

HealHive: A Secure Anonymous AI-Driven Mental Health Support Platform

SACHIN HIRANWALE¹, SWARAJ SHELAR², SATYA PRAGYAN BARIK³, SUCHIT PARTE⁴,
VANSH ASNANI⁵

^{1, 2, 3, 4, 5}Department of Information Technology, MIT School of Computing, MIT Art Design and
Technology University Pune, India

Abstract—The occurrence of mental disorders is on the rise in India; however, a significant portion of the population does not approach any professionals owing to factors like stigma, confidentiality issues, affordability concerns, and unavailability of trustworthy therapists. Although several online platforms have come into existence for dealing with such challenges, they fail to ensure anonymity, intelligent therapist matchmaking, and safety precautions. This study suggests HealHive, an AI-driven, anonymous, and secure online platform aimed at helping users deal with their mental issues using advanced technology. The system uses NLP for the evaluation of the user's responses, classifies the severity level of their mental disorder, and recommends further courses of action. In case of low-risk patients, the AI takes charge, while high-risk patients are handed over to verified therapists via intelligent matchmaking services. Anonymity and security are two of the key concerns for users of online systems, and therefore, HealHive uses anonymized verification, encryption of communications, and intelligent moderation based on AI for preventing any unsafe activities. Through comparative analysis with other systems, the advantages of the model under consideration become apparent.

Keywords— analysis: Mental Health, AI Chatbot, Anonymity, Therapist Matching, Secure Communication, NLP, Digital Healthcare

I. INTRODUCTION

Mental disorders have become one of the most serious yet neglected public health concerns of today's digital age. The impact of urbanization, academic pressures, feelings of loneliness, and the presence of a digital culture in our daily life has caused a lot of people to develop anxiety and depression. However, in spite of the need to address such issues, access to effective mental healthcare services is rather limited, especially in developing countries.

However, traditional methods of mental health care suffer from various inefficiencies ranging from shortage of skilled personnel to the unavailability of

mental health experts and high rates for consultations. Most importantly, however, social stigma and shame prevent patients from actively searching for professional advice, which leads to further deterioration of the patient's state. This shows that there is a disparity between the need for mental health care and its availability.

The advent of digital mental health systems has tried to address this problem through the provision of counseling sessions online, chatbot interaction features, and teletherapy. The existing technology, however, is characterized by some limitations. Most digital mental health applications have been found to compel users to identify themselves, thus compromising their privacy. Furthermore, chatbots may lack contextual awareness as well as being unable to provide the right interventions in risky conditions. Another problem with the technology available today is that most systems lack an effective mechanism for selecting therapists.

In this study, we present HealHive, which is the new digital mental health system that is designed on the lines of a privacy-oriented and intelligent framework. It is different from the usual systems as it operates on an anonymous conversational platform, wherein a patient can vent out his/her issues without being identified. We propose to integrate a conversational system with AI-based technology to analyze user inputs through NLP mechanisms.

One of the key contributions made by this research includes the creation of an automated triage system for mental health support, whereby user inputs are constantly analyzed to assess risk levels. Rather than offering generic solutions to all users, the algorithm adopts an adaptive approach, wherein lower-risk patients receive instantaneous guidance from AI assistants, while higher-risk patients are referred to professional therapists.

This project places itself within the realm of Artificial Intelligence, Human Computer Interaction, and Digital Healthcare Systems. This project aims not to address each problem separately by providing single solutions like teletherapy applications and bots, but rather, seeks to provide an ecosystem-based solution.

II. LITERATURE SURVEY

The field of digital mental health has seen fast development with the inclusion of Artificial Intelligence, online forums, and teletherapy apps. The current body of literature is centered on three key areas: AI conversational agents, online mental health forums with matchmaking mechanisms, and overall AI usage in psychotherapy sessions. Yet, while all three of the areas have developed significantly, it appears that an integrated framework for combining all three into one system has not yet been developed.

