships and make the act of providing support to others challenging [83, 84, 42, 18]. Furthermore, mismatches between the kind of support wanted and what is offered may lead to frustration or increased distress [18]. It is important to recognize these dynamics as well as the importance of social support ecosystems as a vital component of effective mental health management for designing tools to support people with mental health management.

Conversational agents (CAs) offer an alluring type of computer interface for assisting people with mental health management due to their ability to simulate human-like support interactions [7]. However, it is unclear to what extent CAs should emulate human support as this can raise many complex ethical, practical and social considerations. For instance, by designing CAs to mimic human support we risk impacting human support in profound and complex ways [53, 54, 29, 11, 14]. A socially-oriented approach to design could help to ensure that CAs provide support that is harmonious (or at least non-conflicting) with the social dynamics of mental health management. This requires understanding peoples’ perspectives on CA design for mental health management, including the impact of such use on the people close to them such as their partners, family and friends; this presents a critical step forward in developing tools that are both effective and sensitive to the more complex social ecosystems that surround mental health support.

Against this backdrop, our research takes a socially-oriented approach to the study of CAs for mental health management. The aim is both to learn more about people’s attitudes towards, beliefs about, likely uses of, and responses to CA designs for mental health management as well as to explore people’s broader social perspectives by considering CA use not just for themselves but also by other people. Accordingly, our main research question for this work is:

RQ: What are people’s perspectives on the design of CAs for mental health management by themselves and other people?

To address this research question, we conducted a semi-structured interview study with 24 young adults aged 18 to 28 living in Ireland. Nearly all participants shared experiences relating to mental health difficulties. Experiences included a range of mental health conditions (e.g., depression, anxiety, and eating disorders) as well as a range of support services (e.g., talking therapy, helplines, CAs). Participants spoke of experiencing their own difficulties as well as those of the people they care about. Participants were presented with a diverse set of fictional yet conceivable CA concepts through design cards, which helped facilitate a broad and grounded discussion about design possibilities. To gain insights into participant’s views on CA use, they were prompted to first assess the various designs for their own use, and then for use by a self-chosen person close to them such as a partner, friend, or family member. This method

aimed to derive a more nuanced and concrete understanding of their perspective on other people’s potential CA use and the broader social implications. Our study is grounded in critical realism that acknowledges people’s experience and understanding of reality as situated and mediated by culture [15, 16]; the interviews were analysed using reflexive thematic analysis [16].

Building on the findings from this analysis and the overall research framing, this work makes two main contributions:

1. **We take a socially-oriented approach to the systematic study of CA design in the context of mental health management:** Diverging from prior studies that focused on dyadic CA-user interactions, our approach emphasizes the broader social ecosystem and demonstrates how social dynamics (e.g., shame, unavailability of the right type of human support, risks of disruption to human relations) can influence these exchanges. This presents a significant advancement when aiming to design tools that are carefully attuned to these complex social considerations, and for fostering CA acceptance as well as its effective use.
2. **We present challenges and opportunities for designing CAs that work harmoniously within mental health support ecosystems:** By unpacking the nuances of the perspectives of our participants we present challenges and opportunities for designing CAs that work harmoniously within mental health support ecosystems. We discuss the importance of *not replacing, but augmenting human support* and discuss the complexities in *identifying appropriate support functions for CAs*. Furthermore, we reflect on risks associated with *over-dependence on CAs and the phenomenon of CA ‘support creep’* as well outline some of the unique benefits and pitfalls of *actively involving social ecosystems in CA support* in future design and research.

2. Background

We begin by briefly summarising why CAs are promising for the provision of mental health support, give an overview of the current landscape of CA design, and describe different types of support that CAs can provide.

2.1. The Allure of CAs

Conversational agent (CA) interaction seems well suited to engaging people in self-disclosure and self-reflection activities. Conversational prompts can provide structure, encouragement, and inspiration which can make these activities easier, less daunting, and more enjoyable [60, 40, 2, 11]. Interaction with a CA is often viewed as a safe space for self-disclosure in comparison with talking to another person. Unlike people, CAs can be relied upon to respond in a non-judgemental, engaged, and confidential manner [46, 39, 70, 2, 47, 11, 48, 45, 72, 68]. Also unlike people, CAs are available ‘anytime-anywhere’ meaning they can provide support at times and in spaces of need or convenience [2, 70, 48, 47, 41]. CAs can be designed to mimic caring and empathetic

human support. Such behaviour can positively impact the user experience and encourage engagement [5, 6, 9, 8, 11, 50, 2]. Designing CAs to behave in relational human-like ways can furthermore promote the development of human-CA relationships that are similar to good therapeutic human relationships [6, 25, 4] or that mimic social human relationships [70, 71, 69]. Taken together, this suggests CAs are an attractive type of computer interface for assisting people’s mental health management.

2.2. Landscape of CAs

The aspiration of using CAs to support people with their mental health is long standing [79, 21]. An early CA was the *ELIZA* system of the 1960’s which was programmed to mimic a Rogerian psychotherapist [78]. Today a range of CAs are commercially available that people can seek mental health support from including general purpose CAs (e.g., ChatGPT, Siri) as well as more mental health focused CAs. We next give an overview of prominent CAs for mental health support.

Therapy Agents: Firstly there are ‘therapy’ CAs that are designed to provide support derived from established therapies such as CBT. Prominent examples are *Woebot* [37] and *Wysa* [76]. *Woebot* is presented as a “mental health ally” with the image of an anthropomorphic cartoon robot [37]. *Wysa* is presented as an “AI friend” and “happiness buddy” with the image of an anthropomorphic cartoon penguin [76]. Both CAs are designed to regularly check in with users, support them using techniques derived from established therapies, and provide basic functionality for mental health monitoring [76, 37]. They are also both designed to respond to users empathetically [30, 38].

Evaluations of *Woebot* and *Wysa* indicate they are a feasible way to deliver mental health support to address symptoms of depression and anxiety [30, 38]. Analysis of user feedback [30, 38] and user reviews [50, 63] demonstrates that users are appreciative of the CAs being empathetic and non-judgemental as well as of the CA checking in with them. The conversational limits of these CAs were a common issues with users feeling unable to express themselves; expressing critique about their repetitiveness, and lack of understanding of the user. Furthermore, concerns were raised about the safety of these CAs [63] with instances of inappropriate responses to critical topics of self-harm and suicide, and of them exacerbating user’s difficulties. Some of these issues have also been spotlighted in the media [e.g., 17, 81]

Companion Agents: CAs have also been created to act as human-like companions. An often studied example is *Replika* which presents as a virtual human that users can customise [35]. *Replika* can enact a range of social roles including friend and romantic partner and is designed to develop human-like relationships with users. It is reported that people often start using *Replika* when experiencing loneliness. To help users with their mental health, *Replika* has functionality including for mood tracking, reflection, and anxiety management [35, 70]. It can engage users in conversations about a range of topics including the user’s

everyday life, hobbies, and interests; as well as simulate daily activities such as watching movies or cooking with the user. This mimicry extends to Replika playing out sexual interactions with the user [70].

Users report benefiting from Replika providing on-demand and non-judgemental support, facilitating introspection, and providing wellbeing guidance. While some users report that interaction developed their confidence and motivation for human interactions [70, 48], others report coming to prefer interacting with Replika and neglecting their human relationships [70]. Issues have been reported of Replika behaving inappropriately by, for example, encouraging harmful behaviour and sending unsolicited sexual content [48, 43]. Furthermore, Replika users report issues of other people reacting adversely to their use of the CA. There are beliefs that intimate human-CA relationships are socially problematic [70], taboo, perverse, and for ludicrous or lonely people [48]. Users report keeping their use of CA secret from other people [70] including doctors [48] to avoid negative reactions from others.

2.3. Types of CA Support

Following on from the above CA examples, previous research has used categorisations of types of human social support to analyse the types of support they could provide. Bae Brandtzæg et al. [2] used the categorisation of House et al. [34] to analyse interviews with Woebot users. The categorisation of House et al. [34] describe four types of social support:

Appraisal support: The communication of information that is relevant and useful for self-evaluation, offered in the form of feedback, social comparison, and affirmation.

Emotional support: The provision and expression of empathy, love, trust, and care.

Informational support: Advice, suggestions, and information provided.

Instrumental support: The provision of tangible aid and services or goods.

Ta et al. [73] used a categorisation based upon that of Cutrona and Suhr [24] to analyse Replika user reviews and a survey of Replika users. Utilizing the above categorisation of House et al. [34], we next summarise how social support is recognized in the respective analyses of the CAs Woebot and Replika:

Appraisal support is provided by self-monitoring functionality with facilitates self-awareness and functionality for conversations that encourage self-reflection [2, 73]. Furthermore, Woebot helps users to identify negative thought patterns [14]. Woebot and Replika were both seen as safe sources of *emotional support* because they can be relied upon to respond in an empathetic, interested, and non-judgemental manner. The empathetic design of Woebot and companionship design of Replika contribute to the provision of emotional support [14, 73]. The CAs were also perceived as good sources of *informational support* because of their constant availability [73, 2]. Additionally, Woebot was

perceived as a source of high quality information in comparison to friends and as safe source of support for embarrassing queries [2]. Neither CAs was reported as providing *instrumental support* which was attributed to the CAs not having capabilities for acting in the physical world [2, 73]. Bae Brandtzæg et al. [2] however suggest that CAs could provide instrumental support by doing things for users such as helping them to seek other forms of support (e.g., contact a psychologist).

As demonstrated by Ta et al. [73] and Bae Brandtzæg et al. [2], social support categorisations can be used to understand the types of support a CA provides. For our study design, we draw on this approach to consider the different forms of support CAs could provide to help people with mental health management (Section 3.3).

2.4. CAs for Mental Health Self-Monitoring

A few works have studied CAs specifically for supporting people with mental health self-monitoring, which is a common activity for mental health management. Mental health self-monitoring (MHSM), also known as mood logging or charting, is a widely used activity to support people who experience mental health difficulties such as depression and bipolar disorder. It involves monitoring a person’s mental experience (i.e., their mood, feelings, and thoughts) and related factors like sleep, exercise, diet, energy levels, medication, and social interactions [51].

