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

RAJU MAHARJAN, KEVIN DOHERTY, and DARIUS ADAM ROHANI, Department of Health Technology, Technical University of Denmark, Denmark

PER BÆKGAARD, Department of Applied Mathematics and Computer Science, Technical University of Denmark, Denmark

JAKOB E. BARDRAM, Department of Health Technology, Technical University of Denmark, Denmark

The growing commercial success of smart speaker devices following recent advancements in speech recognition technology has surfaced new opportunities for collecting self-reported health and well-being data. Speech-enabled conversational agents (CAs) in particular, deployed in home environments using just such systems, may offer increasingly intuitive and engaging means of self-report. To date, however, few real-world studies have examined users' experiences of engaging in the self-report of mental health using such devices or the challenges of deploying these systems in the home context. With these aims in mind, this article recounts findings from a 4-week "in-the-wild" study during which 20 individuals with depression or bipolar disorder used a speech-enabled CA named "Sofia" to maintain a daily diary log, responding also to the World Health Organization–Five Well-Being Index WHO-5 scale every 2 weeks. Thematic analysis of post-study interviews highlights actions taken by participants to overcome CAs' limitations, diverse personifications of a speech-enabled agent, and unique forms of valuing of this system among users' personal and social circles. These findings serve as initial evidence for the potential of CAs to support the self-report of mental health and well-being, while highlighting the need to address outstanding technical limitations in addition to design challenges of conversational pattern matching, filling unmet interpersonal gaps, and the use of self-report CAs in the at-home social context. Based on these insights, we discuss implications for the future design of CAs to support the self-report of mental health and well-being.

CCS Concepts: • **Human-centered computing** → **User studies**; **Empirical studies in HCI**;

Additional Key Words and Phrases: Conversational user interface, conversational agent, voice user interface, virtual assistant, virtual health assistant, self-reports, mental health, who-5

1 INTRODUCTION

Much of what we know about mental health and well-being is gathered through self-reports drawn from patient diaries and validated health questionnaires [1, 24, 25]. These self-report methods, traditionally administered using pen and paper, have a long history and are considered effective means of monitoring and assessing mental illnesses including depression and bipolar disorder. In recent years many have turned toward technology to facilitate such data collection methods; most often by means of the direct implementation of questionnaires as web and mobile apps.

At the same time, **human-computer interaction (HCI)** researchers working on the design and use of technology to gather self-reports in daily life, often via **ecological momentary assessment (EMA)** methods, have drawn attention to the complexity of the self-report of mental health and well-being as process in itself [2, 26, 73]. While **graphical user interface– (GUI)** based tools can serve as efficient means of collecting textual and visual data of many forms, these are but one medium for doing so, can limit users' capacity for self-expression, pose challenges for the assessment of the validity and reliability of data gathered, and can place a significant burden on users [24, 32, 85].

Speech-enabled CAs, in comparison, may offer the opportunity to obtain richer insight into individuals' experience of mental health and illness—whether for sharing with health professionals [65], to support individuals' own insight [50], or perhaps even to serve as part of health and well-being interventions [28, 86]. Studies have shown that speech, as a primary mode of communication, can engage users in more natural and human-to-humanlike conversations, suggesting the potential of these systems to improve user engagement and self-report response quality while fostering more honest and insightful forms of self-disclosure [23, 50].

Smart speaker devices including Google's Home and Amazon's Alexa have seen a surge in popularity in recent years following advancements in speech recognition technology—rendering these technologies increasingly feasible means of interaction and even leading to increased application of these systems in healthcare [38, 40, 41, 44, 66, 84]. According to one recent survey, 52.0% of 1,004 U.S. adults possess an interest in the use of Conversational Agent (CA)s, while 7.5% have already made use of a CA for a healthcare-related task or inquiry, from inquiring about symptoms of illness (73.0%), to searching for information concerning medication use (45.9%), or seeking care and treatment options (37.7%) [11].

Despite such increased interest, expanded accessibility and even growing adoption of CAs in healthcare, a number of significant technical limitations and challenges to the use of these systems for gathering self-reported data remain. Limitations to date identified by researchers include misinterpretation and failed recognition of user utterances resulting in confusion and further errors [58, 68, 77], CAs' present inability to engage in dynamic conversations [16], the need to specify in advance what users can or cannot say during interaction [20, 30, 65], and many CAs' monotonous, robotic and unnatural voices that have been found to have adverse implications for users' engagement [13, 17, 27, 54].

In response, HCI researchers have undertaken explorations of issues of privacy [45], usability and user satisfaction [16], accessibility [67], and attitudes toward [48] CAs. Several researchers have proposed initial guidelines for the design [19, 57, 77, 88] and evaluation of CAs [39], while

others have begun to explore questions of user experience, trust, feedback, self-reflection and learning during the collection of self-reports of mental and physical health and well-being [42, 69, 86], as well as social engagement [64, 72, 74].

Research in the context of healthcare has investigated CAs' effectiveness in supporting diagnostic performance, symptom disclosure, health intervention, adherence to self-management practices and technical issues in CA dialog management [44, Table 3]. And, although these technologies are currently insufficiently sophisticated to fully support dynamic human-to-humanlike conversational interactions, other healthcare researchers have expressed interest in the anthropomorphic characteristics of CAs as potential means of establishing and maintaining a therapeutic alliance that could facilitate greater disclosure and deeper insight into users' health and well-being [14, 23, 38, 50, 51].

While initial research therefore suggests the potential of CAs to support practices of self-report and self-reflection, this is a highly complex and emerging design space with many outstanding ethical, technical, and medical challenges [84]. We know little about what people living with mental illness make of the idea of employing CAs for the purposes of logging, discussing, reflecting on, reporting or monitoring their mental health, and even less about their lived experiences of engaging with such systems in the real-world, nor how to design for sustainable CA self-report experiences. There exists, therefore, a need to understand, in practice and with representative users, the real-world experience of interacting with a CA to support the self-report of mental health, the challenges of deploying these systems in the homes of users experiencing **affective disorders (ADs)**, and the choices available to designers in overcoming their limitations.

To understand this demographic group's real-world CA self-report experiences and broader perceptions of technology for the self-report of mental health and well-being, we recruited 20 individuals who self-identified as diagnosed with an AD, either depression or bipolar disorder, and conducted a study "in the wild" for the purpose of addressing the following research questions:

RQ1: How do people living with Affective Disorder (AD) experience the use of a speech-enabled CA for the self-report of mental health and well-being in their everyday lives?

RQ2: What are the challenges of employing CAs for the self-report of mental health and well-being in home contexts?

RQ3: Which choices and strategies might designers employ to overcome the current limitations of CAs in support of sustainable practices of self-report?

A 4-week "in the wild" study was designed to enable people living with ADs to experience the verbal self-report of mental health and well-being via CA. During this study, participants kept a daily open-ended conversational diary log of their mental health and well-being, and responded to the WHO-5 questionnaire fortnightly using a speech-enabled CA named *Sofia* deployed via Google Nest device.

Participants' positive responses to the user experience questionnaire (UEQ) and high levels of engagement with the CA despite technical limitations serve as initial evidence for users' willingness to accept the use of these systems for the self-report of mental health and well-being in the home context. Results from thematic analysis of post-study interviews reveal users' experiences as strongly shaped by strategies adopted to overcome CAs' technical limitations, diverse personified perceptions of the agent, the socially contingent nature of self-reporting practice, users' reflections on privacy and security concerns, as well as the CA features considered important for engaging and sustainable self-report experiences.

We discuss implications for the design of CAs for mental health and well-being with respect to the challenges surfaced by this work including challenges of conversational pattern

matching, filling unmet interpersonal gaps, and the use of self-report CAs in the at-home social context.

This work contributes (i) an understanding of factors impacting the self-report experiences and behaviors of people with AD and (ii) implications for the future design of speech-enabled CAs to support the self-report of mental health and well-being.

2 RELATED WORK

This study builds upon and contributes to prior research concerning the study and design of CAs for mental health and well-being, as well as speech-enabled self-report mechanisms.

2.1 CAs for Mental Health and Well-being

The history of CAs for mental healthcare dates back to the development of ELIZA in 1966. This first of many text-based natural language processing programs to come provided pre-determined answers to user queries in mimicry of a Rogerian psychotherapist [90]. Technology has advanced significantly in the years since with the emergence of new forms of interactive voice response, automatic speech recognition, and embodied systems, chatbots, and speech-enabled CAs.

Although CAs remain in many ways nascent technologies, many HCI and healthcare researchers have already begun to explore CAs' potential to support mental health and well-being in a variety of contexts and in various forms, from diagnosis to symptom monitoring, treatment, and intervention [31, 44, 49, 84]. The multi-modal embodied system "Virtual Mindfulness Coach," for example, was developed and evaluated for its effectiveness in training and coaching mindfulness meditation [35]. Student users of this system were found to achieve a higher sense of self-efficacy in regard to the establishment of regular mindfulness practice through coach-based training than via self-administered training using written and audio materials. A remote "Wizard-of-Oz" study of another multi-modal agent designed to provide longitudinal social support to isolated older adults similarly reported high levels of user acceptance and satisfaction with the system [6]. Findings suggested that the system's capacity to address loneliness relied on agents' ability to interact proactively with users rather than waiting passively for users to initiate interaction.

Other studies have demonstrated that human–CA interactions tend to follow prevailing social norms despite users' awareness that their interaction partner is a machine [59, 70]. It has been suggested that this more humanlike conversational mode of interaction may enable users to form a relationship with CAs [7, 87] and that CAs could in turn potentially serve as means to collect more accurate, honest, and insightful information from users [23, 50]. In the broader context of the design of technology to support mental health, CAs' potential capacity to establish relationships with users could eventually facilitate even health-related behavior change among users with ADs [7, 78].

2.1.1 Conversational Agents and Affective Disorders. One in four people worldwide will be affected by an AD in their lifetime [61]. This class of mental health conditions characterized by sudden changes in mood includes depression, bipolar disorder, anxiety, and other mood disorders [33]. Therapies entailing practices of self-report and self-monitoring such as **cognitive behavioral therapy (CBT)** and behavioral activation are known to be effective in treating ADs [21, 79, 89]. The efficacious delivery of treatment, however, is often impacted by challenges including long waiting lists caused by a shortage of mental health professionals, stigma in relation to mental health, and a lack of access related to limited availability as well as inefficient and burdensome means of communication [36].

Initial research suggests that by aiding in the assessment and treatment of mental illnesses including ADs, CAs may prove able to play a role in overcoming these challenges of access and interaction [34, 84]. Fitzpatrick et al., for example, developed a fully automated chatbot named "Woebot" and studied its feasibility, acceptability, and efficacy in delivering a self-help CBT program for college students with symptoms of anxiety and depression [29]. Findings of this study showed that delivery of Cognitive Behavioral Therapy (CBT) through the chatbot significantly reduced participants' symptoms of depression and anxiety. The authors furthermore suggest that process factors were more influential for participants' acceptance of the system than content factors, mirroring traditional therapy.

Denecke et al. conducted a usability study of "SERMO," a chatbot implementing CBT methods for emotional regulation, with four clinicians, nine patients, and eight non-patients [22]. Participants rated the app positively in terms of efficiency, perspicuity, and attractiveness and its hedonic qualities (stimulation and novelty) neutrally. Inkster et al. evaluated the effectiveness of an **artificial intelligence– (AI)** enabled text-based CA named "Wysa" for the elicitation of self-reported symptoms of depression using the patient health questionnaire (PHQ-9) questionnaire as well as improvements to users' mood [36]. Participants reported that the CA proved helpful and a source of encouragement in relation to their mood. Bickmore et al. investigated the usability and effectiveness of a text-based virtual nurse named Elizabeth with 139 patients with depressive symptoms and found that the CA could serve as an effective medium for automated screening and delivery of treatment [8].

While the results of these preliminary studies are promising, therapeutic applications of CAs remain at a very early stage and there are many technical, ethical, and medical challenges that must be overcome before they can be offered as a viable alternative for those in need of support [84]. First, we must in particular develop an understanding of the perceptions and experiences of people living with ADs in relation to the disclosure of their mental health and well-being by means of CA.

2.2 Understanding Speech-enabled Self-report Mechanisms

Much research in recent years has focused on the capacity of mobile devices to support the self-report of mental health and well-being. GUI-based tools have in many cases enabled the efficient collection of textual and visual data of many forms, and yet there persist numerous motivations for considering alternative forms of self-report. In particular, these systems are but one medium of self-expression, often constrain users' responses, pose challenges for the assessment of the validity and reliability of data gathered, and can place a significant burden on users [24, 32, 85].

