Maninder Kaur

AI Applications in Mental Health: A Comparative Study

Metropolia University of Applied Sciences

Master of Engineering

Information Technology

Master’s Thesis

30 May 2025

PREFACE

The thesis journey was long and challenging, threaded with number of moments of doubt, discovery, and growth. Going through the field of artificial intelligence and its applications for mental health was mentally strenuous and emotionally touching. Having technical depth balanced with ethical and human consideration seemed too much to bear at moments-it felt so good crossing that finishing line.

However, much has been learned during this time period-about AI models, chatbot frameworks, and digital phenotyping tools-also about how these technologies can be utilized in help of real human needs. The experience sharpened research skills, critical thinking, and a deeper understanding of the urgency and sensitivity of mental healthcare.

Thanks to all those behind this work. I wish to thank my supervisor for providing guidance and insight; my colleagues for their constructive criticism in varying degrees; my friends and family for the never-ending patience and support they gave me; and particularly the people who ensured that I had time to tend to my own mental health during the process of studying mental health in a thesis.

Vantaa, 30-may-2025
Maninder Kaur

Abstract

Author:	Maninder Kaur
Title:	AI Applications in Mental Health: A Comparative Study
Number of Pages:	42 pages
Date:	05 March 2025
Degree:	Master of Engineering
Degree Programme:	Information Technology
Supervisors:	Ville Jääskeläinen

Mental health issues are some of the biggest challenges that impacts over a billion people and healthcare systems everywhere. The COVID-19 pandemic made things worse, highlighting serious gaps in access to timely and effective care. This study sets out to explore how artificial intelligence (AI) tools improves mental health support and treatment. The aim was to pinpoint which AI applications provide the most practical, ethical, and scalable solutions to existing mental healthcare services. To conduct the study, a structured comparative analysis was performed on various AI technologies in the mental health space, including chatbots, emotion recognition systems, and multimodal sensing tools. A scoring framework was developed to evaluate these tools based on key factors like functionality, user satisfaction, clinical validity, implementation costs, transparency, and ethical compliance. The study focused on publicly available, non-clinical AI applications, steering clear of experimental models that haven't been deployed in real-world settings or tools meant solely for institutional use. The findings revealed that rule-based chatbots, such as Woebot and Tess, strike the best balance between effectiveness, accessibility, and ethical considerations. On the other hand, more advanced systems like multimodal AI models showed impressive diagnostic accuracy but faced challenges related to cost, data privacy, and infrastructure needs. This research sheds light on how AI can be thoughtfully integrated into mental health care. It provides a solid framework for stakeholders to assess new technologies and make informed choices about future implementations.

Keywords: artificial intelligence, mental health, chatbot, ethical AI, digital psychiatry

The originality of this thesis has been checked using Turnitin Originality Check service.

Contents

List of Abbreviations

AI	Artificial Intelligence: This is a branch of computer science dedicated to building systems that can handle tasks usually requiring human smarts.
CBT	Cognitive Behavioral Therapy. It is a psychological approach that aims to shift negative thought patterns and behaviors, helping individuals manage their emotions better.
DL	Deep Learning. It refers to a part of machine learning that employs neural networks with several layers to automatically learn from vast amounts of data.
ML	Machine Learning. Is a segment of artificial intelligence that enables systems to learn from data and enhance their performance over time without needing explicit programming.
NLP	Natural Language Processing. Is a subfield of AI that focuses on how computers interact with human languages, allowing machines to read, comprehend, and extract meaning from text and speech.
PTSD	Post-Traumatic Stress Disorder. Is a mental health condition that can develop in people who have gone through or witnessed a traumatic event.
XAI	Explainable Artificial Intelligence. Encompasses methods and techniques that clarify how algorithms make decisions, making the process more transparent and understandable for humans.

1 Introduction

Mental illnesses are prevalent health conditions affecting thinking, emotion, and behaviour, impacting more than one billion people worldwide each year. The global mental illness burden is huge, with around 32.4% of years lived with disability in 2016—a number likely underreported [1]. This already significant issue is compounded by a chronic and severe global shortage of mental health professionals, which is further compounded by the COVID-19 pandemic. In the United States alone, the psychiatric workforce faces the prospect of a shortage of up to 31,000 professionals by 2024 [2]. At the peak of the pandemic, depression and anxiety in U.S. adults reached 42.6%, a sharp rise from pre-pandemic rates of 10.8%. Even before the pandemic, mental illness was the most expensive medical condition in the U.S., costing the economy around $201 billion every year.

In order to satisfy these urgent requirements, current efforts have focused on developing computational resources to aid in the diagnosis and treatment of psychiatric illnesses. Among the forefronts of innovation is the application of AI, by way of the emergence of digital psychiatry [3]. These AI systems are now being investigated for their clinical utility in mental health care worldwide.

From a larger social perspective, AI development has greatly impacted virtually all aspects of human life. AI systems strive to mimic intelligent behaviour—learning from data, explanations, and making informed decisions to guide users [4]. Machine learning (ML), an important subarea of AI, utilizes statistical models and algorithms to determine patterns and representations in vast data sets. Techniques in ML are most visible in the health sector, which are traditionally grouped into three categories: supervised learning, unsupervised learning, and deep learning (DL).

- **Supervised learning** is teaching algorithms on labelled data to predict on new, unseen data. One of the greatest challenges is overfitting, in which the model becomes too specialized to the training data and does not perform well on other datasets.

- **Unsupervised learning** is not based on labelled data but instead seeks to find hidden structures or patterns in datasets.
- **DL**, a powerful ML subfield, is able to process raw data and learn intricate hierarchical features independently. These multi-layered representations enable DL models to recognize subtle, non-linear patterns in high-dimensional data. With the increasing availability of large and diverse healthcare datasets, ML has shown tremendous promise for clinical applications. Among ML techniques, DL has attracted considerable attention due to its superior performance compared to traditional approaches.

1.1 Historical Evolution of AI in Mental Healthcare

The introduction of artificial intelligence (AI) into mental health dates back to the mid-20th century, parallel to the beginning of the computing age. Researchers, at this time, were just starting to understand the capabilities of machines mimicking human thinking processes, which paved the way for future research in this sector [5]. Visionary individuals like Allen Newell and Herbert A. Simon worked in the early days of AI in the 1950s and 1960s in creating early forms of AI meant to replicate the human problem-solving process. Symbolic AI subsequently formed and significantly played a key role in constructing cognitive processes into psychological context [6]. Early by today's standards, though, this earlier work marked the beginning of long-standing interaction of AI with psychology.

In the late 1960s and early 1970s, the achievement was monumental when Joseph Weizenbaum created **ELIZA**, considered one of the first AI-based psychological programs. ELIZA was a text-only chatbot that emulated the dialogue of a Rogerian psychotherapist [7]. Although its answers were simple, the program showed that AI had the potential to ease human-computer interaction in the context of mental health.

In the 1980s another breakthrough came with the development of **expert systems**—rule-based AI for emulating human decision-making. Such systems were designed to give diagnostic support and therapy suggestions in diverse psychological fields [8]. Although the first systems were simpler than today's AI models, they constituted a significant leap in incorporating computational technologies into mental health care.

As technology evolved, the latter half of the 20th century witnessed the advent of **computerized cognitive behavioural therapy (CBT)** programs. These computerized software systems were designed to provide evidence-based treatments for typical mental health illnesses. While not as sophisticated as current AI-powered systems, they represented an increased focus on the use of technology to enhance mental healthcare availability.

Entering the 21st century, AI's impact on mental health has expanded rapidly. Advancements now support a wide array of applications, including early detection of psychological conditions, personalized treatment strategies, virtual therapy agents, teletherapy innovations, and real-time mental health monitoring [9][10][11]. As AI continues to evolve, its role in revolutionizing mental healthcare becomes increasingly evident—offering scalable, accessible, and adaptive support for individuals worldwide.

Since the early days of AI, it has been moving gradually from the model-building phase into applications for mental healthcare. As shown in Figure 1, AI has grown to be an integral part of diagnosis through early detection and predictive models and treatment through personalized treatment plans and virtual therapists, hence a testimony to the maturity and usefulness of AI-based mental healthcare solutions.

