

12-12-2022

Exploring User Experience with a Conversational Agent to Treat Depression in Youth: A Think-Aloud Study

Florian Onur Kuhlmeier

University of Greifswald, Karlsruhe Institute of Technology, florian.kuhlmeier@kit.edu

Ulrich Gnewuch

Karlsruhe Institute of Technology, ulrich.gnewuch@kit.edu

Lilli Metelmann

University of Tübingen, lilli.metelmann@web.de

Luise Bauch

University of Greifswald, luise.bauch@stud.uni-greifswald.de

Stefan Lüttke

University of Greifswald, stefan.luettk@uni-greifswald.de

Follow this and additional works at: <https://aisel.aisnet.org/sighci2022>

Recommended Citation

Onur Kuhlmeier, Florian; Gnewuch, Ulrich; Metelmann, Lilli; Bauch, Luise; and Lüttke, Stefan, "Exploring User Experience with a Conversational Agent to Treat Depression in Youth: A Think-Aloud Study" (2022). *SIGHCI 2022 Proceedings*. 10.

<https://aisel.aisnet.org/sighci2022/10>

This material is brought to you by the Special Interest Group on Human-Computer Interaction at AIS Electronic Library (AISeL). It has been accepted for inclusion in SIGHCI 2022 Proceedings by an authorized administrator of AIS Electronic Library (AISeL). For more information, please contact elibrary@aisnet.org.

Exploring User Experience with a Conversational Agent to Treat Depression in Youth: A Think-Aloud Study

Florian Onur Kuhlmeier
University of Greifswald, Karlsruhe
Institute of Technology
florian.kuhlmeier@kit.edu

Ulrich Gnewuch
Karlsruhe Institute of
Technology
ulrich.gnewuch@kit.edu

Luise Bauch
University of Greifswald
luise.bauch@stud.uni-greifswald.de

Lilli Feline Metelmann
University of Tübingen
lilli-feline.metelmann@student.uni-tuebingen.de

Stefan Lüttke
University of Greifswald
stefan.luettkel@uni-greifswald.de

ABSTRACT

Conversational agents are a promising digital health intervention that can mitigate help-seeking barriers for youth with depression to receive treatment. Although studies have shown sufficient acceptance, feasibility, and promising effectiveness for adults, not much is known about how youth experience interacting with conversational agents to improve mental health. Therefore, we conducted an exploratory study with 15 youth with to collect data on their interaction with a conversational agent prototype using the think-aloud protocol. We coded the material from the think-aloud sessions using an inductive approach. Our findings provide insights into how youth with depression interacted with the prototype. Participants frequently and controversially discussed the conversational agent's (1) personality and interaction style, (2) its functionality, and (3) the dialogue content with implications for the design of conversational agents to treat depression and future research.

Keywords

conversational agent, depression, user experience, think-aloud

INTRODUCTION

Depression in youth is a large-scale problem with significant individual and socioeconomic costs (Clayborne, Varin, & Colman, 2019). Yet, youth often do not receive treatment due to stigma, uncertainty, and the desire to solve problems on their own (Gulliver, Griffiths, & Christensen, 2010). Conversational agents are considered a promising digital mental health intervention that may mitigate these barriers. Mimicking human conversations may be the key to building a digital working alliance, which is crucial for effective psychotherapy (Darcy, Daniels, Salinger, Wicks, & Robinson, 2021). Studies on conversational agents have shown sufficient user experience and promising efficacy (Abd-Alrazaq, Rababeh, Alajlani, Bewick, & Househ, 2020). However, it

remains largely unknown how youth with depression experience interactions with conversational agents to treat depression. The applicability of related research with adult samples is limited as (1) depression in youth differs from depression in adulthood (Rice et al., 2019) and (2) youth interact with smartphones and conversational agents in different ways than adults (Andone et al., 2016; Huffman, 2014). These differences should be considered when designing conversational agents to treat depression. Therefore, the purpose of this study was to investigate how youth with depression experience interacting with a conversational agent to treat depressive symptoms. The results of this research-in-progress will be used to design a conversational agent to treat depression in youth.