This has led a number of researchers to turn toward the development of speech-enabled methods of self-report. One such research thread concerns work in the healthcare domain focused on speech-enabled embodied CAs as a means of eliciting self-reports of psychological symptoms from patients. Lucas et al., for example, examined the impact of a virtual human interviewer on the disclosure of **post-traumatic stress disorder (PTSD)** symptoms among active military members including the factors influencing self-reporting practices [51]. Results of this study suggested that virtual human interviewers could foster increased patient disclosure of mental health symptoms due in part to the condition of anonymity, reduced stigma, and the CA's capacity to build rapport with users.

DeVault et al. likewise designed a virtual human interviewer to assess mental health conditions including depression and anxiety via automated analysis of verbal and non-verbal behaviors [23]. Findings from this user study suggested that the system was able to engage users in an open-ended conversation as long as 15–25 minutes in duration and that participants were comfortable

sharing intimate information. Finally, Philip et al. contrasted the diagnostic performance of an embodied CA and a psychiatrist in relation to major depressive disorder among 179 outpatient participants [63]. Results from this proof of concept study suggest that CAs might also serve as future means to conduct standardized clinical interviews.

Many of these studies point to the self-report of mental health as a feasible use-case for the deployment of speech-enabled CAs and yet are conducted in lab settings and particular to the design of embodied CAs. Recent advancements in speech recognition technology, NLP, and the accessibility of smartspeaker devices such as Google's Home and Amazon's Echo have introduced the possibility of enabling and studying more "natural" conversational interactions in home settings, as others have begun to explore.

Kocielnik et al., for example, designed an Amazon Alexa Skill named "Robota" to support work activity journaling; asking users to provide 10 open-ended daily reflections [42]. The authors examined how speech-enabled interaction affected workers' reflections and self-learning in comparison to a chat-bot employing the same questionnaire. Results from a 3-week controlled field study showed that speech interaction enabled users to step back and reflect on their work as well as provided opportunities for workplace-related behavior-change, despite many technical limitations. In related work, Quiroz et al. developed an Alexa Skill enabling users to better express their emotions, complete self-assessments for depression and anxiety, and receive suggestions for improving their current state of mind [69]. While the results from this pilot study showed that participants were willing to engage with and trusted the agent when sharing personal information such as depression and anxiety scores, user experience scores indicated that participants considered the system to lack efficiency and novelty. Finally, and in related work, Motalebi et al. examined barriers to the use of Amazon Alexa to implement clinical therapy for patients with PTSD; emphasizing the importance of short dialogues and interactivity for effective therapeutic content delivery [56].

Researchers have therefore also begun to explore the adaptation of commercially available systems to provide more accessible and conversational in-home experiences, including the design and use of CAs to support the self-report of mental health and well-being "in-the-wild." Few studies have yet, however, to generate knowledge of the lived experience of users in relation to the practice of self-report via CA or their perceptions of the use of these systems for this purpose. This knowledge is required to support the ethical and sustainable use of CAs among users with ADs in particular and yet must be approached in light of the present limitations of these technologies.

2.2.1 CAs and Their Limitations. While speech-enabled CAs are gaining widespread adoption in daily use [11, 18, 71], we cannot speak about the potential of CAs to support practices of self-report without mentioning the current limitations of these technologies, which are often significant in nature [16, 52, 55, 76, 77].

First, speech recognition technology is still in its infancy and cannot fully interpret or recognize users' utterances. A recent assessment of automatic speech recognition performance in psychotherapy discourse using Google Cloud's Speech-to-Text service reported a transcription error rate of 25% in general conversation. For depression-related utterances, a sensitivity of 80% was reported along with a positive predictive value of 83%, and for clinician-identified harm-related sentences, the word error rate was 34% [53]. While the authors of this study cautiously suggest that the technology may be feasibly adopted in psychotherapy, numerous other studies have reported instances in which CAs misinterpret or fail to recognize user utterances causing confusion in the conversation and leading to further errors [58, 68, 77].

Second, while some CAs can mimic human-to-humanlike conversations, none are fully capable of engaging in dynamic conversation [16]. Typically, users have to know in advance what they

can or cannot say to the CA [20, 30, 65] which often makes it cognitively demanding for users to interact [65]. Moreover, current commercially available CAs allow a maximum of 12 seconds for each user's response [42] and do not understand pauses in users' utterances, which limits open-ended opportunities for self-report and dialogue.

Third, the monotonous and robotic voice of CAs can undeniably impede users' engagement [54]. Finally, studies have shown that users often establish high expectations for CAs [52], anticipating emotional exchanges, relationship building, and human-to-humanlike conversations [16], which often leads to a drastic decline in use when CAs cannot meet users' expectations [16, 52].

Several researchers have provided initial guidelines to address these limitations in support of the design of effective CAs [19, 43, 57, 77, 88], including for health and well-being management [75]. Studies also report that users themselves apply tactics including hyper-articulation and exaggeration, increased volume, use of different utterances or simplified words, and reformulation strategies, such as addition or substitution, removal, and re-ordering of words to alleviate conversational barriers [15, 58, 62].

It is currently unclear, however, of the extent to which these guidelines for the design of effective dialogue and for overcoming CAs' limitations apply to the mental health context. We know little about how people living with mental illness including ADs might engage with CAs in light of these current technical limitations or in support of the self-report of their mental health and well-being. To address these research gaps, we design and conduct an "in-the-wild" field study.

3 METHOD

A 4-week "in-the-wild" study was designed to enable people living with ADs to experience the self-report of mental health and well-being via CA. Our aims in conducting this study were to generate an understanding of users' perspectives and experiences with respect to the use of CAs for the self-report of mental health and to identify key factors in the design of sustainable and engaging self-report experiences.

3.1 Participant Recruitment

Twenty participants (16 female) with an average age of 27.5 living in three regions of Denmark (Odense, Copenhagen, and Jutland) and with diverse backgrounds were recruited via online platforms including a national patient recruitment site,[1] social media (e.g., Facebook, Twitter), university internal email, and posters placed within the university campus. The study was advertised as a non-clinical study providing users the opportunity to interact with a conversational agent via Google Nest device for the self-report of mental health and well-being. Participants were therefore not screened for ADs. Inclusion criteria for participating in the study required participants to (i) be over the age of 18, (ii) identify as diagnosed with an AD, and (iii) have continuous access to WiFi. Table 1 provides an overview of participants' demographic characteristics. Of the 20 participants, 45% reported that they had other health conditions and 25% reported no prior experience using CAs.

Participants were offered a Google Nest Mini 2 device or a gift card corresponding to DKK 300 (~US$50) for their participation. This study was exempted from ethical approval by the Danish National Committee on Health Research Ethics (Journal no. 21024788) in accordance with Section 14 (2) of the Danish Act on Research Ethics Review of Health Research Projects [60]. However, the highest research ethics standards were adhered to at each step of this process including during the recruitment and engagement of participants [60].

---

[1]http://www.forsoegsperson.dk/.

Table 1. Participant Demographics

ID	Sex	Age	Education	Employment	Diagnosed	Symptoms last experienced	Other health conditions	Technical ability	CA exp.
1	F	25–34	Masters/PhD	Student	<2 years	<a week	None	V. good	Yes
2	F	18–24	Primary	Student	<5 years	<3 mths	Hypothyroidism	V. good	Yes
3	F	18–24	Primary	Employed	<12 mths	<a week	None	Fair	No
4	F	25–34	Bachelors	Student	<5 years	IDR	None	V. good	Yes
5	M	25–34	Masters/PhD	Student	<12 mths	<3 mths	Cancer	V. good	Yes
6	F	25–34	Upper secondary	Student	<12 mths	<a week	Schizophrenia	V. good	Yes
7	F	25–34	Below degree level	Employed	<2 years	<a week	Anxiety, PMDD	V. good	Yes
8	F	18–24	Masters/PhD	Student	<12 mths	<3 mths	None	V. good	No
9	F	25–34	Masters/PhD	Employed	<12 mths	<a week	Scalp psoriasis	V. good	Yes
10	F	25–34	Upper secondary	Employed	<5 years	> 3 mths	None	Good	Yes
11	M	18–24	Masters/PhD	Student	>5 years	IDR	None	Good	Yes
12	F	18–24	Below degree level	Student	<12 mths	<a week	None	Good	Yes
13	F	45–54	Masters/PhD	Employed	<5 years	IDR	None	V. good	Yes
14	F	25–34	Masters/PhD	Student	<5 years	<3 mths	OCD	V. good	No
15	F	25–34	Masters/PhD	Student	<5 years	<3 mths	Migraine	Good	Yes
16	M	25–34	Masters/PhD	Student	<5 years	<3 mths	Asperger Syndrome	V. good	Yes
17	F	25–34	Prefer not to answer	Unemployed	<12 mths	<a week	Pregnancy	V. good	No
18	F	25–34	Bachelors	Employed	<12 mths	<a week	Stress	V. good	Yes
19	M	18–24	Primary	Unemployed	<5 years	<3 mths	None	V. good	Yes
20	F	18–24	Bachelors	Student	<2 years	<a week	None	V. good	No

CA exp. = Prior experience with CAs, PMDD = Premenstrual Dysphoric Disorder, OCD = Obsessive Compulsive Disorder, IDR = I do not remember.

3.2 Conversational Agent Implementation

In order for participants to experience a state-of-the-art speech-based self-reporting practice, we designed a prototype agent named *Sofia* using Dialogflow[2] and deployed it via Google Nest Mini smart speakers.

Sofia combines (a) an open-ended dialog design emulating a human-to-humanlike conversation allowing users to freely express their mental state at any time with (b) fortnightly completion of the WHO-5 health questionnaire. The open-ended dialog design of *Sofia* was adapted from a similar study of CAs for healthcare decision support [23] and comprised three separate questions. The first question asked users to express their current emotional and mental health status, the second asked users to elaborate further upon their mental health and well-being, and the third asked users whether there was anything more they would like to add. We employed the "Wizard of Oz" [3] method involving the participation of the authors, one external HCI researcher, and three university students to contrast multiple dialog designs to reduce the otherwise repetitive nature of the conversation. This process contributed several insights into the design of the conversation. For example, while responding to the open-ended question, "How are you feeling today?," one participant added a follow-up question to their response, saying, "I'm feeling great. How about you?" The final conversation design handled such follow-up questions by responding, "I'm fine." Similarly, in response to *Sofia*'s prompt to continue the conversation, "Please share more about your feelings," users often asked, "What do you want to know?" In line with our

---

[2]https://dialogflow.com.

Table 2. Dialog Management in *Sofia*: Key Intents and Their Purposes and Examples

Intents	Purpose	Example
Greeting	Build rapport	User: Hey Google, talk to *Sofia*. *Sofia*: Good evening, [name of the user].
Response Prompting	Ask for information	*Sofia*: How are you feeling today?
Response Feedback	Acknowledge the user's responses	User: I'm feeling great! *Sofia*: Thanks for sharing that.
Continuation prompts	Ask for more information	*Sofia*: Please share more about your feelings.
Handle user question	Answer user queries	User: What do you want to know? *Sofia*: Anything you want to share.
Confirmation prompts	Ask if there is more to add	User: I had a great time with my friends today. *Sofia*: Do you want to add anything else?
No response error	Ask user for the response	User: [Silence] *Sofia*: Sorry, I didn't get that. Please share more about your feelings.
No match error	Ask user for a response within the provided options	*Sofia*: On the scale of 0 to 5, 0 being "at no time" and 5 being "all of the time," over the last 2 weeks, how often have you felt cheerful and in good spirits? User: I don't know. *Sofia*: You can answer the question on the scale of 0 to 5.
Farewell	Indicate end of the session	User: I don't have anything else to share. *Sofia*: Talk to you next time. Bye!

motivation to keep the conversational flow open-ended, *Sofia* would respond to this query by saying, "Anything you want to share." Table 2 provides an overview of the dialog management as implemented in *Sofia* including key intents, their purpose, and example dialogues.

The World Health Organization-Five Well-Being Index (WHO-5) scale was chosen for use through *Sofia*—an established self-reporting questionnaire requiring respondents to summarize their well-being over the past 2 weeks by responding to five short questions according to a Likert scale. This questionnaire is less invasive than many clinical scales and considered appropriate for a general population group [81]. To render the WHO-5 questionnaire more conversational in nature, slight variations were incorporated into the wording of the preamble and the questions, as previously practiced in digital mental health research [82].

