frontiers | Frontiers in Digital Health

TYPE Original Research
PUBLISHED 07 March 2023
DOI 10.3389/fdgth.2023.1034724

OPEN ACCESS

EDITED BY
Elizabeth Eikey,
University of California, San Diego, United States

REVIEWED BY
Lea Martin,
Ochsner Health System, United States,
Julie Prescott,
The University of Law, United Kingdom

*CORRESPONDENCE
Joonas Moilanen
✉ joonas.moilanen@student.oulu.fi

SPECIALTY SECTION
This article was submitted to Digital Mental Health, a section of the journal Frontiers in Digital Health

RECEIVED 01 September 2022
ACCEPTED 10 February 2023
PUBLISHED 07 March 2023

CITATION
Moilanen J, van Berkel N, Visuri A, Gadiraju U, van der Maden W and Hosio S (2023) Supporting mental health self-care discovery through a chatbot.
Front. Digit. Health 5:1034724.
doi: 10.3389/fdgth.2023.1034724

Supporting mental health self-care discovery through a chatbot

Joonas Moilanen[1]*, Niels van Berkel[2], Aku Visuri[1], Ujwal Gadiraju[3], Willem van der Maden[4] and Simo Hosio[1]

[1]Faculty of Information Technology and Electrical Engineering, Center for Ubiquitous Computing, University of Oulu, Oulu, Finland, [2]Department of Computer Science, Human-Centered Computing, Aalborg University, Aalborg, Denmark, [3]Faculty of Electrical Engineering, Mathematics and Computer Science, Web Information Systems, Delft University of Technology, Delft, Netherlands, [4]Faculty of Industrial Design Engineering, Delft University of Technology, Delft, Netherlands

Good mental health is imperative for one's wellbeing. While clinical mental disorder treatments exist, self-care is an essential aspect of mental health. This paper explores the use and perceived trust of conversational agents, chatbots, in the context of crowdsourced self-care through a between-subjects study ($N = 80$). One group used a standalone system with a conventional web interface to discover self-care methods. The other group used the same system wrapped in a chatbot interface, facilitating utterances and turn-taking between the user and a chatbot. We identify the security and integrity of the systems as critical factors that affect users' trust. The chatbot interface scored lower on both these factors, and we contemplate the potential underlying reasons for this. We complement the quantitative data with qualitative analysis and synthesize our findings to identify suggestions for using chatbots in mental health contexts.

KEYWORDS

conversational agent (CA), chatbot, conversational user interface (CUI), mental health, self-care, human-computer trust, human-computer interaction

1. Introduction

Good mental health is imperative for one's general wellbeing. Conversely, mental disorders cause tremendous social (1) and economic (2) burdens worldwide. Higher education students are especially vulnerable, as they are typically at the peak onset of many mental disorders, such as depression and anxiety (3). However, a staggering number of students suffering from symptoms never seek help, and many seek help far too late in the process (4). To this end, support from one's community has been identified as a valuable avenue to explore as a complementary mechanism to traditional healthcare and clinical interventions (5). However, knowledge is often sparsely shared within the community due to stigma (6). Novel research approaches and support mechanisms with a lower barrier for participation are required to address this. In addition to helping people with existing mental health conditions, it is important to maintain healthy mental wellbeing for those not feeling particularly ill. Support mechanisms have proved effective for preventive approaches as well (7).

One approach currently investigated for mental health is *self-care*. Self-care, in general, refers to how people take care of their wellbeing or a mental health condition on their own, either using the information found online or as instructed by their caretakers (8). A community sharing a similar burden can be an excellent resource for self-care methods. While various other means of serving these methods exist, researchers are currently actively looking into the affordances of chat-based conversational agents, chatbots, due to their inherent relatability and rapidly increasing interaction capabilities (see, e.g., (9,10)).

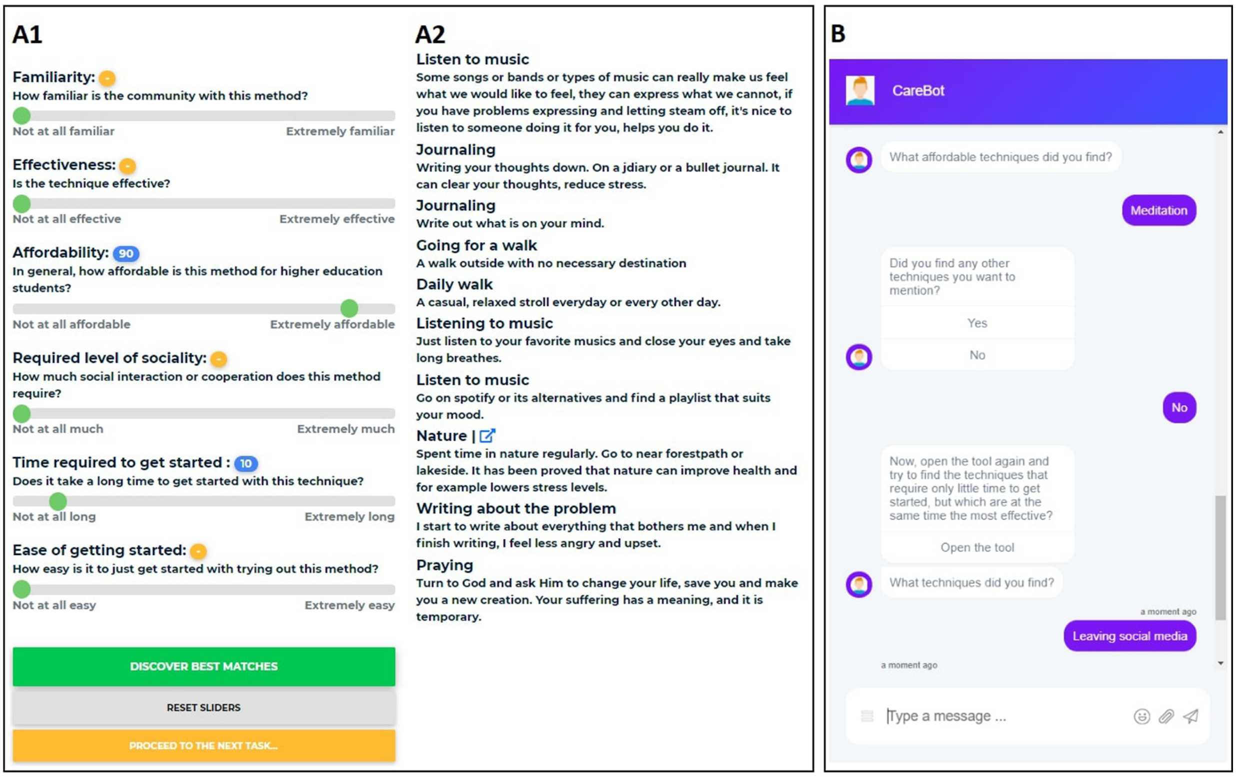
In our earlier work (unpublished in academic venues), we have crowdsourced an extensive list of self-care methods among the higher education community to uncover how students maintain and improve their mental health. These methods include, for example, meditation, spending time with others, volunteering, and working out at a gym, with additional methods presented in Figure 1. The students have also cross-evaluated each other's contributions across a set of specific criteria. In this paper, we used this data to bootstrap a decision support system (DSS) that allows for discovering suitable self-care methods through an online user interface (UI) and by using the same criteria that were used to bootstrap the DSS (see Figure 1). To explore the potential use of chatbots in serving the DSS and trust in the system, we offered the DSS UI to 80 higher education students in a between-subjects study. The study groups consist of two groups of 40 participants through A) a standalone online DSS, and B) the DSS embedded in a narrative served by a conversational interface (see Figure 1).

