DOI(Journal): 10.31703/gssr
DOI(Volume): 10.31703/gssr.2025(X)
DOI(Issue): 10.31703/gssr.2025(X.III)

p-ISSN: 2520-0348

e-ISSN: 2616-793X

www.gssrjournal.com
Global Social Sciences Review
exploring humanity

GSSR

GLOBAL SOCIAL SCIENCES REVIEW

HEC-RECOGNIZED CATEGORY-Y

Volum X, ISSUE III SUMMER (SEPTEMBER-2025)

HumanityPublications
sharing research
www.humanitypublications.com
Karachi | Islamabad | Lahore

Double-blind Peer-review Journal
www.gssrjournal.com

Humanity Publications(HumaPub)
www.humapub.com
Doi: https://dx.doi.org/10.31703

Article Title

The Therapeutic Alliance in AI-Assisted Psychotherapy: A Clinical Evaluation of Human-AI Interaction in Mental Health Care

Abstract

AI is now a new trend in mental health care, with chatbots, digital therapists, and clinical decision-support systems. However, the impact it has on therapeutic alliance, or the process of a mutual and emotional connection between a client and therapist, is not well comprehended. This paper examines the effect of AI-assisted psychotherapy on the formation of alliance and clinical outcomes. Basing its findings on the attachment theory and relational psychoanalysis, as well as the human-computer interaction systems, the study examines how AI facilitates or disrupts the therapy. The scholar concludes that AI is most effectively used as an adjunctive device in a hybrid paradigm of care - human empathy, ethical awareness, and emotional responsiveness are the pillars of successful psychotherapy that cannot be substituted with AI.

Keywords: Human-computer Interaction, Psychotherapy , Hybrid Paradigm

Authors:

Afshan Tabasum: (Corresponding Author)
Independent Scholar, LADO Institute, Vienna, VA 22182, United States.
(Email: afshantabasumhere@gmail.com)

Pages: 321-335
DOI:10.31703/gssr.2025(X-III).28
DOI link: https://dx.doi.org/10.31703/gssr.2025(X-III).28
Article link: https://gssrjournal.com/article/the-therapeutic-alliance-in-aiassisted-psychotherapy-a-clinical-evaluation-of-humanai-interaction-in-mental-health-care
Full-text Link: https://gssrjournal.com/article/the-therapeutic-alliance-in-aiassisted-psychotherapy-a-clinical-evaluation-of-humanai-interaction-in-mental-health-care
Pdf link: https://www.gssrjournal.com/jadmin/Auther/31rvIolA2.pdf

Global Social Sciences Review

p-ISSN: 2520-0348 **e-ISSN:** 2616-793x
DOI(journal):10.31703/gssr
Volume: X (2025)
DOI (volume):10.31703/gssr.2025(X)
Issue: III Summer (September-2025)
DOI(Issue):10.31703/gssr.2025(X-III)

Home Page
www.gssrjournal.com

Volume: X (2025)
https://www.gssrjournal.com/Current-issue

Issue: III-Summer (September 2025)
https://www.gssrjournal.com/issue/10/3/2025

Scope
https://www.gssrjournal.com/about-us/scope

Submission
https://humaglobe.com/index.php/gssr/submissions

Humanity Publications (HumaPub)
www.humapub.com
Doi: https://dx.doi.org/10.31703

Citing this Article

28	The Therapeutic Alliance in AI-Assisted Psychotherapy: A Clinical Evaluation of Human-AI Interaction in Mental Health Care		
Authors	Afshan Tabasum	DOI	10.31703/gssr.2025(X-III).28
		Pages	321-335
		Year	2025
		Volume	X
		Issue	III

Referencing & Citing Styles	
APA	Tabasum, A. (2025). The Therapeutic Alliance in AI-Assisted Psychotherapy: A Clinical Evaluation of Human-AI Interaction in Mental Health Care. *Global Social Sciences Review*, *X*(III), 321-335. https://doi.org/10.31703/gssr.2025(X-III).28
CHICAGO	Tabasum, Afshan. 2025. "The Therapeutic Alliance in AI-Assisted Psychotherapy: A Clinical Evaluation of Human-AI Interaction in Mental Health Care." *Global Social Sciences Review* X (III):321-335. doi: 10.31703/gssr.2025(X-III).28.
HARVARD	TABASUM, A. 2025. The Therapeutic Alliance in AI-Assisted Psychotherapy: A Clinical Evaluation of Human-AI Interaction in Mental Health Care. *Global Social Sciences Review,* X, 321-335.
MHRA	Tabasum, Afshan. 2025. 'The Therapeutic Alliance in AI-Assisted Psychotherapy: A Clinical Evaluation of Human-AI Interaction in Mental Health Care', *Global Social Sciences Review*, X: 321-35.
MLA	Tabasum, Afshan. "The Therapeutic Alliance in Ai-Assisted Psychotherapy: A Clinical Evaluation of Human-Ai Interaction in Mental Health Care." *Global Social Sciences Review* X.III (2025): 321-35. Print.
OXFORD	Tabasum, Afshan (2025), 'The Therapeutic Alliance in AI-Assisted Psychotherapy: A Clinical Evaluation of Human-AI Interaction in Mental Health Care', *Global Social Sciences Review,* X (III), 321-35.
TURABIAN	Tabasum, Afshan. "The Therapeutic Alliance in Ai-Assisted Psychotherapy: A Clinical Evaluation of Human-Ai Interaction in Mental Health Care." *Global Social Sciences Review* X, no. III (2025): 321-35. https://dx.doi.org/10.31703/gssr.2025(X-III).28.

e-ISSN:2616-793x Volume: X (2025) Issue: III-Summer (September 2025) p-ISSN:2520-0348

Global Social Sciences Review
www.gssrjournal.com
DOI:http://dx.doi.org/10.31703/gssr

Pages: 321-335 URL: https://doi.org/10.31703/gssr.2025(X-III).28 Doi: 10.31703/gssr.2025(X-III).28

Title

The Therapeutic Alliance in AI-Assisted Psychotherapy: A Clinical Evaluation of Human-AI Interaction in Mental Health Care

Authors:

Afshan Tabasum :(Corresponding Author)
Independent Scholar, LADO Institute, Vienna, VA 22182, United States.
(Email: afshantabasumhere@gmail.com)

Contents

Abstract

AI is now a new trend in mental health care, with chatbots, digital therapists, and clinical decision-support systems. However, the impact it has on therapeutic alliance, or the process of a mutual and emotional connection between a client and therapist, is not well comprehended. This paper examines the effect of AI-assisted psychotherapy on the formation of alliance and clinical outcomes. Basing its findings on the attachment theory and relational psychoanalysis, as well as the human-computer interaction systems, the study examines how AI facilitates or disrupts the therapy. The scholar concludes that AI is most effectively used as an adjunctive device in a hybrid paradigm of care - human empathy, ethical awareness, and emotional responsiveness are the pillars of successful psychotherapy that cannot be substituted with AI.

