Conversational Agents for Mental Health Self-reports as Experienced by People Living with Mental Illness

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

Self-reports obtained by health questionnaire and diary log are considered effective methods of monitoring and assessing affective disorders (ADs) including depression and bipolar disorder. While these methods are long-established, they represent only one, and potentially limited, means of insight into individuals' experiences. Recent advancements in speech recognition technology and the growing popularity of smart speakers have the potential to enable new opportunities for collecting self-reported health and wellbeing data. Conversational agents (CAs), in particular, may allow for more natural and engaging means of mental health data collection. To date, however, few real-world studies have examined users' experiences of engaging in such practices, nor the impact of critical design choices. With the aim of understanding CAs' potential to support the practice of self-report, we conducted a four-week 'in the wild' study during which 20 individuals with AD used a CA named 'Sofia' to maintain a daily diary log while also responding to the WHO-5 questionnaire every 2 weeks. Findings from a thematic analysis of post-study interviews highlight actions taken by participants to overcome CA limitations, diverse personifications of the agent and acceptance among users' social circle; suggesting the potential of CAs to support mental health and wellbeing despite technical limitations. Participants' social context and privacy challenges, including eavesdropping and data security concerns, were identified as the primary barriers to engaging and effective CA self-report experiences. Based on these insights, we discuss implications for the future design of CAs to support mental health and wellbeing.

CCS Concepts: • **Human-centered computing** → **Personal digital assistants**; **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

ACM Reference Format:
Raju Maharjan, Darius Adam Rohani, Kevin Doherty, Per Bækgaard, and Jakob E. Bardram. 2020. Conversational Agents for Mental Health Self-reports as Experienced by People Living with Mental Illness. *Proc. ACM Meas. Anal. Comput. Syst.* 1, 1, Article 111 (January 2020), 26 pages. https://doi.org/10.1145/1122445.1122456

Authors' addresses: Raju Maharjan, rajm@dtu.dk, Department of Health Technology, Technical University of Denmark, Ørsted Plads, Kgs. Lyngby, DK-2800, Denmark; Darius Adam Rohani, daroh@dtu.dk, Department of Health Technology, Technical University of Denmark, Ørsted Plads, Kgs. Lyngby, DK-2800, Denmark; Kevin Doherty, kevdoh@dtu.dk, Department of Health Technology, Technical University of Denmark, Ørsted Plads, Kgs. Lyngby, DK-2800, Denmark; Per Bækgaard, pgba@dtu.dk, Department of Applied Mathematics and Computer Science, Technical University of Denmark, Richard Pedersens Plads, Kgs. Lyngby, DK-2800, Denmark; Jakob E. Bardram, jakba@dtu.dk, Department of Health Technology, Technical University of Denmark, Ørsted Plads, Kgs. Lyngby, DK-2800, Denmark.

2476-1249/2020/1-ART111 $15.00
https://doi.org/10.1145/1122445.1122456

1 INTRODUCTION

Much of what we know about mental health and wellbeing is gathered through self-reports drawn from patient diaries and validated health questionnaires [1, 17, 18]. 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 towards technology to facilitate such data collection methods; most often consisting of the straightforward 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) apps, have drawn attention to the complexity of the self-report of mental health and wellbeing as a process in itself [3, 19, 61]. 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, often limit users' capacity for self-expression, can pose challenges of data validity and reliability, and often place a burden on users [17, 24, 69].

Conversational Agents (CAs), in comparison, may offer the opportunity to obtain richer insight into the experience of mental health and illness – whether for sharing with health professionals, to support individuals' own insight, or perhaps even to serve as part of health and wellbeing interventions. Studies have shown that speech, as our primary mode of communication, can engage users in more natural and human-human-like conversations, thereby 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 [16, 38, 47]

CAs are becoming increasingly feasible as modes of interaction – a trend supported by increased applications of these systems in healthcare following advancements in speech recognition technology, as well as the growing popularity of smart speaker devices [27, 29, 30, 32, 54]. According to one recent survey, 52.0% of 1,004 U.S. adults possess an interest in the use of CAs, while 7.5% have already made use of such CAs for a healthcare-related task or inquiry such as inquiring about symptoms of illness (73.0%), searching for information concerning medication use (45.9%), or seeking care and treatment options (37.7%) [8].

Despite growing accessibility and the increased likelihood of CAs's continued adoption in healthcare, there is no denying that technical limitations and challenges to the use of these systems for gathering self-reported data remain. Limitations identified by researchers include CAs misinterpreting or not recognizing user utterances, causing confusion in the conversation and leading to further errors [49, 56, 65]; systems' inability to engage in dynamic conversations [12]; the requirement to specify in advance what users can or cannot say to the CAs [15, 22, 53, 53]; and CAs' monotonous, robotic and unnatural tone of voice with implications for users' engagement [10, 13, 20, 44].

Hence, there is a need for HCI research to understand present possibilities as well as the choices available to designers in overcoming these limitations. Much recent HCI research has focused on issues of privacy [33], usability and user satisfaction [12], accessibility [55], and attitudes towards [37] CAs. Others have focused on questions of usability, user experience, trust, feedback, self-reflection and learning during the collection of self-reports of mental and physical health and wellbeing [31, 57, 70] as well as social engagement [52, 60, 62]. HCI researchers have also proposed initial guidelines for the design [48, 65, 72] and evaluation of CAs [28].

Research in healthcare has investigated CAs' effectiveness in diagnostic performance, health symptoms disclosure, health intervention, adherence to self-management practices and issues in CA dialog management [30, Table 3]. Although these technologies are currently insufficiently sophisticated to fully support dynamic human-to-human-like conversational interactions, other healthcare researchers have shown increased interest in the anthropomorphic characteristics of CAs

as a potential means of establishing and maintaining a therapeutic alliance that could encourage users to disclose deeper insights into their health and wellbeing. [11, 16, 27, 38, 39].

