REVIEW OF RESEARCH
ISSN: 2249-894X
IMPACT FACTOR : 5.7631(UIF)
VOLUME - 16 | ISSUE - 1 | OCTOBER - 2026

HUMAN THERAPY AND AI-DRIVEN CHATBOT THERAPY: A COMPARATIVE REVIEW OF EFFECTIVENESS, THERAPEUTIC ALLIANCE, AND PATIENT OUTCOMES

Pravida Menon
Department of Psychology, Arunodaya University.

ABSTRACT:



Psychological treatment has traditionally rested on direct human interaction between a trained therapist and a client, and this model is supported by one of the largest evidence bases in health care. In recent years, artificial intelligence (AI) driven therapy, delivered through conversational agents on smartphones and computers, has emerged as an alternative and as a possible answer to the worldwide gap between the need for mental health support and its supply. This paper examines the shifting paradigm of psychological treatment by comparing human interaction therapy and AI-driven therapy in terms of effectiveness, the therapeutic alliance, accessibility, cost, safety, and patient outcomes. The paper is based on a narrative review of meta-analyses of psychotherapy outcome research, randomized controlled trials of conversational agents such as Woebot, Tess, and Wysa, systematic reviews of mental health chatbots, and recent analyses of large language models in behavioral health care. The review shows that human therapy produces larger and more durable symptom change and rests on a well-studied working relationship, while AI-driven therapy offers continuous availability, low cost, and scalability, with promising but early evidence of effectiveness for mild to moderate symptoms over short periods. The paper proposes a comparative cohort study of adults receiving human therapy, AI-based support, or a combination of both, and presents suggested table templates for its analysis. It concludes that AI-driven therapy is at present best understood as a complement to human care rather than a replacement for it.

KEYWORDS: *Human Interaction, AI-Driven Therapy, Conversational Agents, Chatbots, Therapeutic Alliance, Depression, Anxiety, Patient Outcomes, Digital Mental Health.*

INTRODUCTION

Mental health problems are among the most common health conditions in the world, and depression and anxiety together account for a large share of the global burden of disease. For more than a century, the main form of psychological treatment has been the conversation between a trained therapist and a client. This model of care is supported by an enormous body of research. Meta-analyses covering hundreds of randomized trials show that psychological therapies such as cognitive behavior therapy (CBT) produce meaningful and lasting improvement for the majority of people who receive them (Cuijpers et al., 2023; Wampold & Imel, 2015).

Yet the reach of human therapy remains limited. Trained therapists are scarce, treatment is expensive, waiting lists are long, and stigma prevents many people from seeking help at all. The result is a persistent gap between the number of people who could benefit from psychological support and the number who actually receive it. It is in this context that AI-driven therapy has appeared. Conversational

agents, often called chatbots, can deliver elements of structured therapy through text-based conversation at any hour, at very low cost, and to very large numbers of users at the same time. Products such as Woebot, Wysa, and Tess have attracted large user bases, and the rapid spread of large language models has intensified debate about what artificial intelligence can and cannot do in mental health care (Stade et al., 2024; Torous et al., 2021).

These developments amount to a shifting paradigm in psychological treatment. The question is no longer whether technology will play a role in mental health care, but what that role should be, and how AI-driven therapy compares with the human interaction it partly imitates. The question matters to several groups at once. Patients are already choosing between booking a therapist, downloading an app, or using both. Clinicians are being asked whether these tools are safe to recommend. Health systems are weighing low-cost automated support against scarce professional time. Decisions with real consequences for all of these groups are being made now, often on the basis of marketing claims rather than evidence. The present paper examines the comparison with respect to effectiveness, the therapeutic alliance, practical outcomes, and risks, and proposes a study design for comparing human, AI-driven, and combined support directly.