In this work, we set out to find factors which affect the formed trust between a mental health chatbot and the user. We offer the users hundreds of crowdsourced methods for mental health self-care in an interface embedded in a chatbot conversation and compare that to a traditional web interface. We hypothesize that providing the users with clear instructions and interactive conversation alongside the method discovery could lead to improved trust towards the system.

Our findings highlight that the participants interacting with the chatbot report lower perceived system security and integrity but no significant difference in the overall trust between the DSS group. The human-like behaviour of the chatbot also appears to affect trust for individual participants. Based on our findings, we argue that improving the (perceived) security and integrity of the chatbot will help design more effective chatbots for mental health. Furthermore, we find that using a chatbot for mental health self-care method discovery shows promise, with several participants stating their fondness towards the chatbot. While research in mental health chatbots and their trust is plentiful, we provide contributions to direct comparisons of two systems and how to further improve the trust towards them. In addition, we provide information on how viable these kind of crowdsourced methods are in digital healthcare.

1.1. Related work

Chatbots mimic human conversation using voice recognition, natural language processing, and artificial intelligence. Initial versions of chatbots operated purely through text-based communication, aiming to provide intelligent and human-like replies to its users (11). Over the past decade, chatbots have grown in popularity, and together with voice-activated conversational agents such as Apple's Siri, they have become part of everyday life (11).



Familiarity: -
How familiar is the community with this method?
Not at all familiar — Extremely familiar

Effectiveness: -
Is the technique effective?
Not at all effective — Extremely effective

Affordability: 90
In general, how affordable is this method for higher education students?
Not at all affordable — Extremely affordable

Required level of sociality: -
How much social interaction or cooperation does this method require?
Not at all much — Extremely much

Time required to get started : 10
Does it take a long time to get started with this technique?
Not at all long — Extremely long

Ease of getting started: -
How easy is it to just get started with trying out this method?
Not at all easy — Extremely easy

Listen to music
Some songs or bands or types of music can really make us feel what we would like to feel, they can express what we cannot, if you have problems expressing and letting steam off, it's nice to listen to someone doing it for you, helps you do it.

Journaling
Writing your thoughts down. On a jdiary or a bullet journal. It can clear your thoughts, reduce stress.

Journaling
Write out what is on your mind.

Going for a walk
A walk outside with no necessary destination

Daily walk
A casual, relaxed stroll everyday or every other day.

Listening to music
Just listen to your favorite musics and close your eyes and take long breathes.

Listen to music
Go on spotify or its alternatives and find a playlist that suits your mood.

Nature |
Spent time in nature regularly. Go to near forestpath or lakeside. It has been proved that nature can improve health and for example lowers stress levels.

Writing about the problem
I start to write about everything that bothers me and when I finish writing, I feel less angry and upset.

Praying
Turn to God and ask Him to change your life, save you and make you a new creation. Your suffering has a meaning, and it is temporary.

What affordable techniques did you find?

Meditation

Did you find any other techniques you want to mention?
Yes
No

No

Now, open the tool again and try to find the techniques that require only little time to get started, but which are at the same time the most effective?
Open the tool

What techniques did you find?

a moment ago
Leaving social media

a moment ago

FIGURE 1
Snippets of our two platforms used in the study. (A) The system for mental health self-care discovery with sliders for different criteria (A1) and recommended methods (A2), based on the slider selections. (B) The chatbot with examples of asking found methods and presenting the user with a new task.

1.1.1. Chatbots in mental health

Mental health has been defined by the World Health Organization (WHO) as "*a state of wellbeing in which the individual realizes his or her abilities, can cope with the everyday stresses of life, can work productively and fruitfully, and can make a contribution to his or her community*" (12). Mental health is a suitable context for chatbots due to their ability to provide dynamic interaction without relying on a professional's availability (13), and the potential for chatbots to provide empathic responses (14).

In this article, we specifically focus on self-care for mental health. Self-care is used both to manage long-term conditions and to prevent future illnesses and has been identified as a critical approach to supporting independence, providing control to the patient rather than solely relying on a clinician, and reducing reliance on an overburdened healthcare system (15). The application of chatbots as self-care tools is a relatively under-explored opportunity, with many open questions regarding identifying, monitoring, and evaluating self-care methods. Here, we focus on using a chatbot as a tool for *discovering* self-care solutions in mental health.