RELATED WORK

Conversational agents are digital applications designed to mimic human conversational behavior (Dale, 2016). They are popular in sectors such as customer service, education, and healthcare (Følstad & Brandtzæg, 2017). Using conversational agents to diagnose and treat mental disorders has also been explored to expand current capacities and intervention types, as they are scalable, always available, and easily accessible (Vaidyam, Linggonegoro, & Torous, 2020). Most conversational agents to improve mental health are based on cognitive behavioral therapy – an effective and first-line psychotherapy approach for depression (Oud et al., 2019). The goal of cognitive behavioral therapy is to change dysfunctional thoughts and behaviors (Auerbach, Webb, & Stewart, 2016). Studies on conversational agents to improve mental health have shown sufficient acceptance and feasibility (Abd-Alrazaq et al., 2020; Vaidyam et al., 2020). However, conversational agents have primarily been designed for adults (Vaidyam et al., 2020) and preventing rather than treating mental disorders (Chan et al., 2022; Høiland, Følstad, & Karahasanovic, 2020).

METHODOLOGY

Study Design

The work presented in this article was part of an exploratory study to investigate the needs and preferences of youth for a conversational agent to treat depressive symptoms. First, we conducted a semi-structured interview on (1) problems and coping strategies for depression, (2) attitudes toward conversational agents to treat depression, and (3) design preferences. Second, we collected data on how users experienced interacting with a conversational agent prototype - which is the focus of this paper. We used the think-aloud method (Jaspers, Steen, Bos, & Geenen, 2004), where we asked users to narrate their thoughts and feelings while interacting with the prototype. At the beginning of the trial, the interviewer briefly introduced our conversational agent 'Cady', the content of the dialogue, and explained how the think-aloud method works. The users then interacted with the prototype via a laptop placed in front of them. We recorded the screen of the laptop and the user's audio with an internal function of the laptop. On average, the participants interacted with Cady for 15:38 minutes (range = 11:54 – 20:38). After the study, the recordings were transcribed for further analysis.

Participants

Participants qualified to take part in the study if they were between the ages of 14 and 17 years, owned a smartphone, and had previously been diagnosed with depression. They were excluded if they had suicidal ideation, psychosis, or low cognitive functioning. We advertised our study at local psychotherapists. 15 youth (14-17 years old; $M = 16$, $SD = 1.14$) with symptoms of depression participated in the study at the University of Tübingen. 13 participants identified as female and 2 as non-binary.

Conversational Agent: Cady

Participants were asked to interact with a conversational agent, developed with the prototyping software [botsociety](#), that guides users through a behavioral activation exercise. Figure 1 shows a screenshot demonstrating the interaction between the conversational agent and the user. Behavioral activation is a core component of cognitive behavioral therapy to treat depression in youth (Oud et al., 2019). It aims to increase the engagement of patients with pleasurable activities. The conversation consisted of the following sections: (1) introduction, (2) mood check with adaptive responses, (3) psychoeducation on the relationship between behavior, thoughts, and feelings, (4) finding and planning pleasant activities, (6) advice on how to overcome barriers to performing activities, and (7) feedback and goodbye. The conversation mainly consisted of buttons with predefined answer alternatives and a few free text input fields. We named the conversational agent 'Cady' and did not specify age, gender, or other demographic characteristics to prevent specific

demographic characteristics from influencing the results of the study.

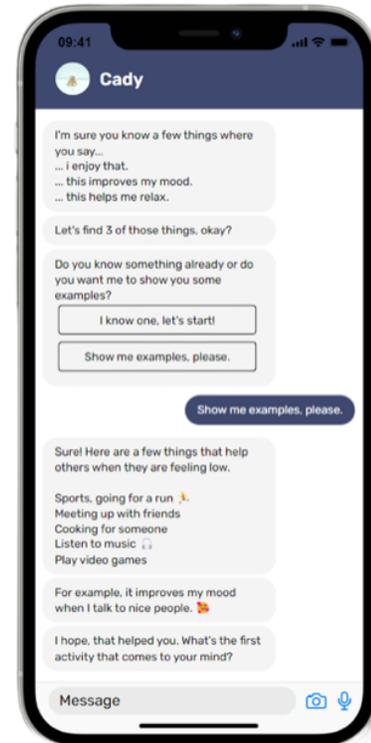


Figure 1. Screenshot of Cady.