While initial research and the growing popularity of smart-speaker systems 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. 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, or how to design for positive CA self-report experiences.

To address these gaps in knowledge, we designed and conducted a study for the purpose of understanding users' real-world CA self-report experiences as well as identifying barriers and enablers to the design of engaging, ethical and effective CAs interactions – addressing the following research questions;

RQ1 How do people living with affective disorders (ADs) perceive the use of a Conversational Agent (CA) for the self-report of mental health and wellbeing?
RQ2 What factors – barriers and enablers – shape their experiences of mental health self-reports via CA?
RQ3 Which CA design choices and strategies do this user group see as most important to support long term conversational self-reports?

To answer these questions, we conducted a four-week 'in the wild' study during which participants kept a daily open-ended conversational diary log of their mental health and emotional wellbeing, and responded to a World Health Organization-Five Well-Being Index (WHO-5) questionnaire fortnightly via a CA named *Sofia* deployed on a Google Nest device.

Participants' positive responses to the User Experience Questionnaire (UEQ) questionnaire and strong engagement with the study despite technical limitations suggest CA's potential for mental health and wellbeing. Thematic analysis of participants' comments provided during post-study interviews further strengthens these findings, as reflected in strategies adopted by participants to overcome technical limitations, diverse personified perceptions of the agent, acceptance among users' social circles and a positive impact on their personal relationships. Users' social context and privacy challenges including eavesdropping and data security concerns are identified as the primary barriers to CA self-reports. From their four-week experience of self-reporting via CA, participants suggested several design choices to improve agents' conversational skills as well as features important for long term CA self-reports.

We discuss implications for the design of CAs for mental health, reflecting on enablers and barriers to engaging conversational design, potential ethical ramifications of CA personification and the value of a relationship-oriented design framing to support sustainable CA-user relationships.

This work contributes towards; (i) An understanding of factors impacting the self-report experiences and behaviors of people with AD, and (ii) Considerations for the future design of engaging and ethical CA self-reports of mental health and wellbeing.

2 RELATED WORK

This study builds upon and contributes to research in the domains of CAs for health and wellbeing as well as speech-based self-report mechanisms.

2.1 Designing Conversational Agents for Health & Wellbeing

Although in many ways CAs remain emergent technologies, many HCI and healthcare researchers have already begun to explore their potential to support health and wellbeing in a variety of

healthcare contexts [23, 32]. Such systems include Interactive Voice Response (IVR), Automatic Speech Recognition (ASR), embodied systems, and chatbot technologies.

IVR and ASR based systems are some of the earliest examples of speech-enabled self-report technologies, and have been successfully deployed in a number of real-world contexts. Azzini et al. [2] for example, developed a telephone-based dialog system to enable hypertensive patients to record their symptoms by calling a toll-free number, thereby mitigating the need for a clinical visit. Levin and Levin [36] evaluated the usability of an ASR-based 'Pain Monitoring Voice Diary' system, finding that users were able to navigate the flexible interface, and that self-reporting efficiency increased with users' experience, both in terms of session duration and avoidance of troublesome dialog scenarios.

Others have explored applications of such technologies to support mental health. Hudlicka [25] developed a multi-modal embodied system named 'Virtual Mindfulness Coach' and evaluated its effectiveness in training and coaching mindfulness meditation. Results from this study demonstrated that students achieved a higher sense of self-efficacy concerning the establishment of regular mindfulness practice through coach-based training than via self-administered training using written and audio materials. Fitzpatrick et al. [21] developed a fully automated chatbot named 'Woebot' and studied its feasibility, acceptability, and efficacy in delivering a self-help Cognitive Behavioral Therapy (CBT) program for college students with symptoms of anxiety and depression. Findings of this study showed that delivery of CBT through the chatbot significantly reduced participants' symptoms of depression and anxiety. The authors suggest that process factors were more influential on participants' acceptance of the system than content factors, mirroring traditional therapy.

Other researchers have focused on the social dimension of CA technologies. Miner et al. [42], for example, compared sentiment tendencies and mirroring behaviors in human-human and human-CA dialogues in mental health settings by employing Relational Frame Theory. They found that while human sentiment-related interaction norms persist in human-agent dialogues, inhibition towards use of obscenity was greatly reduced. Other studies have demonstrated that interaction between human and CA follows common social rules even though users are aware their interaction partner is a machine [50, 58]. It has been proposed that this human-to-human like natural conversational interaction enables users to form a relationship with CAs [5, 71], and in turn that CAs could potentially collect more accurate, honest and insightful self-reports.

Research therefore suggests that CAs may prove able to address such challenges in mental health assessment and treatment as geographic barriers to care, a lack of efficient and asynchronous communication, continuous availability, and stigma around mental health [42]. And yet realizing this potential also hings upon the design of appropriate conversations; a far from simple-task.

2.2 Understanding Speech-based 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 wellbeing. And, although these methods have in many cases proved effective, there is growing evidence of a need to consider alternative forms of self-report. This has led researchers to focus on the development of speech-based methods of self-report. One such research thread entails work in the healthcare domain focused on speech-enabled CAs as a means of eliciting self-reports of psychological symptoms from patients. One such study examined the impact of a virtual human interviewer on the disclosure of Post-traumatic Stress Disorder (PTSD) symptoms among active-duty service members and investigated the factors influencing the resulting self-reporting practices [39]. Results suggested that virtual human interviewers could increase patients' disclosure of mental health symptoms due to anonymity, reduced-stigma, and CA's ability to build rapport with users. DeVault et al. [16] likewise designed a virtual human interviewer to assess mental health conditions including depression and anxiety via automated analysis of

verbal and non-verbal behaviors. Results 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.