Keywords:

Human-computer Interaction, Psychotherapy , Hybrid Paradigm

Introduction

The therapist-client relationship, which is commonly described as the collaborative/emotional bond between therapist and client (Bordin, 1979), is to this day among the strongest predictors of psychotherapy outcomes regardless of modality or diagnosis (Horvath et al., 2011; Flückiger et al., 2018). It involves three key elements, which include consensus on therapeutic objectives, cooperative work on therapeutic activities, and building an affective relationship. These process relationships lead to a feeling of safety, trust, and mutual interaction that allows therapeutic change (Norcross & Wampold, 2019). The superiority of the alliance when compared to the particular therapeutic approach has been proven over the decades to be successful in predicting the success of treatment (Horvath et al., 2011).

Nevertheless, recent years have seen the digital transformation of mental health care, which has been catalyzed by the progress in artificial intelligence (AI), start to redefine the establishment of conventional notions of this relational core. The advent of AI-assisted psychotherapy and digital mental health interventions has cast some of the most basic questions regarding what counts as a working alliance in the scenario where one party to the therapeutic dyad is not a human. With the increased conversational and emotional simulation features of AI systems, there will be an increased blur of human therapists, tools, and autonomous agents (Ho et al., 2022; McDonald et al., 2023).

AI-assisted psychotherapy is a wide area of technology. Woebot, Wysa, and Replika are the most widely known ones, as they use natural language processing and machine learning to provide evidence-based methods of cognitive-behavioral therapy (CBT) and dialectical behavior therapy (DBT) (Fitzpatrick et al., 2017; Inkster et al., 2018). These systems are characterized by communication with users via text messages or voice and provide emotional support, guided reflections, and mood tracking. As an example, Woebot offers the ability to make daily conversational check-ins, recalculating the maladaptive thoughts with the help of the principles of CBT, whereas Wysa suggests mindfulness or self-care exercises, based on sentiment analysis (Inkster et al., 2018). Such systems have an attraction that is based on their accessibility, scalability, and cost-effectiveness. They offer psychological services to people who may otherwise be confronted with obstacles, including financial and geographical isolation, and stigma (Prochaska et al., 2020).

Studies have demonstrated that consumers tend to complain about fewer depressive and anxious symptoms following an extended time of engagement with AI-based applications (Fitzpatrick et al., 2017; Fulmer et al., 2018). In addition, the perceived lack of judgment of AI by many clients is noted, and it is characterized as a safe place to share vulnerability (Sarda et al., 2021). In this regard, AI can be democratization of mental health care provision and the expansion of the scope of clinical interventions. However, other negative consequences of these technological inventions are complicated relational and moral issues. Though fluent in language, AI systems have no real emotional intelligence, awareness, or moral judgment. The empathy they produce is computational in nature; it is produced by the large language models that have been trained with massive datasets of human communication, as opposed to being produced by experience or emotional sensitivity (Ho et al., 2022). As a result, the process of engaging with AI can lead to what other researchers refer to as synthetic empathy, the perceived emotional resonance, which is feigned but does not come naturally (Miner et al., 2019). This leads to the question: Is it possible to have a meaningful therapeutic alliance, or to be able to maintain it, when one of the partners is not emotionally subjective? This issue becomes particularly acute since even the therapeutic alliance is a relationship. In addition to cognitive restructuring or behavior change, psychotherapy is also provided by intersubjective and interpersonal processes like empathy, mirroring, and transference (Mitchell, 1988; Safran & Muran, 2000).

These processes are based on the ability of the therapist to identify, perceive, and react to the emotional conditions of the client and this ability is based on human neurobiology and emotional awareness. However, AI systems can simulate such behaviors by means of affective computing, but they cannot have empathy or compassion. Consequently, according to a number of scholars, AI-assisted therapy can produce a functional alliance, which is helpful in terms of collaborating on tasks, but lacks the emotional depth required to produce lasting therapeutic transformation (Provoost et al., 2020; Ho et al., 2022). However, there has been an early indication that clients can develop attachment to AI systems that are relational. Research shows that people might personify chatbots and endow them with the attributes of human beings, including attentiveness, affection, and empathy (Sarda et al., 2021; Fiske et al., 2019). In other instances, users indicate that they feel much safer sharing distressing events with AI agents than with human therapists because they believe that the latter will judge them (Inkster et al., 2018). This suggests a paradox to AI: the user might develop a bond with a human therapist, which is not necessarily the case with AI, since it does not have the same emotional reciprocity as a human therapist. In this, the

therapeutic alliance will be partially constructed through the perception of the user as opposed to the relational reciprocity. Clinically, this has very deep implications. When clients can obtain therapeutic benefit by interacting with AI, does this change the relationship to mean a subjective phenomenon as opposed to a phenomenon inherently mutual? Or does such an advantage only indicate temporary ease and not permanent change? Moreover, how does one deal with emotional attachment to non-humans, especially when they are vulnerable and need attachment and validation (McDonald et al., 2023)? In addition to the issues of relationships, there are also ethical and professional issues. Data privacy, algorithmic bias, and informed consent make it a more complicated task to implement AI tools into practice (Char et al., 2023).

To illustrate, AI systems can auto-promote negative stereotypes in case they are trained with biased data or jeopardize confidentiality by sharing data with third parties. Furthermore, therapeutic displacement may occur, as clients may replace human therapy with interactions with AI, which may lead to overlooking profound emotional work that may be performed with the help of human interaction and empathy (Ho et al., 2022). It is against this context that the present research aims to critically analyze the potential of AI-assisted psychotherapy to change the relationships between therapist and client during therapy. It examines the two-sidedness of the role of AI, as an access facilitator of mental health and an interferon of the human relational fabric, which is the core of psychotherapy.