Data Analysis

Two coders analyzed the think-aloud sessions using content analysis according to Mayring (2004) with the software [QCAmap](#). We chose inductive coding to achieve the most unbiased and thorough description of the data, which we considered important due to the exploratory nature of the study. Both coders first coded the material independently and then agreed on overarching categories. When codes differ due to the inductive approach, we report the codes of the first coder.

RESULTS

Cady's (1) personality and interaction style, (2) functionality, and (3) dialogue content emerged as the main categories.

Personality and Interaction Style

The personality and interaction style of Cady were frequently discussed. Most participants characterized the friendly and personal interaction style as one of its main strengths. One participant said: *'I like that you communicate in a friendly way, like an internet friend.'* Another participant was pleasantly surprised by Cady's reaction after finding three pleasant activities: *'That's sweet! It is good that you can tell that Cady is happy.'*

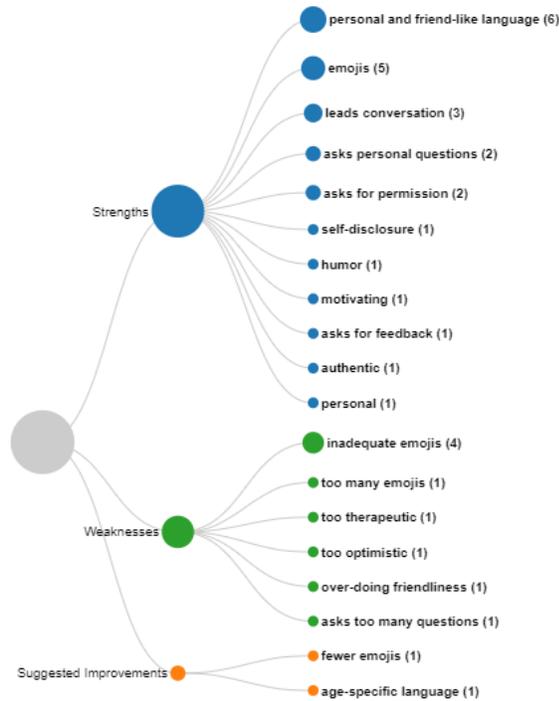


Figure 2. Categories and codes regarding personality and interaction style.

Several participants also appreciated Cady’s use of emoji. Others, however, criticized too many or inadequate emojis. A participant explained their criticism: ‘Here, for example, is a rocket emoji. And I would not use a rocket emoji myself.’ Consequently, this participant would have preferred fewer emojis. Another participant commented on the language as too therapeutic: ‘It was relatively authentic, but you also notice that it is not written for my age. But otherwise, if you see it as a therapy conversation, it was good.’ The participant wished for more age-specific language, such as short sentences: ‘So, this was therapeutic, but if it is supposed to be like a friend, I think it is better if it is written like my age. Short sentences, not completed.’ Furthermore, three participants appreciated that Cady led the conversation, which was perceived as less effort: ‘I think it’s good that the computer leads the conversation, and you don’t have to do too much yourself.’ Two participants were pleased that Cady asked personal questions. For example, a participant approved that Cady had asked if she had experienced becoming less active herself: ‘I just think it’s good that she is asking if I know something like that or haven’t experienced it yet.’ However, one participant concluded that Cady asked too many questions. Although two participants appreciated Cady’s motivational style and humor, another participant thought that Cady was too optimistic. One participant was pleased that Cady asked for feedback after the conversation. Lastly, one participant appreciated that Cady disclosed personal information such as enjoying chats with nice people.

Functionality

The participants pointed out several aspects of Cady’s functionality. Seven participants enjoyed interacting through predefined responses. One participant stated: ‘I liked that there are predefined responses and I don’t need to think of a response every time.’ Six participants appreciated optional reminders to carry out the planned activities, while two would have liked to personalize the reminders. Another participant suggested automatic rather than optional reminders. Four participants appreciated free text responses, while one participant suggested including more free text input. Furthermore, one participant liked how Cady was able to understand their responses: ‘I think it’s really impressive that Cady understands what I write and then responds to it.’ Yet, three other participants were frustrated by inadequate responses. One participant said: ‘The message I sent does not even make sense. But Cady does not realize that.’



Figure 3. Categories and codes regarding functionality.

