Enhancing Mental Health Support for Young Adults Through Conversational AI-Based Screening Tools

[1]Bamidele ADENLE, [2]Wilson SAKPERE, [3]Temitope OPALEYE
[1]bamidelejohnson2@gmail.com/+2348037699149,
[2]Sakpere.wilson@lcu.edu.ng/+2349137792962, [3]opaleyete@neuroaro.gov.ng /+2348060965538

[1-2]Department of Computer Science, Lead City University, Ibadan. Oyo State Nigeria,
[3]School of Psychiatric Nursing Neuropsychiatric Hospital Aro Abeokuta Ogun State

Abstract

Young adults are increasingly vulnerable to depression and anxiety due to educational, social and economic pressures, but stigma and limited access to care often prevent early intervention. This study evaluates the acceptability and performance of a conversational AI-based screening tool that integrates the validated Patient Health Questionnaire-9 (PHQ-9) and Generalized Anxiety Disorder-7 (GAD-7) scales. Using a cross-sectional experimental design, data were collected from young adults aged 18–35 interacted with the natural language chatbot. A total of 950 young adults participated in this study, contributing conversational text and standardized scores. Preprocessed responses were analyzed using a support vector machine (SVM) classifier to distinguish between crisis (+1) and non-crisis (-1) expressions. The model achieved a precision of 84.07%, recall of 83.62%, and F1 score of 83.84%, confirming its reliability in identifying emotional distress. Participants reported greater comfort and openness when communicating with chatbots compared to traditional self-report formats, indicating increased engagement and reduced stigma. The findings highlight the potential of conversational AI to improve early mental health screening and provide accessible, non-judgmental support. The study recommends ethical integration of AI-powered screeners into consultation and telemedicine systems to expand scalable, privacy-preserving mental health care for young adults.

Keywords: Conversational AI, Young adults, Mental health screening, Acceptability, Digital health support

Introduction

Mental health has become one of the most pressure global public health preferences in the 21st century, especially among young adults. There are typically quick changes in biological, psychology, and society throughout the period between adolescence and early adulthood. For individuals between the ages of 18 and 35, this stage includes navigating higher education, employment, relationships and identity formation (Ida Marie & Anne Kathrine, 2023).

These increase vulnerability to conditions such as development challenges, economic pressure, social expectations and external stresses such as the effect of digital media, often anxiety and depression (Khaparde et al., 2023). Research continuously indicates that mental health problems eat for a significant disease burden around the world, and among young people these challenges often appear quietly, undetected or untreated until they become serious (the World Health Organization, 2023). The stigma, limited access to professional care and inadequate awareness and prevents more early recognition and governance. In recent years, technological advances have opened new ways to address these persistent obstacles to care.

People who utilize digital health interventions, especially the artificial intelligence (AI), have attracted attention as a possible scalable, cost effective and user-friendly tool for mental health care. Unlike traditional services, which are often limited by geographical boundaries and resource availability, AI operated approaches can provide immediate, confidential and personal commitment. It is especially relevant for young adults, who are usually more open to digital solutions and have heavy users of mobile and online platforms. In this context, conversational AI as often known as chatbots or dialogue systems appears as a promising medium for screening and support.

These systems are designed to mimic human-like conversation, creating an accessible and non-judgmental environment where users may feel more comfortable disclosing sensitive emotions and experiences (Manole et al., 2025). Standardized screening tools are a crucial way that conversational AI is being used in mental health. The Patient Health Questionnaire-9 (PHQ-9) for depression and the Generalized Anxiety Disorder-7 (GAD-7) for anxiety are two of the most extensively validated tools. Both tools are clinically trustworthy, brief, and simple to use, which makes them appropriate for quick evaluation in a variety of healthcare contexts. and extensively used in diverse settings, including primary care and telehealth (Shen & Zhao, 2024).