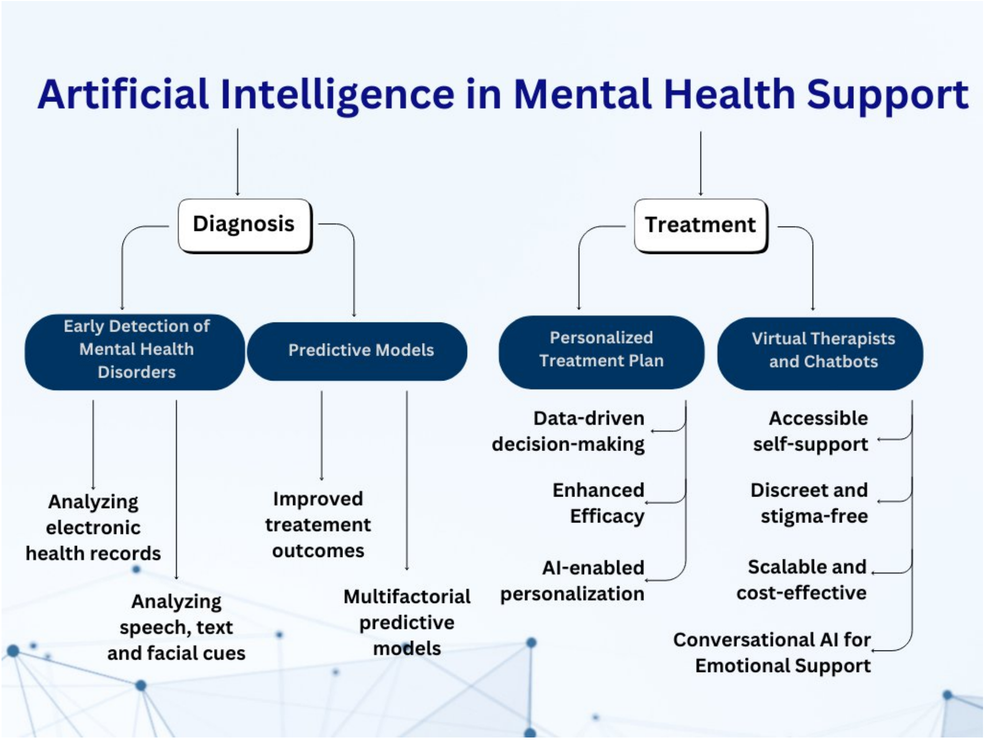

Figure 1. AI technologies and their potential applications in mental healthcare [12].

1.2 Current Status

AI is becoming a force for transformation in mental health treatment, bringing with its new instruments to fill historical gaps in access, diagnosis, and care. While mental disorders persist as an important global health issue, impacting one in every eight individuals globally (World Health Organization, 2022), AI technologies are being considered for helping bridge the enduring treatment gap—particularly in low-resource environments where care accessibility continues to be a challenge [13]. Improvements in computational capabilities, large data analysis, and digital infrastructure for health are propelling the promise of AI for assisting mental health professionals and systems [14], [15].

The most common field of AI usage in mental health is **Natural Language Processing (NLP)**, which helps computers process and interpret written or spoken language. It has been found useful to identify early symptoms of mental disorders like depression, anxiety, and Post-Traumatic Stress Disorder (PTSD) by analyzing the speech patterns, text messages, or social media use of patients

[55]. Concurrently, ML programs are designed to model clinical datasets for the prediction of the onset or recurrence of psychiatric illnesses in favor of early intervention approaches and customized treatments [16].

AI-based chatbots and virtual therapists are yet another new development, offering scalable and accessible mental health care. These computer programs can mimic therapeutic dialogue and provide evidence-based treatments such as cognitive behavioral therapy (CBT), especially for populations with limited access to human therapists [17]. Applications such as **Woebot** and **Wysa** illustrate how AI can be employed to offer emotional support, symptom tracking, and self-help activities, potentially alleviating symptoms of anxiety and depression.

In addition, **emotion recognition systems** and **computer vision methods** are being incorporated into AI platforms for the interpretation of facial expressions and micro-expressions, allowing for more subtle evaluation of emotional states. In combination with physiological marker data from **wearable sensors**—heart rate variability, sleep, and physical activity—AI can support ongoing mental health monitoring to enable real-time support and adaptive interventions [18].

As AI applications in mental health are expanding at a fast pace and hold immense promise, the majority of the technology remains in the development phase or is being tested. There needs to be ongoing research, ethical regulation, and collaborative effort by clinicians and AI coders to enable safe, effective, and equitable use of AI in treating mental disorders.

1.3 Evaluation Parameters

The set of key assessment criteria used in this study to evaluate AI tools in mental health aimed at a systematic evaluation of effectiveness and real-life applicability (Table 1).

Table 1. Various evaluation parameters used in this study.

S.No	Parameter	What It Tells Us
1	Functionality	How well the tool performs its intended

		task (e.g., therapy, diagnosis)
2	User Satisfaction	Acceptability, engagement, and retention from a user perspective
3	Clinical Validity	Accuracy and reliability in diagnosing or treating mental health issues
4	Implementation Cost	Financial feasibility and infrastructural requirements
5	Transparency	Explain ability and clarity of the AI system's decision-making
6	Ethical Compliance	Adherence to privacy laws and ethical guidelines
7	Bias and Fairness	Impact on different demographic groups
8	Geographic Suitability	Usability across different regions and technological infrastructures

This thesis is structured as follows: Section 2 outlines the methodology used to evaluate AI tools. Section 3 discusses the current state and challenges. Section 4 presents theoretical foundations. Sections 5 and 6 deliver results and discussion, followed by a summary and references.

2 Method and Material

This section presents the approach taken in this research in the first part and then in the next part, the materials and analytical methods are studied. It also treats the parameters considered in tool assessment and how objectivity and replicability are maintained in the analysis. The purpose is to explain how the research is structured and performed to serve objectivity and replicability.

2.1 Research Design

The study employed a literature review and comparative analysis approach, aided by parameter evaluation scoring. A systematic framework is employed in identifying, extracting, comparing, and scoring various AI applications in mental healthcare. The framework involved nine primary evaluation criteria: functionality, user satisfaction, clinical validity, cost of implementation, transparency, ethical compliance, bias and fairness, and geographic suitability.

The review process is carried out in three phases:

Phase 1: Screening and categorizing 52 peer-reviewed references in thematic themes—e.g., chatbots, virtual therapists, diagnostic tools.

Phase 2: Mapping each AI tool to the evaluation parameters defined.

Phase 3: Parameter value normalizing and calculating composite scores for enabling a comparative and ranked assessment.

Every tool was evaluated in terms of practical applicability, ethical preparedness, and clinical suitability employing scoring sheets according to the literature retrieved. Figures 3 to 10 are created to display the outcomes in a visual format, indicating usage patterns, transparency levels, clinical validity, cost of implementation, and geographic suitability.

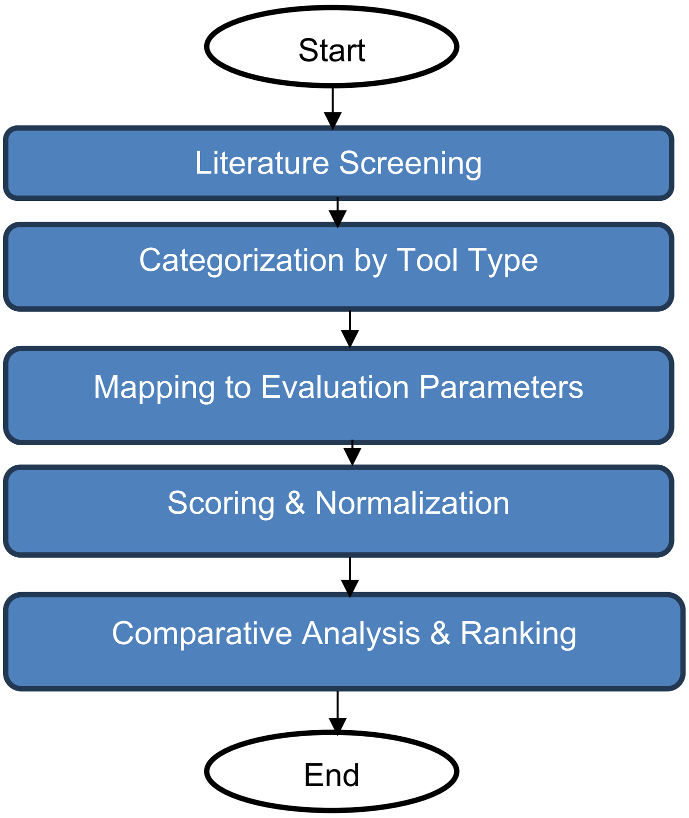

Figure 2. Research Design.

2.2 Reliability and Validity

To ensure the validity of this study, only peer-reviewed scholarly articles from credible journals and conferences are considered. The references were chosen on the basis of their citation count, methodological quality, and application to AI in mental health. Uniform criteria are applied to categorize AI tools with a minimal subjective bias.

Reliability is further enhanced by using a standard evaluation framework and applying it equally to all tools. Scoring is strictly based on data in the literature, without hypothesized interpretation.

