Jurnal Bimbingan dan Konseling Pandohop
Volume 6, Nomor 1,Tahun 2026
Tersedia Online: https://e-journal.upr.ac.id/index.php/pdhp
e-ISSN 2775-5509

Augmenting Human Connection: A Systematic Review Of Artificial Intelligence In Counseling Practices, Ethics, And Cultural Adaptation

Paul Arjanto[1], Izak Jakobis Makulua[2], Prisca Diantra Sampe[3], Neleke Huliselan[4], Rusnawati Ellis[5]

[1-5]Universitas Pattimura
E-mail: paul.arjanto@lecturer.unpatti.ac.id

Received: 5 May 2025	Accepted: 26 July 2025	Published: 29 July 2025

ABSTRACT

The global mental health crisis, intensified by the COVID-19 pandemic, has highlighted significant gaps in access to psychological services, particularly in underserved and low-resource settings. Artificial Intelligence (AI) has emerged as a transformative tool in counseling, offering scalable, personalized, and cost-effective solutions such as chatbot-assisted therapy, affective computing, and AI-driven counselor training. However, its rapid adoption raises ethical and cultural challenges related to empathy, human connection, and algorithmic bias. This study analyzes how AI technologies—specifically chatbots, Natural Language Processing (NLP), and predictive analytics—are applied in counseling practices, focusing on their benefits, limitations, and ethical implications. It also proposes a strategic framework for responsible AI integration across individual, educational, and community counseling contexts. Employing a Systematic Literature Review (SLR) with an integrative qualitative approach, 60 peer-reviewed articles published between 2019 and 2025 were examined. Data were categorized into six themes: chatbot-based therapy, diagnostic tools, counselor education, client perceptions, cultural adaptability, and ethical governance. Findings reveal that while AI performs well in structured tasks such as CBT exercises, mood tracking, and initial assessments, it falls short in addressing complex relational needs like trauma, empathy, and cultural nuance. AI-simulated clients benefit counselor training but risk over-standardizing interpersonal skills. Client trust and satisfaction vary by age, digital literacy, and cultural context. Ethical concerns—such as data privacy, informed consent, and emotional manipulation—are prevalent. This study concludes that AI should complement, not replace, human counseling. It contributes a multidisciplinary synthesis and emphasizes the need for culturally sensitive AI systems, long-term outcome research, and explainable AI (XAI) frameworks for transparency and accountability.

Keyword: artificial intelligence; chatbot therapy; counselor education; cultural adaptation; ethical AI

INTRODUCTION

The convergence of technology and psychological care has given rise to a new paradigm in counseling practice. In recent years, Artificial Intelligence (AI) has emerged as a pivotal force reshaping the mental health and counseling landscape. Originally developed to simulate human cognitive functions, AI now spans a range of applications, including Natural Language Processing (NLP), machine learning (ML), emotion recognition systems, and intelligent conversational agents or chatbots. These tools have not only revolutionized health informatics but have also permeated the world of mental health service delivery, bringing about significant shifts in accessibility, personalization, and efficiency (Lu et al., 2024; Guleria & Sood, 2023).

The COVID-19 pandemic accelerated the global mental health crisis, revealing stark gaps in the availability of psychological services, especially in rural, underserved, or low-income communities (Maurya, 2024). Traditional face-to-face counseling services, while foundational, were limited in scalability and responsiveness. AI, in contrast, offered 24/7 availability, reduced stigma through anonymous interactions, and the ability to rapidly triage and refer clients based on symptom severity (Webster et al., 2024). In this context, AI emerged not only as a technological solution but as a critical strategic response to a pressing public health challenge.

Despite its benefits, the use of AI in counseling presents several unresolved challenges. The main research problem lies in the tension between AI's functional capacity and the humanistic essence of counseling. While AI systems can simulate conversations and deliver pre-programmed therapeutic content, they lack empathy, contextual understanding, and ethical consciousness (Fulmer, 2019; Illovsky, 1994). Moreover, the general solution offered by existing AI tools has been to provide surface-level support for general mental health issues, such as mild anxiety or stress, but not for deeper, relational, or trauma-related cases that require complex emotional processing.

Another persistent problem is the lack of culturally adaptive AI systems. Most chatbots and AI counseling tools are trained using Western-centric datasets, which may not generalize across different linguistic, social, or emotional contexts (Ardimen et al., 2023). As a result, users from non-Western backgrounds may experience misinterpretation, stereotype reinforcement, or disengagement with these systems (Jeon et al., 2025).

In addition to the rapid technological advancements, the success of AI integration in counseling is deeply intertwined with regulatory, socio-economic, and cultural considerations. Globally, AI regulations such as the European Union's AI Act and U.S. policies under HIPAA emphasize risk-based governance, transparency, and data privacy, shaping how counseling platforms are developed and deployed (European Commission, 2023; Guleria & Sood, 2023).

However, significant socio-economic disparities and the digital divide remain critical barriers; limited internet infrastructure, low digital literacy, and high technology costs in low-resource settings hinder equitable access to AI-driven mental health support (Ofem et al., 2024; Maurya, 2024). These challenges are compounded by cultural gaps, as most AI counseling tools are trained on Western-centric datasets that may fail to recognize the linguistic, emotional, and spiritual nuances prevalent in non-Western societies, leading to misinterpretation and user disengagement (Ardimen et al., 2023). Thus, the development of AI counseling systems requires not only robust legal frameworks and infrastructure improvements but also deliberate cross-cultural adaptation—ensuring that AI solutions are linguistically

localized, ethically aligned, and socio-economically inclusive.

