MentalBuddy: Expanding Access to Mental Healthcare Through Conversational Agents

Luke Haliburton
LMU Munich
Munich, Germany
luke.haliburton@ifi.lmu.de

Mari Kruse
LMU Munich
Munich, Germany

Christopher Voit
LMU Munich
Munich, Germany

Albrecht Schmidt
LMU Munich
Munich, Germany

Abstract

Voice interactions with conversational agents are becoming increasingly ubiquitous. At the same time, stigmas around mental health are beginning to break down, but there remain significant barriers to treatment. Mental health conditions are highly prevalent and people fail to receive help due to lack of access, information, or structures. We aim to address these problems by investigating the applicability of voice-based conversational agents for mental health. In this paper, we introduce our first prototype, MentalBuddy, present initial user feedback, and discuss the potential ethical implications of using conversational agents in mental health applications. With proper considerations, conversational interfaces have the potential to create scalable access to mental health prevention, diagnosis, and therapy.

Author Keywords

VCA, VUI, Mental Health, CUI

CCS Concepts

•**Human-centered computing** → **Sound-based input / output;** *Personal digital assistants;* •**Applied computing** → *Health informatics;*

Introduction

Depression is highly prevalent [14], and although the conversation around mental health is evolving, stigmas are still
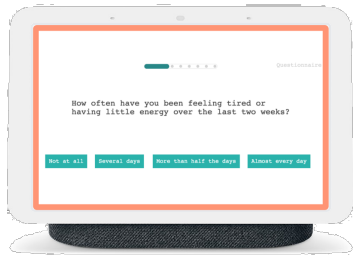
Figure 1: A visualization of MentalBuddy on the Google Nest Hub.

a key reason that people do not seek help [1]. Financial issues, service availability, wait times, and personnel shortages are also key barriers to treatment [1, 8]. Ubiquitous access to healthcare support through intelligent agents is one promising method to combat some of these barriers.

Researchers have identified a need for digital health tools to support home and remote treatment [25]. Independent conversational agents are one potential solution that can scale to huge populations and can collect additional contextual data if they are deployed in smartphones or embedded in homes. Prior work has shown that voice-based mental health evaluations are as effective as traditional paper-based methods [18], and text-based conversational agents have been effective in improving depression symptoms [8].

In this paper, we present a voice-based conversational agent (VCA), MentalBuddy, which is an at-home mental health evaluation tool. MentalBuddy measures depression symptoms in users by talking through a standardized questionnaire and then delivers recommendations. We conducted a pilot evaluation and found that MentalBuddy is highly usable, although trust is a concern and users have individual interaction preferences requiring personalization. We discuss the ethics of using VCAs for highly sensitive conversations about mental health and delivering difficult results to users. Overall, VCAs have the potential to increase access to mental health support in any location.

Background

Despite the potential benefits of implementing conversational user interfaces (CUIs) in healthcare, multiple reviews [4, 12, 9] found that the field is relatively new and more work is needed before they become clinically relevant.

Recent work has investigated the use of CUIs to enhance mental health. Fitzpatrick et al. [5] attempted to make web-based cognitive behavioral therapy more engaging with a CUI, leading to a significant reduction in depression symptoms. Liu et al. [15] developed voice user interface (VUI) prototypes to support family caregivers with therapy exercises and report that although there were benefits to hands-free interaction, there was a high cognitive load requirement and concerns about discussing sensitive topics out loud. Previous research has also focused on preventative mental health care. For example, Muppirishetty et al. [19] and Lee et al. [13] developed VUIs to promote self-compassion. We aim to extend this body of work by enabling easy access to early diagnosis tools through VCAs.

Implementation of MentalBuddy

We created MentalBuddy using Google Assistant[1] on a Google Nest Hub[2]. The Nest Hub has a screen to support users who desire visual support [20]. The primary dialogue flow begins with an introduction of information, proceeds through a standardized questionnaire, and finally delivers the results.

The users are guided through the PHQ-9 questionnaire [11], which was selected because it is short (9 questions) and commonly used [5]. We incorporated a standardized scale as a first step with the intention of adding more natural conversation dynamics in the future. We provided the users with their results using the SPIKES (Setting, Perception, Invitation, Knowledge, Empathy, and Strategy) protocol [2], which physicians use to deliver difficult news to patients.

We designed the personality of MentalBuddy to be friendly and non-judgmental, as well as rational and professional. The aim was to create a welcoming environment where

---

[1] https://developers.google.com/assistant
[2] https://store.google.com/us/product/nest_hub_2nd_gen

users are not afraid to say how they feel, while also emulating the professionalism of a therapist to instill trust.