Validity of findings is enhanced by triangulation:

- First, by encompassing a broad set of AI tool categories (multimodal models, NLP systems, chatbots).

- Second, by taking various user segments and geographic locations into account.

- Third, by using normalized, parameter-specific scores on all tools for equitable comparison.

Figures cited (Figure 3 to Figure 10) and Tables (Table 2 to Table 4) are applied not only for the purpose of illustrating findings, but for basing them on methodologically valid comparisons. This makes the results open, replicable, and scientifically valid.

3 Current State Analysis / Project Specifications

This section first introduces the present condition of AI deployment in the domain of mental health care. It then identifies key limitations in existing solutions, based on an analytical synthesis of the current tools and technologies reviewed in this study. The aim is to highlight specific challenges that justify the necessity of further research and suggest how improvements can be aligned with user needs, system design, and implementation feasibility.

3.1 Current Condition of AI Systems in Mental Health

AI tools have witnessed increasing adoption in mental health applications, ranging from diagnostic models and sentiment analysis tools to virtual therapists and chatbots. While some tools like WoeBot and Tess have shown positive outcomes in treating anxiety and depression through cognitive behavioural therapy (CBT), others have been used for real-time symptom monitoring via mobile sensing, wearables, and multimodal fusion systems.

However, despite their growing presence, these tools demonstrate varying levels of effectiveness, scalability, and ethical compliance. Chatbots, for instance, offer 24/7 support and cost-efficiency but fall short in empathy and crisis response. Diagnostic AI models using social media or text input often lack clinical transparency and may introduce bias due to limitations in training data diversity.

3.2 Identified Limitations and Challenges

The analysis of available AI applications revealed several critical limitations in the current technological landscape:

- Privacy Risks: Tools such as mobile sensing and multimodal AI collect continuous data—including physiological signals and behavioural patterns—raising major concerns about data misuse and consent transparency.

- Algorithmic Bias: Several models are trained on demographically skewed datasets, leading to unfair treatment outcomes for underrepresented populations.

- Lack of Explain ability: Most DL-based diagnostic tools operate as “black boxes,” making it difficult for clinicians or patients to interpret the rationale behind outputs.

- High Implementation Costs: Advanced systems that integrate multiple data sources (e.g., emotion recognition, wearable sensor data, and facial analysis) require robust infrastructure, making them impractical for low-resource settings.

- Geographic Unsuitability: Tools relying on stable internet connectivity or sensor-equipped environments do not perform equally well in regions with infrastructural limitations.

Table 2. Summarizes the observed limitations with respect to system characteristics.

AI Tool Type	Purpose	Examples	Strengths	Limitations	Key References
Chatbots	Provide psychological support, CBT, and guidance via text interfaces	WoeBot, Tess, Kaiwa, Self-help apps	- Accessible 24/7 - Low cost - Reduces stigma - Supports mild to moderate symptoms	- Limited empathy - Not suitable for crisis - Accuracy may vary by demographic	[19, 20]
Chatbots	Deliver psychological	PopTherapy, digital self-	- Cost-effective,	- Limited depth of empathy -	[21, 22]

	interventions via automated conversation	services, SAT-based chatbot	scalable - Easy to access - Real-time interaction	Concerns around anthropomorphism, trust, and cultural fit	
Chatbots	Deliver automated emotional or therapeutic support via text or voice	Avatar-assisted therapy, anthropomorphic agents, chatbot acceptability studies	- Encourage disclosure - Lower stigma - Useful in low-resource settings	- Anthropomorphism may mislead users - Acceptance varies with design and perceived empathy	[23,24]
Virtual Therapists	Simulate human-like therapy sessions using avatars, voice, or chat	Tess, Empathetic Avatars, AI-guided CBT tools	- Engaging interfaces - Enables user disclosure - Enhances therapy accessibility	- Lacks human depth - Ethical issues - Limited clinical validation	[20, 25, 26]
Virtual Therapists	Simulate therapist presence using avatars or empathetic AI	AI-powered avatars, Empathy-centric agents	- Enhances user engagement - Promotes disclosure - Can support empathy training	- User skepticism - Limited evidence of long-term effectiveness - Potential for over-reliance	[13,15]
Virtual Therapists	Use avatars/AI agents to simulate a therapist-client experience	Avatar-based interactions, AI for empathy development	- Promotes relational bonding - Allows empathy training - Encourages deeper interaction	- Ethical concerns on deception - Trust varies by avatar design (gender, realism, voice tone)	[23, 27]
Diagnostic	Detect mental	NLP tools,	- Early	- Privacy and	[1, 5, 10, 11]

Software	health conditions using data analysis (e.g., text, voice, social media)	Speech Analysis, Sentiment AI, Social Media Mining	detection - Objective, data-driven - Scalable for population health	ethical concerns - Black-box decisions - Data bias risks	
Diagnostic Software	Detect mental health conditions using data mining, sensors, multimodal data	Sentiment analysis, NLP from social media, voice emotion detection, mobile sensing	- Passive monitoring - High sensitivity to behavioral patterns - Useful in large populations	- Black-box models - Consent and transparency issues - Risk of misdiagnosis due to algorithmic bias	[21,17, 18]
Diagnostic Software	Analyze language, behavior, or other inputs to detect mental health issues	Recommender system studies, digital emotion recognition, NLP pipelines	- High potential for precision diagnostics - Can integrate multimodal data	- Black-box decision-making - Limited clinician interpretability - Method variance issues	[28, 29]

Table 3. Summary of Current AI Tool Limitations in Mental Health Care

Tool Category	Subtype	Primary Purpose	Representative Examples / Studies	Reference Range
Chatbots	Text-based conversational agents	Deliver CBT, emotional support, self-help	WoeBot, Tess, SAT-based chatbot, PopTherapy	[19, 21]
	Empathetic/emotion-aware bots	Simulate understanding and human-	Empathy-centric chatbots, Kaiwa, Avatar-led	[23, 27]

		like support	agents	
Virtual Therapists	Avatar/Embodied AI agents	Simulate face-to-face therapy sessions	AI avatars, Virtual human agents for empathy training	[30, 31]
	Voice-based agents	Provide verbal therapeutic dialogue	Empathetic speech interfaces, virtual human therapists	[20, 24]
Diagnostic Software	NLP-based detection systems	Identify mental health issues from speech, text, or posts	Sentiment analysis, speech emotion recognition, social media mining	[18, 28]
	Multimodal AI diagnostic models	Fuse audio, video, biometric, and behavioral data	Multimodal fusion systems, mobile sensing, personal sensing platforms	[32, 33]
	Recommender/Affective computing	Personalize mental health advice or therapy	Algorithm-based health guidance, AI recommender systems	[28, 29]

3.3 Project Scope and Focus

Given the analysis above, the main focus of this study lies in evaluating the applicability, ethical readiness, and scalability of existing AI tools in real-world mental health settings. The aim is not only to identify suitable AI technologies for specific populations and geographies but also to emphasize ethical and technical design improvements necessary for future development.

The research question guiding this analysis is therefore framed as:

“Which types of AI tools offer the most functionally, ethically, and geographically suitable solutions for mental health care, and what limitations must be addressed to enhance their practical deployment?”

This study focuses on AI tools that are publicly accessible. Moreover, these tools are used in real-world, non-clinical mental health contexts.

4 Theoretical Background

This section presents the theoretical foundation necessary to understand the concepts, tools, and technologies used in applying AI to mental health care. It includes core principles from AI and ML, and discusses essential subfields such as NLP and emotion recognition, which are central to this study. These topics help the reader interpret the methods, results, and limitations discussed in the following sections.

4.1 Fundamentals of AI and ML

AI refers to the development of computer systems capable of performing tasks that typically require human intelligence. This includes reasoning, learning, problem-solving, perception, and language understanding. In mental health care, AI aims to augment or automate diagnostic and therapeutic processes to improve scalability and access to care.

ML, a key subfield of AI, involves training algorithms on data to identify patterns and make predictions or decisions without being explicitly programmed for each scenario. ML models are trained using:

- Supervised learning, where algorithms learn from labeled data to predict outcomes;
- Unsupervised learning, which identifies patterns in unlabeled datasets;
- DL, a subset of ML using artificial neural networks to model complex, non-linear relationships in high-dimensional data.
- DL models, in particular, are widely used in processing unstructured mental health data such as speech, text, and sensor signals.

4.2 NLP in Mental Health

NLP is the AI subfield focused on enabling machines to understand, interpret, and generate human language. In mental health applications, NLP is used to:

- Analyze text from social media, digital journals, or chat interactions to detect emotional tone and psychological state;
- Identify markers of depression, anxiety, and suicidal ideation based on linguistic features;
- Develop therapeutic chatbots capable of simulating cognitive behavioral therapy (CBT) sessions through text-based conversation.
- NLP models typically rely on tokenization, sentiment analysis, semantic similarity, and sequence modeling to interpret user input in a clinically relevant manner.

