Building Crowd-powered Conversational Systems to Provide On-demand Psychological Support

Tahir Abbas
Eindhoven University of Technology
Eindhoven 5600 MB, Netherlands
Mirpur University of Science & Technology
Mirpur AJK, Pakistan
t.abbas@tue.nl

Vassilis-Javed Khan
Eindhoven University of Technology
Eindhoven 5600 MB, Netherlands
v.j.khan@tue.nl

Ujwal Gadiraju
Delft University of Technology
Delft, The Netherlands
u.k.gadiraju@tudelft.nl

Panos Markopoulos
Eindhoven University of Technology
Eindhoven 5600 MB, Netherlands
p.markopoulos@tue.nl

ABSTRACT

AI based self-help interventions are gaining considerable attention. However, they lack the necessary expertise and emotional intelligence skills to fully engage with a stressed person. More recently, researchers have investigated crowd-based psychological interventions, which proved to be effective in treating mental illnesses. In this paper, we showcase two on-going research projects pertaining to crowd-based counselling and their shortcomings that need to be resolved, before such conversational systems can be deployed in the wild.

KEYWORDS

Crowd-powered conversational systems, stress management, social robotics, chatbots, crowd learning

INTRODUCTION AND BACKGROUND

Advances in artificial intelligence (AI), NLP, and speech recognition have led to the development of a plethora of conversational agents (CAs) for the healthcare domain. Coping with stress is crucial for a

Table 1: Sample dialogue from a previous study where we trained crowd workers to learn motivational interviewing skills using a chatbot [2].

User: I had a quite stressful few weeks really and it kept me from focusing on my studies. I would like to talk to you about that.
Crowd: I'm sorry to hear you'd struggled with your studies; can you outline what's been affecting you this week?
User: I moved from Belgium to the Netherlands 3 months ago to do minor in Textile Engineering so it's totally new city and new the house where I am living in and the subjects are so different. . . it's is very difficult to find ground to walk on almost.
Crowd: I can understand why that would make you feel stressed, you've gone through quite a few changes recently. How have you been managing with everything so far?
User: It would be nice to find free time although I'm already behind on my schedule. I should actually spend even more time on my study so that I will be able to finish it on time.
Crowd: I think you need to not be too hard on yourself! Sometimes having even just half an hour of relax time can help clear your mind and you may find that it makes you more productive afterwards!
User: that's a good tip. I did try that. . . So, I have to go now.. thank you so much for having this conversation with me.
Crowd: You are more than welcome. I hope your stress eases soon. I am always here if you need somebody to speak to.

Figure 1: High-level system architecture of Crowd-of-Oz.

[1]https://www.koko.ai/

healthy lifestyle. Prolonged and high levels of stress in humans can affect several physiological and psychological functions [9, 23]. CAs have recently been proposed to treat stress-related symptoms [24]. Others have proposed the leveraging of embodied CAs to alleviate stress and anxiety related problems among children [7], older adults [20], and teens [10, 21]. On the other hand, chatbots have also been used in the treatment of anxiety and stress related symptoms using variety of underlying interventions such as cognitive behavior therapy [6] and motivational interviewing [19].

The shortcomings of current AI techniques and natural language understanding are not yet capable to deal with the complexities in conversational interactions, which results in breakdown in the conversations [5], failure to provide more detailed and contextualized feedback [19], and inaccuracy of automatic speech recognition [25]. Furthermore, building a fully autonomous CA for delivering psycho-therapeutic solutions requires advancing research in emotional intelligence, affect analysis and computational psychology. Due to these limitations, the adoption rate and adherence is quite low in health care related CAs [4].

We believe that CAs coupled with human intelligence support in real time can circumvent many of the aforementioned challenges. For example, researchers in affective crowdsourcing [16] have already proposed methods to leverage crowdsourcing to deliver positive psychological interventions to people who are stressed [17]. One example is Panoply (currently inactive), a crowd-powered system that leverages the crowds' wisdom to provide on demand emotional support to people in need [18]. Another example is a Koko [1], which is an online mental health intervention app based on the concept of crowdsourced cognitive therapy to combat stressful thoughts [8].