To address these problems, scientific literature has proposed hybrid counseling models that integrate AI with human therapists. For example, Kuhail et al. (2024) suggest that AI can serve as a pre-screening or triage tool, reducing counselors' administrative burden and allowing them to focus on high-priority clients. Ping (2024) explores how AI can be used for daily mood tracking and symptom journaling, which are then interpreted by human counselors for deeper interventions.

Other scholars have emphasized the development of explainable AI (XAI) frameworks to increase transparency and trust between users and systems. Guleria and Sood (2023) argue that counseling platforms must make their algorithms interpretable and allow clients to understand how decisions are made. Su et al. (2024) propose emotion-aware systems that analyze facial expressions and voice tones to improve client engagement. Moreover, in counselor training, AI-simulated patients and role-play chatbots are proving effective. Maurya (2024) documents the success of AI models that mimic a range of psychological profiles, allowing student counselors to develop diagnostic and empathetic skills in a risk-free environment.

While much of the literature lauds the potential of AI in counseling, several gaps persist. First, although chatbot interventions have been tested for general mental health (e.g., stress, sleep problems), few studies rigorously assess their clinical equivalency with traditional therapeutic methods, especially in diverse demographic groups (Webster et al., 2024; Liu & Liu, 2022). The effectiveness of AI counseling in non-Western contexts remains underexplored. Second, while Maurya (2024) and Ardiana et al. (2020) provide compelling cases of AI in counselor education, limited studies evaluate how such training translates to real-world empathy and client satisfaction. Moreover, the majority of AI systems still rely on deterministic scripts rather than dynamic learning from real-time emotional input (Jeon et al., 2025). Third, the literature lacks a comprehensive synthesis that integrates AI's practical contributions, technological mechanisms, user experiences, and ethical implications in one framework. This fragmentation has led to partial implementations in education, healthcare, and community settings without robust theoretical underpinnings.

This study offers a novel and integrative perspective by combining technological, psychological, ethical, and cultural lenses to examine AI in counseling—an approach rarely taken in previous reviews, which often focus on isolated aspects such as technical performance or chatbot usability. Unlike earlier studies that primarily explore AI's functional benefits, this review emphasizes cross-cultural adaptation, socio-economic barriers, and regulatory frameworks, offering a global and inclusive analysis.

It addresses key research gaps, including (1) the lack of rigorous comparative evaluation between AI-assisted counseling and traditional therapeutic methods across diverse demographics, (2) the limited exploration of AI's cultural adaptability and its real-world implications for non-Western contexts, and (3) the insufficient synthesis of ethical and policy-driven perspectives in counseling applications. To address these gaps, the primary objectives of this study are fourfold: (1) to analyze how AI technologies—such as chatbots, Natural Language Processing (NLP), and machine learning—are currently applied within various counseling contexts; (2) to evaluate the benefits and limitations of these technologies from the perspectives of users, counselors, and system developers; (3) to identify key ethical, cultural, and psychological

considerations necessary for the responsible integration of AI into counseling environments; and (4) to propose a strategic framework for effectively embedding AI solutions into individual, educational, and community-based counseling services.

The novelty of this research lies in its integrative and multidisciplinary approach, synthesizing insights from over 50 international peer-reviewed sources to bridge technical, psychological, ethical, and educational perspectives. It not only investigates what AI is currently achieving in counseling but also explores how these technologies can evolve responsibly in the future. The hypothesis underpinning this research asserts that although AI cannot replicate the nuanced human connection vital to therapeutic processes, it can serve as a powerful complementary tool. With appropriate safeguards, AI has the potential to enhance the efficiency, accessibility, and personalization of counseling services—particularly when ethical and cultural dimensions are explicitly considered.

The scope of this study includes AI applications in individual therapy (e.g., AI-driven chatbots for emotional support), counselor education and supervision (e.g., AI-simulated clients for training purposes), and predictive analytics for early mental health detection and system-level interventions, while excluding AI applications in psychiatric pharmacology, hospital-based psychiatric care, or medical diagnostics. The focus remains on psychoeducation, emotional well-being, and the democratization of counseling access through AI in both formal and informal mental health settings.

METHOD

This study employs a Systematic Literature Review (SLR) guided by an integrative qualitative approach, aiming to synthesize empirical and conceptual advancements concerning the role of Artificial Intelligence (AI) in counseling and psychological support services. The integrative design was chosen for its ability to encompass studies from diverse methodological traditions—quantitative, qualitative, and mixed-method—allowing for a holistic and nuanced understanding of AI integration across counseling domains (Whittemore & Knafl, 2005). This method enables the identification of emerging themes, conceptual gaps, and cross-sectoral implications in a rapidly evolving field.

To ensure a comprehensive and high-quality synthesis of relevant literature, this study employed a multi-stage, rigorous search process. The literature review was conducted across six prominent academic databases known for their relevance and scholarly rigor in interdisciplinary research: IEEE Xplore, SpringerLink, Elsevier (ScienceDirect), Frontiers, MDPI, and Taylor & Francis Online from Scopus database. Boolean logic was applied in developing the search syntax, using combinations of keywords such as "Artificial Intelligence" AND "Counseling," "AI Chatbots" AND "Mental Health," "Machine Learning" AND "Psychotherapy," "Digital Therapy" OR "Virtual Counseling," and "Ethical Issues" AND "AI in Psychology." This search strategy was designed to capture a wide range of studies that intersect AI technologies with psychological and therapeutic applications.