Pilot Study: Initial Feedback

We conducted a pilot study with N=10 participants (6 female, 4 male), with an average age of 26.3 (SD=2.54) and a high affinity for technology interaction (M=4.304, SD=0.86). After interacting with MentalBuddy, the participants completed the User Experience Questionnaire (UEQ) [24] and UEQ+ scales for voice interaction [10], and participated in a brief semi-structured interview.

The results of the UEQ and UEQ+ are shown in Figures 2 and 3. MentalBuddy was rated positively in every category except for Trust, indicating that the system is usable, but there are concerns with sharing sensitive information. In the interviews, participants provided varying feedback on the personality of MentalBuddy, which matches prior work showing individual preference in VUI interactions [17]. For example, P3 said that they "*did not feel judged*" since they knew they were interacting with a artificial agent, while P7 found that the "*robotic*" voice did not sound empathetic. Notably these participants also had very high (2.00 — P3) and very low (-1.50 — P7) mean trust scores respectively.

Consequences for CUI Ethics

Trusting MentalBuddy MentalBuddy received a "neutral" Trust score, although trust ratings appeared to vary according to the subjective perceptions of MentalBuddy they expressed in the interviews. The interview responses are in line with literature showing that users do not feel judged while talking to chatbots [6], but prior work has also shown that trust is reduced when participants believe there is a human operator [16]. Although we attempted to give MentalBuddy a trusting personality, it is possible that our partic-
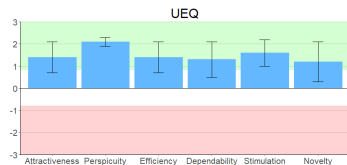

Figure 2: User responses to the UXQ. All categories were positive.
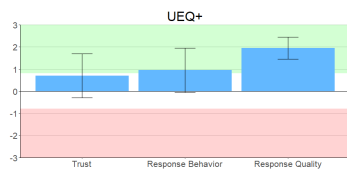

Figure 3: User responses to the UXQ+. The system is rated positively in all categories except "Trust", which is neutral

ipants were not convinced that MentalBuddy was an independent agent.

Handling Sensitive Information. We consulted two Clinical Therapists to understand where they expect MentalBuddy to fit into the mental health process. The therapists saw opportunities for a VCA to act as either a pre-screening tool or as a mediating entity between patients at home and their therapist. In both roles, they imagined that MentalBuddy would share information about the patient with the therapist. There are basic ethical considerations in designing such a system, such as being transparent with how the data is shared and used, but it is also crucial to preserve user agency. Incorporating the right to transparent explanations [7] and the right to be forgotten [23] can both help move the system towards being an ethical partner.

One-Size-Fits-All Questionnaires. Standardized scales like the PHQ-9 [11] are commonly used in digital mental health studies because they are short and have clinical validation. However, Pendse et al. point out that such scales are not equally valid across all identities and cultures [22]. In particular, Borgogna et al. [3] found that the PHQ-9 is less sensitive to depression symptoms for gay men, questioning women, and queer men. Embedding standardized scales in technology is a common practice in HCI that unintentionally embeds biases into the decisions made by that technology. It is therefore crucial to critically analyze the lenses and standards through which the users will be viewed, especially in diagnostic applications. In future iterations, we would aim to leverage the flexibility of voice interaction to create a more dynamic, conversation-based approach to mental health. As suggested by [22], we should incorporate a focus on lived experience and seek to understand more context rather than categorizing users as quickly as possible.

Handling a Crisis. A conversational agent for mental health will likely have to deliver difficult results to some patients. We used the SPIKES protocol in our prototype to deliver results in an empathetic manner, but there are outstanding issues. If MentalBuddy delivers negative results to a user and their condition subsequently worsens, how should the system react? In its current form, the system is not equipped to handle a crisis situation, which is true of most mental health apps [21]. The debate around *who* should respond to mental health crises has increased recently, with calls to increase funding for unarmed crisis intervention teams to replace armed law enforcement for such tasks [22]. It is crucial for MentalBuddy and similar agents to consider crisis situations in the design phase.

Conclusion

In this paper we present MentalBuddy, a VCA that uses a standardized depression questionnaire to evaluate users and makes recommendations using an empathetic protocol. Users found MentalBuddy to be usable but had concerns with trust. We discuss the ethics of using conversational agents for highly sensitive evaluations and communicating difficult information. Overall, conversational agents have the potential to increase access and reduce barriers for users seeking mental health support.