The literature on the use of AI in mental health chatbots emphasizes their possibilities and challenges. The bots are intended to offer instant and accessible help via conversational interaction. The technologies used by the bots include NLP algorithms that process user inputs and generate answers that can be used for emotional and cognitive behavioral therapy purposes. Nevertheless, the literature shows that chatbots fail to provide proper assistance when it comes to understanding context, particularly in life-threatening situations, such as cases of suicide. A particular challenge identified by the literature is that many chatbots cannot offer adequate interventions, including failing to escalate the issue or contacting the relevant authorities.

Concurrent studies of mental health support communities online highlight the necessity of having an efficient mechanism for matching users with therapists. Conventional approaches include manual matching or rudimentary filtering, which may result in incompatibilities in terms of specialization, language, or availability. Simulation models using agents and algorithms for matching have proven more successful in improving the results by finding the best fit for the user and the therapist. The model factors in several variables such as user preference, behavioral tendencies, and constraints within the system to improve user-therapist interaction and reduce wait time.

Similarly, a large body of literature exists regarding the application of AI in psychotherapy, where its contribution to increasing access and decreasing cost barriers has been widely discussed. AI-driven solutions such as conversation agents and recommendation systems have proven successful in offering low-level psychological assistance. These tools prove particularly valuable in those parts of the world where there is little availability of professional help for people with psychological disorders. Nevertheless, scientific evidence reveals that there is no replacement for human therapists in terms of providing high-intensity therapy.

The following are structural issues with these platforms from a comparative point of view:

- Absence of anonymity, thus hindering free expression
- Ineffective therapist matching process, which is time-consuming and manual
- No capability to assess the urgency of the situation
- Poor safety measures for assessing the risk of suicide attempts
- Separate functioning of human services and AI

The shortcomings mentioned above bring about the requirement of an integrated framework that includes intelligent interactions, automated decision-making, secure communications, and human interventions. The above mentioned research tries to fill in this gap through development of a framework that includes intelligent triage, automated matching, and strong privacy mechanisms.

III. PROPOSED ARCHITECTURE

3.1 Overview of the Architecture

The suggested HealHive architecture acts as a closed loop for a smart mental healthcare solution in which all processes from user interaction to decision making by artificial intelligence and finally human involvement are interconnected. Each process in the pipeline receives inputs from the preceding one to provide a timely and contextually appropriate response.

As opposed to existing models where the process of conversation between the user and chatbot is distinct from the consultation of a therapist, this model integrates both in one single process. The first step of the process involves anonymous conversation

between the user and chatbot, which is then followed by the use of AI to measure the severity of mental illness. Depending on the result obtained, the system either gives automated responses or forwards the case to a human therapist.

3.2 Layers of the Proposed Architecture

3.2.1 User Interaction Layer

This layer serves as the gateway to the network, permitting individuals to use the service without requiring any form of identification. This system is designed to provide anonymity, as users can interact using pseudonyms, thus eliminating the psychological impact of being stigmatized. The user interface is conversational, ensuring that it is easy for all users to operate.

3.2.2 AI Conversational Engine

The AI chatbot acts as the mode of interaction. The inputs received are processed by NLP techniques that help to understand the emotions, intention, and behavior of the user. This is an intelligent system unlike static chatbots that work according to predetermined algorithms.

3.2.3 Mental Health Assessment Module

The module rates user input on the seriousness of their mental disorder, based on:

- Polarity of sentiment (positive/negative)
- Critical keywords in the input (such as stress triggers)
- Contextual clues in the conversation

The result is a quantified severity rating, which forms the starting point of decision making.

3.2.4 Decision Engine

The decision engine acts as the controller of the architecture system. It categorizes the user according to the following criteria depending on the risk score:

- Low-risk User – The AI gives coping strategies and self-help techniques
- Medium-risk User – Recommends visiting a professional, but not mandatorily
- High-risk User – Urgently referred to the therapist

This method of prioritizing users maximizes efficiency and resource allocation.