Other researchers have focused on the adaptation of commercially-available systems including Google Home and Amazon Alexa devices, which have built on recent advancements in Natural Language Processing (NLP) and Virtual Assistants (VAs) interfaces to provide more accessible and conversational in-home experiences. Such research includes the design and use of custom agents (Google Actions, Alexa Skills) to investigate speech-enabled CAs' potential to support self-reports of health and well-being 'in-the-wild'. Kocielnik et al. [31] for instance, designed a CA named 'Robota' Amazon Alexa Skill to support work activity journaling; by asking users to provide ten open-ended daily reflections. The authors examined how speech-enabled interaction affected workers' reflections and self-learning in comparison to a chat-bot employing the same questionnaire. The results from a three-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. [57] developed an Alexa Skill that enabled users to better express their emotions, complete self-assessments for depression and anxiety, and receive suggestions to improve their current state-of-mind. While the pilot study results showed that participants were willing to engage and trusted the agent for sharing personal information such as depression and anxiety scores, their user experience scores for the agent demonstrated that participants considered it to lack efficiency and novelty. Motalebi and Abdullah [46] conducted a study to identify the barriers in implementing clinical therapy for patients with PTSD via Amazon Alexa; emphasizing the importance of short dialogues and interactivity for effective therapeutic content delivery.

2.2.1 CAs & Their Limitations. While speech-enabled CAs are gaining widespread adoption in daily use [8, 14, 59], we cannot speak about the potential of CAs to support practices of self-report without mentioning the present limitations of these technologies, which are often significant in nature [45, 64, 65].

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% [43]. While the authors 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 [49, 56, 65].

Secondly, while some CAs can mimic human-to-human-like conversations, none are fully capable of engaging in dynamic conversation [12]. Typically, users have to know in advance what they can or cannot say to the CA [15, 22, 53] which often makes it cognitively demanding for users to interact [53]. Moreover, CAs allow a maximum of 12 seconds for each user's response, 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 [44]. And finally, studies have shown that users often establish high expectations for CAs [40], anticipating emotional exchanges, relationship building, and human-human like conversations [12], which often leads to a drastic decline in use when the CAs cannot meet users' expectations [12, 40].

Several researchers have provided initial guidelines to address these limitations, and support the design of effective CAs [48, 65, 72], including for health and wellbeing management [63]. Studies 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 [49, 51].

Yet, we know less about how people living with mental illness including ADs might engage with CAs to support the self-report of their emotional states, and what CA characteristics they view as important for long term self-reporting practice. Despite the current limitations of CA technologies, a present lack of knowledge concerning users' experiences, and ethical and design challenges, it is also possible that individuals experiencing mental illness may have the most to benefit from speech-enabled self-report technologies. We design and conduct an 'in the wild' field study to explore these possibilities for the first time.

3 METHOD

A four-week 'in the wild' study was designed to enable people living with ADs to experience the self-report of mental health and wellbeing via CA. Our aims in conducting this study were to generate an understanding of users' perspectives and experiences with respect to such use of CAs and to identify essential factors for the design of positive self-report experiences.

3.1 Participant Recruitment

Participants who self-identified as diagnosed with an affective disorder (AD) (unipolar or bipolar disorder) were recruited via online platforms including national patient recruitment sites[1], social media (e.g., Facebook, Twitter), university internal email, and poster. Participants were offered a Google Nest Mini 2 device or a gift card corresponding to DKK 300 (~US$50) for their participation.

The study was exempted from ethical approval by the Danish National Committee on Health Research Ethics in accordance with Danish research standards[2]. The highest research ethics standards were adhered to at each step of this process including during the recruitment and engagement of participants.

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[3] and deployed it via Google Nest Mini[4] smart speakers.

Sofia combines an open-ended dialog design emulating a human-to-human like conversation allowing users to freely express their mental state, with additional completion of the WHO-5 health questionnaire. The WHO-5 scale is an established reporting mechanism requiring respondents to summarize their wellbeing by responding to five short questions using a Likert scale. The combination of these two distinct forms of self-report would also enable us to understand users' perceptions of different variants of conversational self-report.

Table 1 provides an overview of the dialog management as implemented in *Sofia* including key intents, their purpose and example dialogues. Each self-report session started with *Sofia* greeting the user and asking how they were feeling. *Sofia* then took the user through two additional open-ended questions designed to provide further insight into their mental state. Each question had multiple

[1] http://www.forsoegsperson.dk/
[2] Section 14 (2) at https://en.nvk.dk/rules-and-guidelines/act-on-research-ethics-review-of-health-research-projects
[3] https://dialogflow.com
[4] https://store.google.com/product/google_nest_mini

Table 1. Dialog management in *Sofia*: Key intents, 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 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!

variations in order to reduce the otherwise repetitive nature of the conversation. Additionally, every two weeks, *Sofia* asked the users to respond to the WHO-5 questionnaire, following the daily open-ended conversation. *Sofia* stated the preamble, and then asked the WHO-5 questions in random order. In order to render the 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 [67].

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 "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 patient for sharing their information and bids them farewell.

3.3 Study Procedure

The study procedure consisted of 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 motivation and expectations of participants. Each eligible participant was asked to sign a consent form and to provide demographic information including name, email, age, gender, education level, employment, the period during which AD was diagnosed, symptoms last experienced, other known health conditions, technical ability, and prior experience using CA.

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 walked through and provided a handout describing CA's limitations (section 2.2.1), instructions regarding how to interact with *Sofia* (e.g., invocation phrases, repeating questions, ending the conversation) and what *Sofia* can and can not do. Participants were encouraged to use Google Assistant's 'routine'[5] feature or set up a reminder by asking the assistant to remind them to speak to *Sofia* everyday.

3.3.2 In-study. During the four-week study, each participant would speak to *Sofia* regarding their mental state by answering a series of questions every day, and responding to the WHO-5 questionnaire fortnightly. 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. At the end of the study, participants filled out the UEQ [34] questionnaire, yielding a subjective assessment of *Sofia*'s usability as well as their individual experience. The UEQ is a 26-item scale covering six key user experience dimensions; Attractiveness, Perspicuity, Efficiency, Dependability, Stimulation, and Novelty [34].