Using chatbots in mental health care has grown in popularity in recent years (16,17) and point to the opportunity to support users in long-term self-care development and effectively communicate goals in response to prior and new user needs. While chatbots are not suitable to provide the users with actual clinical intervention, they are an excellent way to provide mental health counselling, such as presenting the users with various self-care methods to help them improve their mental health (18). While most research for chatbots offering self-care focuses on young people, it has been shown to be effective for older adults, as well, as is shown by Morrow et al., who present a framework for the design of chatbots on health-related self-care for older adults (9). As is found in the review by Abd-Alrazaq et. al. (17), using chatbots for these kind of purposes can improve their mental health, but should commonly be used as an adjunct to intervention with a healthcare professional. In addition, using chatbots in mental health is not without its risks. One of the most crucial things to be taken into consideration when designing a mental health chatbot is its ability to reply accordingly to the user's messages; for example, poorly managing the responses to suicidal behaviour might lead to serious consequences (16).

Various factors affect the overall effectiveness of a chatbot, but in this research we focus specifically on the user's trust towards the chatbot. A vital prerequisite to offering mental health through chatbots is to build a level of trust between the user and the chatbot. Müller et al. find that a lack of trust in chatbots results in reduced uptake of these digital solutions (19). Furthermore, there are several factors to be taken into account to further enhance the perceived trust for chatbots, most notably, the chatbot's personality, knowledge and cognition have shown to increase trust (20). Previous research shows promise in building trust between a chatbot offering counselling and the user (21) and that the use of conversational interfaces as compared to a conventional web interface can lead to better performance and user experience (22). Recent work by Gupta et al. shows a similar setting to ours and an increase to trust when using a chatbot as opposed to a traditional web interface for housing recommendations (23).

1.1.2. Trust in computer systems

Trust and its formation have been important topics in automated (24,25) and online systems (26,27), as well as specifically in chatbots (28). The definition of trust varies, but for this paper, we define it as "*the attitude that an agent will help achieve an individual's goals in a situation characterized by uncertainty and vulnerability*" (25). An agent can refer to a human individual but also a chatbot. Zhang & Zhang highlight the many factors that affect trust, stating that trusting behaviour is formed from an individual's trust beliefs, attitude towards trust, and trust intention, which are further influenced by, for example, external environmental factors (26). Work by Tolmeijer et al. shows that trust develops slowly, with the user's initial trust impression having a large anchoring effect (29).

To measure trust, we used the "trust in the automation" scale by Jian et al. (24), which is one of the most widely used scales for measuring trust. Several other scales for measuring trust exist, but as most revolve around the same core topics and the scale by Jian offers easily interpretable results, we deemed this scale suitable for our purposes. Extensive research by Nordheim (30) shows that trust in chatbots is formed with factors such as risk, brand, and expertise, which are covered in our survey questions, presented in **Table 1**. In addition, the concept of trust seems to be similar for both human-human and human-machine situations (24).

2. Materials and methods

2.1. Apparatus

2.1.1. Crowdsourcing decision support system

We used a publicly available lightweight crowdsourcing tool developed by Hosio et al. (31) to collect and assess mental health self-care methods. The tool is implemented using HTML,

TABLE 1 Questions of the 'trust in automation' scale used to measure trust.

#	Question
Q1	The system is deceptive
Q2	The system behaves in an underhanded manner
Q3	I am suspicious of the system's intent, action, or outputs
Q4	I am wary of the system
Q5	The system's actions will have a harmful or injurious outcome
Q6	I am confident in the system
Q7	The system provides security
Q8	The system has integrity
Q9	The system is dependable
Q10	The system is reliable
Q11	I can trust the system
Q12	I am familiar with the system

All items use a 7-point Likert scale. These questions are derived from the work by Jian et. al. (24).

Javascript, PHP, and MySQL, and can be deployed on any website using a standard HTML iFrame tag. We will refer to this tool as the Decision Support System or **DSS**. The DSS has three main components; users can search for methods, rank existing methods, and input new methods to the system. A similar system framework has been adapted to other studies, e.g., for crowdsourcing treatments for low back pain (32) and personalized weight-loss diets (33).

In the context of this work, the study participants use the search component. Using the decision support interface, as depicted in **Figure 1A**, participants can search for self-care methods through a configuration of six different sliders that adjust familiarity, effectiveness, affordability, required level of sociality, the time required to get started, and ease of getting started. After the sliders have been adjusted, the tool presents the participant with the mental health self-care methods that best match the criteria configuration. Each of the six characteristics was rated by the users of the tool during the data collection.

Data collection of the shown methods was conducted before this study. The tool was made publicly available, and it was used to collect new mental self-care methods from its users and requested users to rate and validate pre-existing methods in the system. Methods were collected from over 900 participants, and over 30 000 individual ratings for hundreds of different self-care methods were obtained during the study. In addition to these methods, the participants were asked for open feedback on where, how, and why they seek self-care-related information.

As these components of the tool are not the focus of this article, we point the reader to (32) for more information about its functionalities.

2.1.2. Chatbot implementation

In this study, we were interested in exploring whether wrapping the tool in a conversational interface where participants could converse with an agent would affect the perceived trust or other aspects of the system. We purchased a license to BotStar[1] to use as the chatbot. BotStar supports opening external URLs in a full-screen modal popup as part of the conversation flow, which is how we embedded the DSS among the scripted conversation. A snippet of the used conversation script can be seen in **Figure 1B**. The chatbot was fully implemented via BotStar and was launched on a remote WordPress page.

2.1.3. Post-task survey

After completing the three tasks of using the DSS either through the web interface and instructions, or while interacting with the chatbot, each participant responds to a final questionnaire through Google Forms. The final questionnaire contains the trust in automation scale items (see **Table 1**) using a 7-point Likert Scale (1 = Strongly Disagree, 2 = Disagree, 3 = Somewhat Disagree, 4 = Neutral, 5 = Somewhat Agree, 6 = Agree, 7 = Strongly Agree) and three open-ended follow-up questions;

- **F1**: We asked you to search for mental health self-care methods with criteria of your own choice. What are your thoughts on the results?
- **F2**: What kind of support do you expect from a system offering mental health self-care techniques?
- **F3**: What features affected your trust in the system?

