AI-Driven Mental Health Counseling: Opportunities, Challenges, And Ethical Implications

Ms Deepti S, and Dr Rashmi Rangaswamy[*]

[1*]MSc Clinical Psychology student, 2nd year, Department of Psychology, School of Psychological Sciences, Yeshwanthpur Campus, Christ University, Bangalore, Karnataka, India.
[1*]Assistant Professor, Department of Psychology, School of Psychological Sciences, Yeshwanthpur Campus, Christ University, Bangalore, Karnataka, India.

Abstract
Research into artificial intelligence (AI) is reshaping mental health therapy, from virtual psychotherapists to counseling robots. The COVID-19 pandemic underscored the need for digital mental health tools, accelerating the shift to virtual care. Despite a surge in mental health apps and startups, a significant gap in effective care remains, highlighting the need for innovative approaches. We advocate for next-generation technologies in counseling psychology that combine evidence-based practices with patient-centred engagement to create adaptive, scalable, and user-focused solutions. This paper examines AI's role in various aspects of mental health: AI-augmented diagnosis, personalized counseling strategies, ongoing and post-therapy support, relationship counseling, and virtual reality. We explore how AI might replicate or surpass human abilities and raise critical questions: Can AI handle the emotional complexity of mental health care, or might it oversimplify it? Will overreliance on AI diminish the personal connection crucial for effective therapy? What are the ethical concerns around data privacy, consent, and algorithmic bias? If these questions remain unresolved and the research is still evolving, how can AI already be applied to such a vital area as mental health? This analysis seeks to understand if AI can truly revolutionize mental health care while addressing these challenges through peer-reviewed articles, conference proceedings, and theses. Data will be sourced from electronic databases like PubMed, Google Scholar, citation tracking, and internet search engines. Inclusion criteria encompass all studies involving individuals with mental health disorders and interventions utilizing AI or VR, published in English. Excluded are editorials, opinion pieces, and studies unrelated to AI or VR in mental health. The results of the analysis indicate that the integration of AI holds significant promise for enhancing service efficacy and accessibility and shows potential for improving diagnostic accuracy, personalizing treatment, and enhancing therapeutic outcomes. However, challenges such as the need for more robust evidence, ethical concerns regarding data privacy, and the lack of human empathy in AI interactions highlight the importance of using AI as a complementary tool to human expertise. Addressing these challenges through rigorous research and ethical integration is essential.

Keywords: Artificial Intelligence, Mental Health, Counseling, Technology

1. Introduction

Artificial Intelligence (AI) has significantly impacted various sectors, and mental health care is no exception. As mental health issues continue to rise globally, the integration of AI offers a promising future by enhancing the efficiency, accessibility, and personalization of mental health services. This article explores the various opportunities AI presents in the mental health sector while also delving into the challenges that come with its adoption.

AI offers a transformative approach to mental health care, promising to enhance the efficiency, accessibility, and personalization of treatment. By leveraging advanced technologies such as machine learning, natural language processing, and predictive analytics, AI can provide valuable insights and support for mental health professionals and patients alike. AI's ability to process vast amounts of data and identify patterns can lead to earlier diagnoses, more effective treatment plans, and continuous monitoring, thereby improving overall mental health outcomes (Marciano et al., 2023; Abd-alrazaq et al., 2023).

Machine learning and deep learning are at the core of AI applications in mental health. These technologies can analyze large datasets to identify patterns and correlations that may not be apparent to human practitioners. For example, machine learning models can predict the onset of mental health conditions based on historical data, enabling early intervention. Deep learning algorithms can analyze complex data such as medical imaging and genomic information to uncover insights into mental health disorders (Khare et al., 2024).

Natural Language Processing (NLP) allows AI systems to understand, interpret, and generate human language. In mental health, NLP can be used to analyze text and speech for signs of mental health issues. For instance, AI systems can monitor social media posts, written diaries, and spoken interactions to detect changes in mood and behavior that may indicate

mental health problems. NLP can also enhance the capabilities of virtual therapists and chatbots, enabling them to understand and respond to patients more effectively (Oh et al., 2017).

Predictive analytics uses statistical techniques and machine learning algorithms to forecast future outcomes based on historical data. In mental health, predictive analytics can identify individuals at risk of developing mental health conditions, allowing for proactive measures to be taken. For example, predictive models can analyze patterns in electronic health records, wearable device data, and environmental factors to predict the likelihood of a patient experiencing a mental health crisis (Khare et al., 2024).

