Wiley
Human Behavior and Emerging Technologies
Volume 2026, Article ID 3616773, 20 pages
https://doi.org/10.1155/hbe2/3616773

WILEY

Research Article

Therapists and Technology: A Qualitative Study on AI's Role in Counseling

Thseen Nazir

Guidance and Counseling, Faculty of Education, Ibn Haldun University, Istanbul, Turkey

Correspondence should be addressed to Thseen Nazir; thseen.nazir@ihu.edu.tr

Received 28 July 2025; Revised 16 November 2025; Accepted 8 December 2025

Academic Editor: Rajneesh Choubisa

Background: Integrating artificial intelligence (AI) into mental health counseling revolutionizes therapeutic practices, with tools like ChatGPT and Woebot offering scalable and accessible support. While these advancements hold promise, the relational nature of counseling, rooted in empathy and cultural sensitivity, raises questions about the role of AI in maintaining the essence of human connection.

Purpose: This qualitative study explores therapists' attitudes and beliefs about integrating AI in counseling, examining factors that facilitate or hinder acceptance. As AI tools increasingly enter therapeutic settings, understanding therapists' perspectives on their potential, limitations, and ethical implications is crucial for responsible integration.

Methodology: Semistructured interviews were conducted with 20 therapists from diverse professional backgrounds and varying levels of AI exposure. Data were analyzed using reflexive thematic analysis following Braun and Clarke's [1] framework, with findings interpreted through sociotechnical systems theory (STS) and an extended technology acceptance model (TAM).

Findings: The analysis revealed four major themes (1) AI as a complementary tool; (2) ethical concerns regarding data privacy, bias, and transparency; (3) cultural sensitivity challenges; and (4) limitations in replicating therapeutic dynamics. These findings informed a conceptual framework highlighting the conditional nature of AI acceptance.

Conclusions: The findings underscore the need for ethical, culturally sensitive AI systems to complement human therapists rather than replace them. AI acceptance among therapists depends on addressing trust deficits, incorporating diverse cultural perspectives in AI development, and preserving the relational core of therapeutic practice. This study offers theoretical and practical insights for researchers, developers, and practitioners seeking to integrate AI responsibly into mental health services.

Keywords: artificial intelligence; attitude; counselor; ethical issues; therapist

1. Introduction

1.1. Background and Context. Artificial Intelligence (AI) has become a transformative force across numerous disciplines, including healthcare and mental health counseling. In mental health services, AI-powered tools such as chatbots, virtual agents, and emotion-recognition systems are increasingly being adopted as innovative solutions to enhance care delivery [2]. These tools harness machine learning and natural language processing (NLP) advancements to provide personalized and accessible mental health support. Platforms such as Woebot and Wysa, for example, deliver cognitive behavioral therapy (CBT)-based interventions through user-friendly conversational interfaces, demonstrating the growing ability of AI to simulate supportive interactions and provide tailored advice [3–5].

In this study, counseling refers to professional mental health services provided by trained therapists across various modalities, including individual psychotherapy, clinical counseling, school-based counseling, and private practice settings. This encompasses therapeutic work with both adults and children, delivered through diverse theoretical orientations (e.g., cognitive-behavioral, person-centered, and psychodynamic approaches). The term is used broadly

to reflect the heterogeneous nature of therapeutic practice while focusing specifically on mental health counseling rather than other forms of professional counseling (e.g., career or academic advising).

1.2. Benefits and Promise of AI in Mental Health Counseling. One of the key benefits of AI in mental health counseling lies in its ability to enhance accessibility, particularly for underserved populations. In regions where mental health professionals are scarce, AI tools can act as an immediate source of support, bridging the gap between clients and professional care. Research demonstrates that AI-based interventions can reduce barriers to access by providing immediate, stigma-free, and cost-effective support [5, 6]. ChatGPT, for instance, serves as a compelling supplement to psychotherapy and provides an easily accessible, user-friendly resource for individuals experiencing mental health challenges who have not yet sought professional support or undergone psychotherapy [7]. Fitzpatrick et al. [4] noted that AI chatbots can provide 24/7 availability, offering preliminary mental health interventions and promoting self-management strategies for clients experiencing mild to moderate symptoms of anxiety and depression. Similarly, in low-resource settings or during crises such as the COVID-19 pandemic, AI-powered platforms have proven instrumental in maintaining continuity of care [8].

AI shows promise in preliminary support, particularly where stigma prevents help-seeking. Studies indicate that chatbots provide judgment-free spaces for emotional disclosure and may especially appeal to younger, tech-comfortable populations [9, 10]

1.3. Challenges and Concerns. While counseling is traditionally rooted in human connection and empathy, the effectiveness of AI-driven interactions remains debated [11]. The capacity of AI systems to replicate the emotional depth, authenticity, and relational nuance central to therapeutic work has been questioned by both researchers and practitioners [11, 12]. Some studies suggest AI can mimic certain aspects of emotional resonance through advanced NLP and sentiment analysis [13], whereas others highlight fundamental relational deficits, particularly in understanding context-dependent emotional states and nonverbal communication [14, 15].

As AI continues to evolve, understanding how therapists, the primary gatekeepers of therapeutic practice, perceive and engage with this technology is crucial for its effective and ethical integration into counseling. Therapists' perceptions are a key determinant of AI adoption in counseling. While many acknowledge AI's potential to enhance efficiency, skepticism exists about its ability to address complex emotional needs and foster therapeutic rapport. For instance, AI-powered chatbots, such as Woebot and ChatGPT, have shown promise in promoting emotional disclosure and user engagement [13], but they are often criticized for their inability to provide the depth of understanding required in sensitive counseling contexts [14, 16, 17]. Cultural sensitivity poses critical challenges, as most AI systems lack diverse training data and struggle to adapt to multicultural contexts [18, 19], potentially harming therapeutic relationships.

Furthermore, the impact of AI on therapeutic dynamics remains a contentious issue. The therapeutic alliance characterized by trust, empathy, and rapport is central to effective counseling [20]. While AI systems can simulate empathy through NLP and emotional disclosure mechanisms, studies indicate that therapists and clients remain skeptical about their ability to replicate the authenticity and depth of human interactions [15, 21]. This skepticism is compounded by concerns that overreliance on AI could undermine the relational aspects of counseling, which are essential for building trust and promoting meaningful change.

Despite these challenges, AI also holds significant promise in practical applications, particularly training and supervision. Tools such as ChatGPT client simulation (CCS) have been used to help counselors-in-training develop their skills through realistic role-play scenarios, providing immediate feedback and enhancing their practical experience [22–24]. AI's ability to analyze large datasets and monitor client progress offers additional opportunities to complement human expertise, making it a valuable resource in clinical and educational settings. However, its effectiveness depends on addressing the limitations highlighted by therapists, including trust issues, cultural adaptability, and ethical concerns.

1.4. Research Gap and Contribution. Although research on AI in mental health counseling is growing, significant gaps remain in understanding therapists' attitudes and beliefs toward its integration. Existing studies have predominantly focused on the technical capabilities of AI systems and their potential to enhance user engagement and accessibility [25, 26]. However, there is limited research on how therapists, who are the professionals responsible for implementing these technologies, perceive AI, especially in terms of its ethical implications, cultural adaptability, and influence on the therapeutic relationship.

Furthermore, the unique challenges therapists face in multicultural and diverse counseling contexts have been largely overlooked, leaving a critical gap in the literature on how AI systems can be designed to address these complexities. Additionally, while studies have explored the use of AI for training purposes [23, 27], the practical implications of these tools in real-world counseling settings remain underexplored. Most importantly, there is a lack of qualitative research capturing therapists' first-hand experiences and nuanced perspectives, which is essential for understanding the broader implications of AI integration in counseling. Given the absence of prior in-depth qualitative research on therapists' lived experiences with AI, an exploratory, inductive approach is particularly appropriate for generating hypothesis-forming insights and developing theoretical frameworks grounded in practitioners' actual concerns and experiences.

1.5. Study Purpose and Research Questions. This study addresses these gaps by exploring therapists' attitudes and beliefs about integrating AI in counseling, focusing on their perceptions of its potential benefits, limitations, ethical concerns, cultural sensitivity, and impact on therapeutic dynamics. By foregrounding therapists' voices, this research aims to identify factors that facilitate or hinder AI acceptance and to

develop a conceptual framework that can guide future research, policy, and practice.

The research questions guiding this study were developed based on three key considerations (1) gaps identified in the existing literature, particularly the lack of qualitative exploration of therapists' perspectives; (2) theoretical constructs from the technology acceptance model (TAM) and sociotechnical systems (STS) theory, which emphasize perceived usefulness, perceived ease of use, ethical considerations, and human-technology alignment; and (3) consultation with two practicing therapists during pilot testing, who identified key areas of concern that were underrepresented in current research.

The following research questions guided this study:

a. What are therapists' perceptions of AI's potential benefits and limitations in counseling?

b. How do therapists view the ethical implications of integrating AI into counseling, particularly regarding trust, data privacy, and transparency?

c. What are therapists' concerns about AI systems' cultural sensitivity and adaptability in diverse counseling contexts?

d. How do therapists perceive the impact of AI on therapeutic dynamics, including empathy, rapport, and relational trust?

e. What are the practical applications of AI in counseling, and how do therapists evaluate its role in training and clinical practice?

By addressing these questions, this study seeks to provide a comprehensive understanding of how therapists view AI's role in counseling and to develop a theoretically grounded framework that integrates their perspectives with existing models of technology acceptance and organizational change.

2. Methods

2.1. Research Design. This study adopted a qualitative research design using semistructured interviews to explore therapists' attitudes and beliefs about integrating AI into counseling practice. A qualitative approach was chosen because it allows for an in-depth understanding of participants experiences, perceptions, and reasoning processes within a relatively new and complex area of inquiry.

Given the novelty of AI in therapeutic contexts, particularly the emergence of generative AI tools such as ChatGPT, there is little existing research that captures the lived experiences of therapists engaging with these technologies. Qualitative research is especially suitable for examining emerging phenomena where theoretical frameworks remain underdeveloped and where exploratory insights are needed to guide future studies [1]. While existing quantitative studies have primarily focused on AI's efficacy, usability, and user satisfaction, they provide a limited understanding of therapists' professional concerns, ethical reflections, and relational considerations. These dimensions require the kind of nuanced and contextual data that qualitative inquiry is uniquely positioned to reveal. The study was designed to generate, rather than test, hypotheses. Its aim was to develop new conceptual and theoretical insights grounded in the perspectives of practicing therapists. An exploratory and inductive orientation was therefore considered most appropriate to build theory from experience and interpretation rather than from preexisting assumptions.

Reflexive thematic analysis, as described by Braun and Clarke [1], was employed as the analytical method. This approach offers flexibility in identifying and interpreting patterns across data while recognizing the researcher's active role in meaning-making. It also enables attention to both explicit content and underlying meanings within participants' narratives. The method aligns with the study's goal of constructing a theoretically informed yet experience-grounded understanding of how therapists perceive and engage with AI in counseling.