However, the traditional form as paper based or self-administered surveys may feel impersonal and disengaging to young users. Integrating these tools into conversational AI interfaces can reframe the process into a more interactive, natural dialogue that aligns with the communication preferences of today's young adults. Instead of presenting a stable question, such as "for the past two weeks, how many times do you have little interest or joy in doing things?", A chatbot can ask, "Do you recently enjoy your general activities, or have you felt less interested?" Such rewording maintains clinical fidelity while enhancing relatability. While several published studies have explored the use of AI-powered chatbots for mental health support including (Akinboyejo & Asimolowo 2025). On high school students, (Konadu 2025) on general student well-being, and meta-analytic reviews evaluating the effectiveness of conversational tools for youth populations these studies primarily focus on general, or perception support, efficiency. In contrast, this study is uniquely positioned in three key ways.

First, it integrates clinically validated screening instruments (PHQ-9 and GAD-7) directly into the interactive delivery format rather than using fixed questionnaires, thereby transforming traditional assessments into adaptive interactions tailored for young adults. Second, unlike previous work that emphasizes qualitative user experiences, this study combines standardized scores and real conversational text to create a labeled computational dataset, enabling the development and evaluation of an NLP-powered SVM classifier specifically trained to detect emotional distress in natural language interaction.

Third, this research simultaneously evaluates acceptance, engagement, and clinical performance within a controlled, secure, purpose-built web interface hosted on Firebase which is distinct from studies conducted on general messaging platforms or theoretical reviews. These contributions provide a new empirical basis for integrating clinically-based conversational screeners into real-world telehealth systems for young adults, thereby expanding the literature beyond descriptive analysis to practical, deployable, and data-driven innovation for mental health screening.

Aims

The study aims to evaluate the acceptability and integration of conversational AI based delivery of validated screeners for enhancing mental health support among young adults.

Related Study

According to (Yuan et al., 2025). One notable area of growth in mental health intervention is the use of synchronous chat-based online one-on-one psychological therapies, where text-based communication serves as the primary mode of interaction between patients and therapists. A comprehensive review of these interventions underscores their increasing popularity as web-based solutions for individuals experiencing psychological distress. The review presents preliminary evidence suggesting that synchronous text-based therapy is not only effective but also yields comparable outcomes to traditional face-to-face treatments. In some cases, it has even demonstrated better results than waitlist control groups, indicating its viability as a credible alternative to conventional therapy particularly for individuals who value anonymity, flexibility, and accessibility**.**

The study by (Hossain et al., 2025) stated that health professionals are often thin, and they can look for technology based solutions, such as chatbots, some administrative or other loads to reduce the time. Chatbots are available 24/7 without the requirement for human personnel at the same time. Chatbots can answer faster questions than humans and handle a large volume of requests and provide information in a standardized format for a smooth user experience

Recent studies highlight the growing role of conversational agents in mental health care, with evidence suggesting that machine learning (ML) enables chatbots to adapt to past user experiences through session storage, thereby personalizing interactions and customizing therapeutic support for each individual (Aman Singh, et al 2025). This adaptability enhances user engagement and may improve the effectiveness of mental health screening and assistance. Existing literature has consistently shown that AI-driven chatbots can reduce barriers such as stigma and limited access, providing immediate, scalable, and cost-effective support to vulnerable populations, particularly young adults. However, most prior research has focused on intervention outcomes or general user satisfaction, while limited attention has been given to the acceptability of validated screening tools when integrated into conversational platform

Methods

This study employed a cross-sectional experimental design that integrated validated clinical instruments the Patient Health Questionnaire-9 (PHQ-9) for depression and the Generalized Anxiety Disorder-7 (GAD-7) for anxiety within a conversational AI framework. Participants aged 18–35 interacted with the natural language chatbot. A total of 950 young adults participated in this study, contributing conversational text and standardized scores. The data derived from these interactions were pooled to create a labeled dataset for model training and evaluation.

Participants interacted with the chatbot through a secure online platform. Specifically, the chatbot was deployed on a custom web-based interface hosted using Google Firebase. This proprietary, controlled platform (unlike commercial messaging services) ensured a consistent and standardized user experience for all participants.

- **Structured responses**: Numerical scores from PHQ-9 and GAD-7 items.
- **Unstructured responses**: Free-text conversational answers, which may contain subtle indicators of emotional distress.

