DARA: Development of a Chatbot Support System for an Anxiety Reduction Digital Intervention

A Technical Report submitted to the Department of Engineering Systems and Environment

Presented to the Faculty of the School of Engineering and Applied Science

University of Virginia • Charlottesville, Virginia

In Partial Fulfillment of the Requirements for the Degree

Bachelor of Science, School of Engineering

R.X. Schwartz

Spring, 2022

Technical Project Team Members

Annabel Lynch

Disha Patel

Aparna Ramanan

On my honor as a University Student, I have neither given nor received unauthorized aid on this assignment as defined by the Honor Guidelines for Thesis-Related Assignments

Advisor

Laura Barnes, Associate Professor, Department of Engineering Systems and Environment

DARA: Development of a Chatbot Support System for an Anxiety Reduction Digital Intervention

R.X. Schwartz, Aparna Ramanan, Disha Patel, Annabel Lynch, Sonia Baee, Laura Barnes
Department of Engineering Systems and Environment
University of Virginia
Charlottesville, Virginia, USA
res7cd@virginia.edu

***Abstract*— This paper presents the development and preliminary evaluation of a chatbot, "DARA," which is designed to be added to a digital mental health (DMH) cognitive bias intervention targeting anxiety. Although "supportive accountability" (the addition of human coaching to a digital intervention) may help reduce attrition, human support is not always well-suited for DMH interventions targeting anxiety or emotional difficulties (due to the inherent challenge of social interaction among individuals in these populations.) In response, a virtual conversational agent was designed after evaluating support needs with respect to embodiment/non-embodiment, free-text response/quick replies, and domain-free/domain-specific conversational properties. DARA's usability and comprehensiveness were evaluated among 12 subject-matter experts using a suite of quantitative and qualitative usability questions, including the PSSUQ Version 3. Usage data were also recorded from chatbot interactions. Results suggest that chatbots may be particularly appropriate for anxious users, although navigation and lack of depth in chatbot responses may pose concerns. We suggest that these concerns can be resolved by technical improvements and hybridization of the human-chatbot support system. We make future recommendations for the design and integration of the DARA chatbot and the use of chatbots in other DMH interventions targeting anxiety.**

I. BACKGROUND

A. Digital Mental Health Interventions and Chatbots

The prevalence of mental health conditions among the US young adult population has increased over the first decades of the 21st century [1], with a 2020 prevalence of nearly 1 in 3 among the young adult population in the US and 1 in 5 among the entire US population [2]. At the same time, only about half of the US population with a mental health condition has received treatment for their condition [2].

Digital mental health (DMH) interventions, including mobile health (mHealth) and electronic health (eHealth) interventions, represent an avenue for treatment outside of clinical settings and present a scalable opportunity for a greater number of patients [3], [4]. DMH interventions rely on various techniques to support behavior change, including symptom monitoring and cognitive bias training [3].

Attrition (when a participant ends a treatment before achieving optimal response) occurs in DMH interventions for various reasons, including a negative reaction to the intervention, lack of engagement, technical difficulties, or simply not enough time available to complete the program [5], [6]. Attrition remains a concern among many behavior-change technologies, including MindTrails, the system of interest in this study [5], [7], [8], [9].

In order to reduce attrition, many DMH intervention studies have added human coaching, based on a model of "supportive accountability," to their technology-based interventions [10], [11]. The goal of human coaching in a DMH intervention is to help users adhere to the treatment protocol; individuals without professional mental health training serve as coaches, which ideally reduces costs and improves scalability of technology-based interventions. Human coaching has been shown to be more effective than uncoached interventions for symptom reduction, number of completed intervention modules per participant, and rates of intervention completion [12]. However, results from 39 responses to a survey sent by the MindTrails program to the 111 noncompleters of the intervention (N=282) indicate that 59% "did not want to talk on the phone to the coach" as one of their reason(s) for quitting the study [5]. Further, text-based feedback indicated that interaction with a human coach was anxiety-producing for some participants [5]. These outcomes suggest that human coaching may not be the best fit in the context of the MindTrails intervention or more broadly for DMH interventions for anxious individuals.

