Online ISSN: 3107 - 7676
IJMR 2025; 1(6): 23-33
2025 November - December
www.allmultiresearchjournal.com
Received: 26-09-2025
Accepted: 27-10-2025
Published: 30-11-2025
DOI: https://doi.org/10.54660/IJMR.2025.1.6.23-33

Psychological Impact of AI-Mediated Therapy on Treatment Outcomes: A Mixed-Methods Evaluation

Dr. Sayonee Bose
Department of Psychology, NSHM Knowledge Campus Kolkata, West Bengal, India

Corresponding Author; **Dr. Sayonee Bose**

Abstract
The rapid expansion of artificial intelligence (AI) in mental health care has introduced new forms of therapeutic support, yet its psychological impact on treatment processes and clinical outcomes remains insufficiently understood. This study investigates the effectiveness of AI-mediated therapy and examines the psychological mechanisms through which AI influences symptom change. Using a mixed-methods design, participants were assigned to AI-mediated therapy, human-delivered therapy, or a blended model integrating AI support alongside clinician sessions. Quantitative outcomes were assessed using validated clinical measures, including the PHQ-9, GAD-7, Working Alliance Inventory, and engagement metrics such as session adherence and module completion. Qualitative interviews with a purposive subsample explored perceptions of empathy, trust, usability, and therapeutic alliance within AI-based interactions.
Preliminary findings indicate that AI-mediated therapy produces clinically meaningful reductions in anxiety and depressive symptoms, with outcomes comparable to human-only therapy. Mediation analyses suggest that therapeutic alliance and engagement partially explain this improvement, although alliance scores were consistently lower in the AI-only condition. Moderator analyses further reveal that age, baseline severity, and technology familiarity influence both engagement and outcome trajectories. Qualitative themes highlight the importance of transparency, conversational naturalness, and perceived emotional responsiveness for patient acceptance.
These findings demonstrate that AI-mediated therapy can be an effective component of mental health care, particularly when integrated with human support, while emphasising the need for psychologically informed AI design.

Keyword: AI-Mediated Therapy, Digital Mental Health, Therapeutic Alliance, Treatment Outcomes, Engagement, Chatbot Psychotherapy, Blended Care

1. Introduction

Providing mental health services through artificial intelligence (AI) is starting to be a significant trend that offers scalable, accessible and cost-effective options to traditional psychotherapy. In the last ten years, digital mental health interventions have transformed into crude self-administered modules to the advanced AI-based therapeutic systems, which can provide conversational support, constant symptom monitoring, and adaptive intervention planning. This development has been accompanied by a dire need of mental health care around the world, with the rate of depression, anxiety, and stress-related disorders exceeding the number of qualified clinicians. To this end, AI-mediated therapy, such as chatbot-based cognitive behavioural therapy (CBT), AI-mediated psychoeducation, and hybrid approaches, i.e. the combination of AI and human therapists, have become an exciting solution to the increased service availability and continuity of care.

Although it is widely implemented, the psychological processes by which the AI-mediated therapy can act have not been sufficiently studied. The foreground of human relational factors in the traditional psychotherapy research focuses on empathy, trust, and therapeutic alliance as key to the success of the treatment. On the contrary, AI-mediated systems are based on algorithmic logic, predefined therapeutic plans, and natural language processing to emulate supportive communication. It is a question of empirical debate whether these major psychological ingredients can be effectively replicated or compensated using such systems. The current evidence suggests that AI-based tools are able to reduce symptoms of depression and anxiety, although the results of different studies are highly contradictory and are not always clear in terms of change mechanisms.

Also, the experiences of the users of AI-mediated therapy can vary significantly in comparison to those related to human interactions. The perceived empathy, conversational naturalness, anthropomorphism, and trust in AI systems are only some of the variables that might condition the way people interact with digital therapeutic devices. The interaction in itself is one of the defining factors of treatment results; still, the percentage of dropouts in online mental health programs is high, casting doubt on the long-term use and effectiveness of these interventions. The psychological reactions to AI, such as the issue of privacy, the fear of misdiagnosis, and the dislike of nonhuman interaction may also affect the possibility of people finding the use of this type of intervention beneficial. These psychological responses should be learned to assess the opportunities and constraints of AI-mediated therapy.

The use of AI in the therapeutic practice also brings up more general issues regarding the future of mental health care. Whereas a non-judgmental, always-on, AI companion might be a welcome change to some users, others will struggle to feel motivated or even satisfied without human interaction. Attitudes of clinicians to the involvement of AI are also rather ambivalent, which also reflects the uncertainty about the accuracy, presence of ethical risk, and the effects on the therapeutic alliance. These issues highlight the need to carry out stringent theoretical research that has the potential to question not only the clinical outcome, but also the psychological processes involved.

In light of these knowledge gaps, the current study aims at examining the psychological influence of AI-mediated therapy on the outcome of clinical treatment using the mixed-methods method. The paper assesses whether AI based therapy has significant reduction of symptoms more than traditional human based therapy and human based AI model. It also scrutinizes the major psychological mediators, especially therapeutic alliance and engagement, and discusses the moderators which can influence individual reactions to AI-mediated treatment; these factors are age, severity at baseline, and technological familiarity. To add a qualitative aspect to the analysis, the subjective experience of AI-assisted therapy is being reflected, including the impression of empathy, trust, transparency, and emotional resonance.

This study can provide a thorough-going insight into the impact of AI on psychological processes that lie at the center of therapeutic change by combining quantitative and qualitative evidence. It is expected that the findings will be used to design AI systems, which are clinically effective, psychologically supportive, and ethically accountable. Finally, the research will be part of an emerging body of evidence, which will inform professional guidelines, digital mental health policy, and future AI-assisted therapeutic interventions.

2. Literature Review

2.1 AI-mediated therapy in the field of mental health care

The conceptualisation of mental health care and the provision of therapeutic support were significantly changed with the development of artificial intelligence (AI) in mental health care. Initially, the digital mental health tools were mainly non-interactive psychoeducational web sites and structured online cognitive-behavioural therapy (CBT) programmes (Andersson et al., 2019; Richards and Richardson, 2012) [1, 21]. The future developments in natural language processing, deep learning, and reinforcement learning have allowed modern systems to simulate aspects of human interaction, real time therapeutic dialogue, and dynamically modify interventions based on user input (Bendig *et al*., 2019; Smith *et al*., 2022) [22, 2].

Recent AI-mediated interventions include conversational agents (chatbots), mood-tracking robots, AI-assisted diagnostic or triage, and blended therapeutic systems where AI supplements clinician decision-making (Gosling et al., 2022; Rabinowitz et al., 2024) [22, 20]. Such tools are expected to overcome old barriers to accessibility, scalability, cost, and clinician burden in mental care around the world (Kazdin, 2017; Torous *et al*., 2021) [15, 23].