AIMS AND OBJECTIVES

The main aim of the study is to examine the shifting paradigm of psychological treatment by comparing human interaction therapy and AI-driven therapy in terms of effectiveness, therapeutic relationship, and patient outcomes. The specific objectives are:

- To examine the present status of human-delivered psychotherapy and its established evidence base.
- To analyse the development and growth of AI-driven conversational agents in mental health care.
- To compare the effectiveness of human therapy and AI-driven therapy in reducing symptoms of depression and anxiety.
- To study the role of the therapeutic alliance in human interaction and its possible forms in AI-driven settings.
- To compare accessibility, cost, and scalability of human and AI-driven psychological support.
- To identify the safety, privacy, and ethical concerns associated with AI-driven therapy.
- To examine patient outcomes and satisfaction reported in existing comparative and single-arm studies.
- To propose a research design, with suggested tables, for a comparative study of adults receiving human therapy, AI-based support, or a combination of both.
- To suggest suitable measures for the responsible integration of AI-driven tools into psychological care.

REVIEW OF LITERATURE

The effectiveness of human-delivered psychotherapy is supported by one of the largest evidence bases in all of health care. Cuijpers and colleagues (2023) examined 409 randomized trials with 52,702 patients and found that CBT is effective for depression compared with control conditions, with effects holding across delivery formats. Broader syntheses conclude that the major established therapies work roughly equally well for common conditions and that gains are often maintained after treatment ends (Wampold & Imel, 2015). Research on why therapy works has emphasized factors shared across approaches, above all the therapeutic alliance, defined as the emotional bond between client and therapist together with their agreement on the goals and tasks of treatment (Horvath & Greenberg, 1989). Meta-analyses consistently find a moderate relationship between the alliance and outcome: Martin, Garske, and Davis (2000) estimated r = .278 across 79 studies, and Fluckiger and colleagues (2018) confirmed the same estimate, r = .278, across 295 studies with more than 30,000 patients.

The idea of machine conversation about personal problems is older than the personal computer. Weizenbaum (1966) created ELIZA, a simple program that imitated a Rogerian therapist by reflecting the user's statements back as questions, and observed with some alarm how readily people attributed

understanding and empathy to it. Modern conversational agents are far more capable. Most established mental health chatbots, including Woebot, Wysa, and Tess, deliver content derived from evidence-based therapies through structured decision trees and rule-based natural language processing (Vaidyam et al., 2019). A newer generation, based on large language models, can generate open-ended responses and is only beginning to be studied clinically (Stade et al., 2024).

The clinical evidence for AI-driven therapy rests on a small number of trials. Fitzpatrick, Darcy, and Vierhile (2017) randomized 70 young adults with symptoms of depression and anxiety either to Woebot, a text-based conversational agent delivering CBT content, or to an information-only control; after two weeks the Woebot group showed a significant reduction in depression symptoms, while the control group did not. Fulmer and colleagues (2018) randomized 74 university students to the psychological AI agent Tess for two or four weeks or to an information-only control, and found significant reductions in depression and anxiety in the Tess groups. Inkster, Sarda, and Subramanian (2018) analysed real-world data from users of Wysa and found that more frequent users showed greater improvement in self-reported depressive symptoms and described the agent in terms of empathy and support despite knowing it was a machine.

Beyond symptom scores, studies have examined engagement and acceptability. Participants in the Woebot trial interacted with the agent on average about twelve times over two weeks, and their open-ended comments suggested that the conversational format, the sense of being accompanied, and the informal tone contributed as much to acceptability as the therapeutic content itself (Fitzpatrick et al., 2017). Users of Wysa in real-world settings described the agent as empathetic and helpful, and usage patterns showed that a substantial share of conversations took place outside the hours when human services are normally available (Inkster et al., 2018). Engagement of this kind is meaningful because dropout is one of the persistent problems of human therapy: a support that people actually use, even a modest one, can outperform a superior treatment they never attend. At the same time, engagement with an app is easy to start and easy to abandon, and most mental health apps lose the majority of their users within weeks (Torous et al., 2021).

Systematic reviews advise cautious optimism. Abd-alrazaq and colleagues (2020) reviewed twelve studies and found weak evidence that chatbots can reduce depression, distress, and stress, conflicting evidence for anxiety, and very little evidence on safety, since only two studies assessed it. Vaidyam and colleagues (2019) likewise concluded that conversational agents show promise for mild to moderate symptoms but that the evidence base is small and short-term. On the alliance, Henson and colleagues (2019) found that users of digital mental health apps report bond-like experiences of being understood and accepted, although the digital alliance appears thinner than the human one. Safety concerns were documented early: Miner and colleagues (2016) showed that widely used smartphone agents responded inconsistently and often inappropriately to statements about depression, suicide, and violence. The World Health Organization (2021) has issued guidance on the ethics and governance of AI for health, covering autonomy, safety, transparency, accountability, inclusiveness, and responsiveness, and commentators such as D'Alfonso (2020) and Torous and colleagues (2021) have mapped the wider field of AI in mental health.