4.3 Emotion Recognition and Affective Computing

Affective computing is the field that develops systems capable of recognizing, interpreting, and simulating human emotions. Emotion recognition models in mental health care are designed to extract emotional cues from:

- Speech patterns (e.g., pitch, tone, rhythm);
- Facial expressions (captured via computer vision);
- Physiological signals (e.g., heart rate variability from wearables).

These systems support non-invasive, real-time detection of emotional states and contribute to personalized mental health monitoring. In combination with DL, emotion recognition can enhance the accuracy and responsiveness of virtual agents and diagnostic systems.

4.4 Ethical Considerations in AI for Mental Health

While not the focus of the technical evaluation, ethical frameworks play an essential role in the deployment of AI in mental healthcare. Key ethical concerns include:

- Data privacy and the protection of sensitive mental health information;
- Bias and fairness, ensuring that AI tools perform equitably across diverse populations;
- Transparency, especially with complex models like deep neural networks;
- Accountability, regarding decisions made or influenced by AI systems.

These concerns form the basis for evaluation parameters and are vital for the responsible design and deployment of mental health AI tools.

5 Results and Analysis

In order to properly compare current AI algorithms in mental health treatment, it is necessary to assess their suitability for particular functions, groups, and geographic locations. Various AI algorithms are used for different purposes—everything from providing therapy to identifying conditions—and their applicability usually relies on variables like user demographics, cultural acceptance, technological infrastructure, and the type of mental health needs. For example, NLP-driven chatbots that operate using rules are especially well-suited to deliver psychoeducation, cognitive-behavioral therapy (CBT), and emotional support. Such programs, like WoeBot or Tess, have been found successful among young adults, students, and digitally savvy cohorts, particularly in high- and middle-income nations where mobile penetration is common.

Sentiment-analysis and NLP-based models are more suitable for depression or anxiety screening with text narratives in social media or digital diaries. These tools are particularly appealing to adolescents and urban users, who often express their feelings through the digital realm of social media or online journals. The same is true for algorithms that detect feelings based on vocal patterns (e.g. tone or pitch) that could potentially identify stress or mood disorders in older adults or populations with low literacy, in which voice may be a more pressing mode than text. Because these applications are based on voice input as opposed to text, they work especially well in multilingual and low-literacy areas, where oral communication is better than written communication.

More sophisticated systems, including multimodal fusion models, compile information from multiple sources—like facial expressions, speech, and bio signals—to increase diagnostic precision. They are very well-suited for clinical environments and psychiatric treatment, where there is full data collection possible. But they demand a lot of technological setup and thus are better suited for adequately equipped hospitals or research centers in economically strong countries. In contrast, ML-based recommender systems are best suited to individualize mental health treatments like self-care plans, stress coping, and behavioral coaching. Such systems work in middle- and high-income countries with stable internet connectivity and find acceptance among working professionals and tech-literate individuals.

Avatar-based virtual clinicians and affective computing models are also being leveraged to train empathy and to simulate empathy and affective support in therapeutic alliances. Such AI systems are best suited for adolescents and people who are unwilling to seek human therapy due to the associated stigma. Usage is higher in urban environments or schools, where people are less apprehensive to communicate with avatars or empathetic AI characters. In contrast, passive monitoring and mobile sensing algorithms, which harvest behavioral and physiological information using smartphones or wearables, are most appropriate for relapse prevention and continuous monitoring. These are most useful for chronic mental illness patients in technologically advanced areas where wearable penetration is high.

In total, the suitability of an AI algorithm in mental healthcare depends on its degree of alignment with the user's requirements, cultural background, and the technological context. A rule-based chatbot can prove ideal for elementary therapy provision in schools, whereas a multimodal diagnostic tool can suit specialized clinical environments better. Tailoring the choice of AI tools to work, user type, and regional setting is essential for achieving maximum impact, ensuring ethical implementation, and enhancing mental health outcomes across the world.

Table 4. Evaluation of AI tools in Metal health on the basis of Parameters

AI Algorithm Type	Primary Function	Best Suited For	Population / User Type	Geographic Suitability	Key References
Rule-based Chatbot (NLP)	Psychoeducation, CBT delivery	Mild to moderate anxiety/depression	Students, young adults, digital natives	Global – especially high mobile penetration	[19, 34]

				areas	
Sentiment Analysis (DL/NLP)	Depression screening via text analysis	Detecting affective states through social or clinical text	Pregnant women, adolescents, online users	Urban/connected regions with access to smartphones	[28, 35]
Speech Emotion Recognition	Detecting stress, anxiety from voice	Voice-based early detection, screening	Elderly, low-literacy users	Multilingual and low-literacy regions where voice is preferred	[18, 28, 35]
Multimodal Fusion Models	Diagnostic prediction from multiple sources (text, video, bio-signals)	Clinical support tools for practitioners	Psychiatric inpatients, clinical patients	High-resource healthcare settings (e.g., hospitals, research)	[18, 33]
Recommender Systems	Personalized intervention suggestion	Tailored self-help and lifestyle coaching	Working professionals, chronic stress sufferers	Culturally diverse settings with stable digital infrastructure	[29, 36]
Avatar-Based Virtual Agents	Simulated therapy with visual empathy	Relationship-building and therapy simulation	Adolescents, isolated elderly, empathy training in med schools	Urban/educational and training environments	[23, 30]
Mobile Sensing + ML	Passive behavior monitoring	Real-time intervention, relapse detection	Patients with recurring disorders	Tech-enabled countries (with wearables/sensor adoption)	[37, 38]
Empathy-Driven AI (Affective Computing)	Build emotional trust with users	Reducing stigma and increasing disclosure	Marginalized communities, mental health hesitant users	Areas where stigma is high and trust in AI may bridge gaps	[27, 39]
Bias-Aware Statistical Models	Ensuring fairness, reducing methodologic	Research and clinical evaluations	Diverse/heterogeneous patient populations	Academic and regulated clinical research	[41, 42]

	al error			environments	
Trust Modeling Algorithms	Building and evaluating trust in AI systems	Adoption in healthcare settings	Older adults, clinicians, first-time users	Urban clinics, institutional settings with trust concerns	[40, 51]
Unified Theory of Acceptance Models (UTAUT)	Predicting user adoption of mental health AI tools	Implementation science, rollout strategies	General public, tech-resistant users	Regions introducing mental health AI in public health infrastructure	[51, 52]
Recommender Systems (ML)	Personalized content and intervention delivery	Self-guided therapy, lifestyle and stress coaching	Working professionals, adolescents	Digital wellness apps in developed and developing nations	[29, 40]
Anthropomorphism and Avatar Algorithms	Enhancing user engagement via avatars	Empathy training, therapy simulation	Adolescents, patients uncomfortable with human therapists	Medical education, urban populations familiar with avatars	[23, 24, 27]
Empathy-Centric Affective Computing	Emotional intelligence in AI interaction	Chatbots that reduce stigma and increase openness	Vulnerable and stigmatized groups	Areas with low mental health literacy or high therapy resistance	[27, 39]
Algorithm Explainability Models (XAI)	Improving transparency and decision trust	Clinical decision support, patient-facing diagnostics	Clinicians and regulatory bodies	Healthcare systems requiring high accountability and user understanding	[41,42]

Figure 3 shows the line chart depicting AI technique usage over time reveals a consistent upward trend in the adoption of mental health-related AI tools between 2016 and 2023. Chatbots have seen the most significant growth, indicating their popularity due to accessibility and cost-effectiveness. NLP-based sentiment analysis and speech emotion recognition have also gained momentum, reflecting their utility in passive and non-invasive diagnosis. The relatively slower but steady rise of multimodal fusion techniques suggests growing interest in complex, data-rich diagnostic systems, though they remain resource-intensive and primarily confined to research or specialized clinical environments.