In particular, it explores (a) the perception of quality and depth of alliance by clients and therapists based on AI-assisted settings; (b) the mechanisms aiding or preventing the establishment of trust, empathy, and collaboration between human and AI agents; and (c) how the hybrid model of the integration of human and AI agent can be optimally applied to improve therapeutic outcomes. The location of these questions within the known theoretical frameworks, especially that of Bordin (1979) of the tripartite model of the alliance and the relation-constructivist views, offers the present paper to the ongoing debate of the future of psychotherapy in a technologically mediated world. The question of the impact of human-AI exchange on therapeutic relationships is not just an academic experiment but a clinical issue of urgency due to the growing role of digital agents in delivering emotional support to patients in daily life.

Literature Review:

Conceptualizing the Therapeutic Alliance

The therapeutic relationship has been cited to be one of the most influential indicators of treatment outcome in various psychotherapeutic modalities. The idea was originally explained by Bordin (1979) and includes three components, which are inseparable: the agreement of therapeutic goals, task distribution, and the creation of a personal bond. The alliance is the emotional and cooperative system where the therapeutic change takes place. As meta-analytic research has always demonstrated, a good alliance is positively associated with a variety of therapeutic methods, such as cognitive-behavioral, psychodynamic, and integrative ones (Fluckiger et al., 2018; Horvath et al., 2011).

The classic conceptualizations of the alliance focus on the skill of the therapist to offer empathy, authenticity, and acceptance that, conversely, lead to client trust and involvement (Rogers, 1957; Norcross & Wampold, 2019). According to this perspective, the relationship between two individuals is not just a working alliance but a jointly formulated emotional experience that assists the client in his/her exploration of vulnerability and transformation. This is the depth of relationships that is based on the affective attunement and mutual understanding, which is regarded as the sign of successful psychotherapy (Safran & Muran, 2000). Nevertheless, the advent of AI-assisted psychotherapy questions these suppositions.

By definition, AI systems are not conscious, not emotional, and do not possess subjective intentionality. The question is, do such systems have the ability to access the relational and affective aspects of therapeutic alliance? Theorists like Ho et al. (2022) introduce the idea of a functional alliance - a type of perceived relational linkage on the foundations of simulated empathy and behavior responsiveness. In this respect, the clients can feel that the alliance is real, although AI

responses are not based on their emotions but on algorithms (Wampold, 2015). This simulated empathy does not necessarily need to be felt in a phenomenological way but can still have a therapeutic effect because it satisfies the needs of the users to be validated and predictable (Cameron et al., 2023). However, critics claim that this kind of relationship does not have the existential authenticity of human relations, and it can strengthen superficial emotional involvement (McDonald et al., 2023). This argument represents a larger philosophical conflict between functional realism (it is what counts is the effect of empathy) and phenomenological authenticity (it is what counts is the experience of empathy).

AI am in Psychotherapy: Current Applications

The psychotherapy application of AI is on a spectrum of fully autonomous chatbot-based interventions to clinician-assisted tools that aid in decision-making and monitoring. Woebot, Wysa, and Replika are automated chatbots based on natural language processing (NLP) to provide structured interventions, which are mostly oriented on Cognitive Behavior Therapy (CBT) and Dialectical Behavior Therapy (DBT) (Fitzpatrick et al., 2017; Inkster et al., 2018).

The apps take a user through cognitive restructuring exercises, mood monitoring, and mindfulness practices using the conversational frameworks that simulate a supportive conversation. An example is the use of sentiment analysis in Woebot to respond to user tones, and CBT-based coping prompts and motivational interviewing techniques of Wysa (Inkster et al., 2018). In addition to chatbots, AI-enhanced clinical systems can help human therapists through the use of data-driven insights. They are linguistic markers analyzers in the transcript of the session, emotional tone monitors, and predict relapse risk based on behavioral trends (Prochaska et al., 2020; Miner et al., 2019). Speaking of which, applications like Ginger.io or Ellipsis Health claim to infer mental health conditions in real-time based on the passive data collection on smartphones and voice analysis (Greer et al., 2019).

There is some conflicting empirical evidence on the clinical effectiveness of such tools. Depression and anxiety symptoms have been shown to be reduced slightly in those who used AI chatbots after controlled studies (Fitzpatrick et al., 2017; Inkster et al., 2018). Nevertheless, such results are usually characterized by a high turnover rate and a lack of emotional involvement (Fulmer et al., 2018; Bendig et al., 2019). The experience is often reported by the users as useful to self-reflect and control emotions, but inadequate to meet deeper psychological needs (Provoost et al., 2020).

AI-based interventions have obvious constraints, especially in complex disorders that demand sensitive relational knowledge, including trauma, personality disorders, or psychosis (Miner et al., 2019). Although AI is highly consistent, scalable, and available, it does not have the ability to have authentic emotional capture or intersubjective repair (McDonald et al., 2023). However, AI-assisted psychotherapy has great potential in a low-resource setting, as mental health practitioners are limited. It is especially appealing to mental health activities with an international focus because it can be scaled (WHO, 2021).

The Alliance in Digital and Artificial Intelligence Situations

Evidence on the topic of digital therapy, such as online CBT platforms, mobile apps, and teletherapy, points to the idea that even without physical co-presence, the clients still may develop a feeling of therapeutic alliance (Sarda et al., 2021; Berger, 2017). In a treatment approach based on the internet, clients tend to interpret the interface itself as a sympathetic medium and give digital agents human-like qualities (Rehm et al., 2016). This anthropomorphization, in which the user imposes human traits upon non-human systems, is a key factor in the formation of the dynamics of relations in AI-assisted psychotherapy (Ho et al., 2022). The conversational design of the AI and the social cognition processes inherent to users may force them to anthropomorphize (Nass & Moon, 2000).

Users will become more inclined to report trust and emotional comfort when AI agents use warmth, validation, and self-disclosure scripts (Seabrook et al., 2021). Nevertheless, researchers warn that this perception of friendship might be an illusion, based on responsiveness, and not mutuality (Bendig et al., 2019). The illusion of empathy is safe in the short run but may be a

problem in the long run when the users become emotionally dependent or avoid human contact (Provoost et al., 2020). On the part of the therapist, AI tools are ambivalent. Clinician surveys indicate optimism over the possibility of AI to increase efficiency, cut down on administrative activities, and increase diagnostic accuracy (McDonald et al., 2023).

However, most practitioners are worried that the use of algorithmic systems may destroy the relational essence of psychotherapy (Torous and Roberts, 2021). It is also fearful of data privacy, transparency, and ethical responsibility, especially when clients share sensitive data with inhuman agents (Luxton, 2016).

Theoretical Model: Attachment and Relational Approaches.