RESEARCH METHODOLOGY

The present paper, entitled "Human Therapy and AI-Driven Chatbot Therapy: A Comparative Review of Effectiveness, Therapeutic Alliance, and Patient Outcomes," is based on a descriptive and analytical research approach. It reviews and compares published evidence on human-delivered psychotherapy and AI-driven conversational therapy. Secondary sources include peer-reviewed journals, meta-analyses, randomized controlled trials, systematic reviews, books, and reports of the World Health Organization, published between 1966 and 2024.

Building on this review, the paper also proposes a primary study. The proposed study would follow a cohort of adults who are already receiving psychological support in one of three forms: human therapy only, AI-based support only, or a combination of both. No participants would be assigned to any condition; the study would observe people in the care they have already chosen. A proposed sample of approximately 100 adults would complete short, established questionnaires at the start of the study

and again after several weeks. Symptom change would be measured with the Patient Health Questionnaire (PHQ-9) for depression and the Generalized Anxiety Disorder scale (GAD-7) for anxiety (Kroenke et al., 2001; Spitzer et al., 2006). The quality of the working relationship would be measured with the short revised Working Alliance Inventory (Hatcher & Gillaspy, 2006; Horvath & Greenberg, 1989), adapted appropriately for AI-based support, and satisfaction would be measured with the Client Satisfaction Questionnaire (Attkisson & Zwick, 1982). Collected data would be analysed using percentages, averages, tables, and comparisons of change across the three groups. Because data collection has not yet been carried out, the tables below are presented as proposed templates, to be completed with the data of the study.

SUGGESTED TABLES (4–5)

The following five tables are designed for a proposed sample of 100 respondents. They examine the demographic profile of respondents, the type of psychological support received, symptom change over time, the working alliance, and satisfaction. Each table is a proposed template: the figures are left blank and are to be completed with the primary data of the study.

Table 1: Demographic Profile of Respondents (Proposed)

Demographic Category	Respondents	Percentage
Age 18-25 years	-	-
Age 26-35 years	-	-
Age 36-45 years	-	-
Age 46 years and above	-	-
Male	-	-
Female	-	-
Other / prefer not to say	-	-
Total	100	100%

Explanation: The table will present the age and gender distribution of the proposed sample of 100 adults receiving psychological support. It is intended to show whether the three support groups are broadly comparable in background characteristics, which is important for a fair comparison of outcomes.

Source: Proposed table template prepared by the researcher; to be completed with primary data collected for the study.

Table 2: Type of Psychological Support Received (Proposed)

Type of Support	Respondents	Percentage
Human therapy only	-	-
AI-based support only	-	-
Combination of human and AI-based support	-	-
Total	100	100%

Explanation: The table will show the distribution of respondents across the three forms of support studied: human therapy only, AI-based support only, and a combination of both. This distribution defines the three groups whose outcomes the study will compare.

Source: Proposed table template prepared by the researcher; to be completed with primary data collected for the study.

Table 3: Change in Depression Symptoms (PHQ-9) by Type of Support (Proposed)

Type of Support	Mean PHQ-9 at Start	Mean PHQ-9 at Follow-up	Mean Change
Human therapy only	-	-	-
AI-based support only	-	-	-
Combination of human and AI-based support	-	-	-

Explanation: The table will present mean depression scores on the PHQ-9 at the start of the study and at follow-up for each support group, together with the mean change. It is designed to answer the study's central question: how symptom improvement differs between human therapy, AI-based support, and their combination.

Source: Proposed table template prepared by the researcher; PHQ-9 is a validated nine-item depression measure (Kroenke et al., 2001).