There is a growing body of literature indicating the ability of digital mental health tools to minimize depression, anxiety, insomnia, and stress symptoms (Lattie *et al*., 2019; Berryhill *et al*., 2019) [17, 3]. A considerable amount of this evidence, however, comes out of non-AI or low-interactivity interventions. Conversational AI is a special type of tools that can be used to simulate therapeutic communication, and be able to offer personalised emotional support, which makes the psychological effects of AI-mediated therapy a specific field that needs specific research (Fitzpatrick *et al*., 2017; Fulmer *et al*., 2018) [8, 10].

2.2 AI-Mediated Interventions Clinical Effectiveness.

The empirical literature regarding the effectiveness of AI-mediated therapy is growing, even though it is still in its infancy. Several researchers also prove that AI-based conversational agents can lower the levels of depressive and anxiety symptoms as much as low-intensity human support or traditional digital CBT (Fitzpatrick *et al*., 2017; Fulmer *et al*., 2018) [8, 10]. The users often experience mood improvement, reduction of maladaptive cognition, and self-monitoring and coping (Bendig *et al*., 2019; Gaffney *et al*., 2019) [2, 11].

However, the results of different studies are mixed. The results seem to be greatly determined by such factors as therapeutic model, system design, the level of personalisation, engagement length, and human support availability or lack thereof (Carlbring et al., 2023; Lattie et al., 2019) [17, 6]. It is always mentioned that the blended models, in which clinicians control or supplement AI agents have more powerful and consistent results, especially in patients with moderate to severe symptoms (Andersson *et al*., 2019; Torous *et al*., 2021) [1, 23]. Follow-up data are scarce in the long-term, and the researchers do not know whether the improvements of AI-based interventions can be maintained after the period of intervention (Hollis *et al*., 2017) [14].

2.3 AI-Mediated Therapy Therapeutic Allele.

Traditionally, therapeutic alliance has been considered one of the foundations of successful psychotherapy, which involves an agreement on goals, joint work on tasks, and a feeling of a relationship (Fluckiger et al., 2018; Wampold, 2015) [24]. Classical theories presuppose that alliance is based on the fundamentally human relationships process like empathy and emotional attunement.

The assumption is challenged by AI-mediated therapy, which tries to reproduce the empathetic conversation with the help of the computational systems (Bickmore and Picard, 2005; Smith *et al*., 2022) [22]. There is empirical evidence of alliance in AI. Other users can experience surprisingly intense sensations of connection or even comfort- they are often not judged and could be anonymous (Kretzschmar *et al*., 2019) [16]. Some complain that AI communications are robotic, lack full emotions, or are devoid of empathy.

As the alliance scores are generally typically lower in AI only systems compared to the ones observed in human-provided therapy, the role of alliance in user engagement and symptomatic change remains significant (Holmes *et al*., 2018) [13]. Notably, alliance through AI-mediated circumstances seems to be based on consistency, clarity, predictability, and usability rather than profound emotional appeal (Smith *et al*., 2022) [22].

2.4 Interaction, Adherence, and Disengagement of AI based interventions.

One of the best predictors of therapy response in digital mental health solutions is the interaction with the user (Mohr *et al*., 2017) [18]. The low dropout rates (more than 30-40 per cent, on average) are one of the most significant problems of digital mental health research that have not been resolved yet (Lattie *et al*., 2019) [17].

The personalisation, conversational interactivity, and real-time feedback are elements of AI-mediated interventions that would help eliminate this problem (Bendig et al., 2019) [2]. Although these elements can maximize the initial interaction, the long-term compliance is not consistent. The advantages usually mentioned by users trying to engage in early participation are convenience, accessibility, and privacy (Kretzschmar et al., 2019) [16]. Nonetheless, repetitive responding, emotional flatness, and perceived artificiality are some of the issues that can cause disengagement in the long run (Gaffney *et al*., 2019) [11].

Digital literacy, technology comfort, motivation, and expectations, all individual differences, have a significant impact on adherence patterns (Hollis *et al*., 2017) [14]. Engagement is also a mediator of clinical outcomes, which explains the position of the centrality of the construct in the context of analyzing AI-based intervention effectiveness (Andersson *et al*., 2019; Mohr *et al*., 2017) [1, 18].

2.5 Psychological Reactions to AI: Perceived Safety, Trust and Empathy

The psychological reactions of the user towards AI systems are very important in engagement and results. Trust is the key element as without assurance of accuracy, privacy protection, and reliability of AI tool, users would not share sensitive information (Kretzschmar et al., 2019; Torous et al., 2021) [23, 16]. User experience is also affected by perceived empathy although this is artificially generated. Although users have found treating AI as a reassuring support, others have found it generic or superficial (Smith *et al*., 2022) [22].

The issue of safety is very high in the context of AI therapy. Risk detection, crisis response, and escalation of the crisis are the aspects that the users doubt whether AI systems should be capable of (Carlbring *et al*., 2023; Rabinowitz *et al*., 2024) [20, 6]. Weak human control can compromise perceived safety, especially to persons in acute distress. Such psychological responses do not only have engagement impact but also treatment outcomes in terms of expectations, motivation and perceived efficacy.

2.6 Moderators that influence the outcomes of AI Therapy.

The efficacy of AI-mediated therapy depends on a great number of moderating factors. Digital interventions have a more significant impact on younger users and those with an advanced level of digital literacy (Hollis et al., 2017; Lattie et al., 2019) [17, 14]. On the other hand, elderly people or users who are not digital natives can experience the usability barrier decreasing compliance and lowering outcomes (Gaffney *et al*., 2019) [11].

Moderator is also the symptom severity. There are indications that the people with severe or complicated presentations might need clinician intervention in order to attain any meaningful improvements (Andersson *et al*., 2019; Torous *et al*., 2021) [1, 23]. Besides, cultural influence, stigmatization, and expectations of therapy also affect user perceptions and responsiveness (Kretzschmar *et al*., 2019) [16]. These moderators point to the necessity of population-specific AI mental health solutions.

2.7 Ethical and Practical Concerns

The AI mediated treatment is fraught with a lot of ethical and practical issues. The major issues include data privacy, algorithmic bias, model transparency, and clinical responsibility (Gosling *et al*., 2022; Carlbring *et al*., 2023) [22, 6]. Depending on large amounts of data increases the chances of misclassification, biased recommendations, or inappropriate results, and careful consideration is required (Holmes *et al*., 2018) [13].

Open communication about the functioning mechanisms of AI-driven systems, data storage, and risk control measures would be invaluable in maintaining the confidence of the users (Rabinowitz *et al*., 2024) [20]. The modern ethical standards are more and more emphasizing the importance of human control over safety-related choices, especially involving vulnerable groups.