Natural Language Processing (NLP) and Data Processing

The core Natural Language Processing (NLP) architecture utilized is a vector-space model for text classification. This approach transforms raw, unstructured conversational text into a numerical format suitable for a supervised machine learning classifier,

The total dataset used for computational modeling comprised 2,016 individual user responses (data points) derived from the 950 participants. Each response was manually or semi-automatically labeled into two classes:

The responses are then manually or semi-automatically labeled into two classes:

+1 (Distress present) – responses reflecting anxiety, depression, or other markers of emotional discomfort.

-1 (No distress) – responses that do not indicate mental health challenges.

Clinical labeling used data on approved psychological tests (GAD-7 and PHQ-9). Emotion-based labeling, using the circumplex model of affect, was also integrated to enrich the system's capacity for tailored reactions

1. **NLP Preprocessing Pipeline**

Raw data needs to be cleaned and standardized before it can be examined before computational modeling to improve feature quality and reduce noise. The following pipeline was implemented:

2. **Text Normalization:** Conversion to lowercase, removal of punctuation, special characters, and extra whitespaces.
3. **Stop-word Removal:** Common non-informative words (e.g., “and,” “is,” “the”) removed using NLTK’s predefined list.

4. **Tokenization:** Splitting text into individual words or sub-words using a subword-aware tokenizer to preserve context.
5. **Lemmatization and Stemming:** Reduction of words to their root form (e.g., "doing" → "do") using spaCy.
6. **Emoji and Slang Translation:** Mapping of informal expressions to their standard semantic equivalents (e.g., "😊" → "happy," "💔" → "sad").
7. **Anonymization:** Replacement of any identifiable information (names, locations) with neutral placeholders.

Classification Model Development

A Support Vector Machine (SVM) classifier was selected for its proven effectiveness in high-dimensional text classification tasks. The model utilizes the $TF - IDF$ feature vectors (xi) to asses: distress ($\hat{y}i = +1$) and non-distress ($\hat{y}i = -1$).