Table 4: Change in Anxiety Symptoms (GAD-7) by Type of Support (Proposed)

Type of Support	Mean GAD-7 at Start	Mean GAD-7 at Follow-up	Mean Change
Human therapy only	-	-	-
AI-based support only	-	-	-
Combination of human and AI-based support	-	-	-

Explanation: The table will present mean anxiety scores on the GAD-7 at the start of the study and at follow-up for each support group, together with the mean change, allowing anxiety outcomes to be compared across the three forms of support.

Source: Proposed table template prepared by the researcher; GAD-7 is a validated seven-item anxiety measure (Spitzer et al., 2006).

Table 5: Working Alliance and Satisfaction by Type of Support (Proposed)

Type of Support	Mean Alliance Score (WAI-SR)	Mean Satisfaction Score (CSQ-8)
Human therapy only	-	-
AI-based support only	-	-
Combination of human and AI-based support	-	-

Explanation: The table will compare the strength of the working alliance, measured by the short revised Working Alliance Inventory, and client satisfaction, measured by the CSQ-8, across the three support groups. It is intended to test whether the human advantage in the therapeutic relationship, well documented in traditional therapy, also appears when AI-based support is involved.

Source: Proposed table template prepared by the researcher; WAI-SR (Hatcher & Gillaspy, 2006) and CSQ-8 (Attkisson & Zwick, 1982) are established measures.

OVERALL INTERPRETATION

The five proposed tables provide the structure for a direct comparison of human interaction therapy, AI-driven therapy, and combined support. Together they are designed to answer four questions: who receives each form of support, how much symptoms of depression and anxiety change over time in each group, how the working alliance differs when the helper is a person, a program, or both, and how satisfied clients are with each form of care. Because the study has not yet collected data, the tables are templates and contain no figures. Their value at this stage is to make the planned analysis explicit before data collection begins, so that the questionnaires, the sample, and the comparisons are fixed in advance rather than shaped by the results.

RESULTS

As the proposed primary study is still in the design stage, this section reports the main results of the published studies reviewed above, which together constitute the present state of knowledge on the comparison between human and AI-driven therapy.

First, human-delivered psychotherapy is effective. The meta-analysis of Cuijpers and colleagues (2023), covering 409 trials and 52,702 patients, found that CBT reduces depression compared with control conditions, and the broader literature supports the effectiveness of the major therapies across common mental health conditions (Wampold & Imel, 2015). The therapeutic alliance is a consistent predictor of these outcomes, with a meta-analytic correlation of about r = .278 between alliance and outcome (Fluckiger et al., 2018; Martin et al., 2000).

Second, AI-driven conversational agents show early evidence of benefit. In the Woebot randomized controlled trial, participants using the agent showed a significant reduction in depression symptoms over two weeks compared with an information-only control group (Fitzpatrick et al., 2017). In the Tess trial, participants using the agent for two or four weeks showed significant reductions in depression and anxiety compared with controls (Fulmer et al., 2018). Real-world analysis of Wysa found greater improvement among more frequent users (Inkster et al., 2018). At the same time, the meta-analysis of Abd-alrazaq and colleagues (2020) rated the overall evidence as weak, conflicting for anxiety, and nearly silent on safety.

Third, on process and experience, users of digital tools do report alliance-like experiences, but of a thinner kind than the human alliance (Henson et al., 2019). Fourth, on safety, general-purpose smartphone agents have responded inconsistently and sometimes inappropriately to crisis statements (Miner et al., 2016), and purpose-built agents handle crises only through scripted referrals. No published study has yet directly compared human therapy, AI-driven therapy, and combined support in a single design, which is the gap the proposed study is intended to address.

METHODS

The paper adopts a narrative review method. Published studies were identified from major peer-reviewed sources in psychology, psychiatry, and digital health, including the Journal of Medical Internet Research and its sister journals, World Psychiatry, JAMA Internal Medicine, The Canadian Journal of Psychiatry, BJPsych Open, Psychotherapy Research, and the Journal of Consulting and Clinical Psychology, together with institutional reports. Priority was given to randomized controlled trials, meta-analyses, and systematic reviews. Studies were selected for relevance to three questions: the effectiveness of human-delivered psychotherapy, the effectiveness and safety of AI-driven conversational agents in mental health, and the therapeutic alliance in human and digital settings. Findings were organized thematically and compared across the two forms of care. The proposed primary study described in the Research Methodology section follows a cohort design, observing adults already receiving human, AI-based, or combined support, without assigning participants to conditions, and measuring change over time with standardized instruments.