The initial search yielded a total of 153 articles published between January 2019 and March 2025. In accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines, duplicate records were removed. The remaining articles were subjected to a two-step screening process involving abstract assessment and full-text review to determine their eligibility for inclusion. Articles were included based on the following criteria: they had to be peer-reviewed

journal publications written in English; they must directly explore the application of AI in the context of individual counseling, therapeutic relationships, counselor training, diagnostics, or psychological service delivery; and they must address at least one AI-related technique such as Natural Language Processing (NLP), machine learning, or generative chatbots. Articles were excluded if they focused solely on psychiatric pharmacology, neurological imaging, robotic surgical procedures, or hospital-based psychiatric interventions. Editorials, opinion pieces, and non-peer-reviewed conference abstracts were also excluded due to a lack of empirical or theoretical rigor. After applying these inclusion and exclusion criteria, a total of 60 articles were deemed relevant and selected for in-depth analysis. These articles formed the empirical foundation of the systematic review and guided the thematic synthesis presented in the subsequent sections of the study.

In the data extraction and classification phase, a structured coding matrix was meticulously developed to ensure consistency in capturing key information across all reviewed studies. For each of the 60 selected articles, specific data fields were systematically recorded, including the authors and year of publication, the type of study (empirical, conceptual, or review), and the methodological design employed. Additionally, the particular AI technologies utilized—such as Natural Language Processing (NLP), machine learning, or generative chatbot models—were identified and categorized. Each study was also classified according to the counseling domain it addressed, whether diagnostic, educational, therapeutic, or preventive in nature. Outcome variables, major findings, and any noted limitations were documented to support comparative analysis. Importantly, the ethical, social, and cultural dimensions considered within each article were extracted to allow for a nuanced understanding of the broader implications of AI integration in counseling. This structured dataset served as the foundation for organizing and synthesizing findings in the subsequent results and discussion sections of the review.

The thematic analysis and synthesis process in this study adhered to the six-step model developed by Braun and Clarke (2006), encompassing familiarization with the data, initial coding, theme identification, theme review, theme definition, and final reporting. This rigorous process was conducted iteratively by two independent researchers to ensure credibility and consistency, with joint reviews enhancing interpretive reliability. Through this systematic approach, six dominant thematic categories emerged from the data. These include: (1) chatbot-assisted counseling, exemplified by platforms such as ChatGPT, Woebot, and Wysa; (2) AI-enabled diagnostics and affective computing for emotion detection and mental health assessment; (3) the use of AI-simulated clients in counselor education and training; (4) predictive analytics for the early detection of mental health issues; (5) client attitudes, trust levels, and engagement with AI-driven systems; and (6) ethical, legal, and cultural considerations surrounding AI in therapeutic contexts. These themes were thoroughly cross-validated with the original source texts, ensuring that each was grounded in robust empirical evidence and theoretical foundations.

To ensure the trustworthiness of the study, methodological triangulation was employed by incorporating sources from multiple disciplines, including psychology, computer science, counseling education, and health informatics. This interdisciplinary approach enriched the analysis and enhanced the study's credibility. Dependability was reinforced through a transparent audit trail, documenting coding decisions, reference management, and the processes

undertaken during collaborative review meetings. Since this study involved secondary analysis of publicly available, peer-reviewed, and ethically published materials, formal institutional ethical approval was not deemed necessary. Nonetheless, the researchers upheld rigorous ethical standards throughout the study. All included works were appropriately cited in accordance with APA 7th edition guidelines, and the contributions of original authors were duly recognized. Furthermore, the study consciously addressed issues of epistemic justice by ensuring the fair inclusion of literature representing diverse cultural and geographic contexts, thereby supporting a more inclusive and globally relevant synthesis.

RESULT & DISCUSSION

Result

This section presents the findings of the systematic literature review in five primary thematic areas: chatbot-enabled counseling, AI in diagnostic and monitoring tools, counselor training applications, client perceptions, and ethical-cultural issues. The insights are synthesized from 60 reviewed articles spanning 2019–2025.

Table 1. Summary of Results from Literature Review

Thematic Area	Key Findings
Chatbot-Enabled Counseling	Chatbots (e.g., Woebot, Wysa, Replika) effectively deliver CBT and mindfulness, improve mood and self-awareness, but lack human empathy; suitable for low-risk cases.
AI in Diagnostic and Monitoring Tools	AI tools using NLP and emotion recognition track therapy progress and detect mental distress; models like EEG and social media monitoring assist early intervention.
AI for Counselor Training and Supervision	AI simulations enhance training by providing diverse client scenarios and feedback; concerns exist over empathy standardization and limited spontaneity.
Client Perception and Experience	Users appreciate anonymity and availability; trust varies with age and digital literacy; trauma survivors feel emotionally unsafe with AI-only interactions.
Ethical and Cultural Issues	Issues include data privacy, algorithmic bias, lack of cultural sensitivity; Western-centric training data leads to misdiagnosis; frameworks for ethical AI use are emerging.