Finally, each participant took part in a semi-structured interview regarding their experience of *Sofia* (app), their overarching views of the technology and thoughts on the future design of speech-based conversational self-report agents. Interviews lasted between 25 to 65 minutes and were audio-recorded. Due to COVID-19 guidelines including 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 time, one participant (P17) was interviewed via a series of emails.

3.4 Analysis

Following the principles of Braun & Clarke's thematic analysis [6, 7], 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. Finally, participants' responses to the UEQ questionnaire as well as the timestamps of each user's interactions were analyzed using R (v. 3.6.2).

[5] https://support.google.com/googlenest/answer/7029585?co=GENIE.Platform%3DAndroid&hl=en

Table 2. Participant demographics. CA exp. = Prior experience with CAs, PMDD = Premenstrual Dysphoric Disorder, OCD = Obsessive Compulsive Disorder, IDR = I do not remember

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

4 RESULTS

We next present our results in the primary form of qualitative analyses 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 bi-weekly self-reporting of mental health and wellbeing 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 features considered important for engaging and sustainable self-report experiences.

First however, we provide some brief insight into participants' demographics and their engagement with the technology during this 4-week period; as context for our qualitative findings.

4.1 First, Some Context

4.1.1 Demographics. Participants (F=16, M=4) were adults with a mean age of 27.5 from various backgrounds living in different regions of Denmark. 45% of them reported that they had other health conditions and 25% reported that they had no prior experience using CAs. Table 2 provides a summary of participants' demographic characteristics.

4.1.2 Participant Engagement. Analysis of the log timestamps collected from *Sofia* shows that the average participant's rate of engagement in self-report (calculated as the percentage of requested reports completed) was 75%. This is comparable to rates of compliance reported in prior EMA studies using mobile devices [73]. Figure 1 shows a temporal distribution of the usage pattern of each participant over the course of the study.

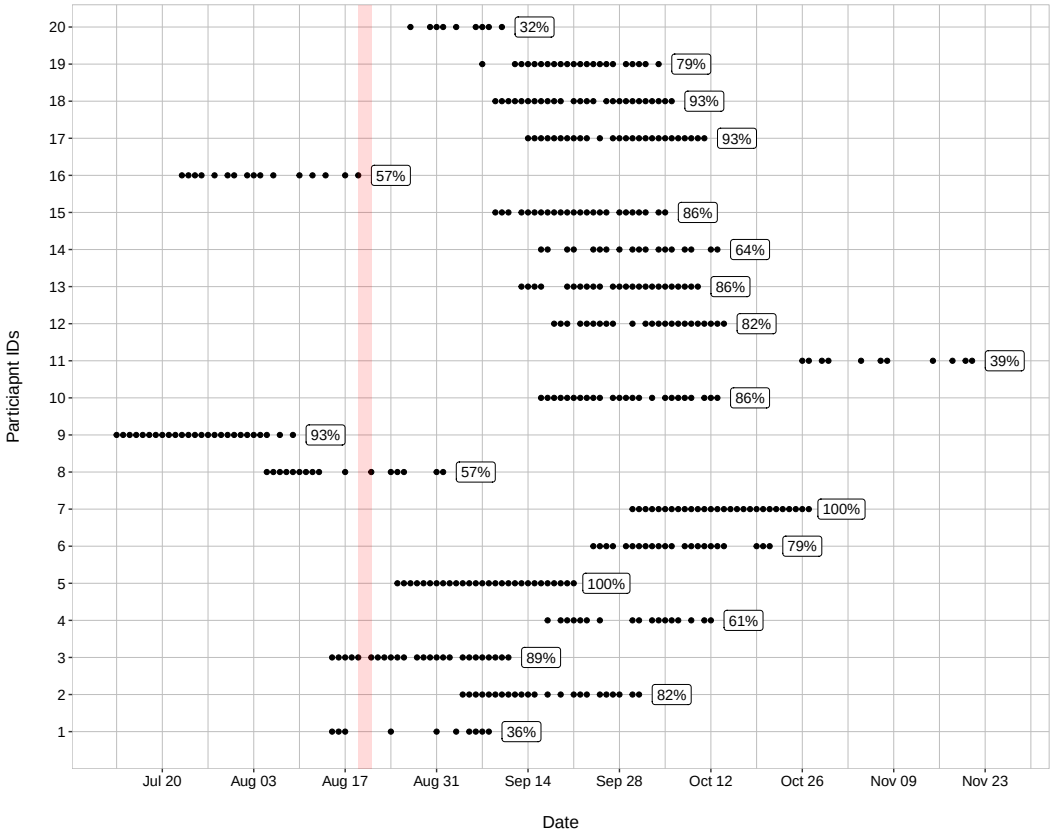

Fig. 1. Participants' engagement with daily self-reports of mental health and wellbeing via *Sofia*. Each dot represents an entry for the day. Annotations reflect each participant's rate of engagement during the study. The red bar indicates Dialogflow's service outage on August 20th, 2020. Participants were not able to use *Sofia* on that day.