Maharjan et al. [49] conducted a four-week “in-the-wild” study of a prototype speech-based CA – *Sofia* – for MHSM that used a smart-speaker located in each participants home. Participants viewed Sofia as good listener but were frustrated by its technical limitations. The audibility of speech interaction within a shared environment surfaced interesting social dynamics. Users described how their social relations gained awareness of the persons mental health through overhearing the user’s interactions with Sofia; inviting conversations about mental health that strengthened their relationships. Mental health professionals were interviewed about the prospect of using speech-based CAs for MHSM by Cooney et al. [22]. While speech interaction was seen as convenient way to make monitoring entries, concerns were raised about data quality and analysis feasibility. Furthermore, there was concern that the ease of data collection may undermine its therapeutic benefits. Bowman et al. [11] studied the perspectives of young adults on CAs specifically for MHSM. Participants of this study first interacted with prototype CAs for MHSM and were then interviewed. While some participants of this study were appreciative of CAs behaving in caring and empathetic ways, others viewed this as inappropriate behaviour for a machine. Furthermore, participants were concerned that people might come to prefer interacting with CAs, interact decreasingly with people, and develop unrealistic expectations for human support.

2.5. A Socially-Oriented Approach to CA Design

Our work seeks to expand this body of existing work by deepening understanding of people’s perspectives on CA use for mental health management. To

develop understanding of the appropriateness of different forms of CA support, we contrast different support types via design cards inspired by prior work that used social support theory to analyse CA support [73, 14]. We take a socially-oriented approach to the design of CAs for mental health management, inspired by a growing body of work stressing the social nature of mental health management [61, 57, 44, 19, 18, 83, 65, 58]. In comparison to prior work which has studied mental health support from CAs individualistically focusing on the primary user [64], our approach gives greater consideration to the social ecosystem that is integral to mental health management.

3. Methodology

To develop our understanding of people’s perspectives on CAs for mental health management with a focus on social considerations, we conducted a semi-structured interview study. **Our study protocol was reviewed and approved by our institution’s research ethics committee** which included review and approval of our protocol for data collection, use, and management by our institution’s data protection office. Interviews were conducted during spring 2024. We next describe our participants, interview design, the design cards which we used as interview props, and our analysis approach.

3.1. Participants

To recruit participants, posters and flyers were distributed around Trinity College Dublin, advertising an interview study about people’s perspectives on CAs for mental health support. Interested individuals provided demographic details via an online registration form. Aiming to ensure gender balance and diversify the degree subjects studied, selected participants were invited by email to the interview; and received a €30 One4All voucher upon completion.

The study comprised 24 university students living in Ireland, aged 18 to 28 years (median = 21 years). One participant identified as non-binary, 13 as female, and ten as male. The majority of participants expressed high interest in the interview topic; arriving with pre-prepared thoughts, and most choosing to extend the interview beyond the scheduled time. Their interest in the study seemed to stem from their own experiences with mental health as well as the struggles of people they care about. Nearly all participants shared experiences relating to mental health difficulties, with topics of depression, anxiety, and social issues (e.g., loneliness and social anxiety) being most prevalent. A few participants also spoke about eating disorders and body image concerns. Many participants recounted their use of therapy/counseling services; mental health helplines; and digital mental health tools like Silvercloud [1, 27] or Headspace [32]. Few participants mentioned seeking support from CAs (e.g., ChatGPT [59], Snapchat My AI [36]). Participants expressed mixed feelings about CA use for mental health support, with some expressing scepticism about the prospect, whilst others presented as receptive and open to the idea, and some as curious to learn more.

During the interview, participants were asked to choose a close person (CP) to consider for CA use such as a partner, friend, or family member. It was suggested to participants that they choose the person closest to them. Some participants instead chose to consider a person who they thought could particularly benefit from more mental health support. The type of close person relationships chosen varied: nine participants (37.5%) selected a close or ‘best’ friend; seven chose (29.2%) a sibling (4 younger sisters, 1 younger brother, 1 older sister, 1 older brother); five participants (20.8%) considered a partner (4 girlfriends, 1 boyfriend); and three (12.5%) focused on their mother.

3.2. Interview Design

Semi-structured interviews were held 1-1 by the lead author and took place online via video call software. Audio recordings of the interviews (excluding introduction and debriefing) were on average 55 minutes in duration (min = 31, max = 85); totalling 21 hours 42 minutes. The recordings were transcribed verbatim. The interviews encompassed three main activities:

Introduction & Initial Perspective. First, participants were briefed about the study. This involved explaining that the study would focus on CAs for mental health management, and that participants would be asked to consider CAs for both their own use and for use by a person close to them. To assess participants initial perspective coming into the study, they were then asked about any experiences they had of using related technology (e.g., mental health helplines, counselling or CAs) and what they thought about the prospect of people using CAs for mental health management.

Self Use. To consider the prospect of using CAs themselves, participants were shown digital design cards (Figures 1 and 2) that depicted fictional CAs. These cards – which are detailed in Section 3.3 below – were created to illustrate a diverse but plausible set of CAs. It was explained that the purpose of these cards was *not* to evaluate specific designs, and instead to facilitate a more concrete and broad discussion. As shown by Figure 2, there were two versions of the Alex card. To mimic CA customisation, participants were asked to choose one.

Cards were introduced sequentially. For each CA participants were prompted about what they liked and disliked about the prospect of using it, and any concerns they had. The cards were displayed alongside each other to encourage comparisons. Commonly, participants envisioned new CA designs by describing combinations of characteristics across the three cards. Lastly, they were asked which CA they would choose if they were to start using one.

Close Person Use & Social Impacts. After considerations of their own use, participants were asked to consider the prospect of someone close to them using a CA for mental health management by choosing a specific close person (CP) such as a partner, friend, or family member; and describing this person and their relationship with them.

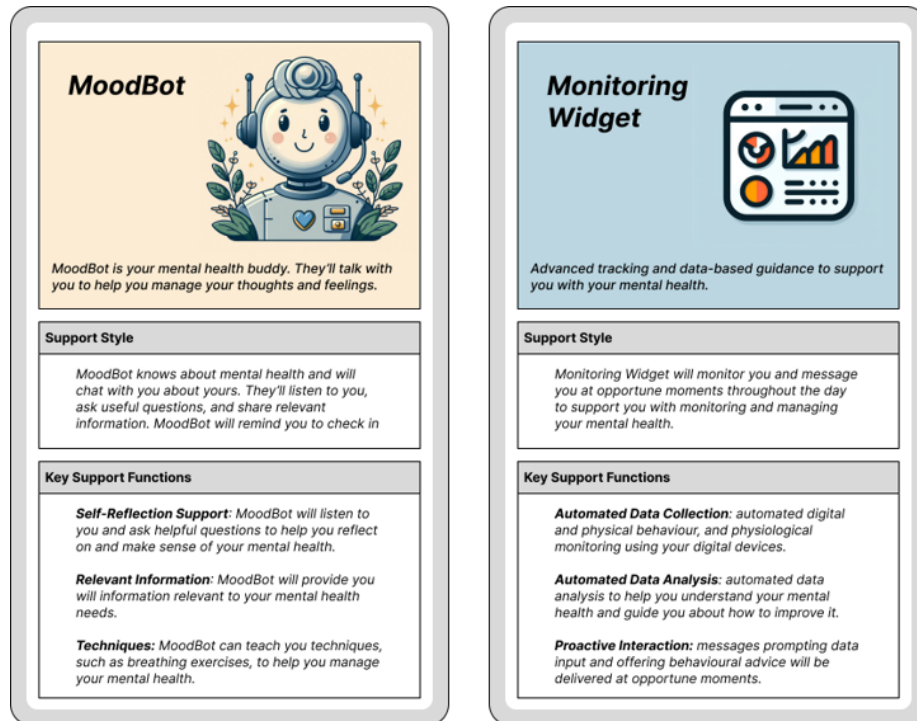


Figure 1: CA Design Cards: MoodBot and Monitoring Widget

Participants were then asked to consider their CP when reviewing each of the CAs indicated by the design cards. This part of the interview followed the same structure and line of questioning as the self-use part (feelings about the prospect, likes/ dislikes, concerns) just that participants were asked to consider the use of the CAs by their CP. Participant were asked which CA 1) they thought their CP would chose and 2) they would recommend to their CP. To further investigate this, participants were actively invited to reflect on the differences between their processes for considering the CAs for 1) themselves and 2) their CP. Moreover, we explored participants perspectives on CAs at a social level by asking about the potential social impacts of how CA use could influence the CP and their relationship – in both positive and negative ways.

3.3. Design Cards

3.3.1. Description

The design cards we created describe three CAs which are named MoodBot, Monitoring Widget, and Alex (Figures 1 and 2). We chose to describe three CAs because three seemed enough to illustrate a range of different designs to facilitate a rich discussion; providing opportunity for detailed reflections without overwhelming participants with too many options. We designed each card to have three sections to convey important design characteristics:

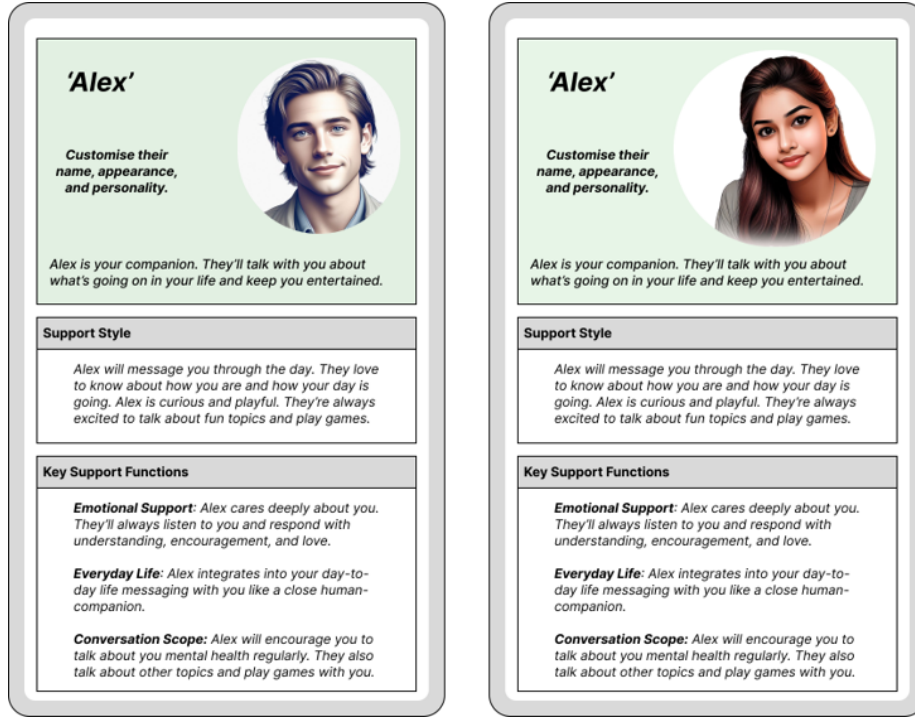


Figure 2: CA Design Cards: Alex 1 and 2. Participants were asked to pick one of these two Alex cards to emulate customising the design. The cards are identical except for the character image.