The conceptual framework on how to comprehend the alliance in AI-assisted psychotherapy can be placed at the crossroads of the attachment theory (Bowlby, 1988) and the relational psychoanalysis (Mitchell, 1988). Both schools of thought theorize the therapeutic relationship as a form of attachment bond, with the therapist serving as a safe haven and secure base, which facilitates exploration and emotion regulation. In this context, the relationship between the alliance is not just a work contract; it is an emotionally-charged attachment interaction. Once clients are communicating with AI-based therapeutic systems, the pattern of attachment can be triggered. Users who view the AI as sensitive and predictable can develop what Provoost et al. (2020) refer to as the para-social attachments, which are emotional bonds similar to human attachment bonds but which happen with the artificial beings. The bonds are particularly acute when it comes to people with insecure or avoidant attachment styles and who might also feel less threatened or judged by AI systems than by human therapists (Ho et al., 2022).

In a relational psychoanalytic perspective, the AI is used as a reflection of the client's relational expectations. This ability to ruminate on oneself can be supported in some ways by the AI and the fact that it is always available and non-judgmental, although due to the absence of intersubjectivity, the AI can never truly recognize the other person or repair (Mitchell, 1988; Aron, 1996). Such an imbalance raises the ethical basis of practice in the field of therapy: Is it ever possible to create a system without moral capacity or with emotional sensitivity to take part in the healing of relational trauma? As a result, the therapeutic alliance of AI-assisted psychotherapy needs to be redefined as a hybrid construct - a part cognitive collaboration, part simulated affective exchange. This hybrid nature can only be understood by incorporating the field of social cognition, attachment theory, and affective neuroscience. Finally, AI-assisted psychotherapy may serve as a special laboratory to test what is fundamentally human in therapeutic relationships, namely empathy, presence, and the ability to be truly emotionally attuned.

Methodology:

Research Design

This research used a qualitative phenomenological research design in its attempt to investigate the lived experiences of both the therapist and the client involved in AI-assisted psychotherapy. Phenomenology, as defined by van Manen (1990), aims at discovering the nature of human experience in rich and descriptive descriptions as opposed to statistical generalization. Such a method was especially applicable to the current study since the therapeutic relationship, and in particular, AI, is a highly subjective, relational, and interpretive phenomenon. Although quantitative approaches may be used to assess satisfaction or symptom remedies, they do not numerically reflect the subtle emotional, ethical, and existential aspects of human-AI relationality (Smith, Flowers, and Larkin, 2009).

Phenomenological orientation provided the opportunity to critically analyze the experience of empathy, trust, and emotional resonance when one of the partners of the alliance is not human. The focus on including the first-person narratives of the participants in the study was aimed at shedding light on the way in which conventional ideas about the therapeutic connection are redefined in the digital realm. Furthermore, this design was selected to address the gap in the existing literature, as indicated by Ho et al. (2022) and McDonald et al. (2023), since most available articles provide an overview of AI therapy based on user satisfaction

surveys or outcome measures, disregarding the experiential aspect of the formation of the alliance.

Participants

There were two groups of participants in the research:

1. Clinicians (n = 10) – Licensed psychologists and psychotherapists who already use AI-based apps like Wysa, Limbic, or Woebot in their clinical practice. These subjects were chosen according to their level of experience with hybrid or AI-based therapy models, so that they were exposed to both classic and AI-based therapeutic dynamics.

2. Clients (n = 15) were Adults aged between 20 and 45 and had actively used AI-assisted therapy platforms for at least six weeks. This time was selected because it was necessary to make sure that participants had enough time to build a significant interactional pattern with the AI system.

The recruiting telehealth and digital therapy sites will be used to recruit individuals who will be based in the United Kingdom and the Czech Republic between February and June 2025. Purposive and snowball methods of recruitment were used (Creswell & Poth, 2018), and the first invitations were issued through the influence of professional networks and mental health community forums. Both groups had to meet the inclusion criteria of being fluent in English and being interested in contemplating emotional experiences in the course of therapy. Informed consent was given by all the participants. The National University of Modern Languages in Islamabad had an ethical committee (Institutional Review Board, IRB) which granted ethical approval of the research and adhered to the ethical standards of the British Psychological Society (BPS, 2021) in terms of confidentiality, data protection, and psychological well-being.

Data Collection

The information was gathered via semi-structured and in-depth interviews that were carried out online and through the platforms of secure video conferencing (Zoom or Microsoft Teams). The interview period was between 60 and 90 minutes, and the participants took advantage of the time to expand on their subjective experiences without straying too much from the main research themes. The interview guide was based on the previous research on digital therapeutic alliances (Sarda et al., 2021; Provoost et al., 2020) and revolved around four main domains:

1. Perceived empathy - How the participants felt understood and responsive by the AI or the human therapist through the use of AI tools.
2. Trust and reliability -A Measure of how much the participants placed trust in the guidance offered by the AI or its data-driven recommendations.
3. Emotional resonance - The degree of emotional attachment or identification that participants had with the system, or whether they felt understood.
4. Formation and maintenance of alliances - How the existence of AI influenced the perception of partnership, objective, and relationship richness. All the interviews were recorded and transcribed word-for-word.

To make data accurate, member checking was carried out by sending transcripts to participants so that they could clarify or elaborate on their answers (Lincoln & Guba, 1985). To record non-verbal and situational information that aided in inferring the meaning, field notes were kept.

Data Analysis

Thematic Analysis (TA) was adopted to analyze the collected data in accordance with the six-step model of Braun and Clarke (2021). The method is selected due to its flexibility and the ability to reveal patterns of meaning in qualitative data, and also be close to the experiences of the participants. The analysis was done in the following steps:

1. Familiarization: The transcripts were read repeatedly in order to get the researcher familiar with the data and record initial observations.
2. First Coding: Inductive coding of segments of text pertinent to empathy, trust, relational dynamics, and human-AI interaction did not require the use of existing categories.
3. Theme Development: Codes were clustered to form candidate themes representing recurrent patterns of experience, including: simulated empathy, functional trust, emotional distance, and relational augmentation.
4. Themes: Reviewing the Themes. Themes were polished by considering the coherence of

groups of participants and consistency with theoretical constructs of the therapeutic alliance.

5. Defining and Naming Themes: The themes were explained and conceptualized based on the tripartite model of Bordin (1979), goals, tasks, and bond to understand how these factors are expressed in AI-based settings.
6. Reporting: The last step entailed the synthesis of the themes into a logical story with the support of representative quotes by a therapist and clients.