Another participant pointed out that the conversation felt impersonal/ not tailored to them and their needs. Four participants pointed out that Cady sent its messages too fast. Two participants suggested lowering Cady’s typing speed and including short breaks between messages. On the contrary, one participant was happy with the speed of the messages, while another participant felt that the pause between messages was too long. One participant suggested a typing indicator to make pauses more like human-to-human chats. Four participants liked the simple, easy to use, and familiar user chat interface, while one participant suggested an option to personalize the user interface.

Dialogue Content

Participants frequently commented on the content of the dialogue, primarily its selection and presentation. The participants noted how useful and easy to implement the content was. One participant said: *‘I think the prototype has already helped me, so that I can become more active again’*. The participants also appreciated the mood check-in. One participant stated: *‘I think it's good that Cady asks me how I'm doing.’* Another person was disappointed that there was no option to just talk about something that was on their mind and would have preferred such a feature: *‘If you want to get something off your chest, it is good to have the opportunity right at the beginning, during the conversation or at the end, so that you read and participate, but can still say something afterwards.’*

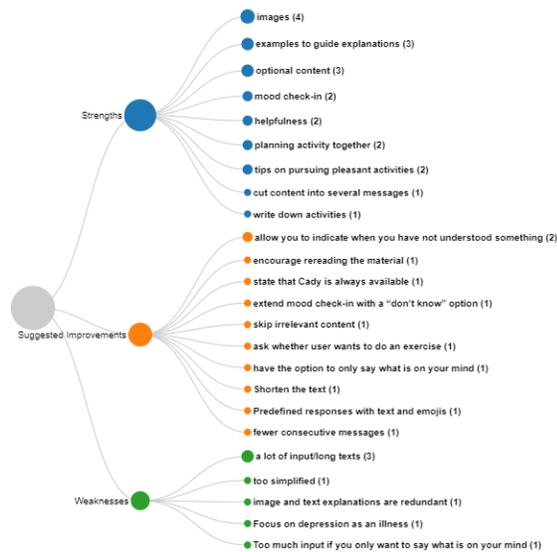


Figure 4. Categories and codes regarding the dialogue content.

Four participants suggested personalized content. For example, they would like Cady to ask if the user wants to learn about specific a topic (e.g., behavioral activation) or if the user is affected by the presented content to determine its relevance. Two participants appreciated that Cady chooses and plans activities with them and that Cady provides tips on how to carry out the activities. Four participants enjoyed the images and examples, which helped to understand the content. The participants also noted that it was helpful that they had to text Cady three activities they find pleasant. However, a participant pointed out that Cady made choosing, planning, and doing a pleasant activity seem easier than it really is: *‘This is also the case with my therapist, they simplify a lot. Like, it is easy to do something pleasant. But sometimes it is difficult to do anything at all.’* In addition, one participant stated that breaking the content down into several messages rather than a single text made it easier to understand and more interactive. However, three participants said that there was too much information. One participant suggested that Cady should encourage users to read the content again to better

memorize it. Two participants proposed to improve the content by shortening the text, receiving fewer consecutive messages, and including emojis and text in every predefined response.

DISCUSSION

In our exploratory study with 15 youth with symptoms of depression, we found that conversational agents may serve as an appropriate digital health intervention, as indicated by the various strengths that the participants pointed out during the think-aloud sessions. Our results also revealed several challenges that provide implications for the design of conversational agents for this target group.

First, our results show contradictory and controversial discussions about several design aspects. For example, language and emoji use was frequently and contradictorily discussed, which emphasizes that personality and interaction style are crucial. This result underscores the need to thoroughly design and evaluate the character of the conversational agent. However, it remains largely unknown, how to provide a conversational agent with personality traits through language use, emojis, and gifs. Despite promising first attempts to guide and systematize this process (Ahmad, Siemon, Gnewuch, & Robra-Bissantz, 2022; Nißen et al., 2022), comprehensive and specific guidelines are needed. Our results also show that the success of conversational agents is not only based on technological advances in natural language processing, consistent with previous research on chatbots in other domains (Schuetzler et al., 2020); success also depends on sufficiently meeting user’s requirements and preferences beyond conversational abilities. Contradictions regarding the personality and interaction style indicate the difficulty in satisfying the needs of all users, suggesting the use of personalization (Kocaballi et al., 2019). For example, designers could provide users with the option to disable the use of emojis.