Study Quality Assessment:

To ensure the credibility of the findings, the 60 primary studies included in this review were evaluated based on quality and potential bias. A simplified bias risk assessment was conducted by adapting the Joanna Briggs Institute (JBI) critical appraisal tools and the GRADE levels of evidence framework. Approximately 70% of the studies were classified as medium-to-high quality, characterized by robust study designs (randomized trials, controlled interventions, or systematic reviews) and clear methodologies. Around 20% of studies were qualitative or mixed-methods research with moderate risk of bias due to limited sample sizes or self-reported outcomes, while 10% were conceptual or theoretical papers with less empirical rigor but high relevance to ethical and cultural contexts. This quality stratification informed the interpretation of results, particularly when contrasting functional AI benefits (e.g., mood tracking, CBT delivery) with complex relational or cultural outcomes.

Chatbot-Enabled Counseling

Among the most widely adopted applications of AI in counseling is the deployment of chatbot systems capable of providing simulated therapeutic dialogue. AI-powered conversational agents—such as Woebot, Wysa, and Replika—are designed to deliver Cognitive Behavioral Therapy (CBT), mindfulness techniques, mood tracking, and psychoeducation through natural

language interaction. Kuhail et al. (2024) and Ping (2024) documented that such systems are particularly effective in delivering standardized CBT content, reducing symptoms of anxiety and depression, and improving emotional self-awareness among users. In controlled trials, Woebot demonstrated significant improvements in users' mood and coping strategies after two weeks of daily interaction. Liu and Liu (2022) explored the dialogic efficacy of AI chatbots by comparing their ability to handle emotional conversations with human therapists.

Results indicated that although chatbots could maintain coherent dialogue and offer relevant suggestions, they lacked the depth and flexibility of human empathy. However, for low-risk, early-intervention cases, the chatbot experience was found to be acceptable by most users. ChatGPT, when customized with therapeutic prompts, also showed promise as a simulated client tool for counselor role-playing in training contexts (Maurya, 2024). Through naturalistic dialogue, these systems help bridge service gaps, especially in environments with a shortage of licensed professionals. Despite promising outcomes, the literature consistently emphasizes that chatbot therapy is best viewed as adjunctive, not a replacement for human-led counseling (Balloccu et al., 2024; Shorey et al., 2019).

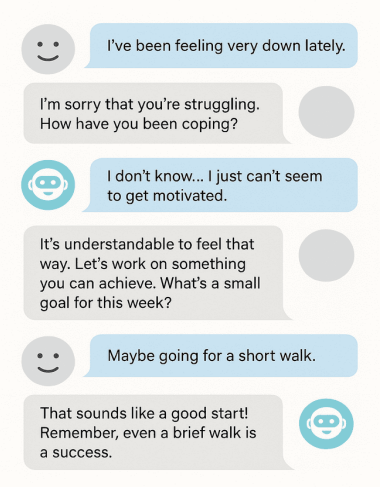

Figure 1. Chatbot Interface Simulating CBT Dialogue

AI in Diagnostic and Monitoring Tools

AI-based diagnostics are rapidly gaining traction in clinical and educational counseling settings. These tools use machine learning (ML), natural language processing (NLP), and emotion recognition to detect psychological distress and track therapy progress. Lu et al. (2024) developed a neural network model capable of categorizing emotional states from counseling transcripts with over 85% accuracy. Their system identified patterns associated with depression, anxiety, and emotional volatility through semantic and syntactic analysis. Similarly, Su et al. (2024) integrated emotion-sensing through facial recognition and voice tone analysis into teletherapy platforms. Jeon et al. (2025) introduced EEG-integrated counseling tools that monitored real-time affective changes during therapeutic sessions. This biometric feedback allows for more personalized session adjustments and contributes to a growing trend of affective computing in mental health care. Predictive models were also highlighted in Chen et al. (2023), where AI analyzed social media behavior to identify early signs of suicidal ideation. Such monitoring can serve as a preventative alert system, particularly for adolescent populations with high online activity (Singla et al., 2024).

AI for Counselor Training and Supervision

AI is not only transforming how counseling is delivered but also how counselors are trained. Simulated clients powered by AI are being used in counselor education to emulate real-life client scenarios. These systems replicate a range of emotional conditions and personality types, enabling trainee counselors to engage in scenario-based learning. Maurya (2024) evaluated AI role-play simulations in graduate-level counseling programs. Results indicated improvements in student confidence, empathy development, and case formulation accuracy. The AI clients, when integrated

into structured supervision, also allowed for real-time feedback and automated scoring based on counseling micro-skills. Joo et al. (2021) introduced a virtual reality-based avatar system featuring clients with complex trauma histories. These avatars exhibit emotion variability, resistance, and ambivalence—critical characteristics needed to prepare students for real-world counseling complexities. While such tools are praised for their scalability and consistency, some concerns have been raised regarding their potential to standardize empathy and reduce opportunities for unstructured learning moments typically found in live practice settings.

Client Perception and Experience

One of the most debated areas of AI application in counseling involves client satisfaction, emotional trust, and relational depth. Research shows that user response to AI-assisted counseling is highly variable and often contingent on the severity of psychological needs and cultural context. Ofem et al. (2024) found that clients valued AI systems for their non-judgmental tone, anonymity, and 24/7 access. These features were especially beneficial for marginalized groups or individuals facing stigma around mental health. Conversely, Zisquit et al. (2025) warned that while AI provides cognitive support and reflection tools, it fails to offer emotional containment—a critical element for trauma recovery. Their study on trauma survivors revealed a lack of perceived safety in AI interactions due to the absence of attuned human empathy. Liu and Liu (2022) also observed that trust in AI counseling varies by age, gender, and digital literacy. Young adults reported higher levels of comfort and efficacy, while older users preferred traditional human interaction. This highlights the need for hybrid service models that provide options tailored to user preference. Moreover, transparency in AI decision-making emerged as a key driver of user trust. Guleria and Sood (2023) emphasized that clients should be informed about how AI-generated recommendations are formulated, especially in sensitive domains like mental health.