Wearable devices and the Internet of Things (IoT) can continuously monitor physiological and behavioral data, providing real-time insights into an individual's mental health. These technologies can track sleep patterns, physical activity, heart rate, and other biomarkers, alerting users and healthcare providers to potential issues. Wearable devices can also facilitate the collection of large-scale data for research and analysis, contributing to the development of more effective mental health interventions (Bourla et al., 2024).

AI has been making significant strides in the field of mental health, offering innovative solutions to enhance patient care, improve diagnostic accuracy, and provide accessible support. Examples of AI applications in mental health include AI-powered chatbots like Woebot, which provides cognitive behavioral therapy (CBT) through conversational interactions, and predictive analytics platforms like Ginger, which analyzes data to predict mental health crises and offer proactive support. AI-driven apps like Wysa offer CBT exercises and mindfulness techniques, while sentiment analysis tools like Mindstrong monitor smartphone usage patterns to detect mood changes. Virtual reality therapy, as used by Limbix, provides exposure therapy for conditions like PTSD, and AI-assisted diagnostic tools like Quartet Health improve diagnostic accuracy and facilitate access to care (Dosovitsky et al., 2024; Mehta et al., 2020).

While the potential benefits of AI in mental health care are significant, several critical questions and ethical concerns need to be addressed. Ethical and privacy concerns are paramount, as sensitive information about an individual's mental health must be handled with the utmost care to prevent misuse and protect patient confidentiality. AI systems can sometimes inherit biases present in the data they are trained on, leading to inaccurate diagnoses and treatment recommendations, particularly for marginalized groups. Ensuring that AI algorithms are trained on diverse and representative datasets is crucial to avoid perpetuating existing inequalities in mental health care. Furthermore, while AI can provide valuable support, it cannot replace the human touch that is often essential in mental health care. The therapeutic relationship between a patient and a mental health professional is built on trust, empathy, and understanding, which AI cannot fully replicate. Regulatory and standardization issues also present challenges, as the rapid advancement of AI in mental health care has outpaced regulatory frameworks. Establishing clear guidelines and standards is necessary to ensure the safe and effective use of AI in mental health (Bourla et al., 2024).

The integration of AI into mental health care presents a promising avenue for addressing the significant gaps in current care delivery systems. However, it is essential to approach this integration with caution, ensuring that AI technologies are developed and implemented in a manner that prioritizes patient well-being, ethical considerations, and the preservation of the human elements of therapy. This analysis seeks to understand if AI can truly revolutionize mental health care while addressing these challenges, ultimately aiming to provide adaptive, scalable, and user-focused solutions for individuals with mental health needs. The objective of this research is to explore the transformative potential of Artificial Intelligence (AI) in mental health care, focusing on its ability to enhance the efficiency, accessibility, and personalization of treatment. By examining various AI technologies and their applications in diagnosing, monitoring, and treating mental health conditions, this study aims to identify the opportunities and challenges associated with integrating AI into mental health services. The research seeks to provide insights into the ethical, practical, and regulatory considerations necessary for the successful implementation of AI-driven solutions in mental health care. How can Artificial Intelligence (AI) improve the efficiency, accessibility, and personalization of mental health care while addressing the ethical, practical, and regulatory challenges associated with its adoption?

2. Methods
2.1 Selection Criteria
This research will include studies that focus on the application of AI technologies, such as machine learning, deep learning, virtual reality, and chat bots, in the diagnosis and treatment of mental health disorders, relationship counseling, VR. Studies must provide empirical data on the efficacy, accuracy, or effectiveness of these AI tools, with relevant metrics like accuracy, sensitivity, specificity, user satisfaction, and symptom reduction. Only peer-reviewed studies published in English.

Research that does not involve AI technologies or focuses on non-mental health disorders will be excluded. This includes theoretical papers, opinion pieces, and studies lacking empirical data or significant methodological rigor. Unpublished manuscripts, conference abstracts, and studies with unrepresentative samples that do not provide sufficient context for understanding the generalizability of the findings will also be excluded. These criteria ensure the research maintains high quality.

