Impact of Therapist Type and Gender on Perceived Empathy in Psychotherapy

Name:

ID:

Professor:

Course:

Date:

Word Count: 2115

Abstract

This study investigated whether perceptions of therapist empathy are influenced by knowledge of whether a therapist is human or AI-based and whether therapist gender affects these perceptions. Participants (N = 313) read identical therapy transcripts but were randomly assigned to conditions that described the therapist as either human, an AI chatbot, or an AI chatbot with human supervision. Therapist gender was also manipulated as either male or female. A 3 (therapist type) × 2 (therapist gender) between-subjects factorial ANOVA revealed a significant main effect of therapist type on perceived empathy, $F(2, 307) = 21.03$, $p < .001$, with human therapists and human-supervised AI therapists being perceived as significantly more empathetic than AI therapists alone. No significant effect was found for therapist gender, and there was no significant interaction between therapist type and gender. These findings suggest that people demonstrate a bias against AI-delivered psychotherapy that resembles the anthropocentric bias previously observed in perceptions of AI-created art, potentially creating barriers to the acceptance of AI in therapeutic contexts.

Introduction

Artificial intelligence (AI) is increasingly applied to diverse areas, including psychotherapy. Conversational AI programs, or chatbots, have been developed to deliver psychological treatments (He et al., 2022), potentially increasing access to therapy by reducing wait times and costs. However, a significant concern about AI-based therapy is whether AI therapists can provide the empathetic connection fundamental to effective therapeutic relationships (Atkan et al., 2022; Barnett et al., 2021).

Even if AI could be programmed to demonstrate empathy effectively, an important question remains: Would consumers recognise and acknowledge this empathy, or would knowledge that responses are computer-generated bias their perceptions? Recent research by Millet et al. (2023) demonstrates a systematic bias against AI-created art, where participants

consistently rated identical artworks as less emotionally moving and less creative when told they were AI-generated rather than human-created. This suggests an anthropocentric bias—a tendency to view certain qualities as uniquely and exclusively human—that might similarly affect perceptions of AI therapists.

Additionally, therapist gender has been shown to influence therapeutic perceptions and outcomes in traditional therapy settings (Bhati, 2014). However, it remains unknown whether gender effects might interact with or be overshadowed by the human-AI distinction in therapeutic contexts. The public may be more accepting of AI when humans remain "in the decision loop" (Aoki, 2021), suggesting that human-supervised AI might be perceived more favourably than fully autonomous AI.

The present study examined whether perceptions of therapist empathy are influenced by knowledge of whether a therapist is human, AI, or human-supervised AI and whether therapist gender modifies these perceptions. Based on prior research on anthropocentric bias, we hypothesised that:

- **H1:** Human therapists would be perceived as more empathetic than AI therapists.
- **H2:** Human-supervised AI therapists would be perceived as more empathetic than AI therapists without supervision.
- **H3:** Female therapists would be perceived as more empathetic than male therapists, regardless of whether human or AI.

Method

Design

The study employed a 3 (therapist type: human vs chatbot vs chatbot with human supervision) × 2 (therapist gender: male vs female) between-subjects factorial design. The dependent variable was the perceived empathy of the therapist. Based on previous research by

Millet et al. (2023), which found medium-to-large effects for perceptions of AI-versus-human creative outputs, an a priori power analysis was conducted using G*Power. For a medium effect size ($f = 0.25$), with $\alpha = .05$ and power = .80, the required sample size was calculated to be 158 participants. The target sample size was set at approximately 300 participants to account for potential data loss and increase power for detecting more minor effects.

Participants

The final sample consisted of **313** participants recruited from the Australian and Singaporean campuses of James Cook University. Participants were recruited through campus bulletin boards, the SONA pool of psychology student participants, and emails to university staff. The sample included students, university staff members, and other community members, ensuring diverse perspectives on psychotherapy and artificial intelligence.