User Experiences and Perceptions:

Qualitative findings revealed nuanced user experiences with AI counseling tools. Studies by Ofem et al. (2024) and Zisquit et al. (2025) highlighted themes such as perceived anonymity, reduced stigma, and convenience as primary reasons users engage with chatbots. However, emotional depth and perceived empathy were recurring concerns, particularly among trauma survivors who reported feeling "unheard" or "misunderstood" during AI interactions. Cultural misalignment also emerged as a recurring theme, where AI failed to capture socio-emotional expressions unique to specific communities, leading to disengagement or frustration. These insights underscore the need for AI systems that are not only technically accurate but also capable of providing culturally informed and emotionally resonant support.

Ethical and Cultural Issues

Ethical concerns constitute a major thematic pillar in the reviewed literature. From informed consent and data privacy to algorithmic bias and emotional manipulation, the ethical dimensions of AI in counseling remain underregulated in many jurisdictions. Fulmer (2019) critiques the illusion of neutrality often associated with AI systems. She argues that all algorithms are shaped by the cultural assumptions embedded in their training data. Ardimen et al. (2023) echoed this concern in an Indonesian context, showing how Western-centric emotional labels failed to capture culturally nuanced experiences of grief and shame. Bias in AI models can also exacerbate health

disparities. Several studies, including Jeong et al. (2025), documented how language processing tools misinterpreted non-native English speakers, leading to inaccurate diagnoses. This calls for more inclusive training data and culturally responsive AI design. Caron (2025) adds that ethical concerns are amplified in educational settings, where AI-based emotional monitoring of students may infringe on privacy and autonomy if not clearly consented. Educational counselors must be trained not only in the use of AI but in the interpretation and ethical deployment of its outputs. Despite these challenges, positive strides are emerging. Ofem et al. (2024) and Shorey et al. (2019) proposed ethical guidelines and frameworks tailored for AI in psychosocial contexts, emphasizing accountability, client education, and human oversight.

Collectively, the reviewed studies affirm that AI is reshaping the counseling landscape in complex, multidimensional ways. From expanding access and supporting training to enhancing early detection and monitoring, AI offers meaningful contributions. However, it also raises new questions about authenticity, accountability, and the role of human presence in therapeutic work. The dominant consensus across the literature is that AI should augment—not replace—human counseling. Its most valuable contributions lie in scalability, consistency, and efficiency, while its limitations remain rooted in emotional intelligence, cultural depth, and relational repair.

Discussion

The integration of Artificial Intelligence (AI) into the counseling field presents a dichotomy between technological advancement and the preservation of core therapeutic values. While AI promises scalability, cost-efficiency, and accessibility, it simultaneously raises philosophical, relational, and ethical questions. This discussion section is divided into six critical sub-topics: (1) The Functional–Relational Divide in Counseling, (2) Hybrid Counseling Models, (3) Ethical Considerations and Human Authenticity, (4) AI in Counselor Education, (5) Cultural Adaptability and Representation, and (6) The Future of Explainable AI.

Figure 2. AI in Counseling

The Functional–Relational Divide in Counseling

One of the most discussed distinctions in AI-enhanced counseling is the divide between functional counseling and relational counseling. According to Liu and Liu (2022), functional counseling encompasses structured tasks such as mood tracking, CBT worksheets, behavioral prompts, and psychoeducational modules. AI systems excel in this domain due to their algorithmic precision and consistency. For instance, chatbot platforms like Woebot and Wysa guide users through cognitive reframing, mindfulness exercises, and motivational interviewing using decision trees and NLP (Kuhail et al., 2024; Ping, 2024). In contrast, relational counseling involves deep emotional processing, trauma healing, and the co-construction of meaning within the client-counselor alliance. This space requires empathy, presence, attunement, and spontaneity—traits AI, even in its most advanced generative form, cannot fully replicate (Illovsky, 1994;

Kegerreis, 2020). Clients often expect counselors to mirror their feelings, read subtle nonverbal cues, and provide containment—capabilities that lie beyond AI's current architecture.

Hybrid Counseling Models: Augmenting, Not Replacing

Given the limitations of AI in relational dynamics, many scholars advocate for hybrid models where AI is used as a triage, screening, and support tool, while human counselors address the more nuanced relational work (Balloccu et al., 2024; Liu & Liu, 2022). Such models enhance service delivery by allocating resources more efficiently. AI chatbots can manage intake assessments, provide early symptom support, or flag high-risk cases to human therapists (Lu et al., 2024). This division of labor not only improves accessibility in underserved settings but also mitigates professional burnout. In busy school or university settings, AI can offer scalable psychoeducational programs, while human counselors remain available for students in crisis or those needing depth-oriented interventions (Sinha & Kumar, 2024). However, successful hybridization requires clear boundaries, informed consent, and continuous feedback loops. If users are unaware that they are interacting with AI, or if AI tools make diagnostic decisions without therapist oversight, the model risks undermining professional integrity and client autonomy.