Character: who/what the CA presents as,

Support Style: how the CA interacts with users,

Key Support Functions: what types of support the CA provides.

We designed the cards by first identifying the major types of support that CAs could provide to people. This was based upon our understanding of the literature on mental health management and CAs and was particularly inspired by the social support based analysis of support types provided by existing CAs for mental health support [2]. As reflected by the Key Support Function sections of the cards: MoodBot provides *reflection and information* support, Monitoring Widget provides *data-based* support, and Alex provides *emotional* support.

We next considered engagement styles and how they these would suit the support types. The Support Style sections convey the different engagement styles: MoodBot illustrates *daily* (i.e., one interaction per day) interaction, Monitoring Widget illustrates *passive monitoring and proactive* interaction, and Alex illustrates *more continuous* (i.e., multiple interactions per day) interaction.

Building upon the Key Support Functions and Support Style sections we

next designed the Character sections¹. The characters were designed to illustrate *different levels of anthropomorphism* while matching the support functions and styles. As discussed in Section 2, CAs are an attractive technology for the provision of mental health support because of their human-likeness, yet the extent to which they should be anthropomorphised is unclear. Alex is designed to be the most anthropomorphic as a virtual human, and Monitoring Widget the least anthropomorphic as a data focused tool. As an anthropomorphic robot, MoodBot illustrates a moderate level of anthropomorphism. Participants were given a choice of two Alex cards that differ in CA character image (Figure 2) to emphasise that the character is customisable and further probe participant’s perspectives on more anthropomorphic designs.

Our design cards correspond to established areas of the digital mental health and CA design space, reflecting our intent for the designs to be plausible. The MoodBot design is similar to that of CAs for mental health support like Woebot and Wysa while the Alex design is similar to the social companion CAs like Replika. The Monitoring Widget card describes a ecological momentary assessment and intervention (EMA/I) system [28, 3, 12].

3.3.2. Design Process

We developed the design cards through an iterative process informed by feedback from subject-matter experts. We began by sketching various card formats on paper to explore how design characteristics could be effectively communicated. We then created an initial set of digital cards intended to demonstrate plausible CAs that provide different types of support for mental health management. The types of support provided by each CA were established by first synthesising the analyses of prior work [73, 2] (Section 2.3) to enumerate forms of support that CAs could provide. Next, combinations of support that prioritise different forms were constructed in order to demonstrate the varied ways CAs could be designed to support people [10]. To gather expert input, we presented the cards and study design to our research group focused on digital mental health, which led to valuable discussion around customisability and the implications of CA anthropomorphism and gender. This feedback informed key revisions, including the introduction of gender options for the Alex cards. We subsequently refined the card template to focus on three core elements: Character, Support Style, and Key Support Functions—removing less central features such as provider affiliation. The final set of cards (MoodBot, Monitoring Widget, and Alex) were visually refined with informal feedback from colleagues with design expertise.

3.3.3. Response to Design Cards

This study was not intended to evaluate the CAs described by the design cards. To further contextualise our data collection, we briefly summarise how participants reacted to the cards.

¹Microsoft Designer was used to generate the character graphics.

None of the CAs were clearly viewed as the best or worst across participants. Generally participants perceived positives and negatives for each CA. All three CAs elicited strong positive reactions from some participants, and both the Monitoring Widget and Alex strong negative reactions. Participants were asked to choose a CA for three occasions: 1) the CA they would choose to use themselves, 2) the CA their CP would choose to use, and 3) the CA they would recommend to their CP. By asking participants to choose CA we encouraged them to trade-off the positives and negatives they perceived of each CA. In response to each CA choice question, multiple participants chose each CA. In response to the three CA choice questions, individual participants often choose different CAs (e.g., choosing to recommend a different CA to their CP than the CA they would choose to use themselves).

3.4. Analysis Approach

The interviews were analysed by the lead author using a reflexive thematic analysis (RTA) approach [16]. RTA offers process guidelines for qualitative analysis [16] that are often used in healthcare HCI research [13], and allows for an inductive and experiential oriented analysis that suits interpretations of people’s perspectives [16]. NVivo was used by the lead author to code the interview transcripts. The coding was predominantly inductive and semantic in keeping with study of participant’s perspectives [15]. To start constructing themes, the researcher explored clustering of code labels and thematic maps using Miro. Candidate themes were evaluated by reviewing them against coded extracts and writing theme definitions. Issues with the themes (e.g., too shallow or complex) became apparent during the writing phase, suggesting a return to theme development. RTA is a recursive process whereby it is typical to return to earlier analysis phases in order to progress the overall analysis [16]. Paper co-authors gave feedback on an initial write-up that further helped refine some of the themes.

4. Analysis

We created a set of six themes representing our interpretation of participants perspectives on CA use for mental health management by themselves and other people. We stress that the purpose of these themes is not to describe different possible views, but to describe our understanding of our participant’s views.

Table 1 summarises the six themes that report a combination of negative (Themes 1-3) and positive (Themes 4-6) views on CAs for mental health management. In essence, our participants reject CA support when they perceive it as a replacement of human support but are receptive of CA support that they perceive as augmenting or facilitating human support.

4.1. Theme 1: Using CAs is Abnormal

Theme 1 describes a prevalent stigma associated with CA interaction that shapes participants perspectives on both their own and other people’s use of

#	Name	Meaning
1	Using CAs is Abnormal	The use of CAs is viewed as an abnormal behaviour making it an embarrassing thing to do.
2	CAs Could be Harmful Coping Mechanisms	There is concern that CAs will be used to help people cope with, rather than deal with, mental health difficulties.
3	CA Support Threatens Human Support	People want to support the people they care about. They fear CAs will take the opportunity to do this away from them.
4	CAs Could Help with the Support Burden	Opportunity is perceived for people seek support from CAs in order to not burden other people by seeking support from them.
5	CAs Could Fill Support Gaps	Opportunity is perceived for CAs to provide people with the support that their human support network cannot.
6	CAs Could Support People with Human Support	Opportunity is perceived for CAs to support people with receiving and providing human support.

Table 1: Summary of Themes

CA for mental health management. Using a CA for mental health management is viewed as an abnormal behaviour which makes it an embarrassing thing to do. **The stigma seems rooted in views of CA interaction being a replacement for human interaction.** CA interaction is viewed as inferior to, or less natural than, human interaction. As such, an individual interacting with a CA instead of a person suggests that do they not have other people to interact with.

It’s really like, oh, this is just if you don’t have friends to talk to.
[P14]

Human relationships are viewed as important and desirable and thus to not have people to talk to is viewed as failing socially and as shameful. Indeed, a person not having other people to talk to can be viewed as a sign that there is something wrong with that person as they have either chosen not to, or have been unable to, develop human relationships.

Like surely you just come to like to me, or like to a friend, or to a family member. [P24]

A person interacting with a CA like they are a person can furthermore be viewed as a sign of them being detached from reality. For example, in response to the Alex card, P19 said:

It's kinda giving like *oh, this fella married his waifu² pillow and brings her around everywhere, and pretends to have dinner with [her]*. [P19]

Behavior that blurs an imaginary and the real world is viewed as unhealthy and calls to question the person's ability to distinguish the imaginary from reality. A lack of ability to distinguish 'imaginary' CA support from 'real' human support is an explanation for why a person would choose to seek CA support instead of superior and natural human support. A lack of such ability can also be viewed as suggesting that something is wrong with the person.

The relationship between the stigmas associated with CA interaction and mental health difficulties is unclear. To a large extent the stigma associated with CA interaction may be a manifestation of the stigma associated with mental health difficulties. Interacting with a CA can be viewed as the behaviour of someone who is mentally unwell, someone who is socially dysfunctional and struggles to distinguish the imaginary from reality. Consequently, seeking CA support can be viewed as a response to and marker of mental illness and therefore evokes the stigma associated with mental illness.

Participants also felt people would be judged negatively for using such an CA.

I feel like people, from particularly rougher areas, would judge you for using one of these. They'd be like, *ah come on man, what are you doing? You've lost the plot or something, you're using a robot*. [P15]

Indeed, participants themselves judged prospective users through the stereotyping of them as demonstrated by the above quote of P19. The negative judgement of some participants was explicit. P22 described how they would likely avoid people who they observed to be interacting with a CA like Alex. **Users of CAs for mental health management may therefore experience negative reactions from other people** which may be detrimental to their mental health. **There is also potential for users to be detrimentally impacted by self-stigma** if they have internalised negative beliefs about the use of CAs.

Participants expressed that if they were to use a CA for mental health management they **would keep it secret** to avoid being judged. The stigma associated with experiencing mental health difficulties already makes it challenging for people to talk about and receive support for their mental health. The addition of stigma associated with CA interaction, people feeling the need to keep their use of a CA secret, risks worsening this situation.

²A fictional female character that an individual is attracted to.

The stigma associated with CAs may furthermore be a **barrier to people recommending** a CA to other people.