2.2 Procedure
This systematic review adhered to the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines.

Figure 1

PRISMA flow diagram of selection process used in the present study
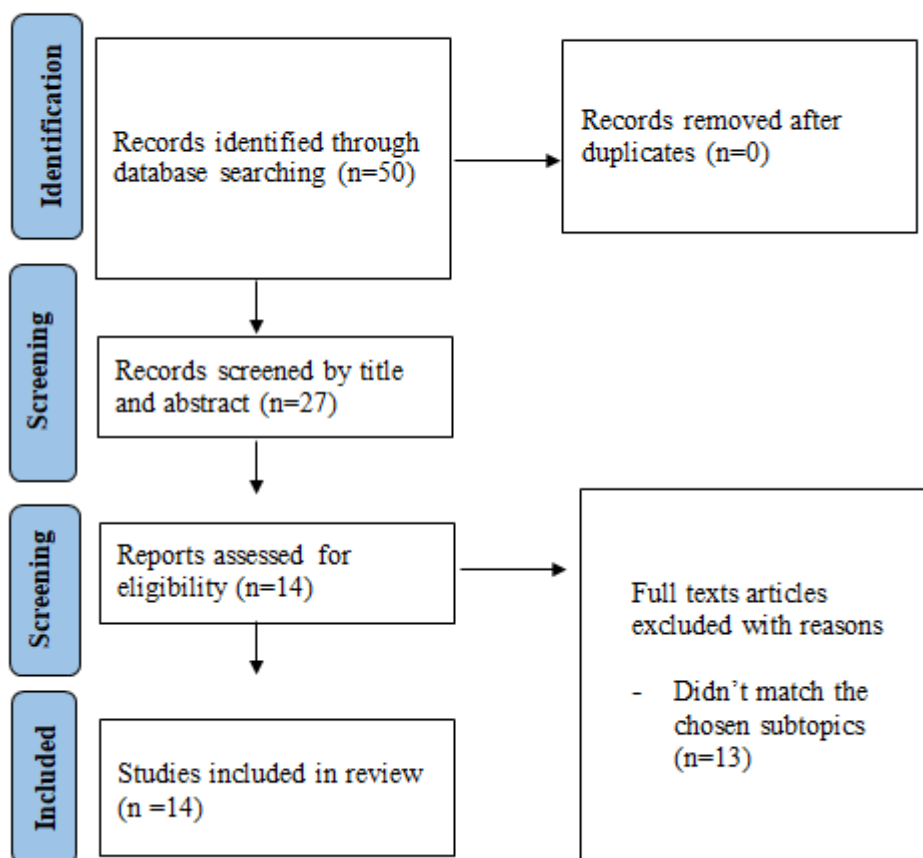
- Screening

The initial screening involved evaluating the titles and abstracts of these records to determine their relevance to the research objectives. This process led to the selection of 27 records for further assessment. The screening criteria focused on the application of AI technologies, such as machine learning, deep learning, VR, and chatbots, in the diagnosis and treatment of mental health disorders and relationship counseling. Studies are needed to provide empirical data on the efficacy, accuracy, or effectiveness of these AI tools, using metrics such as accuracy, sensitivity, specificity, user satisfaction, and symptom reduction. Or empirical data on how AI didn't help or the repercussions of it. Only peer-reviewed studies published in English were considered.

- Selection Process

From the 27 records screened, 14 reports were assessed for eligibility through a more detailed examination. This stage involved a thorough review of the full texts to ensure that each study met the inclusion criteria. Studies were included if they focused on AI's application in mental health, provided empirical data, and reported relevant metrics. Ultimately, 14 studies were included in the review, as they met all the selection criteria.

The selection process is depicted in Figure 1 (PRISMA flow diagram). No records were removed due to duplication, as the initial search did not yield duplicates. Thirteen full-text articles were excluded because they did not match the chosen subtopics or lacked empirical data relevant to the research focus.

2.3 Bias Prevention
● Maintained a neutral mindset to avoid confirmation bias, neither for nor against AI.
● Ensured consistency in evaluating each study through a standardized screening protocol.
● Established clear inclusion and exclusion criteria before the review to maintain a comprehensive and objective assessment.
● Conducted regular cross-checks to verify that selected studies adhered to the established criteria.
● Sought feedback from peers or mentors to confirm the appropriateness of the selected studies and exclusion decisions.
● Adhered strictly to these measures to ensure the review was free from personal bias

3.Review of Literature
3.1 AI-Augmented Diagnosis for Mental Disorder
Marciano et al. (2023) conducted a review focusing on the application of machine learning (ML) to improve the diagnostic and treatment capabilities for Autistic Spectrum Disorder (ASD). The study aimed to provide a comprehensive overview of current ML research in ASD, addressing the complex and heterogeneous nature of the disorder that complicates diagnosis through traditional screening tools. These tools target behaviors such as stereotypical and repetitive actions, gait, facial emotion expressions, and speech assessments. The review systematically categorized articles into various ASD research fields, including behavior, gait, speech, facial emotion expression, neuroimaging, genetics, and metabolomics. Despite not specifying the number of participants or their geographical locations, the review highlighted the potential of ML to enhance analyses in these areas, potentially simplifying and expediting the diagnostic process and enabling earlier access to therapies. The study's limitations include the unspecified number of reviewed studies and participant details. The conclusion emphasizes the increasing acceptance and implementation of AI and ML techniques, which can process large datasets and extract valuable insights, positioning ML as a promising tool for future ASD research and potentially improving screening, diagnostic, and therapeutic methods in clinical practice).