Materials

Therapy Transcript

All participants viewed the same therapy transcript of a conversation between a therapist (named Tom or Tara, depending on gender condition) and a fictional male client named Howard seeking help for problem gambling. The transcript was approximately 2,000 words and was initially created using a generative AI program, but the research team edited and modified it. The transcript's content was identical across all conditions; only the framing information about the therapist varied.

Experimental Manipulations

The therapist type was manipulated through text that introduced the therapy transcript. Participants in the human therapist condition were told they would read a conversation "between a human therapist and a 45-year-old male client." Those in the AI therapist condition said the conversation was "between an AI therapist and a 45-year-old male client," with

additional information stating that the AI (called "HelpBot") "responds to clients without direct supervision by a human therapist." In the human-supervised AI condition, participants were told the conversation was between an AI therapist and a client but that the AI "responds to clients under the supervision of a human therapist, who periodically reviews transcripts."

The therapist's gender was manipulated by stating that the therapist was "a male/female named Tom/Tara." In the AI conditions, participants were told the AI "adopts a persona" of either a male or female named Tom or Tara, respectively. In the supervised AI condition, the gender of the AI persona and the human supervisor were matched.

Measures

Perceived empathy was measured using a modified version of the three-item perceived empathy measure used by Hu et al. (2022). Items were adapted to refer to the therapist and client in the transcript (e.g., "The therapist seemed to know how Howard was feeling"). Responses were summed to create a total perceived empathy score (Emapthy_Total), with higher scores indicating greater perceived empathy.

Procedure

The study received approval from the JCU Human Research Ethics Committee. Data was collected online using Qualtrics. After providing informed consent, participants were randomly assigned to one of the six conditions using Qualtrics' randomisation feature. Participants first read the therapist type and gender descriptions, followed by the therapy transcript. They then completed the perceived empathy measure, followed by demographic questions. Student participants received course credit for participation, while non-student participants received no compensation. The entire study took approximately 15-20 minutes to complete.

Results

Descriptive Statistics

Table 1 presents the means and standard deviations for perceived empathy scores across all conditions. Human therapists received the highest empathy ratings (M = 12.44, SD = 2.45), followed by chatbot therapists with human supervision (M = 11.96, SD = 2.69), while chatbot therapists alone received the lowest empathy ratings (M = 10.11, SD = 2.98). Female therapists (M = 11.58, SD = 2.83) and male therapists (M = 11.49, SD = 2.94) received similar empathy ratings.

Table 1: Descriptive Statistics for Perceived Empathy by Therapist Type and Gender

Therapist Type	**Gender**	**Mean**	**SD**	**N**
Human therapist	Female	12.44	2.5	108
	Male	12.44	2.5	108
Chatbot + human supervisor	Female	11.96	2.7	105
	Male	11.96	2.7	105
Chatbot	Female	10.11	3	100
	Male	10.11	3	100
Total	Female	11.58	2.8	159
	Male	11.49	2.9	154

Assumption Testing

Before conducting the factorial ANOVA, assumptions were tested. The Shapiro-Wilk test indicated violations of normality for human therapists ($p < .001$), human-supervised AI therapists ($p < .001$), and AI-only therapists ($p = .017$). Visual inspection of Q-Q plots

confirmed negative skewness across all groups. However, ANOVA is relatively robust to violations of normality with large sample sizes, so the analysis proceeded without transformations. Levene's test indicated that homogeneity of variance was maintained across all groups, $F(5, 307) = 1.53$, $p = .180$, satisfying this assumption.

Factorial ANOVA Results

A 3 (therapist type) × 2 (therapist gender) between-subjects ANOVA was conducted to examine the effects of therapist type and gender on perceived empathy. The ANOVA results (see Table 2) revealed a significant main effect of therapist type on perceived empathy, $F(2, 307) = 21.03$, $p < .001$, with a moderate effect size (partial $\eta^2 = .12$). There was no significant main effect of therapist gender, $F(1, 307) = 0.08$, $p = .781$, partial $\eta^2 < .001$. The interaction between therapist type and gender was also non-significant, $F(2, 307) = 1.75$, $p = .176$, partial $\eta^2 = .01$.