2.2. Participants

2.2.1. Sampling Strategy and Rationale. This study employed purposive sampling to recruit 20 therapists, ensuring diversity in professional experience, cultural backgrounds, therapeutic settings, and exposure to AI applications in counseling. Purposive sampling was chosen to maximize variation and capture a range of perspectives relevant to the research questions [28]. In qualitative research, sample size adequacy is determined not by statistical power but by information richness, data saturation, and analytical depth [1, 29]. For reflexive thematic analysis, sample sizes typically range from 10 to 30 participants, with the specific number depending on the scope of the study, heterogeneity of the sample, and depth of analysis [30].

In this study, theoretical saturation was reached after 17 interviews, meaning that subsequent interviews (18–20) yielded no substantively new themes or insights, though they provided richer elaboration of existing themes. This aligns with established guidelines suggesting that saturation in interview studies typically occurs between 12 and 20 participants when the research questions are focused and the analytical approach is thorough [31, 32].

Furthermore, the goal of this study was not to achieve statistical generalizability but to provide a rich, contextual understanding of therapists' attitudes and to develop conceptual insights that can inform theory and practice. The sample size of 20 is appropriate for these objectives and is consistent with published qualitative studies in similar domains [10, 21]

2.2.2. Sample Characteristics. The sample comprised 12 women and eight men, with a mean age of 42.5 years (SD = 8.4), ranging from 30 to 60 years. Participants represented varied counseling environments, including clinical, school-based, and private practice settings, with professional experience spanning from 5 to 25 years (M = 15.8 years).

Theoretical orientations represented in the sample included CBT, person-centered therapy, psychodynamic

therapy, integrative approaches, and systemic family therapy. Participants worked with diverse client populations, including both adults and children/adolescents.

The sample included therapists from diverse cultural and geographical contexts including individuals of Middle Eastern, European, South Asian, and North American backgrounds. While the study did not conduct a comparative cultural analysis, this diversity ensured that perspectives on cultural sensitivity and adaptability reflected varied lived experiences. Ethnic composition included: eight participants identifying as White/European, six as Middle Eastern, four as South Asian, and two as mixed heritage. This diversity was intentionally sought to capture concerns about AI's cultural limitations from multiple standpoints.

Among the participants, 40%($n = 8$) identified having prior experience using AI-based tools in their therapeutic practice, such as chatbots or virtual agents for initial assessments, client progress tracking, or psychoeducation. The remaining 60% ($n = 12$), while not actively using AI in therapy, expressed familiarity with its potential applications through professional training, academic exposure, or informal exploration of tools like ChatGPT.

2.2.3. Rationale for Treating Samples as a Single Group. Given the heterogeneity of the sample, with participants varying in experience, setting, cultural background, and AI exposure, a deliberate decision was made to treat them as a single analytic group rather than conducting subgroup comparisons. This choice emerged from several interconnected considerations about the nature and purpose of the study. The primary aim was exploratory in nature, seeking to map the breadth of attitudes and concerns across the counseling profession as a whole, rather than testing for differences between specific subgroups. This broad focus appeared most appropriate given the nascent stage of AI integration in counseling and the limited prior qualitative research in this area. Preliminary analysis revealed something noteworthy: the major themes that emerged, such as viewing AI as a complementary tool, ethical concerns, cultural sensitivity issues, and questions about therapeutic dynamics, appeared consistently across all participant subgroups. While individual therapists emphasized different aspects based on their unique experiences, the core concerns appeared to be shared across the profession. This thematic coherence suggested common professional values and priorities that transcended individual differences in background or experience.

From a practical standpoint, splitting the sample into meaningful subgroups would have created analytical challenges. Breaking down 20 participants into categories such as high versus low AI exposure or grouping them by cultural background would have resulted in cell sizes too small to support robust comparative analysis. Such subdivision would have risked making interpretive claims based on insufficient data within each subgroup.

Perhaps most importantly, the methodological approach itself shaped this decision. In reflexive thematic analysis, heterogeneity is not viewed as a problem to be controlled or eliminated. Rather, it serves as an interpretive resource that enriches understanding. Diverse perspectives contribute to more nuanced thematic development and help ensure that findings reflect the complexity of real-world therapeutic practice. As Braun and Clarke [1] emphasize, the goal is not to eliminate variation but to understand and thoughtfully interpret it.

This approach does not mean the diversity within the sample was ignored. Throughout the Results section, contrasting perspectives are deliberately highlighted where they illuminate important variations within themes. By selecting illustrative quotes that demonstrate this range of viewpoints, the analysis honors the diversity of participants while maintaining analytical coherence. This approach captures both the shared professional concerns that unite therapists and the individual variations in emphasis and interpretation that reflect their unique positions and experiences.

Table 1 presents the demographic characteristics of the participants, highlighting their age, gender, years of experience, therapeutic setting, and exposure to AI in counseling practices.

2.3. Data Collection

2.3.1. Ethical Approval and Informed Consent. Before initiating the data collection process, ethical approval was obtained from the Ethical Committee of Ibn Haldun University, ensuring the study adhered to all ethical guidelines for research involving human participants and the principles outlined in the Declaration of Helsinki. Participation in the study was entirely voluntary, and written informed consent was obtained from all participants prior to their involvement. Confidentiality was assured, and participants were informed that their identities and personal details would remain anonymous and that data would be used solely for research purposes.

Detailed information was provided regarding how the collected data would be stored, used, and reported in future academic publications. Participants were told that there were no anticipated risks associated with participation and that their contribution would help advance understanding in this emerging area of research.

To preserve confidentiality, each participant was assigned a unique identifier (P1–P20), and no personal or identifying information was collected beyond basic demographic details. All audio recordings and interview transcripts were stored securely in password-protected digital files accessible only to the researcher. The data were analyzed and presented in aggregated form, ensuring that individual responses could not be traced back to any participant.

2.3.2. Interview Procedure. A total of 20 semistructured interviews were conducted to gather in-depth insights into therapists' attitudes and beliefs about the integration of AI in counseling. Of these, 14 interviews were conducted via Zoom, and six were conducted face-to-face based on the participants' preferences and comfort.

While research suggests that video-mediated and face-to-face interviews can produce comparable data quality in qualitative research [33], we acknowledge that communication modality may subtly influence rapport, nonverbal cue interpretation, and conversational depth. However, no systematic differences in response depth or thematic content

TABLE 1: Demographic characteristics of participants.

Participant ID	Gender	Age (years)	Years of experience	Setting	Theoretical orientation	Exposure to AI in practice
P1	Female	30	5	Clinical	CBT	Yes
P2	Female	35	8	Private	Person-centered	No
P3	Female	38	10	School	Integrative	Yes
P4	Female	40	12	Clinical	CBT	No
P5	Female	42	15	Private	Psychodynamic	No
P6	Female	45	17	Clinical	CBT	Yes
P7	Female	47	18	Private	Integrative	Yes
P8	Female	50	20	School	Person-centered	No
P9	Female	55	22	Clinical	Systemic	No
P10	Female	58	25	Private	CBT	Yes
P11	Female	60	7	Clinical	Psychodynamic	Yes
P12	Female	32	6	School	CBT	No
P13	Male	33	8	Private	Integrative	No
P14	Male	37	12	Clinical	CBT	Yes
P15	Male	41	16	Private	Person-centered	No
P16	Male	44	18	School	Systemic	Yes
P17	Male	48	21	Clinical	Psychodynamic	No
P18	Male	52	23	Private	CBT	Yes
P19	Male	56	19	Clinical	Integrative	No
P20	Male	59	24	Private	Person-centered	Yes

Note: AI Exposure indicates whether participants had prior experience using AI tools in their therapeutic practice.

were observed across interview modes during analysis. Both formats allowed for rich, reflective dialogue, and participants appeared equally comfortable sharing sensitive professional concerns. Future research with larger samples might explore this issue more systematically.

Before each interview, participants were provided with a detailed description of the study's aims and procedures. Rapport was intentionally built through informed conversation, acknowledgment of participants' expertise, and reassurance about confidentiality. Permission to audio-record the interviews was sought and obtained from all participants. Each interview lasted between 45 and 60 min, depending on the depth and elaboration of the participant's responses.

2.3.3. Interview Design and Pilot testing. Interview questions were designed based on three sources:

i. Gaps identified in the literature: As discussed in the Introduction, prior research has focused predominantly on technical capabilities and user satisfaction, with limited attention to therapists' ethical reasoning, cultural concerns, and relational priorities. Our questions were designed to address these underexplored dimensions.

ii. Theoretical frameworks: Questions were informed by STS (e.g., alignment between AI and professional values) and extended TAM (e.g., perceived usefulness, ethical risk, cultural adaptability, relational fit).

iii. Pilot testing with two practicing therapists: These individuals were not part of the final sample of 20 and were recruited separately to provide feedback on question clarity, relevance, and comprehensiveness. Both pilot participants were licensed mental health counselors with 10+ years of experience. Their feedback led to refinements in wording (e.g., avoiding jargon) and the addition of prompts related to cultural sensitivity, which they identified as underrepresented in our initial draft.

The interviews were guided by open-ended questions designed to elicit comprehensive responses while allowing flexibility for follow-up questions and probing. The interviewer (the author) maintained a neutral, nonleading stance, using reflective listening and open-ended prompts to encourage elaboration without imposing predetermined interpretations.

Key interview questions included (see Appendix A for the complete interview guide):

i. "What are your general thoughts on integrating AI in counseling?"

ii. "How do you perceive AI's role in enhancing or hindering the therapeutic process?"

iii. "What ethical concerns, if any, do you associate with using AI in therapy, particularly regarding privacy and confidentiality?"

iv. "How do you think cultural factors influence the effectiveness of AI tools in diverse counseling contexts?"

v. "What are the potential benefits and limitations of using AI tools such as chatbots or virtual agents in your practice?"

vi. "Have you had any direct experiences with AI-based tools in counseling, and if so, how would you describe these experiences?"

vii. "In what ways, if any, do you think AI could be used in counselor training or supervision?"

This approach ensured a structured yet flexible data collection process, allowing participants to express their opinions and share their experiences freely.

2.3.4. Data Recording and Transcription. All interviews were audio-recorded with participants' explicit consent. Recordings were transcribed verbatim by a professional transcription service, with transcripts subsequently reviewed by the author for accuracy. Each transcript was cross-checked against the original audio to ensure completeness and to capture nuances such as pauses, emphasis, and emotional tone where relevant. Transcripts were then anonymized and uploaded to NVivo 14 for systematic analysis.

2.4. Data Analysis

2.4.1. Analytical Approach. The data analysis was conducted systematically using NVivo, a qualitative data analysis software that facilitated the organization, coding, and interpretation of the interview material. The analysis followed the reflexive thematic analysis framework proposed by Braun and Clarke's [1], which guided the process through six iterative and interconnected phases. The researcher began by immersing herself in the data through repeated reading and re-reading of the transcripts to gain a deep familiarity with participants' narratives. From this immersion, initial codes were generated by systematically identifying meaningful segments within the text that captured relevant features of the data. These codes were then examined and grouped into broader patterns of meaning, forming the basis for potential themes.

In the next phase, themes were carefully reviewed and refined to ensure both internal coherence and clear distinction from one another. Each theme was then defined and named to capture its conceptual essence and analytical significance. Finally, the findings were written up, integrating the emergent themes with existing theoretical perspectives and relevant literature to produce a cohesive interpretation of the data.