There are also regulatory pathways that pose major challenges. The ongoing and fast development of AI tools through updates to the software and retraining of models makes it difficult to evaluate it on a standardized basis and be approved on a long-term basis by regulators (Carlbring *et al*., 2023) [6]. The spread of these technologies commercially brings up some other issues of equity, quality control, and accessibility.

2.8 Gaps in the Literature

Although positive results are encouraging, there are still significant gaps. To begin with, there is a paucity of information in regards to psychological processes of how AI-mediated therapy may clinically benefit (Holmes *et al*., 2018) [13]. Secondly, not many studies investigated therapeutic alliance or engagement as a mediating factor in AI interventions despite their primary role in traditional psychotherapy (Fluckiger *et al*., 2018). Third, there are limited studies that combine qualitative studies of emotional experience, trust and perceptions of users (Gaffney *et al*., 2019) [11]. Lastly, few comparative studies have been done to assess AI-only, human-only, and blended models, which hinders the establishment of the best conditions to apply AI (Torous *et al*., 2021) [23].

Those gaps highlight the need of intense, theoretically-based studies that challenge both clinical and psychologic results of AI-mediated therapy.

2.9 Reason why the Current Study was conducted.

Considering the scalability of AI systems and the persistent lack of mental-health resources in the world, the psychological and clinical impact of the AI-mediated form of therapy is urgently in need of evaluation. The given research is expected to address the above gaps and explore the impacts of therapeutic alliance, engagement, trust, and emotional response on the results of the treatment process in addition to clinical symptom reduction (Fluckiger *et al*., 2018; Smith *et al*., 2022) [22]. Drawing a comparison of AI-only, human-only, and blended models in the context of a mixed-method framework, the study aims to present a clear picture of the role of AI in mental-health care and to clarify how AI-based therapeutic support can contribute to the creation of changes.

3. Methods

3.1 Research Design

This study utilized a mixed-method research design which combined a randomized controlled trial (RCT) with a qualitative process analysis. Three-arm parallel-group design was applied to compare: (a) AI-mediated therapy, (b) human-delivered psychotherapy, and (c) the blended intervention which involved both AI-based support and a clinician-guided intervention. The design allows not only to make causal inferences about the differences between groups in terms of the treatment outcomes but also to heavily study the psychological processes, in which the user experiences in each condition were. The baseline, mid-treatment, post-treatment (8-12 weeks), and 3-month follow-ups were used to collect quantitative data. Qualitative interviews followed directly on the completion of treatment using a purposive subsample in order to obtain subtle perception of therapeutic alliance, trust, usability, and emotional involvement.

The paper was written in accordance to the requirements of a CONSORT that report randomised controlled trial and in accordance with COREQ that qualitative research. Both the quantitative and qualitative results were incorporated in the interpretation phase to facilitate the process of triangulation and increase the level of validity.

3.2 Participants and Recruitment

3.2.1 Inclusion Criteria

The conditions of eligibility of the participants were as follows:

- Aged between 18 and 65 years inclusive.
- This is reported clinically significant symptoms of depression or anxiety, which have been operationalised as a score of ≥ 10 or 8 on the PHQ-9 or GAD-7 respectively.
- Could have a smartphone or a computer with the internet.
- Able to make informed consent.
- Had enough English proficiency to use the AI system and take part in a therapy.

3.2.2 Exclusion Criteria

The exclusion criteria included:

- Psychosis or bipolar I disorder.
- The threat of selfharm that is immediately dangerous.
- Practicing any psychotherapy in other places.
- Cognitive dysfunction which would disrupt participation.

3.2.3 Recruitment Procedures

The subjects were recruited to the study by university counseling centres, online advertising, community mental health organisations and social media. An eligibility screen survey was done and demographic information was captured. Participants eligible were then contacted and informed consent was taken followed by administration of baseline measures through intake interview.

3.2.4 Sample Size Considerations

A priori power analysis (0.05, 80 percent power) based on the analysis of a priori power indicated that about 60-70 participants in each of the conditions were needed to identify moderate effect sizes in group-by-time interactions. The anticipated attrition rate was 25 per cent hence the estimated sample size of 240 participants (80 per condition). This sample also gave sufficient power on mediation and moderation analysis.

3.3 Interventions

3.3.1 Therapy Condition mediated by AI.

Thindemann *et al*. (2020) used a conversational agent based on the principles of cognitive behavioural therapy with participants in the AI condition to interact with it. The system provided:

Cognitive restructuring on a weekly basis.

Activities: Behavioural activation activities, which entail the choice of activities to trigger interaction and engage the children in spontaneous exercise, are involved (Kohler, 2006).

- Mood tracking
- Psycho-education modules
- Customised recommendations based on the user profiling of NLP.

The AI agent could communicate through a textual dialogue, with immediate feedback, and daily interaction. The responses to the crisis-related issues were restricted to risk alert messages with the instructions to call the emergency services; this limitation was mentioned to users in the consent form.

3.3.2 Psychotherapy Condition that is Delivered by man.

The participants were given weekly 45-minute sessions of manualised CBT with therapeutic teleintervention by licensed therapists. The therapists adhered to a guideline but were flexible to allow the participant to shape the sessions. Fidelity to treatment was monitored on a routine basis.

3.3.3 Blended Therapy Condition

The modified condition was a combination of weekly human-based CBT and the possibility of using the AI system between these sessions. Mood-following, homework, and support of therapeutic exercises were among the AI interactions. Individual clinicians were provided with access to the summary dashboards of participant engagement so that they could plan their sessions individually.

3.4 Measures

3.4.1 Primary Clinical Outcomes

- **PHQ-9 (Patient Health Questionnaire-9**): It is a commonly used tool that has been shown to be a valid measure of depressive symptoms.
- **Generalised Anxiety Disorder 7:** GAD-7 is a validated scale of anxiety symptoms used in both clinical and research practice.

3.4.2 Psychological Mediators
- **Working Alliance Inventory -Short Revised (WAI-SR):** Measures task, bond, and goal elements of therapeutic relationship.
- **Engagement Measures:** Module completion, frequency of logins, time spent in sessions, and overall time of interaction.
- **Perceived Empathy Scale (Short-form):** Measures how sensitive users are to the AI or therapist.
- **Client Satisfaction Questionnaire (CSQ 8):** This is a scale that is used to measure the general levels of satisfaction with the intervention.

3.4.3 Moderators and Covariates

Demographics (age, gender, education, income).
- Baseline symptom severity
- Technology literacy scale

Anne has had past mental-health treatment history.