Second, a counterintuitive finding was that several participants liked to interact with Cady by selecting predefined responses. Intuitively, we would have assumed that users prefer more human-like communication, which has also been demonstrated in customer service (Diederich, Brendel, Lichtenberg, & Kolbe, 2019). Some participants explained their preference because the predefined responses were more efficient. It seems that implementing predefined responses is advantageous for highly structured dialogues such as teaching behavioral activation. On the other hand, more flexible dialogues to address individual issues are complex to implement in a highly structured way. Future research should address how the advantages of predefined responses and advanced natural language processing can be combined to deliver the best user experience. Third, a high degree of anthropomorphism was salient in the think-aloud sessions. Participants repeatedly used pronouns (e.g., he/she) to refer to Cady even though it was specifically designed to be gender neutral (i.e., a gender-neutral name and a robot avatar). In addition,

different personality traits were attributed to Cady. Most importantly, one participant had the impression that their message made Cady happy. Interestingly, the diverse statements indicate that there are different ways in which anthropomorphism functions. Personalization appears to be an appropriate strategy to address these different ways of increasing anthropomorphism.

Third, our participants showed different perceptions of conversational capabilities. Although one participant was impressed that Cady understood what they were saying, others criticized inadequate responses. Human characteristics, such as expectations and previous experience with conversational agents, may influence the user experience. In line with the agenda for conversational agent research developed by Diederich et al. (2022), future research should investigate how user characteristics influence the interaction with conversational agents.

LIMITATIONS

Our study has three main limitations. First, our sample consists predominantly of female participants. Although depression is substantially more prevalent among women (Seedat et al., 2009), future studies need to include male participants to provide a more holistic understanding of youth's user experience with conversational agents to treat depression. Second, our conversational agent prototype focused on behavioral activation, which is only one part of cognitive behavioral therapy. Although most results appear generalizable, think-aloud data on other therapeutic exercises is needed to test generalizability. Lastly, Think-aloud has proven valuable in providing rich and detailed data, but does provide an overall assessment. Future research should therefore complement a think-aloud approach with open-ended questions, established user experience questionnaires and interaction data.

CONCLUSION

This article provides insight into how youth with symptoms of depression interact with a conversational agent designed to deliver therapeutic content. Our results suggest that the participants enjoyed the interaction and discussed (1) Cady's personality and interaction style, (2) its functionality, and (3) the dialogue content. Our next step will be to fully analyze and publish our study on needs and preferences for a conversational agent to treat depression in youth. In our future research, we will extend the prototype to several components of cognitive behavioral therapy and complement the think-aloud with semi-structured interviews and user experience surveys to follow a mixed methods approach.