This study adopted the reflexive form of thematic analysis as described by Braun and Clarke [1], which recognizes the active and interpretive role of the researcher in meaning-making. Unlike codebook thematic analysis, which emphasizes inter-rater reliability and treats themes as objectively discoverable, reflexive thematic analysis positions the researcher as a coconstructor of knowledge. In this approach, coding is seen not as a mechanical exercise but as an interpretive act that reflects both the researcher's theoretical lens and engagement with the data.

Consequently, traditional inter-rater reliability measures such as Cohen's kappa or Krippendorff's alpha were not employed, as they are not epistemologically compatible with the reflexive paradigm. Instead, analytical rigor was achieved through processes designed to enhance credibility and interpretive depth. Two researchers independently coded a subset of five transcripts and then engaged in in-depth discussions to compare interpretations, refine coding decisions, and strengthen conceptual clarity. The purpose of this collaboration was not to achieve numerical agreement but to enrich the analytical process through reflexive dialogue. When differences in interpretation emerged, they were resolved through discussion, allowing both coders to articulate their reasoning and reach a shared understanding.

The primary coder, who is an assistant professor in guidance and counseling with expertise in therapeutic practice and technology integration, conducted the main analysis. The secondary coder, a doctoral researcher in educational psychology trained in qualitative research, contributed to the credibility checks and reflexive discussions. Neither researcher participated in the pilot testing phase, ensuring an independent and critical engagement with the data. This collaborative yet interpretive approach strengthened the depth, transparency, and trustworthiness of the overall analysis.

2.4.2. Coding Process. The analysis began with repeated readings of the transcripts to achieve in-depth familiarity with participants' responses. During open coding, meaningful data segments were identified and labeled with initial codes such as "AI as a complementary tool," "ethical concerns," "cultural sensitivity," and "therapeutic dynamics." NVivo facilitated the organization, retrieval, and traceability of codes, enabling transparent analytical decisions. Next, codes were reviewed for patterns and grouped into subthemes. For instance, codes related to privacy, data security, algorithmic bias, and transparency were classified under the broader theme of "Ethical Concerns." Similarly, codes addressing AI's use in role-play, supervision, and client monitoring were grouped under "Practical Applications." Themes were then refined iteratively to ensure they were distinct, internally coherent, and aligned with the research objectives. To enhance confirmability, rich descriptions and verbatim participant quotes were incorporated, offering contextual depth to each theme.

2.4.3. Reflexivity and Audit Trail. Reflexivity formed an integral part of this study, guiding the researcher's continuous self-awareness and critical reflection throughout the research process. It involved a conscious examination of how personal background, professional identity, and underlying assumptions might have shaped both data collection and interpretation [1]. The author, a practicing counselor and educator, brought to the research a professional understanding of therapeutic processes and human interaction. However, she deliberately maintained an open and questioning stance, remaining receptive to insights that challenged her expectations. For instance, although initial assumptions suggested that participants might express strong resistance to the

integration of AI in counseling, the analysis instead revealed a more nuanced and conditional acceptance of AI—an insight that emerged organically through engagement with the data rather than from preconceived notions.

To ensure transparency and rigor, a detailed audit trail was maintained throughout the analytic process. This record documented the evolution of the study, including early impressions and memos written during the familiarization phase, the reasoning behind how specific codes were organized into broader themes, and reflections on decisions related to naming and defining thematic boundaries. It also captured moments of interpretive shift, illustrating how understanding deepened and transformed over time as the analysis progressed.

Maintaining this audit trail served not only as a methodological safeguard but also as a means of enhancing the study's credibility and dependability. By providing a clear record of the analytic process, it allows readers and future researchers to trace the reasoning that underpinned each interpretive decision. In qualitative inquiry, such documentation stands as evidence of systematic and rigorous analysis and contributes to the overall trustworthiness of the findings [34].

2.4.4. Ensuring Rigor. The analysis culminated in four central themes that encapsulated the participants' core experiences and perspectives: Perceptions of AI as a complementary tool, ethical concerns, cultural sensitivity and adaptability, and therapeutic dynamics. Each theme was carefully articulated to reflect its conceptual essence and supported by multiple sources of evidence drawn from the participants' narratives. Together, these themes portray a comprehensive picture of how therapists understand and negotiate the role of AI within counseling practice.

The use of a rigorous and reflexive analytical approach strengthened the credibility and trustworthiness of the findings. This process allowed for a nuanced interpretation that moved beyond surface-level attitudes, revealing the complex ways in which therapists balance professional ethics, cultural awareness, and therapeutic connection when considering the integration of AI into their work.

3. Results

The reflexive thematic analysis of interview data yielded four major themes that encapsulate therapists' attitudes and beliefs toward integrating AI in counseling practice. These themes directly address the five research questions posed in this study and illuminate the interplay between technological affordances and the sociocultural values embedded in therapeutic work. Table 2 synthesizes the thematic structure, providing an overview of major themes, constituent subthemes, representative quotations, frequency of expression across participants, and alignment with research questions. Frequencies indicate the approximate number of participants (out of 20) who articulated concerns or perspectives related to each subtheme, rather than total coded instances. This approach reflects the interpretive, meaning-centered focus of reflexive thematic analysis [1].

To enhance interpretive clarity, the themes are presented according to their predominant evaluative orientation toward AI. Themes framed as "positive" reflect areas in which therapists perceived AI as useful, supportive, or potentially beneficial within clearly bounded roles. Themes framed as "negative" capture domains where therapists expressed ethical, relational, or cultural concerns that constrain AI's appropriateness for therapeutic practice. Although participants' experiences were often nuanced rather than strictly dichotomous, this organizational clarification highlights the directional valence of therapists' perspectives while preserving the complexity of their accounts. In this study, Theme 1 reflects predominantly positive evaluations of AI, whereas Themes 2–4 represent predominantly negative or cautionary perspectives.

Table 2 presents the major themes, their corresponding subthemes, illustrative participant quotations, the frequency with which these themes emerged, and their relevance to the study's research questions.

3.1. Structural Organization of Findings. The results are presented according to major themes that directly correspond to the research questions. When differences appeared among participants in their previous experience with AI, cultural background, or therapeutic approach, these distinctions were emphasized through varied narratives and interpretive explanations. This method reflects both the shared professional principles that connect therapists and the individual viewpoints shaped by their specific experiences and contexts [29]

3.1.1. Theme 1: Perceptions of AI as a Complementary Tool. Across all participant subgroups, AI was conceptualized as an adjunctive resource rather than a substitutive technology. This framing reflects what we term bounded acceptance: a willingness to incorporate AI into peripheral or administrative domains of practice, coupled with firm resistance to its deployment in relationally intensive therapeutic contexts. Notably, this pattern held regardless of prior AI exposure, suggesting that professional socialization and therapeutic values may exert a stronger influence on attitudes than direct technological experience.

3.1.2. Efficiency and Accessibility Enhancement. Therapists identified operational efficiency and expanded service reach as AI's primary contributions. Participants with prior AI experience ($n = 8$) articulated these benefits with greater specificity, drawing on concrete use cases: "AI can handle routine tasks like scheduling or providing initial assessments, which gives us more time to focus on deeper therapeutic work." (P4, AI-experienced, clinical setting, CBT orientation).

Participants without direct AI exposure ($n = 12$) acknowledged similar affordances but tempered their optimism with concerns about quality-access trade-offs: "I can see how AI might help with accessibility issues, but I worry about quality being sacrificed for convenience." (P5, no AI experience, private practice, psychodynamic orientation).

TABLE 2: Themes, subthemes, and research question alignment.

Major theme	Subthemes/codes	Example statements	Codes frequency (approx.)	Related RQ	Theme definition
Perceptions of AI as a complementary tool	Efficiency and accessibility; AI in training; AI for underserved populations; administrative support	"AI can handle routine tasks like scheduling or providing initial assessments, which gives us more time to focus on deeper therapeutic work." – P4	28 instances across 16 participants	RQ1, RQ5	Therapists see AI as useful for supporting but not replacing human counselors, particularly in administrative, training, or accessibility contexts.
Ethical concerns	Privacy and data security; Transparency; algorithmic bias; client trust	"I worry about data breaches and how AI systems might misuse client information." – P7	31 instances across 18 participants	RQ2	Therapists express ethical reservations related to data use, transparency of AI processes, and potential embedded biases.
Cultural sensitivity and adaptability	Lack of diverse datasets; cultural misinterpretation; inclusivity; representation	"AI struggles to adapt to the unique cultural contexts of clients."—P9	24 instances across 15 participants	RQ3	Therapists highlight the lack of cultural competence in AI tools, emphasizing the need for inclusivity and diverse training data.
Therapeutic dynamics	Empathy limitations; Non-verbal cue absence; emotional intelligence; rapport and trust	"Empathy is at the heart of counseling, and I do not think AI can truly replicate that."—P1	35 instances across 19 participants	RQ4	Therapists believe AI cannot replicate the relational depth of therapy, particularly empathy, emotional nuance, and non-verbal understanding.

This divergence suggests that while both groups recognized AI's instrumental utility, those without hands-on experience maintained greater skepticism about potential unintended consequences—a pattern consistent with technology acceptance research indicating that perceived risk moderates perceived usefulness [35].

3.1.3. Training and Professional Development Applications. AI-assisted training emerged as a particularly promising application domain, garnering support across experience levels and theoretical orientations. Participants viewed simulation-based learning as a low-stakes environment for skill acquisition: "I have used AI-based simulations in training, and they are great for helping trainees practice basic skills before working with real clients." (P8, AI-experienced, school setting, Person-centered orientation). However, several participants emphasized the limitations of simulated practice, particularly regarding the development of contextual sensitivity and real-time clinical judgment. This distinction between technical skill development (where AI shows promise) and contextual competence development (where human supervision remains essential) aligns with situated learning theory's emphasis on authentic practice communities [36].

3.1.4. Boundaries of Complementarity. Universal across all participants was the conviction that AI cannot and should not replace core therapeutic functions. This unanimity is striking given the sample's heterogeneity and represents perhaps the study's most robust finding: "While AI is useful for certain tasks, it cannot replace the depth and nuance that human counselors bring to the therapeutic process." (P15, No AI experience, private practice, Person-centered orientation) "I see AI as a bridge rather than a substitute. It helps in making counseling services more scalable and accessible." (P2, AI-experienced, private practice, Person-centered orientation). The metaphor of AI as a "bridge" recurred across multiple interviews, suggesting a shared conceptual model wherein technology serves as an intermediary or facilitator rather than an autonomous agent. This framing has important implications for human-AI collaboration design, suggesting that systems perceived as "partners" rather than "replacements" may encounter less professional resistance.

In response to RQ1 and RQ5, participants articulated a nuanced acceptance of AI for administrative automation, initial screening, psychoeducation delivery, and skills training domains characterized by relatively structured, protocol-driven tasks. However, they categorically rejected AI's suitability for complex therapeutic interventions requiring relational depth, contextual adaptation, and emotional attunement. This pattern reflects what STS describes as task-technology fit [37]: AI is accepted where its capabilities align with task requirements but resisted where task complexity exceeds its affordances.

3.2. Theme 2: Ethical Concerns. Ethical concerns constituted the most consistently expressed source of resistance to AI integration, transcending all participant subgroups. These concerns were not abstract or hypothetical but grounded in therapists' professional obligations as defined by ethical codes, regulatory frameworks, and their own moral commitments to client welfare. The prominence of this theme underscores that technology acceptance in healthcare contexts cannot be reduced to usability or efficiency considerations alone; ethical alignment is a prerequisite for adoption.