These findings, as well as unresolved questions about human coaching in other areas of the DMH field [13], led to the design of the DARA chatbot as an alternative means of support. Notably, this method would ideally scale more easily with more participants, and would allow human coaches to step in if users indicated a *preference* for human coaching, rather than leaving human coaching as the only option [5]. Chatbot-based support also affords round-the-clock accessibility [14], and would ideally answer some proportion of basic questions, allowing human coaches to focus on more demanding questions (such as those focusing on individualization of treatment).

Chatbots have been included as a delivery method for other DMH interventions [14], [15], [16], [17], with some evidence that chatbot interventions are associated with improved treatment outcomes, particularly in relation to anxiety, well-being, and stress. However, the chatbot we propose in this study is not an intervention delivery method: rather, it supplements human coaching and is designed to encourage the user to continue with the study by responding to common questions.

B. The MindTrails Project

MindTrails, or the MindTrails Project [18], is a web-based, public research program developed by a team of clinical psychologists, computer scientists, and engineers at the University of Virginia [7]. It offers free multi-session training programs to individuals with emotional difficulties, such as anxiety or depression, and has been delivered to thousands of people in more than 75 countries around the world [5], [7], [9], [19], [20]. Through a variety of activities for cognitive bias modification for interpretation training (CBM-I), the platform promotes healthier, flexible thinking patterns for anxious individuals [21].

Attrition in MindTrails has remained a significant problem: more than 80% of participants have dropped out of versions of the program before the sixth session in an eight-session intervention [7], [9]. In one MindTrails study, a subgroup of participants had access to a human coach with whom they were able to discuss problems or concerns about the study [5]. About half of this group never responded to coaches' attempts to contact them, though about 30% of the group completed the full intervention with their coaches. These findings further encouraged us to implement a virtual conversational agent to assist human coaches.

II. Methods

A. DARA Chatbot Design

The transition to conversational agent-aided support entailed consideration of three major design choices: *embodiment*; *user response method*; and *domain selection*.

1) Embodiment: Embodied conversational agents (CAs) are virtual characters that have natural human tendencies. In particular, these characters are aware of conversational qualities beyond the task at hand, and can either communicate verbally or through changes in facial expressions, hand gestures, or body language [22], [23]. A non-embodied CA will lack these nonverbal reactions and will solely communicate via text [24]. When implementing DARA, we decided to create a non-embodied CA (a chatbot) that had a friendly icon (Fig. 1), potentially supporting an empathetic reaction to the chatbot regardless of a lack of embodiment. This choice was justified due to a lack of support for the utility of embodied CAs in health settings [22], [25], [26].

2) User response method: Free-text user response methods require a natural language processing component in order for the CA to respond in turn [24], [27]. On the other hand, a CA that provides "quick replies" (sometimes known as "suggestion chips") responds based on a finite set of possible user responses [27]. Other DMH chatbots offer a variety of free-text responses and quick replies [14], [15], [16], [17]. We had higher confidence in quick replies: this method limits the flexibility of the chatbot, but may also improve reliability and the ease of troubleshooting issues with the chatbot [24].

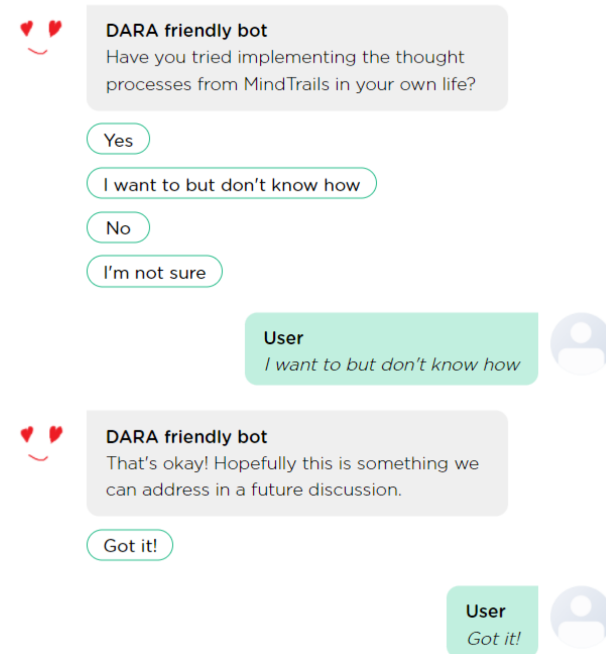

Fig. 1. Screen capture of an interaction with the DARA chatbot, showing a question prompt (top) and informational prompt (bottom).

