ChatGPT-4's Ability to Alleviate Momentary Stress Immediately by Using CBT Techniques: A Randomized Controlled Trial

Leilani Kuroki

(Senshu University)

Key Words: AI, CBT, Stress Management

This might surprise many, but the first attempt to build a therapy chatbot dated back to the 1960s. It was an unintelligent program called ELIZA, approximating a Rogerian therapist (Bassett, 2018). While the developer himself found the idea "perverse", today, AI chatbots have become far more intelligent and some people are serious about the idea. So, are today's intelligent generative AI chatbots capable of at least some aspects of the job?

Purpose: This study aimed to examine the immediate effect of a ChatGPT-4 response using Cognitive-Behavioral Therapy (CBT) techniques on momentary stress, using a randomized controlled trial in an online experiment setting.

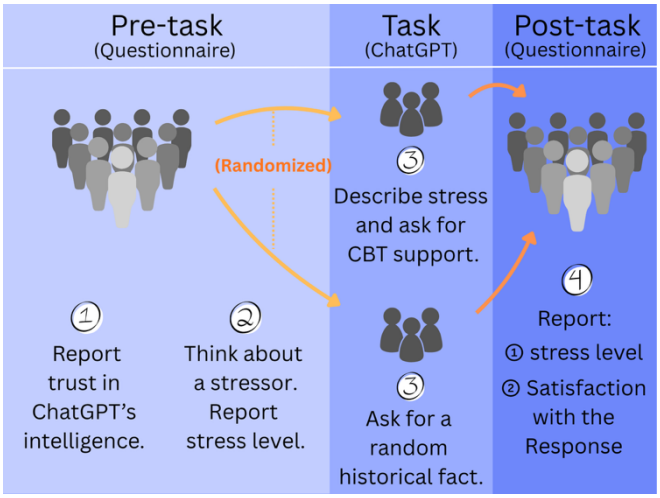
Method

Participants: There were 71 participants and 68 valid responses. Participants were proficient in either English or Japanese and were recruited mainly through social media and the University of Wolverhampton.

Design: This was an online randomized controlled experiment. The treatment group asked ChatGPT for CBT-based support, while the control group asked for a random historical fact.

Measure: Stress Numerical Rating Scale-11, a single-item (0 to 10) measure for momentary stress (Karvounides et al., 2016).

Procedure: Participants accessed the Qualtrics questionnaire with questions and instructions without monitoring. The flow of the experiment is illustrated in Figure 1 below. Certain limits were put in place to control for variability: the prompts given to both groups requested a response of 25 sentences; no follow-ups were allowed (one question, one response); and the participants had to return to the questionnaire within 10 minutes.

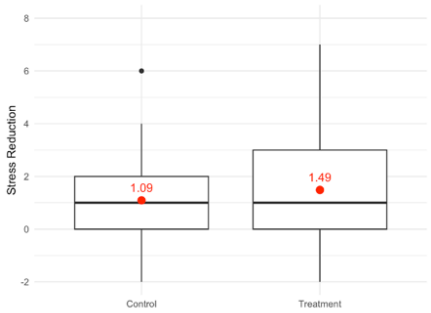


(Fig. 1)

To access the data:
https://osf.io/59hv4

Results

Did momentary stress levels decrease for both groups? Yes. For the control group, the mean stress level decreased by 1.09 (SD = 1.72, $p < .001$, $d = 0.54$, 95% CI [0.48, 1.70]), or 17%. For the treatment group, the mean stress level decreased by 1.49 (SD = 1.93, $p < .001$, $d = 0.79$, 95 % CI [0.82, 2.15]), or 23%. This is illustrated in Figure 2 below.



(Fig. 2)

Did the treatment group show a greater stress reduction than the control group? Yes, but based on the two-sample t-test, the difference was not statistically significant ($t(66) = 0.89$, $p = .4$, 95% CI [-0.49, 1.28]). This could be due to the high variance in the treatment group, which decreased the statistical power.

Did trust in ChatGPT's intelligence correlate with the level of stress reduction in the treatment group? No. The simple linear regression analysis did not demonstrate that ($r^2 = 0.04$, $p = 0.26$).

Discussion

Why would reading history decrease momentary stress? This could be the distraction effect, as expected before the data collection (Nolen- Hoeksema & Morrow, 1993). This implies that the result of this current study does not demonstrate that the ChatGPT treatment effect is statistically different from the distraction effect.

Why was there such variance in the treatment group? One reason could be that this treatment was less efficacious for some stressors. This was suggested by some of the qualitative feedback.

How could future studies improve? They could improve by limiting the type of stressor(s), increasing the sample size, and improving participation quality.

References

Bassett, C. (2018). The computational therapeutic: exploring Weizenbaum's ELIZA as a history of the present. *AI & Society*, 34(4), 803–812. https://doi.org/10.1007/s00146-018-0825-9

Karvounides, D., Simpson, P. M., Davies, W. H., Khan, K. A., Weisman, S. J., & Hainsworth, K. R. (2016). Three studies supporting the initial validation of the stress numerical rating scale- 11 (Stress NRS-11): A single item measure of momentary stress for adolescents and adults. *Pediatric Dimensions*, 1(4). https://doi.org/10.15761/pd.1000124

Nolen-Hoeksema, S., & Morrow, J. (1993). Effects of rumination and distraction on naturally occurring depressed mood. Cognition & Emotion, 7(6), 561–570. https://doi.org/10.1080/02699939308409206