2.1.4. System overview

The full study system setup consists of the following components, and is presented in detail in **Figure 2**:

- **Prolific**: The study is deployed on the Prolific[2] crowdsourcing platform, where participants are given instructions and links to proceed to their study tasks.
- **DSS Platform**: The decision support system is deployed as a web interface on our remote server.
- **Chatbot**: The chatbot is self-hosted using BotStar on our remote server. For the chatbot participant group, the DSS platform is opened inside the chatbot using embedded web views.
- **Final Questionnaire**: The final questionnaire for the participants is deployed using Google Forms.

2.2. Experimental setup and protocol

For this article, the two study groups are named and referred to as follows:

- **CB group**: Group using the chatbot that wraps the online DSS
- **WEB group**: Group using only the online DSS.

Participants were recruited from Prolific, an online crowdsourcing platform. The participants were pre-filtered to higher education students using the platform's quality control mechanisms. Participants were rewarded USD2.03–USD2.54 based on a task duration of 11–15 min. Participants are anonymous; thus, no approval for human subject research was needed beyond our project-wide approval from our University's ethics board. Participants were asked to give consent at the beginning of the study and could terminate their participation at any point of the study.

We asked participants from both groups to look for self-care methods three times; first to explore methods that are the most affordable, then methods requiring only a little time to get started but which are at the same time the most effective, and finally methods that would suit the participants' own needs the best. In the CB group, these instructions were given by the chatbot, and in the WEB group, the instructions were given on top of the web page in a simple notice box.

In the CB group, we focused on making the narrative realistic and neutral tone. To this end, the chatbot walks the participants through the process of discovering self-care methods, providing them with

[1] https://botstar.com/

[2] https://www.prolific.co/

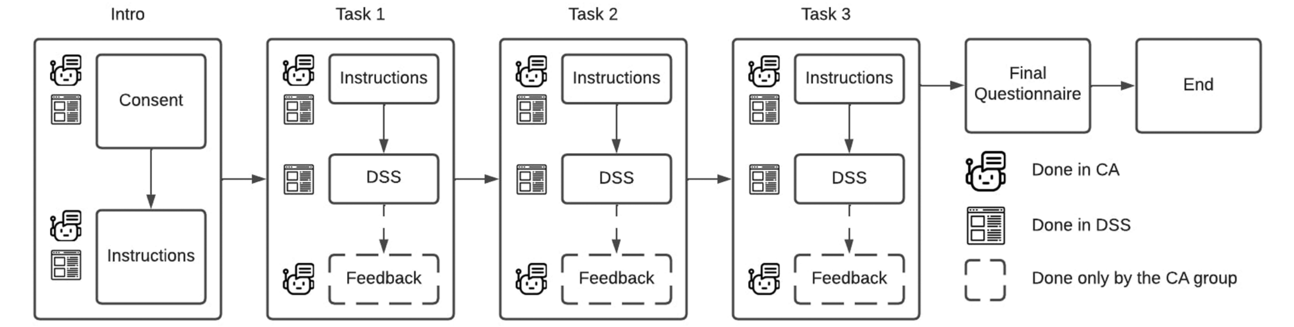


FIGURE 2
Flow diagram of the study design for the two study groups. The WEB group does every part of the study in the DSS, while the CB group switches from the chatbot to embedded DSS only for task completion. The chatbot briefly asks, what methods the user found from the CB group.

detailed instructions on what to do next (Figure 1B). To add a level of human-like conversation between the chatbot and the participant, the participant can be called by a nickname. The chatbot greets them at the beginning and expresses their gratitude at the end of the session. In the middle of the conversation, the participants are prompted to complete the same tasks as in the other experimental condition. At the end of each task, the chatbot asks what methods the participant found during this round. To ensure comparability between conditions, participants use the same tool for identifying a suitable self-care method.

After completing the three aforementioned tasks of identifying self-care methods, participants were directed to the final questionnaire. To evaluate the trust and credibility in both conditions, we use the 12-item questionnaire for trust between people and automation proposed by Jian et al. (24) presented in Table 1. This scale has been created using large amounts of empirical data and helps us understand how different system characteristics affect users' trust. In addition, we ask three open-ended questions to determine what the users think of the methods recommended for them, what kind of support they expect from a system offering mental health self-care methods, and which factors affected their trust in the system. This study setup is presented in Figure 2

3. Results

3.1. Participant demographics

43 of the participants identify themselves as male and 37 female. The average age was 23.26 (SD = 4.59) years. 72 of the participants reside in Europe, the two most represented countries being Portugal ($N = 20$) and the UK ($N = 13$). 52 participants were undergraduate students, 24 graduate students, and 5 doctoral students. The mean age for the two groups was 24.1 (SD = 5.89) years for the CB group versus 22.43 (SD = 2.44) for the WEB group.

3.2. Quantitative analysis

Scores for the trust in automation scale are presented in Figure 3. The CB group reports approximately half a point lower values for both the security and integrity question than the WEB group with a significant difference ($p < 0.05$). For the other questions, the difference between the two groups is similar, but with $p > 0.05$, so these findings are not statistically significant. In addition, we found a significant difference in positive trust, with the CB group reporting approximately 0.5 lower scores compared to the WEB group ($p < 0.05$).

3.2.1. Trust scores

Trust scores for the *trust in automation* survey were determined by taking the combined means of the positively worded items (*Q6-Q12*) and reverse-scored negatively worded items (*Q1-Q5*). Similarly, trust scores for positive and negative items were determined. For the overall and positive trust scores, a higher value indicates a bigger trust, whereas, for the negative trust scores, a lower value signifies a bigger trust (24).

Positive trust: CB: 4.07 (SD = 1.25), WEB: 4.57 (SD = 0.95). $t = -1.98$, $p = 0.03$, *Cohen's d*: 0.45.

Negative trust: CB: 2.56 (SD = 1.20), WEB: 2.7 (SD = 1.33). $t = -0.51$, $p = 0.31$, *Cohen's d*: 0.11.

Overall trust: CB: 4.64 (SD = 1.12), WEB: 4.87 (SD = 0.98). $t = -0.96$, $p = 0.17$, *Cohen's d*: 0.22.