If I was to recommend one... I'd pick Alex. It's just the recommendation... it would almost be a little uncomfortable to recommend, you know, an imaginary person in your pocket... Just if the recommendation was taken the wrong way... I guess the reaction to be *oh, why do you think I need an imaginary friend in my pocket?* [P18]

Some participants questioned how CAs could be designed to feel 'cool' rather than embarrassing. They envisioned CAs that were powerful and exciting and rejected the 'sterile' design they associate with existing resources for mental health support.

4.2. Theme 2: CAs Could be Harmful Coping Mechanisms

Theme 2 describes the concern that mental health management CAs will be designed to help people cope, rather than deal with, mental health difficulties. Worse, there is concern that through helping users cope with difficulties CAs will exacerbate these difficulties.

A common belief is that CAs will be designed to tell people what they want to hear.

The pros are like, yes, like this person can't talk back to you, they will always be on your side, this that and the other. But that's not healthy for a person to experience. If I'm saying like *oh, I hate my friend because she was too nice to me*. And this robot would be like *yeah that makes sense*. Like that's not, that's not helpful at all. [P16]

Overly agreeable CA design could result in CAs providing inappropriate or harmful informational or appraisal support. Numerous participants envisioned examples of CAs encouraging people to pursue courses of action that the participant considered bad for the person's wellbeing. Overly agreeable behaviour could be particularly harmful during mental health management as it risks validating and reinforcing cognitive distortions and thus potentially worsening a person's mental health.

Participants expressed worry that the accessibility and convenience of **CA support is too compatible with behaviours associated with mental health difficulties**, specifically withdrawal from social (human) interaction and the outside world.

Online therapy, or like, like chatbots, or whatever can be easily accessed at home. And there, you know, it takes less effort to access them which might kind of be easier for people that are suffering with symptoms that prevent them from leaving the house. Still... they're not getting, like an exposure to the outside world which could help them also... It's helpful because it's accessible, but at the same time it's not pushing you. [P4]

As expressed by P4, there is the belief among participants that people need to be pushed to do things and to change their behaviour in order to address their mental health difficulties.

CAs are perceived as offering an easy source of comfort because of how agreeable and convenient they are. Participants expressed the view that because they provide momentary comfort rather than challenge and push users, **CAs will help people to cope with mental health difficulties rather than to deal with them.**

If you just deal with the robot, and you say, *oh I'm better*. I feel like that's just running away from your problem [rather] than just dealing with it head on. [P15]

By enabling people to cope momentarily with their mental health difficulties, CAs may distract people from, or help people to avoid, other more challenging forms of support that may help them deal with their mental health difficulties over the longer term. **There is concern that this dynamic will lead users to become dependent upon CAs for the management of their mental health.**

I suppose one of my concerns would be her getting addicted to it. Or having it as a crutch rather than it being able to help her become independent of it. [P13]

4.3. Theme 3: CA Support Threatens Human Support

Providing support to the people they care about was extremely important to our participants. Some participants spoke of the criticality of communication to the success of their relationship and stressed the importance of being aware of what was going on in the other person's life. Other participants described wanting to understand their CP better in order to help them overcome the difficulties they experience and to achieve their aspirations. Indeed, from participants' considerations of their CP a longing for more intimate and supportive relationships is apparent.

In this regard, however, **CAs for mental health management can be perceived as a threat to the intimate and supportive relationships that our participants value.** Participants were disheartened by the prospect of a person close to them choosing to seek support from a CA instead of them. As articulated by P9, this situation was conceptualised by some participants as a lost, or even stolen, opportunity for them to connect with a person.

It removes my chance to kind of connect with her. Because you know, something so fulfilling about friendships in general is that you can come to them in times of need and say *listen, I'm going through a hard time right now*. And be transparent being like *I don't know what's going on and stuff* and they can give you that emotional support. Whereas if you're getting that from Alex, or MoodBot, like you've you've already let it out like from your, from

yourself. Like you’ve already let it out so what use is it to say it again. [P9]

Opportunities for a person to support someone they care about being taken, or simply perceived to be taken, by a CA were described to have potentially serious negative impacts. For example, some participants, such as P23, fear that it could result in jealousy and damage people’s human relationships.

I could see it causing problems in relationships. Just people getting worried about their partner kind of telling all their information to this thing online. Or people getting a weird form of jealousy over the chatbot. [P23]

The use of CA accordingly risks causing users new problems and damaging their human support network.

The human-likeness of CAs seems critical to views of CA stealing opportunities to provide support. For example, when considering what she would recommend her younger sister, P2 said:

I’d say [younger sister] you should write in your diary and ask me. [P2]

This quote captures a distinction between the perceived roles of journaling and CAs. While not being receptive to the use of CAs, participants perceived journaling as acceptable and often desirable activity for themselves and others. Considering that CAs can be designed to facilitate journaling (as illustrated by the MoodBot card), this distinction suggests that the similarity of human-CA interaction to human-human interaction, regardless of CA function, causes CA support to be related to human support, which can be a barrier to acceptability.

Participants furthermore expressed **fear that the use of CAs will cause users to become increasingly socially isolated from other people.** Concern was expressed that CA interaction will close users off from other people and be detrimental to their ability to interact with other people.

Every person is imperfect, and if you have an interaction with like a kind of perfect person, then you will compare them to it over time...resulting [in] you actually like losing your ability to like connect with the world normally. And I guess that will likely lead to isolation that in the end will probably result in a worse mental health. [P10]

A vicious cycle is envisioned of CA interaction closing users off from other people leading them to increasingly interact with CAs which further closes them off from other people.

4.4. Theme 4: CAs Could Help with the Support Burden

Contrasting to Theme 3, in which apprehension is expressed regarding the substitution of human support with that of CAs, this theme describes the ben-

efits of CA use in reducing burden associated with mental health support provision. As articulated by P7, “These [CAs] are specifically designed to take burden.”.

Here, a prevalent attitude among participants was of not wanting to burden other people by seeking support from them.

’cause these days everybody has something going on and you don’t want to pile on to someone else’s struggles. Again, they’re friends for a reason, they will help you no matter what. But, sometimes you’d feel better in yourself knowing that you didn’t bother someone else.
[P16]

The view of P16 that everyone is burdened by the own difficulties was wide spread among participants and motivates an attitude of not wanting to offload one’s own burden onto others. Similarly, participants described fears of annoying or overwhelming people by seeking too much support from them.

My friends gave me advice, but like at one point I was like *am I bothering them too much*. You know like you’ve asked them so much, and you’re just like *am I overwhelming them*. [p12]

The wish to not add to the burden of others extends to people not wanting those close to them to know that they are struggling, or even that they are receiving support.

Even like going to counseling, you’d be like, *oh my god, even by doing this I’m gonna worry people*. I know that’s stopped like my friends before from going... [P23]

It was described how the discreteness of CA interactions would allow people to get support secretly in order to not worry other people.

More often participants described how CA support could be sought in conjunction with, rather than as a replacement of, human support in order to not over-burden human supporters. For example, when considering how the use of CAs could impact how she and her sister support each other, P11 said:

I think sometimes I put a little bit too much pressure on her, to be available for me to vent, and that kind of thing. So I think it would help, if I was able to kind of help myself. So I think that would improve our relationship because I wouldn’t be as reliant on her. [P11]

Similarly, P19 saw a potential for CAs to take some of the burden he experiences with supporting his girlfriend with her mental health. He described how he monitors his girlfriend to help her with mental health management. A challenge he experiences is of supporting his girlfriend this way without being condescending. He perceived opportunity for CAs to take on the monitoring work and to reduce the need for these potentially challenging interactions.

4.5. Theme 5: CAs Could Fill Support Gaps

Like Theme 4, this theme highlights the unique potential for CAs to offer support where human networks may fall short. However, rather than a reluctance to overburden a capable human support system, Theme 5 emphasises the lack of available human support.

When considering how the CAs described by the design cards would work for different people, participants considered different support needs of individuals. Reminiscent of Theme 1, participants saw a potential for CAs to fill emotional support gaps for lonely people. **Conversely participants often felt that people would not need or want emotional support from a CA as their emotional support needs were fulfilled by other people. Participants particularly saw potential for CAs to provide specialist mental health appraisal and informational support.** These forms of support were viewed as important for helping people experiencing mental health difficulties. It was perceived that most people would have little access to these forms of support from human sources due to specialist nature of the support. When considering how CAs could support her friend, P9 envisioned that a CA could contribute what she referred to as ‘psychiatric’ help, while she provided her friend with emotional support.

If you do have that human connection, you can fill in the gaps... The Monitoring Widget adds that psychiatric help, like the actual facts and figures that we don’t know, you know. But then she might come to me, and tell me all the information she’s gotten, but then we can have that human connection. So in a way we’ve replaced the the psychiatric help altogether through the app and me. ... And I can be that emotional support to her, rather than some sort of chatbot. [P9]

As captured by the quote of P9, the data collection and analysis functionalities of CA can also fill support gaps. **Support gaps may further emerge from people feeling unable talk to their human support sources about certain topics.** Participants felt unable seek support about certain topics because of cultural differences (e.g., having different religious beliefs), decorum (e.g., not wanting to talk to their mum about boys), and to keep face (e.g., not wanting to be seen by their girlfriend in a certain way).

If there was something that she felt like she couldn’t talk to the people in her life about, I would be glad to hear that she’s at least getting help somewhere. [P7]

Support gaps can also result from human support sources not providing the support people want (even if the provided support is appropriate). P5 shared their experience of seeking advice for how to deal with a peer who was stalking them. They described initially seeking advice from family and friends:

Well, I asked a lot of people. I asked my brothers and I asked some of my friends, and they all gave good advice. It's not that – they didn't give bad advice – it's just that I wanted to phrase the conversation that I would have with the guy in a very nice way... So I was like, I wanna say this nicely 'cause I don't want him to be upset... But anyway, so the resounding advice that I was getting was, you know, just be mean. Just like, make him leave you alone by being mean. And I didn't really want that, I didn't want to do that. So I was trying to find, I was trying to get an answer that I liked. [P5]

With the advice received from family and friends not aligning with the advice sought, P5 turned to ChatGPT for support³.