Abd-alrazaq et al. (2023) conducted an umbrella review of 15 systematic reviews to synthesize evidence on AI models' performance in diagnosing mental disorders, including AD, MCI, SCZ, BD, ASD, OCD, PTSD, and psychotic disorders, with accuracy ranging from 21% to 100%. The studies featured diverse participants from multiple high-income countries, using AI classifiers like SVM, DT, RF, CNN, and K-NN, and data types such as neuroimaging, genetic data, EEG measures, and neuropsychological tests. Evaluation metrics included sensitivity, specificity, and accuracy, though inconsistently reported. Limitations included heterogeneity in AI classifiers, inconsistent reporting, small sample sizes, and reliance on neuroimaging. The review concluded that while AI holds potential for accurate, rapid diagnoses, there is a need for better reproducibility, generalizability, and clinical validation, advocating for more systematic reviews and ethical exploration of AI diagnostic tools (Abd-alrazaq et al., 2023).

Khare et al. (2024) conducted a study investigating the application of Artificial Intelligence (AI) in diagnosing and treating psychiatric disorders such as depression, schizophrenia, bipolar disorder, Autism Spectrum Disorder (ASD), Alzheimer's Disease (AD), Attention Deficit Hyperactivity Disorder (ADHD), Post-traumatic Stress Disorder (PTSD), and Obsessive-Compulsive Disorder (OCD). The review includes various AI techniques such as Machine Learning (ML), Deep Learning (DL), and robotics, highlighting their potential to provide enhanced diagnostic accuracy, personalized treatment, and continuous monitoring capabilities. The methods utilized involved neuroimaging, electronic health records (EHRs), sensor-based tracking, and multimodal data integration to predict and categorize psychiatric disorders. Although the number of participants and their geographical locations were not specified, the study identifies significant limitations, including ethical concerns, biases, the need for rigorous regulatory review, and the absence of human empathy in AI applications. The review concludes that while AI has a transformative role in mental healthcare, further research and collaboration among healthcare professionals, technologists, and ethicists are essential to address these limitations and maximize AI's potential in mental health).

Bourla et al. (2024) explore the transformative potential of new technologies, such as digital phenotyping, imaging, and biomarkers, in understanding and managing PTSD. The study investigates the impact of technological advancements like computerized measurement tools on traditional clinical practices and professional ethics. An extensive PubMed search synthesized literature into prediction and assessment categories, involving diagnostic, screening, and monitoring methods, with data extraction and quality appraisal by two independent reviewers. The findings indicate a growing body of studies contributing to PTSD digital phenotype development using passive biometric sensor data and active data from Ecological Momentary Assessment (EMA). However, integrating AI, connected devices, and remote monitoring in PTSD care requires careful adaptation to the diverse needs of physicians and patients. The study emphasizes the importance of involving patients, caregivers, and health professionals in the design and evaluation of these tools to ensure their effectiveness and acceptance, highlighting the need for further research to address ethical and practical challenges in psychiatric practice (Bourla et al., 2024).

3.2 AI - Therapy, Post-Therapy, checking clients' progress.
The study by Mehta et al. (2020) investigates the acceptability and effectiveness of Youper, an AI-based mobile app for treating anxiety and depression. The sample comprised 4517 paying users who consented to their data being used for research. The geographic location of the participants was not specified. User ratings on a 5-star scale and retention statistics over the first four weeks characterized the app's acceptability. To examine the effectiveness, longitudinal measures of anxiety and depression symptoms were analyzed, and the cumulative regulation hypothesis was tested by predicting symptom reduction based on successful emotion regulation attempts. Results indicated that users rated the app highly (mean 4.36 stars), and 42.66% were retained by week four. Anxiety and depression symptoms showed significant decreases in the first two weeks of app use, with anxiety improvements sustained in the subsequent two weeks, while depression symptoms slightly increased. Successful emotion regulation attempts significantly predicted greater reductions in anxiety and depression symptoms. The study concludes that Youper is a low-cost, self-guided treatment accessible to users who might not otherwise obtain mental health care, demonstrating the app's acceptability and effectiveness. Limitations include the lack of geographic data and the need for further study in a randomized clinical trial to validate these findings.