3.3. Qualitative analysis

To obtain a more in-depth understanding of the underlying reasons behind participants' differences in trust, we conducted a thematic analysis of the participants' feedback responses F1–F3 as collected directly after using the system. This analysis was conducted by using conventional content analysis (34) in which two of the paper's authors tagged the feedback responses in a shared online document with key themes present in the given response. In case of disagreement in categorizing participant responses, a third author was included in the discussion. We present the three themes that we identified to affect participant trust; "personal experiences," "perceived reliability," and "presentation of results." These themes are present in both two study groups and we found no significant difference to how often each theme appeared between the CB and WEB groups.

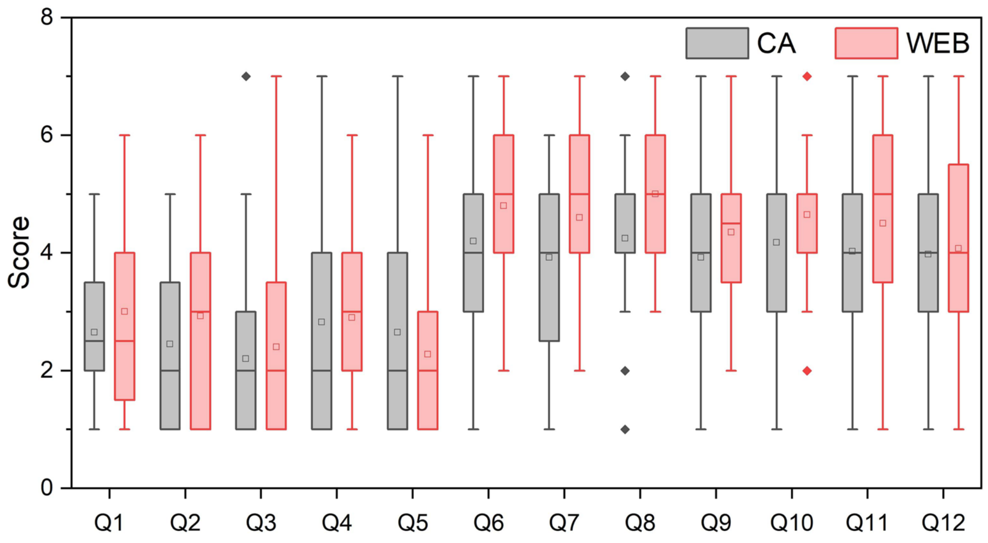


FIGURE 3
Scores of the trust in automation scale items for both study groups. Student's t-test shows significant difference for Q7 (*The system provides security*): CB: 3.93 (SD = 1.58), WEB: 4.6 (SD = 1.34), $t = -2.07$, $p = 0.042$, Cohen's d = 0.46, and Q8 (*The system has integrity*): CB: 4.25 (SD = 1.46), WEB: 5 (SD = 1.18), $t = -2.53$, $p = 0.014$, Cohen's d = 0.57. As the performed t-tests were independent of each other, there was no need for Bonferroni corrections.

3.3.1. Personal experiences

The ability to relate presented results to one's own experiences strongly affected participants' trust. Being presented with familiar results leads to participants being more receptive to methods they had not seen before. One participant discusses this notion in terms of establishing initial trust in the system; "*I found that my trust was established when I saw many techniques that I use, like mindfulness, running, etc. I think it was how familiar I was with the techniques shown that established my trust.*" (P21). On the other hand, one participant that had poor experiences with some of the methods suggested by the system expressed that the inclusion of these methods negatively affected their trust level; "*disagreeing with some of the suggestions made me question how good it was*" (P25).

In addition to relying on their own experience in assessing and evaluating presented methods, several participants also expressed their wish for a future system to incorporate their own experiences to provide more accurate suggestions. "*I expect the system to be able to assist me in choosing the most suitable techniques for me, giving me the right information and listening to my needs*" (P14). Participants' responses indicate that they are willing to put in the additional effort required to provide this data if it would result in more valuable suggestions; "*A short questionnaire done previously so that the system can make more accurate suggestions.*" (P05). Such an approach may also help present the mental health self-care methods a participant already has experienced differently.

3.3.2. Perceived reliability

While our quantitative results indicate that participants typically trusted the results presented by the system, our analysis of the open-text responses also highlights why some participants indicated a lower level of trust in the presented results. A couple of participants highlighted a lack of information on the people who contributed to the system's data as an obstacle to building trust. "*Not knowing much about the 'community' behind the system.*" (P26) and "*You cannot know precisely who is suggesting what.*" (P47) highlight these perspectives.

In line with the aforementioned theme of personal experiences, our participant responses indicate a tight balance between novel suggestions to which participants might express some uneasiness and well-known, established ideas that participants could dismiss as too obvious. For example, one participant highlighted that they had high trust in the presented results but that the methods were not extremely helpful to them; "*I thought the techniques would show me something 'new' that I have not heard of yet, but I have already tried most of them. I also think the techniques are too focused on depression and anxiety, which I guess is what most people suffer from, but I was expecting something more… groundbreaking?*" (P75). On the other hand, other participants commented that the lack of novelty did not affect their perceived usefulness or trust in the system; "*I think the results were rather expected, but that does not make them worse in any way. The system provided simple solutions and great ideas overall.* (P27), indicating that being familiar with the suggestions increases the perceived reliability of the overall system. Even as participants may have experience with some of the presented self-care methods, this did not necessarily deter them from trying out any of the other methods; "*Some of the techniques do not work for me, but some I have not thought about it and might be good to give them a try.*" (P31).

3.3.3. Presentation of results

Lastly, a large part of the participant sample showed how results were presented as a trust-affecting factor. Several participants mentioned a lack of uniformity in presenting self-care methods as having a negative effect. While this lack of uniformity directly results from the crowdsourced nature of the self-care methods and our deliberate choice not to edit participant contributions, some of these issues can be addressed relatively quickly in future iterations. For example, one participant commented on the capitalization of the methods "*The results […] were not written uniformly (some lacked capitalization, other did not, which is fine but weird)*" (P10). Similarly, another participant highlighted grammatical errors, as well as inconsistency in capitalization, as a trust-impeding factor in "*poor capitalization/non-standard grammar*" (P18). Although participants identified these errors as problematic, this lack of 'strict' grammar and spelling also highlighted to the participants that these methods were contributed by other users; "*it seems that a lot of the options were submitted by users, the spelling was wrong in some places too*" (P34)

Without explicitly being asked to do so, several participants commented positively on how the chatbot presented the system to the user; "*The fact that the system used a very humane, calm and warm language. The fact that it called me by nickname also affected my trust in the system.*" (P14). Similarly, another participant mentioned that their trust was positively affected by "*[…] the aesthetic and the reassuring phrases*" (P42). While not made explicit by the participants, the sensitive nature of mental health can play a significant role in this expressed sentiment.