4.1.3 Participant Perceived Experience. Participants' responses to the UEQ questionnaire upon conclusion of the study also provide insight into their experience of use. Figure 2 reveals participants' impressions of their self-report experience in terms of ratings of Attractiveness ($U = 1.09, SE = 0.44$), Dependability ($U = 0.74, SE = 0.37$), Efficiency ($U = 0.98, SE = 0.34$), Novelty ($U = 0.71, SE = 0.38$), Perspicuity ($U = 2.09, SE = 0.33$), and Stimulation ($U = 0.38, SE = 0.36$). These results show that participants held positive perceptions of the CA in terms of perspicuity, attractiveness and efficiency, and neutral views with respect to dependability, novelty and stimulation factors. Combined, these results suggest that participants held positive views of the use of CAs for self-report although they

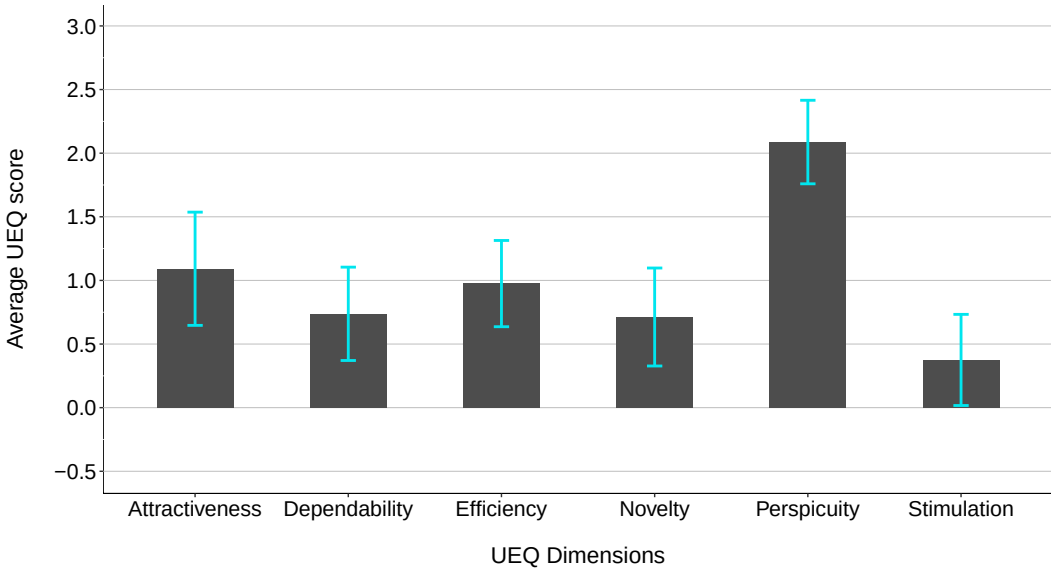

Fig. 2. Participant's mean UEQ score in each dimension with error bars reflecting margin of error. Values between −0.8 and 0.8 represent a more or less neutral evaluation of the corresponding scale, values > 0.8 represent a positive evaluation and values < −0.8 represent a negative evaluation.

were aware of the technology's technical limitations – findings further detailed in the results of our qualitative analysis.

4.2 Theme 1. Engaging With & Getting Around the Technology

The theme we first discuss is one which informed all others, and stems from users' frequent comments to the need to get around the technology in order 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 & 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-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 the prior literature 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 did adopt 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 it's starting phase - I let it do it's 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 loose 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 briefly;

> *"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 *"[i]t 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 concise diary entry which 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 that really quite quick"* [P7]. This willingness to adapt can only be interpreted in light of the value users associate with their interactions with the technology – which, 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. "Someone" (or Something?) "Who Listens"

Engaging in the self-report of mental health and wellbeing 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 towards 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 the *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 it 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 CA, 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 them 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, the *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, on the other hand, 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 the CA 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 ones 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'."*. Whereas, 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 [9, 26, 68]. As P14 noted:

> *"I talk in my head and not really out loud. And I think, 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 and they're all the time there and 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, 35, 66], 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 app 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 app about their emotions. P1's partner, for example, told P1 that app 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 they 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 the *Sofia* was 'cool' and expressed its potential to support mental health and wellbeing. 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 of 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 two 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' 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 her 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. We also have more details about how it is going. "*

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

4.5 Theme 4. Personal Privacy & Data Security

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 final theme highlights.

4.5.1 Away From Prying Eyes & 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. In order 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*.

4.5.2 Hands Off Our Data. In addition to the concerns regarding 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 patients' data were carefully explained both in-person and within 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 lead researcher'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 consisted of a vulnerable exercise requiring high levels of social trust and personal courage. We conclude this paper by moving from participants' reflections on their experience of the technology to implications for the future design of CAs to support the self-report of health and wellbeing.

4.6 Participants' Design Recommendations

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 discussed ways to enable this.

4.6.1 Improving Conversation Design . Participants made several design recommendations to improve 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 which 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"*.

(ii) Probing Further. 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 conversation 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 subjects like okay 'How do you feel from 1 to 10?', it's like 'okay 1. . . 5. . . 3'. . . "* [P9].

(iii) Guiding More. Participants finally suggested probing their responses further to engage them in a more 'natural' conversation. They stressed that the CA does not need to understand everything but could employ strategies to support a richer conversation. P20 provided an example; *"For e.g., 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 key words and follow up on the basis of those keywords"*.

Many participants also suggested alternate conversation designs. P18 commented, 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 and having to be the one leading the story through the entire way" [P19].

4.6.2 Provide Space 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 for self-reflection, arguing that it 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 with friends, family, and 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 the 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 the 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].

5 DISCUSSION

We conclude by examining enablers and barriers to positive CA self-report experiences as well as ethical and conceptual reflections on the design and use of CAs to support mental health and wellbeing.

5.1 Enablers and Barriers to User Engagement

Engagement is a prominent goal for the design of many systems within HCI, and has been touted as one of the potential advantages of speech-enabled systems. One way to support design for engagement in the practice of self-report via CA is therefore to begin to understand its enablers and barriers.

Many participants of this study demonstrated high levels of engagement (See Sections 4.1.2, 4.1.3), which often reflected broad acceptance of the technology as well as positive perceptions of the CA's potential to support mental health and wellbeing. As motivations for their engagement in self-report via an agent, many participants, in line with prior findings [31, 33], pointed to the convenience offered by the CA's hands-free experience; *"I love it because it's like an interface you talk directly to. It's super easy to use. You don't have to open your laptop and go to a specific page. I can just go home, open the door and talk to Sofia, super easy"* [P1].