3.2.1. Privacy and Data Security. Client data confidentiality emerged as the paramount ethical concern, with near-universal expression across the sample: "I worry about data breaches and how AI systems might misuse client information." (P7, AI-experienced, private practice, integrative orientation).

Intriguingly, participants with direct AI experience expressed heightened rather than diminished concern, having observed commercial AI platforms' data collection practices firsthand. This suggests that familiarity with AI systems' backend operations may actually amplify privacy anxieties, contradicting simple exposure-based models of technology acceptance.

The specificity of concerns varied by cultural context. Participants from regions with stringent data protection regimes (e.g., European backgrounds) referenced regulatory frameworks like GDPR, while those from contexts with weaker protections expressed anxiety about the absence of adequate oversight.

3.2.2. Transparency and Algorithmic Explainability. The opacity of AI decision-making processes, often termed the "black box" problem, troubled therapists across all settings: "There is a lack of transparency about how AI makes decisions. Clients have a right to understand how their data is being used." (P12, AI-experienced, school setting, CBT orientation).

Participants framed this as both an informed consent issue (clients cannot meaningfully consent to processes they cannot understand) and a clinical quality concern (therapists cannot responsibly implement interventions whose logic they cannot evaluate). This dual framing aligns with explainable AI (XAI) scholarship emphasizing that interpretability is not merely a technical desideratum but an ethical imperative in high-stakes domains [38].

3.2.3. Algorithmic Bias and Equity. Concerns about embedded biases were particularly pronounced among culturally diverse participants and those serving marginalized populations: "Even with the best intentions, biases can creep into AI algorithms, which could have serious consequences in counseling." (P18, no AI experience, private practice, CBT orientation, South Asian background)

Participants demonstrated awareness of research documenting algorithmic discrimination in healthcare contexts [39], expressing concern that AI-driven assessments or recommendations might systematically disadvantage already vulnerable groups. This concern extends beyond technical accuracy to questions of algorithmic justice: whether AI systems perpetuate or mitigate existing social inequities.

3.2.4. Client Trust and Therapeutic Alliance. "Clients are often hesitant to trust AI with their personal information, and I think this skepticism is valid." (P3, AI-experienced, school setting, integrative orientation)

Several participants noted that clients' own reservations about AI represented a practical barrier to implementation, regardless of therapists' views. This observation highlights the importance of multistakeholder perspectives in technology evaluation; acceptance by practitioners is necessary but insufficient if service users resist engagement.

In response to RQ2, participants identified data privacy vulnerabilities, algorithmic opacity, embedded biases, and trust erosion as critical ethical barriers to AI integration. Importantly, these concerns were not framed as technological problems amenable to engineering solutions alone but as challenges requiring robust governance frameworks, regulatory oversight, and ongoing ethical deliberation. The findings suggest that AI acceptance among therapists is fundamentally conditional on demonstrated ethical trustworthiness—a threshold that current systems have not yet met in participants' assessments.

3.3. Theme 4: Therapeutic Dynamics. Cultural limitations of AI systems emerged as a particularly salient concern, especially among participants from non-Western backgrounds ($n = 10$ of 10 culturally diverse participants) and those working in multicultural settings ($n = 14$). This theme illuminates how AI development practices often centered in Western, predominantly English-speaking contexts create systems that inadequately serve global and culturally diverse populations. The consistency of this concern across different participant groups suggests that cultural responsiveness is increasingly recognized as a core competence in therapeutic practice, one that current AI tools conspicuously lack.

3.3.1. Homogeneous Training Data and Representation Gaps. Participants across backgrounds identified Western-centric biases in AI development as a fundamental limitation: "Most AI systems are developed with Western data, which might not be effective for clients from different cultural backgrounds." (P14, no AI experience, clinical setting, CBT orientation, Middle Eastern background).

These observations align with critical AI scholarship documenting how machine learning systems trained predominantly on WEIRD (Western, Educated, Industrialized, Rich, Democratic) populations fail to generalize across cultural contexts [40]. Participants articulated this not as a minor technical deficiency but as a barrier to culturally competent practice—a professional standard that has become increasingly central to therapeutic training and ethics [41].

3.3.2. Risks of Cultural Misinterpretation and Harm. Beyond mere ineffectiveness, participants described potential harms arising from cultural mismatches: "There is a risk of cultural misinterpretation, which could harm the therapeutic relationship." (P5, no AI experience, private practice, psychodynamic orientation, European background) "Clients from minority groups might not feel understood if AI fails to address their cultural values and beliefs." (P11, AI-experienced, clinical setting, psychodynamic orientation, mixed heritage).

These concerns reflect an understanding that culturally inappropriate interventions are not merely unhelpful but actively harmful, potentially pathologizing normative cultural practices or invalidating clients' lived experiences. This framing positions cultural sensitivity not as an enhancement

to AI systems but as a baseline requirement for safe deployment.

3.3.3. Calls for Inclusive Development Practices. Rather than rejecting AI wholesale, most participants advocated for more inclusive development approaches: "For AI to work in multicultural settings, it needs to be trained on diverse datasets that reflect real-world diversity." (P16, AI-experienced, school setting, systemic orientation). This solution-oriented stance suggests that cultural concerns represent a barrier to acceptance rather than an insurmountable obstacle. Participants implied that AI systems developed through participatory, culturally informed design processes might overcome current limitations—an insight consistent with value-sensitive design approaches emphasizing stakeholder involvement [42].

In response to RQ3, therapists identified cultural insensitivity as both a technical limitation (inadequate training data diversity) and a potential source of iatrogenic harm (misinterpretation and cultural invalidation). This concern was more pronounced though not exclusive to participants from non-Western backgrounds, suggesting that positionality shapes the salience of cultural considerations. The findings underscore that AI systems cannot be culturally "neutral"; they embody the values, norms, and knowledge systems of their developers and training data. For therapeutic AI to achieve global applicability, development practices must intentionally center cultural diversity and epistemic pluralism.

3.4. Therapeutic Dynamics. Skepticism regarding AI's capacity to replicate the relational foundations of therapy was universal (20/20 participants), representing the study's most consistent finding. However, the articulation of this skepticism varied across therapeutic orientations, with person-centered and psychodynamic practitioners emphasizing relational authenticity most emphatically, while CBT-oriented therapists acknowledged greater latitude for AI in structured intervention delivery. This variation suggests that therapeutic modality influences perceptions of where the "nonnegotiable" human elements of therapy lie.

3.4.1. Empathy and Affective Attunement. Participants across all orientations questioned AI's capacity for genuine empathy, distinguishing between simulated empathic responses and authentic empathic presence: "Empathy is at the heart of counseling, and I don't think AI can truly replicate that." (P1, no AI experience, clinical setting, CBT orientation). Person-centered and psychodynamic therapists articulated this concern most forcefully, reflecting their theoretical emphasis on the therapeutic relationship as the primary mechanism of change.

This skepticism reflects a common-sense understanding of empathy as involving subjective emotional experience—a phenomenological quality that participants viewed as fundamentally irreplicable by computational systems. However, this framing also reveals a potential conceptual limitation: participants appeared to operate with a unitary conception of empathy, overlooking distinctions that have become central in affective computing research.

Contemporary scholarship distinguishes between cognitive empathy (understanding others' mental states), affective empathy (sharing others' emotions), and empathic concern (motivation to alleviate others' distress) [43]. AI systems may plausibly approximate cognitive empathy through sentiment analysis and theory-of-mind modeling, even if they lack phenomenal consciousness or moral agency. Participants' categorical dismissal of AI empathy may thus reflect conflation of these distinct constructs.

3.4.2. Nonverbal Communication and Embodied Presence. The absence of nonverbal attunement capabilities was highlighted as a critical limitation: "The therapeutic relationship relies on subtle, nonverbal cues, which AI simply cannot pick up on." (P20, no AI experience, private practice, Person-centered orientation)

Participants emphasized that much therapeutic communication occurs through paralinguistic features (tone, pacing, and silence) and embodied expressions (posture, gesture, and gaze) that text-based AI systems cannot access. While multimodal AI architectures incorporating video and audio analysis are under development, participants demonstrated limited awareness of these advances, suggesting a knowledge gap regarding current AI capabilities.

3.4.3. Emotional Depth and Dynamic Responsiveness. Participants distinguished between scripted responsiveness (pattern-matching based on training data) and authentic emotional engagement: "AI can simulate certain responses, but it does not have the emotional depth needed to understand clients truly." (P17, AI-experienced, clinical setting, psychodynamic orientation) This distinction points to concerns about AI's capacity for contextual sensitivity, the ability to recognize when standard protocols should be abandoned in favor of client-specific, moment-to-moment improvisation. Participants viewed this flexibility as essential to competent practice, particularly when addressing crisis situations, complex trauma, or rapid emotional shifts.

In response to RQ4, participants unanimously viewed AI as incapable of replicating the empathic, adaptive, and nonverbally attuned presence essential to therapeutic alliance. This was the most consistent finding across all participant subgroups, suggesting that relational authenticity represents a "red line" in therapists' conceptualization of acceptable AI roles. However, the framing of this limitation in absolute terms AI cannot be empathic may reflect both legitimate concerns about current technological capabilities and potentially overgeneralized assumptions about AI's fundamental constraints. The Discussion section engages more critically with this tension.

3.5. Cross-Cutting Observations and Sample Heterogeneity. The strong thematic convergence across diverse participants suggests that professional socialization—the process through which individuals internalize occupational norms and values—powerfully influences technology attitudes. Therapeutic training, regardless of modality, emphasizes

relational skills, ethical responsibility, and cultural humility. These commitments appear to create a shared evaluative framework structuring how therapists assess AI, even among practitioners with divergent backgrounds. This has implications for technology implementation: AI systems aligning with core professional values encounter less resistance than those threatening them, regardless of technical sophistication landscape

4. Discussion

This study examined therapists' attitudes and beliefs toward the integration of AI into counseling practices, revealing a nuanced landscape of conditional acceptance shaped by professional values, ethical commitments, and relational priorities. Rather than encountering uniform resistance or uncritical enthusiasm, we found that therapists engage in sophisticated evaluative reasoning, welcoming AI where it demonstrably serves therapeutic goals without compromising core values while resisting deployment in domains requiring relational depth, cultural attunement, or complex ethical judgment.

The four identified, perceptions of AI as a complementary tool, ethical concerns, cultural sensitivity and adaptability, and therapeutic dynamics converged remarkably across a heterogeneous sample, suggesting that certain professional commitments may be constitutive of therapeutic identity regardless of individual training or context. This section situates these findings within contemporary scholarship on human-AI collaboration, therapeutic alliance theory, and the ethical governance of healthcare AI. We also develop a conceptual framework synthesizing our results with existing TAMs, propose design principles for culturally responsive therapeutic AI, and critically examine assumptions embedded in both participants' views and current AI discourse.