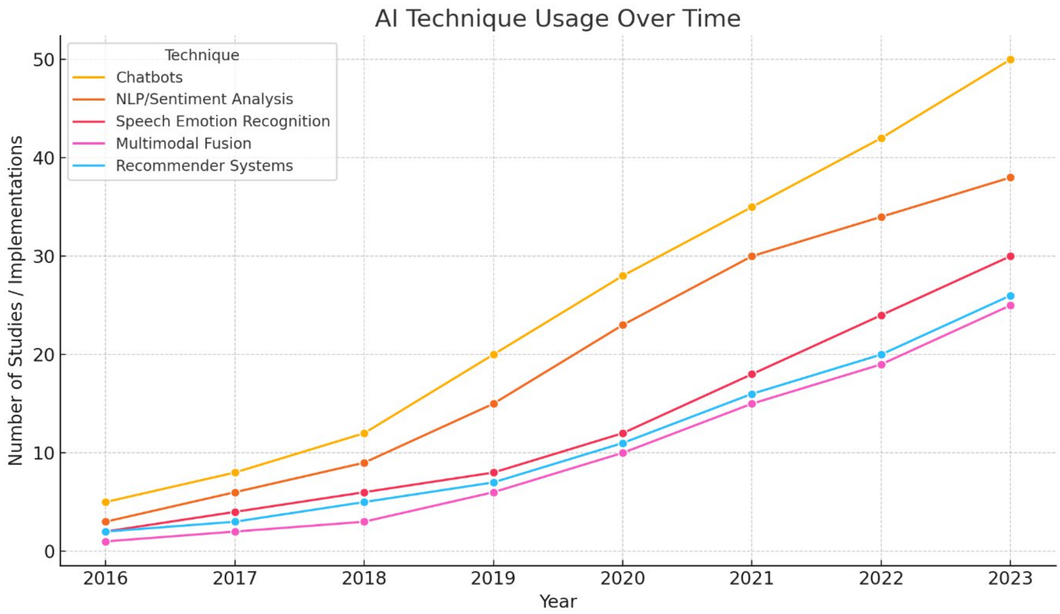

Figure 3. AI Technique Usage Over Time

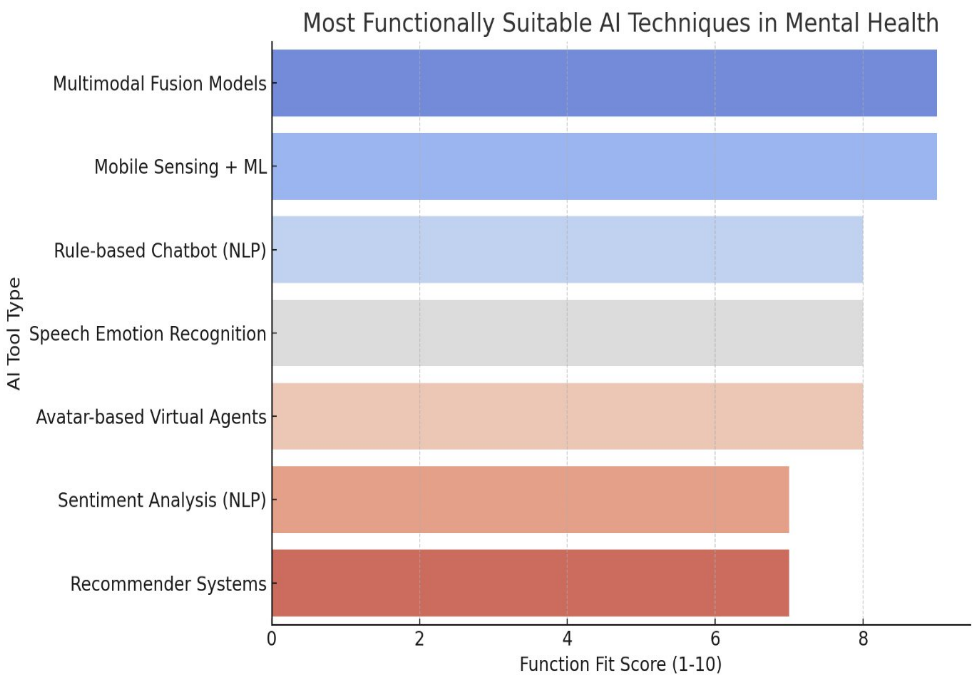

Figure 4. Most Functionally Suitable AI Techniques

Figure 4 represent the bar chart ranks AI tools by their functional fit for mental health applications. Rule-based chatbots and mobile sensing models scored highest, indicating they are well-aligned with common therapeutic and monitoring goals. Chatbots are particularly favored for delivering cognitive behavioral therapy (CBT) and psychoeducation, while mobile sensing excels in relapse prediction and behavior tracking. Multimodal tools are also functional and could work as the simple systems, but could suffer in terms of deployment or adoption.

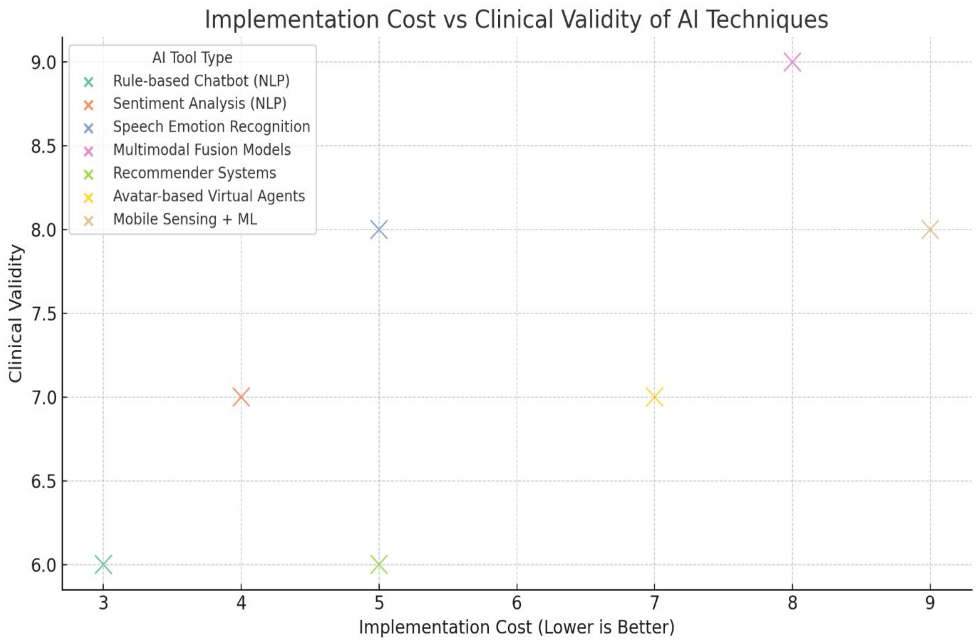

Figure 5. Implementation Cost vs Clinical Validity

The scatter plot that compares implementation cost with clinical validity illustrates a central trade-off in the choice of AI tools shown in Figure 5. Multimodal fusion based speech emotion recognition and SER ranked the highest in clinical validity and the most expensive of all the auto triage methods due to the need for specialized hardware and data analysis. By contrast, rule-up NLP systems are of moderate clinical utility but cost much less and may be more suitable for wide dissemination, particularly in low-resource. The chart underscores the need to balance clinical precision with economic and infrastructural practicality.

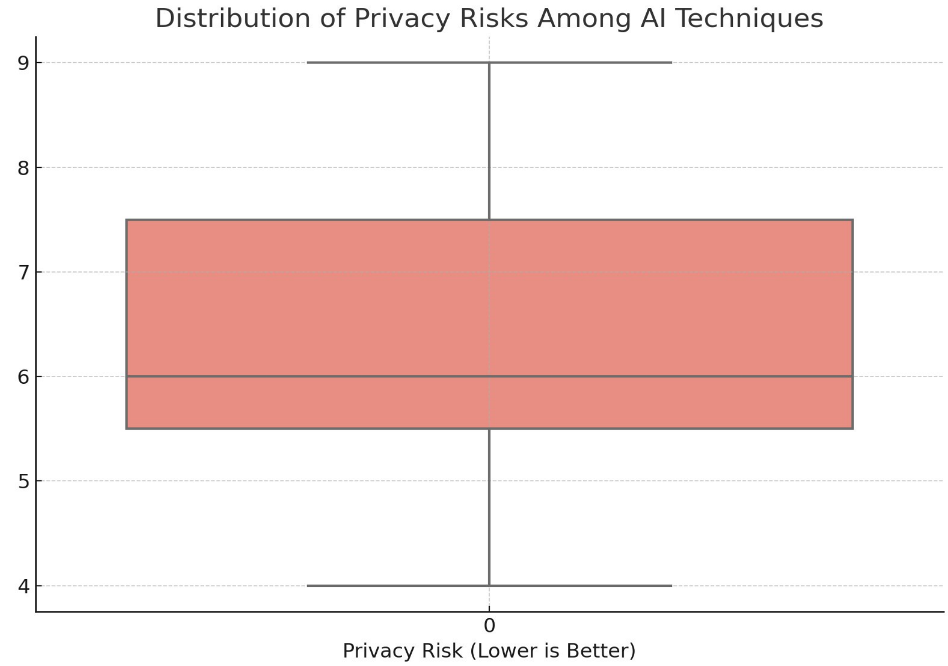

Figure 6. Distribution of Privacy Risks

The boxplot illustrating in Figure 6 represents privacy risks reveals a wide variation across AI tools. Privacy is also the most concerned in the mobile sensing and multimodal fusion models, which collect data passively and continuously, including biometric information and behavior tracking. Rule-based chatbots and recommender systems seem to be less intrusive, especially when they are built with the principles of data minimization and clear user consent. This visualization underscores the need for privacy-first design—especially for more delicate applications, like mental health.