While participants typically found the presented discovery interface useful, they also highlighted that additional information would be helpful. In particular, a couple of participants highlighted that the tool could do more to help them on their way once a method had been selected; "*Other than listing the options, as the tool had greatly done, I would also love a description on how to get started with each technique.*" (P21).

4. Discussion

4.1. Using chatbots in a self-care discovery system

Chatbots are forecast to ease the looming resource crisis in healthcare and automate many customer service functions in general. Mental health is a high-impact and sensitive domain. Thus, any interactions must be safe, secure, and confidential. To this end, crowdsourcing systems can complement 'official' clinical care: a repository of user-contributed self-care methods is simply another way to structure people's ideas and content, much like an online forum. However, before deploying this kind of crowdsourced system into real use, significant work has to be done to ensure it works as intended. In our research we did not receive any major feedback on potentially harmful or undesired methods, due to researches having gone through the list of methods beforehand. To take this into account in similar systems, we believe, that a moderator or automated recognition could increase the system's safety to be viable in mental healthcare.

Our study compared a standalone decision support interface to a setup wrapped in a conversation with an artificial agent. We found that the version wrapped within a conversation with a chatbot led to a lower trust for the system's security and integrity. While the exact reasons for this cannot directly be identified, we speculate that one explanation might be the privacy concerns that are frequent among online mental health services (20, 35). Participants could not be identified from the conversations with the chatbot, but we asked the users whether they would like to be called by a nickname to make the chatbot more humane and empathetic and act like most chatbots online do. Participants might have felt this makes them more identifiable and interferes with their privacy needs. This might also be due to the study design, as the nickname was asked only from the CB group. Other possible explanations for the lower scores could be the preference to use such a system without additional guidance and the general uncertainty towards the chatbot. Subsequent studies, perhaps using a within-subjects design, could help determine why the chatbot seemed to degrade these scores.

This study gives a clear direction to create a broadly accessible system for helping people to maintain and improve their mental health. With an active user base, new self-care methods could be introduced and ranked as they continuously interact with the system. Integrating a self-care discovery system within the conversation can make it more approachable and easily accessible (36). We believe that interaction with the chatbot can also improve its overall usability and performance, thus potentially increasing the effect it can have on the user's mental health. To achieve this, the trust towards the system using the chatbot needs to be improved.

4.2. Transparency, integrity, and security

One of the other factors affecting chatbot trust is transparency, i.e., sharing the limitations of the chatbot with the users to help them predict different outcomes and conversations, with other affecting factors including dialogue, interface, expressions, and conversational styles (28, 37). Here, transparency was also clearly an issue with the content itself: Users wanted more details about who articulated, assessed, and helped build the knowledge base of self-care methods. As our qualitative analysis shows, this negatively affected the trust for both of the two groups. Much uncertainty comes from the recommended self-care methods and how those have been added to our system.

Second, the difference between the integrity and security scores might also come from the fact that the CB group uses two different systems instead of one for the WEB group. Participants' experienced that if the service they are using is fragmented to multiple interfaces, this could potentially create more points for attack for privacy intrusions. When the number of used systems, applications, websites, and similar increases, the risk of being exposed to data breaches increases. Some might feel uncomfortable sharing private information if it is required to share it with multiple sources. Thus, we believe that implementing this method discovery directly in the conversation could lead into better trust, without the need to do that in a separate system.

4.3. Chatbot behaviour and humanity

Chatbot behaviour and human likeness are essential factors in forming trust in chatbots (9, 20, 38). There is evidence that the personality of the user makes a difference in how trust between the user and the chatbot is formed (19, 37). In line with our findings, previous studies have frequently mentioned human likeness as a key aspect of trust (9, 38). Indeed, in our study, six participants directly mentioned how the chatbot and its human-likeness positively affected their trust in the system. Some contradicting evidence emerged in a study by Folstad et al. (38) on chatbots in customer service, where the majority of participants preferred human-likeness and personalized chatbots for building trust. However, some participants referred to the uncanny valley effect. To conclude, it is essential to keep a chatbot identifiable as non-human when developing trustworthy systems. Our chatbot was designed to be identifiable as non-human from the beginning with qualities such as its introduction, speech patterns, and name, *CareBot*. A

We believe our results might have been significantly different with a different kind of bot personality, but having the chatbot play as neutral of a role as possible could yield the best results in the mental health context. Naturally, chatbots with varying personalities are excellent avenues to explore in future work.

4.4. Chatbot as a factor affecting trust

Although some participants did mention that the chatbot affected their trust in the systems, the amount was lower than expected. Most participants focused on describing the factors affecting their trust in the self-care discovery tool while disregarding the chatbot. While we made sure to keep the connections between the two systems seamless by, for example, implementing the self-care discovery system within the chatbot window instead of a separate web-browser tab, this is something to focus on more in the future. An excellent way to handle this would be to have the chatbot ask for the criteria and give the recommended methods to the user directly, without needing a tool of its own. This could make the system easier to use while also giving more possibilities to explain the methods and their practical usage to the participants, one feature requested by the users and discussed in the *presentation of the results* section. Transparency should remain, and the sources of the recommendations offered should always be explained and referred to.

As it stands, the created chatbot might not yet be trustworthy enough for self-care method discovery, as the standalone DSS version enjoys larger overall trust. While we got significant results only for the security and integrity between the two conditions, those are crucial factors for a successful chatbot and might lead to users not being comfortable using the chatbot due to, for example, security concerns. Neglecting these factors can lead to significantly lower trust compared to standalone systems. This gives us a clear direction for improving the chatbot.

