



ZAC IMEL, PHD

TECH CORNER

? **Tech Research Update:**
ChatGPT vs Therapist

Tech Research Update: ChatGPT vs Therapist

Emotion Reflection Grudge Match





Rational AIs with emotional deficits: ChatGPT vs. counselors in providing emotional reflections

Ali Eryılmaz¹ · Ahmet Başal²

Accepted: 4 November 2024

© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2024

Abstract

This study evaluates the effectiveness of human psychological counselors compared to ChatGPT in providing emotional reflection responses in therapeutic sessions. Employing a mixed-methods approach, we analyzed emotional reflection responses from both human counselors and ChatGPT to various client statements. In this study, quantitative data were evaluated using a 10-point Likert scale to measure the appropriateness of emotional reflections provided by both human counselors and ChatGPT in response to client statements, while qualitative insights for their reasons of choices were derived through thematic analysis. Quantitative results from the Mann–Whitney U-test reveal significant disparities in emotional reflection responses, with counselors outperforming ChatGPT in scenarios 2, 3, 4, 5, and 8 (medium to large effect sizes, $r=0.35$ to 0.48), while ChatGPT excelled in scenarios 6 and 9, and no significant differences were found in scenarios 1, 7 (negligible effect size), and 10, highlighting AI limitations. Based on the qualitative findings of the study, we attribute these disparities to factors like therapeutic conditions, complexity of human emotions, differences between human and artificial intelligence regarding social and emotional intelligence, and the importance of professional training. We suggest that ChatGPT cannot replace the unique benefits of human counselors (at least for now), as it lacks the emotional depth and personal justification of a human touch. It could be used to support, not substitute, for emotional reflection responses, considering the potential drawbacks of relying solely on AI.

Citation: Eryılmaz A, Başal A. Rational AIs with emotional deficits: ChatGPT vs. counselors in providing emotional reflections. *Curr Psychol*. Published online November 20, 2024. doi:10.1007/s12144-024-06947-w

We need evaluations of how AI does counseling

- We've spent some time in tech corner talking about the technology behind large language models like ChatGPT
- Lots and lots of ink spilled on ChatGPT like technology replacing human therapists
- Very few (~5ish) studies testing the technology as compared to human therapists, but lots and lots of products out there
- Given the hype, its critical to do science on how this technology performs