Speech, as a more natural form of interaction, was also described as allowing users to express their emotions more freely and spontaneously compared to other means of self-report; *"Speaking is much easier because you can just let the words flow and you don't have to think about it"* [P14]. Still other participants spoke to feeling heard – a quality that we believe is fundamental for sustainable long term CA-user relationships. Users' appropriation of the technology into their lives and for their own purposes therefore reflects diverse forms of value associated with the technology. P14 and P6, for example, even took the smart speaker with them when away from home; *"So for three weeks I was in the sort of caravan and actually it worked fine. I just had to reset it to my, to the Wi-Fi, and that was not a problem"*[P14].

And yet, as barriers to their engagement, participants often mentioned technical limitations which prevented users from completing their self-reports, social factors, and privacy concerns including eavesdropping and data security. Our findings therefore reflect participants' willingness to engage with CAs for the self-report of mental health and wellbeing, but also raise questions about the readiness of CAs for real-world use.

When thinking about engagement as a design goal it is therefore important to resist a narrow focus on engagement as an aim in itself, and to think about how the technology best supports participants' own health and wellbeing goals, and how users are best able to appropriate the technology into their own lives. This is particularly important for users living with AD and other vulnerable populations.

5.2 Stigma, Vulnerability & Design Ethics

Participants' involvement in this study required supporting high levels of trust given the potentially vulnerable nature of the population group, the stigma surrounding mental illness, and the novel nature of the technology itself.

Participants' perceptions of apps as a good listener, a companion, and a tool for emotional venting and self-talking suggest the potential of CAs to prove of meaningful value to users, and yet also raises questions concerning the potential ethical ramifications of CA personification. Such connections and associations may, for example, serve as a means of overcoming stigma, or strengthening it, should that stigma become attached to the technology itself.

It is also therefore important to question the extent to which these agents are truly capable of serving as companions, and indeed the extent to which we desire them to do so. The therapeutic, and healthcare literature more broadly, highlights the role of relationship and rapport as mediating factors in the efficacy of care. Can a CA, for example, play the role of a therapist? This research suggests that users may find value in connecting to agents to a greater extent than even we as developers imagined given the limited nature of this agent, and yet also underlines the need for future research if these technologies are to be implemented and deployed in ethical ways.

5.3 Conversational Self-Report | A Relationship-Oriented Design Framing

One of the as-of-yet unspoken questions underlining this framing of technology concerns whether self-report is best framed as a conversation at all. And indeed, where the value lies in doing so. The term 'report' may be read as implying a power imbalance which did not necessarily reflect how all participants of this study conceived of their relationship to *Sofia*.

Indeed, participants' descriptions of *Sofia* itself ranged from 'this round little thing' to 'this little person in my life;' more often than not leaning in the direction of humanization and personification of the agent. As these devices grow ever more ubiquitous, researchers are beginning to question, as in this work, ways of understanding and designing for increasingly meaningful and longitudinal forms of interaction. This is not a trivial change in orientation, and one which has featured strongly in our thinking around and framing of the design of *Sofia*. As designers and developers of these systems

we are increasingly called upon to consider questions pertaining to the kinds of relationships we hope to embody and support via these systems, and the expectations we intend to set for users in this respect. Do we aspire to realize, for example, professional, casual, therapeutic or wholly transactional relationships, and which design choices realize these modes not only of interacting, but relating?

5.4 Desired Futures

> *"If Sofia can do, not to automate the life, to be someone to talk to. Yes, I can do that everyday. I can trust to talk to Sofia everyday. I don't know if you have seen the movie, 'Her' [41]. Something like that. The 'Her' thing is not to automate your life or to remind you anything. Google Assistant can do that. Someone to talk with, not to talk with to get any feedback, just to talk with to review your day and to see okay what you're actually doing."* [P16]

Looking forward, as smart-speaker devices become more mainstream, it is worth considering the futures we desire, and how we might design to sustainably support them. What role do we want these devices to play in our lives?

Although most participants expressed hesitation in considering health recommendations provided by a CA, several mentioned that they would welcome non-intrusive recommendations that reminded them of the things they could do, although added that a CA should grant them broad scope to develop and make their own decisions; *"You can remind sometimes. It should be not like 'Okay, here you have a breathing exercise'. Like 'Do you want to do a breathing exercise to improve your mental health?' and then you can say 'Yes' or 'No'. I think it's important that you still have the possibility to say no"* [P14]. Other participants suggested incorporating humor and health advice into conversations with *Sofia*.

These possibilities raise additional questions. Do we, for example, want agents to provide recommendations for our health and wellbeing, or prove truly capable of holding rich conversations, or might we prefer uncaring machines, who may also ironically be best placed to provide an experience of care? These are the kinds of questions raised by this study, as one of the first to engage such a participant group in the real-world use of CA technologies, and we hope will serve to fuel future research efforts, as we continue to reflect on the implications of growing adoption of these technologies,

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, most participants mentioned that their mental, physical, and social situations at the time of the study was different than usual, which means that the study results may not translate to this population's more typical context. Some participants might have engaged more often with CA 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].

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 [33]. Inconsistency in the pre- and post-study interview methods, as described in Section 3.3.3, could also have induced bias in the interview data.

Our study includes a small and homogeneous sample size ($N = 20$) from one country and hence may not generalize to other contexts.