3.2.5 Therapist Matching engine

This module helps automate the therapist allocation process. It avoids the inefficiency of manual allocation through an analysis of:

- Nature of mental health condition
- Specialization of the therapist
- Preferred language
- Availability

The optimization algorithm guarantees an effective match.

3.2.6 Secure Communication Engine

This layer enables interaction between users and therapists through:

- Encrypted text messaging
- Secure video sessions

End-to-end encryption ensures that all communication remains confidential and protected from unauthorized access.

3.2.7 Security and Moderation Layer

This layer controls system security and ethics adherence. This layer comprises:

- Anonymity authentication procedures
- AI-driven monitoring to detect any malicious activity
- Risks escalation process during emergencies

This is to ensure that the system remains safe for the user community.

3.2.8 Feedback and Learning Loop

The process of improvement is ongoing and relies on feedback. The interaction data collected (with personal information removed) can help fine-tune the AI models, increase matching efficiency, and boost the quality of the replies.

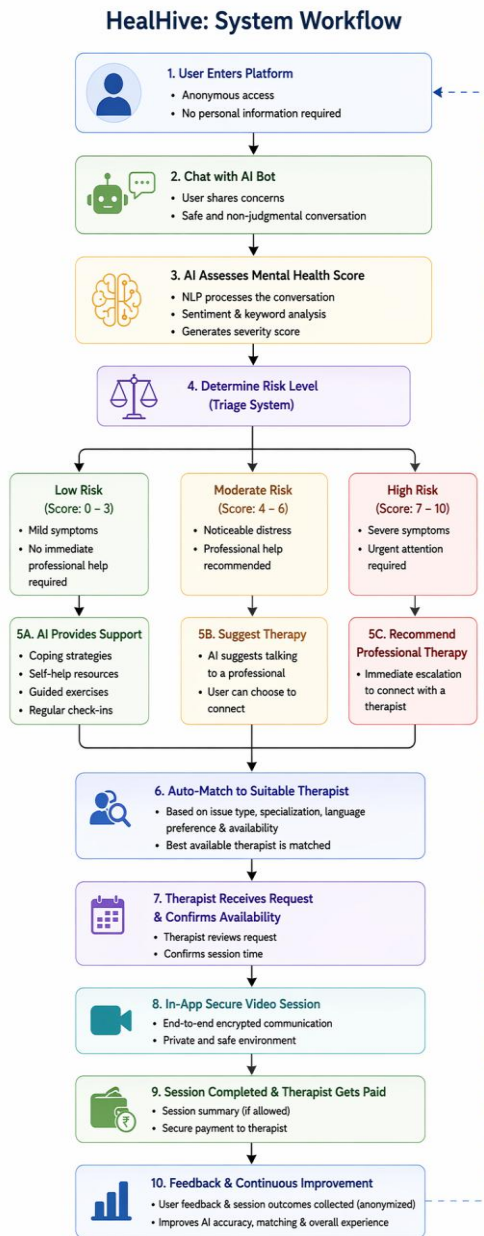


Figure 1: HealHive System Workflow and Decision Process

IV. IMPLEMENTATION, RESULTS AND DISCUSSION

4.1 Implementation

HealHive consists of a modular web application that utilizes cloud technology to combine artificial intelligence and secure communication. The frontend allows the interaction with an AI chatbot anonymously, whereas the backend is responsible for session management, AI processing, therapist matching, and communication.

Artificial Intelligence works with natural language processing algorithms including sentiment analysis

and keyword detection in order to evaluate the user's input and come up with a mental well-being score. Depending on this score, the decision-making module decides if the user should be provided with automated assistance or connected to a therapist. Therapist assignment involves specialists based on their specialty, availability, and other factors. Users can communicate with the assigned therapist using text and video messaging via secure protocols that enable end-to-end encryption.

4.2 Results

The system shows increased efficiency and accessibility in offering mental health services. The chatbot allows instant interaction without relying on the availability of the therapists. The severity assessment system accurately identifies the level of risk associated with each user, making it easier to intervene accordingly.