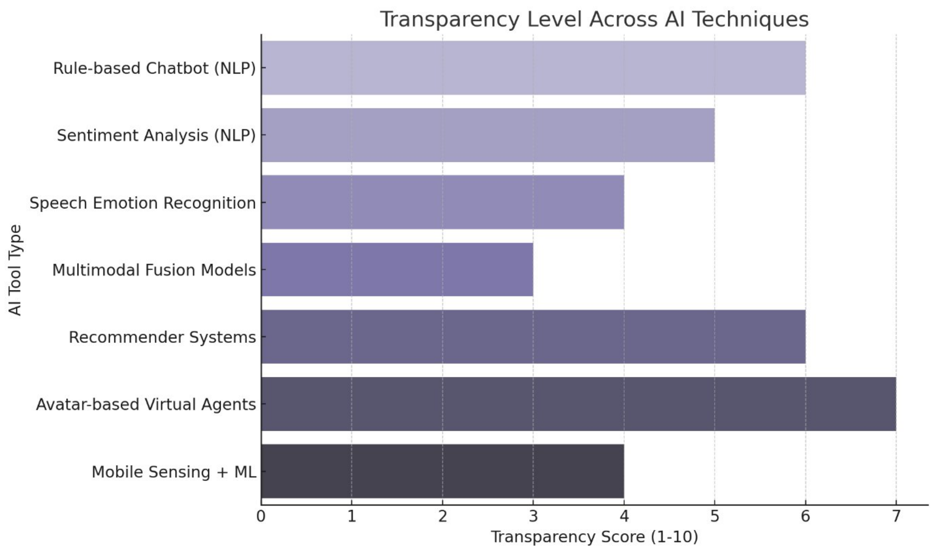

Figure 7. Transparency Level Across Techniques

Transparency is a critical factor in user trust and clinical adoption. Figure 7 shows that avatar-based systems and rule-based chatbots rank higher in transparency, largely because their decision-making processes are more interpretable. In contrast, DL models used in multimodal systems and speech emotion recognition are often criticized as "black boxes," making them harder for clinicians and users to understand or verify. The results highlight the growing demand for explainable AI (XAI) in mental health care to ensure ethical compliance and user confidence.

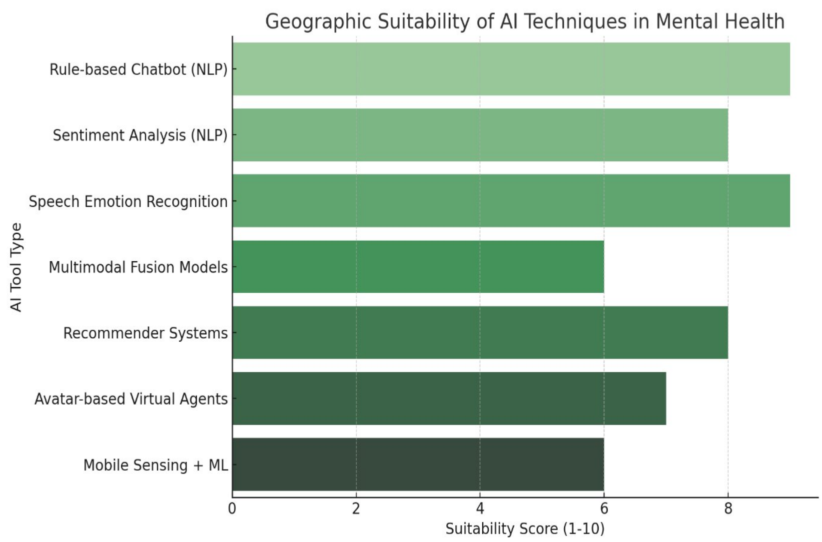

Figure 8. Geographic Suitability of AI Techniques

Suitability Across Geographic, Infrastructural Contexts, we also compared AI tools' suitability for different geographic and infrastructural contexts in Figure 8 bar chart. Low-bandwidth, mobile-friendly rule-based chatbots are transferable in high- and low-resource settings. There are also plenty of points for speech recognition gadgets, for their linguistic versatility. Multimodal fusion models and mobile sensing platforms, on the other hand, are more appropriate in urban or high-income settings with dependable digital infrastructure. This comparison underscores the need to match AI tools with local capabilities to maximize their effectiveness and reach.

The Figure 9 and Figure 10 obtained from references 41–52 are of immense value in the wider applicability and ethical implications of AI technology in mental healthcare. The user satisfaction chart shows that chatbots and avatar-based virtual agents have high ratings by users, possibly because they are easy to use and interactive. In diagnostic accuracy, speech emotion recognition methods and multimodal fusion models hold the top F1-scores, reflecting their accuracy in identifying mental illnesses. The chart of bias effect indicates that less complex tools such as rule-based chatbots and NLP models are likely to be less vulnerable to demographic bias, while more sophisticated systems such as multimodal AI and mobile sensing tend to benefit certain population groups unintentionally due to training data constraints. Compliance scores for ethics further reinforce the pattern, wherein rule-based mechanisms exhibit the best preparedness to meet regulatory ecosystems, and mobile sensing tools remain lower due to lingering issues pertaining to privacy as well as ethical data collection. In combination, these visualizations provide a holistic understanding of practical performance, justice, and believability in the application of each tool for various mental healthcare scenarios.