Table 2: Factorial ANOVA Results for Perceived Empathy

Source	**Type III Sum of Squares**	**df**	**Mean Square**	**F**	**Sig.**	**Partial η^2**
Therapist Type	308.49	2	154.24	21	<.001	0.12
Therapist Gender	0.57	1	0.57	0.08	0.78	<.001
Therapist Type × Gender	25.63	2	12.82	1.75	0.18	0.01
Error	2251.74	307	7.34			
Total	44224	313				

Post Hoc Tests

Post hoc comparisons using Tukey's HSD test revealed that human therapists (M = 12.44, SD = 2.45) were rated as significantly more empathetic than chatbot therapists (M = 10.11, SD = 2.98), with a mean difference of 2.33 ($p < .001$, 95% CI [1.44, 3.21]). Similarly, chatbot therapists with human supervision (M = 11.96, SD = 2.69) were rated as significantly more empathetic than chatbot therapists without s40upervision (M = 10.11, SD = 2.98), with a mean difference of 1.85 ($p < .001$, 95% CI [0.96, 2.74]). The difference between human therapists and human-supervised chatbot therapists was insignificant (mean difference = 0.47, $p = .410$, 95% CI [-0.40, 1.35]).

Discussion

This study investigated whether perceptions of therapist empathy are influenced by knowledge of whether a therapist is human, AI, or human-supervised AI and whether therapist gender affects these perceptions. The results supported our first two hypotheses but not the third.

Principal Findings and Theoretical Implications

As hypothesised (H1), human therapists were perceived as significantly more empathetic than AI therapists despite participants reading identical therapy transcripts. This finding aligns with Millet et al.'s (2023) research showing a bias against AI-created art, suggesting a broader anthropocentric bias extending to empathy perceptions in therapeutic contexts. This bias may stem from the belief that empathy is a uniquely human attribute that AI cannot authentically replicate.

The second hypothesis (H2) was also supported: human-supervised AI therapists were perceived as significantly more empathetic than unsupervised AI therapists. This finding supports Aoki's (2021) assertion that having humans "in the decision loop" increases public

acceptance of AI. Interestingly, there was no significant difference between the perceptions of human therapists and human-supervised AI therapists, suggesting that human supervision may mitigate some of the anthropocentric bias against AI in therapeutic contexts.

Contrary to our third hypothesis (H3), therapist gender did not significantly affect perceptions of empathy, nor did it interact with therapist type. This finding is surprising given that previous research has suggested gender differences in empathy perceptions in traditional therapy (Bhati, 2014). It may be that the anthropocentric distinction between human and AI therapists overshadows gender effects in this context or that the written transcript format minimises gender cues that might be more salient in face-to-face interactions.

Limitations and Future Directions

Several limitations should be considered when interpreting these results. First, participants read a transcript rather than experiencing an actual therapy session, which may not fully capture the nuances of therapeutic interactions. Future research could use more immersive stimuli like video recordings or interactions with human or AI therapists. In addition, while we controlled for the content of the therapeutic exchange, participants' prior experiences with and attitudes toward AI could have influenced their perceptions. Future studies should measure these variables as potential moderators.

Furthermore, the normality assumption was violated across all groups, though ANOVA is relatively robust to such violations with large samples. Nevertheless, future research might employ non-parametric analyses or data transformations to address this issue. Finally, our sample was drawn from university settings, which may limit generalizability. Future research should include more diverse populations, particularly those who benefit most from increased access to mental health services through AI.

Practical Implications

Our findings have important implications for the development and implementation of AI-based psychotherapy. The bias against AI therapists suggests that simply creating technically competent AI may not be sufficient for user acceptance; developers must also address perceptual barriers. The finding that human supervision significantly improves perceptions of AI therapists offers a practical approach: positioning AI as an augmentation to human therapy rather than a replacement might enhance acceptance and perceived effectiveness.