To improve the level of credibility and reliability, the study employed peer debriefing involving two independent qualitative researchers who were conversant with psychotherapy research. The reflexive journaling ensured the track of what the researcher assumed about technology and relationality (Smith et al., 2009). The data were managed and coded with NVivo software (v.14). With this intensive methodological procedure, the research aimed to identify the phenomenological nature of the therapeutic relationship during AI-assisted psychotherapy to shed some light on how trust, empathy, and collaboration are bargained within human-AI relational domains.

Results

The interpretation of the interviews demonstrated three general themes that summarized the experiences of participants in forming an alliance and being emotionally involved in AI-assisted psychotherapy. These themes, which include Functional Empathy and Perceived Understanding, Alliance Without Attachment and Hybrid Therapeutic Models, all demonstrate the dynamic nature of therapeutic relationships in the digital mental health environment..

Functional Empathy and Perceived Understanding

One of the key discoveries was about the experience of the participants, where there was what clinicians referred to as algorithmic empathy. Clients often asserted that AI tools like Wysa or Woebot made them feel understood, supported, and emotionally validated, although they recognize that they are artificial. One client reflected: “It didn’t judge me. It just responded calmly. I literally became more open than I am with my therapist at times. This is a paradox of feeling safe emotionally in a non-sentient relationship, and it occurred very often. Clients also projected empathy towards the nonjudgmental attitude and linguistic consistency of AI. The conversational tone of the chatbot, reflective phrasing, and feedback helped the chatbot to seem to understand what users thought it did.

This was, however, perceived negatively by clinicians. Some of them complained that clients were putting emotional nuances on automatic answers. One psychologist noted: They see the light where there is none- it is not compassion, it is recognition of patterns disguised as concern. This is reminiscent of the study by Ho et al. (2022), who reported the anthropomorphic attitude of users towards AI-based counseling devices and understood algorithmic feedback as actual emotional sensitivity. However, even in this study, therapists also recognized therapeutic value in such projections: when clients can feel heard, they can be more open, which helps them to better self-reflect. Interestingly, clients tended to put the neutrality of AI in contrast to the perceived biases of human therapists. As one participant put it: I do not fear being judged with the help of the bot. It’s just there to listen.”

This implies that functional empathy, the simulation of comprehension by affectively tuned language, may have a legitimate use in the beginning of therapeutic interaction, especially with clients having a high social anxiety or fear of judgment.

Figure 1

Thematic Map Thematic Map Outlining the Essential Dimensions of Functional Empathy and Algorithmic Understanding in AI-Assisted Therapy

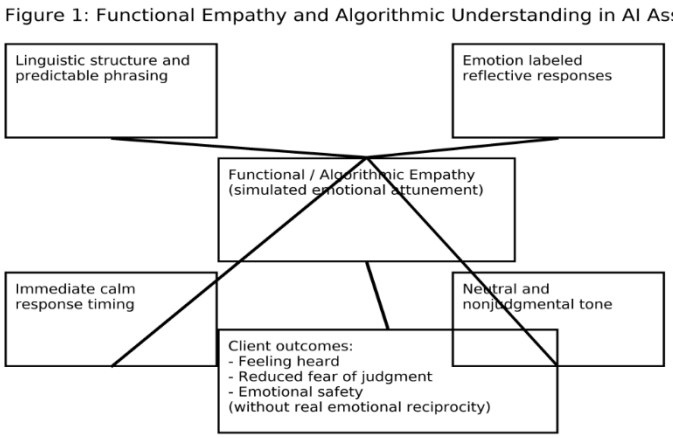

Figure 2 represents the way in which the perception of empathy among clients will come to be the result of the language structure, timing of responses, and emotional labelling by the AI, even though there will not really be the feeling of affective reciprocity.

Alliance Without Attachment

The second significant theme was Alliance Without Attachment, which was also an expression of the instrumental but emotionless character of the human-AI collaboration. The clients reported AI to be useful to particular therapy-related tasks (including setting goals, monitoring moods, and cognitive restructuring) but not applicable to more profound relational or emotional work. One of the therapists stated that AI is excellent with CBT check-ins but not grief or uncertainty. Such a difference is similar to the model of a therapeutic alliance introduced by Bordin (1979) that consists of goals, tasks, and bond dimensions.

Although the AI system was useful in the goal and task aspect, by providing reminders and other psychoeducational materials to monitor the progress, the bond aspect was not developed. Some clients have reported having had one-dimensional interactions with AI, which they called flat, predictable, or repetitive. They admitted that although the chatbot could be a reflection of their language, it was not as spontaneous and emotional as a real conversation. One participant commented: It hearkens, and is not listening—it is talking to an echo. Clinicians noted that AI is currently at its best with regard to structured interventions like Cognitive Behavioral Therapy (CBT) or some form of mindfulness-related check-ins, where the depth of relations is not as important.

Nevertheless, ambivalence tolerance, empathy, or emotional holding, which are necessary in situations such as trauma therapy or grief counseling, were found to be insufficient with AI. Regardless of these shortcomings, the stability and availability of the AI were of value to some clients. It is always there, to know, as I said, it is no appointments, no judgment. This dependability provided a feeling of predictability that added to a pragmatic type of alliance - functional but devoid of feelings. On the whole, this theme displays a type of instrumental alliance that is not the same as the standard relations between humans. AI is not intersubjective, but structure and support, which do not provide emotional transformation like therapeutic attachment.

Hybrid Therapeutic Models

The last theme was the propensity of participants towards hybrid therapy, where people can use their human skills, but with the effectiveness of AI. Clinicians and clients shared an interest in AI-assisted, but not AI-replacing, models of the therapeutic process. Clients appreciated the ability of AI tools to offer between-session continuity (daily check-ins, progress summaries, or reminders) that kept them accountable and in

touch with their treatment objectives. One client shared: It assists in monitoring my moods, and thereafter I speak about the patterns with my therapist. It feels like teamwork." Clinicians considered AI as a partner, which would complement therapeutic work.

One of the therapists called AI a co-therapist that could expand their reach: Neither does it substitute me--it stretches me over. I am able to concentrate on the more substantial items, and AI deals with the framework. This example of a hybrid model is a representative of a complementary partnership, whereby human therapists and AI systems jointly help to promote client welfare. AI gives order, surveillance, and access, whereas therapists give compassion, moral discretion, and insight.

The application of AI raised new ethical and professional issues: data privacy, informed consent, and possible over-dependence on the feedback of the machine. However, the majority of interviewees considered that these problems can be controlled as part of a controlled system. In general, the results indicate the best implementation of AI-assisted therapy in human-based clinical frameworks that incorporate emotional authenticity and technological consistency. The new alliance is not a substitute for human connection, but an extension that increases accessibility and continuity of care.