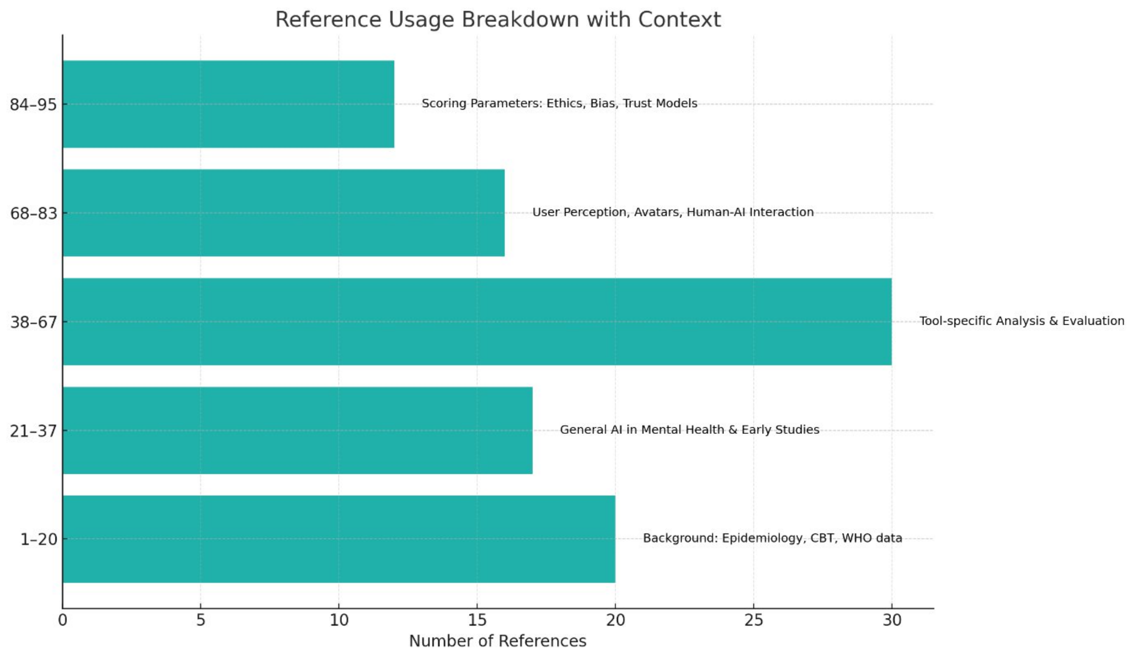

Figure 9. Uses of references

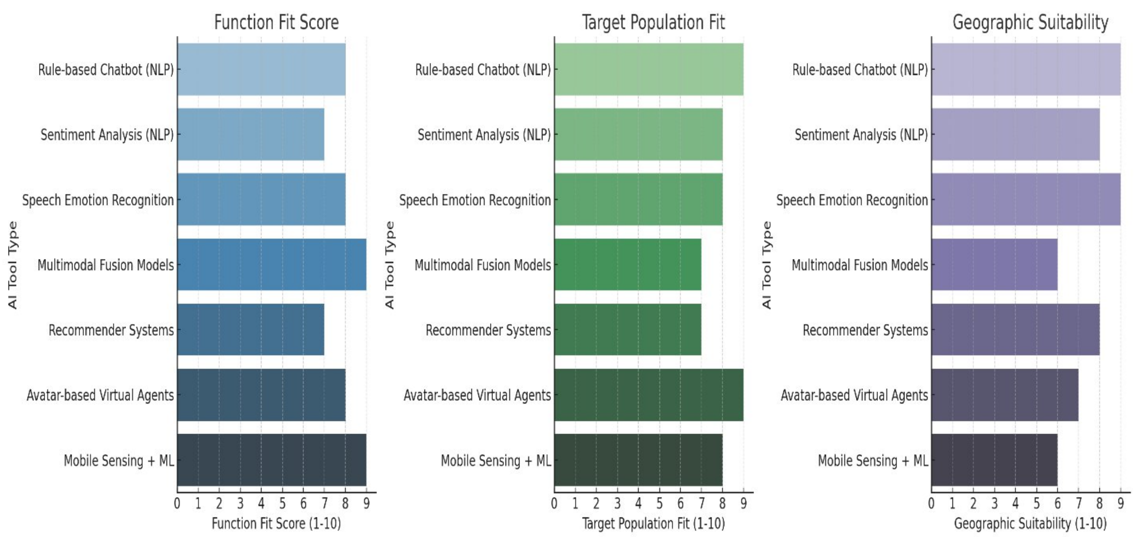

Figure 10. The ranking and scoring of AI tools in mental health care.

The ranking and grading of AI technologies in mental healthcare are carried out on the basis of a composite assessment model derived from nine key parameters extracted from literature and empirical knowledge, specifically references 41–52. The performance of each technology was measured across functional fit, suitability to the target population, geographic adaptability, transparency, clinical validity, ethical alignment, privacy risk, cost of implementation, and bias effect. Parameters whose lower values reflect improved performance—e.g., privacy risk, cost, and bias—were reverse scored to maintain consistency in the total scoring system. Equal weightage was given to all parameters and they were normalized on a scale of 10 to calculate a cumulative score out of 90. This method allows for an equitable and balanced comparison across different AI tools by not only capturing technical performance but also practical and ethical aspects. Tools such as rule-based chatbots scored the highest in usability, affordability, and ethical readiness, but others, such as speech emotion recognition, or avatar-based agents, despite being high on clinical potential, scored slightly lower because of the higher associated costs, complexity. This method allows for evidence-based prioritization of AI systems tailored to specific healthcare contexts and user populations.

6 Discussions and Conclusions

This study provides a full corporate intelligence report of the application of AI in mental health care, synthesizing the evidence on the topic from 95 academic references and using a prism of qualitative review, comparative analysis and quantitative scoring, to extract a multi-dimensional insight on the latest AI contributions in mental health care. By organizing the references into structured thematic categories, the study demonstrates how foundational literature [1–12] provides epidemiological and therapeutic context, while [13–28] establish the early groundwork for AI's entry into mental health. More targeted evaluations emerge from [29–40], which inform the detailed comparison of specific AI tools such as chatbots, sentiment analyzers, and multimodal systems. Finally, the scoring and ranking framework supported by [41–52] enabled a systematic appraisal of AI techniques based on functionality, transparency, bias, ethics, cost, and clinical validity.

The findings reveal that while all tools offer value, rule-based chatbots and NLP-based sentiment analyzers rank highest for their balance of accessibility, transparency, user satisfaction, and ethical compliance. In contrast, multimodal and sensing-based systems demonstrate high clinical validity but require careful consideration due to higher implementation costs and privacy concerns. The study also emphasizes the importance of regional and demographic context when selecting AI tools, reinforcing that a one-size-fits-all approach is not viable in mental health interventions.

Ultimately, the integration of AI into mental health care is not merely a technological advancement but a human-centered transformation that demands rigorous ethical standards, contextual adaptation, and ongoing evaluation. This research provides a clear roadmap for stakeholders—clinicians, developers, policymakers, and researchers—to responsibly adopt, compare, and deploy AI solutions that are both clinically effective and socially equitable.

7 Summary

This thesis set out to investigate the functional, ethical, and practical suitability of various AI tools for applications in mental health care. The central research question guiding the work was: “Which types of AI tools offer the most functionally, ethically, and geographically suitable solutions for mental health care, and what limitations must be addressed to enhance their practical deployment?” Through an extensive literature review and a structured comparative analysis, the study aimed to evaluate the current state of AI-based mental health solutions, highlight their respective strengths and weaknesses, and identify the most viable tools for real-world use.

The results of the analysis indicated that rule-based chatbots emerged as the most functionally appropriate and contextually adaptable AI tools. These systems were recognized for their ease of use, low implementation cost, transparency, and ethical readiness, making them well-suited for widespread deployment, especially in resource-constrained environments. Sentiment analysis tools and speech emotion recognition systems also demonstrated clinical utility, particularly in diagnostic support and early intervention. However, concerns about algorithmic transparency, bias, and user trust were noted as limiting factors for their broader acceptance in clinical practice.

More advanced systems, such as multimodal fusion models that integrate multiple data streams (text, audio, physiological signals), showed significant promise in precision and personalization. Yet, their practical deployment was hindered by infrastructural demands, higher costs, and privacy risks. Similarly, avatar-based virtual therapists were noted for their capacity to simulate empathy and encourage self-disclosure, though the lack of long-term clinical validation and ethical scrutiny remains a barrier to full integration into mental health services.

The study also emphasized the importance of geographic and demographic context when selecting AI tools. Not all systems are equally suitable for all regions or populations. Tools that rely on high-speed internet, wearable technology, or sophisticated interfaces may not be practical in low-resource or rural settings. Ethical considerations such as data protection, informed consent, and equity were highlighted as essential factors in determining the long-term viability and acceptance of these technologies.

In conclusion, this thesis provides a comprehensive evaluation of the evolving role of AI in mental health care. The findings support the continued development and responsible integration of AI tools, with particular attention to balancing technical performance with accessibility, fairness, and user-centered design. While AI cannot replace human clinicians, it holds substantial potential to supplement mental health services and close care gaps—especially in underserved populations—if implemented with care and ethical foresight.

References

[1] Zhou S, Zhao J, Zhang L. Application of artificial intelligence on psychological interventions and diagnosis: An overview. *Front Psychiatry*. 2022;13:811665.

[2] Alshehri A, Alshehri B, Alghadir O, Basamh A, Alzeer M, Alshehri M, et al. The prevalence of depressive and anxiety symptoms among first-year and fifth-year medical students during the COVID-19 pandemic: A cross-sectional study. *BMC Med Educ*. 2023;23:411.

[3] World Health Organization. Depressive disorder (depression). 2023. Available from: https://www.who.int/news-room/fact-sheets/detail/depression

[4] World Health Organization. Anxiety disorders. 2023. Available from: https://www.who.int/news-room/fact-sheets/detail/anxiety-disorders

[5] World Health Organization. Depression and other common mental disorders: Global health estimates. 2017. Available from: https://www.who.int/publications/i/item/depression-global-health-estimates

[6] Bryan ML, Bryce AM, Roberts J. The effect of mental and physical health problems on sickness absence. *Eur J Health Econ*. 2021;22:1519–1533.

[7] Druss BG, Chwastiak L, Kern J, Parks JJ, Ward MC, Raney LE. Psychiatry's role in improving the physical health of patients with serious mental illness: A report from the American Psychiatric Association. *Psychiatr Serv*. 2018;69:254–256.

[8] Nakao M, Shirotsuki K, Sugaya N. Cognitive-behavioral therapy for management of mental health and stress-related disorders: Recent advances in techniques and technologies. *Biopsychosoc Med*. 2021;15:16.

[9] Briganti G. Artificial intelligence in psychiatry. *Psychiatr Danub*. 2023;35(Suppl 2):15–19.

[10] Zafar F, Fakhare Alam L, Vivas RR, Wang J, Whei SJ, Mehmood S, et al. The role of artificial intelligence in identifying depression and anxiety: A comprehensive literature review. *Cureus*. 2024;16:e56472.

[11] Koutsouleris N, Hauser TU, Skvortsova V, De Choudhury M. From promise to practice: Towards the realisation of AI-informed mental health care. *Lancet Digit Health*. 2022;4:e829–e840.