This study, conducted by (Dosovitsky et al., 2024), investigates the usage patterns of a mental health chatbot named Tess, designed to provide support for depression. The study analyzed interactions of 354 users with the Tess depression modules from July 2017 to September 2018. These users engaged in text-based conversations with Tess through Facebook Messenger. The study aimed to describe the overall utilization of Tess, understand participant flow between and within the modules, and offer recommendations for chatbot design. The methods involved analyzing descriptive statistics of user interactions, including the number of messages, characters typed, completion rates, and time spent per module. Tess's algorithm is organized into distinct modules focusing on different treatment modalities. The results showed that users sent a total of 6220 messages, with large heterogeneity in engagement across different modules. This variability appeared to be influenced by the length, complexity, content, and style of questions within the modules. The study concludes that while Tess effectively engaged users, improvements in module design could enhance user experience. Limitations include the lack of demographic data and the focus on interactions within a single platform

The study by Oh, Lee, Ko, and Choi (2017) examines the application of a chatbot for psychiatric counseling in mental healthcare services, emphasizing emotional dialogue analysis and sentence generation. Conducted by researchers from the School of Computing at Korea Advanced Institute of Science and Technology (KAIST) in Daejeon, Republic of Korea, the study highlights the need for more accurate and continuous emotion recognition to enhance user satisfaction in mental health care. Although specific participant numbers and geographic locations are not provided, the study focuses on developing a conversational service that integrates methodologies such as high-level natural language understanding (NLU) and multi-modal emotion recognition. The chatbot framework includes continuous monitoring of emotional changes, personalized dialogue systems, and ethical clinical psychological responses. Results from early implementations indicated positive user reception and effectiveness in promoting self-awareness and behaviour change. However, limitations include the need for extensive corpora for training word-embedding models and the absence of long-term observational data. The study concludes that the integration of emotional intelligence and ethical response mechanisms in chatbot services could significantly improve mental health care access and treatment effectiveness.

3.3 AI-Driven Personalized Counseling Strategies.
Oh, Lee, Ko, and Choi (2017) conducted a study to investigate the potential of AI-driven chatbots in enhancing mental health treatment planning, focusing on personalized counseling and immediate intervention. The research addresses the challenge of limited access to mental health professionals due to high costs and inadequate availability, proposing chatbots as a viable solution. This model enables the chatbot to generate contextually appropriate responses by learning important keywords in a sentence. The study demonstrates the effectiveness of personalized dialogues, especially in urgent situations like suicidal impulses, where chatbots can provide immediate support and alert relevant individuals. Continuous monitoring of mood swings and emotional states is another significant benefit, aiding in the development of targeted interventions. However, limitations include potential biases in AI responses and the need for more varied and human-like interactions to avoid repetitiveness. The study concludes that AI-driven chatbots can significantly enhance treatment planning by providing real-time data and immediate support, offering valuable tools for mental health professionals in addressing diverse psychological needs. Further research is needed to validate these findings and explore the broader application of AI in mental health interventions.

Yuksel and Kocaballi (2022) explored the effectiveness of the Couple Therapy Chatbot (CTC), a conversational agent on Facebook Messenger, for CBT-based couple therapy. Drawing from prior adaptations like CBCT on Amazon Alexa, the study used a Wizard-of-Oz experiment where 12 University of Technology Sydney participants believed they were interacting with an automated system controlled by a human wizard. The CTC guided structured conversations on relationship frustrations, core beliefs, and behaviours using Ellis' ABC model of CBT. Participants appreciated the emotional outlet but found the responses too generic and lacking empathy, highlighting the need for more personalized interactions. The study concluded that while CAs can support couple therapy, response flexibility and personalization

improvements are necessary. Limitations included a small sample size, convenience sampling, and potential Wizard-of-Oz bias.

3.4 AI in relationship counseling

Troitskaya and Batkhina (2022) evaluated the effectiveness of the iCognito Relationship Program, a self-help mobile application utilizing chatbot technology, in enhancing couple relationships in Russia. The study employed a between-group experimental design with 58 female participants over two weeks, comparing the iCognito program to a bibliotherapy control condition. Relationship indicators such as satisfaction, tenderness, constructive communication, commitment, self-efficacy, communicative skills, self-esteem, and conflict levels were assessed. Results indicated significant improvements in satisfaction, tenderness, communication, commitment, self-efficacy, communicative skills, and self-esteem among iCognito users, with reduced conflict levels and medium effect sizes. Participants expressed high satisfaction with the technology and a positive attitude towards using a chatbot for personal issues. The study highlights the potential of the iCognito program to enhance relationship satisfaction and communication skills, outperforming traditional self-help books. However, limitations include the small sample size and short intervention duration, affecting generalizability and sustainability of results.