For these reasons, we exclusively implemented quick replies in this version of DARA.

3) Domain selection: A domain-free (or "open-domain") CA allows either participant to determine the conversation direction and respond to a wide range of topics [27]. Conversely, domain-specific CAs operate in a single domain such as medicine or customer service [27]. We noted that a domain-free CA would introduce too large of a field to process on a preplanned basis and would not be suitable for directed problem-solving. For this reason, the questions within DARA were split into three domains: Technical Issues; Usability and Knowledge Issues; and Implementation Issues. These domains were derived from the failure points noted among users of the human-supported (Version 4) MindTrails program [5]; the failure points originated from the Efficiency Model of Support [28]. Technical Issues refer to smaller bugs and procedural issues that users may encounter on the MindTrails platform; Usability and Knowledge Issues refer to how user-friendly the platform is and how well users understand its purpose; and Implementation Issues refer to how to incorporate lessons learned during the intervention into everyday life. We plan to include prompts for Engagement Issues (one of the original failure points) once MindTrails user behavior in this area is better understood [5].

4) Implementation details: The DARA chatbot (Fig. 1) was implemented using Juji, a chatbot platform that supports IDE-based and visual chatbot design [29]. The final DARA chatbot consisted of 51 different prompts, with a general user flow from an introduction, to Technical Issues, to Usability and Knowledge Issues, to Implementation Issues,

TABLE I

SUMMARY OF CHATBOT CONTENT ORDERED BY GENERAL QUESTION FLOW

Domain	Sample prompt	Question prompts (multiple quick replies)	Informational prompts (single quick reply)
Introduction	"Hello, I am DARA! My role is to help with any issues you have with using the program, issues with motivation, or technical issues. I'm not trained as a therapist, and coaching isn't meant to be phone therapy. My role is to support you as you go through the MindTrails program. Today we will be receiving feedback and asking questions about your previous experience with MindTrails. Let's begin! Note: I don't take free response text, it's less work for you and me :) Please click the buttons below this chat!"	0	1
Technical Issues	"My apologies, could you tell me more about what kind of technical issue you are facing?"	8	4
Usability and Knowledge Issues	"Usability/knowledge issues are any problems or barriers that participants have in completing the training (CBM) sessions. Did you find the program confusing at all or have any questions about the study?"	20	1
Implementation Issues	"Overall, do you think MindTrails can improve on the realistic nature of the scenarios in the program?"	6	10
Conclusion	"Thank you so much for the feedback - you helped me to know more about MindTrails issues! Our goal is to learn every day, so your feedback will help us develop a better version of MindTrails for you and your peers. Please feel free to close your browser at this point, as the further responses will not be recorded."	0	1

to a conclusion (Table I). A user could continue to ask more questions within the same domain, but once the user progressed to the next domain, they could not return to the previous domain. Due to branching, all prompts were not necessarily visited in a single session or in a particular order.

Of the 51 DARA prompts, 34 were questions (those that asked the user to select between more than one quick reply). The remaining 17 prompts were informational, only allowing a single quick reply (e.g. "Got it!" or "Continue").

B. Data Collection

Quantitative and qualitative data were collected to measure interactions with the DARA chatbot: chatbot session length and chatbot session contents were collected directly from the Juji platform. Additionally, a post-interaction survey was constructed: the goal of the post-interaction survey was to have participants assess the usability of DARA and provide additional feedback on the user experience. We selected the 16-question PSSUQ Version 3 as the main usability questionnaire [30]. This usability questionnaire was selected due to its extensive testing, high reliability, and suitability for small sample sizes [31]. 10 additional quantitative and 7 additional qualitative usability questions were added to better understand other aspects of DARA that were not covered by the PSSUQ.

C. Usability Study

A group of subject-matter experts tested the DARA chatbot: participants were recruited from a population that was (1) currently or recently participating in the design and administration of MindTrails (2) not involved in the design of DARA, and (3) willing to participate in a 45-minute usability study (Fig. 2). Due to their familiarity with anxious individuals and the MindTrails platform, these participants were considered to be particularly well-suited to assess DARA in a preliminary usability study.