7 CONCLUSION

This study provides initial insights into the perceptions, experiences, and behaviors of people living with affective disorder (AD) following four weeks of CA use to self-report mental health and well-being. Results suggest that CAs can serve as engaging tool for emotional expression, despite outstanding technical limitations, social challenges and privacy concerns. Participants' high engagement and positive responses to the UEQ questionnaire suggest users' positive perception and acceptance of CA for mental health self-reports. Thematic analysis of participants' comments provided during post-study interviews revealed that users' perceptions of the agent as a good listener, companion, and tool for emotional venting and self-talking enabled them to express their emotions more naturally, made them feel heard, and facilitated positive relationships with their loved ones. In addition to the technical limitations preventing users from completing self-reports, we find that users' social context, and privacy challenges including eavesdropping and data security concerns were the primary barriers to mental health self-reports via CA. To overcome technical limitations and fully express their mental state, participants applied strategies including making multiple entries, adapting their speech and preparing their reflections in advance. Based on their self-reporting experiences, participants suggested improving the conversation design by varying questions more often, probing, and guiding the conversation as well as suggesting ways to reflect on their data to support engaging and sustainable mental health self-reports via CA. Reflecting on our findings, we discuss several challenges in designing CAs to support mental health reporting, which we hope will guide the future design of CAs for mental health and wellbeing.

ACKNOWLEDGMENTS

This project is financially supported by the Novo Nordisk Foundation, Denmark as part of the "SOFIA" project and the Copenhagen Center for Health Technology.

REFERENCES

[1] Patricia A Areán, Kien Hoa Ly, and Gerhard Andersson. 2016. Mobile technology for mental health assessment. *Dialogues in clinical neuroscience* 18, 2 (2016), 163.

[2] Ivano Azzini, Daniele Falavigna, Toni Giorgino, Roberto Gretter, Silvana Quaglini, Carla Rognoni, and Mario Stefanelli. 2003. Automated spoken dialog system for home care and data acquisition from chronic patients. *Studies in health technology and informatics* 95 (2003), 146–151.

[3] 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.

[4] JC Bennett. 1991. The irrationality of the catharsis theory of aggression as justification for educators' support of interscholastic football. *Perceptual and motor skills* 72, 2 (1991), 415–418.

[5] Timothy Bickmore, Amanda Gruber, and Rosalind Picard. 2005. Establishing the computer–patient working alliance in automated health behavior change interventions. *Patient education and counseling* 59, 1 (2005), 21–30.

[6] Ann E Blandford. 2013. Semi-structured qualitative studies. Interaction Design Foundation.

[7] Virginia Braun and Victoria Clarke. 2006. Using thematic analysis in psychology. *Qualitative research in psychology* 3, 2 (2006), 77–101.

[8] 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/

[9] Kimberly J Callicott and Hija Park. 2003. Effects of self-talk on academic engagement and academic responding. *Behavioral Disorders* 29, 1 (2003), 48–64.

[10] 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 2020 CHI Conference on Human Factors in Computing Systems (CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3313831.3376789

[11] 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 32*. 1–5.

[12] 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 2019 on Designing Interactive Systems Conference (DIS '19)*. Association for Computing Machinery, New York, NY, USA, 1557–1569. https://doi.org/10.1145/3322276.3322332

[13] 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 2020 CHI Conference on Human Factors in Computing Systems (CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3313831.3376569

[14] 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 and uHealth* 6, 9 (2018), e174.

[15] 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*. 72–82.

[16] 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 2014 international conference on Autonomous agents and multi-agent systems*. 1061–1068.

[17] Kevin Doherty, Andreas Balaskas, and Gavin Doherty. 2020. The Design of Ecological Momentary Assessment Technologies. *Interacting with Computers* (2020).

[18] Kevin Doherty and Gavin Doherty. 2018. Engagement in HCI: conception, theory and measurement. *ACM Computing Surveys (CSUR)* 51, 5 (2018), 1–39.

[19] 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*. 1–15.

[20] 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, USA, Article 6, 9 pages. https://doi.org/10.1145/3405755.3406120

[21] 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 mental health* 4, 2 (2017), e19.

[22] Anushay Furqan, Chelsea Myers, and Jichen Zhu. 2017. Learnability through Adaptive Discovery Tools in Voice User Interfaces. In *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems (CHI EA '17)*. Association for Computing Machinery, New York, NY, USA, 1617–1623. https://doi.org/10.1145/3027063.3053166

[23] Hannah Gaffney, Warren Mansell, and Sara Tai. 2019. Conversational Agents in the Treatment of Mental Health Problems: Mixed-Method Systematic Review. *JMIR Mental Health* 6, 10 (2019), e14166.

[24] 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. *Perspectives on Psychological Science* 11, 6 (2016), 838–854.

[25] Eva Hudlicka. 2013. Virtual training and coaching of health behavior: Example from mindfulness meditation training. *Patient education and counseling* 92, 2 (2013), 160–166.

[26] Philip C Kendall. 1991. Guiding theory for therapy with children and adolescents. *Child and adolescent therapy: Cognitive-behavioral procedures* 2 (1991).

[27] Junhan Kim, Sun Park, Lionel Robert, et al. 2019. Conversational Agents for Health and Wellbeing: Review and Future Agendas. (2019).

[28] A Baki Kocaballil, 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 32*. 1–12.

[29] 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. *Journal of medical Internet research* 21, 11 (2019), e15360.

[30] 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 2020 CHI Conference on Human Factors in Computing Systems*. 1–8.

[31] 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 2018 Designing Interactive Systems Conference (DIS '18)*. Association for Computing Machinery, New York, NY, USA, 881–894. https://doi.org/10.1145/3196709.3196784

[32] Liliana Laranjo, Adam G Dunn, Huong Ly Tong, Ahmet Baki Kocaballi, Jessica Chen, Rabia Bashir, Didi Surian, Blanca Gallego, Farah Magrabi, Annie YS Lau, et al. 2018. Conversational agents in healthcare: a systematic review. *Journal of the American Medical Informatics Association* 25, 9 (2018), 1248–1258.