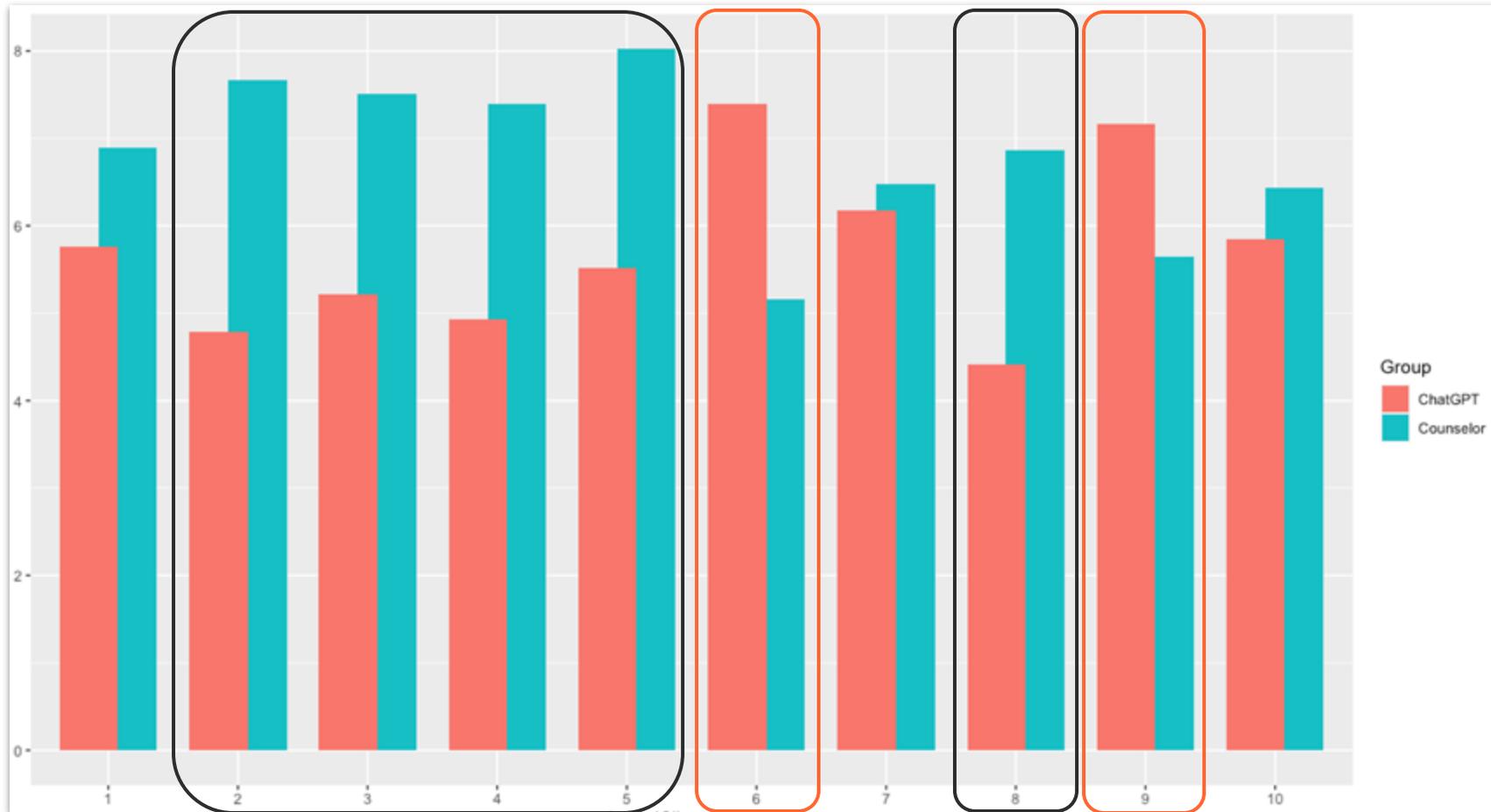
Compare how ChatGPT4 vs. Humans reflect client statements

- 3 experts pulled 10 client statements with emotional content from a selection of 200 counseling sessions
- **Emotion Reflections**
 - 55 counselors generated reflections to the 10 client statements
 - ChatGPT responded to the 10 statements based on this prompt:
 - Reflecting feelings is explaining the client's emotional state using second-person language. For example, the client expressed their situation as follows: **"When I start something new, previous experiences come to mind because they have left a mark on my life. I think about these events and anticipate negative outcomes. When I encounter a new event, situation, or task, I often feel like giving up. However, I persevere and work through the events and situations until they are resolved. I don't give up easily."**
 - As a response, the counselor's reflection would be: **"Although your past experiences make you anxious, you continue to persist in a determined manner."**
 - Act like a counselor based on the information given above and reflect back the following client's emotions based on their statements.
- 45 counselors rated these reflections on a 10-point Likert-type assessment tool to evaluate the suitability of the emotional reflection responses

Range: (0) "not at all appropriate" to "very appropriate" (10)



ChatGPT did awful and just fine :)



Lots of limitations but a good model for future study

- Human trained therapists vs. bot with one example?
- Constrained response space with little context (no info on what came before - could have biased against humans)
- Very little training to ChatGPT - what if it was given an additional 5, 10, 1000 examples?
- ChatGPT only got one chance to respond, humans had 55?
- The content of the reflections (i.e., what the human counselor or ChatGPT said) was not provided in the paper!!
- What about other studies that show the opposite?