At the beginning of the the 45-minute session, participants signed a consent form (UVA IRB-SBS #2220) and were familiarized with the supportive relationship of DARA to the MindTrails intervention. Once the consent form was completed, participants were given 15 minutes to interact with DARA. Participants were instructed to complete as many interaction sessions as needed to understand the design and flow of the DARA chatbot. Upon completion of their 15-minute DARA interaction, participants were given another 25 minutes to complete the post-interaction survey.

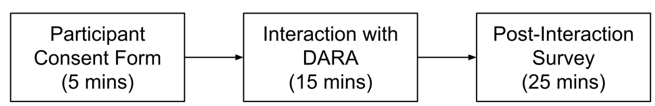

Fig. 2. Flowchart of the preliminary usability study, in which 12 MindTrails subject-matter experts interacted with the DARA chatbot and provided feedback via a post-interaction survey.

III. RESULTS

A. Quantitative Results

Twelve subject-matter experts participated in the study, in the roles of student researcher, licensed psychologist, developer, research assistant, faculty, project coordinator, and website maintainer (some roles were identical between participants). This group conducted a total of 23 sessions with the DARA chatbot, ranging from 0-14 minutes, with an average session lasting 5 minutes. 2 sessions consisted of merely starting the chatbot and viewing the start screen for under one minute. The majority of participants (10) conducted one or two sessions each, while 2 participants conducted four sessions each. In all, participants interacted with DARA chatbot prompts a total of 520 times (Table II).

The overall PSSUQ score of DARA (out of 7, higher scores indicating worse usability) was 2.64, with a System

Quality (SysQual) score of 1.88, an Information Quality (InfoQual) score of 3.19, and an Interface Quality (IntQual) score of 2.94. The inter-rater reliability of the PSSUQ questionnaire was calculated using the irrNA package in R in order to reliably impute missing values [32]. The intra-class correlation coefficient ICC(2,k) was found to be 0.895 ($p<0.001$), with a 95% CI of [0.781, 0.964], indicating good to excellent reliability [33], [34].

The additional 10 quantitative usability questions were also analyzed using the ICC(2,k) irrNA model and found to have an ICC of 0.399 (p=0.06) and a 95% CI of [-0.153, 0.783], indicating poor reliability. Due to poor reliability and the small sample size, we did not pursue further analysis of this non-PSSUQ quantitative question set.

B. Qualitative Results

The responses to the 7 qualitative questions were collaboratively coded by three of the authors at the phrase-level. The tags used were "positive," "negative," and "interesting," and tags could overlap with each other. After an iterative and collaborative classification exercise, four of the authors grouped feedback into three categories: "Initial Opinions," "Ease of Use," and "Effectiveness."

1) Initial Opinions: Participants uniformly described the chatbot personality using positive terms, including "cute," "friendly," and "very pleasant," as well as "very professional." 25% of participants stated that DARA had the potential to provide good accessibility or round-the-clock support for MindTrails users. Additionally, DARA was observed by 58.3% of participants as more ideal than human support for individuals with anxiety. One participant mentioned that DARA produced an environment with "no fear of negative evaluation regarding asking 'silly' or 'obvious' questions," which would help users "anxious about talking to strangers." However, the majority of participants (75%) did not express full confidence in the chatbot, with one participant stating "if the chatbot does not address the concerns of the user adequately, some individuals may prefer to have their specific needs met via human support."

2) Ease of Use: Participant responses regarding chatbot ease of use varied widely, ranging from very positive interactions with DARA to more critical feedback. When participants were asked what they liked about DARA, 58.3% complemented the usability of the interface, using terms including "clean/concise," or "easy to use/navigate." One participant noted "I don't have to type to explain/ask things," referring to the quick replies feature. Another participant mentioned that "[DARA] gave answers in a way that I think was a lot better than any other bot I have recently used." However, one participant described DARA as having a "TON of text for [them] to read" and that "...being able to receive information through conversation may be easier." One participant also suggested it would be beneficial to "have an option… to modify [their] previous [quick replies]."

58.3% of respondents suggested that the DARA chatbot would be less stressful than human interactions, while 25% mentioned that the effect on stress could be mixed, and 8.3% did not believe DARA would be less stressful than human interactions. An interesting point brought up by one participant was that given their experience with DARA, they believe that "it will be less stressful for those high in social anxiety at first but more stressful for people who get frustrated when they can't immediately find information." As general feedback, one participant also mentioned that DARA "randomly 'restarted'" their conversation partway through.