The ($\hat{y}i = sign\big(f(xi)\big) where\ f(xi) = \omega^{T} xi + b$

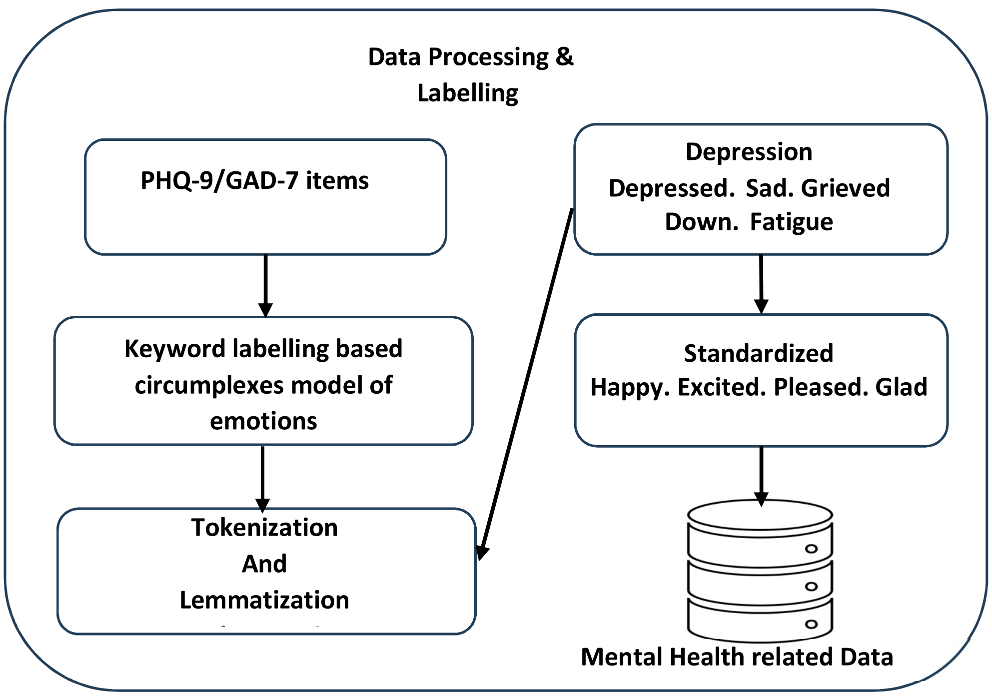

Fig 1: Conceptual Diagram

Label Encoding: Clinical labeling and emotion -based labeling were two main marking techniques. Data on approved psychological tests, including GAD-7 and PHQ-9 scale, are used in clinical labeling. The method maps linguistic content to a degree of mental health problems, from moderate anxiety to severe depression to evaluate each lesson entrance, which is according to this psychological score or diagnosis. These labels provide a dataset with a medical foundation that can be used ... Unlike, emotion -based labeling takes a more complex, loving approach Using the circumplex model of affect, which divides emotional states into valence (positive vs. negative) and arousal (high vs. low), emotional descriptions like "hopeless," "anxious," "agitated," "calm," or "motivated" are attached to each entry.

This process improves the capacity of the system for tailor -made reactions according to a wide range of emotions, which increases the possibility of sympathy and participation. In situations where ambiguity or sarcasm can occur, manual comments are performed. These include compound signals like "I'm doing okay, but I've been crying every night" or edge instances like "I'm totally fine," where sarcasm may imply suffering. To guarantee consistent and contextually aware categorization, these cases are cross-examined by expert reviewers. In addition, the Lexicon-based spirit assessment (from NRC feelings Lexicon, for example), was integrated into enchanted labeling techniques to help with a challenging them ... As a result, in the final dataset, each entrance consists of two labels: a primary binary flag that engages itself (0 or 1) and the risk of the risks.

Results

The Support Vector Machine (SVM) classifier effectively distinguishes between two response classes: crisis (+1) and non-crisis (−1). The achieved robustness and accuracy are significantly attributable to the rigorous Natural Language Processing (NLP) Preprocessing Pipeline implemented on the conversational data. This essential methodology involved a sequence of steps—including text normalization, stop-word removal, tokenization, lemmatization and stemming (using spaCy), as well as slang and emoji translation —to standardize linguistic variations and mitigate noise. The successful reduction of noise and normalization of features through this pipeline ultimately enhanced the quality of the Term Frequency–Inverse Document Frequency $TF - IDF$ vector representation, which served as the input for the SVM, resulting in the high classification performance observed

The classifier achieved a precision of 84.07%, recall of 83.62%, and F1 score of 83.84%, confirming its robustness in detecting emotional distress in conversational text. The SVM captured both explicit indicators of depression and anxiety (e.g., "I feel depressed", "I can't sleep") and implicit emotional cues in everyday manifestations. Compared to traditional text analysis techniques, SVM demonstrated strong performance in high-dimensional feature spaces derived from conversational data.

Discussion of Findings

The findings show that young adults showed higher levels of engagement and comfort when interacting with conversational AI tools compared to traditional self-assessment formats. Many participants felt more comfortable disclosing sensitive feelings on a chatbot, which is consistent with previous research suggesting that AI-mediated dialogue reduces stigma and promotes openness in mental health discussions (Supriyono et al., 2024). Furthermore, the dialogue-based presentation of the PHQ-9 and GAD-7 items was perceived as less clinical and more natural, which increased user trust and the authenticity of the answers. Analysis of SVM output indicates that a large proportion of free-text responses contain micro markers of emotional distress – subtle linguistic cues that are often overlooked in static survey formats.

These insights reinforce the potential of conversational AI as a scalable and empathetic tool for early mental health screening among young adults. The results also highlight how machine learning models can increase user engagement in digital spaces by adopting natural language styles commonly used by younger populations.