4.5. Limitations and future research

To gain a more in-depth understanding of how users form trust in the specific DSS used here, a larger sample could be beneficial. We also compared results only between two groups; one group using the WEB interface wrapped in conversation and another group using only the WEB interface but were missing a study condition where the full interaction is conducted through the chatbot. The DSS interface was included in both groups. Even though the DSS interface was presented to the CB group embedded in the chatbot interface, it remains a web interface at its core. In future studies, it would be important to see how a conversational agent performs without the need of external interfaces outside of the chatbot. For this study, due to technical limitations, this feature was not yet implemented.

The original research by Jian et al. (24) shows, that trust, and distrust (positive and negative trust) have a negative correlation, and thus there is no need to develop separate scales for the two conditions. The scale used is their proposed way to measure overall trust, but their findings also suggest that calculating the positive and negative trust scores might be unnecessary. We are also aware of other existing measurements and standardized surveys for trust. In our work we decided to use a widely accepted and used survey to gain first insights to how trust might change when using a conversational agent for mental health self-care intervention. However, in future research, it might be beneficial to use surveys made to measure trust in a health or web-based context instead of a more generalized one (39,40).

As mentioned before, previous research shows that the human-likeness of the chatbot might significantly affect the formed trust between them and the user. This could be especially important in mental health applications. To best compare the use of chatbots within self-care discovery systems, chatbots with differing personalities and levels of anthropomorphism could be used.

5. Conclusion

We presented an exploration of using a chatbot for self-care method discovery, specifically focusing on perceived trust in the system. The use of a chatbot was compared to a traditional standalone web interface. We found significant differences between security, integrity, and the positive trust between the two conditions, and that trust is affected mainly by personal experiences, perceived reliability, and presentation of results. To improve the trust for the chatbot further, more attention is needed for its security and integrity, which could be done by, for example, implementing the self-care method discovery directly within the conversation.

Although the results for the trust survey showed lower trust in all categories for the CB group, several students mentioned the chatbot to positively affect their trust in the system. We believe improvements to the chatbot, especially to increase its security and integrity, could indeed increase the trust to the same level as the WEB group. Using a chatbot for self-care method discovery could make this kind of system more easily accessible, easier to use, and overall increase the user experience if the overall trust towards it is high enough.

Data availability statement

The datasets presented in this study can be found at https://doi.org/10.6084/m9.figshare.22193803.v1.

Ethics statement

Ethical review and approval was not required for the study on human participants in accordance with the local legislation and institutional requirements. The patients/participants provided their written informed consent to participate in this study.

Author contributions

All authors contributed substantially to the conception and design of the study. JM and AV contributed to the data analysis of the results. JM and NvB contributed substantially to the writing of the paper. UG, WvdM and SH provided critical reviews and significant contributions to the manuscript. All authors contributed to the article and approved the submitted version.

Funding

This research is connected to the GenZ strategic profiling project at the University of Oulu, supported by the Academy of Finland (project number 318930), and CRITICAL (Academy of Finland Strategic Research, 335729). Part of the work was also carried out with the support of Biocenter Oulu, spearhead project ICON.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

1. Boardman J. Social exclusion, mental health–how people with mental health problems are disadvantaged: an overview. *Ment Health Soc Incl.* (2011) 15 (3):112–21. doi: 10.1108/20428301111165690

2. Soeteman DI, Roijen LH, Verheul R, Busschbach JJ. The economic burden of personality disorders in mental health care. *J Clin Psychiatry*. (2008) 69:259. doi: 10.4088/JCP.v69n0212

3. Kruisselbrink Flatt A. A suffering generation: six factors contributing to the mental health crisis in north American higher education. *Coll Q.* (2013) 16(1):1–17. https://eric.ed.gov/?id=EJ1016492

4. Hartrey L, Denieffe S, Wells JS. A systematic review of barriers, supports to the participation of students with mental health difficulties in higher education. *Ment Health Prev*. (2017) 6:26–43. doi: 10.1016/j.mhp.2017.03.002

5. Naslund JA, Aschbrenner KA, Marsch LA, Bartels SJ. The future of mental health care: peer-to-peer support, social media. *Epidemiol Psychiatr Sci*. (2016) 25:113–22. doi: 10.1017/S2045796015001067

6. Martin JM. Stigma and student mental health in higher education. *High Educ Res Dev*. (2010) 29:259–74. doi: 10.1080/07294360903470969

7. Barnett JE, Baker EK, Elman NS, Schoener GR. In pursuit of wellness: the self-care imperative. *Prof Psychol Res Pr*. (2007) 38:603a. doi: 10.1037/0735-7028.38.6.603

8. Richards K, Campenni C, Muse-Burke J. Self-care and well-being in mental health professionals: the mediating effects of self-awareness and mindfulness. *J Ment Health Couns*. (2010) 32:247–64. doi: 10.17744/mehc.32.3.0n31v88304423806

9. Morrow DG, Lane HC, Rogers WA. A framework for design of conversational agents to support health self-care for older adults. *Hum Factors*. (2021) 63:369–78. doi: 10.1177/0018720820964085

10. Qiu S, Gadiraju U, Bozzon A. Ticktalkturk: conversational crowdsourcing made easy. *Conference Companion Publication of the 2020 on Computer Supported Cooperative Work and Social Computing*. New York, NY, USA: Association for Computing Machinery (2020). p. 53–57.

11. Laranjo L, Dunn AG, Tong HL, Kocaballi AB, Chen J, Bashir R, et al. Conversational agents in healthcare: a systematic review. *J Am Med Inform Assoc*. (2018) 25:1248–58. doi: 10.1093/jamia/ocy072

12. WH Organization, et al. *Promoting mental health: concepts, emerging evidence, practice: summary report*. World Health Organization (2004).

13. Miner A, Chow A, Adler S, Zaitsev I, Tero P, Darcy A, et al. Conversational agents and mental health: theory-informed assessment of language and affect. *Proceedings of the Fourth International Conference on Human Agent Interaction*. New York, NY, USA: Association for Computing Machinery (2016). p. 123–130.