There are mainly two challenges for implementing CAs based on crowdsourced therapy: a) Existing crowd-powered CAs tools take enormous amount of time to give empathetic responses to stressed peoples. For instance, Panoply takes on average 9 min to 2 h to receive a first response from crowd workers and other registered users, which is not suitable for a two-way live synchronous conversation; b) Secondly, it is not known how can one train a non-expert worker holistically to deliver positive psychological support? Such training is challenging since it requires a plethora of skills ranging from understanding a person's thoughts and feelings to deciding what actions to undertake based on specific problems.

To solve aforementioned challenges, we built two systems: Crowd of Oz (CoZ) and Trainbot. CoZ [3] is a system where Softbank's Pepper robot broadcasts a live audio-video (AV) feed of a user who is stressed to synchronous group of workers who respond to user utterances in real time. CoZ was developed based on the principles of Real-time crowdsourcing (RTC), which is an area in Human Computation research where online workers carry out tasks under real time constraints [11]. Although CoZ was effective in surmounting latency issues [1], invoking an inexperienced pool of workers to deliver positive psychological support can be detrimental. To tackle this problem, we developed Trainbot, which is a rule-based chatbot built on the principles of motivational interviewing to train

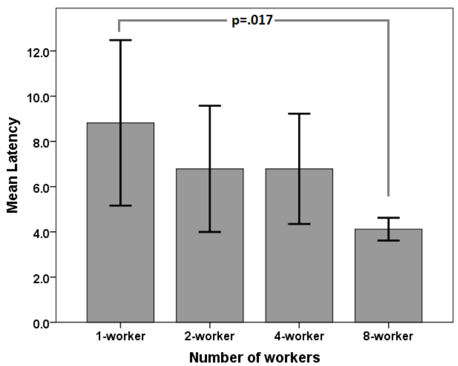

Figure 2: Mean Latency: significant difference in the average response time was observed between 1-worker and 8-worker conditions

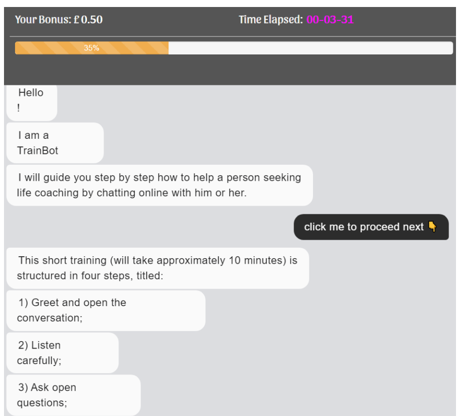

Figure 3: Trainbot interface

armature workers for providing emotional support to people in need. In the next section, we describe these systems and their studies in detail.

PROJECT DESCRIPTIONS AND RESEARCH QUESTIONS

Crowd of Oz

Fig. 1 shows a high-level system architecture of CoZ. We used Softbank's Pepper robot, which can exhibit some social behaviors e.g. gestures, head movements, and adjust its head pose to track users. Workers operate the robot remotely from a web interface. When a user talks to the robot, her voice is converted into text. The transcribed text is displayed on the Pepper's Tablet and also forwarded to the worker's interface. Additionally, a live audio-video (AV) stream from Pepper's front camera is broadcasted to workers. Workers can see the transcribed text along with the AV stream. They can compose a message either by typing or using the speech-to-text (STT). Their message is then spoken by Pepper through an animated speech.

In a recent study with CoZ [1], we investigated the effect of the number of workers simultaneously controlling the speech of the robot to the response latency and conversation quality, in the context of stress mitigation -i.e. the robot would be acting as a coach for a stressed user. We recruited N=1, 2, 4, & 8 workers in the queue for our investigation. Since this was our first study in leveraging the crowd as a coach to alleviate stress, we opted to hire a professional actress who played the role of a stressed university student. The professional actress improvised dialogues in all conditions and each condition was executed 5 times (total of 20 sessions). Our results indicate that increasing the target number of workers can improve the response latency linearly (Fig. 2) but at the expense of financial cost. Additionally, we did not find any difference in the quality of conversation (we use appropriateness as a measure for conversational quality [26]) in all conditions.