Acknowledgments

REFERENCES

[1] L H Andrade, J Alonso, Z Mneimneh, J E Wells, A Al-Hamzawi, A Al-Hamzawi, G Borges, E Bromet, R Bruffaerts, G de Girolamo, R de Graaf, S Florescu, O Gureje, H R Hinkov, C Hu, Y Huang, I Hwang, R Jin, E G Karam, V Kovess-Masfety, D Levinson, H Matschinger, S O'Neill, J Posada-Villa, R Sagar, N A Sampson, C Sasu, D J Stein, T Takeshima, M C Viana, M Xavier, and R C Kessler. 2014. Barriers to mental health treatment: results from the WHO World Mental Health surveys. *Psychological medicine* 44, 6 (April 2014), 1303–1317. DOI: http://dx.doi.org/10.1017/s0033291713001943

[2] Walter F. Baile, Robert Buckman, Renato Lenzi, Gary Glober, Estela A. Beale, and Andrzej P. Kudelka. 2000. SPIKES—A Six-Step Protocol for Delivering Bad News: Application to the Patient with Cancer. *The Oncologist* 5, 4 (Aug. 2000), 302–311. DOI: http://dx.doi.org/10.1634/theoncologist.5-4-302

[3] Nicholas C. Borgogna, Rachel E. Brenner, and Ryon C. McDermott. 2021. Sexuality and gender invariance of the PHQ-9 and GAD-7: Implications for 16 identity groups. *Journal of Affective Disorders* 278 (Jan. 2021), 122–130. DOI: http://dx.doi.org/10.1016/j.jad.2020.09.069

[4] Caterina Bérubé, Theresa Schachner, Roman Keller, Elgar Fleisch, Florian v Wangenheim, Filipe Barata, and Tobias Kowatsch. 2021. Voice-Based Conversational Agents for the Prevention and Management of Chronic and Mental Health Conditions: Systematic Literature Review. *Journal of Medical Internet Research* 23, 3 (March 2021), e25933. DOI:http://dx.doi.org/10.2196/25933

[5] 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 (June 2017), e7785. DOI: `http://dx.doi.org/10.2196/mental.7785`

[6] Asbjørn Følstad, Cecilie Bertinussen Nordheim, and Cato Alexander Bjørkli. 2018. What Makes Users Trust a Chatbot for Customer Service? An Exploratory Interview Study. In *Internet Science*, Svetlana S. Bodrunova (Ed.). Vol. 11193. Springer International Publishing, Cham, 194–208. DOI: `http://dx.doi.org/10.1007/978-3-030-01437-7_16` Series Title: Lecture Notes in Computer Science.

[7] Bryce Goodman and Seth Flaxman. 2017. European Union Regulations on Algorithmic Decision-Making and a "Right to Explanation". *AI Magazine* 38, 3 (Oct. 2017), 50–57. DOI: `http://dx.doi.org/10.1609/aimag.v38i3.2741` Number: 3.

[8] Becky Inkster, Shubhankar Sarda, and Vinod Subramanian. 2018. An Empathy-Driven, Conversational Artificial Intelligence Agent (Wysa) for Digital Mental Well-Being: Real-World Data Evaluation Mixed-Methods Study. *JMIR mHealth and uHealth* 6, 11 (Nov. 2018), e12106. DOI: `http://dx.doi.org/10.2196/12106`

[9] William R. Kearns, Nai-Ching Chi, Yong K. Choi, Shih-Yin Lin, Hilaire Thompson, and George Demiris. 2019. A Systematic Review of Health Dialog Systems. *Methods of Information in Medicine* 58, 6 (Dec. 2019), 179–193. DOI: `http://dx.doi.org/10.1055/s-0040-1708807` Publisher: Georg Thieme Verlag KG.

[10] Andreas M. Klein, Andreas Hinderks, Martin Schrepp, and Jörg Thomaschewski. 2020. Construction of UEQ+ scales for voice quality: measuring user experience quality of voice interaction. In *Proceedings of the Conference on Mensch und Computer (MuC '20)*. Association for Computing Machinery, New York, NY, USA, 1–5. DOI: `http://dx.doi.org/10.1145/3404983.3410003`

[11] K. Kroenke, R. L. Spitzer, and J. B. Williams. 2001. The PHQ-9: validity of a brief depression severity measure. *Journal of General Internal Medicine* 16, 9 (Sept. 2001), 606–613. DOI:`http://dx.doi.org/10.1046/j.1525-1497.2001.016009606.x`