Vowels, Francois-Walcott, and Darwiche (2024) examined ChatGPT's potential in relationship counseling, focusing on its therapeutic capabilities and user experiences. The study included 20 participants (8 men, 12 women) aged 25-62 years (M = 41.8, SD = 11.2), who had been in relationships for an average of 11.2 years (SD = 8.52). Using Qualtrics and Prolific for participant recruitment and scheduling, the methodology involved a 1-hour audio-only Zoom session with ChatGPT followed by a post-interaction interview. Participants highlighted four main themes: positive future outlook, clarity in relationship issues, overly clinical interaction style, and the beneficial therapeutic setting. ChatGPT was found to provide helpful, impartial, and non-judgmental advice, though some participants desired more personalized responses. The study's limitations include a small sample size, self-reported data, and a single-session format. Future research should include larger, diverse samples and assess long-term effects. Overall, ChatGPT offers a valuable, accessible platform for relationship counseling but needs improvements in personalization.

Rathnayaka, Mills, Burnett, De Silva, Alahakoon, and Gray (2024) explored the efficacy of the iCognito Relationship Program, a self-help application for couple relationships utilizing chatbot technology. Conducted in Russia, the study involved 58 female participants and employed a between-group experimental design to compare the iCognito app with a bibliotherapy control. Over two weeks, various relationship indicators, including satisfaction, communication, and conflict levels, were assessed. Results indicated significant improvements in these indicators for the iCognito group, with medium effect sizes. Participants reported high satisfaction with the app, although some noted its limited personalization. The study concluded that the iCognito chatbot effectively enhances relationship satisfaction and communication, outperforming traditional self-help books. Limitations included the small sample size and short intervention period, necessitating further research to confirm the findings and explore long-term effects.

Rathnayaka et al. (2024) developed a Behavioral Activation (BA)-based AI chatbot for mental health support, addressing the limitations of existing chatbots by providing personalized, recurrent emotional support and remote monitoring. The study's conceptual framework was developed in three phases: reviewing existing BA studies, enriching constructs with data from online support groups, and validating with expert panels. The framework emphasizes continuous engagement, emotion detection, mood tracking, and activity planning. Pilot study results showed positive feedback on the chatbot's functionality and user interface, with participants noting improvements in mood awareness and self-care. Limitations included privacy concerns and the need for better personalization. The study concluded that the BA-based AI chatbot effectively supports mental health, but further refinements are needed for broader application.

3.5 VR in Counseling

Zhu (2022) conducted a study examining the application of virtual reality (VR) technology in psychological counseling and therapy. The research involved an advanced search on CNKI for "virtual reality and psychotherapy," yielding 171 results. Zhu thoroughly analyzed these findings, noting that VR, as an advanced human-computer interactive interface, allows users to immerse themselves in a computer-generated natural environment, creating highly realistic simulations. Initially used for preventing childhood phobias and PTSD, VR's applications have expanded to include autism, intellectual disability, and brain damage recovery. Despite some limitations, the integration of VR in mental health treatment has progressed significantly due to technological advancements. VR combined with counseling can enhance psychiatric symptom improvement, patient motivation, and limb coordination, leading to greater psychological satisfaction and improved quality of life. The study underscores VR's potential in mental health care and highlights the need for further research to address shortcomings (Zhu, 2022).

2. Discussion

The transformative potential of artificial intelligence (AI) in mental health care, illustrates both its capabilities and its limitations. AI has made significant strides in five key areas: diagnostic tools, therapy applications, chatbots, virtual reality (VR), and personalized counseling. Across these domains, AI demonstrates the ability to enhance diagnostic accuracy,

personalize treatment, and improve therapeutic outcomes. For instance, AI-powered diagnostic tools and machine learning models excel in analyzing large datasets to identify patterns that traditional methods might overlook. This capability facilitates earlier and more accurate diagnoses of mental health conditions such as Autism Spectrum Disorder (ASD), Alzheimer's Disease (AD), and Post-Traumatic Stress Disorder (PTSD).