4.1. AI as Complementary Infrastructure: Reconceptualizing the Human-Technology Relationship

4.1.1. Beyond Replacement Anxieties. Participants' framing of AI as a "bridge" or "complement" rather than a substitute reflects what we term bounded acceptance: a conditional openness to AI integration within carefully circumscribed domains. This finding challenges simplistic narratives that position healthcare professionals as either resistant technophobes or enthusiastic early adopters [44]. Instead, therapists demonstrated evaluative sophistication, distinguishing between tasks where AI's capabilities align with therapeutic needs (administrative automation, initial screening, and psychoeducation) and those where its limitations preclude responsible use (complex case formulation, relational attunement, and crisis intervention).

4.1.2. Training and Skill Development: A promising Application Domain. AI's role in training emerged as the least contested application, garnering support across all participant subgroups. This enthusiasm aligns with growing evidence that simulation-based learning environments enhance skill acquisition in healthcare contexts [45]. Recent studies demonstrate that AI-powered role-play simulations enable counselors-in-training to practice therapeutic responses in low-stakes environments, receiving immediate feedback on verbal and nonverbal communication patterns [23, 24].

However, participants also articulated an important limitation: simulations cannot fully replicate the contextual complexity, unpredictability, and emotional intensity of actual therapeutic encounters. This observation resonates with situated learning theory's emphasis on authentic practice communities [36] and suggests that AI training tools are most appropriately positioned as supplements to rather than replacements for supervised clinical experience.

4.1.3. Accessibility and the Ethics of Scalability. Participants acknowledged AI's potential to expand mental health service reach, particularly for geographically isolated or resource-constrained populations. This benefit aligns with documented evidence that AI chatbots like Woebot and Wysa increase access to evidence-based interventions for individuals who might not otherwise seek or afford professional care [4, 5]

Yet accessibility gains raise ethical tensions that participants implicitly recognized: Does providing AI-mediated care to underserved populations constitute equitable access, or does it create a two-tiered system where privileged individuals receive human care whereas marginalized groups receive "good enough" technological substitutes? This concern echoes broader debates about technological solutionism in social policy [46], the tendency to address structural inequities through technical interventions rather than resource redistribution.

4.2. Ethical Governance as a Prerequisite for AI Acceptance

4.2.1. The Trust Deficit: Privacy, Transparency, and Professional Responsibility. Ethical concerns constituted the most consistently expressed barrier to AI acceptance, transcending all participant subgroups. This finding underscores that in healthcare contexts, technology adoption cannot be reduced to usability or efficiency considerations; ethical trustworthiness is a nonnegotiable prerequisite [47].

Participants' anxieties about data privacy, algorithmic opacity, and embedded biases were not abstract philosophical concerns but grounded in concrete professional obligations. Therapeutic ethics codes mandate absolute confidentiality protection, informed consent for all interventions, and competent, nondiscriminatory practice [48]. AI systems that cannot guarantee these standards violate foundational professional commitments, regardless of their technical sophistication or potential benefits.

Intriguingly, participants with direct AI experience expressed heightened rather than diminished ethical concerns. This pattern contradicts simple exposure-based models of technology acceptance and suggests that familiarity with AI systems' backend operations, data collection practices, algorithmic decision processes, security vulnerabilities may amplify rather than allay privacy anxieties. This finding has important implications for implementation strategies: transparency about AI system operations, rather than opacity justified by proprietary concerns, may be essential for building professional trust.

4.2.2. The Black Box Problem: Explainability as an Ethical Imperative. The opacity of AI decision-making, what participants described as a "lack of transparency" emerged as a critical concern. This aligns with the explainable AI (XAI) movement's emphasis that interpretability is not merely a technical desideratum but an ethical requirement in high-stakes domains [38, 49].

Participants framed algorithmic opacity as both an informed consent violation (clients cannot meaningfully consent to processes they cannot understand) and a clinical quality concern (therapists cannot responsibly implement interventions whose logic they cannot evaluate). This dual framing is particularly important: it positions explainability not as a concession to technologically unsophisticated users but as essential for professional accountability.

Recent developments in mechanistic interpretability [50] offer some promise for understanding neural network decision processes, but these methods remain nascent. Until AI systems can provide clinically meaningful explanations for their recommendations—not merely statistical correlations but causal reasoning grounded in therapeutic theory—they are unlikely to gain full professional acceptance.

4.2.3. Algorithmic Bias and the Imperative of Justice-Oriented Design. Concerns about embedded biases were particularly pronounced among culturally diverse participants and those serving marginalized populations. This observation aligns with growing evidence that AI systems trained on nonrepresentative datasets systematically disadvantage already vulnerable groups [39, 51]

In mental health contexts, algorithmic bias can manifest in multiple ways: misdiagnosis of culturally normative behaviors as pathological, provision of culturally inappropriate coping strategies, failure to recognize culture-specific idioms of distress, or differential accuracy across demographic groups. These failures are not merely technical limitations but are actively harmful, potentially exacerbating health disparities rather than ameliorating them.

Participants' emphasis on this concern suggests awareness of these risks and underscores the importance of justice-oriented AI design development practices that center equity considerations from inception rather than treating them as post hoc corrections [52]. This includes (1) intentional diversification of training datasets across cultural, linguistic, and demographic dimensions; (2) participatory design involving diverse stakeholders; (3) algorithmic auditing for disparate impact; and (4) ongoing monitoring for bias after deployment.

Importantly, addressing bias is not simply a matter of data diversification. It requires grappling with deeper questions about what constitutes "good" therapeutic outcomes and whether these definitions encode culturally specific values. For instance, Western therapeutic traditions often emphasize individual autonomy and emotional expressiveness—values that may conflict with collectivist or stoic cultural frameworks. AI systems trained to optimize Western-defined outcomes may inadvertently pathologize alternative value systems.

4.3. Cultural Sensitivity: From Technical Problem to Epistemological Challenges

4.3.1. Beyond Dataset Diversification. Participants identified cultural insensitivity as a critical limitation, emphasizing that most AI systems are trained on Western datasets and embody Western therapeutic assumptions. While accurate, this framing risks treating cultural responsiveness as a purely technical problem solvable through data diversification, adding more non-Western training examples until the model "learns" cultural variations.

This approach, while necessary, is insufficient. Cultural competence in therapy is not merely knowledge of different cultural practices but involves epistemic humility recognizing that one's own cultural framework is partial and that alternative frameworks may be equally or more valid [53]. It requires ongoing negotiation, mistake acknowledgment, and relationship repair processes fundamentally incompatible with algorithmic pattern matching.

Moreover, cultural "difference" is not a static feature that can be encoded and retrieved but emerges through interaction. What counts as culturally appropriate therapeutic response depends on the specific client, their unique positioning relative to various cultural identities, the therapeutic relationship's history, and the immediate situational context. AI systems, which operate through predefined categorical logics, struggle with this kind of emergent, context-dependent meaning-making.

4.3.2. Toward Culturally Humble AI Design. Rather than pursuing the likely impossible goal of culturally omniscient AI systems, designers might instead aim for culturally humble AI systems that explicitly acknowledge their limitations, defer to human judgment on culturally complex matters, and facilitate rather than replace culturally responsive practice.

This could involve:

1. Transparent cultural positioning: AI systems explicitly stating what cultural contexts they were trained on and where their competence is limited.

2. Cultural checkpoints: Systems that flag potential cultural misunderstandings and prompt human oversight rather than proceeding autonomously.

3. Client-driven customization: Interfaces allowing clients to adjust AI responses based on their cultural preferences and values.

4. Participatory development: Engaging diverse communities not merely as data sources but as codesigners with authority over how their cultures are represented.

This approach acknowledges that cultural responsiveness is not a technical feature to be implemented but an ongoing ethical commitment requiring human judgment, relationship, and accountability—precisely the qualities participants emphasized as irreplaceable.

4.3.3. Postcolonial Perspectives on Therapeutic AI. Participants' concerns about Western-centric AI development invite engagement with postcolonial AI critique [19, 54]. This scholarship examines how AI systems often reproduce colonial patterns of knowledge extraction, epistemic violence, and cultural homogenization, positioning Western knowledge as universal while marginalizing indigenous and non-Western epistemologies.

Truly decolonial therapeutic AI would not merely include diverse data but would pluralize its foundational assumptions about what constitutes mental health, therapeutic success, and appropriate intervention. This might mean, for instance, AI systems that recommend community-based or spiritually grounded healing modalities rather than assuming that Western psychotherapy represents the gold standard. Such radical pluralism remains largely absent from current therapeutic AI development.

4.4. Rethinking Empathy: Beyond the Binary of Presence and Absence

4.4.1. The Categorical Dismissal and Its Limitations. Participants' unanimous skepticism about AI's empathic capabilities represents this study's most robust finding and also its most conceptually fraught. Therapists operated with an implicit model of empathy as a unitary, phenomenologically grounded capacity requiring conscious emotional experience. From this perspective, AI systems lacking subjective experience and phenomenal consciousness are categorically incapable of empathy, regardless of their behavioral sophistication.

This framing, while intuitively compelling, may not fully account for the conceptual complexity of empathy or the empirical capabilities of contemporary AI systems. Cognitive science distinguishes between multiple empathy constructs: cognitive empathy (understanding others' mental states), affective empathy (sharing others' emotions), and empathic concern (motivation to alleviate others' distress) [43]. These components are dissociable both neurologically and functionally, raising the question: Must all be present for therapeutic benefit?

Recent research suggests that AI systems can approximate cognitive empathy through advanced NLP, sentiment analysis, and theory-of-mind modeling [55, 56]. Large language models (LLMs) demonstrate the capacity to infer emotional states from text, generate contextually appropriate supportive responses, and adapt interaction patterns based on inferred user needs. While these capabilities do not involve the phenomenal experience of emotion, they may nonetheless provide functionally equivalent support for certain therapeutic purposes.

4.4.2. Simulated Empathy and Authentic Presence: A False Dichotomy. Participants distinguished sharply between "simulated" AI responses and "authentic" human empathy. This binary framing assumes that authenticity requires genuine emotional experience that supportive responses are valuable only if they arise from actual felt empathy rather than computational pattern-matching.

However, therapeutic alliance research complicates this assumption. Studies indicate that clients' perception of being understood and cared for predicts therapeutic outcomes more strongly than therapists' actual internal states [57]. This suggests that what matters therapeutically is the relational effect of empathic communication, not necessarily its phenomenological origins.

This observation does not erase important differences between human and AI empathy. Human empathy, even when professionally performed, emerges from embodied social experience, cultural learning, and potential for genuine emotional resonance capacities AI systems currently lack. But it does suggest that the boundary between "authentic" and "simulated" empathy may be less absolute than participants assumed.

4.4.3. Multimodal Affective Computing: Expanding AI's Empathic Repertoire. Participants emphasized AI's inability to interpret nonverbal cues, body language, facial expressions, and tone—as a fundamental limitation. This assessment accurately describes text-based chatbots but does not reflect the state of the art in affective computing.

Multimodal AI systems now integrate video, audio, and physiological signals to infer emotional states with considerable accuracy [58]. Systems can detect subtle facial micro expressions, analyze vocal prosody, and even estimate autonomic arousal from peripheral sensors. While these capabilities remain imperfect and raise additional privacy concerns, they demonstrate that AI's sensory limitations are not permanent but actively being addressed through technological development.

4.4.4. Towards a Nuanced Framework: Empathic Sufficiency vs. Equivalence. Rather than asking whether AI "has empathy" a binary, ontologically charged question that is unlikely to yield a satisfactory resolution, it is more productive to consider a pragmatic alternative: for which therapeutic purposes can AI demonstrate empathic sufficiency? In other words, when does AI offer enough affective attunement to provide meaningful therapeutic benefit, even if it falls short of the empathic depth associated with human practitioners?