The purpose of combining these two distinct forms of self-report (open-ended and WHO-5 Likert scale) was to enable us to understand users' experiences of different variants of conversational self-report rather than to measure changes in users' well-being or emotions over time. While significant validation studies would be essential to enable the use of self-reports obtained through either the WHO-5 questionnaire or open-ended diary for clinical use, it is important to understand the relationship between different conversational design strategies and user's self-reporting experiences to inform future CA designs, given that we are likely to soon see increased conversion of these questionnaires for delivery via CA following growing adoption of these systems in healthcare [11].

Sofia provides fallback re-prompts following two types of errors: (i) "no response" in which case the respondent takes too long to respond and (ii) "no match" in which case *Sofia* does not understand the respondent. Respondents have three opportunities to respond to a question following these errors. After three re-prompts, *Sofia* ends the conversation. If needed, users can ask *Sofia* to repeat the question by stating either "repeat" or "what was the question?" During the fortnightly

WHO-5 questionnaire, users can also ask for help by stating "help" or "what are my options," in which case *Sofia* repeats the preamble ("You can answer the question on the scale of 0 to 5; 0 being at no time, and 5 being all of the time"). Respondents can also end the conversation at any time by voicing the phrase "quit" or "stop." Upon completion of the session, *Sofia* thanks the user for sharing their information and bids them farewell.

3.3   Study Procedure

This study in its complete form comprised a between-group design contrasting the outcomes of two distinct self-report methods, *Sofia* and a separate text-based web app, with respect to the characteristics of the self-reported data and users' engagement. In this article, however, we focus on the qualitative analysis of post-study interviews conducted with the CA users alone.

The study procedure comprised the following three phases.

3.3.1   Pre-study. During individual pre-study sessions, each participant was provided with a participant information sheet detailing the study's motivations and expectations of participants. This handout also described the CA's limitations (e.g., the 8- to 12-second limit on each open-ended response), instructions regarding how to interact with *Sofia* (e.g., invocation phrases, repeating questions, ending the conversation), what *Sofia* can and cannot do as well as the data collected though *Sofia* during the study, and informed participants of the research team's access to the self-report transcripts.

Each eligible participant was asked to sign a consent form and provide demographic information including their name, email, age, gender, education level, employment, year when diagnosed with AD, symptoms last experienced, other known health conditions, technical ability, and prior experience using CAs.

Participants were then familiarized with *Sofia* by taking them through the set-up process, which involved pairing the Google Nest device with users' smartphone devices and creating an account with *Sofia*, as well as a sample conversation. While creating a user account, users were asked to share their details such as name, email address, and profile picture with *Sofia*.

Participants were informed that *Sofia* did not actively trigger notifications to remind participants to engage in the self-report of their mental health. They were, however, encouraged to use Google Assistant's "routine"[3] feature, ask the assistant to remind them to speak to *Sofia* everyday, or use other methods of their own preference to set up a daily reminder.

3.3.2   In-study. During the 4-week study, each participant was expected to invoke *Sofia* to self-report their mental state: by answering a series of three questions every day, and responding to the WHO-5 questionnaire fortnightly. Each self-report session began with the participant invoking *Sofia*. *Sofia* then greeted the user and asked how they were feeling. The agent next took the user through two additional open-ended questions designed to gather additional insight into their mental state. Every 2 weeks, *Sofia* also asked the user to respond to the WHO-5 questionnaire, following conclusion of the daily open-ended conversation. Data collected through *Sofia* included automatically transcribed participant responses to the open-ended questions as well as the WHO-5 questionnaire and timestamps for each question and response.

3.3.3   Post-study. Upon conclusion of the study, participants filled out the UEQ [46] questionnaire, yielding a subjective assessment of *Sofia*'s usability and their individual experience.[4]

---

[3]https://support.google.com/googlenest/answer/7029585?co=GENIE.Platform%3DAndroid&hl=en.
[4]The UEQ is a 26-item, seven point Likert scale questionnaire covering six key user experience dimensions; Attractiveness, Perspicuity, Efficiency, Dependability, Stimulation, and Novelty [46].

Finally, each participant took part in a semi-structured interview regarding their experience of *Sofia*, their overarching views of the technology and thoughts on the future design of speech-enabled conversational self-report agents. Interviews lasted between 25 to 65 minutes and were audio recorded. Due to COVID-19 guidelines regarding social distancing and travel restrictions, 50% of the pre-study sessions were conducted online, and the Google Nest device shipped directly to participants; 85% of the post-study interviews were conducted online. Due to constraints on their personal time, one participant (P17) was interviewed via a series of emails.

3.4 Analysis

Following the principles of Braun and Clarke's thematic analysis [9, 10], authors 1 and 2 analyzed the interview transcripts together, generated initial codes by inductive process, and grouped related codes and supporting quotes into categories to form candidate themes. These candidate themes were then iteratively reviewed and refined with the additional involvement of the third author to produce final themes. Participants' responses to the UEQ questionnaire were analyzed using the data analysis tool provided[5] in Microsoft Excel (v. 16.16.27). Participants' log data, including self-reports and their timestamps were additionally analyzed in R (v. 3.6.2). Participants' open-ended self-reports were summarized in a word cloud following removal of stop-words, punctuation, symbols and any words appearing less than 5 times in the corpus using the "Quanteda" (v. 2.1.1) library in R [5].

4 RESULTS

We next present our findings from qualitative analysis of post-study interviews conducted with 20 people living with AD following their participation in a 4-week "in-the-wild" study involving daily and fortnightly self-reporting of mental health and well-being via CA. Our findings reveal users' experience as strongly shaped by actions taken to overcome the technological limitations of CAs, diverse personified perceptions of a self-report agent, the socially contingent nature of self-reporting practice, users' reflections on privacy and security concerns, as well as the CA's conversational features considered important for engaging and sustainable self-report experiences.

First, however, we provide quantitative insight into participants' engagement with the technology and experience of the CA as context for our qualitative findings.

4.1 Participant Engagement

Analysis of participants' engagement with the CA by means of interaction log data resulted in a global average engagement rate (calculated as the percentage of 560 total self-report sessions) of 75% (SD = 5.91)—a figure comparable to rates of "compliance" reported in prior Ecological Momentary Assessment (EMA) studies using mobile devices [91]. Figure 1 shows the temporal distribution of each participant's usage pattern over the course of the study.

During the study, 20,154 words from 418 daily open-ended self-reports were collected. On average, a participant's daily self-report contained 48.22 (SD = 30.48) words, lasting 46.96 seconds (SD = 31.33). The shortest self-report consisted of 4 words while the longest contained 225 words. Figure 2 shows the words most frequently used by participants when responding to the open-ended questions eliciting expression of their emotional state. More than half (59.33%) of participants engaged with the system during the evening hours of 18:00 to 21:00 as shown in Figure 3.

As shown in Figure 4, participants' responses to the UEQ questionnaire upon conclusion of the study reveal an overall positive impression of their self-report experience in terms of attractiveness

---

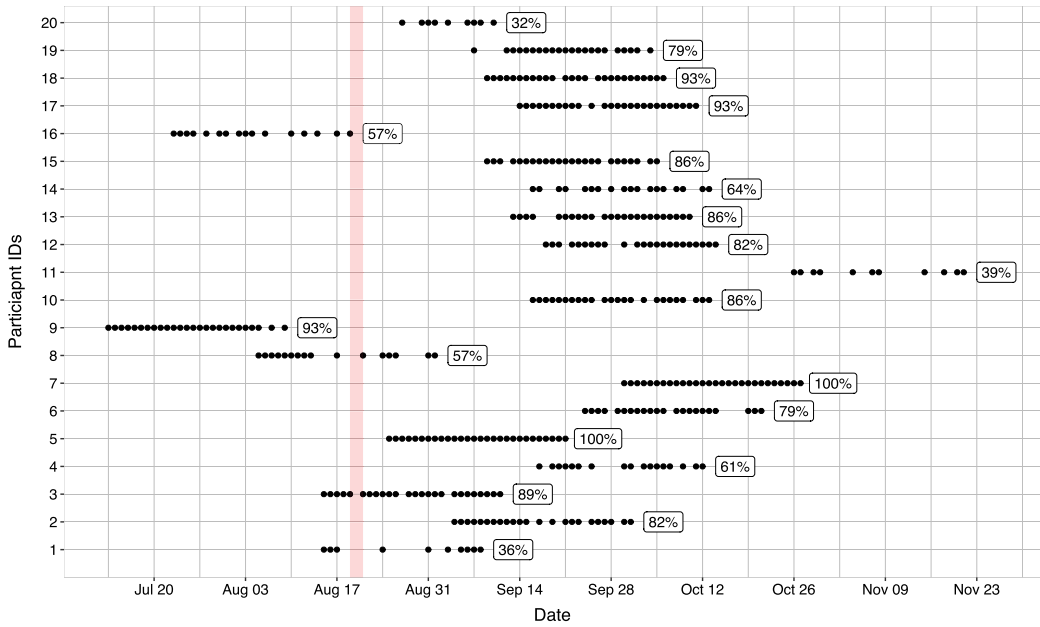[5]https://www.ueq-online.org/Material/Data_Analysis_Tools.zip.

Fig. 1. Participants' engagement with the daily self-report of mental health and well-being via *Sofia*. Each dot represents an entry for the day. Annotations reflect each participant's rate of engagement calculated as the percentage of 28 self-report sessions over the course of the study. The red bar indicates Dialogflow's service outage on August 20, 2020. Participants were not able to use *Sofia* on that day.

Fig. 2. A word cloud generated from the participants' open-ended self-reports of their mental health and well-being.
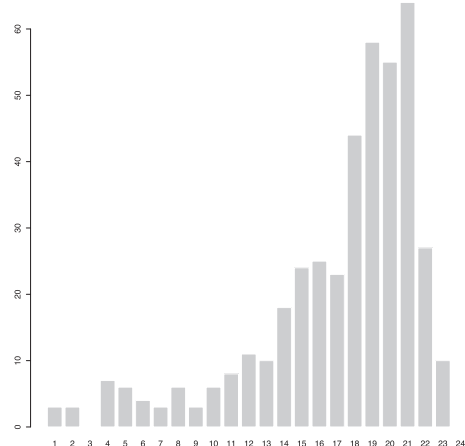

Fig. 3. Self-report frequency plotted against time of day. The *x*-axis represents the time of day in 24-hour format, and the *y*-axis self-report frequency.

(U = 1.09, SD = 1.01), perspicuity (U = 2.09, SD = 0.75), efficiency (U = 0.98, SD = 0.77), dependability (U =0.74, SD = 0.84), stimulation (U = 0.38, SD = 0.82), and novelty (U = 0.71, SD = 0.88).

These data serve as context for our thematic analysis of interviews conducted with participants following their engagement in this 4-week "in-the-wild" study; the primary findings of which we present next.
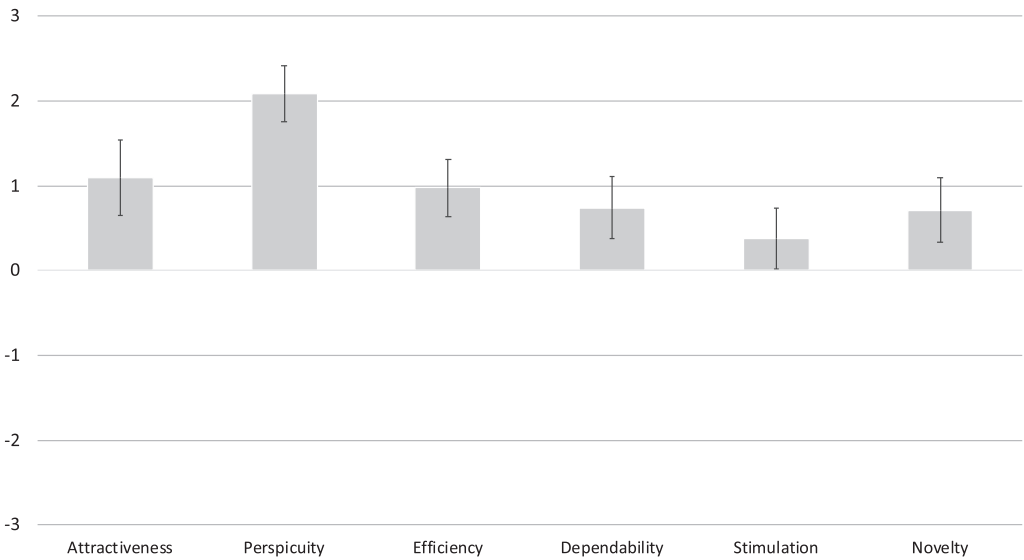
Fig. 4. Participants' mean UEQ score with respect to each dimension. Error bars reflect a 95% confidence interval, the *y*-axis represents scores ranging from −3 (horribly bad) to +3 (extremely good), and the *x*-axis illustrates the six dimensions of the scale.