Trainbot

Trainbot (Fig. 3) is a rule-based conversational interface built on top of chat-bubble[2] framework that leverages Motivational Interviewing (MI) theory, which is a powerful counseling approach for treating anxiety, depression, and other mental problems. [14]. We structured Trainbot based on the following MI topics: 1) greeting and opening the conversation; 2) reflective listening; 3) showing empathy; 4) asking open questions; 5) affirming the user's strengths/coping skills; 6) wrapping-up the conversation. After training workers on a specific topic, Trainbot confirms by asking "did you understand the topic?" and if a worker answers in the negative, Trainbot provides an elaborate explanation with more examples. At the end of a topic, Trainbot presents workers with short quizzes to solve. Upon answering a question correctly, Trainbot continues to a new topic. If they fail to answer a question, Trainbot presents them with two options – to either retake the quiz or read the instructions again.

[2] https://github.com/dmitrizzle/chat-bubble

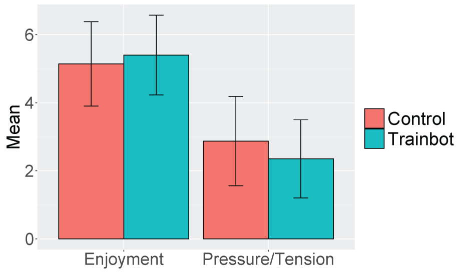

Figure 4: The treatment group took less retakes while solving quizzes

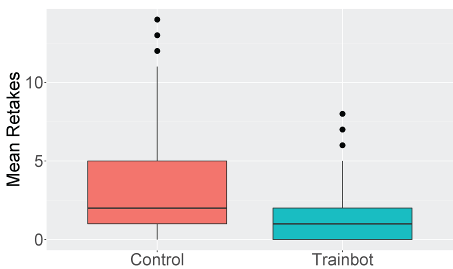

Figure 5: The treatment group felt less pressure as compared to control group The treatment groupusing Intrinsic Motivation Inventory ([13])

We conducted a between-subjects experimental study on Prolific, wherein a group of workers (N=200) received training on motivational interviewing via either a Trainbot or a conventional web interface. We found that workers in the Trainbot group: 1) felt less pressure (Fig. 4), retook fewer quizzes (Fig. 5); 2) provided psychological interventions that were rated consistently higher by psychologists than the control group; 3) felt a higher self-efficacy in helping deal with stress management after the training task (Fig 6).

Research Questions

We derived following research questions based on the two studies: **R1:** How could the system handle a failure of the crowd to produce a response within an acceptable delay? **R2:** What is the effect of video and audio on the perceived privacy of users interacting with the robot? **R3:** How can we ensure reliable and accountable crowd-powered counseling?

PROPOSED SOLUTIONS

R1: One solution that has been developed to support artificial conversational agents is to 'buy time' with conversational fillers or acknowledgement tokens [27]. These conversational fillers can be simple (e.g., "well" and "uh") or complex (e.g., "that is an interesting question, let me think about it"). This way, the crowd would still be handling the essence of the dialogue while CoZ would aim to yield a more positive user experience and a better perception of system latency.

R2: Previous study showed that embodied conversational agent can collect more sensitive information compare to its counterpart (dis-embodied agent) [22]. However, deploying a social robot operated by anonymous crowd-workers in a realistic scenario entails addressing privacy. According to Rueben et al. [22], two types of privacy is relevant to CoZ: physical privacy and informational privacy. Physical privacy refers to preserving one's own personal space or territory. In context of CoZ, this can be achieved through privacy veils or avatars to hide the identity of the person talking to a robot and altering the voice of the person using audio modulation techniques. On the other hand, the perception of informational privacy can be improved by training workers to avoid requesting personal information or using AI to filter out such queries. In a recent preliminary study, we deployed CoZ in a trip planning task to study the privacy aspects. In one condition, the robot broadcasts both audio and video to crowd workers as opposed to broadcasting only the participants' audio cues in the other condition. Interestingly, participants in the video condition did not feel very intimidated while participants in No-Video condition were critical about what crowd workers could do with their data. However, this was small study in a laboratory setting with only 14 participants. Further study is needed to understand the privacy perception of participants in the wild with CoZ.