In therapeutic contexts, AI-driven applications and chatbots provide scalable solutions that can complement traditional therapies by offering real-time support and feedback. Virtual reality has emerged as a promising tool for immersive therapy, especially for exposure-based treatments of conditions like phobias and PTSD. AI-driven personalized counseling strategies can tailor interventions to individual needs, delivering immediate support and tracking emotional states to guide therapeutic processes effectively.

However, the literature also reveals several challenges and limitations associated with AI in mental health care. A significant concern is the need for more robust, generalizable evidence to confirm the effectiveness of AI tools. Many studies face issues such as small sample sizes, lack of demographic diversity, and inconsistent evaluation metrics, which limit the reliability and applicability of findings. Ethical concerns are also prominent, particularly regarding data privacy and the risk of AI systems perpetuating biases present in their training data. The absence of human empathy in AI interactions poses another challenge, as these tools cannot fully replicate the nuanced, personal interactions essential to effective therapy.

Overall, while AI shows considerable promise in enhancing mental health care, it is crucial to address these challenges through rigorous research, ethical considerations, and careful integration into clinical practice. AI should be viewed as a complementary tool rather than a replacement for human therapists, with ongoing refinement and adaptation needed to ensure that these technologies meet the evolving needs of patients and practitioners.

4.1 Ethics – How to use AI

Artificial intelligence (AI) holds promise as a valuable support tool in mental health services, educational guidance, and career counseling. However, the American Counseling Association (ACA) warns that AI should not replace human counselors (ACA, 2024). The ACA's AI Working Group has issued guidelines to help understand the value and limitations of using AI tools in mental health.

4.1.1 Guidelines and Considerations:
- Technical Shortcomings and Biases: AI algorithms can inherit biases and fallibilities from their human creators, risking culturally insensitive care and providing false or inaccurate information (Fulmer, 2024).
- Privacy and Confidentiality: Ensure AI tools comply with federal and state privacy laws to protect client confidentiality (ACA, 2024).
- Professional Judgment: AI cannot holistically consider a client's complex personal history and cultural context. It should only be used as a supportive tool, not as a replacement for professional counselors (ACA, 2024).
- Crisis Response: Avoid using AI for crisis response. Instead, use crisis hotlines and emergency services (ACA, 2024).
- Informed Use: Clients should be informed about what AI can and cannot offer to make educated decisions about incorporating it into their counseling (Fulmer, 2024).

4.1.2 Government Regulation:

Executive Order on AI Regulation: President Biden signed an Executive Order to establish new standards for AI safety and security, protect privacy, advance equity and civil rights, and promote innovation. This order aims to govern AI technology to mitigate risks, including worker displacement, data privacy issues, and national security concerns (White House, 2023).

4.2 Conclusion

The integration of artificial intelligence (AI) into mental health care, educational guidance, and career counseling holds significant potential to enhance the efficacy and accessibility of these services. However, the literature underscores the critical need to approach AI with caution, adhering to ethical guidelines and professional standards. While AI tools can support diagnostic accuracy, therapeutic interventions, and client engagement, they must be used as adjuncts to, rather than replacements for, human expertise. Ensuring that AI applications are free from biases, uphold client privacy, and are designed with client involvement is essential for maintaining trust and effectiveness in counseling practices. Moreover, ongoing training for counselors and transparent AI practices are pivotal to addressing the ethical challenges and ensuring that AI benefits are maximized while minimizing risks. By embracing AI's potential within a framework of rigorous ethical considerations and professional judgment, mental health and counseling practices can evolve to better meet the diverse needs of clients, enhancing overall care and support in a responsible and informed manner.

4.3 Limitations:
- Limited generalizability due to the small sample sizes in included studies.
- Inconsistent evaluation metrics across different studies, making comparisons difficult.

- Possible biases in AI algorithms due to the nature of training data.
- Challenges in maintaining a neutral stance to avoid confirmation bias.
- Necessity of rigorous research to validate the efficacy of AI tools in mental health care.
- Focused on only a few areas of counseling, leaving other counseling areas needing further exploration.
- Limited existing research on certain AI applications in counseling, highlighting the need for more comprehensive studies.