4.2 Theme 1: Engaging with and Getting around the Technology

The theme we first discuss is one that informed all others, and stems from users' frequent comments to the need to get around the technology to fully engage with it. These comments relate for the most part not to choices made in the design of this particular CA but to the more general limitations of commercially available smart-speaker technology. These limitations undoubtedly had an impact on participants' self-report experiences, and yet the lengths to which participants went to overcome these limitations and maintain their engagement highlights the value participants saw in their use of the technology.

4.2.1 Frustration, Rejection, Illusion, and Technical Limitations. Limitations described by participants include the CA's tendency to frequently interrupt and thereby impair their self-report experiences; due to both the CA's inability to recognize pauses between user utterances and the 12 second limit imposed on user responses. This would mean that participants were often interrupted in the middle of a sentence by the following question before they had completed their desired response.

These experiences understandably proved frustrating for users, "I'm, like, 'No, bitch, you cut me off. So, shut up…'" [P9], and led to some participants quitting the conversation early, "Okay, fine, that's it for today" [P18]. Several participants compared their CA interactions with a human-to-human conversation and in turn expressed feelings of rejection following repeated interruptions and the inability to fully recount their emotional experience:

> I know it's not a human being. I know it's this little round little thing, but still it's, like, I'm trying to be personal here, and then you're interrupting me. It feels like a rejection. [P9]

> I felt, like, a bit like when you talk with a friend, and they don't want to listen to you. [P8]

P8, in this example, relates their interrupted self-report experience to a conversation with a person who does not care, leading to a sense of disconnect. Similar accounts were shared by other participants who noted that it was important for them to feel listened to and cared for, despite their awareness that the CA could not understand their emotions nor in reality "care" about what they shared:

> Of course, the software doesn't care. But, you want the illusion that it cares about you. [P5]

Many of these limitations of CA technologies are well known and well documented in prior literature (see Section 2.2.1). What is most striking about participants' accounts as shared during this study, however, is the extent to which these technical limitations were recounted in emotional terms and to which this reflects a desire by participants to relate to the technology. This theme is therefore furthermore supported by the strategies employed by participants to overcome these challenges.

4.2.2 Strategies to Overcome CA Limitations. Participants described their development of a number of strategies to maintain their engagement;

(i) Making Multiple Entries. At the pre-study session, participants were given some insight into the limitations of the technology, and adding multiple entries was offered as a possible strategy for overcoming these challenges, should users encounter them. Only a small number of participants adopted this strategy, however; either accepting the technology's limitations and sympathizing with the CA or, interestingly, contrasting the CA's 12-second limitation with a human interlocutor's potentially limited capacity to listen to others:

> With my frustration, I took a deep breath and thought it's just a technology. It's in its starting phase—I let it do its thing. [P3]

> I guess that there's also, like, a limit to how, how long she can listen, like another person would also have. [P12]

Other participants noted that they did not find this a feasible strategy, commenting that they would lose their train of thought should they start a new session; "The problem is that you lose your momentum or train of thought when you're interrupted like that" [P9]. P2, in contrast, compared their interactions to a human-to-human conversation and thought that it would be rude to ask *Sofia* to listen again, "when you're done talking to her and she is, like, 'Thanks for sharing that. Have a nice day'...It would feel rude to ask her again."

(ii) Adapting One's Speech. Other participants adopted more active strategies. This included changing their natural speech patterns by stretching out their utterances, using filler words, and speaking faster or more concisely:

> I tried different tricks, like, for instance, I tried to make some songs like 'hmmm leeet meee thiiink,' but it seems like my process was too slow for *Sofia*. So at one point, I started giving very short answers so I won't be cut off yet could answer. [P1]

Although considered effective, this approach was also certainly viewed by participants as a compromise: "My speech is a bit artificial, not very smooth, and didn't feel fluent. I had to change the way I talk to fit what *Sofia* can accept" [P8].

(iii) Preparing in Advance. Another strategy spontaneously adopted by participants was to reflect in advance on what they desired to share with the CA. P16 for example, recounted how they would divide their response into three parts to match the design of *Sofia*'s dialog flow:

I realized, okay two sentences for the first question, two sentences for the second question, and one sentence for the last question. Just five sentences overall.

Interestingly, some participants commented that this process of formulating their thoughts in advance in itself helped them to reflect on their day. P11 viewed this process as "meditative," commenting, that "it helps you organize things in your head…It's, like, 'you time,' you know, I believe it helps you clear your mind." P6 compared the experience to composing a brief diary entry that served to focus their reflections; "It's not just three pages of how I saw something nice and I had a nice cup of coffee…It really is an emphasis on three things which really matter to me."

Although for some participants such tactics led them to feel their responses were increasingly shallow, "She expects an answer in few seconds, so I'll be, like, 'tired,' 'decent,' 'not sure.' So in that sense, the answer might be a bit less thought out" [P2]. Others showed a surprisingly significant willingness to adapt, at times expressed in overtly human terms,

I started thinking about what I want to say before I called her up. Normally, when people talk about feelings, they need space and time for pauses. I knew my *Sofia* friend could not really do that. So we had to do it on her terms. [P7]

A number of participants spoke to their increased capacity to adapt to the technology's limitations over time, despite the undoubted impact on their experience: "It was kind of hard in the beginning as she is really sensitive to pauses and doesn't give you any space to think. But my brain adapted [to] that really quite quick" [P7]. This willingness to adapt can only be interpreted in light of the value users associated with their interactions with the technology—which, in turn and despite these limitations, appears to be closely related to their perceptions of the agent itself, as reflected in our second theme.

4.3 Theme 2: "Sometimes All You Need Is Someone Who Listens"

Engaging in the self-report of mental health and well-being by any means is an exercise in vulnerability. And many participants spoke to this effect, commenting how they would refrain from sharing their emotions even with close friends and family due to the social stigma attached to their illness. P7 and P14 noted the need to think about the potential repercussions of disclosure, listeners' reactions, as well as the potential for others to demonstrate disinterest in the subject, each of which demotivated their sharing of emotion:

You know, my dysphoria (PMDD) often makes it so that I think that I'm a bother to people, like 'It's really annoying to listen to you, bitch'… [P7]

Another barrier to self-expression recounted by participants pertained to the potential for listeners to become pre-occupied with trying to understand their emotions in depth, requiring them to repeatedly explain what they are experiencing, when most of all they wished simply to be heard, "When you say you feel sad and empty, you don't really want the focus to be on describing how emptiness feels. You just want to say, 'I'm feeling sad…'" [P6]. Strikingly, participants also frequently commented on their struggle with listeners' keen desire to "solve" their problems rather than listening to them. They explained that they usually do not require a solution to their problems, and due to the chronic nature of their conditions, often know how to cope with their situation and ongoing mental state:

Sometimes all you need is someone who listens than have an answer…because there's not necessarily anything to solve, you just need someone to share your thoughts and feelings so that you are not alone. [P2]

This desire to be listened to, rather than to have their problems solved, goes some way toward explaining the value users saw in this particular CA implementation, although interpretations and perceptions of the CA itself were strikingly diverse.

4.3.1   CA as Good Listener. Throughout the interviews, participants' comments echoed a desire for "emotional support instead of emotional counseling" [P7], a role for which *Sofia* was often appropriated, as "someone" who made them feel heard. For many participants, therefore *Sofia* was simply a good listener, who allowed them to express their emotions freely:

> It sounds a bit stupid to say but I could say that I'm glad someone listens, except you know there isn't actually someone that listens. But it feels like it. [P2]

> She is a good listener. Sometimes a listener is exactly what you need in a situation like that. [P6]

Participants valued the non-judgmental nature of conversations with *Sofia*, which neither provided unwanted solutions nor meant that they had to worry about potential judgment and repercussions for sharing their feelings: "Since I knew *Sofia* doesn't mind, I felt in some instances, in a twisted way, I felt like she cared, like she was always there no matter how negative or positive I was feeling" [P7]. They furthermore appreciated *Sofia*'s compassionate feedback and mentioned that it would provide a sense of comfort and empathy when feeling depressed:

> When you have a diagnosis like this, you know that only talking doesn't make it go away, but sometimes getting assurance, 'Okay, it's okay to feel, it's okay to experience what you experienced'—that helps a lot. [P6]

4.3.2   CA as Machine Companion. Other participants employed a vocabulary of companionship when describing *Sofia*, associating the technology's value with its ability to fill a felt gap in their social interactions:

> When you are depressed, your social circle tends to restrain more and more. So, you've less opportunity to talk to someone. I felt like I was talking to someone, even though *Sofia* is not very smart. Just because of this feeling, I think it is very useful for people with depression. [P1]

For some participants, the CA's value, therefore stemmed from its constant availability: "I do have friends to talk to. But a friend could be busy. When you need to talk it out, it's in your reach even though it's a machine" [P5]. Others spoke of the CA as also more approachable and trustworthy—a perception of the CA as a harmless machine meaning that it could not turn against users no matter what they shared:

> I have massive trust issues. I grew up thinking that whoever you tell anything, they will use it against you. But, since I know that *Sofia* is a machine, she doesn't really think that much of herself and she's just there sitting on my table. That kind of makes it a little bit more approachable. [P7]

And yet, for P4, talking to a CA about their emotions could equally serve to make them feel lonely, given that the CA could not "truly" understand their feelings: "Sometimes she made me feel a bit lonely, somehow, because you're just reminded that you're talking to a machine, who has no capacity to understand what you're actually feeling...something that kind of tries to mimic a human being, but it's not convincing."

4.3.3   CA as Human. Other participants went to greater lengths to personify the CA, with both positive and negative implications and associations.

Participants with positive perceptions of their self-report experiences tended to personify the CA as a friend, therapist, or talking diary. Although noting their awareness that the CA was not human, participants would often comment that the anthropomorphic characteristics of the CA granted them the impression of a conversation with a person. P10, for example, who personified *Sofia* as a friend, commented, "I think it's the whole, having a voice and having a name, I keep referring to it as her instead of it, even though I'm aware that it's a system and not a person."

P14 personified *Sofia* as an older friend and mentioned that although she could also be their age, *Sofia*'s voice made them think of someone older: "Its more like an older friend. I mean, I'm also quite young, I'm 25, so maybe that's why? Maybe it's also the voice that reminds me of someone a bit older than I am." P15, however, viewed the CA as a "talking diary" and compared their self-reporting practice to talking to their dog: "I compare *Sofia* to my dogs. Because a lot of the times, especially when I was growing up, if I was upset I went outside, sat down and talked to my dog, who had no freaking clue what I was talking about, and it made me feel better."

In contrast, participants with negative perceptions of their self-report experiences personified *Sofia* as an elderly woman, an uncaring friend, and a teenager trying to act grown-up. Referring to the CA's tendency to frequently interrupt her mid-sentence, P9 personified *Sofia* as a friend who does not care: "It's kind of, like, you know, that friend who like sits with the phone when you try to talk about something deep, like, 'Okay, fine, just say you don't want to listen to me….'" Likewise, P3 personified the CA as an impatient old lady, an impression informed by the 12-second response limitation and tone of voice: "*Sofia* is like a very impatient woman, and she's probably quite old. She's very monotoned, there's no variation."

These various personifications also shaped participants' responses to the agent, including their self-reporting behaviors. A positive personification of the CA motivated P6 to find and share positive emotions even when negative feelings were abundant; "You feel, like, 'Oh, I'm talking to someone and I don't just want to be negative. I want to say something positive.'" However, for P9, a negative perception of the agent led them to vent their frustrations directly to the device: "Sometimes I guess it took my emotions out. It was probably not her causing the reaction, but because I was annoyed about something, and then I took it out on her. Sorry, *Sofia*."