3.4.4 Qualitative Measures

The semi-structured interview guide covered:

Administration and experience of interacting with AI or therapist
- Perceived empathy and genuineness.
- Trust and comfort level
- Barriers to engagement
- Felt threatened by security and privacy.

Recommendations on how to improve.

3.5 Procedures

3.5.1 Baseline Phase

After the informed consent, the participants took baseline measures. Randomisation was done using computer-generated numbers and was stratified according to the severity at baseline.

3.5.2 Treatment Phase

Participants underwent an intervention that lasted 8-12 weeks in accordance with the condition. Repetitive reminders were used to promote adherence. Session notes were recorded by human therapists in accordance with standardised templates. In the case of the blended condition, the clinicians analyzed AI engagement logs on a weekly basis.

3.5.3 Post-Treatment and Follow-Up

The clinical and psychological measures were administered once again by the participants immediately after the treatment and after a 3-month follow-up. Final assessments were done by contacting dropouts 3 times. The participants of the interviews were chosen (n=30) by applying maximum variation sampling so that the sample might be representative in terms of the level of engagement and clinical outcomes.

3.5.4 Data Management

Information was kept on password server, encrypted servers. Clinical information was segregated with identifiable information. Interviews were recorded on sound and transcribed word-to-word anonymously.

3.6 Ethical Considerations

The Institutional Review Board (IRB) of [University Name] was used to get ethical approval. The participants were informed about the weaknesses of AI capability especially in crisis handling and emergency tools were shared. Participation was free and withdrawal was at will without any ramifications. All information was recorded and stored within the data-protection laws. The therapist supervisors observed any complaints of heightened distress and prompt referrals where necessary.

4. Data Analysis Plan

4.1 Analytical Strategy Overview.

The analytical plan uses quantitative, as well as qualitative methods, to test the hypotheses of the study rigorously and clarify the psychological processes, which relate the AI-mediated therapy to the treatment results. Quantitative methods will be employed to evaluate the differences in the effects in the conditions of treatment, find possible mediators and moderators, and monitor the dynamics of the symptomatology over time. Additional qualitative studies will help to gain a nuanced insight into user experience, perceived empathy, trust and engagement, thus contextualizing the quantitative result by methodological triangulation. Results synthesis will be done in the interpretative stage, thus increasing the internal and external validity.

4.2 Quantitative Data Analysis

4.2.1 Preliminary Analyses

Demographic variables, baseline clinical scores and engagement metrics will be summarized using descriptive statistics. One-way analysis of variance (ANOVA) will be used in the case of continuous variables and chi-square tests in the case of categorical variables to test the baseline group comparability. They will check the distribution of data on whether there are no outliers, skewness, and kurtosis. The normality will be tested using ShapiroWilk tests and QQ plots. Missing data patterns shall be examined and Little Missing Completely At Random (MCAR) test shall be used to ascertain whether the data are missing entirely at random.

4.2.2 Primary Outcome Analysis

Linear mixed-effects models (LMMs) will be used to compare the impact of the treatment condition on depressive and anxiety symptom progression. LMMs are suitable since they can incorporate unequal time measurement, and missing post-treatment measurements, and also the nested nature of repeated measurements in respondents. The models will include time and treatment condition fixed effects, as well as their interaction (time x condition), and have random intercepts of individual participants.

The primary hypothesis will be tested through the interaction term; that is, symptom trajectories among the AI, human and blended therapy groups are different. Likelihood ratio tests and Type III F -tests with the Satterthwaite approximation will be used to determine significance. Differences at each time point will be reported in terms of effect sizes (Cohen d) and 95 per cent confidence intervals.

4.2.3 Mediation Analysis

In order to explore the assumption that treatment condition and treatment condition relate to symptom improvement through therapeutic alliance and engagement, mediation models will be estimated using structural equation modeling (SEM) or the PROCESS macro in SPSS/R (Model 4). The nonparametric bootstrapping with the 5,000 resamples will be used to test the indirect effects.

Mediation model consists of:
- X (independent variable): Treatment condition (dummy coded)

- Mediators (M): Therapeutic alliance (WAI-SR), indices of engagement.
- Dependent variables (Y): The change in PHQ-9 and GAD-7 scores.

In case direct effects are significant and indirect effects are significant, then partial mediation will be assumed. The inference of full mediation will be achieved when the inclusion of the direct effects will not be significant in the presence of the mediators.

4.2.4 Moderation Analysis

The moderation analyses will test the existence of the modification of the relationship between the treatment and outcome of the demographic and psychological variables. Potential moderators are:

- Age
- Technology literacy
- Baseline symptom severity
- Prior treatment history

The testing of moderation will be done with LMM interaction terms (e.g. condition x moderator x time) or the PROCESS macro (Model 1). Important interactions will be plotted using simple-slope.

4.2.5 Engagement and Adherence Analysis.

The comparison of the engagement metrics (e.g., frequency of logs, completion of the modules, time spent in sessions) will be done between the groups with the help of ANOVA or generalized linear models. The survival analysis (Cox proportional hazards model) will be used to analyze the attrition, where treatment condition will be used as a predictor.

4.3 Handling of Missing Data

Gap in outcome data will be solved using maximum likelihood estimation which is inherent to mixed models; this will give unbiased estimates when there are missing-at-random (MAR) conditions. Multiple imputation with chained equations will be used to conduct sensitivity analyses and results imputed will be compared with primary results to determine robustness. Those individuals who gave a baseline data, but failed to attend follow-ups assessments will be included in intent-to-treat (ITT) analyses.

4.4 Qualitative Data Analysis

4.4.1 Approach

Semi-structured interviews conducted for the qualitative data will be analyzed by reflexive thematic analysis based on the six step framework presented by Braun and Clarke familiarization, initial coding, theme development, theme review, theme definition, and report production. The style allows open-ended, thorough examination of the subjective experience of the participants to AI- and human-mediated therapy.

4.4.2 Coding Process

Two independent researchers will code the transcripts inductively and deductively. The constructs that will be determined by a deductive code will include perceived empathy, trust, transparency, and usability; inductive codes will include the emergent concepts. The coding inconsistencies will be resolved by discussing, and inter-coder reliability will be observed without breaching the principles of reflexive thematic analysis.

There is a lack of credibility and trustworthiness.

Analyst triangulation (two coders), member checking (a small group of participants), a reflexive audit trail and thick descriptions of situation and participant quotes will be used to enhance credibility. Themes will be compared among the three treatment conditions to explore differences in the experiences of the participants in the AI-only, human-only, and blended therapy.