14. Morris RR, Kouddous K, Kshirsagar R, Schueller SM. Towards an artificially empathic conversational agent for mental health applications: system design and user perceptions. *J Med Internet Res*. (2018) 20:e10148. doi: 10.2196/10148

15. Lucock M, Gillard S, Adams K, Simons L, White R, Edwards C. Self-care in mental health services: a narrative review. *Health Soc Care Community*. (2011) 19:602–16. doi: 10.1111/hsc.2011.19.issue-6

16. Vaidyam AN, Wisniewski H, Halamka JD, Kashavan MS, Torous JB. Chatbots and conversational agents in mental health: a review of the psychiatric landscape. *Can J Psychiatry*. (2019) 64:456–64. doi: 10.1177/0706743719828977

17. Abd-Alrazaq AA, Rababeh A, Alajlani M, Bewick BM, Househ M. Effectiveness and safety of using chatbots to improve mental health: systematic review and meta-analysis. *J Med Internet Res*. (2020) 22:e16021. doi: 10.2196/16021

18. Cameron G, Cameron D, Megaw G, Bond R, Mulvenna M, O'Neill S, et al. Towards a chatbot for digital counselling. *Proceedings of the 31st International BCS Human Computer Interaction Conference (HCI 2017) 31*. Sunderland, UK: BCS Learning and Development Ltd. (2017). p. 1–7.

19. Müller L, Mattke J, Maier C, Weitzel T, Graser H. Chatbot acceptance: a latent profile analysis on individuals' trust in conversational agents. *Proceedings of the 2019 on Computers and People Research Conference*. New York, NY, USA: Association for Computing Machinery (2019). p. 35–42.

20. Wang W, Siau K. Living with artificial intelligence–developing a theory on trust in health chatbots. *Proceedings of the Sixteenth Annual Pre-ICIS Workshop on HCI Research in MIS*. San Francisco, CA: Association for Information Systems (2018).

21. Pesonen JA. 'Are you ok?' Students' trust in a chatbot providing support opportunities. *International Conference on Human-Computer Interaction*. Springer (2021). p. 199–215.

22. Abbas T, Khan VJ, Gadiraju U, Markopoulos P. Trainbot: a conversational interface to train crowd workers for delivering on-demand therapy. *Proceedings of the AAAI Conference on Human Computation and Crowdsourcing*. Vol. 8. Palo Alto, CA, USA: Association for the Advancement of Artificial Intelligence (2020). p. 3–12.

23. Gupta A, Basu D, Ghantasala R, Qiu S, Gadiraju U. To trust or not to trust: how a conversational interface affects trust in a decision support system (2022).

24. Jian JY, Bisantz AM, Drury CG. Foundations for an empirically determined scale of trust in automated systems. *Int J Cogn Ergon*. (2000) 4:53–71. doi: 10.1207/S15327566IJCE0401_04

25. Lee JD, See KA. Trust in automation: designing for appropriate reliance. *Hum Factors*. (2004) 46:50–80. doi: 10.1518/hfes.46.1.50.30392

26. Zhang X, Zhang Q. Online trust forming mechanism: approaches and an integrated model. *Proceedings of the 7th International Conference on Electronic Commerce*. New York, NY, USA: Association for Computing Machinery (2005). p. 201–209.

27. Cheshire C. Online trust, trustworthiness, or assurance? *Daedalus*. (2011) 140:49–58. doi: 10.1162/DAED_a_00114

28. Rheu M, Shin JY, Peng W, Huh-Yoo J. Systematic review: trust-building factors and implications for conversational agent design. *Int J Hum Comput Interact*. (2021) 37:81–96. doi: 10.1080/10447318.2020.1807710

29. Tolmeijer S, Gadiraju U, Ghantasala R, Gupta A, Bernstein A. Second chance for a first impression? Trust development in intelligent system interaction. *Proceedings of the 29th ACM Conference on User Modeling, Adaptation and Personalization*. New York, NY, USA: Association for Computing Machinery (2021). p. 77–87.

30. Nordheim CB, Følstad A, Bjørkli CA. An initial model of trust in chatbots for customer service—findings from a questionnaire study. *Interact Comput*. (2019) 31:317–35. doi: 10.1093/iwc/iwz022

31. Hosio S, Goncalves J, Anagnostopoulos T, Kostakos V. Leveraging wisdom of the crowd for decision support. *Proceedings of the 30th International BCS Human Computer Interaction Conference 30*. Sunderland, UK: BCS Learning and Development Ltd. (2016). p. 1–12.

32. Hosio SJ, Karppinen J, Takala EP, Takatalo J, Goncalves J, Van Berkel N, et al. Crowdsourcing treatments for low back pain. *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. New York, NY, USA: Association for Computing Machinery (2018). p. 1–12.

33. Hosio SJ, van Berkel N, Oppenlaender J, Goncalves J. Crowdsourcing personalized weight loss diets. *Computer*. (2020) 53:63–71. doi: 10.1109/MC.2019.2902542

34. Hsieh HF, Shannon SE. Three approaches to qualitative content analysis. *Qual Health Res*. (2005) 15:1277–88. doi: 10.1177/1049732305276687

35. Kraus M, Seldschopf P, Minker W. Towards the development of a trustworthy chatbot for mental health applications. *International Conference on Multimedia Modeling*. Springer (2021). p. 354–366.

36. Ahmed A, Ali N, Aziz S, Abd-Alrazaq AA, Hassan A, Khalifa M, et al. A review of mobile chatbot apps for anxiety and depression and their self-care features. *Comput Methods Programs Biomed Update*. (2021) 1:100012. doi: 10.1016/j.cmpbup.2021.100012

37. Bickmore TW, Picard RW. Establishing and maintaining long-term human-computer relationships. *ACM Trans Computer Human Interact*. (2005) 12:293–327. doi: 10.1145/1067860.1067867

38. Følstad A, Nordheim CB, Bjørkli CA. What makes users trust a chatbot for customer service? An exploratory interview study. *International Conference on Internet Science*. Springer (2018). p. 194–208.

39. Sillence E, Blythe JM, Briggs P, Moss M. A revised model of trust in internet-based health information and advice: cross-sectional questionnaire study. *J Med Internet Res*. (2019) 21:e11125. doi: 10.2196/11125

40. Rowley J, Johnson F, Sbaffi L. Students—trust judgements in online health information seeking. *Health Informatics J*. (2015) 21:316–27. doi: 10.1177/1460458214546772