Figure 2

The Hybrid Therapeutic Alliance Conceptual Model of AI-Assisted Psychotherapy

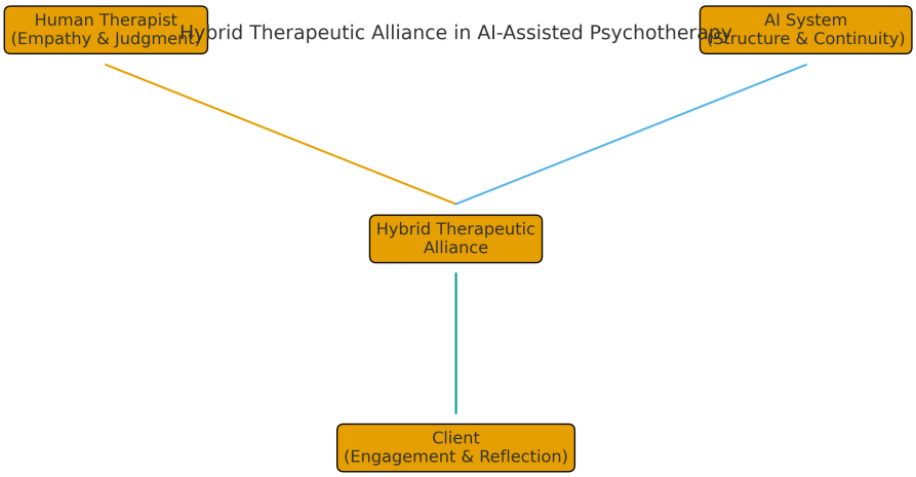

Figure 2 shows that the interaction among human empathy, AI structure, and client engagement is described in a model that shows that a symbiotic relationship between a clinician and AI increases the therapeutic outcomes without reducing the connections between a human and AI.

Discussion

This research proposes that AI-based psychotherapy helps to establish a functional but emotionless therapeutic relationship, which is consistent with the growing amount of evidence that suggests that digital technologies can be used to facilitate, and not to replace, the depth of the human relationship (Fiske et al., 2019; Ho et al., 2022). The interviews also indicated that clients often find empathy in the AI-generated responses, a phenomenon that clinicians described as algorithmic empathy. Such a perception, however, is cognitively built, but not co-experienced emotionally. That is, although clients might experience being heard by an AI system, they feel that, because of linguistic simulation and predictable feedback that is emitted by the AI system, they are heard as opposed to being moved by the affective resonance.

This perceived empathy/experienced empathy difference resonates with a previous study by Bickmore and Picard (2005), who established that users tend to project social and emotional traits upon conversational agents that express a consistent, polite, and context-sensitive dialogue

style. In the same way, Lucas et al. (2014) also found that people are more willing to disclose to virtual agents because they do not feel judged as actively, which means that emotional safety can be attained even when there is no empathic reciprocity. But the results of the current research point to the fact that the mentioned safety is only surface-level and does not incorporate the intersubjective aspect of human-to-human therapeutic interactions (Wampold & Imel, 2015).

These findings lead to an increasing acknowledgement that AI systems can conceptualize the functional aspects of the therapeutic relationship (especially setting goals and managing tasks), but fail to provide the bond aspect of the tripartite model of Bordin (1979). Customers appreciated the efficiency, the availability, and the neutrality of AI, especially when it comes to cognitive-behavioral and psychoeducational settings. But clients and clinicians stressed that emotional interactions with AI were either flat, repetitive, or detached. This evidence highlights why AI can at best be seen as a continuation of the therapist, but it cannot ever substitute the interpersonal aspects of the therapeutic relationship that are based on attunement, empathy and co-presence (Norcross & Lambert, 2019).

Clinical Practice Implications.

These findings have many implications to clinical training and therapeutic delivery. Instead of viewing AI as a replacement of human interaction, clinicians should consider it as an addition tool, which improves therapeutic processes with the help of structure, monitoring, and accessibility (Greene et al., 2021). Clients can be assisted by AI-based systems to ensure continuity between sessions, monitor symptoms, or reinforce cognitive restructuring tasks, and thus increase adherence and engagement. Nevertheless, the excess use of AI may lead to the loss of the centrality of the therapist-client relationship that is the most stable predictor of therapeutic outcomes regardless of the model (Horvath et al., 2011).

Thus, the therapists have to gain digital relational competence, a new set of skills that incorporates the mastery of moving between emotional involvement, boundaries, and expectations in the AI-mediated care. This skill involves the ability to handle the attachment of clients to AI technology, to identify transference-like reactions to nonhuman agents, and to use digital data as part of therapeutic contemplation (Riva et al., 2023). As an example, clinicians can consider AI-gathered data (e.g. mood tracking or journaling insights) to be prompts to be further discussed during the face-to-face sessions. In addition, the research claims that hybrid therapeutic models (in which artificial intelligence is engaged on routine or structured interactions and therapists are occupied with relational and interpretive work) can be the most effective and ethical AI-assisted care arrangement (Vaidyam et al., 2019).

These models are consistent with the concept of augmentation, and not replacement, whereby technology increases human capacity of therapeutic work, without displacing empathy or judgment. According to one of the clinicians participating in the study, AI does not substitute me; it makes me reach further. The modules on AI literacy and ethics must be included in the clinical training programs and professional societies to allow practitioners to critically evaluate digital tools, comprehend the algorithmic biases, and preserve confidentiality in digital environments. With the growth of digital mental health, therapists need to be the moral and relationship-based foundation of care, and technology must only be used as a tool to enhance, not replace, human connection.

Ethical Considerations

Introduced into psychotherapy, the use of AI presents a complicated set of ethical, legal, and relationship issues. Research subjects in this research actively raised their voices on issues of data privacy, informed consent, and emotional dependency, which require the active regulation and professional concern. Informed consent is one of the significant ethical concerns. Customers tend to overrate the therapeutic abilities of AI and think that a reply with coherent emotions indicates that the client deeply understands (McDonald et al., 2023). These misconceptions may result in either false trust or shunning of human therapy in cases where there is an emotional distress. Thus, when starting treatment, clinicians need to clearly define the scope and limitations of AI tools and, in particular, they should focus on introducing AI as a

supportive tool, not a replacement of professional decision-making or empathy (Luxton, 2016). Confidentiality and data security also become a major issue.