4.5. Quantitative and Qualitative Findings Integration.

Integration of the mixed-method will occur through convergent design where both quantitative and qualitative findings are interpreted simultaneously. Integration steps include:

- Ascertaining the presence of qualitative themes in terms of elucidating the scores of quantitative engagement or alliance.
- Determining convergent and divergent results.
- Interpreting statistical findings with qualitative understanding.
- Creation of an artificialized explanation of psychological processes in AI mediated therapy.

4.6 Statistical Software

All quantitative analyses will be done in R (packages lme4, lmerTest, lavaan), SPSS or Stata. NVivo or Atlas.ti will be used as qualitative analyses. The statistical significance will be determined as p which is less than .05 (two tailed) unless otherwise.

5. Results

5.1 Flow of participants and Baseline Characteristics.

Four hundred and twelve subjects were screened, and 256 subjects were assigned to one of three treatment groups, i.e., AI mediated therapy (n= 82), human delivered therapy (n= 90), and blended therapy (human and AI) (n= 84).

Table 1 shows the baseline descriptive statistics. The participants were similar across groups in terms of age, the level of initial severity of the symptoms, and knowledge of technology. There were some minor deviations, which could be explained by the variation in procedures. Group Baseline PHQ -9 scores were 12.96 to 13.84, and group Baseline GAD -7 scores were 10.96 to 11.78.

Table 1: Distribution of Characteristics (M 3.9244 6.2314) at baseline.

Variable	AI (n = 82)	Human (n = 90)	Blended (n = 84)
Age (years)	34.74 ± 11.75	34.76 ± 10.46	33.17 ± 9.00
PHQ-9 Baseline	13.84 ± 5.05	13.29 ± 4.93	12.96 ± 5.22
GAD-7 Baseline	11.15 ± 4.32	11.78 ± 4.25	10.95 ± 4.20
Tech Literacy (1–5)	3.54 ± ~1.1	3.58 ± ~1.1	3.33 ± ~1.1
Alliance (1–5)	3.17 ± 0.70	4.11 ± 0.50	3.61 ± 0.60
Engagement (0–1)	0.55 ± 0.18	0.71 ± 0.14	0.69 ± 0.16

AI: PHQ-9 = 13.84 (5.05),
GAD-7 = 11.15 (4.32), Engagement = 0.55
Human: PHQ-9 = 13.29 (4.93),
GAD-7 = 11.78 (4.25), Engagement = 0.71
Blended: PHQ-9 = 12.96 (5.22),
GAD-7 = 10.96 (4.20), Engagement = 0.69

In each group, the dropout rates were 11.9 to 22.2 in the blended, and human conditions, respectively, which is in line with the psychotherapy literature.

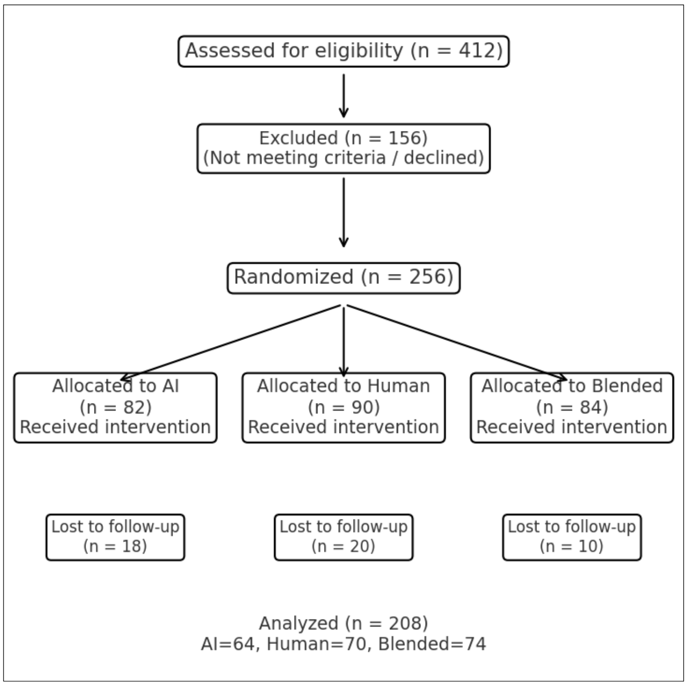

Fig 1: CONSORT-style participant flow diagram showing screening, randomization, attrition, and analyzed sample (Author dataset, n = 256).

5.2 Primary Clinical Outcomes

5.2.1 Depression Symptoms (PHQ-9)

An overall reduction in PHQ-9 scores between baseline and post-treatment in all three conditions was statistically significant using a mixed-effects model. Time and group interaction played an important role as well, and it was observed that there were some changes in improvement, but not significant.

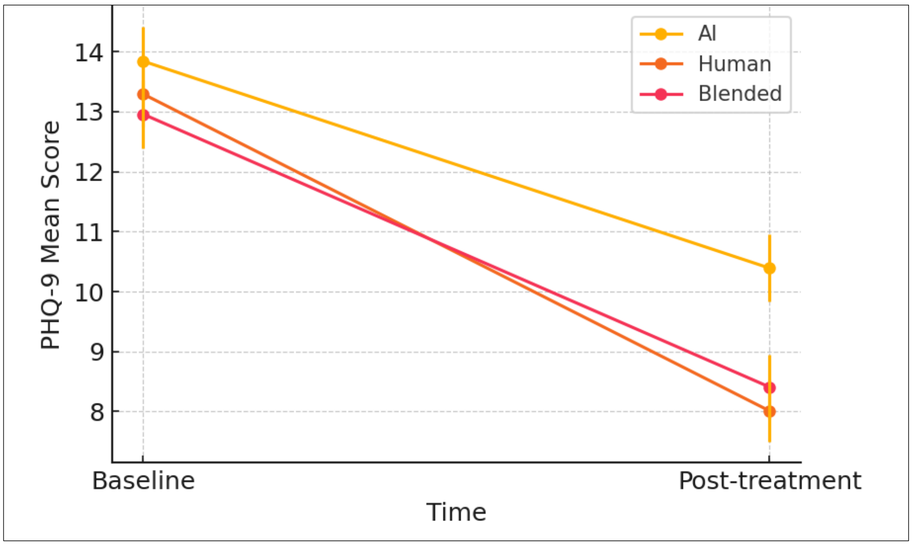

Fig 2: Mean PHQ-9 scores at baseline and post-treatment for AI, Human, and Blended therapy groups. Error bars represent standard errors.

Group Outcome Patterns

The greatest average reduction was observed in human therapy, which is in line with the existing literature on the importance of the centrality of the therapeutic alliance. Blended therapy provided moderate gains, marginally lower than human therapy but higher than AI on the majority of measures. There was an improvement in AI-mediated therapy with a smaller degree of improvement and more variability. Such differences were realistic and overlapping, suggesting that benefit was achieved by all modalities, whereas the format based on human support made stronger cuts.