3) Effectiveness: 58.3% of participants mentioned that DARA would be effective for answering basic questions. Specifically, one participant said it helped them "navigate technical questions, and better understand general aspects about the MindTrails program." However, 33.3% of participants mentioned that DARA responses were often rigid, preventing them from gaining additional insight on a topic. One participant stated that they "prefer talking [to a human] and getting more authentic answers," whereas "chatbots are generally not as good at answering complex questions." Another participant mentioned that there was "no opportunity for further clarification if the response to my question doesn't satisfy my needs." Additionally, 33.3% of participants noted DARA's failure to recommend therapeutic resources: one participant suggested that DARA could "[provide] a list of resources at the end of the conversation." To describe their general experience of DARA, 8.3% of participants preferred a chatbot being the only form of external coaching, 33.3% preferred only a human coach, and 58.3% preferred a chatbot/human hybrid.

IV. DISCUSSION AND LIMITATIONS

A. Discussion

The contents of the DARA interactions and survey responses indicated that the subject-matter expert evaluation was comprehensive and mathematically rigorous. 49 of 51 DARA prompts were tested by the twelve participants (96% coverage), and the PSSUQ inter-reliability coding value for the PSSUQ usability scoring was 0.895, indicating good to excellent reliability. These results suggest that the free-form navigation of a chatbot (i.e. mimicking normal usage, as opposed to a structured evaluation) and PSSUQ questionnaire can be effectively combined to evaluate the suitability of new DMH intervention additions among small groups of subject-matter experts.

The relative strength of DARA's System Quality score (1.88/7.00) compared to its slightly worse performance on the Information Quality and Interface Quality scores (3.19/7.00 and 2.94/7.00) support the qualitative results which indicated that DARA was particularly streamlined, easy to learn, and easy to use. Qualitative survey responses also indicated that the DARA chatbot is high in accessibility and supports users' ability to ask basic or even embarrassing questions, both of which are traits that are desirable for an anxious population. The chatbot was also perceived as inviting and friendly, which could potentially drive inter-

TABLE II

SUMMARY OF INTERACTIONS ORDERED BY GENERAL QUESTION FLOW.

Domain	Prompts explored by participants (Total prompts in domain)	Total interactions among all prompts in domain
Introduction	1 (1)	26
Technical Issues	12 (12)	122
Usability and Knowledge Issues	20 (21)	184
Implementation Issues	15 (16)	170
Conclusion	1 (1)	18

action among anxious users. However, participants noted the inflexibility of the chatbot, the length of prompts, and confusion regarding navigation within the chatbot, which may limit the ability of anxious users to effectively take advantage of chatbot support.

Other studies that have included chatbots in health interventions have also noted the conversational limitations of chatbots [14], [15], although some users expressed a preference for quick relies instead of free-text inputs [16]. Additionally, the length of a chatbot conversation has been cited as a concern in other studies [15], while desirable chatbot features may include regular outreach, personalization, and variation in prompts [17]. Participants in other studies have also reported a lack of fear in asking chatbots "the slightest question" [35]. Glitches within chatbot interfaces have also been mentioned in the literature [15].

Based on our results, we recommend that future chatbots for anxious individuals include better navigation assistance, including expectation-setting for chatbot themes, testing of optimal prompt content and quick replies, and the ability to return to an earlier point in the conversation. It would also be desirable for chatbots to route participants to human coaching when participants are confused, frustrated, or simply interested in more in-depth responses. Although our participants appreciated the reliability and straightforwardness of the "quick replies" user response method, free-text inputs may be desirable to understand how to improve chatbot functionality or answer new questions. In the case of MindTrails, free-text inputs would be helpful for understanding the last failure point, Engagement Issues, which is likely an important contributor to attrition [8]. This area is relatively underexplored among MindTrails users and there remains more user feedback we would need to collect before we can define prompts in this area.

B. Limitations

Although this version of DARA does not accept free-text responses, the Juji interface did not allow the free-text entry box to be removed. For this reason, both the DARA chatbot and the study leaders had to explain to participants that free-text entry was not allowed. In a large-scale evaluation (and in the eventual implementation of the chatbot within a DMH system) this issue should be resolved to improve usability.