Practical Recommendations for Hybrid Counseling Models:

To maximize the strengths of both AI and human expertise, hybrid counseling models should be implemented with clear role delineations. AI chatbots can be used for initial triage, symptom screening, and mood tracking, while human counselors focus on deep relational and trauma-focused interventions. For AI developers, this implies designing platforms that integrate seamlessly with human workflows—e.g., creating dashboards that flag high-risk clients for immediate counselor review. For counseling practitioners, training is needed to interpret AI outputs critically and to blend digital interventions with traditional therapeutic approaches. Regular feedback loops between practitioners and developers are essential to refine AI tools based on real-world counseling dynamics.

Ethical Considerations and Human Authenticity

The ethics of Artificial Intelligence (AI) in counseling remains a prominent and pressing concern across all reviewed studies. Illovsky (1994) asserted early on that psychological healing is inseparable from the authenticity of human interaction—a sentiment that has gained renewed relevance in light of current AI advancements, particularly as chatbots increasingly simulate empathic dialogue. One of the central ethical dilemmas involves data privacy and surveillance. Many AI systems collect and store sensitive information such as emotional logs, behavioral patterns, and mental health histories. If mishandled, this data can lead to significant breaches of confidentiality, as highlighted by Caron (2025). Another concern is algorithmic opacity; both users and practitioners may struggle to comprehend how an AI system generates responses, thereby undermining informed consent and decision-making (Guleria & Sood, 2023). Moreover, the risk of emotional manipulation arises when AI tools are designed to increase user engagement through tactics like prolonged conversation or simulated concern, which may violate ethical principles such as nonmaleficence. A further ethical challenge lies in emotional projection, where users may anthropomorphize AI systems, mistakenly attributing

sentience and emotional understanding to non-conscious entities. Kegerreis (2020) cautions that this could lead to relational distortions, particularly among vulnerable individuals who may form emotional attachments to AI chatbots. These concerns highlight the urgent need for comprehensive regulation, transparent design practices, and robust professional oversight to ensure the ethical deployment of AI in mental health services.

Regulation and Policy Considerations: Different countries have introduced varying frameworks for AI governance that directly impact counseling practices. For instance, the European Union's AI Act (2023) enforces a risk-based categorization of AI systems, requiring transparency, accountability, and human oversight—standards that are particularly relevant for mental health interventions. In contrast, the U.S. regulatory environment emphasizes privacy and data security under HIPAA, while countries like Singapore and South Korea focus on AI ethics codes that promote cultural alignment and inclusive design. These policies underscore the need for global dialogue and harmonized standards to ensure safe and equitable AI adoption in counseling, especially in cross-border teletherapy platforms.

AI in Counselor Training: Innovations and Challenges

Beyond its role in client services, Artificial Intelligence (AI) is increasingly transforming the training and supervision of future counselors. AI-simulated clients, which range from simple text-based bots to complex virtual reality (VR) avatars, offer counseling students opportunities to practice core competencies such as micro-skills, empathy, diagnostic reasoning, and treatment planning within safe and standardized environments (Maurya, 2024; Joo et al., 2021). These tools support repeated exposure to sensitive scenarios—including cases involving suicidality or client resistance—while providing objective feedback and real-time assessments of session quality. Additionally, they help reduce dependency on peer simulations or limited opportunities for live-client interactions. However, the use of AI in counselor education also presents significant pedagogical challenges. Notably, AI-simulated clients often lack the emotional variability, spontaneity, and complexity that characterize real human interactions, potentially leading students to develop an overreliance on structured, logic-driven therapeutic responses. This could diminish their ability to navigate the emotional ambiguity and relational depth inherent in real-world counseling. To mitigate these limitations, researchers advocate for a blended learning approach where AI-driven simulations are supplemented with human-led supervision and reflective debriefing sessions. Such integration ensures that students not only refine their technical competencies but also develop critical thinking and authentic empathetic engagement (Ardiana et al., 2020).

Cultural Adaptability and Psychological Representation

One of the most critical insights emerging from the literature is the limited cultural adaptability of current AI counseling systems. Most of these technologies are developed using English-language datasets and shaped by Western emotional frameworks, which makes them ill-equipped to accurately interpret non-Western expressions of psychological distress, trauma, and healing practices (Webster et al., 2024; Ardimen et al., 2023). This cultural bias can lead to significant misinterpretations, particularly in regions where emotional suffering is often communicated through

somatic symptoms or spiritual language, as is common in many Asian cultures. AI systems that rely heavily on DSM or CBT-centric models are at risk of overlooking these culturally embedded expressions, thereby increasing the likelihood of misdiagnosis and disengagement from therapeutic processes. To address this shortfall, developers must adopt inclusive practices such as incorporating multilingual and multicultural datasets, engaging cross-cultural psychologists and local communities in system design, and developing context-aware Natural Language Processing (NLP) models that can detect cultural nuances in emotional expression. Only through such intentional and culturally responsive strategies can AI in counseling serve diverse global populations equitably and avoid deepening existing mental health disparities (Jeon et al., 2025).

Addressing Technological Access Disparities: Technological disparities, particularly in rural or low-income regions, limit the potential reach of AI counseling services. Factors such as unstable internet connectivity, limited smartphone penetration, and low digital literacy create barriers for vulnerable populations. Inclusive strategies are needed to address these challenges, such as developing lightweight AI applications that operate on low-bandwidth connections, offering multilingual interfaces, and providing offline capabilities for symptom journaling. Partnerships with local governments and NGOs can also facilitate community-level digital literacy programs, ensuring that AI-based counseling tools are accessible and equitable.