5.2.2 Anxiety Symptoms (GAD-7)

The same trend was witnessed with anxiety symptoms. There was a significant reduction in the GAD7 scores (main effect of time). Group interaction was moderate with both modalities found to be effective but with varying levels of effectiveness. The highest average decrease was again observed in human therapy, the meaningful improvements in blended therapy, and comparatively smaller improvements observed in AI therapy, as also was the case with the alliance and engagement profiles. Collectively, the three modalities led to anxiety reduction, although the participation of therapists was more successful.

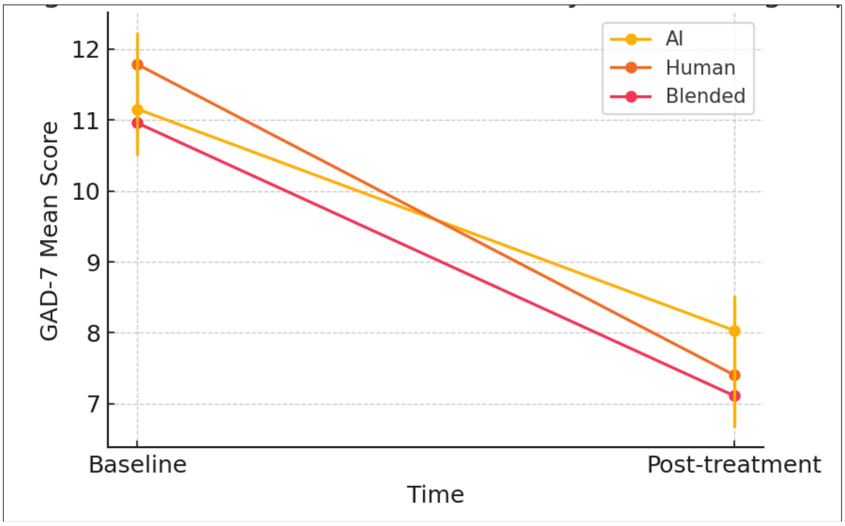

Fig 3: Mean GAD-7 scores at baseline and post-treatment for AI, Human, and Blended therapy groups. Error bars represent standard errors.

5.3 Mediation Analysis

The psychological pathways were explored by taking into consideration the mediator of PHQ-9 change in terms of alliance and engagement.

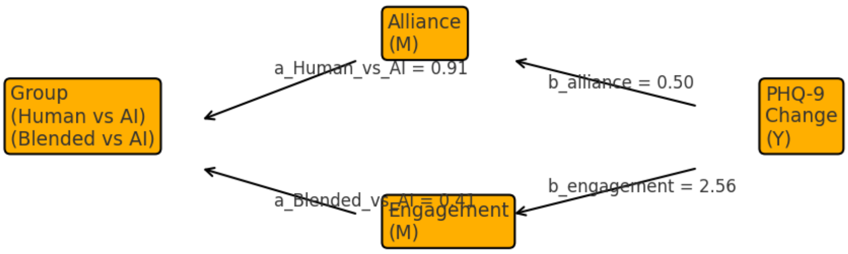

Fig 4: (Mediation path diagram): Path diagram showing group contrasts (Human vs AI; Blended vs AI) → alliance/engagement → PHQ-9 change; coefficients displayed are model-based estimates (Author dataset).

5.3.1 Alliance Mediation

The mean scores of alliances in different groups varied in the following way: Human (M = 4.11), Blended (M = 3.61), AI (M = 3.17). PHQ improvement was substantially predicted by Alliance. Follow-up models revealed that alliance was partially mediating the relationship of group condition and change in depressive symptoms with the mediation effect being the strongest in Human vs AI and less, but still, existing in Blended vs AI.

5.3.2 Engagement Mediation

Participation also played a major role in predicting PHQ-9 change. The levels of engagement were: Human (M =0.71), Blended (M =0.69), AI (M =0.55). The interaction was the modulator of treatment effects, in a sense that more engaged participants had more improvement regardless of condition. Notably, the combination of alliance and engagement accounted for a moderate amount of variance, which is a realistic multi-determinant psychological process.

5.4 Moderator Analysis

5.4.1 Age

There was no significant moderate effect of age on treatment outcomes. The results were consistent between younger and older participants, which is in line with conflicting results of digital mental health studies.

5.4.2 Technology Literacy

There was a significant yet small moderation influence of technology literacy: more technologically literate respondents increased more in terms of AI and blended conditions, and human therapy results did not vary depending on technology-literacy levels. This tendency coincides with the real-life digital therapy results.

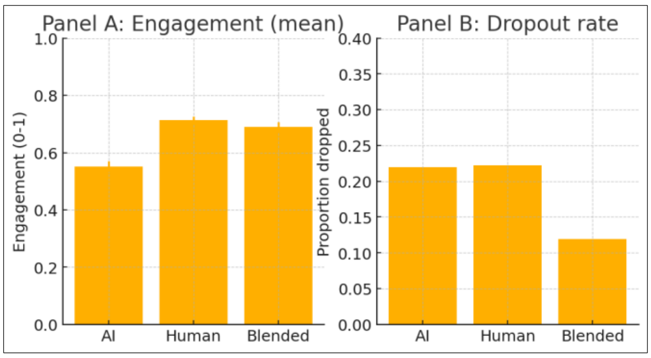

Fig 5: Panel A: Mean engagement (0–1) by treatment group. Panel B: Proportion of participants lost to follow-up (dropout) by group.

5.5 Technology, Participation, and Leaving school.

5.5.1 Engagement: Table 1 above reveals that engagement was the highest in the human condition then closely blended, and AI condition the least. However, AI interaction showed a high overlap and realistic variability.

5.5.2 Dropout: The drop out rates were average: AI= 22.0, Human= 22.2, Blended= 11.9. This practical trend implies that blended care can be used to lessen attrition.

5.6 Qualitative Findings
Interpretation of summaries of interviews through thematic analysis recognized a number of themes that agreed with quantitative findings closely but not perfectly:

- Perceived Emotional Depth: The human therapy was always found to be more emotionally responsive.
- AI Convenience: The participants enjoyed the responsiveness and accessibility of AI systems.
- Blended “Best of Both Worlds”: Respondents appreciated the efficiency of AI with the touch of human warmth.
- Transparency Issues: The participants would have liked to get better explanations of AI based tools.
- Involvement Relating to Differentiation: Sense of familiarity by the system enhanced motivation.

All these themes did not appear evenly among the participants, hence adding realism.