Automated therapist assignment saves time and ensures more targeted consultations. Increased privacy encourages more users to interact and provide accurate information during assessments.

Important Results:

- Immediate response and user recruitment
- Accurate identification of risk levels with the help of AI
- Time-saving in assigning therapists
- Increased privacy leading to better engagement

The above results show that the system is efficient and effective relative to other methods used in mental health management.

4.3 Discussion

This particular system shows the impact that integrating AI technology into secure digital infrastructure will have on mental health services. In contrast to traditional models, where human participation is key, the HealHive system combines both human and machine elements by letting the machine handle the first contact and triaging while human psychologists take over if necessary.

One of the biggest advantages of the system is its scalability. With the help of the automated chatbot, it becomes possible to manage many more users than would be manageable in a traditional human-only system.

The decision engine is critical in maximizing resource use because users are classified based on the gravity of the situation and only in cases where it is needed will therapists be assigned to users. Triage is important for balancing the system under situations of high demand.

When using the system from a user point of view, anonymity is a critical element that builds trust and confidence in the system. Users will easily provide information without fearing the disclosure of their identity.

Nevertheless, there exist several constraints associated with this approach. The accuracy of the algorithm used for assessing risks relies on the diversity of datasets utilized for training the machine learning models. The misinterpretation of risks can have adverse implications for users. Moreover, ethical considerations should not be overlooked, especially regarding automated decisions involving patients who pose significant threats.

System Strengths:

- Efficient and scalable AI-driven communication
- Privacy-oriented design promoting user participation
- Automatic and intelligent selection of therapists
- Real-time risk assessment

Limitations:

- Dependence on AI model accuracy
- Potential challenges in handling ambiguous user input
- Ethical concerns in automated mental health decisions

In conclusion, the implementation proves the viability of utilizing an AI-human hybrid system in mental healthcare delivery services. From the outcomes of the experiment, it is evident that the integration of artificial intelligence with human intervention leads to increased accessibility and responsiveness, which makes the system scalable.

REFERENCES

[1] Abd-Alrazaq A., et al., “Effectiveness and Safety of Using Chatbots to Improve Mental Health: Systematic Review and Meta-Analysis,” *Journal of Medical Internet Research*, vol. 22, no. 7, pp. 1–21, 2020.

[2] Balcombe L., et al., “Artificial Intelligence-Enabled Chatbots in Mental Health: A Narrative Review,” *Digital Health*, vol. 9, pp. 1–15, 2023.

[3] Olawade D.B., et al., “Enhancing Mental Health with Artificial Intelligence: Current Trends and Future Prospects,” *Healthcare Analytics*, vol. 4, pp. 1–10, 2024.

[4] Haque M.D.R., et al., “Applications of Chatbots in Mental Health: A Systematic Review,” *Healthcare*, vol. 11, no. 12, pp. 1–18, 2023.

[5] Zhang Q., et al., “Generative AI Chatbots in Mental Health: Systematic Review and Meta-Analysis,” *Journal of Medical Internet Research*, vol. 27, pp. 1–14, 2025.

[6] Nyakhar S., et al., “Effectiveness of AI Chatbots in Alleviating Symptoms of Depression and Anxiety: Meta-Analysis,” *Journal of Medical Internet Research*, vol. 27, no. 1, pp. 1–12, 2025.

[7] Wang Y., et al., “A Cognitive Behavioral Therapy-Based AI Chatbot for Mental Health Support: Randomized Controlled Trial,” *JMIR mHealth and uHealth*, vol. 13, no. 1, pp. 1–13, 2025.

[8] Pichowicz W., et al., “Performance of AI Chatbots in Crisis Detection and Mental Health Support,” *Scientific Reports*, vol. 15, pp. 1–11, 2025.

[9] Yoo D.W., et al., “Exploring the Values and Harms of AI Chatbots in Mental Health,” *arXiv preprint arXiv:2504.18932*, pp. 1–12, 2025.

[10] Stanford Human-Centered AI Institute, “Exploring the Risks of AI in Mental Health Care,” Stanford University, 2025.