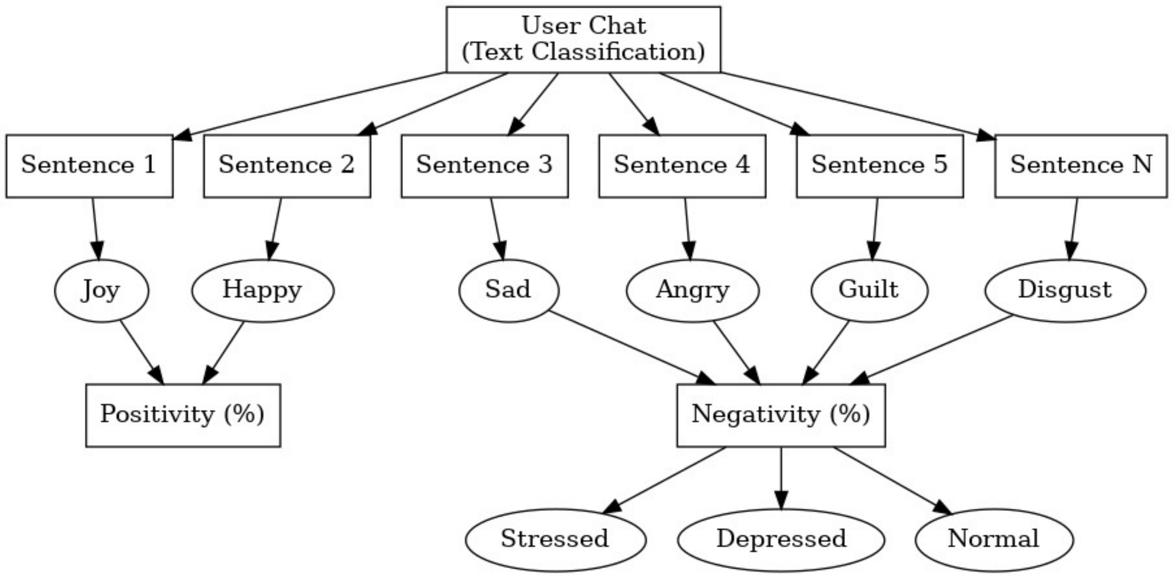

Figure 2: Mental State Identification with User Text

A support medical chatbot was suggested to provide relief to people from different stress stages by distributing emotional identification techniques. It takes recording texts from the user in chat and treats it in different ways. Within the text there is a list of emotions including happiness, shame, happy, hatred, fear, anger, crime and sadness. In addition, it determines the positivity and negativity percentage of each chat material on emotions

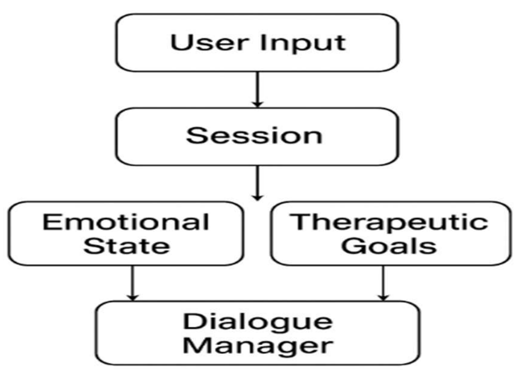

Figure 3: User-Centered Therapeutic Model

The dialogue manager acts as the main workflow manager of the interactive AI Chatbot, and decides how to start, continue and end the dialogue. This item is required to convert stable question-response exchange to conversion streams that are lively, sympathetic and relevant consciousness. In order to offer significant support and participation, user entrances, sessions are required, emotional status and therapeutic purposes, in the diagnosis of anxiety and depression in young people. In order to preserve a memory of past conversation states, emotional patterns, and user, emotional trends and previous discussions. The Dialogue Leader receives information about the user's previous interaction, such as the emotional points, the symptoms mentioned, and pre -medical treatment, when the user interacts with Chatbot. For example, if a user had previously acknowledged having a panic attack, a calming intervention or follow-up inquiry based on previous chats could be triggered in subsequent conversations that suggest heightened concern.

Table 1: Examples of Empathetic Chatbot Responses to User Emotional Triggers

User Input	System Classification	Chatbot Response
I have been having trouble sleeping	Mild Depression	I'm really sorry you are feeling this way. You are not alone, and there are steps we take together
I feel hopeless sometimes	Moderate Depression	That sounds really difficult. Please know that support is available and you don't have to go through it alone
I am no longer sure what to do	Moderate Depression	Is okay to feel overwhelmed. Let's take it one step at a time. Would you like some options that can help right now?
Nothing matters to me anymore	Severe Depression	Thank you for sharing that. It's important that you are talking about this. You are not alone; help is always available
I'm always tired and can't concentrate	Mild Depression	These could be signs of stress or emotional exhaustion. Would you like some tips to help improve your focus and rest?
	Moderate Anxiety	That must feel very lonely. I'm here to support you, and there are also