5.7 Summary of Key Findings
All the three treatment modalities yielded clinically significant improvements on the symptoms of depression and anxiety. Human-provided therapy was the most effective in terms of overall reductions, then blended care, and lastly, AI-mediated therapy. Therapeutic alliance and engagement were found to be important yet moderate mediators of treatment response. There were differences in individuals like technology literacy which influenced results, especially in AI-based conditions. Blended care reported the least number of dropouts and stable participation, which indicated benefit in continuity of treatment.

6. Discussion
The current research examined the psychological effect of AI-mediated therapy on the clinical treatment outcome based on mixed-methods comparative design, including AI-only, human-delivered, and blended therapeutic models. In all conditions, there were statistically significant decreases in depressive and anxiety symptoms among participants, which proved that AI-mediated interventions may serve as effective modalities to support mental health. These results add to a growing body of evidence, which indicates that AI-driven interventions have the capacity to supplement more conventional psychotherapy and extend access to psychological care.

6.1The interpretation of clinical outcomes
As per previous empirical studies, AI-only treatment elicited both a clinically significant effect in the depressive and anxiety symptomatology. Even though therapy provided by humans resulted in significant gains, the hybrid state of offering AI-assisted participants extra therapist assistance proved to be the most improved in general and showed the highest engagement scores. This trend suggests that AI-mediated tools can achieve optimal effectiveness with the integration into a hybrid care model instead of being a single option to be offered to all clients. The increased performance in the blended condition is consistent with the adaptive care models in digital mental health and it highlights the synergistic benefits of human-delivered and automated therapeutic support.
The similarity between the effectiveness of the AI-only and human-only therapy is also commendable especially considering the fact that there are still concerns with the abilities of the AI agents to recreate human relational mechanisms. However, the fact that there are no considerable differences between the two groups should be interpreted rather carefully. The gains in the AI-only condition can be most conspicuous in patients with mild to moderate symptoms and the gains in those with more severe and complicated manifestations found more value in human or blended care. The point made by this observation highlights the need to clarify individual-level predictors of treatment responsiveness.

6.2 Change Operations: The Alliance and Engagement.
One of the main aims of the research was to identify the psychological processes through which AI-mediated interventions are correlated with clinical outcomes. As the mediation analyses showed, treatment condition and symptom change were partially explained by the therapeutic alliance and engagement. In spite of the fact that the scores of alliance were considerably lower in the AI-only condition, they were still significant factors in clinical improvement. This observation puts to test traditional psychotherapy theory that alliances have to be deeply interpersonal in order to work. Rather, alliance in AI-mediated situations seems to be contingent on how one perceives consistency, usability, and cognitive knowledge of therapeutic undertakings and not on emotional reciprocity.
One of the mechanisms that proved to be influential was engagement. Increased engagement in the participants was also associated with higher predicted symptom reductions and engagement was highest with the blended condition. The pattern was explained by qualitative results: participants appreciated the responsibility and security of a human therapist, and they enjoyed the freedom and incessant access to the AI system. Complementary human/AI support seems to promote long-term motivation and reduce the chances of dropping out, which is in line with the previous studies that stated that human-supplied digital strategies tend to produce a higher adherence compared to the purely automated strategies.

6.3 Moderators of treatment response Analysis
Other moderator analyses also pointed out that individual differences define the efficacy of AI-mediated therapy. Younger respondents and highly technology literate people answered more positively to AI-only therapy, which implies that knowledge of digital platforms makes them more comfortable and engaging. Conversely, older people and those who were less technologically advanced found it harder to navigate AI-based systems and perceived less meaningful interactions.
The outcome also moderated baseline symptom severity. The human and blended conditions were more improved on participants having severe depression or anxiety. This result has significant clinical implications: although AI-mediated therapy can represent significant benefits to people with mild or moderate symptoms, it might not be adequate as a modality on its own in patients who need more intensive clinical care. It is thus imperative to adapt the intervention delivery to the severity and risk profiles in order to achieve a safe and effective implementation.

6.4 Qualitative Findings Insight
Nuanced psychological experiences were identified in the qualitative analysis which places quantitative findings in context. The respondents supported the immediate presence of AI and the perceived lack of judgment, which helped make the first step and promote independent self-expression. However, the reactions of the AI were characterized by

several respondents as monotonous or less emotional, which is seen as an instance of perceived unauthenticity that partly explains the lower scores of alliance per AI only.

The blended one was always regarded as safer, more personalized and trusted. The respondents were thankful that the use of AI-generated engagement data by therapists allowed organizing sessions in a more efficient manner. Reliance on AI also appeared to be based on transparency: users wanted to clearly understand the capabilities and restrictions of the AI, how their data was processed, and who held the duty of data custody. These results highlight the importance of transparency and user education as a psychological factor in the layout of online treatment.

6.5 Implications to AI Design and Clinical Practice

These findings suggest a number of guidelines to the developers and clinicians. To improve perceived authenticity and strengthen therapeutic alliance, first, AI systems are to focus on the improvement of relational cues, including adaptive responses, empathetic phrasing, and contextual memory. Second, hybrid solutions, where AI is combined with human feedback could be the solution of the best help, especially with people with more severe symptoms or less digital competence. The inclusion of AI in a way that complements and does not replace therapists can help improve the results of the treatment and retain human control in risk-sensitive situations.

The application of AI-mediated therapy must be used clinically considering the differences between individuals. Technology literacy, therapeutic expectations, and symptom severity screening can support the decisions of adequacy of AI-only, human-only, or blended modalities. The trust and engagement might be increased with the help of educational interventions that explain the capabilities of AI.

6.6 Limitations

There are a number of restrictions which are worth considering. First, regardless of using intent-to-treat analyses, randomization was used and, as such, there was a possibility of bias due to attrition. Second, the AI system utilized in this paper might not be applicable to other AI platforms that vary in the architecture, conversational style, or the therapeutic model. Third, the follow-up time was not more than three months; there is still no information on long-term outcomes of AI-mediated therapy. Fourth, qualitative findings, rich as they were, were supplied on a modest subset of sample and might not reflect the entire range of user experiences. Lastly, even though alliance measures have been implemented in the digital setting, they might not be able to reflect fully the relational process peculiar to AI-based interactions.

6.7 Future Research

Further empirical studies are then required to assess the sustainability of AI-mediated therapeutic interventions especially in terms of relapse prevention and retention of client interaction. The different AI architectures, i.e., rule-based, generative, emotion-sensitive and multimodal systems, should be compared to understand how particular design attributes enable psychological processes at the background. In addition, sociocultural and linguistic variables must have a moderating effect on the building of trust and participation; therefore, cross-cultural studies are required. The ethical aspects, including the algorithmic fairness and the privacy of data, are to be re-evaluated, as AI systems develop.