4.3.4 CA as Blank Slate. Finally, a number of participants spoke of emotional venting and self-talking as useful practices enabled by the smart speaker device. These accounts reflect interpretation of the CA as a blank slate. P11 and P14 considered self-reporting via CA a self-talking exercise and stated that doing so helped them to clear their mind [12, 37, 83]. As P14 noted,

> Once you say stuff out loud, it just changes how you think about certain things…when I have them in my head, it just sounds like they are super important. If I just talk out loud, then suddenly it becomes less important and I realize that they're just thoughts and they're not really like who I am.

Emotional venting has likewise been regarded as an effective means of finding relief by releasing strong or repressed emotions [4, 47, 80], a practice on which P8 commented,

> I almost see it as, you know, when you get super upset, you go out and scream, you go very far away and you just scream at the wind, cows or whatever. And you know that you're never going to get anything back, like, no one is going to respond to you, but it feels good to just express.

These various perceptions of *Sofia* suggest the value of different framings and implementations of mental heath agents and multiple paths forward for their future design for varied yet valuable

purposes. The prevalence of personification in itself furthermore underlined the extent to which participants perceived self-report as a socially contingent practice.

4.4 Theme 3: Self (Report) Is Social

While all forms of self-report, and, indeed, human behavior, are to one degree or another socially contingent in nature, participants' experiences of *Sofia* in particular revealed a diverse variety of social associations. Many of these accounts stem, at least in part from the spoken nature of the interaction itself, and in turn its more open nature.

4.4.1 A Private Practice Rendered Social. Eight of the twenty (40%) participants in this study mentioned that they engaged with CA in an open environment (e.g., a co-living space) where the presence of others' was likely to have an effect. And yet this was not the only way in which social considerations were seen to manifest. The choice of a number of participants to make their self-reporting a private practice was also informed by their social circle's perception of the CA, as well as their relationships to their loved ones.

The self-report of mental health and well-being is most typically a private practice, due at least in part to the social stigma attached to many mental health conditions, and the potential lack of privacy when engaging with a CA therefore creates a unique self-reporting context. Participants P17 and P20 commented that it felt "weird" and "awkward" to talk to *Sofia* when other family members were around: "It felt a bit weird talking to a speaker and was especially weird when my husband was home. It was unnatural in general but especially when I had someone else listening at the same time" [P17]. For P4, this experience was awkward because it created confusion when their partner did not know that they were talking to the CA,

> Talking to *Sofia* is little awkward because because if my boyfriend wasn't aware of what I was doing, he's like, 'What? Do you want something from me?'…like, 'No, no. I'm talking to the device.' Yeah, 'I'm talking to this other thing in our house.'

P6 shared that although it was also strange for them to talk to *Sofia* initially, it became natural due to their partner's support—serving as an interesting example of the implicit value of such a technology in providing an opportunity for others to demonstrate care:

> At first, I felt like this is something personal. Like, this is something I should hide, like a diary under the pillow or something. But, well, first of all, my husband has been really supportive and he is, like, 'Remember your *Sofia*, remember to talk to her.' So it became more natural to talk to her.

Of course, reporting practices varied among participants, as intended by a study designed to allow users to appropriate the technology as they saw fit. P11 noted that they were comfortable sharing their feelings in the presence of others, "I don't mind…sharing these things with my friends, or anyone to be honest," whereas P3 commented that they refrained entirely from self-reporting when friends were around: "I'd have friends over so during those times, I didn't log a diary—You don't want to let everyone know about deep personal life."

4.4.2 A New Member of the Social Circle. Other participants went further and spoke of *Sofia* as entering into their social circle. A number of participants commented that their social circle held a positive view of the CA and supported the idea of talking to *Sofia* about their emotions. P1's partner, for example, told P1 that *Sofia* could help them when they were feeling down: "My boyfriend told me that maybe I should use *Sofia* when I was down. He told me that I shouldn't be closing myself to the world, I should be talking. He said it might be useful."

P6 spoke openly about the study with friends and family and described their friends as very curious about the technology, which quickly became a topic of conversation in their daily lives:

> I remember a friend who said, 'So how's it going with *Sofia*?,' and they started joking and, like, 'So when is she going to tell you about her?' It's quite interesting how she became this little person in my life somehow.

P6 continued that their father was equally excited about the technology and shared that he had in turn also expressed an interest in purchasing a smart-speaker, as a potential means of addressing loneliness. However, they also wondered whether their father would be able to set up the device:

> I talked to my dad; he was really interested. He's, like, 'How does someone put a voice in this tiny box?' Like, he doesn't understand it's just code and such. I asked him if you would buy something like this, and he said, yeah he probably would, because he's also alone a lot of the times. But he also said, 'If I have to use it with apps and phones, I can't figure out how to do with them…'

P18's friends similarly thought that *Sofia* was "cool" and expressed its potential to support mental health and well-being. Their partner commented, however, that, although such a system could help those particularly isolated, it would not be able to replace a more human connection:

> "My boyfriend is a therapist himself and he said that that might be smart. It could do something for very, very, very lonely people. They could feel like, 'Well, there's someone who is interested in my day.' But we also talked about, like, we don't think an artificial intelligence can ever replace a real-life person.

4.4.3 An Influence on Close Personal Relationships. Finally, and perhaps most surprisingly, a number of participants spoke of the influence of *Sofia* on their intimate and personal relationships. Several noted that openly describing their emotions to the CA also led to conversations with their partners and served to strengthen their relationships.

P4 commented that after hearing her response to the fortnightly WHO-5 questionnaire, her partner realized for the first time just how she had been feeling during those 2 weeks: "When he heard me responding to *Sofia*, he was, like, 'Oh so this week is that shit? I didn't know…,'" creating a "different situation," "an intervention in your life where you're, like, 'So this is how you're actually feeling,'" despite the fact that "of course, we see each other every day, we live together. I do tell him if I'm tired or mad or whatever and try to keep him in the loop about my moods. But it's also just like sometimes you're not aware of it yourself."

This potential of the technology to shape the social fabric of a household was reflected in several other participants' comments. P6 mentioned that their mental condition could make it hard for them to explain exactly how they are feeling and that their tone of voice when speaking to *Sofia* therefore became the medium by which their significant other could best understand how they are actually feeling:

> Sometimes I will talk to her and not like I said anything specific but more like he [P6's husband] could feel from, how I was talking to *Sofia*, my tone of voice. Some days he would peek his head in and be like, 'Do you want to talk about anything? Are you okay?,' because I can be feeling extremely bad, but I'll look okay.

P14 spoke of a moment of realization during which she and her partner noticed that although they talked about their daily activities, these conversations did not necessarily reflect their feelings:

> My boyfriend does not share his emotions very often; neither do I. So sometimes when I was talking to *Sofia* and he (P14's boyfriend) noticed that I had some different feelings

which he wasn't aware of. It was quite good in that way. Now we are speaking more about feelings.

Many participants' comments therefore reflected an open and welcoming attitude toward *Sofia*. And yet, this did not preclude attention to questions of privacy and security, as our next theme highlights.

4.5   Theme 4: Personal Privacy and Data Security

In line with prior research [45], participants' engagement in self-report was also shaped by their perceptions of personal privacy and data security. Although open to interpretation as pragmatic in nature, the implications of these concerns were not straightforward, as this theme highlights.

4.5.1   Away from Prying Eyes and Ears. A number of participants recounted privacy concerns stemming from their own and their wider social circle's perception of the technology. The concern most-often shared related to the potential for the smart-speaker device to continuously listen and record conversations, even when not activated. To protect their privacy, participants described either turning off the smart speaker's mic or unplugging the smart speaker entirely. Participants P12 and P14, for example, both commented on the potential of the technology itself to serve as a source of anxiety and subsequently chose to keep the microphone off at all times, only turning it on when speaking to *Sofia*: "I felt a bit paranoid, so I was, like, turning the speaker off" [P12].

A small number of participants commented on their decision, in general, to resign from engaging with the challenge of navigating questions of privacy, due to a perception that it was broadly impossible to avoid pervasive collection of their data. P16, for example, mentioned that they did not mind sharing their data as long as they got something in return: "I'm not really concerned about privacy because I cannot stop them (tech companies) from collecting the data. But if I'm sharing something, I need something in return. For example, I really like that every month Google sends you a timeline and gives you how many kilometers you have walked."

4.5.2   Hands Off Our Data. In addition to their concerns regarding potential eavesdropping, participants expressed an interest in knowing and closely controlling how their data would be handled, citing understandable fears that their data might be shared with a third party or become publicly available.

Although appropriate limitations on the use of their self-reported data were carefully explained both in-person and through participant materials, these privacy concerns understandably had an impact on certain participants' self-reporting practices. P9 commented that they "probably held back, a great deal. Because of the whole, I know that it's not ending up in cyberspace for anyone to see, but there's still that thought that is a bit more out of my control than if I have it in a physical diary." Despite describing herself as an open book, P18 also commented that knowledge of the research team's access to her transcripts led her to refrain from fully recounting her mental states: "I knew that you would get the transcript of what I told her. Now, I'm not really a private person and I'm pretty much an open book, but I did refrain from saying stuff that I would have said if I knew you wouldn't be able to read it."

These concerns are not new, yet are particularly important to consider, and of particular ethical significance, in the lives of this population group, whose involvement in this study and engagement with this technology required trust and courage.

4.6   Theme 5: Designing for Conversation and Reflection

Many participants spoke of the value of the technology as allowing them to fully express their emotions and yet also as one of the ways in which the system was currently most limited.

Additionally, they highlighted the potential need to reflect on their self-reported data and dis-
cussed ways to enable this.

4.6.1 Designing for Conversation. Participants made several design recommendations to im-
prove the CA's conversational skill and allow users to express their emotions more fully.

(i) Varying More Often. Participants recommended incorporating greater variation in questions,
including voice characteristics (e.g. tone and intonation) and randomizing their presentation more
often so that the questions did not feel repetitive: "No one's going to ask you the same kind of
question in the same way with the same intonation every time. So having a few different questions
and rotating them would be better" [P18]. In contrast, however, P6 noted that the repetitiveness
of certain questions provided consistency in interaction that they found important: "What's really
important when you suffer from a condition like schizophrenia is consistency . . . And in that regard,
I almost find a sort of comfort in *Sofia*. She is like an anchor, which you can use to ground yourself,
because she always says the same thing."
Participants perceived the fortnightly WHO-5 questionnaire as a useful addition of variety to the
daily open-ended questions: "I was happy when those questionnaires came up—it was something
different at least" [P5], and several suggested also adding more discrete questions to the conversa-
tion design: "She could toss in some of that once in a while; that wouldn't actually be bad" [P4].
While many agreed that closed-ended questions can be more efficiently answered, P9 and P10 also
interestingly noted that responding to the discrete questions according to a pre-defined scale made
them feel more like study subjects: "You feel a bit more like a study subject, like, okay, 'How do
you feel from 1 to 10?; it's like, 'okay, 1 . . . 5 . . . 3 . . .'" [P9].

(ii) Probing Further. Participants suggested probing their responses in greater depth to further
engage them in a more "natural" conversation. They stressed that the CA does not need to un-
derstand everything but could employ strategies to support a richer conversation. P20 provided
an example, "For example, if I say, 'It was a very busy day, very overwhelming,' then she could
ask, like, 'How did you deal with that?,' 'How could you improve?' So it'd be best if there was an
Artificial Intelligence (AI) which could pull out some keywords and follow up on the basis of those
keywords."

(iii) Guiding More. Many participants also suggested alternate conversation designs. P18 com-
mented, for example, that when depressed, they would tend to have negative thoughts all the time
and would appreciate a conversation that encourages them to talk about something positive in
their lives: "She could ask me, like, 'Can you tell me a positive thing that happened to you today?'
I can always tell you about bad things happening in my day. It would be better if I was turned away
from that a little bit" [P18]. P15 shared similar views but argued that the conversation should delve
into both positive and negative emotions, including the reasons behind those emotions. P12 and
P19 likewise advocated enabling users to select the topic of the conversation, envisioning a set of
topics from which users could choose, allowing users to lead the conversation and reflect more
deeply on their emotions: "If it had a certain set of questions for different topics that would help
people reflect a lot more than having to be the one leading the story through the entire way" [P19].

4.6.2 Designing for Reflection. The practice of reporting entails reflection—a point a number
of participants highlighted—and the design of a self-report experience might therefore equally be
construed as structuring a practice of reflection. Participants' comments for improvements in turn
pertained not only to the current conversational design but also to the potential to support further
reflection on their own reported data. Many participants discussed the need for a visual tool, such

as a mobile or web app, to display their emotional well-being trends in support of self-reflection, arguing that this would be the easiest means of searching and examining their information.