I don't have anyone to talk to		professionals who care and want to listen.
I think I might need help, but I'm scared	Neutral	It's completely okay to feel that way. Reaching out takes courage, and I can guide you on what steps to take next

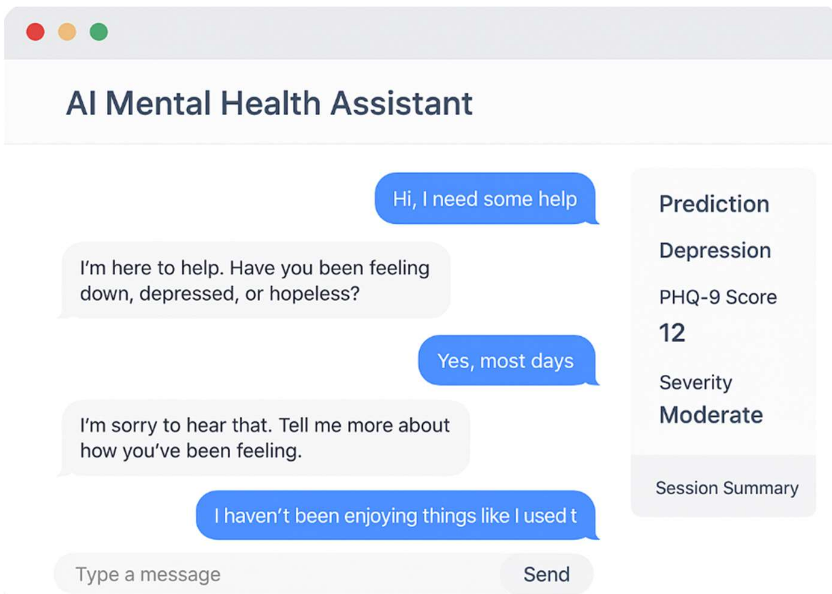

Fig: 4 Live web-based chatbot interface showing user system interaction and automated emotional-state prediction

The interface demonstrates a real-time conversation between the user and the AI mental-health assistant, showing natural dialogue, contextual understanding, and automated prediction of depressive symptoms. The right-side panel displays the model's PHQ-9 score and severity level, confirming the system's ability to classify moderate emotional distress within a live, web-based environment.

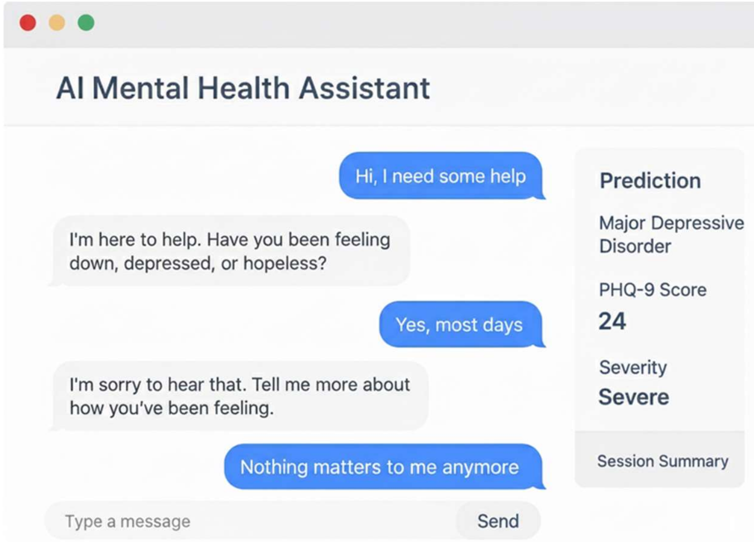

Fig: 5 Representative Screenshot of the Live Conversational AI Interface Highlighting Real-time Severe Anxiety Classification.

Figure 5, illustrates a critical anxiety-focused exchange. The user input, "My thoughts won't stop and I can't breathe properly," is processed in real-time. The prediction panel confirms the system's capability by classifying the distress as Generalized Anxiety Disorder with Severe intensity (GAD-7 Score 19), showcasing the chatbot's immediate clinical utility for early screening.