Therapists in this study consistently described AI as sufficiently empathic for tasks such as delivering psychoeducation, offering basic emotional validation (e.g., acknowledging when a client feels overwhelmed), providing information about crisis resources, conducting routine check-ins or progress monitoring, and supporting structured skill-building exercises like CBT thought records. These functions were viewed as appropriate because they rely on predictable emotional cues and standardized responses, allowing AI to perform competently within its current computational boundaries.

In contrast, participants viewed AI as emphatically insufficient for work requiring profound emotional nuance or relational complexity. This included processing complex trauma, navigating client ambivalence or resistance, engaging in culturally attuned interactions within diverse contexts, managing rupture and repair in therapeutic relationships, and responding to crises that demand sensitive, context-dependent judgment. Such tasks were perceived as

fundamentally reliant on human intuition, embodied presence, and the capacity for deep attunement—qualities not adequately replicable through current AI systems. This reframing acknowledges AI's present limitations while remaining open to the possibility that empathic capacities may expand as technology develops. It shifts attention away from abstract debates about whether AI truly "possesses" empathy and instead emphasizes a functional, context-sensitive evaluation of the therapeutic roles AI can competently and safely fulfill.

4.5. A Conceptual Framework for Conditional AI Acceptance in Counseling

4.5.1. Extending the Technological Acceptance Model. The TAM traditionally asserts that technology adoption is shaped primarily by users' perceptions of usefulness and ease of use [59]. However, in therapeutic contexts, our findings indicate that acceptance depends on several additional dimensions that extend beyond TAM's original scope. Therapists emphasized the importance of perceived ethical trustworthiness, referring to their confidence that AI systems can safeguard client privacy, maintain data security, operate transparently, and avoid harmful algorithmic biases. They also highlighted perceived cultural responsiveness, which reflects beliefs about whether AI can recognize, respect, and adapt to diverse cultural identities, values, and communication norms. A third key factor is perceived relational sufficiency, or the degree to which therapists believe AI is capable of providing enough empathic attunement, emotional understanding, and therapeutic presence to be appropriate for particular intervention tasks. These dimensions operate not as isolated considerations but as dynamically interrelated judgments. For example, ethical reservations may be mitigated when therapists perceive AI as culturally responsive, particularly if they view it as more equitable than human practitioners who may carry their own biases. Conversely, even highly useful technologies may be rejected if they fall short on relational sufficiency in situations where therapeutic alliance and emotional depth are essential.

4.5.2. Sociotechnical Alignment and Professional Values. STS argues that organizational effectiveness depends on the alignment between technical and social subsystems [60]. In the context of AI in counseling, our findings reveal a significant misalignment between these domains. The technical subsystem consists of AI tools developed primarily to maximize efficiency, scalability, and cost reduction values consistent with broader healthcare economics but often misaligned with therapeutic ethics. By contrast, the social subsystem reflects therapists' professional commitments to relational authenticity, cultural humility, ethical responsibility, and the primacy of client welfare values that do not lend themselves easily to standardization or optimization. According to STS, successful integration requires "joint optimization," in which technological innovations enhance efficiency without undermining the human-centered values that define professional practice. Achieving such alignment would entail designing AI systems that embed ethical considerations directly into their architectures, allowing therapists to remain central decision-makers through human-in-the-loop structures and ensuring that transparency and explainability are foundational features rather than afterthoughts. It also requires participatory development processes in which therapists act as active codesigners empowered to challenge or reject system features that conflict with professional standards or compromise therapeutic integrity.

4.5.3. The Conditional Nature of AI Acceptance. Bringing together these insights, we argue that therapists' willingness to accept AI in counseling is fundamentally conditional, depending on whether AI systems can demonstrate alignment with four core professional commitments. First, AI must support beneficence, providing genuine benefit to clients in ways that go beyond symptom reduction to include the promotion of well-being, autonomy, and dignity. Second, it must uphold justice, contributing to equitable access and outcomes rather than reinforcing or deepening existing disparities across cultural, socioeconomic, or demographic groups. Third, it must respect persons, honoring clients' autonomy, cultural identities, and their right to meaningful human connection while avoiding reductive or dehumanizing forms of interaction. Finally, AI must remain faithful to relationship, enhancing rather than undermining the therapeutic alliance and serving to support, rather than substitute for, the relational work of human practitioners. Therapists in this study expressed openness to AI when these conditions appeared to be met; conversely, resistance emerged not from technophobia but from ethical vigilance whenever such alignment was absent or uncertain.

4.6. Design Principles for Culturally Responsive, Ethically Trustworthy Therapeutic AI. Drawing on our findings and the conceptual framework developed above, we propose a set of guiding principles for the future development of AI in counseling. First, AI systems should embody epistemic humility by clearly acknowledging their limits, particularly in areas such as cultural competence and relational depth. Rather than appearing as all-knowing assistants, they should transparently communicate the cultural contexts in which they were trained and the areas where their understanding may be insufficient. Second, human oversight and the capacity to override AI recommendations must remain central; decisions related to diagnosis, crisis assessment, or intervention planning should always require human judgment, with AI serving strictly as a decision-support tool rather than an autonomous actor. Third, AI tools should be built on privacy-preserving architectures, employing strategies such as data minimization, federated learning, and differential privacy, while giving clients meaningful control over what information is collected and how it is used. Fourth, systems must undergo regular algorithmic auditing and bias monitoring to identify disparate impacts across demographic groups, with audit results made publicly accessible and with the expectation that biased systems be retrained or, when necessary, withdrawn from use rather than superficially adjusted. Fifth, development processes should embrace participatory and justice-oriented design, involving therapists,

clients, and culturally diverse communities as genuine codesigners with real decision-making authority, rather than as symbolic consultants. Finally, AI should aim for relational augmentation rather than replacement, supporting therapists by freeing time for deeper engagement, enriching cultural understanding, and identifying clinically relevant patterns, rather than attempting to simulate or substitute the therapeutic relationship itself.

5. Limitations

5.1. Sample Composition and Generalizability. Although the sample was intentionally heterogeneous, it consisted exclusively of practicing therapists. This composition introduces a potential source of bias, as therapists may harbor concerns about professional displacement that intensify their skepticism toward AI, independent of the technology's actual capabilities. Their perspectives may therefore overemphasize risks related to relational disruption or loss of professional identity. Incorporating clients as participants could offer a crucial comparative perspective, particularly in light of emerging findings suggesting that clients sometimes prefer AI-generated responses when they are unaware of the source [61]. Future research would benefit from multistakeholder designs that bring together therapists, clients, and AI developers to evaluate the same technological tools. Such triangulation would illuminate where professional concerns align with or diverge from the lived experiences and expectations of those receiving or creating AI-supported interventions.

5.2. The Limits of Self-Report. This study relies on interview data that capture participants' stated beliefs and attitudes, which do not always correspond to actual behavior. Research in implementation science indicates that even when individuals express skepticism about a technology, such skepticism does not necessarily prevent adoption, especially when use becomes organizationally mandated or when external incentives make adoption advantageous [62]. As a result, it remains unclear whether therapists' concerns would translate into resistance in real-world implementation contexts. Future research could incorporate observational methods, implementation trials, or experimental vignette designs to assess whether the reservations therapists articulate are borne out in practice, or whether they adapt their perspectives as AI becomes increasingly embedded in clinical workflows.

5.3. Technological Dynamism and Temporal Specificity. The findings reflect attitudes toward AI as it existed in late 2024 and early 2025. Given the rapid rate of technological progress such as improvements in multimodal processing, context-aware reasoning, and culturally adaptive fine-tuning, some limitations identified by participants may become less applicable as capabilities evolve. Because therapist attitudes are likely to shift in response to more advanced and reliable systems, future research should examine how perceptions change over time and whether initial skepticism diminishes when new forms of AI address shortcomings that were evident at the time of this study.

5.4. Free vs. Premium AI Model Differences. Although participants discussed AI tools broadly, the study did not differentiate between free and premium versions of LLMs. Therapists' perceptions may vary substantially depending on which tier they have experience with, as free models often have more limited capabilities in areas such as coherence, cultural responsiveness, and relational attunement. Consequently, some negative evaluations may reflect the constraints of lower-capability systems rather than inherent limitations of therapeutic AI. Future research should explicitly compare therapist impressions across model tiers to better understand how capability differences shape professional acceptance.

5.5. Alternative Interpretations of Empathy Skepticism. The study interpreted participants' skepticism about AI empathy as reflecting both an accurate recognition of current technological limitations and, at times, an overgeneralization of these constraints. However, an alternative interpretation is that therapists were expressing a deeper ontological position: that empathy inherently requires phenomenal consciousness, subjective experience, or moral agency, and therefore cannot be meaningfully attributed to computational systems regardless of future advances. From this perspective, no degree of behavioral sophistication or affective simulation would qualify as genuine empathy. If this interpretation holds, debates about AI empathy may not be resolvable through empirical demonstration of functional equivalence but may instead reflect fundamentally different philosophical commitments regarding the nature of mind and relational presence. In this view, the eventual acceptance of AI in counseling may depend not only on technological improvement but also on conceptual negotiation about which qualities are essential to therapeutic work and which can be augmented or replicated by artificial agents.

6. Conclusion

This study explored therapists' attitudes and beliefs toward integrating AI in counseling, offering valuable insights into the opportunities and challenges of this rapidly evolving technology. The findings highlight a nuanced perspective: while therapists recognize AI's potential to enhance accessibility, streamline routine tasks, and support counselor training, they remain cautious about its limitations in addressing the relational and ethical complexities of counseling practice.

Therapists emphasized the complementary role of AI in routine functions, such as scheduling, initial assessments, and psychoeducation, which could help bridge gaps in mental health care, particularly for underserved populations. However, the study also underscored significant concerns regarding AI's capacity to replicate human qualities essential to counseling, such as empathy, trust, and cultural sensitivity. Ethical issues, including data privacy, algorithmic bias, and the transparency of AI systems, were identified as critical barriers to integration, reflecting a need for robust regulatory frameworks and privacy-preserving technologies.

Cultural adaptability emerged as another crucial challenge, with therapists pointing out the limitations of current

AI systems in addressing diverse cultural contexts. These findings suggest that for AI to be effectively integrated into counseling, its development must prioritize inclusivity, representation, and cultural responsiveness. Moreover, the therapeutic alliance, characterized by nonverbal communication and emotional intelligence, remains an irreplaceable component of effective counseling—one that AI systems currently cannot fully replicate.

Ultimately, this research highlights the importance of positioning AI as a supportive tool that complements, rather than replaces, human counselors. By addressing the ethical, cultural, and relational challenges identified, AI has the potential to enhance mental health care delivery without compromising the foundational principles of counseling. Future research should focus on clients' perspectives, the longitudinal effects of AI integration, and the development of AI systems that foster deeper relational dynamics. Through these efforts, the counseling field can embrace the opportunities presented by AI while maintaining its commitment to empathy, trust, and human connection.