Given the limited sample size and subject-matter expert participant group in this preliminary study, it is unclear whether the results will generalize to a larger group of anxious individuals. Additional testing is also needed to validate the usability of the chatbot when it is implemented within the MindTrails platform. One notable difference is that this study *required* participants to interact with the DARA chatbot: MindTrails users may or may not choose to interact with a chatbot on a voluntary basis. Given the specific nature of MindTrails failure points, it is not clear which changes would need to be made to the DARA chatbot to generalize results to other DMH interventions.

V. CONCLUSION

This paper presents DARA, a non-embodied, domain-specific, quick reply (non-free-text) conversational agent which is a proposed addition to the MindTrails digital mental health intervention platform. DARA is intended to provide support to users at scale in order to handle basic questions that would be encountered by human coaches. The paper provides background on chatbots in DMH interventions, design objectives for DARA, the methodology for a preliminary usability study among twelve subject-matter experts, results from the study, a discussion, and study limitations. Participants found the chatbot to be highly usable based on the PSSUQ, and expressed favorable reactions to the personality, accessibility, and streamlined nature of the chatbot. Participants also noted that the chatbot could be inflexible, wasn't a good fit for in-depth advice, and could be hard to navigate. In response, we suggest that several improvements can be made to DARA, including better navigation and user expectation-setting, hybridization of human and chatbot support, and the addition of free-text responses in order to better capture new issues. As a next step, usability studies of DARA should be conducted among users of MindTrails.

ACKNOWLEDGMENTS

We would like to thank the MindTrails subject-matter experts for their participation in this study. Research reported in this publication was supported by the National Institute of Mental Health under Award Number MH113752.

REFERENCES

[1] J. M. Twenge, A. B. Cooper, T. E. Joiner, M. E. Duffy, and S. G. Binau, "Age, period, and cohort trends in mood disorder indicators and suicide-related outcomes in a nationally representative dataset, 2005–2017," *Journal of Abnormal Psychology*, vol. 128, no. 3, pp. 185–199, 2019. [Online]. Available: https://doi.org/10.1037/abn0000410

[2] "Mental Illness - National Institute of Mental Health (NIMH)." [Online]. Available: https://www.nimh.nih.gov/health/statistics/mental-illness

[3] S. Michie, L. Yardley, R. West, K. Patrick, and F. Greaves, "Developing and Evaluating Digital Interventions to Promote Behavior Change in Health and Health Care: Recommendations Resulting From an International Workshop," *Journal of Medical Internet Research*, vol. 19, no. 6, p. e7126, Jun. 2017. [Online]. Available: https://doi.org/10.2196/jmir.7126

[4] A. E. Kazdin and S. L. Blase, “Rebooting Psychotherapy Research and Practice to Reduce the Burden of Mental Illness,” *Perspectives on Psychological Science: A Journal of the Association for Psychological Science*, vol. 6, no. 1, pp. 21–37, Jan. 2011. [Online]. Available: https://doi.org/10.1177/1745691610393527

[5] A. Werntz, A. L. Silverman, H. Behan, S. K. Patel, M. Beltzer, M. O. Boukhechba, L. Barnes, and B. A. Teachman, “Lessons Learned: Providing Supportive Accountability in an Online Anxiety Intervention,” *Behavior Therapy*, Dec. 2021. [Online]. Available: https://doi.org/10.1016/j.beth.2021.12.002

[6] I. D. Schmidt, N. R. Forand, and D. R. Strunk, “Predictors of Dropout in Internet-Based Cognitive Behavioral Therapy for Depression,” *Cognitive therapy and research*, vol. 43, no. 3, pp. 620–630, Jun. 2019. [Online]. Available: https://doi.org/10.1007/s10608-018-9979-5

[7] J. L. Ji, S. Baee, D. Zhang, C. P. Calicho-Mamani, M. J. Meyer, D. Funk, S. Portnow, L. Barnes, and B. A. Teachman, “Multi-session online interpretation bias training for anxiety in a community sample,” *Behaviour Research and Therapy*, vol. 142, p. 103864, Jul. 2021. [Online]. Available: https://doi.org/10.1016/j.brat.2021.103864