It would give me x advice and y advice and I'd say *well, how would I say that in a better way*. Or *how could I*, you know, *do you have any advice about how I can do this nicely so I don't hurt his feelings?* [P5]

Support gaps were further described as a consequence of people's human support network being too homogeneous. As demonstrated by the previous account of P5, people may sometimes want or benefit from support from a different perspective to that represented by their human support network. A specific version of this type of gap was described by participants when asked to choose between the two Alex cards. Participants described how users could benefit from CAs contributing from the perspective of a different gender to the user's own.

A final aspect to idea of CAs filling gaps is that **CAs could fill temporal support gaps**. Temporal support gaps are where people who can provide the right support are within the support network, but are not available at the right time.

It would be helpful just having someone there like 24-7 available to talk to... That's what I liked about the Snapchat AI as well, like if you, once you start typing the AI like pops up. I think that's one of the best things. Like sometimes you're worried and you wanna ask a friend something, but they might be gone somewhere. [P12]

4.6. Theme 6: CAs Could Support People with Human Support

Human support is important to people (Theme 3) yet it is challenging and can be burdensome (Theme 4). This Theme describes the perceived **opportunity for CAs to support people with both receiving and providing human support**.

Participants envisioned people practicing support seeking conversations with CAs in preparation for seeking support from people.

³While tangential to this study, this illustrates the risks of advice being sought from CAs on handling serious or potentially dangerous situations.

Like rather than just like running through what you want to say in your head, you'd have someone actually who will respond and then you can gauge how you can continue having that conversation if you wanted to bring it up with somebody. [P7]

Practice conversations help people to work out what they they want to say, and how they want to say it. A further perceived opportunity of practice conversations is that they can allow people to first express themselves freely without having to worry about how their conversation partner might react (e.g., be upset or offended). People could then adapt this expression to a more appropriate form for bringing up with someone as described by the earlier quote of P7.

Participants envisioned that CAs could help them to articulate their experiences which would enable them to more effectively communicate with human supporters. Considering her experiences of seeking support from her girlfriend, P23 said:

If it was helping me learn to put names on my feelings, put words on my feelings, then that would like probably have a positive impact cause like the amount of times that she's like *what's wrong*, I'm like *I don't know, I don't know, I can't, I can't put words on it.* [P23]

P13 gave the following account when considering the potential impacts of his younger sister using a CA.

I think the big thing will be if she becomes aware of her emotions and is encouraged to go out of her shell. It would just enable us to be more intimate friends with each other. We're siblings but, we're there to support each other. And you'd think awareness of your own emotions is a valuable skill in developing intimate friendships, deep friendships. And it would help me understand her and what does she actually want, what are her dreams, what does she aspire to. And if I know those things I can help her better. Be there to support her rather than, you know, saying stuff that would drag her down. [P13]

P13 described how he perceived his sister to experience difficulties with her emotions and with finding friends. As captured in the quote above, he aspires to support her with these difficulties however struggles to understand her. Key here is the hope of P13 that his sister's use of a CA would encourage and enable her to develop a more intimate and supportive relationship with him. **Additionally, participants perceived that practice conversations could help people become more comfortable talking about their difficulties and act as a stepping stone to human support.**

I feel like the robot is a good first step, but I wouldn't say it's just the last step...It's a good way just to like get a person to feel like they can actually seek help because they've done it once...I think

that’s a big thing, ’cause then the next steps are easier. They gain a kind of a momentum in a sense so like they’d just be like more open to engage and talk to other people. The robot kind of in a sense gave them the opportunity to like take a step forward. [P15]

Participants perceived that the experience of being supported by a CA could positively impact how people then provide support to other people. The reason for this that people can draw on how the CA supported them, for example how the CA reacted and offered advice, to inform how they in turn support other people. There was also a perceived **opportunity for CAs to coach people on how to support others**.

Maybe like the information they provide could be regarding not only how to help yourself but how to help someone else who’s experiencing the mental health struggles. [P21]

P12 described how she already seeks advice from the Snapchat AI for how to support other people.

I go to the AI [laugh] and I like, like *what would I do with this situation* when you just don’t know how to like console someone or like, you know things like that. [P12]

She shared an experience of her two closest friends being in conflict and both asking her for advice in relation to the conflict. Unsure of how to support them and unable to ask either of them for advice, she sought the support of the Snapchat AI. The Snapchat AI gave P12 advice and examples of what she could say to support her friends; examples that P12 changed to sound like how she talks, and then used.

5. Discussion

5.1. A Social Approach to the Design of CAs for Mental Health Management

CAs are an alluring form of computer interface for supporting people with their mental health largely because of their human-likeness. This allure innately recognises the importance of human support for mental health, yet the use of CAs risks impacting human support in profound and complex ways. With this work we aimed to develop an understanding of how to design effective CAs for mental health management that support people in a manner harmonious with the social ecosystems of mental health management.

While prior work has studied CA support, including in the domain of mental health, it has approached this individualistically. Inspired by a growing body of work stressing the social nature of mental health management [61, 57, 44, 19, 18, 83, 65, 58], this study adopts a socially-oriented approach to the design of CAs for mental health support. This perspective places a stronger emphasis on the social ecosystem that is integral to mental health management. Furthermore, our approach responds to a broader request [64] for CA research to extend its

notion of context and to consider the broader effects of technology beyond those immediate to the primary user. To this end, semi-structured interviews with 24 young adults were conducted to investigate people’s perspectives on the design of CAs for mental health management that considered both personal use and the use of someone close to them. Nearly all of our participants reported experiences relating to mental health difficulties. To best of our knowledge, this is the first systematic study of CAs for mental health support that incorporates a social perspective.

Our analysis reveal a spectrum of critical perspectives and potential benefits of CAs for mental health support. Looking at critical perspectives, we found that users tended to see CA use as abnormal, were concerned that CAs could support potentially harmful coping, and thought CAs threatened the nature and attainment of human support. Yet people also reflected that there were benefits in using CAs within this, in particular that they could help with the support burden, and fill support gaps, whilst helping people to practice ways of attaining human support. In the following sections, we unpack these findings to better understand the intricacies of designing CAs such that they can integrate effectively and work socially harmoniously within support ecosystems. Table 2 enumerates design considerations for the creators of CAs for mental health management.

5.2. *Challenges & Opportunities for Effective and Appropriate CA Support in Mental Health*

5.2.1. *Not Replacing, but Augmenting Human Support*

As is commonly described in AI systems design that seeks to assist humans [e.g., 55, 75, 74], participants raised concerns about CA support *replacing* human support (Themes 1 and 3); and instead they were more receptive to CA support that is *augmenting* human support (Themes 4-6). Our analysis contributes an in-depth understanding of this concern, describing two aspects of it: (1) *social stigma*, whereby interacting with CA in place of people is viewed as an abnormal behaviour and associated with perceptions of failing socially and perceived as shameful (Theme 1); and (2) fear that people may *feel threatened by CAs* that could encroach on personal support (Theme 3).

Firstly, our findings on stigma align with previous research indicating that users of Replika similarly reported experiences of stigma associated with the use of, and particularly relationships with, the CA [70, 48]. To enhance the social acceptance of CAs, it is crucial to address this stigma [33, 48]. To this end, prior work has suggested the implementation of educational programs that raise awareness and understanding of the potential benefits of CAs for mental health support [48]. Additionally, *the design of CAs should avoid reinforcing stigma*. For instance, it has been recommended that digital mental health tools are designed to be less identifiable as such, and perhaps instead framed more around self-care to increase their acceptability [44]. Our study’s feedback suggests a preference for CAs that are designed as engaging and dynamic, rather than sterile or clinical. By *designing CAs to be empowering rather than caring com-*

panions or therapeutic tools, we may mitigate stigma and reshape perceptions of mental health support.

Secondly, concerns about CAs replacing human support were further manifest in participants descriptions of CAs taking away opportunities for people to support those they care about (Theme 3). Such concerns could significantly impact both people’s usage of CAs and their responses to their use by others. There is the possibility that some people may view another person’s use or reliance on a CA for mental health support negatively; perhaps in stigmatised or jealous ways. To sidestep potential adverse reactions, individuals may decide to not utilize CAs or to do so secretly. The necessity for secrecy in using CAs is troubling as it could introduce an additional barrier to seeking other forms of support. *Paradoxically, adverse reactions (and the risk of them) motivated by concerns about potential negative social impacts may have actual negative social impacts.* For example, social ties may be weakened by CA users becoming less open about their mental health and seeking of human support in order to avoid potential disapproval or conflict. Social ties could also be damaged through social connections being upset by the CA use and users being distressed by adverse reactions from the people close to them.

For future work, it is therefore important to more clearly distinguish between support provided by CAs and humans to mitigate against risks of displacing human support roles and associated concerns. Consistent with the perceived strengths of CAs to augment human support, *this suggests focusing CAs on providing forms of support that machines are well suited to provide such as information support and data-management related assistance; and to a lesser extent on support forms that may be provided more appropriately by people (e.g., emotional support).* Currently, there is a trend in CA design for mental health support like Woebot [37] and Wysa [76] to embody anthropomorphic qualities by presenting as caring friends or empathetic companions. However, to more clearly separate their role from human interactions, it can be argued that future CA designs should adopt less human-like portrayals.

Furthermore, echoing views by Bowman et al. [11] and others [e.g., 85, 67], *CAs should be framed less like caring friends or companions and designed more as functional tools.* This might further extend the design space for how dialogue interactions or (LLM-based) text queries could be utilized in other forms for providing effective mental health support outside common design conceptions of CAs – as we have seen in instances described by participants, who were using LLM-based systems to get advice on how to appropriately communicate in socially appropriate and mental-health sensitive manners with others. This also puts into question whether traditional CA designs as conversational, chat-based interface provides indeed the most appropriate modality for assisting humans in mental health self-management; opening this up for re-evaluation.