AI-based systems are known to regularly gather sensitive emotional information, which can be stored or handled by any third-party provider, which can be found as a potential vulnerability (Char et al., 2018). Clinicians have to make sure that the platforms meet the data protection standards, including the General Data Protection Regulation (GDPR) of the EU and have to inform the clients about the data-sharing patterns. The ethics of professional associations (e.g., APA, BPS) put growing importance on transparency, encryption, and autonomy on the side of the client in the realm of digital care. A less obvious yet equally important ethical threat is an emotional addiction to AI. The presence of chatbots at any time may create overdependence, especially in people who have attachment disorders as multiple clients mentioned in the study. Although accessibility contributes to the support, it also can negatively affect self-regulation or postpone intervention by human therapists when more sophisticated emotional processing is needed (Fiske et al., 2019). Clinicians need to be friendly and restrictive at the same time, allowing the client to differentiate between supportive functionality and relational depth. Lastly, AI-mediated therapy raises the question of therapeutic responsibility and accountability.

With the potential of an AI system to deliver feedback contributing to the emotional state of a client, the issue of liability becomes questionable, be it of the clinician, developer or the institution. To solve such problems, technologists, clinicians, and ethicists should join forces to come up with clear models of collective responsibility and moral control. Overall, the paper emphasizes that although AI assisted psychotherapy has potential to increase accessibility, structure and engagement, it is incapable of emulating the human therapeutic practice in terms of its emotional, ethical and relational aspects. Given the empathy, ethical issues, and human understanding, clinicians should be at the heart of the therapeutic process whereby technology use in mental health care is guided by empathetic and ethical issues.

Conclusion

The results of the present research indicate the potential of AI-aided psychotherapy as a transformative process but also suggest that it has a major implication on the traditional therapeutic relationship. The AI systems have significant potential to increase access to mental health care through providing immediate, structured, and scaleable interventions. The respondents indicated that AI tools have the potential to encourage behavior change and offer a stable form of guidance, especially within the framework of structured therapeutic programs, including CBT check-ins, mood monitoring, and psychoeducation. These advantages come in handy especially to clients who find it difficult to deal with availability, stigma, or anxiety about human interaction. Nevertheless, the findings also show that there is a certain limit to the AI capabilities in the therapeutic relationship. Although AI can be empathetic with empathetic word choices and a conversational reaction, its ability to empathize and respond to emotions, feelings, and relationships is deficient, as well as the depth of emotion, intuition, and presence that define effective human therapy.

The relationship that clients develop with the AI is mainly functional based on routine, predictability, and cognitive involvement as opposed to that which is truly relational. Emotional reciprocity, subtle interpretation, and common vulnerability are still highly human traits, which the present-day AI system is not capable of imitating. Based on these strengths and weaknesses, hybrid therapeutic models are seen to hold the key to the future of mental health care. These models utilize the efficiency, consistency and reach of AI systems and maintain the ethical judgement, compassion and emotional intelligence of human clinicians. Thoughtfully incorporated, AI can serve as an additive tool and not a substitute and contribute to creating a more balanced and responsive therapeutic ecosystem that can increase care without undermining the nature of the human relationship.

Future Directions

The future studies are to utilize multimodal and longitudinal designs to study long-term

sustainability of AI-assisted alliances and their effects on long-term outcomes (relapse prevention and emotional regulation). Combining neurophysiological and behavioral outcomes such as affective synchrony, eye-tracking, or neural correlates of empathy may add more insight into the effects of AI interaction on emotional processes and therapeutic bonding in the long run.

References

Aron, L. (1996). *A meeting of minds: Mutuality in psychoanalysis*. Analytic Press.
Google Scholar Worldcat Fulltext

Bendig, E., Erb, B., Schulze-Thuesing, L., & Baumeister, H. (2019). The next generation: Chatbots in clinical psychology and psychotherapy to foster mental health. *Zeitschrift für Psychologie, 227*(4), 227–239. https://doi.org/10.1027/2151-2604/a000386
Google Scholar Worldcat Fulltext

Berger, T. (2017). The therapeutic alliance in internet interventions: A narrative review and suggestions for future research. *Psychotherapy Research, 27*(5), 511–524. https://doi.org/10.1080/10503307.2015.1119908
Google Scholar Worldcat Fulltext

Bickmore, T. W., & Picard, R. W. (2005). Establishing and maintaining long-term human-computer relationships. *ACM Transactions on Computer-Human Interaction, 12*(2), 293–327. https://doi.org/10.1145/1067860.1067867
Google Scholar Worldcat Fulltext

Bordin, E. S. (1979). The generalizability of the psychoanalytic concept of the working alliance. *Psychotherapy: Theory, Research & Practice, 16*(3), 252–260. https://doi.org/10.1037/h0085885
Google Scholar Worldcat Fulltext

Bowlby, J. (1988). *A secure base: Parent-child attachment and healthy human development*. Basic Books.
Google Scholar Worldcat Fulltext

Braun, V., & Clarke, V. (2021). *Thematic analysis: A practical guide*. Sage Publications.
Google Scholar Worldcat Fulltext

British Psychological Society. (2021). *Code of ethics and conduct*. BPS Press.
Google Scholar Worldcat Fulltext

Cameron, G., et al. (2023). Simulated empathy in conversational agents: A systematic review. *Journal of Medical Internet Research, 25*, e43088.
Google Scholar Worldcat Fulltext

Char, D. S., McDonald, A., & Wykes, T. (2023). Artificial intelligence, ethics, and mental health: Navigating bias and trust in digital psychiatry. *Frontiers in Psychiatry, 14*, 1158294.
Google Scholar Worldcat Fulltext

Char, D. S., Shah, N. H., & Magnus, D. (2018). Implementing machine learning in health care—Addressing ethical challenges. *New England Journal of Medicine, 378*(11), 981–983. https://doi.org/10.1056/NEJMp1714229
Google Scholar Worldcat Fulltext

Creswell, J. W., & Poth, C. N. (2018). *Qualitative inquiry and research design: Choosing among five approaches* (4th ed.). Sage Publications.
Google Scholar Worldcat Fulltext

Fiske, A., Henningsen, P., & Buyx, A. (2019). Your robot therapist will see you now: Ethical implications of embodied artificial intelligence in psychiatry, psychology, and psychotherapy. *Journal of Medical Internet Research, 21*(5), e13216. https://doi.org/10.2196/13216
Google Scholar Worldcat Fulltext

Fitzpatrick, K. K., Darcy, A., & Vierhile, M. (2017). Delivering cognitive behavior therapy to young adults with symptoms of depression and anxiety using a chatbot: Randomized controlled trial. *JMIR Mental Health, 4*(2), e19. https://doi.org/10.2196/mental.7785
Google Scholar Worldcat Fulltext