REFERENCES

1. Abd-Alrazaq, A. A., Rababeh, A., Alajlani, M., Bewick, B. M., & Housch, M. (2020). Effectiveness and Safety of Using Chatbots to Improve Mental Health: Systematic Review and Meta-Analysis. *Journal of Medical Internet Research*, 22(7).
2. Ahmad, R., Siemon, D., Gnewuch, U., & Robra-Bissantz, S. (2022). Designing Personality-Adaptive Conversational Agents for Mental Health Care. *Information Systems Frontiers*.
3. Andone, I., Blaszkiewicz, K., Eibes, M., Trendafilov, B., Montag, C., & Markowitz, A. (2016). How age and gender affect smartphone usage. In *Proceedings of the 2016 ACM international joint conference on pervasive and ubiquitous computing: adjunct* (pp. 9–12).
4. Auerbach, R. P., Webb, C. A., & Stewart, J. G. (2016). *Cognitive behavior therapy for depressed adolescents: A practical guide to management and treatment*. Routledge.
5. Chan, W. W., Fitzsimmons-Craft, E. E., Arielle C. Smith, Firebaugh, M. L., Fowler, L. A., Bianca DePietro, ... Jacobson, N. C. (2022). The Challenges in Designing a Prevention Chatbot for Eating Disorders: Observational Study.
6. Clayborne, Z. M., Varin, M., & Colman, I. (2019). Systematic Review and Meta-Analysis: Adolescent Depression and Long-Term Psychosocial Outcomes. *Journal of the American Academy of Child & Adolescent Psychiatry*, 58(1), 72–79.
7. Dale, R. (2016). The return of the chatbots. *Natural Language Engineering*, 22(5), 811–817.
8. Darcy, A., Daniels, J., Salinger, D., Wicks, P., & Robinson, A. (2021). Evidence of Human-Level Bonds Established With a Digital Conversational Agent: Cross-sectional, Retrospective Observational Study. *JMIR Formative Research*, 5(5), e27868.
9. Diederich, S., Brendel, A. B., Lichtenberg, S., & Kolbe, L. (2019). Design for fast request fulfillment or natural interaction? Insights from an experiment with a conversational agent.
10. Diederich, S., Brendel, A., Morana, S., & Kolbe, L. (2022). On the Design of and Interaction with Conversational Agents: An Organizing and Assessing Review of Human-Computer Interaction Research. *Journal of the Association for Information Systems*, 23(1), 96–138.
11. Følstad, A., & Brandtzæg, P. B. (2017). Chatbots and the new world of HCI. *Interactions*, 24(4), 38–42.
12. Gulliver, A., Griffiths, K. M., & Christensen, H. (2010). Perceived barriers and facilitators to mental health help-seeking in young people: a systematic review. *BMC Psychiatry*, 10(1), 113.
13. Høiland, C. G., Følstad, A., & Karahasanovic, A. (2020). Hi, Can I Help? Exploring How to Design a Mental Health Chatbot for Youths. *Human Technology*, 16(2), 139–169.
14. Huffman, S. (2014). OMG! Mobile voice survey reveals teens love to talk. Retrieved October, 14.

15. Jaspers, M. W. M., Steen, T., Bos, C. van den, & Geenen, M. (2004). The think aloud method: a guide to user interface design. *International Journal of Medical Informatics*, 73(11), 781–795.
16. Kocaballi, A. B., Berkovsky, S., Quiroz, J. C., Laranjo, L., Tong, H. L., Rezazadegan, D., ... Coiera, E. (2019). The Personalization of Conversational Agents in Health Care: Systematic Review. *Journal of Medical Internet Research*, 21(11), e15360.
17. Mayring, P. (2004). Qualitative content analysis. *A companion to qualitative research*, 1(2), 159–176.
18. Nißen, M., Rügger, D., Stieger, M., Flückiger, C., Allemann, M., Wangenheim, F. v., & Kowatsch, T. (2022). The Effects of Health Care Chatbot Personas With Different Social Roles on the Client-Chatbot Bond and Usage Intentions: Development of a Design Codebook and Web-Based Study. *Journal of Medical Internet Research*, 24(4), e32630.
19. Oud, M., de Winter, L., Vermeulen-Smit, E., Bodden, D., Nauta, M., Stone, L., ... Stikkelbroek, Y. (2019). Effectiveness of CBT for children and adolescents with depression: A systematic review and meta-regression analysis. *European Psychiatry: The Journal of the Association of European Psychiatrists*, 57, 33–45.
20. Rice, F., Riglin, L., Lomax, T., Souter, E., Potter, R., Smith, D. J., ... Thapar, A. (2019). Adolescent and adult differences in major depression symptom profiles. *Journal of Affective Disorders*, 243, 175–181.
21. Schuetzler, R. M., Grimes, G. M., & Scott Giboney, J. (2020). The impact of chatbot conversational skill on engagement and perceived humanness. *Journal of Management Information Systems*, 37(3), 875–900.
22. Seedat, S., Scott, K. M., Angermeyer, M. C., Berglund, P., Bromet, E. J., Brugha, T. S., ... Kessler, R. C. (2009). Cross-national associations between gender and mental disorders in the World Health Organization World Mental Health Surveys. *Archives of General Psychiatry*, 66(7), 785–795.
23. Vaidyam, A. N., Linggonegoro, D., & Torous, J. (2020). Changes to the Psychiatric Chatbot Landscape: A Systematic Review of Conversational Agents in Serious Mental Illness: Changements du paysage psychiatrique des chatbots: une revue systématique des agents conversationnels dans la maladie mentale sévère. *The Canadian Journal of Psychiatry*, 0706743720966429.