STATEMENT OF THE PROBLEM

Psychological therapy delivered through human interaction is effective but cannot reach everyone who needs it. The shortage of trained therapists, the cost and time demands of treatment, geographical barriers, and stigma together leave a large proportion of people with common mental health problems without any support. AI-driven therapy has emerged in response to this gap, and millions of people already use conversational agents for emotional support. However, the evidence for these tools is young, short-term, and limited, while their risks, including privacy, safety in crisis situations, bias, and the absence of professional accountability, are not yet well governed. Patients, clinicians, and health systems therefore face a practical problem: decisions about adopting AI-driven therapy are being made faster than the evidence on which they should rest. There is a need for research that compares human interaction therapy, AI-driven therapy, and their combination directly, measuring symptom change, the working relationship, and patient satisfaction in the same study. The present

paper responds to this problem by reviewing the comparative evidence and proposing a design for such a study.

SCOPE AND LIMITATIONS

Scope of the Study

The review covers peer-reviewed research and major institutional reports on human-delivered psychotherapy and AI-driven conversational therapy published between 1966 and 2024, with emphasis on depression and anxiety, the most common reasons for seeking psychological support. It considers effectiveness, the therapeutic alliance, accessibility, cost, safety, and patient outcomes. The proposed primary study is designed for adults receiving psychological support in community settings, and its findings are intended to be useful for students, researchers, clinicians, and providers considering the integration of AI-based tools into psychological care.

Limitations

The study has certain limitations. First, the AI evidence base is small and recent: few randomized trials exist, follow-up periods are short, and samples are often young, educated, and self-selected. Second, most trials compare AI agents with waiting lists or information-only controls rather than with active human treatment, which limits direct comparison. Third, much of the available research is conducted or funded by the developers of the products studied. Fourth, the technology is changing faster than studies can be completed, and findings about one generation of tools may not apply to the next. Fifth, the proposed study is observational: because participants are not randomly assigned to human, AI-based, or combined support, differences between groups may partly reflect differences in the people who choose each form of care. Finally, this paper is a narrative review rather than a systematic one, and aims to map the evidence rather than to pool effect sizes.

DISCUSSION

The comparison developed in this paper suggests that human interaction therapy and AI-driven therapy are not two versions of the same thing competing for the same role. Human therapy is a proven treatment built on a genuine relationship. Its central mechanism, the alliance between client and therapist, has been studied for decades and predicts outcomes reliably across hundreds of studies (Fluckiger et al., 2018; Wampold & Imel, 2015). Human therapists read tone, silence, and what is not said; they exercise judgment in ambiguous and high-risk situations; and they carry professional and legal responsibility for their clients. These capacities matter most precisely where need is greatest.

AI-driven therapy offers a different set of strengths. It is available at any hour, costs very little per user, has no waiting list, and can reach people who would never consult a therapist. The trial evidence, though early, shows real short-term improvements in depression and anxiety for users of structured agents (Fitzpatrick et al., 2017; Fulmer et al., 2018; Inkster et al., 2018). Many users engage with these tools more readily than they would with a person, and some disclose more freely to a listener that cannot judge them (Henson et al., 2019). At the same time, the relationship with a machine remains a simulation of understanding rather than understanding itself, and no current system can assess risk or take responsibility for safety (Miner et al., 2016; Stade et al., 2024).

A further point concerns what should count as a patient outcome. Symptom reduction is the most common measure, and the most important, but it is not the only one. Satisfaction with care, the sense of being heard, the acquisition of coping skills, functional improvement in daily life, and continued engagement with support all matter to patients. These dimensions can be measured with brief validated instruments, such as the PHQ-9 and GAD-7 for symptoms, the Working Alliance Inventory for the relationship, and the CSQ-8 for satisfaction (Attkisson & Zwick, 1982; Hatcher & Gillaspy, 2006; Kroenke et al., 2001; Spitzer et al., 2006). A serious comparison of human and AI-driven therapy must examine all of these dimensions rather than symptom scores alone, because the two forms of care may succeed on different ones.