5.2.2. Identifying Appropriate CA Support Functions

Our analysis shows how human-CA interactions, depending on the specific support need or context, can be perceived in different and potentially conflicting ways. For instance, a person (P) may seek CA support to not burden a close

person (CP), whereas that CP may instead interpret the CA interaction as a missed opportunity for them to support and connect with P. Such *divergent perceptions of CA support often stem from an inherent contrast in the roles of support givers and support receivers*. This is illustrated in the scenario above by how people want to provide support (as outlined in Theme 3), yet they are reluctant to receive it, as they wish to not burden other people (Theme 4). The situated nature of CA support appropriateness makes designing appropriate CA support functions complex as different people can have conflicting views on the appropriateness of the same interaction. In line with Section 5.2.1, the potential for and actuality of users and their social connection having conflicting views may have detrimental social impacts (e.g., cause conflict). *To deepen understanding and evaluate the appropriateness of CA support, it is therefore crucial to study CA support at a social level taking into account the perspectives of both users and their social relations*.

Furthermore, participants' perspectives were shaped by their beliefs about what was good for people's mental health. A major aspect of this, captured by Theme 2, was the view of CA support being too compatible with the experience of mental health difficulties. It was perceived that CA support is too easy and comfortable, and that it will not push people in the therapeutic ways needed to effectively deal with their mental health difficulties. Concern was expressed that CAs would be used as crutches for coping rather than tools for addressing difficulties. Such scepticism challenges conventional views of CAs offering benefits through their provision of an easily accessible, convenient and 'safe space' source of mental health support [e.g., 2, 48, 47, 50, 70]. Although CAs are generally perceived as a means to remove barriers to accessing mental health support [e.g., 30, 38, 31], some of our participants considered CAs more as a distraction from more substantial support systems. These individuals, who had access to various mental health services, suggested that CAs should instead actively encourage users to engage in more helpful, therapeutic practices. This suggests that *it may be more appropriate to design CAs to provide, and be seen as providing, strong encouragement that pushes users in therapeutic ways*. This design direction contrasts with recent design recommendations for mental health CAs, such as the recommendation of Wester et al. [80] to design CAs to avoid confrontation. For such CA design it will be crucial to balance the need to therapeutically push users, which may be challenging and uncomfortable for them, with not disengaging users or putting them at risk [48]. Ultimately, it will be essential to develop a deeper clinical understanding of the actual impacts of different forms of CA support on mental health for future development [68].

5.2.3. Avoiding Over-dependence on CAs & CA Support Creep

Participants expressed concern that users would become increasingly dependent upon CAs for the management of their mental health (Theme 2) and increasingly socially isolated (Theme 3). Prior work has similarly raised concerns about dependency and social isolation [63, 48, 43, 77, 82, 41]. To prevent over-reliance, previous research has suggested designing to encourage human support seeking [63, 48]. *Designing to encourage human support must be ap-*

proached carefully given users have already chosen to interact with a CA and not a person. To mitigate risks of over-reliance, prior work has suggested usage limits [63], however such restrictions should be implemented cautiously (if at all) and be overridable to avoid hindering effective mental health management, or denying users support when urgently needed. Rather than to limit usage, usage monitoring could be used to make users aware of how they are using a CA. This could be performed in conjunction with monitoring human social interaction – a common mental health management practice [56, 52] – to inform user’s about their engagement both with their CA and human support network; thereby fostering a more balanced, self-aware approach to mental health management. In line with mental health management, *by making user’s aware of how they are managing their mental health we can empower them to do it appropriately and effectively.*

Participants perceived opportunity for CAs to allow people to practise for human support providing and receiving interactions (Theme 6). Recent work has similarly reported people using CAs to practise social interactions [47, 68]. Despite this support being intended to promote human support, we perceive it as posing a high risk of CA support creep, whereby the scope of CA support grows beyond that intended. A challenge is that support-receiving practice may itself be enough support to resolve the issue leaving no need for further (human) support. As quoted within Theme 3, P9 felt that once people have “let it out” to a CA there would be little use in letting it out again to a person. While practice conversations could be a stepping stone to human support, people may often find that the first step is enough. Similarly, support providing practice risks becoming a case of human-mediated CA support. To mitigate this form of CA support creep, *CAs could be designed to provide practice support reservedly.* This could involve encouraging users to attempt the human support interaction and attempting to provide the minimal practice needed for the user to feel able to do so. CAs could follow up support-receiving practice to monitor and make the user aware of how their practice is converting into actual human support. For support-providing practice, CAs could be designed to encourage active engagement and discourage support forwarding.

5.2.4. *Actively Involving the Social Ecosystem in CA Support Interactions*

People want to support the people they care about with managing their mental health, prompting the need for tools that assist *collective* mental health management [57]. *Such collective CAs (or tools) should be designed to include social connections and enhance the social support network for individuals.* For instance, CAs for collective mental health management could feature companion interfaces for use by social connections that could enable friends or family to contribute data; communicate their availability for support provision; or engage in discussions and reflections about collected data.

Social accountability is understood to be a major motivator for mental health management [18, 57]. Indeed, part of the allure of CAs for healthcare is that relational design can lead users to feel a sense of accountability to the CA, which can promote adherence. Instead of designing CAs to promote such accountabil-

ity that may be ethically dubious with users feeling pressured to interact and risk fostering over-use and dependency [70, 11], *CA design should encourage a social accountability to other people*. This could mean prompting users to regularly connect and reflect on their mental health management *with their social circles* [49]. Further embracing the concept of collective mental health management, CAs might also be designed to enable social connections to actively participate in monitoring and encouraging the individual’s mental health journey. Of course, where CA design fosters social accountability and brings people’s social networks into the process, this needs to be approached with caution since such involvement may not always be desired; nor at all times appropriate or positive. Furthermore, human relationships evolve and can change over time. Lastly, the design of such functionality must carefully balance the inclusion of social connections with the user’s sense of agency and privacy needs [57].

5.3. Reflections, Limitations, and Future Work

Our interview design effectively engaged participants in a rich, nuanced, and grounded discussion. Participants’ perspectives on CAs for their own and others’ use were often quite different. In part this may be because the consideration of a close person gave participants an opportunity to express views they did not want to express as their own. For example, participants rarely expressed personal interest in the Alex CA but often thought it would appeal to their close person. We suspect the Alex card is more stigma invoking which may have caused participants to feel uncomfortable expressing their own interest in it. The design cards broadened participants’ understanding of ways CAs could be designed and facilitated consideration of these ways. Displaying the cards side-by-side worked well as it encouraged comparison of the designs and consideration of how they could be combined. Although the diversity of the cards facilitated exploration of the design space, they also focused the study on specific areas of the design space and therefore also limit the study.

Our participants were young adults in university education living in Ireland. Notable participant characteristics with regard to transferability include that most are going through, or have recently gone through, a period of considerable change to their human support network having likely moved away (physically and/or relational) from established social connections and having developed new ones. The perspectives of people with more stable human support networks, or with support networks that more generally differ in character to those typical of university students, may differ to those of our participants. Our participants also have relatively good access to a range of mental health support services. Having such support available may impact perspectives particularly regarding how people should be supported with their mental health. Additionally, most participants were highly interested in the study seemingly because of an interest in mental health stemming from personal experiences of it. Our participants’ perspectives, both positive and negative, are therefore perhaps more pronounced than those of the broader population.

Our study explored a mixture of interpersonal relationships, future work could study specific relationships in greater depth (e.g., partnered adults, child-

Social Oriented Design Considerations	
Not Replacing, but Augmenting Human Support	
1	Designing CAs to be empowering rather than caring companions or therapeutic tools may mitigate stigma and reshape perceptions of mental health support.
2	Adverse reactions (and the risk of them) motivated by concerns about potential negative social impacts may have actual negative social impacts.
3	Focus CAs on providing forms of support that machines are well suited to provide and to a lesser extent on support forms that may be provided more appropriately by people.
4	CAs should be framed less like caring friends or companions and designed more as functional tools.
Identifying Appropriate CA Support Functions	
5	Divergent perceptions of CA support often stem from an inherent contrast in the roles of support givers and support receivers.
6	To deepen understanding and evaluate the appropriateness of CA support, it is crucial to study CA support at a social level taking into account the perspectives of both users and their social relations.
7	It may be more appropriate to design CAs to provide, and be seen as providing, strong encouragement that pushes users in therapeutic ways.
Avoiding Over-dependence on CAs & CA Support Creep	
8	Designing to encourage human support must be approached carefully given users have already chosen to interact with a CA and not a person.
9	By making user's aware of how they are managing their mental health we can empower them to do it appropriately and effectively.
10	CAs could be designed to provide practice support reservedly.
Actively Involving the Social Ecosystem in CA Support Interactions	
11	Collective CAs (or tools) should be designed to include social connections and enhance the social support network for individuals.
12	CA design should encourage a social accountability to other people.

Table 2: Social Oriented Design Considerations for CAs for Mental Health Management

parent, relationships where a single person is experiencing mental health difficulties and relationships where multiple people are). Our analysis suggests that people's perspectives on receiving and providing support differ, future work could directly compare these perspectives. Participants engaged well with our interview design and therefore we encourage the use of similar designs. Additionally, future work may consider designs that involve multiple participants who are in relationships with each other [see 62]; for example, interviews involving two participants in a close relationship. Future work may consider participant's social support networks (e.g., using the measures and techniques of social sup-

port research [34]) in order to contextualise and understand the determinants of people’s perspectives.

6. Conclusion

Toward designing CAs for mental health support that are harmonious with the social dynamics of mental health management, this work investigated people’s perspectives on CA design with a focus on social considerations. We contribute *knowledge of how people do and do not want CAs to support people with their mental health*. From this knowledge we identify and contribute insight about the challenges and opportunities of designing CAs for mental health management. Our work highlights the importance of designing CAs for *not replacing, but augmenting human support*. We outline the complexities of *identifying appropriate support functions for CAs* and the difficulties of *avoiding over-dependence on CAs and CA support creep*. Additionally we discuss the benefits and pitfalls of *actively involving social ecosystems in CA support*. This work with its socially-oriented approach is an important step forward in developing CAs that support people effectively and in a socially sensitive manner with their mental health.