R3: Due to anonymity and lack of fair communication between worker and requester (who post the task) on crowdsourcing platform, one can expect to receive inferior counselling, which is devastating

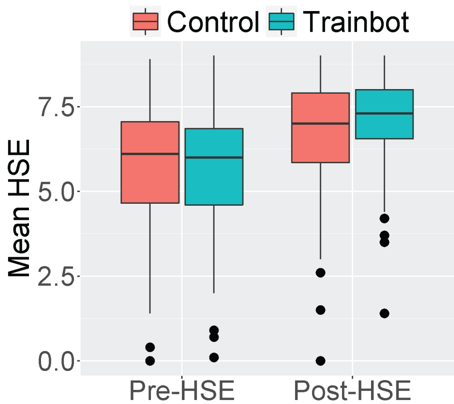

Figure 6: A significant difference between Pre- (helping skills before commencing the training) and Post- (after accomplishing the training) helping self-efficacy (HSE) was observed in both MI-based interventions. The workers in the treatment group scores relatively higher than the control group after the training task

for a stressed person who is seeking sincere help. To increase accountability, one can augment CA with external support from human – a concept known as supportive accountability [15]. Although, supportive accountability stresses on how human coaches can impact on adherence to eHealth interventions, we can apply this theory to make workers more accountable. For instance, one aspect of accountability is the social presence, which refers to "the implicit or explicit expectation that an individual may be called upon to justify his or her actions or inactions" [12]. This can be achieved with external set of workers who have some experience in psychology to provide external feedback to lay workers' helping skills. Also, akin to freelancing business where workers' performance is shown to requesters, we can also show the performance of workers (coaching skills, acquired training score) to people seeking help.

CONCLUSION

We present two systems for stress management: CoZ and Trainbot. The former is a novel system to elicit crowd generated counselling in real-time, to sustain a conversation between a robot and a stressed person. We highlighted that the latency and privacy issues need to be resolved for the CoZ to provide counselling to actual stressed people. The latter is a conversational system built on the principles of MI for training untrained workers to provide productive and meaningful counseling. We are working towards a comprehensive toolkit for delivering on-demand crowdsourced psychotherapy. Such an emotional tool is an elusive societal need that is exacerbated in difficult times – as witnessed during the COVID-19 pandemic.

REFERENCES

[1] Tahir Abbas, Vassilis-Javed Khan, Ujwal Gadiraju, Emilia Barakova, and Panos Markopoulos. 2020. Crowd of Oz: a crowd-powered social robotics system for stress management. *Sensors* 20, 2 (2020), 569.

[2] Tahir Abbas, Vassilis-Javed Khan, Ujwal Gadiraju, and Panos Markopoulos. 2020. Trainbot: A Conversational Interface to Train Crowd Workers for Delivering On-Demand Therapy. In *Proceedings of the Eighth AAAI Conference on Human Computation and Crowdsourcing, HCOMP 2020, 26-29 October 2020, Hilversum, the Netherlands*. AAAI Press.

[3] Tahir Abbas, Vassilis-Javed Khan, and Panos Markopoulos. 2020. CoZ: A crowd-powered system for social robotics. *SoftwareX* 11 (2020), 100421.

[4] Murray Aitken, Jennifer Lyle, et al. 2015. Patient adoption of mHealth: use, evidence and remaining barriers to mainstream acceptance. *Parsippany, NJ: IMS Institute for Healthcare Informatics* (2015).

[5] Zahra Ashktorab, Mohit Jain, Q Vera Liao, and Justin D Weisz. 2019. Resilient chatbots: repair strategy preferences for conversational breakdowns. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. 1–12.

[6] Eileen Bendig, Benjamin Erb, Lea Schulze-Thuesing, and Harald Baumeister. 2019. The next generation: chatbots in clinical psychology and psychotherapy to foster mental health–a scoping review. *Verhaltenstherapie* (2019), 1–13.

[7] Molly K. Crossman, Alan E. Kazdin, and Elizabeth R. Kitt. 2018. The influence of a socially assistive robot on mood, anxiety, and arousal in children. *Professional Psychology: Research and Practice* 49, 1 (2 2018), 48–56.

[8] Adam C Jaroszewski, Robert R Morris, and Matthew K Nock. 2019. Randomized controlled trial of an online machine learning-driven risk assessment and intervention platform for increasing the use of crisis services. *Journal of consulting and clinical psychology* 87, 4 (2019), 370.

[9] Marian Joëls, Zhenwei Pu, Olof Wiegert, Melly S Oitzl, and Harm J Krugers. 2006. Learning under stress: how does it work? *Trends in cognitive sciences* 10, 4 (2006), 152–158.