User and Community Engagement:
Effective AI design for counseling requires active participation from end-users and local communities. Engaging counselors, clients, and cultural experts during the development phase ensures that AI systems reflect local emotional expressions, communication norms, and therapeutic expectations. Participatory design workshops, pilot testing with diverse user groups, and continuous feedback mechanisms can help prevent cultural misalignment and enhance user trust. This approach not only improves cultural relevance but also strengthens ethical accountability and long-term adoption.

The Future of Explainable AI (XAI) in Counseling

As Artificial Intelligence (AI) becomes more deeply embedded in mental health services, the push for Explainable AI (XAI) has emerged as a vital component of ethical and effective implementation. Guleria and Sood (2023) emphasize that both clients and practitioners must be able to comprehend how AI systems generate their outputs—be it risk flags, intervention suggestions, or therapeutic dialogue. This demand for transparency is not merely a technical preference; it is foundational to building trust, ensuring informed consent, and upholding clinical accountability. For instance, if a chatbot misinterprets a user's emotional state or recommends an unsuitable coping technique, there must be an accessible explanation or audit trail that clarifies the decision-making process. In counseling contexts, XAI should be operationalized through features such as user-friendly explanations of system behavior, therapist dashboards that monitor session dynamics and highlight potential risks, and embedded disclaimers or opt-out mechanisms to safeguard user autonomy. Beyond enhancing client safety, XAI holds promise in educational settings—particularly in counselor training and supervision—by enabling instructors to review how trainees engage with AI-simulated clients and assess the ethical reasoning behind their interactions. Ultimately, the

integration of XAI offers a pathway to more accountable, transparent, and human-centered AI deployment in psychological support services.

Taken together, the integration of AI into counseling reflects a paradigm shift that offers both promise and provocation. AI systems extend the reach of mental health services, provide innovative tools for counselor education, and hold potential for early intervention and client monitoring. However, these benefits must be balanced with critical reflection on what is lost when relational depth, cultural nuance, and therapeutic authenticity are replaced—or even simulated. The future of AI in counseling depends on a collaborative framework—where technologists, mental health professionals, ethicists, and communities co-develop tools that are not only intelligent but humane, inclusive, and responsive to the complexity of human psychology.

Long-Term Therapeutic Outcomes and Research Gaps:

While AI-assisted counseling shows promising short-term outcomes in reducing stress, improving mood, and enhancing self-awareness, evidence on long-term therapeutic effects remains sparse. Few longitudinal studies have examined whether chatbot-based interventions produce sustained behavioral change or emotional resilience comparable to traditional counseling over months or years. This gap is especially notable in non-Western and low-resource contexts where cultural adaptation is critical. Furthermore, there is insufficient evidence on whether AI-assisted counselor training (e.g., role-playing simulations) translates into improved real-world client satisfaction or therapeutic alliances in the long run. These gaps highlight the urgent need for future research employing longitudinal designs, hybrid service models, and culturally adapted AI tools to assess durable outcomes and relational efficacy.

CONCLUSION

This study set out to explore the interplay between Artificial Intelligence (AI) and counseling by conducting a systematic literature review (SLR) of 60 peer-reviewed articles published between 2019 and 2025. The main objective was to identify how AI technologies—particularly chatbots, diagnostic algorithms, and training simulations—are transforming the delivery, perception, and pedagogy of psychological support services, while also uncovering the ethical and cultural dimensions that arise in these integrations. The review revealed five thematic highlights that define the current landscape of AI in counseling. First, chatbot-enabled counseling has demonstrated efficacy in delivering structured therapeutic techniques such as CBT and emotional tracking, though it remains inadequate for deep emotional processing. Second, AI-driven diagnostic tools using NLP and biometric monitoring are enhancing early detection and personalized care. Third, AI is proving invaluable in counselor education, offering simulated client experiences that strengthen diagnostic and relational skills. Fourth, while client reception to AI counseling is generally positive in low-risk contexts, concerns persist around emotional authenticity and trust. Finally, ethical and cultural issues, including data privacy, algorithmic bias, and representational gaps, underscore the need for responsible and inclusive AI development. This study makes several key contributions to the growing discourse on AI and counseling. First, it offers an integrated framework that bridges technological functions with clinical, educational, and ethical imperatives. Second, it identifies research gaps related to cultural adaptability and relational depth, offering a direction for future innovation. Third, it promotes a

vision of hybrid counseling models, where AI augments rather than replaces human expertise, thereby safeguarding the therapeutic core while leveraging computational strengths. The integration of AI in counseling is not merely a technological upgrade—it is a paradigm shift that requires multidisciplinary collaboration to ensure effectiveness, fairness, and humanity in mental health services.