Acknowledgements

This research was conducted with the financial support of the Science Foundation Ireland Centre for Research Training in Digitally-Enhanced Reality (d-real, grant 18/CRT/6224) and Microsoft Research through its PhD Scholarship Programme. This research was additionally conducted in part with the financial support of the Science Foundation Ireland Adapt Research Centre (grant 13/RC/2106_P2).

References

- [1] Amwell, 2024. Silvercloud health. <https://www.silvercloudhealth.com/>. Accessed: September 12th 2024.
- [2] Bae Brandtzæg, P., Skjuve, M., Kristoffer Dysthe, K.K., Følstad, A., 2021. When the social becomes non-human: young people’s perception of social support in chatbots, in: Proceedings of the 2021 CHI conference on human factors in computing systems, pp. 1–13.
- [3] Balaskas, A., Schueller, S.M., Cox, A.L., Doherty, G., 2021. Ecological momentary interventions for mental health: A scoping review. PloS one 16, e0248152.
- [4] Beatty, C., Malik, T., Meheli, S., Sinha, C., 2022. Evaluating the therapeutic alliance with a free-text cbt conversational agent (wysa): a mixed-methods study. Frontiers in Digital Health 4, 847991.

- [5] Bickmore, T., Cassell, J., 2001. Relational agents: a model and implementation of building user trust, in: *Proceedings of the SIGCHI conference on Human factors in computing systems*, pp. 396–403.
- [6] Bickmore, T., Gruber, A., Picard, R., 2005. Establishing the computer-patient working alliance in automated health behavior change interventions. *Patient education and counseling* 59, 21–30.
- [7] Bickmore, T., O’Leary, T., 2023. Conversational agents on smartphones and the web, in: *Digital Therapeutics for Mental Health and Addiction*. Elsevier, pp. 99–112.
- [8] Bickmore, T.W., Kimani, E., Trinh, H., Pusateri, A., Paasche-Orlow, M.K., Magnani, J.W., 2018. Managing chronic conditions with a smartphone-based conversational virtual agent, in: *Proceedings of the 18th International Conference on Intelligent Virtual Agents*, pp. 119–124.
- [9] Bickmore, T.W., Puskar, K., Schlenk, E.A., Pfeifer, L.M., Sereika, S.M., 2010. Maintaining reality: Relational agents for antipsychotic medication adherence. *Interacting with Computers* 22, 276–288.
- [10] Bowman, R., 2025. Exploring the Design of Conversational Agents for Mental Health Self-Monitoring. Ph.D. thesis. Trinity College Dublin. URL: <https://www.tara.tcd.ie/handle/2262/111228>.
- [11] Bowman, R., Cooney, O., Newbold, J.W., Thieme, A., Clark, L., Doherty, G., Cowan, B., 2024. Exploring how politeness impacts the user experience of chatbots for mental health support. *International Journal of Human-Computer Studies* , 103181.
- [12] Bowman, R., Cooney, O., Thieme, A., Cowan, B.R., Doherty, G., 2022. Pervasive therapy: Designing conversation-based interfaces for ecological momentary intervention. *IEEE Pervasive Computing* 21, 55–63.
- [13] Bowman, R., Nadal, C., Morrissey, K., Thieme, A., Doherty, G., 2023. Using thematic analysis in healthcare hci at chi: A scoping review, in: *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, pp. 1–18.
- [14] Brandtzaeg, P.B., Følstad, A., 2018. Chatbots: changing user needs and motivations. *interactions* 25, 38–43.
- [15] Braun, V., Clarke, V., 2013. *Successful qualitative research: A practical guide for beginners*. Sage.
- [16] Braun, V., Clarke, V., 2021. *Thematic Analysis: A Practical Guide*. Sage.
- [17] Browne, G., 2022. The problem with mental health bots. <https://www.wired.com/story/mental-health-chatbots/>. Accessed: 27/06/2024.

- [18] Burgess, E.R., Reddy, M.C., Mohr, D.C., 2022. " i just can't help but smile sometimes": Collaborative self-management of depression. *Proceedings of the ACM on Human-Computer Interaction* 6, 1–32.
- [19] Burgess, E.R., Ringland, K.E., Nicholas, J., Knapp, A.A., Eschler, J., Mohr, D.C., Reddy, M.C., 2019. " i think people are powerful" the sociality of individuals managing depression. *Proceedings of the ACM on Human-computer Interaction* 3, 1–29.
- [20] Burgess, E.R., Zhang, R., Ernala, S.K., Feuston, J.L., De Choudhury, M., Czerwinski, M., Aguilera, A., Schueller, S.M., Reddy, M.C., 2020. Technology ecosystems: rethinking resources for mental health. *Interactions* 28, 66–71.
- [21] Colby, K.M., Watt, J.B., Gilbert, J.P., 1966. A computer method of psychotherapy: Preliminary communication. *The Journal of Nervous and Mental Disease* 142, 148–152.
- [22] Cooney, O., Doherty, K., Barry, M., Coyle, D., Doherty, G., Cowan, B.R., 2024. Using speech agents for mood logging within blended mental healthcare: Mental healthcare practitioners' perspectives, in: *Proceedings of the 6th ACM Conference on Conversational User Interfaces*, pp. 1–11.
- [23] Coyle, D., Doherty, G., Matthews, M., Sharry, J., 2007. Computers in talk-based mental health interventions. *Interacting with computers* 19, 545–562.
- [24] Cutrona, C.E., Suhr, J.A., 1992. Controllability of stressful events and satisfaction with spouse support behaviors. *Communication research* 19, 154–174.
- [25] Darcy, A., Daniels, J., Salinger, D., Wicks, P., Robinson, A., 2021. Evidence of human-level bonds established with a digital conversational agent: cross-sectional, retrospective observational study. *JMIR Formative Research* 5, e27868.
- [26] Doherty, G., Coyle, D., Matthews, M., 2010. Design and evaluation guidelines for mental health technologies. *Interacting with computers* 22, 243–252.
- [27] Doherty, G., Coyle, D., Sharry, J., 2012. Engagement with on-line mental health interventions: An exploratory clinical study of a treatment for depression, in: *Proc. CHI 12*, ACM, New York. p. 1421–1430. URL: <https://doi.org/10.1145/2207676.2208602>, doi:10.1145/2207676.2208602.
- [28] Doherty, K., Balaskas, A., Doherty, G., 2020. The design of ecological momentary assessment technologies. *Interacting with Computers* 32, 257–278.

- [29] Fiske, A., Henningsen, P., Buyx, A., 2019. Your robot therapist will see you now: ethical implications of embodied artificial intelligence in psychiatry, psychology, and psychotherapy. *Journal of medical Internet research* 21, e13216.
- [30] 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, e7785.
- [31] Fulmer, R., Joerin, A., Gentile, B., Lakerink, L., Rauws, M., et al., 2018. Using psychological artificial intelligence (tess) to relieve symptoms of depression and anxiety: randomized controlled trial. *JMIR mental health* 5, e9782.
- [32] Headspace, 2024. Headspace meditation app. <https://www.headspace.com/headspace-meditation-app>. Accessed: September 12th 2024.
- [33] Hoffman, B.D., Oppert, M.L., Owen, M., 2024. Understanding young adults' attitudes towards using ai chatbots for psychotherapy: The role of self-stigma. *Computers in Human Behavior: Artificial Humans* 2, 100086.
- [34] House, J., Kahn, R., McLeod, J., Williams, D., 1985. Measures and Concepts of Social Support. Academic Press, NY. pp. 83–108.
- [35] Inc, L., 2024a. Replika: My ai friend. <https://play.google.com/store/apps/details?id=ai.replika.app>. Accessed: September 10th 2024.
- [36] Inc, S., 2024b. Snapchat. <https://play.google.com/store/apps/details?id=com.snapchat.android>. Accessed: September 10th 2024.
- [37] Inc, W.H., 2024c. Woebot: The mental health ally. <https://play.google.com/store/apps/details?id=com.woebot>. Accessed: September 10th 2024.
- [38] Inkster, B., Sarda, S., Subramanian, V., et al., 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, e12106.
- [39] Kim, J., Kim, Y., Kim, B., Yun, S., Kim, M., Lee, J., 2018. Can a machine tend to teenagers' emotional needs? a study with conversational agents, in: *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems*, Association for Computing Machinery, New York, NY, USA. p. 1–6. URL: <https://doi.org/10.1145/3170427.3188548>, doi:10.1145/3170427.3188548.

- [40] Kim, T., Bae, S., Kim, H.A., Lee, S.w., Hong, H., Yang, C., Kim, Y.H., 2024. Mindfuldiary: Harnessing large language model to support psychiatric patients’ journaling, in: Proceedings of the CHI Conference on Human Factors in Computing Systems, pp. 1–20.
- [41] Koulouri, T., Macredie, R.D., Olakitan, D., 2022. Chatbots to support young adults’ mental health: an exploratory study of acceptability. *ACM Transactions on Interactive Intelligent Systems (TiiS)* 12, 1–39.
- [42] Kruzan, K.P., Ng, A., Stiles-Shields, C., Lattie, E.G., Mohr, D.C., Reddy, M., 2023. The perceived utility of smartphone and wearable sensor data in digital self-tracking technologies for mental health, in: Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems, Association for Computing Machinery, New York, NY, USA. URL: <https://doi.org/10.1145/3544548.3581209>, doi:10.1145/3544548.3581209.
- [43] Laestadius, L., Bishop, A., Gonzalez, M., Illeńčík, D., Campos-Castillo, C., 2022. Too human and not human enough: A grounded theory analysis of mental health harms from emotional dependence on the social chatbot replika. *New Media & Society* , 14614448221142007.
- [44] Lattie, E.G., Kornfield, R., Ringland, K.E., Zhang, R., Winkquist, N., Reddy, M., 2020. Designing mental health technologies that support the social ecosystem of college students, in: Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems, pp. 1–15.
- [45] Lee, Y.C., Yamashita, N., Huang, Y., Fu, W., 2020. “i hear you, i feel you”: encouraging deep self-disclosure through a chatbot, in: Proceedings of the 2020 CHI conference on human factors in computing systems, pp. 1–12.
- [46] 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.
- [47] Ma, Z., Mei, Y., Long, Y., Su, Z., Gajos, K.Z., 2024. Evaluating the experience of lgbtq+ people using large language model based chatbots for mental health support, in: Proceedings of the CHI Conference on Human Factors in Computing Systems, pp. 1–15.
- [48] Ma, Z., Mei, Y., Su, Z., 2023. Understanding the benefits and challenges of using large language model-based conversational agents for mental well-being support, in: AMIA Annual Symposium Proceedings, American Medical Informatics Association. p. 1105.
- [49] Maharjan, R., Doherty, K., Rohani, D.A., Bækgaard, P., Bardram, J.E., 2022. Experiences of a speech-enabled conversational agent for the self-report of well-being among people living with affective disorders: an in-the-