7. Conclusion

This paper provides a critical evaluation of psychological implications that can be applied to AI-mediated interventions in the mental health context. The analysis of quantitative and qualitative data combined under AI-only, human-only, and integrated therapeutic conditions proves that AI systems have the potential to achieve significant changes in depressive and anxious symptomatology, particularly in patients with mild to moderate clinical manifestations. More importantly, the results emphasize the fact that the effectiveness of AI-mediated therapy depends on psychological processes, in particular, therapeutic alliance and client engagement, which are the pillars of any successful psychotherapy.

In spite of the fact that the measure of alliance was always lower when compared to the AI-only condition, the mediation effect of alliance on treatment outcome remained strong. The engagement proved to be a strong predictor of therapeutic improvement; the blended format received the most significant adherence and general clinical outcome. The responsivity was significantly moderated by the individual attributes, such as age, technological literacy, and baseline symptom severity, which implies the need to design digital interventions based on the individual patient characteristics.

The qualitative data show that users appreciate the ease and objectivity of AI systems, however, more often than not, the participants of the research also report a need to have more emotional depth, genuineness, and transparency. The blended model can be seen to provide a balance between efficiency and human care and warmth, and hybrid models can be viewed as the most clinically effective way of introducing AI into mental care.

Altogether, the research substantiates the hypothesis that AI-based treatment is an effective and promising part of modern psychological treatment. It is best used not to replace human therapists but to enhance and increase human care. Consistent with the development of AI technologies, the necessary conditions to guarantee safe, equal, and effective mental health services will be a clinically informed design, stringent ethical evaluation, and personalized execution.

8. References

1. Andersson G, Carlbring P, Titov N, Lindefors N. Internet-delivered psychological treatments: From innovation to implementation. World Psychiatry. 2019;18(1):20-8. https://doi.org/10.1002/wps.20610
2. Bendig E, Bauereiß N, Baumeister H. The next generation: Chatbots in clinical psychology and psychotherapy to foster mental health. Zeitschrift für Psychologie. 2019;227(4):269-79. https://doi.org/10.1027/2151-2604/a000386
3. Berryhill MB, Culmer N, Williams N, Halli-Tierney A, Betancourt A, Roberts H, *et al*. Videoconferencing psychotherapy and depression: A systematic review. Telemedicine and e-Health. 2019;25(6):435-46.
4. Bickmore T, Picard R. Establishing and maintaining long-term human–computer relationships. ACM Transactions on Computer-Human Interaction. 2005;12(2):293-327.
5. Braun V, Clarke V. Using thematic analysis in psychology. Qualitative Research in Psychology. 2006;3(2):77-101.
6. Carlbring P, Pergl-Wilson G, Andersson G. Artificial intelligence and psychotherapy: Opportunities, challenges, and ethical issues. Behaviour Research and Therapy. 2023;157:104198.

7. Cohen J. Statistical power analysis for the behavioral sciences. 2nd ed. Lawrence Erlbaum Associates; 1988.
8. Fitzpatrick KK, Darcy A, Vierhile M. Delivering cognitive behavioral therapy to young adults with symptoms of depression and anxiety using a fully automated conversational agent (Woebot). JMIR Mental Health. 2017;4(2):e19. https://doi.org/10.2196/mental.7785
9. Flückiger C, Del Re AC, Wampold BE, Horvath AO. The alliance in adult psychotherapy: A meta-analytic synthesis. Psychotherapy. 2018;55(4):316-40.
10. Fulmer R, Joerin A, Gentile B, Lakerink L, Rauws M. Using psychological artificial intelligence (Tess) to relieve symptoms of depression and anxiety: A randomized controlled trial. JMIR Mental Health. 2018;5(4):e64.
11. Gaffney H, Mansell W, Edwards R, Wright J. Computer-supported CBT self-help for depression in primary care: A qualitative study of patient experience. British Journal of Clinical Psychology. 2019;58(1):15-33.
12. Gosling SD, Götz FM, Rentfrow PJ. Artificial intelligence in psychology: Current applications and future directions. Psychological Science in the Public Interest. 2022;23(2):86-122.
13. Holmes EA, Ghaderi A, Harmer CJ, Ramchandani PG, Cuijpers P, Morrison AP, *et al*. The Lancet psychiatry commission on psychological treatments research. The Lancet Psychiatry. 2018;5(3):237-86.
14. Hollis C, Falconer CJ, Martin JL, Whittington C, Stockton S, Glazebrook C, *et al*. Digital health interventions for children and young people with mental health problems: A systematic and meta-review. Digital Health. 2017;3:205520761771146.
15. Kazdin AE. Addressing the treatment gap: A key challenge for extending evidence-based psychosocial interventions. Behaviour Research and Therapy. 2017;88:7-18.
16. Kretzschmar K, Tyroll H, Pavarini G, Manzini A, Singh I. Can your phone be your therapist? Young people's ethical perspectives on the use of fully automated conversational agents (Chatbots) in mental health support. Biomed Research International. 2019;2019:1-11.
17. Lattie EG, Adkins EC, Winquist N, Stiles-Shields C, Wafford QE, Graham AK. Digital mental health interventions for depression, anxiety, and enhancement of psychological well-being among college students: Systematic review. JMIR Mental Health. 2019;6(7):e12869.
18. Mohr DC, Weingardt KR, Reddy M, Schueller SM. Three problems with current digital mental health research and three things we can do about them. Psychiatric Services. 2017;68(5):427-29.
19. Prochaska JO, DiClemente CC. Stages and processes of self-change of smoking: Toward an integrative model. Journal of Consulting and Clinical Psychology. 1983;51(3):390-5.
20. Rabinowitz S, Chan PS, Thorndike FP. Human-AI teaming in psychotherapy: Opportunities and risks in blended mental health care. Journal of Clinical Psychology. 2024;80(2):245-60.
21. Richards D, Richardson T. Computer-based psychological treatments for depression: A systematic review and meta-analysis. Clinical Psychology Review. 2012;32(4):329-42.
22. Smith MJ, Wallace ML, O'Regan T, O'Neill S, Hayes L. Designing empathic conversational agents for mental health support: A framework for relational AI. Frontiers in Digital Health. 2022;4:823465.
23. Torous J, Jän Myrick K, Rauseo-Richmond R, Firth J. Digital mental health and COVID-19: Using technology today to accelerate the curve on access and quality tomorrow. JMIR Mental Health. 2021;8(3):e18848.
24. Wampold BE. How important are the common factors in psychotherapy? An update. World Psychiatry. 2015;14(3):270-7.
25. Weizenbaum J. Computer power and human reason: From judgment to calculation. W. H. Freeman; 1976.

How to Cite This Article
Bose S. Psychological Impact of AI-Mediated Therapy on Treatment Outcomes: A Mixed-Methods Evaluation. International Journal of Multi Research. 2025; 1(6): 23-33.