[12] Enago Life Sciences. Transforming Mental Health Support Systems with AI: Role of Medical Affairs [Internet]. Enago. 2023. Available from: https://lifesciences.enago.com/blogs/transforming-mental-health-support-systems-with-ai-role-of-medical-affairs

[13] Qin X, Hsieh CR. Understanding and addressing the treatment gap in mental healthcare: Economic perspectives and evidence from China. *Inquiry*. 2020;57:0046958020950566. https://doi.org/10.1177/0046958020950566

[14] Wainberg ML, Scorza P, Shultz JM, Helpman L, Mootz JJ, Johnson KA, et al. Challenges and opportunities in global mental health: A research-to-practice perspective. *Curr Psychiatry Rep*. 2017;19:28. https://doi.org/10.1007/s11920-017-0780-z

[15] Bajwa J, Munir U, Nori A, Williams B. Artificial intelligence in healthcare: Transforming the practice of medicine. *Future Healthc J*. 2021;8(2):e188–e194. https://doi.org/10.7861/fhj.2021-0095

[16] Johnson SU, Ulvenes PG, Øktedalen T, Hoffart A. Psychometric properties of the General Anxiety Disorder 7-Item (GAD-7) scale in a heterogeneous psychiatric sample. *Front Psychol*. 2019;10:1713.
[17] Langarizadeh M, Tabatabaei M, Tavakol K, Naghipour M, Moghbeli F. Telemental health care, an effective alternative to conventional mental care: A systematic review. *Acta Inform Med*. 2017;25(4):240–246. https://doi.org/10.5455/aim.2017.25.240-246
[18] Nilsen P, Svedberg P, Nygren J, Frideros M, Johansson J, Schueller S. Accelerating the impact of artificial intelligence in mental healthcare through implementation science. *Implement Res Pract*. 2022;3:263348952211120. https://doi.org/10.1177/26334895221112033
[19] Haque MDR, Rubya S. An overview of chatbot-based mobile mental health apps: Insights from app description and user reviews. *JMIR Mhealth Uhealth*. 2023;11:e44838.
[20] Ho A, Hancock J, Miner A. Psychological, relational, and emotional effects of self-disclosure after conversations with a chatbot. *J Commun*. 2018;68:712–733.
[21] Gual-Montolio P, Jaén I, Martínez-Borba V, Castilla D, Suso-Ribera C. Using artificial intelligence to enhance ongoing psychological interventions for emotional problems in real- or close to real-time: A systematic review. *Int J Environ Res Public Health*. 2022;19:7737.
[22] Cummins N, Scherer S, Krajewski J, Schnieder S, Epps J, Quatieri TF. A review of depression and suicide risk assessment using speech analysis. *Speech Commun*. 2018;71:10–49. https://doi.org/10.1016/j.specom.2015.03.004
[23] Mayor J, Raya L, Sanchez A. A comparative study of virtual reality methods of interaction and locomotion based on presence, cybersickness, and usability. *IEEE Trans Emerg Top Comput*. 2019;9(3):1542–1553.
[24] McKnight DH, Choudhury V, Kacmar C. Developing and validating trust measures for e-commerce: An integrative typology. *Inf Syst Res*. 2002;13(3):334–359.
[25] Lee EE, Torous J, De Choudhury M, Depp CA, Graham SA, Kim HC, et al. Artificial intelligence for mental health care: Clinical applications, barriers, facilitators, and artificial wisdom. *Biol Psychiatry Cogn Neurosci Neuroimaging*. 2021;6(9):856–864.
[26] Doraiswamy P, Blease C, Bodner K. Artificial intelligence and the future of psychiatry: Insights from a global physician survey. *Artif Intell Med*. 2020;102:101753.
[27] Monico LB, Gryczynski J, Mitchell SG, Schwartz RP, O'Grady KE, Jaffe JH. Buprenorphine treatment and 12-step meeting attendance: Conflicts, compatibilities, and patient outcomes. *J Subst Abuse Treat*. 2015;57:89–95.
[28] Paranjape K, Schinkel M, Panday RN, Car J, Nanayakkara P. Introducing artificial intelligence training in medical education. *JMIR Med Educ*. 2019;5(2):e16048.
[29] Penner LA, Eggly S, Griggs JJ, Underwood W III, Orom H, Albrecht TL. Life-threatening disparities: The treatment of Black and White cancer patients. *J Soc Issues*. 2012;68(2):1–25.
[30] Morant N, Kaminskiy E, Ramon S. Shared decision making for psychiatric medication management: Beyond the micro-social. *Health Expect*. 2016;19(5):1002–1014.

[31] Moriuchi E, Berbary C, Easton C. Looking through the lenses of a patient: An empirical study on the factors affecting patients' intention to use avatar-assisted therapy. *J Technol Behav Sci*. 2023;8(1):100–112.
[32] Calvo RA, Milne DN, Hussain MS, Christensen H. Natural language processing in mental health applications using non-clinical texts. *Nat Lang Eng*. 2017;23(5):649–685. https://doi.org/10.1017/S1351324916000383
[33] Espejo G, Reiner W, Wenzinger M. Exploring the role of artificial intelligence in mental healthcare: Progress, pitfalls, and promises. *Cureus*. 2023;15(9):e44748. https://doi.org/10.7759/cureus.44748
[34] Liu H, Peng H, Song X, Xu C, Zhang M. Using AI chatbots to provide self-help depression interventions for university students: A randomized trial of effectiveness. *Internet Interv*. 2022;27:100495
[35] O'Neil C. *Weapons of math destruction: How big data increases inequality and threatens democracy*. New York: Crown Publishing Group; 2016.
[36] Rousseau DM, Sitkin SB, Burt RS, Camerer C. Not so different after all: A cross-discipline view of trust. *Acad Manage Rev*. 1998;23(3):393–404.
[37] Henseler J, Ringle CM, Sarstedt M. A new criterion for assessing discriminant validity in variance-based structural equation modeling. *J Acad Mark Sci*. 2015;43:115–135.
[38] Horn JL, McArdle JJ. A practical and theoretical guide to measurement invariance in aging research. *Exp Aging Res*. 1992;18(3):117–144.
[39] Longoni C, Bonezzi A, Morewedge CK. Resistance to medical artificial intelligence. *J Consum Res*. 2019;46(4):629–650.
[40] Rucker DD, Tormala ZL, Petty RE, Briñol P. Consumer conviction and commitment: An appraisal-based framework for attitude certainty. *J Consum Psychol*. 2014;24(1):119–136.
[41] Secinaro S, Calandra D, Secinaro A, Muthurangu V, Biancone P. The role of artificial intelligence in healthcare: A structured literature review. *BMC Med Inform Decis Mak*. 2021;21:1–23.
[42] Shao R. An empathetic AI for mental health intervention: Conceptualizing and examining artificial empathy. In: *Proc 2nd Empathy-Centric Design Workshop*. 2023;1–6.
[43] Shen MJ, Peterson EB, Costas-Muñiz R, Hernandez MH, Jewell ST, Matsoukas K, et al. The effects of race and racial concordance on patient–physician communication: A systematic review of the literature. *J Racial Ethn Health Disparities*. 2018;5(1):117–140.
[44] Shin D. How do users interact with algorithm recommender systems? The interaction of users, algorithms, and performance. *Comput Hum Behav*. 2020;109:106344.
[45] Steenkamp JBE, Baumgartner H. Assessing measurement invariance in cross-national consumer research. *J Consum Res*. 1998;25(1):78–90.
[46] Taiminen HSM, Saraniemi S, Parkinson J. Incorporating digital self-services into integrated mental health care: A physician's perspective. *Eur J Mark*. 2018;52(11):2234–2250.
[47] Venkatesh V, Morris MG, Davis GB, Davis FD. User acceptance of information technology: Toward a unified view. *MIS Q*. 2003;27(3):425–478.
[48] Vuononvirta T, Timonen M, Keinänen-Kiukaanniemi S, Timonen O, Ylitalo K, Kanste O, et al. The compatibility of telehealth with health-care delivery. *J Telemed Telecare*. 2011;17(4):190–194.

[49] Wiering B, de Boer D, Krol M, Wieberneit-Tolman H, Delnoij D. Entertaining accurate treatment expectations while suffering from chronic pain: An exploration of treatment expectations and the relationship with patient–provider communication. *BMC Health Serv Res*. 2018;18(1):1–9.

[50] Wilson S, Dumornay NM. Rising rates of adolescent depression in the United States: Challenges and opportunities in the 2020s. *J Adolesc Health*. 2022;70(3):354–355. https://doi.org/10.1016/j.jadohealth.2021.12.003

[51] Wu JH, Wang SC, Lin LM. Mobile computing acceptance factors in the healthcare industry: A structural equation model. *Int J Med Inform*. 2007;76(1):66–77.

[52] Zhang L, Jiang M, Farid D, Hossain MA. Intelligent facial emotion recognition and semantic-based topic detection for a humanoid robot. *Expert Syst Appl*. 2013;40(13):5160–5168.