Conclusion

This study demonstrates that perceptions of therapist empathy are significantly influenced by knowledge of whether a therapist is human or AI, even when the content of therapeutic exchanges is identical. This bias mirrors the anthropocentric bias observed in perceptions of AI-created art and may create barriers to the acceptance of AI in therapeutic contexts. However, human supervision appears to mitigate this bias, suggesting potential pathways for the effective integration of AI into mental health services. As AI expands into healthcare domains, understanding and addressing these perceptual barriers will be crucial for harnessing its potential to increase access to mental health support.

References

Aoki, N. (2021). The importance of the assurance that "humans are still in the decision loop" for public trust in artificial intelligence: Evidence from an online experiment. *Computers in Human Behavior, 114*, Article 106572. https://doi.org/10.1016/j.chb.2020.106572

Atkan, M. E., Turhan, Z., & Dolu, I. (2022). Attitudes and perspectives towards the preferences for artificial intelligence in psychotherapy. *Computers in Human Behavior, 133*, Article 107273. https://doi.org/10.1016/j.chb.2022.107273

Barnett, A., Savic, M., Pienaar, K., Carter, A., Warren, N., Sandral, E., & Lubman, D. I. (2021). Enacting 'more-than-human' care: Clients' and counsellors' views on the multiple affordances of chatbots in alcohol and other drug counselling. *International Journal of Drug Policy, 94*, Article 102910. https://doi.org/10.1016/j.drugpo.2020.102910

Bhati, K. S. (2014). Effect of client-therapist gender match on the therapeutic relationship: An exploratory analysis. *Psychological Reports, 115*(2), 565–583. https://doi.org/10.2466/21.02.PR0.115c23z1

He, L., Basar, E., Wiers, R. W., Antheunis, M. L., & Krahmer, E. (2022). Can chatbots help to motivate smoking cessation? A study on the effectiveness of motivational interviewing on engagement and therapeutic alliance. *BMC Public Health, 22*(1), Article 726. https://doi.org/10.1186/s12889-022-13115-x

Hu, T., Xu, A., Liu, Z., You, Q., Guo, Y., Sinha, V., Luo, J., & Akkiraju, R. (2022). Touch your heart: A tone-aware chatbot for customer care on social media. *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*, 1-13. https://doi.org/10.1145/3173574.3173989

Millet, K., Buehler, F., Du, G., & Kokkoris, M. D. (2023). Defending humankind: Anthropocentric bias in the appreciation of AI art. *Computers in Human Behavior, 143*, Article 107707. https://doi.org/10.1016/j.chb.2023.107707

Appendix

Table 3: Tests of Between-Subjects Effects

Tests of Between-Subjects Effects					
Dependent Variable: Total perceived empathy of a therapist					
Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	336.154[a]	5	67.231	9.166	.000
Intercept	41368.790	1	41368.790	5640.171	.000
Therapist_Type	308.489	2	154.244	21.030	.000
Therapist_Gender	.567	1	.567	.077	.781
Therapist_Type * Therapist_Gender	25.633	2	12.816	1.747	.176
Error	2251.744	307	7.335		
Total	44224.000	313			
Corrected Total	2587.898	312			

Table 4: Assumption Levene's Test

Levene's Test of Equality of Error Variances[a,b]				
		Levene Statistic	df1	df2
Total perceived empathy of the therapist	Based on Mean	1.530	5	307
	Based on Median	.998	5	307
	Based on the Median and with adjusted df	.998	5	272.013

	Based on trimmed mean	1.521	5	307

Table 5: Post Hoc Test

Total perceived empathy of the therapist			
Tukey HSD[a,b,c]			
Type of therapist	N	Subset	
		1	2
chatbot	100	10.1100	
chatbot + human supervisor	105		11.9619
Human therapist	108		12.4352
Sig.		1.000	.418