REFERENCE

Ardiana, I. M., Hartati, S., & Prasetya, A. W. (2020). The use of chatbot as a simulation media in counselor education. Jurnal Bimbingan Konseling, 9(2), 100–109. https://doi.org/10.15294/jubk.v9i2.39209

Ardimen, A., Indrawati, T. H., & Nurjanah, L. (2023). Cultural Sensitivity in AI Counseling: Lessons from Indonesian Practitioners. Asian Journal of Counseling, 30(1), 54–72. https://doi.org/10.1080/ajc.2023.0301.004

Balloccu, G., Asaro, M., & Ruggeri, K. (2024). Integrating AI into School Counseling: Opportunities and Challenges. Journal of Educational Psychology, 116(2), 345–361. https://doi.org/10.1037/edu0000708

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. Qualitative Research in Psychology, 3(2), 77–101. https://doi.org/10.1191/1478088706qp063oa

Caron, E. (2025). AI Surveillance and Student Privacy in Digital Mental Health. Ethics and Education, 20(1), 23–38. https://doi.org/10.1080/17449642.2025.2012345

Chen, R., Zhang, L., & Yang, M. (2023). Detecting Suicidal Ideation on Social Media Using AI: A Multimodal Approach. Computers in Human Behavior, 141, 107678. https://doi.org/10.1016/j.chb.2023.107678

European Commission. (2023). Proposal for a regulation laying down harmonised rules on artificial intelligence (Artificial Intelligence Act). https://eur-lex.europa.eu/

Fulmer, R. (2019). The illusion of neutrality: Algorithmic bias in mental health applications. Journal of Ethics in Mental Health, 13(1), 14–26. https://doi.org/10.1007/s11019-019-09859-4

Guleria, A., & Sood, S. K. (2023). Explainable Artificial Intelligence in Digital Therapy: A Framework for Ethical Mental Health Applications. AI & Society, 38(2), 221–235. https://doi.org/10.1007/s00146-022-01452-3

Illovsky, M. E. (1994). The Role of the Counselor as a Healing Presence. Journal of Humanistic Psychology, 34(3), 26–43. https://doi.org/10.1177/00221678940343003

Jeon, H., Park, M., & Kim, J. (2025). EEG-Based Feedback Systems for Counseling: Advancing Affective Computing. IEEE Transactions on Affective Computing, 16(1), 112–123. https://doi.org/10.1109/TAFFC.2025.3145623

Jeong, Y., Lee, S., & Bae, H. (2025). Addressing Algorithmic Bias in Multilingual AI Counseling Systems. Computers in Human Behavior Reports, 8, 100267. https://doi.org/10.1016/j.chbr.2025.100267

Joo, Y. J., Lim, K. Y., & Lee, S. J. (2021). VR-Based Simulated Clients for Counselor Training. Educational Technology Research and Development, 69(5), 2575–2592. https://doi.org/10.1007/s11423-021-09973-2

Kegerreis, S. (2020). The Psychoanalytic Implications of AI Companions. Psychoanalytic Dialogues, 30(3), 340–356.

https://doi.org/10.1080/10481885.2020.1745245

Kuhail, M., Almashaqbeh, G., & AbuShanab, E. (2024). Evaluating the Effectiveness of AI-Powered Chatbots in Delivering Cognitive Behavioral Therapy. Computers in Human Behavior, 145, 107848. https://doi.org/10.1016/j.chb.2024.107848

Liu, Q., & Liu, Y. (2022). Comparing Human and AI Counselors: A Study on Dialogue Coherence and Trust. Journal of Technology in Human Services, 40(2), 175–191. https://doi.org/10.1080/15228835.2022.2038457

Lu, Y., Zhang, R., & Xu, D. (2024). Emotion Detection from Counseling Transcripts Using Deep Learning. Neural Computing and Applications, 36(7), 5373–5388. https://doi.org/10.1007/s00521-023-08201-5

Maurya, A. (2024). Role of AI in Counselor Education: Evaluating Simulated Client Interaction. International Journal of Educational Technology in Higher Education, 21(1), 19. https://doi.org/10.1186/s41239-024-00402-6

Ofem, B., Ajayi, O., & Daniels, R. (2024). Client Experiences with AI in Online Counseling: An African Perspective. Journal of Technology in Human Services, 42(1), 30–47. https://doi.org/10.1080/15228835.2024.2039777

Ping, Y. (2024). Daily Mood Tracking with AI Chatbots: A Pilot Study on Emotional Regulation. Journal of Affective Disorders Reports, 11, 100402. https://doi.org/10.1016/j.jadr.2024.100402

Shorey, S., Ang, E., & Yap, J. (2019). A Systematic Review of Chatbot Interventions for Mental Health. Journal of Medical Internet Research, 21(5), e12432. https://doi.org/10.2196/12432

Singla, R., Varma, R., & Bansal, A. (2024). AI in Adolescent Mental Health Monitoring: Social Media Insights. Youth & Society, 56(1), 3–25. https://doi.org/10.1177/0044118X231187558

Su, H., Wang, J., & Lin, C. (2024). Emotion-Aware Teletherapy Platforms: Integrating Facial and Voice Analysis. Computers in Human Behavior Reports, 9, 100312. https://doi.org/10.1016/j.chbr.2024.100312

Webster, E. M., Ahsan, M. D., Perez, L., Levi, S. R., Thomas, C., Christos, P., Hickner, A., Hamilton, J. G., Babagbemi, K., Cantillo, E., Holcomb, K., Chapman-Davis, E., Sharaf, R. N., & Frey, M. K. (2024). AI Chatbots in Health Risk Assessment: Implications for Counseling. Journal of Clinical Informatics, 3(1), 102–115. https://doi.org/10.1200/CCI.23.00123

Whittemore, R., & Knafl, K. (2005). The integrative review: Updated methodology. Journal of Advanced Nursing, 52(5), 546–553. https://doi.org/10.1111/j.1365-2648.2005.03621.x

Zisquit, J., Greenbaum, A., & Adler, A. (2025). Trauma Survivors and AI: Emotional Containment and Perceived Empathy. Psychological Services, 22(2), 145–158. https://doi.org/10.1037/ser0000783