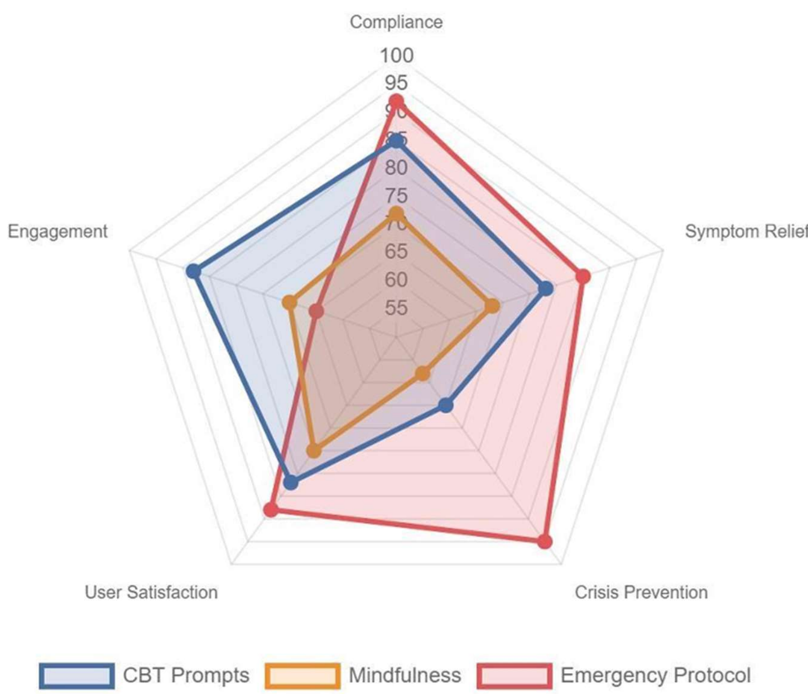

Figure 6: Radar Chart comparing four key dimensions of intervention effectiveness (normalized scores)

- The confusion matrix presented here illustrates how successfully the chatbot was able to categorize **True Positives (diagonal cells)**: Cases where the model correctly predicted the emotional state.
- **Off-diagonal cells**: Misclassifications (e.g., anxiety classified as depression)
- **570** users correctly identified as Anxiety.
- **560** users correctly identified as Depression.
- **550** users correctly identified as Neutral.

users into three emotional states: neutral, depression, and anxiety.

Table 2: Performance Evaluation Metrics for Classification Model

Metric	Calculation	Result
Accuracy	(950 + 700) / (950 + 700 + 180 + 186) = 1,650 / 2,016	0.8186 (81.86%)
Precision	950 / (950 + 180) = 950 / 1,130	0.8407 (84.07%)
Recall	950 / (950 + 186) = 950/ 1,136	0.8362 (83.62%)
F1-Score	2 × (0.8407 × 0.8362) / (0.8407 + 0.8362)	0.8384 (83.84%)

Accuracy

Accuracy is the ratio of the model's total number of accurate predictions (i.e., the sum of true positives and true negatives) to the total number of cases assessed.

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN}$$

Precision

The number of users who truly experienced anxiety or depression is determined by precision.

$$\text{Precision} = \frac{TP}{TP+FP}$$

With an accuracy percentage of 84.07%, the chatbot's affirmative diagnoses showed a high degree of dependability. This is important because in a mental health setting, a false positive could cause people to get unduly alarmed or become overly concerned. Those who are flagged for emotional support are more likely to be the ones who need it if it is done with precision.(C et al., 2025)

Recall (Sensitivity or True Positive Rate)

The chatbot's recall indicates the proportion of real cases of anxiety or despair that it properly identified.

$$\text{Recall} = \frac{TP}{TP+FN}$$

The chatbot's recall score of 83.62% shows that it correctly identified a significant percentage of real-world cases of anxiety and sadness. This is crucial for mental health applications, since a missed chance for early intervention may result from a false negative result, which fails to identify a risk.

Conclusion and Future Work

First, a labelled dataset of emotional interactive exchanges was made using NLP (Natural Language Processing (NLP) techniques, which allowed the model to recognize and classify the symptoms expressed under the Chatbot calls. Secondly. a Support Vector Machine (SVM) classifier was developed to distinguish between distress and non-distress responses with high accuracy, ensuring reliable symptom recognition.