Others, however, also raised the possibility of adapting the voice interface to support reflection. P14, for example, reflected on their use of a mobile app for this purpose and mentioned that they often ignored the data shown by the app. Via verbal reflection, they stated, it would feel more like confronting a problem, which they would find more meaningful than visually examining their data: "If someone is telling me, like, 'Hey man, I've noticed that the last couple of days you have been really stressed out, is there anything wrong?' or 'What's going on?,' it's more confronting... and I think that gives me way more, like, meaning than if I just see that in the app."

For P15, reflecting via CA could also provide value by granting them more scope to disagree with advice provided. Compared to the potential for confrontation to arise with friends, family, or therapist, they felt this approach could better support behavior change: "I would be way more likely to listen to a machine, because even if it tells me something like a recommendation or tells me what to do, if I don't want to do it, I'm not gonna do it." Participants discussed different ways in which the CA could present data verbally for users to reflect upon. They debated whether the CA should automatically announce such data every session or on-demand. Those who supported automated voicing of the data recommended that the information should be announced casually at the end of the session. Others cautioned that the idea of presenting data without users' request could prove intrusive depending upon the user's mental state: "It's kind of a double edged sword, because it can be intrusive if you don't want it. If you don't want any feedback, and then *Sofia* tells you, 'Oh it's been really bad the past three weeks,' that's the last thing that you want to hear" [P8].

We conclude this article with further reflection on these themes and their implications for design.

5 DISCUSSION

This article presents initial insight into the experiences of people living with AD of a speech-enabled CA for the self-report of mental health and well-being in an at-home setting. Our findings indicate that CAs' conversational features have the capacity to support engaging and sustainable self-report experiences, revealing users' experiences as strongly shaped by strategies adopted to overcome CAs' technical limitations, diverse personified perceptions of the agent, the socially contingent nature of self-reporting practice, and users' reflections on privacy and security concerns.

In light of these findings, we further reflect on the challenges of deploying CAs in the homes of users with AD and discuss implications for the future design of conversational agents to support the self-report of mental health and well-being (see Table 3).

5.1 The Challenge of Conversational Pattern Matching

Participants in this study demonstrated high levels of engagement in the practice of self-report, as with *Sofia*, throughout the 4-week period (see Section 4.1); many describing *Sofia* as efficient, easy to use, and an attractive medium for the self-report of mental health and well-being. This, despite frequent recounting of moments of frustration, disappointment, and even occasional anger in relation, most often, to the technical limitations of the technology. Participants recounted as particularly frustrating those limitations of the CA experience that presented barriers to their self-expression; including the imposition of an 8- to 12-second time limit on their responses.

Much prior HCI research has highlighted the challenge of user burden in the design of diverse self-report technologies [24, 32, 85]. And yet what was therefore most surprising in this instance were the lengths to which participants went, often spontaneously, to adopt strategies to maintain

Table 3. Challenges to the Self-report of Mental Health and Well-being via CA and Design Implications

Challenges	Design Implications
The Challenge of Conversational Pattern Matching	Tailor opportunities for self-expression through continuous probing and guiding of the conversation.
The Challenge of Filling the Right Gap	Attend to and provide space for human experiences of sociality, connectedness, empathy and compassion, while allowing users to appropriate technology in the ways they see fit.
The Challenge of The At-Home Social Context	Suggest transparent communication of CA privacy policies and data practices in addition to educating users about privacy settings, to build trust between users and these systems.

their engagement; from making multiple entries to engaging in long periods of prior reflection to providing brief responses in the moment (see Section 4.2.2).

These acts of perseverance and persistence demonstrate a willingness to adapt to the conversational patterns of an agent that one might expect during conversation with a human interlocutor. In this instance, participants were required to adapt to machine speech and dialogue patterns in service of the conversation. In future, a conversational agent might do likewise—leveraging this parallel human capacity to adapt to and match conversational patterns—to keep the conversation going, in the moment and over time.

This is a task we highlight as not only future but present design challenge, suggesting that designers might not only work to overcome CA limitations but also design for adaptation, and in turn sustainable and engaging self-report experiences, within these constraints. In line with participants' comments (see Section 4.6), this might entail first and foremost designing for a conversational pace and pattern congruent with the system's capabilities, by, for example tailoring opportunities for self-expression through continuous probing and guiding of the conversation.

5.2 The Challenge of Filling the Right Gap

Participants of this study recounted not only diverse strategies for interaction but also diverse personifications of an agent: from blank slate to machine, friend, good listener, companion, and tool for emotional venting and self-talking. Although a participant group whom might be classified as vulnerable in regards to their mental health, participants of this study therefore also demonstrated significant resilience in relation to their ability to appropriate the technology to fill the unmet gap in social interaction that they felt it might best—even were that simply "being listened to."

We were surprised, as authors, at the diverse forms of value participants found (and often creatively so) in a technology comprising a number of technical limitations. This practice of appropriation, while intriguing, also represents a significant challenge for the designer who might wish to provide a consistent user experience. For any attempt to shape such intense appropriation of a technology is additionally complicated by the contingency of the act of personification upon both human and machine.

We might then strive instead to support diverse appropriation of these systems in line with users' own needs—particularly in the context of design for the subjective, personal and interpersonal experience of mental health and well-being. Participants in this study, for example, shared a sense of being heard while talking to *Sofia*—a quality fundamental to realizing sustainable long term

CA-user relationships and that could prove additionally beneficial for encouraging positive behavior change as reported in prior research (e.g., References [7, 78])—if authentically felt and granted.

Personification of technology also raises, however, ethical questions in relation to the deployment of systems for the purposes of social interaction; from possibilities for stigmatization to disconnectedness. Morover, deploying CAs for the self-report of mental health and well-being in an ethical fashion will therefore also require attending to and providing space for human experiences of sociality, connectedness, empathy, and compassion, while allowing users to appropriate technology in the ways they see fit.

5.3   The Challenge of the At-Home Social Context

The at-home context in which this system was deployed led to a variety of unexpected findings related to the extent to which not only use of the CA was shaped by social context but also how the CA shaped social context and even intimate relationships. This poses both challenges and opportunities for designers when it comes to the integration of CAs within the unique social context of home use. The unpredictability of this context casts doubt on and raises ethical challenges for the enactment and shaping of new possibilities from opportunities for family members to demonstrate care to creating confrontations with and through data.

Introducing a system for the self-report of mental health into a home context must itself be considered an act of vulnerability requiring trust and courage and shaped by perceptions of CA as harmless machine or eavesdropping device. In this study, we found that participants' privacy concerns primarily stemmed from their distrust of the smart-speaker devices. As a measure to protect their personal privacy, many participants turned off the speaker when not in use, while others held back during self-reporting. Although similar findings have been reported in prior studies (e.g., References [45, 65]), the need for users with mental illness to trust the technology is even more significant as privacy and data security concerns could have additional adverse effects on their mental health and well-being (see Section 4.5.1).

We therefore suggest transparent communication of CA privacy policies and data practices in addition to educating users about privacy settings to build trust between users and these systems. Participants were made aware of the data practices adhered to in this study, including which data were to be collected and who would have access (see Section 3.3.1). Providing this information enabled participants to interact with the system according to their own discretion. While the designers of CAs cannot speak for the privacy policies and data practices of smart-speaker manufacturers and operators, they can support users' privacy concerns by educating them about the privacy settings of these devices to help build trust between users and their devices.

6   LIMITATIONS

This study was conducted during the COVID-19 outbreak. Most participants were affected by different levels of lock-downs, social distancing guidance, and travel restrictions. As a result, many participants mentioned that their mental, physical, and social situations at the time of the study were different than usual, which means that the study results may not translate to this population's more typical context. Some participants might, for example, have engaged more often with *Sofia* as they spent more of their time at home: "Now with Corona, there is not so much else to do in the evenings…and I live by myself, so it's nice that I can have like a daily talk with *Sofia*" [P14].

As described in Section 3.3.3, participants in this study were interviewed either online or in-person, and the method of interview could have affected how participants expressed their experiences. P6 for example commented that "I think it's actually really nice we're doing this on a Zoom meeting. If we have been like in person, I don't think I would have been so open about it. I would have been a bit more shy if we had to sit face to face and talk about this."

All study participants were recruited online and compensated with a Google Nest device for their participation. This may have attracted participants who were more likely to be early adopters of the technology, as suggested in prior literature [45]. Finally, our study comprises a small sample size (N = 20) from a single country and hence may not generalize to other contexts.

7 CONCLUSION

This article presents insight into the experiences of people living with AD of a speech-enabled CA for the self-report of mental health and well-being in an at-home setting over a period of 4 weeks. Results from thematic analysis of post-study interviews indicate that CAs' conversational features have the capacity to support engaging and sustainable self-report experiences, revealing users' experiences as strongly shaped by strategies adopted to overcome CAs' technical limitations, diverse personified perceptions of the agent, the socially contingent nature of self-reporting practice, and users' reflections on privacy and security concerns. Participants' positive responses to the UEQ and high levels of engagement with a CA despite technical limitations furthermore serve as initial evidence of users' willingness to accept the use of these systems for the self-report of mental health and well-being in the home context. Based on these findings, we discuss implications for the design of CAs for mental health and well-being with respect to the challenges surfaced by this work including challenges of conversational pattern matching, filling unmet interpersonal gaps, and the use of self-report CAs in the at-home social context.

REFERENCES

[1] Patricia A. Areán, Kien Hoa Ly, and Gerhard Andersson. 2016. Mobile technology for mental health assessment. *Dialog. Clin. Neurosci.* 18, 2 (2016), 163. https://doi.org/10.31887/DCNS.2016.18.2/parean

[2] Jakob E. Bardram, Mads Frost, Károly Szántó, Maria Faurholt-Jepsen, Maj Vinberg, and Lars Vedel Kessing. 2013. Designing mobile health technology for bipolar disorder: A field trial of the monarca system. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems.* 2627–2636.

[3] L. Frank Baum. 2008. *The Wonderful Wizard of Oz.* Oxford University Press.

[4] J. C. Bennett. 1991. The irrationality of the catharsis theory of aggression as justification for educators' support of interscholastic football. *Percept. Motor Skills* 72, 2 (1991), 415–418. https://doi.org/10.2466/pms.1991.72.2.415

[5] Kenneth Benoit, Kohei Watanabe, Haiyan Wang, Paul Nulty, Adam Obeng, Stefan Müller, and Akitaka Matsuo. 2018. quanteda: An R package for the quantitative analysis of textual data. *J. Open Source Softw.* 3, 30 (2018), 774. https://doi.org/10.21105/joss.00774

[6] Timothy Bickmore and Justine Cassell. 2005. Social dialongue with embodied conversational agents. In *Advances in Natural Multimodal Dialogue Systems.* Springer, 23–54. https://doi.org/10.1007/1-4020-3933-6_2

[7] Timothy Bickmore, Amanda Gruber, and Rosalind Picard. 2005. Establishing the computer–patient working alliance in automated health behavior change interventions. *Patient Educ. Counsel.* 59, 1 (2005), 21–30. https://doi.org/10.1016/j.pec.2004.09.008

[8] Timothy W. Bickmore, Suzanne E. Mitchell, Brian W. Jack, Michael K. Paasche-Orlow, Laura M. Pfeifer, and Julie O'Donnell. 2010. Response to a relational agent by hospital patients with depressive symptoms. *Interact. Comput.* 22, 4 (2010), 289–298. https://doi.org/10.1016/j.intcom.2009.12.001

[9] Ann E. Blandford. 2013. Semi-structured qualitative studies. Interaction Design Foundation. Retrieved Jan 30, 2020 from https://www.interaction-design.org/literature/book/the-encyclopedia-of-human-computer-interaction-2nd-ed/semi-structured-qualitative-studies.

[10] Virginia Braun and Victoria Clarke. 2006. Using thematic analysis in psychology. *Qual. Res. Psychol.* 3, 2 (2006), 77–101. DOI:10.1191/1478088706qp063oa

[11] Ava Mutchler Bret Kinsella. 2019. Voice Assistant Consumer Adoption in Healthcare. Retrieved Jan 30, 2020 from https://voicebot.ai/voice-assistant-consumer-adoption-report-for-healthcare-2019/.