4.4 Recommendations

- Continuous Training: Counselors should receive ongoing training in AI applications, focusing on confidentiality and privacy (Williams, 2024).
- Client-Centered Design: Involve clients and counselors in AI tool design to ensure real-world relevance (ACA, 2024).
- Competence in AI: Providers must understand AI fundamentals, subfields, and applications before use (Fulmer, 2024).
- Understand AI Fundamentals: Learn about algorithms, AI in daily life, and subfields like machine learning and NLP (D'Alfonso, 2020).
- Evaluate Technologies: Remain open to and critically evaluate new technologies before incorporating them (Fulmer, 2024).
- Maintain Human Element: Ensure AI enhances but does not replace, the human aspect of counseling.
- Address AI Bias: Recognize and mitigate potential biases in AI tools (Fulmer et al., 2021).
- Understand Automation Impact: Stay informed on how AI affects the job market and career counseling practices (Bessen et al., 2020).
- Advocate for Transparency: Promote transparent AI algorithms that are accessible, interpretable, and maintainable (Yudkowsky & Bostrom, 2011).
- Informed Consent: Inform clients about AI tools and obtain explicit consent.
- Ensure Data Privacy: Use AI tools that comply with privacy laws and ensure data security.
- Empower Clients: Encourage clients to discuss their use of AI tools in counseling.
- Use AI in Supervision: AI can enhance supervision by analyzing supervisee performance.
- Complement Clinical Judgment: Avoid relying solely on AI for diagnosis; use it to support, not replace, clinical expertise.
- Conduct Research: Engage in research on AI's impact on counseling practices and its effectiveness across diverse populations.

References

3. Abd-Alrazaq, A., Alajlani, M., Alhuwail, D., Schneider, J., Al-Kuwari, S., Shah, Z., Hamdi, M., & Househ, M. (2020). Artificial intelligence in the fight against COVID-19: Scoping review. JMIR. Journal of Medical Internet Research/Journal of Medical Internet Research, 22(12), e20756. https://doi.org/10.2196/20756

4. Abd-Alrazaq, A., Alhuwail, D., Schneider, J., Toro, C. T., Ahmed, A., Alzubaidi, M., Alajlani, M., & Househ, M. (2022). The performance of artificial intelligence-driven technologies in diagnosing mental disorders: an umbrella review. Npj Digital Medicine, 5(1). https://doi.org/10.1038/s41746-022-00631-8

5. Bourla, A., Mouchabac, S., El-Hage, W., & Ferreri, F. (2018). e-PTSD: an overview on how new technologies can improve prediction and assessment of Post-traumatic Stress Disorder (PTSD). European Journal of Psychotraumatology, 9(sup1). https://doi.org/10.1080/20008198.2018.1424448

6. Dosovitsky, G., Pineda, B. S., Jacobson, N. C., Chang, C., Escoredo, M., & Bunge, E. L. (2020). Artificial Intelligence chatbot for depression: Descriptive Study of usage. JMIR Formative Research, 4(11), e17065. https://doi.org/10.2196/17065

7. GeeksforGeeks. (2024, July 2). AI in Mental Health Opportunities and Challenges. GeeksforGeeks. https://www.geeksforgeeks.org/ai-in-mental-health-opportunities-and-challenges/

8. Google Scholar. (n.d.). https://scholar.google.com/scholar?hl=en&as_sdt=0%2C5&q=AI-Augmented+counseling+Therapy++for+Rare+Mental+Disorders&btnG=

9. KeQue Zhu "The use of VR technology in counseling and therapy", Proc. SPIE 12329, Third International Conference on Artificial Intelligence and Electromechanical Automation (AIEA 2022), 1232911 (7 September 2022); https://doi.org/10.1117/12.2646793

10. Marciano, F., Venutolo, G., Ingenito, C. M., Verbeni, A., Terracciano, C., Plunk, E., Garaci, F., Cavallo, A., & Fasano, A. (2021). Artificial Intelligence: the "Trait D'Union" in Different Analysis Approaches of Autism Spectrum Disorder Studies. Current Medicinal Chemistry, 28(32), 6591–6618. https://doi.org/10.2174/0929867328666210203205221

11. Mehta, A., Niles, A. N., Vargas, J. H., Marafon, T., Couto, D. D., & Gross, J. J. (2021). Acceptability and Effectiveness of Artificial Intelligence Therapy for Anxiety and Depression (YOUPER): longitudinal observational study. JMIR. Journal of Medical Internet Research/Journal of Medical Internet Research, 23(6), e26771. https://doi.org/10.2196/26771

12. Recommendations for practicing counselors and their use of AI. (n.d.). www.counseling.org. https://www.counseling.org/resources/research-reports/artificial-intelligence-counseling/recommendations-for-practicing-counselors
13. WISH. (2024, May 5). The Digital Mental Health Revolution: Transforming Care through Innovation and Scale-Up - WISH. https://wish.org.qa/research-report/the-digital-mental-health-revolution-transforming-care-through-innovation-and-scale-up/