Conclusively, this study adds to the expanding field of digital mental health interventions by demonstrating how conversational AI driven by machine learning and natural language processing (NLP) can revolutionize the way unmet youth mental health needs are addressed. If the suggested system is improved, integrated with current telehealth platforms, and has robust ethical protections, it has the potential to be a scalable and sustainable way to increase access to mental health care globally. Future research should focus on increasing the system through the integration of speech conversation and sound -based emotions, as it will provide more flexibility for users who prefer to talk to writing and enriching the system's ability to capture emotional shades from vocal signals. Strengthening cooperation with mental health professionals is also necessary to ensure that the results of Chatbot are clinically valid and the response libraries are constantly processed with the medical material frozen in professional practice. In addition, it will be important to establish a strong moral structure, to guarantee the user's privacy with frequent review processes and strong data protection protocols, maintain openness and ensure complete compliance with regulatory standards. These reforms will not only increase the reliability and projection of the system, but will also create a trust required for widespread adoption in digital mental health care.

References

C, I. G., J, J. M., A, K. E., & Euodial, A. (2025). *Ai Powered Chatbot For Mental*. *10*(4), 703–708.

Cho, Y. M., Rai, S., Ungar, L., Sedoc, J., & Guntuku, S. C. (2023). An Integrative Survey on Mental Health Conversational Agents to Bridge Computer Science and Medical Perspectives. *E*MNLP 2023 - 2023 Conference on Empirical Methods in Natural Language Processing, Proceedings, 11346–11369. https://doi.org/10.18653/v1/2023.emnlp-main.698

DIng, Y., Chen, X., Fu, Q., & Zhong, S. (2020). A Depression Recognition Method for College Students Using Deep Integrated Support Vector Algorithm. IEEE Access, 8, 75616–75629. https://doi.org/10.1109/ACCESS.2020.2987523

Hossain, M. M., Hossain, M. S., Mridha, M. F., Safran, M., & Alfarhood, S. (2025). Multi task opinion enhanced hybrid BERT model for mental health analysis. Scientific Reports, 15(1), 1–20. https://doi.org/10.1038/s41598-025-86124-6

Khaparde, A. A., Das, R., & Bhargava, R. (2023). Transformer Based Approach for Depression Detection. Proceedings - International Conference on Developments in ESystems Engineering, DeSE, 2023-Janua, 369–374. https://doi.org/10.1109/DeSE58274.2023.10099629

Lakizadeh, A., & Zinaty, Z. (2021). A Novel Hierarchical Attention-based Method for Aspect-level Sentiment Classification. Technology Journal of Artificial Intelligence and Data Mining, 9(1), 87–97.

Manole, A., Cârciumaru, R., Brînza□, R., & Manole, F. (2025). An Exploratory Investigation of Chatbot Applications in Anxiety Management: A Focus on Personalized Interventions. Information (Switzerland), 16(1), 1–33. https://doi.org/10.3390/info16010011

Shen, C., & Zhao, Z. (2024). Detection of Negative Emotions and Depression in Social Networks Based on Bert-LSTM Model (Issue Iciaai). Atlantis Press International BV. https://doi.org/10.2991/978-94-6463-540-9_91

Supriyono, Wibawa, A. P., Suyono, & Kurniawan, F. (2024). Advancements in natural language processing: Implications, challenges, and future directions. Telematics and Informatics Reports, 16(September), 100173. https://doi.org/10.1016/j.teler.2024.100173

Vuyyuru, A., Praveena, T. L., Sharma, A., Yelagandula, M., & Nelli, S. (2024). Mental Health Therapist Chatbot Using NLP. 2024 International Conference on Artificial Intelligence and Quantum Computation-Based Sensor Applications, ICAIQSA 2024 - Proceedings, 1–6. https://doi.org/10.1109/ICAIQSA64000.2024.10882362

Yuan, A., Colato, E. G., Pescosolido, B., Song, H., & Samtani, S. (2025). Improving Workplace Well-being in Modern Organizations: A Review of Large Language Model-based Mental Health Chatbots. ACM Transactions on Management Information Systems, 16(1). https://doi.org/10.1145/3701041