7. Implications

The findings of this study indicate that therapists approach AI not as uncritical adopters or resistant technophobes but as discerning professionals who evaluate emerging technologies through the lens of core therapeutic values. Their conditional acceptance embracing AI where it expands access, enhances efficiency, or supports structured tasks yet resisting it where it risks undermining ethical or relational foundations demonstrates that successful integration depends on sociotechnical alignment rather than technological advancement alone.

These results challenge both technological determinism, which assumes AI will transform therapy regardless of professional judgment, and therapeutic exceptionalism, which presumes that counseling is insulated from technological influence. Instead, the evidence supports a collaborative model of human–AI interaction, one that recognizes the respective strengths and limitations of practitioners and machines and respects the diverse needs and preferences of clients.

For AI developers, the implications are straightforward: technologies that prioritize efficiency and scalability while neglecting privacy protections, cultural responsiveness, or relational authenticity are likely to face enduring professional resistance. Systems designed with epistemic humility, clear operational transparency, and a focus on augmenting rather than replacing human judgment are far more likely to be viewed as ethically and clinically acceptable.

For therapists, the challenge lies in maintaining ethical vigilance while remaining open to the potential benefits of AI-assisted care. This requires ongoing engagement with technological developments and a willingness to critically reassess assumptions as AI capabilities evolve, rather than dismissing emerging tools based solely on their current limitations.

For policymakers and healthcare leaders, the findings highlight the necessity of governance frameworks that protect client welfare and uphold professional ethics. Effective oversight will require robust regulatory structures, mechanisms for algorithmic auditing, and participatory design processes that include therapists, clients, and marginalized communities as meaningful contributors.

Ultimately, the incorporation of AI into counseling will unfold through negotiation rather than inevitability. Its trajectory will depend on sustained dialogue among technologists, practitioners, clients, ethicists, and policymakers about what forms of technologically supported therapy are ethically permissible, clinically effective, and socially desirable.

8. Future Research Directions

Future research should adopt a broader, multistakeholder perspective to deepen understanding of AI's role in counseling. Comparative analyses involving therapists, clients, and AI developers would illuminate where their perceptions of the same AI tools converge or diverge, offering insights into shared concerns as well as mismatched expectations. Longitudinal studies are also needed to track how therapists' attitudes evolve as they gain more direct experience with AI, helping to determine whether familiarity leads to increased acceptance, heightened caution, or more nuanced evaluations of AI's strengths and limitations. Beyond attitudinal research, implementation science approaches could examine how AI integration actually unfolds within real-world therapeutic environments, revealing the informal workarounds, resistances, and emergent practices that shape adoption. Cross-cultural comparative studies would further contribute by exploring how attitudes vary across national, cultural, and healthcare system contexts, thereby identifying sociocultural factors that facilitate or hinder acceptance. Additionally, outcome-focused evaluations should investigate whether AI-augmented therapy produces superior, equivalent, or inferior results relative to human-only interventions, and for which client populations or problem types these differences emerge. Finally, there is a pressing need for continued philosophical and applied ethics scholarship to develop coherent frameworks for the responsible governance of therapeutic AI, ensuring that technological innovation remains aligned with professional values and client welfare.

Appendix A: Semistructured Interview Guide

The following interview guide was used to explore therapists' general experiences, perspectives, and reflections on emerging uses of Artificial Intelligence (AI) in counseling. Questions were intentionally broad and open-ended to allow participants to guide the direction of the conversation. Prompts were used flexibly and only when needed to encourage elaboration or clarify meaning. Prompts were used flexibly and only when needed to support elaboration or clarify meaning; they were not asked verbatim or uniformly across participants.

Section 1: Professional Background and Context

1. To begin, could you tell me a bit about your background as a therapist and the kind of work you do?

a. Types of clients or settings you work in.

b. How you typically approach your practice.

c. Any experiences with technology in general (not necessarily AI).

2. How, if at all, have you come across AI or digital tools in your professional or personal life?

a. General familiarity.

b. Initial impressions.

c. Any examples that come to mind?

Section 2: General Views and Initial Reactions

3. When you hear about AI being used in counseling or mental health, what thoughts or feelings come up for you?

a. What seems interesting or surprising?

b. Any hesitations, uncertainties, or curiosities?

4. Can you describe any experiences, observations, or conversations that have shaped how you think about AI in therapy?

a. Encounters with clients.

b. Professional discussions.

c. Media or public narratives.

Section 3: Perceived Possibilities and Challenges

a. In your view, what kinds of roles, if any, might AI play in the broader landscape of counseling?

a. Any areas where it might fit naturally?

b. Situations where it might not fit as well?

b. Are there particular situations in your work where the introduction of AI feels imaginable or unimaginable?

a. Types of tasks

b. Types of clients

c. Clinical moments or contexts.

Section 4: Interactions, Communication, and Practice Realities

7. How do you imagine AI might influence the ways therapists and clients interact if it were used in some capacity?

a. Communication

b. Clinical judgment

c. Flow of sessions

8. Thinking about your everyday practice, what changes positive, negative, or otherwise do you imagine could arise if AI tools were introduced?

a. Workflow

b. Emotional or cognitive load.

c. Support or disruption.

Section 5: Professional Values and Considerations

9. What comes to mind when you think about how AI aligns or does not align with the values you hold as a therapist?

a. What aspects feel compatible?

b. What aspects feel more difficult to reconcile?

10. Are there types of client needs, identities, or contexts where you feel AI might be more or less appropriate?

a. Cultural or individual differences.

b. Experiences of vulnerability or distress.

Section 6: Future Possibilities and Reflections

11. Looking ahead, how do you imagine the future relationship between AI and the counseling profession?

a. Opportunities you could envision.

b. Concerns or cautions you think are important.

c. Hopes for how these systems could evolve.

12. Is there anything we have not discussed that you think is important when considering AI in counseling?

Data Availability Statement

The data supporting this study's findings are available on request from the corresponding author, Thseen Nazir. The data are not publicly available due to restrictions, for example, they contain information that could compromise the privacy of research participants.

Ethics Statement

All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki Declaration and its later amendments

or comparable ethical standards. The study was approved by the Ethical Committee of Ibn Haldun University with reference number E-71395021-050.04-50849 dated 20.12.2024.

Consent

Written informed consent was obtained from all participants prior to their inclusion in the study dated 25.12.2024. Consent was obtained in written form by the principal investigator. Participants were fully informed about the purpose of the research, the voluntary nature of their participation, their right to withdraw at any time without penalty, and the procedures for maintaining anonymity and confidentiality. The consent process included a clear explanation of how their responses would be used for research purposes and the intention to publish the findings in an academic journal. Participants were also assured that no identifying information would be collected and that all data would be analyzed in aggregate form only. All participants were adults (18 years or older), and no vulnerable populations were included. The consent form also clarified that there were no foreseeable risks associated with participation and that data would be stored securely and used solely for academic purposes.

Conflicts of Interest

The author declares no conflicts of interest.

Author Contributions

Thseen Nazir is solely responsible for all aspects of this research, including the conceptualization and design of the study, data collection, analysis, and interpretation. The author also conducted the literature review, developed the research methodology, and wrote the entire manuscript. All revisions and final approvals were completed by the author, ensuring the integrity and accuracy of the research presented.

Funding

No funding was received for this manuscript.

Acknowledgments

The author would like to express gratitude to Ibn Haldun University for their valuable support in making this study possible.

References

[1] V. Braun and V. Clarke, "One Size Fits All? What Counts as Quality Practice in (Reflexive) Thematic Analysis?," *Qualitative Research in Psychology* 18, no. 3 (2021): 328–352, https://doi.org/10.1080/14780887.2020.1769238.

[2] D. B. Olawade, O. Z. Wada, A. Odetayo, A. C. David-Olawade, F. Asaolu, and J. Eberhardt, "Enhancing Mental Health With Artificial Intelligence: Current Trends and Future Prospects," *Journal of Medicine, Surgery, and Public Health* 3 (2024): 100099, https://doi.org/10.1016/j.glmedi.2024.100099.

[3] P. Chandra, G. Joshi, and R. Bhagwat, "ChatGPT's Evolution in Reshaping Cognitive Behavioral Therapy," *IEEE Engineering Informatics* 2023 (2023): 1–9, https://doi.org/10.1109/IEEECONF58110.2023.10520423.

[4] K. K. Fitzpatrick, A. Darcy, and M. Vierhile, "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, no. 2 (2017): e19, https://doi.org/10.2196/mental.7785.

[5] B. Inkster, S. Sarda, and V. Subramanian, "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, no. 11 (2018): e12106, https://doi.org/10.2196/12106.

[6] S. Graham, C. Depp, E. E. Lee, et al., "Artificial Intelligence for Mental Health and Mental Illnesses: An Overview," *Current Psychiatry Reports* 21, no. 11 (2019): 116, https://doi.org/10.1007/s11920-019-1094-0.

[7] P. Raile, "The Usefulness of ChatGPT for Psychotherapists and Patients," *Humanities and Social Sciences Communications* 11, no. 1 (2024): 47, https://doi.org/10.1057/s41599-023-02567-0.

[8] V. Kumar, A. R. Ashraf, and W. Nadeem, "AI-Powered Marketing: What, Where, and How?," *International Journal of Information Management* 77 (2024): 102783, https://doi.org/10.1016/j.ijinfomgt.2024.102783.

[9] S. Affandi, M. I. Ishaq, A. Raza, Q. Talpur, and R. Ahmad, "AI Assistant Is my New Best Friend! Role of Emotional Disclosure, Performance Expectations and Intention to Reuse," *Journal of Retailing and Consumer Services* 82 (2025): 104087, https://doi.org/10.1016/j.jretconser.2024.104087.

[10] T. Nadarzynski, O. Miles, A. Cowie, and D. Ridge, "Acceptability of Artificial Intelligence (AI)-Led Chatbot Services in Healthcare: A Mixed-Methods Study," *Digital Health* 5 (2019): https://doi.org/10.1177/2055207619871808.

[11] J. Sedlakova, *Ethical and Epistemic Challenges of Conversational AI in Mental Healthcare: Interdisciplinary Inquiry for Responsible Human-AI Interaction* (University of Zurich, 2024), https://www.zora.uzh.ch/id/eprint/269480/1/269480.pdf.

[12] F. Tong, R. Lederman, S. D'Alfonso, K. Berry, and S. Bucci, "Digital Therapeutic Alliance With Fully Automated Mental Health Smartphone Apps: A Narrative Review," *Frontiers in Psychiatry* 13 (2022): https://doi.org/10.3389/fpsyt.2022.819623.

[13] J. S. Park, J. O'Brien, C. J. Cai, M. R. Morris, P. Liang, and M. S. Bernstein, "Generative Agents: Interactive Simulacra of Human Behavior," in *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology* (pp. 1–11, https://doi.org/10.1145/3586183.3606763.

[14] A. Eryılmaz and A. Başal, "Rational AIs With Emotional Deficits: ChatGPT vs. Counselors in Providing Emotional Reflections," *Current Psychology* 43, no. 45 (2024): 34962–34977, https://doi.org/10.1007/s12144-024-06947-w.

[15] A. Kottler, *On Being a Therapist Iillustrated* (Oxford University Press, 2022).

[16] R. Fulmer, T. Davis, C. Costello, and A. Joerin, "The Ethics of Psychological Artificial Intelligence: Clinical Considerations," *Counseling and Values* 66, no. 2 (2021): 131–144, https://doi.org/10.1002/cvj.12153.