[33] Josephine Lau, Benjamin Zimmerman, and Florian Schaub. 2018. Alexa, are you listening? privacy perceptions, concerns and privacy-seeking behaviors with smart speakers. *Proceedings of the ACM on Human-Computer Interaction* 2, CSCW (2018), 1–31.
[34] 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.
[35] C Leslie. 2008. Boxing is the best way to stop violence in kids. *The Guardian Politics Blog* (2008).
[36] Esther Levin and Alex Levin. 2006. Evaluation of spoken dialogue technology for real-time health data collection. *Journal of medical Internet research* 8, 4 (2006), e30.
[37] Irene Lopatovska and Harriet Williams. 2018. Personification of the Amazon Alexa: BFF or a mindless companion. In *Proceedings of the 2018 Conference on Human Information Interaction & Retrieval*. 265–268.
[38] Gale M Lucas, Jonathan Gratch, Aisha King, and Louis-Philippe Morency. 2014. It's only a computer: Virtual humans increase willingness to disclose. *Computers in Human Behavior* 37 (2014), 94–100.
[39] 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. *Frontiers in Robotics and AI* 4 (2017), 51.
[40] 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 2016 CHI Conference on Human Factors in Computing Systems (CHI '16)*. Association for Computing Machinery, New York, NY, USA, 5286–5297. https://doi.org/10.1145/2858036.2858288
[41] Ellison Megan, Spike Jonze, and Vincent Landay. 2014. Her. Warner Bros. Pictures (United States/Germany) Sony Pictures Releasing.
[42] Adam Miner, Amanda Chow, Sarah Adler, Ilia Zaitsev, Paul Tero, Alison Darcy, and Andreas Paepcke. 2016. Conversational agents and mental health: Theory-informed assessment of language and affect. In *Proceedings of the fourth international conference on human agent interaction*. ACM, 123–130.
[43] 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 Digital Medicine* 3, 1 (2020), 1–8.
[44] 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 medicine* 176, 5 (2016), 619–625.
[45] Roger K Moore. 2013. Spoken language processing: where do we go from here? In *Your Virtual Butler*. Springer, 119–133.
[46] 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, USA, 347–351. https://doi.org/10.1145/3240925.3240933
[47] Nasim Motalebi, Eugene Cho, S Shyam Sundar, and Saeed Abdullah. 2019. Can Alexa be your Therapist? How Back-Channeling Transforms Smart-Speakers to be Active Listeners. In *Conference Companion Publication of the 2019 on Computer Supported Cooperative Work and Social Computing*. 309–313.
[48] Christine Murad, Cosmin Munteanu, Benjamin R Cowan, and Leigh Clark. 2019. Revolution or Evolution? Speech Interaction and HCI Design Guidelines. *IEEE Pervasive Computing* 18, 2 (2019), 33–45.
[49] 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 2018 CHI Conference on Human Factors in Computing Systems*. ACM, 6.
[50] 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*. 111–112.
[51] 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 2016 CHI Conference on Human Factors in Computing Systems (CHI '16)*. Association for Computing Machinery, New York, NY, USA, 4921–4932. https://doi.org/10.1145/2858036.2858478
[52] 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*. ACM, 640.
[53] 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 2018 CHI Conference on Human Factors in Computing Systems*. 1–13.
[54] Simon Provoost, Ho Ming Lau, Jeroen Ruwaard, and Heleen Riper. 2017. Embodied conversational agents in clinical psychology: a scoping review. *Journal of medical Internet research* 19, 5 (2017), e151.
[55] 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 2017 CHI*

Conference Extended Abstracts on Human Factors in Computing Systems. 2853–2859.

[56] 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.

[57] 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 2020 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2020 ACM International Symposium on Wearable Computers*. 494–496.

[58] 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.

[59] 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/

[60] Lazlo Ring, Lin Shi, Kathleen Totzke, and Timothy Bickmore. 2015. Social support agents for older adults: longitudinal affective computing in the home. *Journal on Multimodal User Interfaces* 9, 1 (2015), 79–88.

[61] 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 2020 CHI Conference on Human Factors in Computing Systems*. 1–13.

[62] Joel Sebastian and Deborah Richards. 2017. Changing stigmatizing attitudes to mental health via education and contact with embodied conversational agents. *Computers in Human Behavior* 73 (2017), 479–488.

[63] Ji Youn Shin and Jina Huh-Yoo. 2018. Designing Everyday Conversational Agents for Managing Health and Wellness. (2018).

[64] Ben Shneiderman. 2000. The Limits of Speech Recognition. *Commun. ACM* 43, 9 (Sept. 2000), 63–65. https://doi.org/10.1145/348941.348990

[65] Bernhard Suhm. 2003. Towards best practices for speech user interface design. In *Eighth European Conference on Speech Communication and Technology*.

[66] 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. *Journal of Aggression, Maltreatment & Trauma* 29, 5 (2020), 611–627.

[67] 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.

[68] Kimberly RH Treadwell and Philip C Kendall. 1996. Self-talk in youth with anxiety disorders: States of mind, content specificity, and treatment outcome. *Journal of consulting and clinical psychology* 64, 5 (1996), 941.

[69] Niels Van Berkel, Denzil Ferreira, and Vassilis Kostakos. 2017. The experience sampling method on mobile devices. *ACM Computing Surveys (CSUR)* 50, 6 (2017), 1–40.

[70] 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 2020 CHI Conference on Human Factors in Computing Systems (CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3313831.3376561

[71] Adam Waytz, John Cacioppo, and Nicholas Epley. 2010. Who sees human? The stability and importance of individual differences in anthropomorphism. *Perspectives on Psychological Science* 5, 3 (2010), 219–232.

[72] Zhuxiaona Wei and James A Landay. 2018. Evaluating Speech-Based Smart Devices Using New Usability Heuristics. *IEEE Pervasive Computing* 17, 2 (2018), 84–96.

[73] 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. *Journal of medical Internet research* 19, 4 (2017), e132.