The shifting paradigm is therefore best described not as a replacement but as a redistribution of roles. The evidence points toward a division of labor in which AI-driven tools provide first-line

structured support for mild to moderate difficulties, reach underserved populations, and support people between therapy sessions, while human therapists remain essential for severe or complex conditions, crisis situations, and the relational work at the heart of psychological change. The arrival of large language models makes this framing more urgent: as AI conversation becomes more natural and more widely used, the need for careful comparative evidence and clear governance grows accordingly (Stade et al., 2024; World Health Organization, 2021). The proposed study, comparing human, AI-based, and combined support in a single cohort, is one step toward the evidence such decisions require.

RECOMMENDATIONS

On the basis of the review, several recommendations can be made. Mental health services should treat AI-driven tools as a complement to human care, using them for first-line support, psychoeducation, mood tracking, and between-session practice, with clear pathways to human professionals when risk appears. Developers should build crisis detection and referral into every mental health product, and should submit their products to independent clinical evaluation before making effectiveness claims. Researchers should conduct direct comparative studies of human, AI-based, and combined support using standardized outcome measures and longer follow-up periods, and should report safety outcomes routinely. Regulators and professional bodies should develop standards for privacy, data protection, bias testing, and accountability in AI-driven mental health tools, in line with the principles set out by the World Health Organization (2021). Clinicians should be trained to understand what these tools can and cannot do, so they can guide patients who are already using them. Finally, users should be informed clearly, at the point of use, that an AI agent is not a therapist and is not a substitute for professional help in a crisis.

CONCLUSION

Psychological treatment is undergoing a genuine shift. Human interaction therapy remains the standard against which alternatives must be measured: it is effective, durable in its effects, and grounded in a working relationship that research has validated across decades. AI-driven therapy has moved from a laboratory curiosity to a mass-used source of support in less than a decade, and early evidence shows that it can reduce symptoms of depression and anxiety in the short term for people with mild to moderate difficulties. Its decisive advantages are availability, cost, and scale; its decisive limitations are the absence of genuine understanding, unresolved safety and privacy concerns, and a thin evidence base. The fair conclusion from present evidence is that AI-driven therapy can extend the reach of psychological support to people who would otherwise go without, that it cannot yet replace the judgment or the relationship of a trained therapist, and that the most valuable question for the next decade is not which of the two prevails, but how each can best be used, separately and together, to improve patient outcomes. The proposed comparative study is designed as one contribution to answering that question.

REFERENCES

1. Abd-alrazaq, A. A., Rababeh, A., Alajlani, M., Bewick, B. M., & Househ, M. (2020). Effectiveness and safety of using chatbots to improve mental health: Systematic review and meta-analysis. Journal of Medical Internet Research, 22(7), e16021. https://doi.org/10.2196/16021
2. Attkisson, C. C., & Zwick, R. (1982). The client satisfaction questionnaire: Psychometric properties and correlations with service utilization and psychotherapy outcome. Evaluation and Program Planning, 5(3), 233-237. https://doi.org/10.1016/0149-7189(82)90074-X
3. Cuijpers, P., Miguel, C., Harrer, M., Plessen, C. Y., Ciharova, M., Ebert, D., & Karyotaki, E. (2023). Cognitive behavior therapy vs. control conditions, other psychotherapies, pharmacotherapies and combined treatment for depression: A comprehensive meta-analysis including 409 trials with 52,702 patients. World Psychiatry, 22(1), 105-115. https://doi.org/10.1002/wps.21069
4. D'Alfonso, S. (2020). AI in mental health. Current Opinion in Psychology, 36, 112-117. https://doi.org/10.1016/j.copsyc.2020.04.005