Flückiger, 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
Google Scholar Worldcat Fulltext

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
Google Scholar Worldcat Fulltext

Greene, M., Thorp, L., & Williamson, A. (2021). Human-centered AI in clinical psychology: Augmentation over automation. *Frontiers in Psychology, 12*, 742510. https://doi.org/10.3389/fpsyg.2021.742510
Google Scholar Worldcat Fulltext

Greer, B., Robotham, D., & Lovell, K. (2019). Digital interventions for mental health: The role of big data and artificial intelligence. *Journal of Mental Health, 28*(3), 229–235.
Google Scholar Worldcat Fulltext

Ho, A., Hancock, J. T., & Miner, A. S. (2022). Psychological, relational, and ethical issues in AI therapy. *Perspectives on Psychological Science, 17*(2), 497–512.
Google Scholar Worldcat Fulltext

Horvath, A. O., Del Re, A. C., Flückiger, C., & Symonds, D. (2011). Alliance in individual psychotherapy.

Psychotherapy, 48(1), 9–16. https://doi.org/10.1037/a0022186
Google Scholar Worldcat Fulltext

Inkster, B., Sarda, S., & Subramanian, V. (2018). An empathy-driven, conversational artificial intelligence agent (Wysa) for digital mental wellbeing: Real-world data evaluation. *JMIR mHealth and uHealth, 6*(11), e12106. https://doi.org/10.2196/12106
Google Scholar Worldcat Fulltext

Lincoln, Y. S., & Guba, E. G. (1985). *Naturalistic inquiry.* Sage Publications.
Google Scholar Worldcat Fulltext

Lucas, G. M., Gratch, J., King, A., & Morency, L. (2014). It's only a computer: Virtual humans increase willingness to disclose. *Computers in Human Behavior, 37*, 94–100. https://doi.org/10.1016/j.chb.2014.04.043
Google Scholar Worldcat Fulltext

Luxton, D. D. (2016). *Artificial intelligence in behavioral and mental health care.* Academic Press.
Google Scholar Worldcat Fulltext

McDonald, A., Reeve, A., & Wykes, T. (2023). Trust and transparency in AI mental health tools: A clinician perspective. *Frontiers in Psychiatry, 14*, 1160054.
Google Scholar Worldcat Fulltext

McDonald, N., Mökander, J., & Floridi, L. (2023). Reimagining psychotherapy with artificial intelligence: Ethical and clinical considerations. *Frontiers in Psychology, 14*, 1103256.
Google Scholar Worldcat Fulltext

McDonald, R., Byrne, C., & Choudhury, M. (2023). Algorithmic trust in mental health apps: Ethical responsibilities in AI-mediated care. *AI & Society, 38*(1), 89–101.
Google Scholar Worldcat Fulltext

Miner, A. S., Milstein, A., & Hancock, J. (2019). Talking to machines about personal mental health problems. *JAMA, 321*(2), 135–136. https://doi.org/10.1001/jama.2017.14151
Google Scholar Worldcat Fulltext

Mitchell, S. A. (1988). *Relational concepts in psychoanalysis: An integration.* Harvard University Press.
Google Scholar Worldcat Fulltext

Norcross, J. C., & Lambert, M. J. (2019). Psychotherapy relationships that work III. *Psychotherapy, 56*(3), 350–365.
Google Scholar Worldcat Fulltext

Norcross, J. C., & Wampold, B. E. (2019). *Relationships that work* (3rd ed.). Oxford University Press.
Google Scholar Worldcat Fulltext

Prochaska, J. O., Wright, J. A., & Velicer, W. F. (2020). Using digital tools to extend the reach of mental health care. *American Psychologist, 75*(8), 1099–1111.
Google Scholar Worldcat Fulltext

Provoost, S., Lau, H. M., Ruwaard, J., & Riper, H. (2020). Embodied conversational agents in clinical psychology: A scoping review. *Journal of Medical Internet Research, 22*(8), e15161. https://doi.org/10.2196/jmir.6553
Google Scholar Worldcat Fulltext

Riva, G., Wiederhold, B. K., & Mantovani, F. (2023). The new frontiers of digital mental health: From technological innovation to relational competence. *Frontiers in Psychology, 14*, 1172892. https://doi.org/10.1089/cyber.2020.29183.gri
Google Scholar Worldcat Fulltext

Safran, J. D., & Muran, J. C. (2000). *Negotiating the therapeutic alliance: A relational treatment guide.* Guilford Press.
Google Scholar Worldcat Fulltext

Sarda, S., Das, S., & Inkster, B. (2021). Measuring therapeutic alliance in text-based therapy with AI chatbots. *Frontiers in Digital Health, 3*, 721571.
Google Scholar Worldcat Fulltext

Sarda, S., Dasgupta, A., & Inkster, B. (2021). Exploring therapeutic alliance in digital mental health interventions. *Frontiers in Digital Health, 3*, 696235.
Google Scholar Worldcat Fulltext

Smith, J. A., Flowers, P., & Larkin, M. (2009). *Interpretative phenomenological analysis: Theory, method, and research.* Sage Publications.
Google Scholar Worldcat Fulltext

Torous, J., & Roberts, L. W. (2021). The ethics of digital mental health. *Current Opinion in Psychology, 44*, 72–77.
Google Scholar Worldcat Fulltext

Vaidyam, A. N., Wisniewski, H., Halamka, J. D., Kashavan, M. S., & Torous, J. (2019). Chatbots and conversational agents in mental health: A review. *Current Psychiatry Reports, 21*(7), 36. https://doi.org/10.1177/0706743719828977
Google Scholar Worldcat Fulltext

van Manen, M. (1990). *Researching lived experience: Human science for an action sensitive pedagogy.* State University of New York Press. https://doi.org/10.4324/9781315421056
Google Scholar Worldcat Fulltext

Wampold, B. E. (2015). How important are the common factors in psychotherapy? *World Psychiatry, 14*(3), 270–277. https://doi.org/10.1002/wps.20238
Google Scholar Worldcat Fulltext

Wampold, B. E., & Imel, Z. E. (2015). *The great psychotherapy debate: The evidence for what makes psychotherapy work* (2nd ed.). Routledge.
Google Scholar Worldcat Fulltext

World Health Organization. (2021). *Global report on artificial intelligence in health: Ethics and governance*. WHO Press.
Google Scholar Worldcat Fulltext