[8] A. Pratap, E. C. Neto, P. Snyder, C. Stepnowsky, N. Elhadad, D. Grant, M. H. Mohebbi, S. Mooney, C. Suver, J. Wilbanks, L. Mangravite, P. J. Heagerty, P. Areán, and L. Omberg, “Indicators of retention in remote digital health studies: a cross-study evaluation of 100,000 participants,” *npj Digital Medicine*, vol. 3, no. 1, pp. 1–10, Feb. 2020, number: 1 Publisher: Nature Publishing Group. [Online]. Available: https://doi.org/10.1038/s41746-020-0224-8

[9] N. Hohensee, M. J. Meyer, and B. A. Teachman, “The Effect of Confidence on Dropout Rate and Outcomes in Online Cognitive Bias Modification,” *Journal of Technology in Behavioral Science*, vol. 5, no. 3, pp. 226–234, Sep. 2020. [Online]. Available: https://doi.org/10.1007/s41347-020-00129-8

[10] D. C. Mohr, P. Cuijpers, and K. Lehman, “Supportive Accountability: A Model for Providing Human Support to Enhance Adherence to eHealth Interventions,” *Journal of Medical Internet Research*, vol. 13, no. 1, p. e30, Mar. 2011. [Online]. Available: https://doi.org/10.2196/jmir.1602

[11] J. Borghouts, E. Eikey, G. Mark, C. D. Leon, S. M. Schueller, M. Schneider, N. Stadnick, K. Zheng, D. Mukamel, and D. H. Sorkin, “Barriers to and Facilitators of User Engagement With Digital Mental Health Interventions: Systematic Review,” *Journal of Medical Internet Research*, vol. 23, no. 3, p. e24387, Mar. 2021. [Online]. Available: https://doi.org/10.2196/24387

[12] H. Baumeister, L. Reichler, M. Munzinger, and J. Lin, “The impact of guidance on Internet-based mental health interventions — A systematic review,” *Internet Interventions*, vol. 1, no. 4, pp. 205–215, Oct. 2014. [Online]. Available: https://doi.org/10.1016/j.invent.2014.08.003

[13] J. A. Himle, A. Weaver, A. Zhang, and X. Xiang, “Digital Mental Health Interventions for Depression,” *Cognitive and Behavioral Practice*, vol. 29, no. 1, pp. 50–59, Feb. 2022. [Online]. Available: https://doi.org/10.1016/j.cbpra.2020.12.009

[14] K. H. Ly, A.-M. Ly, and G. Andersson, “A fully automated conversational agent for promoting mental well-being: A pilot RCT using mixed methods,” *Internet Interventions*, vol. 10, pp. 39–46, Dec. 2017. [Online]. Available: https://doi.org/10.1016/j.invent.2017.10.002

[15] K. K. Fitzpatrick, A. Darcy, and M. Vierhile, “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*, vol. 4, no. 2, p. e7785, Jun. 2017. [Online]. Available: https://doi.org/10.2196/mental.7785

[16] B. Inkster, S. Sarda, and V. Subramanian, “An Empathy-Driven, Conversational Artificial Intelligence Agent (Wysa) for Digital Mental Well-Being: Real-World Data Evaluation Mixed-Methods Study,” *JMIR mHealth and uHealth*, vol. 6, no. 11, p. e12106, Nov. 2018. [Online]. Available: https://doi.org/10.2196/12106

[17] S. Gabrielli, S. Rizzi, G. Bassi, S. Carbone, R. Maimone, M. Marchesoni, and S. Forti, “Engagement and Effectiveness of a Healthy-Coping Intervention via Chatbot for University Students During the COVID-19 Pandemic: Mixed Methods Proof-of-Concept Study,” *JMIR mHealth and uHealth*, vol. 9, no. 5, p. e27965, May 2021. [Online]. Available: https://doi.org/10.2196/27965

[18] “MindTrails - About.” [Online]. Available: https://mindtrails.virginia.edu/calm/public/about

[19] B. Teachman, Apr. 2022, personal communication.