5. Fitzpatrick, K. K., Darcy, A., & Vierhile, M. (2017). Delivering cognitive behavior therapy to young adults with symptoms of depression and anxiety using a fully automated conversational agent (Woebot): A randomized controlled trial. JMIR Mental Health, 4(2), e19. https://doi.org/10.2196/mental.7785
6. Fluckiger, C., Del Re, A. C., Wampold, B. E., & Horvath, A. O. (2018). The alliance in adult psychotherapy: A meta-analytic synthesis. Psychotherapy, 55(4), 316-340. https://doi.org/10.1037/pst0000172
7. Fulmer, R., Joerin, A., Gentile, B., Lakerink, L., & Rauws, M. (2018). Using psychological artificial intelligence (Tess) to relieve symptoms of depression and anxiety: Randomized controlled trial. JMIR Mental Health, 5(4), e64. https://doi.org/10.2196/mental.9782
8. Hatcher, R. L., & Gillaspy, J. A. (2006). Development and validation of a revised short version of the Working Alliance Inventory. Psychotherapy Research, 16(1), 12-25. https://doi.org/10.1080/10503300500352500
9. Henson, P., Wisniewski, H., Hollis, C., Keshavan, M., & Torous, J. (2019). Digital mental health apps and the therapeutic alliance: Initial review. BJPsych Open, 5(1), e15. https://doi.org/10.1192/bjo.2018.94
10. Horvath, A. O., & Greenberg, L. S. (1989). Development and validation of the Working Alliance Inventory. Journal of Counseling Psychology, 36(2), 223-233. https://doi.org/10.1037/0022-0167.36.2.223
11. Inkster, B., Sarda, S., & Subramanian, V. (2018). An empathy-driven, conversational artificial intelligence agent (Wysa) for digital mental well-being: Real-world data evaluation mixed-methods study. JMIR mHealth and uHealth, 6(11), e12106. https://doi.org/10.2196/12106
12. Kroenke, K., Spitzer, R. L., & Williams, J. B. (2001). The PHQ-9: Validity of a brief depression severity measure. Journal of General Internal Medicine, 16(9), 606-613. https://doi.org/10.1046/j.1525-1497.2001.016009606.x
13. Martin, D. J., Garske, J. P., & Davis, M. K. (2000). Relation of the therapeutic alliance with outcome and other variables: A meta-analytic review. Journal of Consulting and Clinical Psychology, 68(3), 438-450. https://doi.org/10.1037/0022-006X.68.3.438
14. Miner, A. S., Milstein, A., Schueller, S., Hegde, R., Mangurian, C., & Linos, E. (2016). Smartphone-based conversational agents and responses to questions about mental health, interpersonal violence, and physical health. JAMA Internal Medicine, 176(5), 619-625. https://doi.org/10.1001/jamainternmed.2016.0400
15. Spitzer, R. L., Kroenke, K., Williams, J. B., & Lowe, B. (2006). A brief measure for assessing generalized anxiety disorder: The GAD-7. Archives of Internal Medicine, 166(10), 1092-1097. https://doi.org/10.1001/archinte.166.10.1092
16. Stade, E. C., Stirman, S. W., Ungar, L. H., Boland, C. L., Schwartz, H. A., Yaden, D. B., Sedoc, J., DeRubeis, R. J., Willer, R., & Eichstaedt, J. C. (2024). Large language models could change the future of behavioral healthcare: A proposal for responsible development and evaluation. npj Mental Health Research, 3, 12. https://doi.org/10.1038/s44184-024-00056-z
17. Torous, J., Bucci, S., Bell, I. H., Kessing, L. V., Faurholt-Jepsen, M., Whelan, P., Carvalho, A. F., Keshavan, M., Linardon, J., & Firth, J. (2021). The growing field of digital psychiatry: Current evidence and the future of apps, social media, chatbots, and virtual reality. World Psychiatry, 20(3), 318-335. https://doi.org/10.1002/wps.20883
18. Vaidyam, A. N., Wisniewski, H., Halamka, J. D., Kashavan, M. S., & Torous, J. B. (2019). Chatbots and conversational agents in mental health: A review of the psychiatric landscape. The Canadian Journal of Psychiatry, 64(7), 456-464. https://doi.org/10.1177/0706743719828977
19. Wampold, B. E., & Imel, Z. E. (2015). The great psychotherapy debate: The evidence for what makes psychotherapy work (2nd ed.). Routledge.
20. Weizenbaum, J. (1966). ELIZA: A computer program for the study of natural language communication between man and machine. Communications of the ACM, 9(1), 36-45. https://doi.org/10.1145/365153.365168
21. World Health Organization. (2021). Ethics and governance of artificial intelligence for health. World Health Organization.