[12] Kimberly J. Callicott and Hija Park. 2003. Effects of self-talk on academic engagement and academic responding. *Behav. Disord.* 29, 1 (2003), 48–64. https://doi.org/10.1177/019874290302900102

[13] Julia Cambre, Jessica Colnago, Jim Maddock, Janice Tsai, and Jofish Kaye. 2020. Choice of voices: A large-scale evaluation of text-to-speech voice quality for long-form content. In *Proceedings of the CHI Conference on Human Factors in Computing Systems (CHI'20)*. Association for Computing Machinery, New York, NY, 1–13. https://doi.org/10.1145/3313831.3376789

[14] Gillian Cameron, David Cameron, Gavin Megaw, Raymond Bond, Maurice Mulvenna, Siobhan O'Neill, Cherie Armour, and Michael McTear. 2018. Best practices for designing chatbots in mental healthcare—A case study on iHelpr. In *Proceedings of the 32nd International BCS Human Computer Interaction Conference (HCI'18)*. BCS Learning & Development Ltd., Swindon, GBR, Article 129, 1–5. https://doi.org/10.14236/ewic/HCI2018.129

[15] Yi Cheng, Kate Yen, Yeqi Chen, Sijin Chen, and Alexis Hiniker. 2018. Why doesn't it work? Voice-driven interfaces and young children's communication repair strategies. In *Proceedings of the 17th ACM Conference on Interaction Design and Children (IDC'18)*. Association for Computing Machinery, New York, NY, USA, 337–348. https://doi.org/10.1145/3202185.3202749

[16] Minji Cho, Sang-su Lee, and Kun-Pyo Lee. 2019. Once a kind friend is now a thing: Understanding how conversational agents at home are forgotten. In *Proceedings of the Designing Interactive Systems Conference (DIS'19)*. Association for Computing Machinery, New York, NY, USA, 1557–1569. https://doi.org/10.1145/3322276.3322332

[17] Dasom Choi, Daehyun Kwak, Minji Cho, and Sangsu Lee. 2020. "Nobody speaks that fast!" an empirical study of speech rate in conversational agents for people with vision impairments. In *Proceedings of the CHI Conference on Human Factors in Computing Systems (CHI'20)*. Association for Computing Machinery, New York, NY, 1–13. https://doi.org/10.1145/3313831.3376569

[18] Arlene E. Chung, Ashley C. Griffin, Dasha Selezneva, and David Gotz. 2018. Health and fitness apps for hands-free voice-activated assistants: Content analysis. *JMIR mHealth uHealth* 6, 9 (2018), e174. https://doi.org/10.2196/mhealth.9705

[19] Leigh Clark, Nadia Pantidi, Orla Cooney, Philip Doyle, Diego Garaialde, Justin Edwards, Brendan Spillane, Emer Gilmartin, Christine Murad, Cosmin Munteanu, Vincent Wade, and Benjamin R. Cowan. 2019. What Makes a Good Conversation? Challenges in Designing Truly Conversational Agents. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI'19)*. Association for Computing Machinery, New York, NY, USA, Paper 475, 1–12. https://doi.org/10.1145/3290605.3300705

[20] Eric Corbett and Astrid Weber. 2016. What can I say? Addressing user experience challenges of a mobile voice user interface for accessibility. In *Proceedings of the 18th International Conference on Human-Computer Interaction with Mobile Devices and Services*. Association for Computing Machinery, New York, NY, USA, 72–82. https://doi.org/10.1145/2935334.2935386

[21] Pim Cuijpers, Tara Donker, Annemieke van Straten, J. Li, and Gerhard Andersson. 2010. Is guided self-help as effective as face-to-face psychotherapy for depression and anxiety disorders? A systematic review and meta-analysis of comparative outcome studies. *Psychol. Med.* 40, 12 (2010), 1943–1957. https://doi.org/10.1017/S0033291710000772

[22] Kerstin Denecke, Sayan Vaaheesan, and Aaganya Arulnathan. 2021. A mental health chatbot for regulating emotions (SERMO)-concept and usability test. *IEEE Trans. Emerg. Top. Comput.* 9, 3 (2021), 1170–1182. DOI: 10.1109/TETC.2020.2974478

[23] David DeVault, Ron Artstein, Grace Benn, Teresa Dey, Ed Fast, Alesia Gainer, Kallirroi Georgila, Jon Gratch, Arno Hartholt, Margaux Lhommet, et al. 2014. SimSensei kiosk: A virtual human interviewer for healthcare decision support. In *Proceedings of the International Conference on Autonomous Agents and Multi-agent Systems*. 1061–1068. https://dl.acm.org/doi/10.5555/2615731.2617415

[24] Kevin Doherty, Andreas Balaskas, and Gavin Doherty. 2020. The design of ecological momentary assessment technologies. *Interact. Comput.* 32, 3 (2020), 257–278. https://doi.org/10.1093/iwcomp/iwaa019

[25] Kevin Doherty and Gavin Doherty. 2018. Engagement in HCI: Conception, theory and measurement. *ACM Comput. Surv.* 51, 5 (2018), 1–39. https://doi.org/10.1145/3234149

[26] Kevin Doherty, José Marcano-Belisario, Martin Cohn, Nikolaos Mastellos, Cecily Morrison, Josip Car, and Gavin Doherty. 2019. Engagement with mental health screening on mobile devices: Results from an antenatal feasibility study. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI'19)*. Association for Computing Machinery, New York, NY, USA, Paper 186, 1–15. https://doi.org/10.1145/3290605.3300416

[27] Mateusz Dubiel, Martin Halvey, Pilar Oplustil Gallegos, and Simon King. 2020. Persuasive synthetic speech: Voice perception and user behaviour. In *Proceedings of the 2nd Conference on Conversational User Interfaces (CUI'20)*. Association for Computing Machinery, New York, NY, Article 6, 9 pages. https://doi.org/10.1145/3405755.3406120

[28] Orla Duffy, Jonathan Synnott, Roisin McNaney, Paola Brito Zambrano, and W. George Kernohan. 2021. Attitudes toward the use of voice-assisted technologies among people with parkinson disease: Findings from a web-based survey. *JMIR Rehabil. Assist. Technol.* 8, 1 (2021), e23006. https://doi.org/10.2196/23006

[29] Kathleen Kara Fitzpatrick, Alison Darcy, and Molly Vierhile. 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 Ment. Health* 4, 2 (2017), e19. https://doi.org/10.2196/mental.7785

[30] Anushay Furqan, Chelsea Myers, and Jichen Zhu. 2017. Learnability through adaptive discovery tools in voice user interfaces. In *Proceedings of the CHI Conference Extended Abstracts on Human Factors in Computing Systems (CHI EA'17)*. Association for Computing Machinery, New York, NY, 1617–1623. https://doi.org/10.1145/3027063.3053166

[31] Hannah Gaffney, Warren Mansell, and Sara Tai. 2019. Conversational agents in the treatment of mental health problems: Mixed-method systematic review. *JMIR Ment. Health* 6, 10 (2019), e14166. https://doi.org/10.2196/14166

[32] Gabriella M. Harari, Nicholas D. Lane, Rui Wang, Benjamin S. Crosier, Andrew T. Campbell, and Samuel D. Gosling. 2016. Using smartphones to collect behavioral data in psychological science: Opportunities, practical considerations, and challenges. *Perspect. Psychol. Sci.* 11, 6 (2016), 838–854. https://doi.org/10.1177/1745691616650285

[33] Richard Harrington, M. Rutter, E. Taylor, and L. Hersov. 1994. Affective disorders. In *A Clinician's Handbook of Child and Adolescent Psychiatry*, 110.

[34] Simon Hoermann, Kathryn L. McCabe, David N. Milne, and Rafael A. Calvo. 2017. Application of synchronous text-based dialogue systems in mental health interventions: Systematic review. *J. Med. Internet Res.* 19, 8 (2017), e267. https://doi.org/10.2196/jmir.7023

[35] Eva Hudlicka. 2013. Virtual training and coaching of health behavior: Example from mindfulness meditation training. *Patient Educ. Couns.* 92, 2 (2013), 160–166. https://doi.org/10.1016/j.pec.2013.05.007

[36] Becky Inkster, Shubhankar Sarda, and Vinod Subramanian. 2018. An empathy-driven, conversational artificial intelligence agent (Wysa) for digital mental well-being: Real-world data evaluation mixed-methods study. *JMIR mHealth uHealth* 6, 11 (2018), e12106. https://doi.org/10.2196/12106

[37] Philip C. Kendall. 2012. Guiding theory for therapy with children and adolescents. In *Child and adolescent therapy: Cognitive-behavioral procedures*, P. C. Kendall (Ed.). Guilford Press, 3–24.

[38] Junhan Kim, Sun Park, and Lionel Robert. 2019. Conversational Agents for Health and Wellbeing: Review and Future Agendas. In *Proceedings of the Identifying Challenges and Opportunities in Human–AI Collaboration in Healthcare at the 22th ACM Conference on Computer Supported Cooperative Work and Social Computing (CSCW'19)*. Austin, Texas.

[39] A. Baki Kocabalil, Liliana Laranjo, and Enrico Coiera. 2018. Measuring user experience in conversational interfaces: A comparison of six questionnaires. In *Proceedings of the 32nd International BCS Human Computer Interaction Conference (HCI'18)*. BCS Learning & Development Ltd., Swindon, GBR, Article 21, 1–12. https://doi.org/10.14236/ewic/HCI2018.21

[40] Ahmet Baki Kocaballi, Shlomo Berkovsky, Juan C. Quiroz, Liliana Laranjo, Huong Ly Tong, Dana Rezazadegan, Agustina Briatore, and Enrico Coiera. 2019. The personalization of conversational agents in health care: Systematic review. *J. Med. Internet Res.* 21, 11 (2019), e15360. https://www.jmir.org/2019/11/e15360.

[41] A. Baki Kocaballi, Juan C. Quiroz, Liliana Laranjo, Dana Rezazadegan, Rafal Kocielnik, Leigh Clark, Q. Vera Liao, Sun Young Park, Robert J. Moore, and Adam Miner. 2020. Conversational agents for health and wellbeing. In *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems*. 1–8.

[42] Rafal Kocielnik, Daniel Avrahami, Jennifer Marlow, Di Lu, and Gary Hsieh. 2018. Designing for workplace reflection: A chat and voice-based conversational agent. In *Proceedings of the Designing Interactive Systems Conference (DIS'18)*. Association for Computing Machinery, New York, NY, 881–894. https://doi.org/10.1145/3196709.3196784

[43] Raina Langevin, Ross J. Lordon, Thi Avrahami, Benjamin R. Cowan, Tad Hirsch, and Gary Hsieh. 2021. Heuristic evaluation of conversational agents. In *Proceedings of the CHI Conference on Human Factors in Computing Systems (CHI'21)*. Association for Computing Machinery, New York, NY, Article 632, 15 pages. https://doi.org/10.1145/3411764.3445312

[44] Liliana Laranjo, Adam G. Dunn, Huong Ly Tong, Ahmet Baki Kocaballi, Jessica Chen, Rabia Bashir, Didi Surian, Blanca Gallego, Farah Magrabi, A. Lau, and E. Coiera. 2018. Conversational agents in healthcare: A systematic review. *J. Am. Med. Inf. Assoc.* 25, 9 (2018), 1248–1258. https://doi.org/10.1093/jamia/ocy072

[45] Josephine Lau, Benjamin Zimmerman, and Florian Schaub. 2018. Alexa, are you listening? Privacy perceptions, concerns and privacy-seeking behaviors with smart speakers. *Proc. ACM Hum.-Comput. Interact.* 2, CSCW, Article 102 (November 2018), 31 pages. https://doi.org/10.1145/3274371

[46] Bettina Laugwitz, Theo Held, and Martin Schrepp. 2008. Construction and evaluation of a user experience questionnaire. In *Symposium of the Austrian HCI and Usability Engineering Group*. Springer, 63–76.

[47] C. Leslie. 2008. Boxing is the best way to stop violence in kids. *The Guardian Politics Blog*. https://www.theguardian.com/politics/blog/2008/apr/14/boxingisthebestwaytosto.