[12] Liliana Laranjo, Adam G Dunn, Huong Ly Tong, Ahmet Baki Kocaballi, Jessica Chen, Rabia Bashir, Didi Surian, Blanca Gallego, Farah Magrabi, Annie Y S Lau, and Enrico Coiera. 2018. Conversational agents in healthcare: a systematic review. *Journal of the American Medical Informatics Association* 25, 9 (Sept. 2018), 1248–1258. DOI: `http://dx.doi.org/10.1093/jamia/ocy072`

[13] Minha Lee, Sander Ackermans, Nena van As, Hanwen Chang, Enzo Lucas, and Wijnand IJsselsteijn. 2019. Caring for Vincent: A Chatbot for Self-Compassion. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–13. DOI:`http://dx.doi.org/10.1145/3290605.3300932`

[14] Grace Y. Lim, Wilson W. Tam, Yanxia Lu, Cyrus S. Ho, Melvyn W. Zhang, and Roger C. Ho. 2018. Prevalence of Depression in the Community from 30 Countries between 1994 and 2014. *Scientific Reports* 8, 1 (Feb. 2018), 2861. DOI: `http://dx.doi.org/10.1038/s41598-018-21243-x` Number: 1 Publisher: Nature Publishing Group.

[15] Yun Liu, Lu Wang, William R. Kearns, Linda Wagner, John Raiti, Yuntao Wang, and Weichao Yuwen. 2021.

Integrating a Voice User Interface into a Virtual Therapy Platform. In *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, 1–6. https://doi.org/10.1145/3411763.3451595

[16] 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 (Aug. 2014), 94–100. DOI: http://dx.doi.org/10.1016/j.chb.2014.04.043

[17] Qianli Ma, Ronggang Zhou, Chenyang Zhang, and Zhe Chen. 2021. Rationally or emotionally: how should voice user interfaces reply to users of different genders considering user experience? *Cognition, Technology & Work* (Sept. 2021). DOI: http://dx.doi.org/10.1007/s10111-021-00687-8

[18] Omtawan Mangkang and Jae Young Yun. 2019. A Comparative Study of Self-Diagnosis User Interfaces for Depression :Focusing on Usability of Paper-Based, Text-Based and Voice-Based Conversational Interface. In *Proceedings of HCI Korea 2019 COnference*. Korean HCI Society, Korea, 262–267. https://www.dbpia.co.kr/journal/articleDetail?nodeId=NODE08008118

[19] P. Muppirishetty and Minha Lee. 2020. Voice User Interfaces for mental healthcare: Leveraging technology to help our inner voice. In *Proceedings of the 3rd ACM Conference on Computer-Supported Cooperative Work and Social Computing*. ACM, Online, 5.

[20] Chelsea M. Myers, Anushay Furqan, and Jichen Zhu. 2019. The Impact of User Characteristics and

Preferences on Performance with an Unfamiliar Voice User Interface. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–9. DOI: http://dx.doi.org/10.1145/3290605.3300277

[21] Emma M. Parrish, Tess F. Filip, John Torous, Camille Nebeker, Raeanne C. Moore, and Colin A. Depp. 2021. Are mental health apps adequately equipped to handle users in crisis? *Crisis: The Journal of Crisis Intervention and Suicide Prevention* (2021), No Pagination Specified–No Pagination Specified. DOI: http://dx.doi.org/10.1027/0227-5910/a000785 Place: Germany Publisher: Hogrefe Publishing.

[22] Sachin R Pendse. 2022. From Treatment to Healing:Envisioning a Decolonial Digital Mental Health. (2022), 23.

[23] Jeffrey Rosen. 2011. The Right to Be Forgotten. *Stanford Law Review Online* 64 (2011), 88. https://heinonline.org/HOL/Page?handle=hein.journals/slro64&id=89&div=&collection=

[24] Martin Schrepp, Jörg Thomaschewski, and Andreas Hinderks. 2017. Construction of a Benchmark for the User Experience Questionnaire (UEQ). (June 2017). DOI:http://dx.doi.org/10.9781/ijimai.2017.445 Accepted: 2021-08-30T08:57:55Z Publisher: International Journal of Interactive Multimedia and Artificial Intelligence (IJIMAI).

[25] Ryan Tennant. 2021. Supporting Caregivers in Complex Home Care: Towards Designing a Voice User Interface. (May 2021). https://uwspace.uwaterloo.ca/handle/10012/16963 Accepted: 2021-05-10T20:10:27Z Publisher: University of Waterloo.