[17] E. Lopes, G. Jain, P. Carlbring, and S. Pareek, "Talking Mental Health: A Battle of Wits Between Humans and AI," *Journal of*

Technology in Behavioral Science 9, no. 4 (2024): 628–638, https://doi.org/10.1007/s41347-023-00359-6.

[18] E. Go and S. S. Sundar, "Humanizing Chatbots: The Effects of Visual, Identity and Conversational Cues on Humanness Perceptions," *Computers in Human Behavior* 97 (2019): 304–316, https://doi.org/10.1016/j.chb.2019.01.020.

[19] Y. Ofosu-Asare, "Cognitive Imperialism in Artificial Intelligence: Counteracting Bias With Indigenous Epistemologies," *Ai & Society* 40, no. 4 (2025): 3045–3061, https://doi.org/10.1007/s00146-024-02065-0.

[20] W. Meissner, *The Therapeutic Alliance* (Yale University Press, 1996).

[21] T. A. Creed, P. B. Kuo, R. Oziel, et al., "Knowledge and Attitudes Toward an Artificial Intelligence-Based Fidelity Measurement in Community Cognitive Behavioral Therapy Supervision," *Administration and Policy in Mental Health and Mental Health Services Research* 49, no. 3 (2022): 343–356, https://doi.org/10.1007/s10488-021-01167-x.

[22] B. Chen, W. Zheng, L. Zhao, and X. Ding, "Leveraging Large Language Models to Assist Philosophical Counseling: Prospective Techniques, Value, and Challenges," *Humanities and Social Sciences Communications* 12, no. 1 (2025): 335, https://doi.org/10.1057/s41599-025-04657-7.

[23] R. K. Maurya, "Using AI Based Chatbot ChatGPT for Practicing Counseling Skills Through Role-Play," *Journal of Creativity in Mental Health* 19, no. 4 (2024): 513–528, https://doi.org/10.1080/15401383.2023.2297857.

[24] Y. K. Dwivedi, L. Hughes, E. Ismagilova, et al., "Artificial Intelligence (AI): Multidisciplinary Perspectives on Emerging Challenges, Opportunities, and Agenda for Research, Practice and Policy," *International Journal of Information Management* 57 (2021): 101994, https://doi.org/10.1016/j.ijinfomgt.2019.08.002.

[25] K. Chemnad and A. Othman, "Digital Accessibility in the Era of Artificial Intelligence—Bibliometric Analysis and Systematic Review," *Frontiers in Artificial Intelligence* 7 (2024): https://doi.org/10.3389/frai.2024.1349668.

[26] Y. K. Dwivedi, L. Hughes, E. Ismagilova, et al., "Artificial Intelligence (AI): Multidisciplinary Perspectives on Emerging Challenges, Opportunities, and Agenda for Research, Practice and Policy," *International Journal of Information Management* 57 (2021): 101994, https://doi.org/10.1016/j.ijinfomgt.2019.08.002.

[27] A. O. Ajlouni, A. S. Almahaireh, and F. A.-A. Whaba, "Students' Perception of Using ChatGPT in Counseling and Mental Health Education: The Benefits and Challenges," *International Journal of Emerging Technologies in Learning (IJET)* 18, no. 20 (2023): 199–218, https://doi.org/10.3991/ijet.v18i20.42075.

[28] M. Q. Patton, *Qualitative Research & Evaluation Methods: Integrating Theory and Practice* (SAGE Publications, 4th ed. edition, 2015).

[29] K. Malterud, V. D. Siersma, and A. D. Guassora, "Sample Size in Qualitative Interview Studies," *Qualitative Health Research* 26, no. 13 (2016): 1753–1760, https://doi.org/10.1177/1049732315617444.

[30] V. Braun and V. Clarke, *Successful Qualitative Research: A Practical Guide for Beginners* (SAGE Publications, 2013).

[31] G. Guest, A. Bunce, and L. Johnson, "How Many Interviews Are Enough?," *Field Methods* 18, no. 1 (2006): 59–82, https://doi.org/10.1177/1525822X05279903.

[32] M. M. Hennink, B. N. Kaiser, and V. C. Marconi, "Code Saturation Versus Meaning Saturation: How Many Interviews Are Enough?," *Qualitative Health Research* 27, no. 4 (2017): 591–608, https://doi.org/10.1177/1049732316665344.

[33] M. M. Archibald, R. C. Ambagtsheer, M. G. Casey, and M. Lawless, "Using Zoom Videoconferencing for Qualitative Data Collection: Perceptions and Experiences of Researchers and Participants," *International Journal of Qualitative Methods* 18 (2019): 1–8, https://doi.org/10.1177/1609406919874596.

[34] Y. S. Lincoln and E. G. Guba, *Naturalistic Inquiry* (SAGE Publications, 1985).

[35] V. Venkatesh, M. G. Morris, G. B. Davis, and F. D. Davis, "User Acceptance of Information Technology: Toward a Unified View1," *MIS Quarterly* 27, no. 3 (2003): 425–478, https://doi.org/10.2307/30036540.

[36] J. Lave and E. Wenger, *Situated Learning: Legitimate Peripheral Participation* (Cambridge University Press, 1991), https://doi.org/10.1017/CBO9780511815355.

[37] D. L. Goodhue and R. L. Thompson, "Task-Technology Fit and Individual Performance," *MIS Quarterly* 19, no. 2 (1995): 213–236, https://doi.org/10.2307/249689.

[38] A. B. Arrieta, N. Díaz-Rodríguez, J. Del Ser, et al., "Explainable Artificial Intelligence (XAI): Concepts, Taxonomies, Opportunities and Challenges Toward Responsible AI," *Information Fusion* 58 (2020): 82–115, https://doi.org/10.1016/j.inffus.2019.12.012.

[39] Z. Obermeyer, B. Powers, C. Vogeli, and S. Mullainathan, "Dissecting Racial Bias in an Algorithm Used to Manage the Health of Populations," *Science* 366, no. 6464 (2019): 447–453, https://doi.org/10.1126/science.aax2342.

[40] J. Henrich, S. J. Heine, and A. Norenzayan, "The Weirdest People in the World?," *Behavioral and Brain Sciences* 33, no. 2–3 (2010): 61–83, https://doi.org/10.1017/S0140525X0999152X.

[41] D. W. Sue, D. Sue, H. A. Neville, and L. Smith, *Counseling the Culturally Diverse: Theory and Practice* (Wiley, 8th ed. edition, 2019).

[42] B. Friedman and D. G. Hendry, *Value Sensitive Design: Shaping Technology With Moral Imagination* (MIT Press, 2019), https://doi.org/10.7551/mitpress/7585.001.0001.

[43] J. Decety and J. M. Cowell, "The Complex Relation Between Morality and Empathy," *Trends in Cognitive Sciences* 18, no. 7 (2014): 337–339, https://doi.org/10.1016/j.tics.2014.04.008.

[44] E. Topol, *Deep Medicine: How Artificial Intelligence Can Make Healthcare Human Again* (Basic Books, 2019).

[45] D. A. Cook, R. Hatala, R. Brydges, et al., "Technology-Enhanced Simulation for Health Professions Education: A Systematic Review and Meta-Analysis," *JAMA* 306, no. 9 (2011): 978–988, https://doi.org/10.1001/jama.2011.1234.

[46] E. Morozov, *To Save Everything, Click Here: The Folly of Technological Solutionism* (PublicAffairs, 2013).

[47] E. Vayena, A. Blasimme, and I. G. Cohen, "Machine Learning in Medicine: Addressing Ethical Challenges," *PLoS Medicine* 15, no. 11 (2018): e1002689, https://doi.org/10.1371/journal.pmed.1002689.

[48] ACA, *ACA Code of Ethics* (American Counseling Association, 2014), https://www.counseling.org/resources/aca-code-of-ethics.pdf.

[49] C. Rudin, "Stop Explaining Black Box Machine Learning Models for High Stakes Decisions and Use Interpretable

Models Instead," *Nature Machine Intelligence* 1, no. 5 (2019): 206–215, https://doi.org/10.1038/s42256-019-0048-x.

[50] C. Olah, N. Cammarata, L. Schubert, G. Goh, M. Petrov, and S. Carter, "Zoom in: An Introduction to Circuits," *Distill* 5, no. 3 (2020): e00024.001, https://doi.org/10.23915/distill.00024.001.

[51] R. Benjamin, *Race After Technology: Abolitionist Tools for the New Jim Code* (Polity Press, 2019).

[52] S. Costanza-Chock, *Design Justice: Community-Led Practices to Build the Worlds We Need* (MIT Press, 2020), https://doi.org/10.7551/mitpress/12255.001.0001.

[53] J. N. Hook, D. E. Davis, J. Owen, E. L. Worthington Jr., and S. O. Utsey, "Cultural Humility: Measuring Openness to Culturally Diverse Clients," *Journal of Counseling Psychology* 60, no. 3 (2013): 353–366, https://doi.org/10.1037/a0032595.

[54] S. Mohamed, M. T. Png, and W. Isaac, "Decolonial AI: Decolonial Theory as Sociotechnical Foresight in Artificial Intelligence," *Philosophy & Technology* 33, no. 4 (2020): 659–684, https://doi.org/10.1007/s13347-020-00405-8.

[55] A. Perry, "AI Will Never Convey the Essence of Human Empathy," *Nature Human Behaviour* 7, no. 11 (2023): 1808–1809, https://doi.org/10.1038/s41562-023-01675-w.

[56] L. Seitz, "Artificial Empathy in Healthcare Chatbots: Does It Feel Authentic?," *Computers in Human Behavior: Artificial Humans* 2, no. 1 (2024): 100067, https://doi.org/10.1016/j.chbah.2024.100067.

[57] R. Elliott, A. C. Bohart, J. C. Watson, and D. Murphy, "Therapist Empathy and Client Outcome: An Updated Meta-Analysis," *Psychotherapy* 55, no. 4 (2018): 399–410, https://doi.org/10.1037/pst0000175.

[58] L. Shu, J. Xie, M. Yang, et al., "A Review of Emotion Recognition Using Physiological Signals," *Sensors* 18, no. 7 (2018): 2074, https://doi.org/10.3390/s18072074.

[59] D. Davis, "Technology Acceptance Model: TAM," *AS: Information Seeking Behavior and Technology Adoption* 5 (1989): 205–219.

[60] E. Mumford, "The Story of Socio-Technical Design: Reflections on Its Successes, Failures and Potential," *Information Systems Journal* 16, no. 4 (2006): 317–342, https://doi.org/10.1111/j.1365-2575.2006.00221.x.

[61] G. Jain, S. Pareek, and P. Carlbring, "Revealing the Source: How Awareness Alters Perceptions of AI and Human-Generated Mental Health Responses," *Internet Interventions* 36 (2024): 100745, https://doi.org/10.1016/j.invent.2024.100745.

[62] T. Greenhalgh, J. Wherton, C. Papoutsi, et al., "Beyond Adoption: A New Framework for Theorizing and Evaluating Nonadoption, Abandonment, and Challenges to the Scale-Up, Spread, and Sustainability of Health and Care Technologies," *Journal of Medical Internet Research* 19, no. 11 (2017): e367, https://doi.org/10.2196/jmir.8775.