[48] Irene Lopatovska and Harriet Williams. 2018. Personification of the Amazon Alexa: BFF or a mindless companion. In *Proceedings of the Conference on Human Information Interaction & Retrieval*. Association for Computing Machinery, New York, NY, USA, 265–268. https://doi.org/10.1145/3176349.3176868

[49] Christopher A. Lovejoy. 2019. Technology and mental health: The role of artificial intelligence. *Eur. Psychiatr.* 55 (2019), 1–3. DOI:10.1016/j.eurpsy.2018.08.004

[50] Gale M. Lucas, Jonathan Gratch, Aisha King, and Louis-Philippe Morency. 2014. It's only a computer: Virtual humans increase willingness to disclose. *Comput. Hum. Behav.* 37 (2014), 94–100. https://doi.org/10.1016/j.chb.2014.04.043

[51] Gale M. Lucas, Albert Rizzo, Jonathan Gratch, Stefan Scherer, Giota Stratou, Jill Boberg, and Louis-Philippe Morency. 2017. Reporting mental health symptoms: Breaking down barriers to care with virtual human interviewers. *Front. Robot. AI* 4 (2017), 51. https://doi.org/10.3389/frobt.2017.00051

[52] Ewa Luger and Abigail Sellen. 2016. "Like having a really bad PA": The gulf between user expectation and experience of conversational agents. In *Proceedings of the CHI Conference on Human Factors in Computing Systems (CHI'16)*. Association for Computing Machinery, New York, NY, 5286–5297. https://doi.org/10.1145/2858036.2858288

[53] Adam S. Miner, Albert Haque, Jason A. Fries, Scott L. Fleming, Denise E. Wilfley, G. Terence Wilson, Arnold Milstein, Dan Jurafsky, Bruce A. Arnow, W. Stewart Agras, et al. 2020. Assessing the accuracy of automatic speech recognition for psychotherapy. *npj Digit. Med.* 3, 1 (2020), 1–8. https://doi.org/10.1038/s41746-020-0285-8

[54] Adam S. Miner, Arnold Milstein, Stephen Schueller, Roshini Hegde, Christina Mangurian, and Eleni Linos. 2016. Smartphone-based conversational agents and responses to questions about mental health, interpersonal violence, and physical health. *JAMA Internal Med.* 176, 5 (2016), 619–625. https://doi.org/10.1001/jamainternmed.2016.0400

[55] Roger K. Moore. 2013. Spoken language processing: Where do we go from here? In *Your Virtual Butler*, R. Trappl (Eds.). Lecture Notes in Computer Science, vol. 7407. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-37346-6_10

[56] Nasim Motalebi and Saeed Abdullah. 2018. Conversational agents to provide couple therapy for patients with PTSD. In *Proceedings of the 12th EAI International Conference on Pervasive Computing Technologies for Healthcare (PervasiveHealth'18)*. Association for Computing Machinery, New York, NY, 347–351. https://doi.org/10.1145/3240925.3240933

[57] Christine Murad, Cosmin Munteanu, Benjamin R. Cowan, and Leigh Clark. 2019. Revolution or evolution? Speech interaction and HCI design guidelines. *IEEE Perv. Comput.* 18, 2 (2019), 33–45. DOI : 10.1109/MPRV.2019.2906991

[58] Chelsea Myers, Anushay Furqan, Jessica Nebolsky, Karina Caro, and Jichen Zhu. 2018. Patterns for how users overcome obstacles in voice user interfaces. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, Paper 6, 1–7. https://doi.org/10.1145/3173574.3173580

[59] Clifford Nass, Jonathan Steuer, Ellen Tauber, and Heidi Reeder. 1993. Anthropomorphism, agency, and ethopoeia: Computers as social actors. In *INTERACT'93 and CHI'93 Conference Companion on Human Factors in Computing Systems (CHI'93)*. Association for Computing Machinery, New York, NY, USA, 111–112. https://doi.org/10.1145/259964.260137

[60] The Danish National Committee on Health Research Ethics. 2018. Act on Research Ethics Review of Health Research Projects. Retrieved May 7, 2021 from https://en.nvk.dk/rules-and-guidelines/act-on-research-ethics-review-of-health-research-projects.

[61] World Health Organization. 2001. The world health report 2001: Mental health: New understanding, new hope.

[62] Hannah R. M. Pelikan and Mathias Broth. 2016. Why that nao? How humans adapt to a conventional humanoid robot in taking turns-at-talk. In *Proceedings of the CHI Conference on Human Factors in Computing Systems (CHI'16)*. Association for Computing Machinery, New York, NY, 4921–4932. https://doi.org/10.1145/2858036.2858478

[63] Pierre Philip, Jean-Arthur Micoulaud-Franchi, Patricia Sagaspe, Etienne De Sevin, Jérôme Olive, Stéphanie Bioulac, and Alain Sauteraud. 2017. Virtual human as a new diagnostic tool, a proof of concept study in the field of major depressive disorders. *Scientific Reports* 7 (2017), 42656. https://doi.org/10.1038/srep42656

[64] Martin Porcheron, Joel E. Fischer, Stuart Reeves, and Sarah Sharples. 2018. Voice Interfaces in Everyday Life. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI'18)*. Association for Computing Machinery, New York, NY, USA, Paper 640, 1–12. https://doi.org/10.1145/3173574.3174214

[65] Alisha Pradhan, Kanika Mehta, and Leah Findlater. 2018. "Accessibility came by accident": Use of voice-controlled intelligent personal assistants by people with disabilities. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. 1–13. https://doi.org/10.1145/3173574.3174033

[66] Simon Provoost, Ho Ming Lau, Jeroen Ruwaard, and Heleen Riper. 2017. Embodied conversational agents in clinical psychology: A scoping review. *J. Med. Internet Res.* 19, 5 (2017), e151. https://doi.org/10.2196/jmir.6553

[67] Amanda Purington, Jessie G. Taft, Shruti Sannon, Natalya N. Bazarova, and Samuel Hardman Taylor. 2017. "Alexa is my new BFF": Social roles, user satisfaction, and personification of the amazon echo. In *Proceedings of the CHI Conference Extended Abstracts on Human Factors in Computing Systems*. 2853–2859. https://doi.org/10.1145/3027063.3053246

[68] Aung Pyae and Tapani N. Joelsson. 2018. Investigating the usability and user experiences of voice user interface: A case of Google home smart speaker. In *Proceedings of the 20th International Conference on Human-Computer Interaction with Mobile Devices and Services Adjunct*. 127–131. https://doi.org/10.1145/3236112.3236130

[69] Juan C. Quiroz, Tristan Bongolan, and Kiran Ijaz. 2020. Alexa depression and anxiety self-tests: A preliminary analysis of user experience and trust. In *Adjunct Proceedings of the ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the ACM International Symposium on Wearable Computers*. 494–496. https://doi.org/10.1145/3410530.3414374

[70] Byron Reeves and Clifford Ivar Nass. 1996. *The Media Equation: How People Treat Computers, Television, and New Media Like Real People and Places.* Cambridge University Press, Cambridge, UK.

[71] Edison Research. 2020. The Smart Audio Report Winter 2019 from npr and Edison Research. Retrieved Jan 30, 2020 from https://www.edisonresearch.com/the-smart-audio-report-winter-2019-from-npr-and-edison-research/.

[72] Lazlo Ring, Lin Shi, Kathleen Totzke, and Timothy Bickmore. 2015. Social support agents for older adults: Longitudinal affective computing in the home. *J. Multimodal User Interfaces* 9, 1 (2015), 79–88. https://doi.org/10.1007/s12193-014-0157-0

[73] Darius A. Rohani, Andrea Quemada Lopategui, Nanna Tuxen, Maria Faurholt-Jepsen, Lars V. Kessing, and Jakob E. Bardram. 2020. MUBS: A personalized recommender system for behavioral activation in mental health. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. 1–13. https://doi.org/10.1145/3313831.3376879

[74] Joel Sebastian and Deborah Richards. 2017. Changing stigmatizing attitudes to mental health via education and contact with embodied conversational agents. *Comput. Hum. Behav.* 73 (2017), 479–488. https://doi.org/10.1016/j.chb.2017.03.071

[75] Ji Youn Shin and Jina Huh-Yoo. 2018. Designing everyday conversational agents for managing health and wellness.

[76] Ben Shneiderman. 2000. The limits of speech recognition. *Commun. ACM* 43, 9 (September 2000), 63–65. https://doi.org/10.1145/348941.348990

[77] Bernhard Suhm. 2003. Towards best practices for speech user interface design. In *Proceedings of the 8th European Conference on Speech Communication and Technology*.

[78] Anja Thieme, Jayne Wallace, Thomas D. Meyer, and Patrick Olivier. 2015. Designing for mental wellbeing: Towards a more holistic approach in the treatment and prevention of mental illness. In *Proceedings of the British HCI Conference*. 1–10. https://doi.org/10.1145/2783446.2783586

[79] Lucy Tindall, Antonina Mikocka-Walus, Dean McMillan, Barry Wright, Catherine Hewitt, and Samantha Gascoyne. 2017. Is behavioural activation effective in the treatment of depression in young people? A systematic review and meta-analysis. *Psychol. Psychother.: Theory Res. Pract.* 90, 4 (2017), 770–796. https://doi.org/10.1111/papt.12121

[80] Franca Tonnaer, Maaike Cima, and Arnoud Arntz. 2020. Explosive matters: Does venting anger reduce or increase aggression? Differences in anger venting effects in violent offenders. *J. Aggr. Maltreat. Trauma* 29, 5 (2020), 611–627. https://doi.org/10.1080/10926771.2019.1575303

[81] Christian Winther Topp, Søren Dinesen Østergaard, Susan Søndergaard, and Per Bech. 2015. The WHO-5 well-being index: A systematic review of the literature. *Psychother. Psychosom.* 84, 3 (2015), 167–176. https://doi.org/10.1159/000376585

[82] John Torous, Patrick Staples, Meghan Shanahan, Charlie Lin, Pamela Peck, Matcheri Keshavan, and Jukka-Pekka Onnela. 2015. Utilizing a personal smartphone custom app to assess the patient health questionnaire-9 (PHQ-9) depressive symptoms in patients with major depressive disorder. *JMIR Mental Health* 2, 1 (2015), e8. https://doi.org/10.2196/mental.3889

[83] Kimberly R. H. Treadwell and Philip C. Kendall. 1996. Self-talk in youth with anxiety disorders: States of mind, content specificity, and treatment outcome. *J. Consult. Clin. Psychol.* 64, 5 (1996), 941. https://doi.org/10.1037//0022-006x.64.5.941

[84] Aditya Nrusimha Vaidyam, Hannah Wisniewski, John David Halamka, Matcheri S. Kashavan, and John Blake Torous. 2019. Chatbots and conversational agents in mental health: A review of the psychiatric landscape. *Can. J. Psychiatr.* 64, 7 (2019), 456–464. https://doi.org/10.1177/0706743719828977

[85] Niels Van Berkel, Denzil Ferreira, and Vassilis Kostakos. 2017. The experience sampling method on mobile devices. *ACM Comput. Surv.* 50, 6 (2017), 1–40. https://doi.org/10.1145/3123988

[86] Jinping Wang, Hyun Yang, Ruosi Shao, Saeed Abdullah, and S. Shyam Sundar. 2020. Alexa as coach: Leveraging smart speakers to build social agents that reduce public speaking anxiety. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. 1–13. https://doi.org/10.1145/3313831.3376561

[87] Adam Waytz, John Cacioppo, and Nicholas Epley. 2010. Who sees human? The stability and importance of individual differences in anthropomorphism. *Perspect. Psychol. Sci.* 5, 3 (2010), 219–232. https://doi.org/10.1177/1745691610369336

[88] Zhuxiaona Wei and James A. Landay. 2018. Evaluating Speech-Based Smart Devices Using New Usability Heuristics. *IEEE Pervas. Comput.* 17, 2 (2018), 84–96. DOI : 10.1109/MPRV.2018.022511249

[89] Laura A. Weiss, Gerben J. Westerhof, and Ernst T. Bohlmeijer. 2016. Can we increase psychological well-being? The effects of interventions on psychological well-being: A meta-analysis of randomized controlled trials. *PLoS ONE* 11, 6 (2016), e0158092. https://doi.org/10.1371/journal.pone.0158092

[90] Joseph Weizenbaum. 1966. ELIZA-a computer program for the study of natural language communication between man and machine. *Commun. ACM* 9, 1 (January 1966), 36–45. https://doi.org/10.1145/365153.365168

[91] Cheng K. Fred Wen, Stefan Schneider, Arthur A. Stone, and Donna Spruijt-Metz. 2017. Compliance with mobile ecological momentary assessment protocols in children and adolescents: A systematic review and meta-analysis. *J. Med. Internet Res.* 19, 4 (2017), e132. https://doi.org/10.2196/jmir.6641