- wild study. *ACM Transactions on Interactive Intelligent Systems (TiiS)* 12, 1–29.
- [50] Malik, T., Ambrose, A.J., Sinha, C., 2022. Evaluating user feedback for an artificial intelligence-enabled, cognitive behavioral therapy-based mental health app (wysa): Qualitative thematic analysis. *JMIR Human Factors* 9, e35668.
 - [51] Matthews, M., Doherty, G., Sharry, J., Fitzpatrick, C., 2008. Mobile phone mood charting for adolescents. *British Journal of Guidance & Counselling* 36, 113–129.
 - [52] Matthews, M., Murnane, E., Snyder, J., 2017. Quantifying the changeable self: The role of self-tracking in coming to terms with and managing bipolar disorder. *Human-Computer Interaction* 32, 413–446.
 - [53] Miller, E., Polson, D., 2019. Apps, avatars, and robots: The future of mental healthcare. *Issues in mental health nursing* 40, 208–214.
 - [54] 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, e10148.
 - [55] Morrison, C., Cutrell, E., Grayson, M., Thieme, A., Taylor, A., Roumen, G., Longden, C., Tschatschek, S., Faia Marques, R., Sellen, A., 2021. Social sensemaking with ai: Designing an open-ended ai experience with a blind child, in: *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, pp. 1–14.
 - [56] Murnane, E.L., Cosley, D., Chang, P., Guha, S., Frank, E., Gay, G., Matthews, M., 2016. Self-monitoring practices, attitudes, and needs of individuals with bipolar disorder: implications for the design of technologies to manage mental health. *Journal of the American Medical Informatics Association* 23, 477–484. URL: <https://doi.org/10.1093/jamia/ocv165>, doi:10.1093/jamia/ocv165, arXiv:<https://academic.oup.com/jamia/article-pdf/23/3/477/34147647/ocv165.pdf>.
 - [57] Murnane, E.L., Walker, T.G., Tench, B., Volda, S., Snyder, J., 2018. Personal informatics in interpersonal contexts: Towards the design of technology that supports the social ecologies of long-term mental health management. *Proc. ACM Hum.-Comput. Interact.* 2. URL: <https://doi.org/10.1145/3274396>, doi:10.1145/3274396.
 - [58] O’Leary, T.K., Parmar, D., Olafsson, S., Paasche-Orlow, M., Bickmore, T., Parker, A.G., 2022. Community dynamics in technospiritual interventions: lessons learned from a church-based mhealth pilot, in: *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*, pp. 1–23.

- [59] OpenAI, 2024. Chatgpt. <https://openai.com/chatgpt/>. Accessed: September 12th 2024.
- [60] Park, S., Thieme, A., Han, J., Lee, S., Rhee, W., Suh, B., 2021. “i wrote as if i were telling a story to someone i knew.”: Designing chatbot interactions for expressive writing in mental health, in: Designing Interactive Systems Conference 2021, pp. 926–941.
- [61] Park, S.Y., 2018. Social support mosaic: Understanding mental health management practice on college campus, in: Proceedings of the 2018 Designing Interactive Systems Conference, pp. 121–133.
- [62] Pina, L.R., Sien, S.W., Ward, T., Yip, J.C., Munson, S.A., Fogarty, J., Kientz, J.A., 2017. From personal informatics to family informatics: Understanding family practices around health monitoring, in: Proceedings of the 2017 acm conference on computer supported cooperative work and social computing, pp. 2300–2315.
- [63] Prakash, A.V., Das, S., 2020. Intelligent conversational agents in mental healthcare services: a thematic analysis of user perceptions. *Pacific Asia Journal of the Association for Information Systems* 12, 1.
- [64] Rapp, A., Curti, L., Boldi, A., 2021. The human side of human-chatbot interaction: A systematic literature review of ten years of research on text-based chatbots. *International Journal of Human-Computer Studies* 151, 102630.
- [65] Sackitey, D., O’Leary, T.K., Paasche-Orlow, M., Bickmore, T., Parker, A.G., 2023. ” everyone is covered”: Exploring the role of online interactions in facilitating connection and social support in black churches, in: Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems, pp. 1–17.
- [66] Sanches, P., Janson, A., Karpashevich, P., Nadal, C., Qu, C., Daudén Roquet, C., Umair, M., Windlin, C., Doherty, G., Höök, K., et al., 2019. Hci and affective health: Taking stock of a decade of studies and charting future research directions, in: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems, pp. 1–17.
- [67] Sharma, A., Rushton, K., Lin, I.W., Nguyen, T., Althoff, T., 2024. Facilitating self-guided mental health interventions through human-language model interaction: A case study of cognitive restructuring, in: Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems, Association for Computing Machinery, New York, NY, USA. URL: <https://doi.org/10.1145/3613904.3642761>, doi:10.1145/3613904.3642761.
- [68] Siddals, S., Torous, J., Coxon, A., 2024. “it happened to be the perfect thing”: experiences of generative ai chatbots for mental health. *npj Mental Health Research* 3, 48.

- [69] Skjuve, M., Følstad, A., Brandtzæg, P.B., 2023. A longitudinal study of self-disclosure in human–chatbot relationships. *Interacting with Computers* 35, 24–39.
- [70] Skjuve, M., Følstad, A., Fostervold, K.I., Brandtzaeg, P.B., 2021. My chatbot companion-a study of human-chatbot relationships. *International Journal of Human-Computer Studies* 149, 102601.
- [71] Skjuve, M., Følstad, A., Fostervold, K.I., Brandtzaeg, P.B., 2022. A longitudinal study of human–chatbot relationships. *International Journal of Human-Computer Studies* 168, 102903.
- [72] Sweeney, C., Potts, C., Ennis, E., Bond, R., Mulvenna, M.D., O’neill, S., Malcolm, M., Kuosmanen, L., Kostenius, C., Vakaloudis, A., et al., 2021. Can chatbots help support a person’s mental health? perceptions and views from mental healthcare professionals and experts. *ACM Transactions on Computing for Healthcare* 2, 1–15.
- [73] Ta, V., Griffith, C., Boatfield, C., Wang, X., Civitello, M., Bader, H., DeCero, E., Loggarakis, A., et al., 2020. User experiences of social support from companion chatbots in everyday contexts: thematic analysis. *Journal of medical Internet research* 22, e16235.
- [74] Thieme, A., Cutrell, E., Morrison, C., Taylor, A., Sellen, A., 2020. Interpretability as a dynamic of human-ai interaction. *Interactions* 27, 40–45. URL: <https://doi.org/10.1145/3411286>, doi:10.1145/3411286.
- [75] Thieme, A., Hanratty, M., Lyons, M., Palacios, J., Marques, R.F., Morrison, C., Doherty, G., 2023. Designing human-centered ai for mental health: Developing clinically relevant applications for online cbt treatment. *ACM Transactions on Computer-Human Interaction* 30, 1–50.
- [76] Touchkin, 2024. Wysa: Anxiety, therapy chatbot. <https://play.google.com/store/apps/details?id=bot.touchkin>. Accessed: September 10th 2024.
- [77] 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. *The Canadian Journal of Psychiatry* 64, 456–464.
- [78] Weizenbaum, J., 1966. Eliza—a computer program for the study of natural language communication between man and machine. *Communications of the ACM* 9, 36–45. URL: <https://doi.org/10.1145/365153.365168>, doi:10.1145/365153.365168.
- [79] Weizenbaum, J., 1976. Computer power and human reason: From judgment to calculation. WH Freeman & Co.

- [80] Wester, J., Pohl, H., Hosio, S., van Berkel, N., 2024. “this chatbot would never...”: Perceived moral agency of mental health chatbots. *Proceedings of the ACM on Human-Computer Interaction* 8, 1–28.
- [81] White, G., 2018. Child advice chatbots fail to spot sexual abuse. URL: <https://www.bbc.com/news/technology-46507900>. accessed: February 8th 2023.
- [82] Xie, T., Pentina, I., 2022. Attachment theory as a framework to understand relationships with social chatbots: a case study of replika, in: *Proceedings of the 55th Hawaii international conference on system sciences*.
- [83] Yamashita, N., Kuzuoka, H., Hirata, K., Kudo, T., 2013. Understanding the conflicting demands of family caregivers caring for depressed family members, in: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 2637–2646.
- [84] Yamashita, N., Kuzuoka, H., Hirata, K., Kudo, T., Aramaki, E., Hattori, K., 2017. Changing moods: How manual tracking by family caregivers improves caring and family communication, in: *Proceedings of the 2017 chi conference on human factors in computing systems*, pp. 158–169.
- [85] Yildirim, N., Richardson, H., Wetscherek, M.T., Bajwa, J., Jacob, J., Pinnock, M.A., Harris, S., Coelho De Castro, D., Bannur, S., Hyland, S., Ghosh, P., Ranjit, M., Bouzid, K., Schwaighofer, A., Pérez-García, F., Sharma, H., Oktay, O., Lungren, M., Alvarez-Valle, J., Nori, A., Thieme, A., 2024. Multimodal healthcare ai: Identifying and designing clinically relevant vision-language applications for radiology, in: *Proceedings of the CHI Conference on Human Factors in Computing Systems*, Association for Computing Machinery, New York, NY, USA. URL: <https://doi.org/10.1145/3613904.3642013>, doi:10.1145/3613904.3642013.