[10] Junhan Kim, Yoojung Kim, Byungjoon Kim, Sukyung Yun, Minjoon Kim, and Joongseek Lee. 2018. Can a Machine Tend to Teenagers' Emotional Needs?. In *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems - CHI '18*. ACM Press, New York, New York, USA, 1–6. https://doi.org/10.1145/3170427.3188548

[11] Walter S Lasecki, Christopher Homan, and Jeffrey P. Bigham. 2014. Architecting Real-Time Crowd-Powered Systems. *Human Computation* 1, 1 (2014), 67–93. https://doi.org/10.15346/hc.v1i1.5

[12] Jennifer S Lerner and Philip E Tetlock. 1999. Accounting for the effects of accountability. *Psychological bulletin* 125, 2 (1999), 255.

[13] Edward McAuley, Terry Duncan, and Vance V Tammen. 1989. Psychometric properties of the Intrinsic Motivation Inventory in a competitive sport setting: A confirmatory factor analysis. *Research quarterly for exercise and sport* 60, 1 (1989), 48–58.

[14] William R Miller and Stephen Rollnick. 2012. *Motivational interviewing: Helping people change.* Guilford press.

[15] David Mohr, Pim Cuijpers, and Kenneth Lehman. 2011. Supportive accountability: a model for providing human support to enhance adherence to eHealth interventions. *Journal of medical Internet research* 13, 1 (2011), e30.

[16] Robert Morris. 2011. Crowdsourcing workshop: the emergence of affective crowdsourcing. In *Proceedings of CHI'11*. ACM.

[17] Robert R Morris and Rosalind Picard. 2012. Crowdsourcing collective emotional intelligence. *arXiv preprint arXiv:1204.3481* (2012).

[18] Robert R Morris, Stephen M Schueller, and Rosalind W Picard. 2015. Efficacy of a web-based, crowdsourced peer-to-peer cognitive reappraisal platform for depression: randomized controlled trial. *Journal of medical Internet research* 17, 3 (2015).

[19] SoHyun Park, Jeewon Choi, Sungwoo Lee, Changhoon Oh, Changdai Kim, Soohyun La, Joonhwan Lee, and Bongwon Suh. 2019. Designing a Chatbot for a Brief Motivational Interview on Stress Management: Qualitative Case Study. *Journal of medical Internet research* 21, 4 (2019), e12231.

[20] Lihui Pu, Wendy Moyle, Cindy Jones, and Michael Todorovic. 2018. The Effectiveness of Social Robots for Older Adults: A Systematic Review and Meta-Analysis of Randomized Controlled Studies. *The Gerontologist* (6 2018).

[21] Emma J Rose. 2017. Designing for engagement: Using participatory design to develop a social robot to measure teen stress. *Proceedings of the 35th ACM International Conference on the Design of Communication* (2017), 1–10.

[22] Matthew Rueben, Cindy M Grimm, Frank J Bernieri, and William D Smart. 2017. A taxonomy of privacy constructs for privacy-sensitive robotics. *arXiv preprint arXiv:1701.00841* (2017).

[23] Joachim Taelman, Steven Vandeput, Arthur Spaepen, and Sabine Van Huffel. 2009. Influence of mental stress on heart rate and heart rate variability. In *4th European conference of the international federation for medical and biological engineering*. Springer, 1366–1369.

[24] Aditya Nrusimha Vaidyam, Hannah Wisniewski, John David Halamka, Matcheri S Kashavan, and John Blake Torous. 2019. Chatbots and conversational agents in mental health: a review of the psychiatric landscape. *The Canadian Journal of Psychiatry* 64, 7 (2019), 456–464.

[25] Adam P Vogel and Angela T Morgan. 2009. Factors affecting the quality of sound recording for speech and voice analysis. *International journal of speech-language pathology* 11, 6 (2009), 431–437.

[26] Nick Webb, David Benyon, Preben Hansen, and Oil Mival. 2010. Evaluating Human-Machine Interaction for Appropriateness. *LREC Conference Proceedings* 2, 1 (2010), 84–92.

[27] Don H. Zimmerman. 1993. Acknowledgment Tokens and Speakership Incipiency Revisited. *Research on Language & Social Interaction* 26, 2 (4 1993), 179–194. https://doi.org/10.1207/s15327973rlsi2602{_}4