[20] J. W. Eberle, M. Boukhechba, J. Sun, D. Zhang, D. Funk, L. Barnes, and B. Teachman, “Shifting Episodic Prediction With Online Cognitive Bias Modification: A Randomized Controlled Trial,” PsyArXiv, Tech. Rep., Jul. 2020, type: article. [Online]. Available: https://psyarxiv.com/dg7z8/

[21] C. Beard, “Cognitive bias modification for anxiety: current evidence and future directions,” *Expert review of neurotherapeutics*, vol. 11, no. 2, pp. 299–311, Feb. 2011. [Online]. Available: https://doi.org/10.1586/ern.10.194

[22] S. Provoost, H. M. Lau, J. Ruwaard, and H. Riper, “Embodied Conversational Agents in Clinical Psychology: A Scoping Review,” *Journal of Medical Internet Research*, vol. 19, no. 5, p. e6553, May 2017. [Online]. Available: https://doi.org/10.2196/jmir.6553

[23] J. Cassell, “Embodied conversational interface agents,” *Communications of the ACM*, vol. 43, no. 4, pp. 70–78, Apr. 2000. [Online]. Available: https://doi.org/10.1145/332051.332075

[24] A. A. Abd-alrazaq, M. Alajlani, A. A. Alalwan, B. M. Bewick, P. Gardner, and M. Househ, “An overview of the features of chatbots in mental health: A scoping review,” *International Journal of Medical Informatics*, vol. 132, p. 103978, Dec. 2019. [Online]. Available: https://doi.org/10.1016/j.ijmedinf.2019.103978

[25] S. ter Stal, L. L. Kramer, M. Tabak, H. op den Akker, and H. Hermens, “Design Features of Embodied Conversational Agents in eHealth: a Literature Review,” *International Journal of Human-Computer Studies*, vol. 138, p. 102409, Jun. 2020. [Online]. Available: https://doi.org/10.1016/j.ijhcs.2020.102409

[26] K. Loveys, G. Sebaratnam, M. Sagar, and E. Broadbent, “The Effect of Design Features on Relationship Quality with Embodied Conversational Agents: A Systematic Review,” *International Journal of Social Robotics*, vol. 12, no. 6, pp. 1293–1312, Dec. 2020. [Online]. Available: https://doi.org/10.1007/s12369-020-00680-7

[27] M. McTear, “Conversational AI: Dialogue Systems, Conversational Agents, and Chatbots,” *Synthesis Lectures on Human Language Technologies*, vol. 13, no. 3, pp. 1–251, Oct. 2020, publisher: Morgan & Claypool Publishers. [Online]. Available: https://doi.org/10.2200/S01060ED1V01Y202010HLT048

[28] S. M. Schueller, K. N. Tomasino, and D. C. Mohr, “Integrating Human Support Into Behavioral Intervention Technologies: The Efficiency Model of Support,” *Clinical Psychology: Science and Practice*, vol. 24, no. 1, pp. 27–45, 2017. [Online]. Available: https://doi.org/10.1111/cpsp.12173

[29] “Juji.” [Online]. Available: https://juji.ai/login

[30] J. Sauro and J. R. Lewis, “Chapter 8 - Standardized usability questionnaires,” in *Quantifying the User Experience (Second Edition)*, Boston, Jan. 2016, pp. 185–248. [Online]. Available: https://doi.org/10.1016/B978-0-12-802308-2.00008-4

[31] T. S. Tullis and J. N. Stetson, “A Comparison of Questionnaires for Assessing Website Usability,” *UPA 2004*, p. 12, 2004.

[32] M. Brueckl and F. Heuer, “Coefficients of Interrater Reliability – Generalized for Randomly Incomplete Datasets,” Apr. 2022. [Online]. Available: https://cran.r-project.org/web/packages/irrNA/irrNA.pdf

[33] L. G. Portney and M. P. Watkins, *Foundations of clinical research: applications to practice*. Harlow: Prentice Hall, 2009, oCLC: 438786352.

[34] T. K. Koo and M. Y. Li, “A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research,” *Journal of Chiropractic Medicine*, vol. 15, no. 2, pp. 155–163, Jun. 2016. [Online]. Available: https://doi.org/10.1016/j.jcm.2016.02.012

[35] B. Chaix, J.-E. Bibault, A. Pienkowski, G. Delamon, A. Guillemassé, P. Nectoux, and B. Brouard, “When Chatbots Meet Patients: One-Year Prospective Study of Conversations Between Patients With Breast Cancer and a Chatbot,” *JMIR Cancer*, vol. 5, no. 1, p. e12856, May 2019. [Online]. Available: https://doi.org/10.2196/12856