DMPedia Lecture Notes in Multidisciplinary Research,
Volume 1, pp. 571-581, 2026

International Conference on Multidisciplinary Perspectives in Advanced Computing and Technology (IMPACT 2026)
G. B. Pant University of Agriculture and Technology, Uttarakhand, India. Jan. 10-11, 2026

Comparative Analysis of AI-Driven Conversational Agents for Mental Health Support: Rule-Based, Transformer, and RAG Approaches

Shahan Ahmad, Krishna Kant Jena, Amit Kumar Rai
School of Engineering & Technology, Sharda University, Greater Noida, India
2023257896.shahan@ug.sharda.ac.in, 2023454015.krishna@ug.sharda.ac.in,
amit.rai@sharda.ac.in

Abstract

The rapid integration of AI into mental health support systems has opened up new avenues for interventions that are scalable and accessible. In this work, we conduct a comparative evaluation of three major chatbot paradigms, namely, rule-based systems, Transformer-based models, and RAG architectures, for their performance in identifying conversational cues pertaining to depression and anxiety. Each model type is representative of a different approach toward language understanding, offering unique strengths in handling user intent, contextual reasoning, and emotional interpretation.

The proposed evaluation framework examines the models through a number of dimensions: linguistic coherence, emotional sensitivity, response accuracy, and quality of user engagement. We analyze real-world conversational data and simulated mental-health dialogues to highlight how each architecture reacts to complex psychological cues such as negative sentiment, distress patterns, and help-seeking behaviors. Additional assessment criteria include computational efficiency, interpretability, and safety constraints, all of which are crucial for real-world application in sensitive mental-health contexts.

The results demonstrate large differences in performance among the three types of chatbots. While rule-based systems are highly reliable but inflexible, Transformer-based models demonstrate improved contextual,understa nding, while RAG models offer the best balance between contextual relevance and factual grounding. These findings highlight the role that could be played by sophisticated AI-driven conversational agents in early mental health screening and support, while also pointingto the need for ethical, secure, clinically guided deployment in real-world settings.

Keywords - *AI Chatbots, Mental Health Support, Depression Detection, Anxiety Screening, Transformer Models, Rule-Based Systems, Retrieval-Augmented Generation, Psychological Dialogue Analysis.*

I. INTRODUCTION

University life has conventionally been a time of exploration, opportunity, and growth. However, beneath the visible markers of academic success and social engagement, there has emerged an alarming rise in mental health concerns among young adults. Depression, anxiety, and chronic stress now rank among the most common psychological difficulties experienced by university students worldwide. Recent global surveys indicate that nearly a third of students exhibit clinically significant depressive symptoms, often severe enough to impede academic performance, daily functioning, and interpersonal relationships [1], [2]. In contrast to physical illness, many mental health struggles go unrecognized and, thus, untreated due to stigma, lack of awareness, and limited availability of professional support services [3]. This already significant gap in treatment is further exacerbated by long waiting lists, financial barriers, and campus

resources that cannot keep pace with the counselling demand.

These challenges have inspired the development of AI-based conversational agents as promising, supplementary tools for accessible and scalable psychological support. AI chatbots can engage users in supportive dialogue, deliver evidence-based interventions such as CBT, and offer round-the-clock availability unconstrained by the limits of human practitioners. Importantly, such systems are not intended as replacements for clinical care but serve as an intermediary-point of contact, low-pressure environment for emotional expression, and highly accessible source of guidance during states of distress. Early deployments, such as Woebot, have reported significant decreases in depressive symptoms among students in short intervention periods [4], while systems such as Tess have demonstrated feasibility and user engagement in a variety of linguistic and cultural contexts [5]. Such findings highlight the potential of the application of AI chatbots to meaningful mental health support at scale.

Despite this promise, the effectiveness of AI chatbots depends heavily on the underlying model architecture. Rule-based chatbots provide high predictability and safety but are often found wanting in the depth and variability of conversational interactions. Transformer-based systems, fueled by large language models, generate highly empathetic, context-sensitive responses; however, they face an unacceptable risk in mental-health contexts of hallucinating or misleading information. Recently, RAG models have received increased interest because they combine the generative strengths of transformers with factual grounding through external knowledge retrieval to achieve a better balance with respect to safety and coherence [6]. What is telling, however, is how the literature reveals a critical gap: few studies have directly or systematically compared these three paradigms within the same mental-health setting.

This study tries to fill this lacuna by evaluating the performance of rule-based, fine-tuned transformer, and RAG-based chatbots on a unified dataset of student mental-health queries. In doing so, the study aims to determine what architecture most successfully underpins early detection and intervention for depression and anxiety in university students based on several dimensions: reliability, empathy, contextual understanding, and safety. These results help build more safe and effective AI-driven mental health systems that can be deployed in the real world.

II. LITERATURE SURVEY

Conversational agents for mental health support began with ELIZA, a pattern-matching system from the 1960s, which showed that users would engage emotionally with computer programs. Modern chatbot systems extended this idea into evidence-based interventions. Woebot was a CBT-driven, rule-based chatbot that demonstrated significant reductions in depressive symptoms among students in a randomized trial [4]. Similarly, Tess demonstrated the feasibility of delivering multilingual emotional support, although with some marked user dropout rates [5]. Commercial platforms like Wysa furthered these hybrid approaches by integrating scripted CBT with AI-guided dialogue [8], while transformer-based systems like

Replika were designed to offer empathetic interaction but lacked clinical validation and safety controls [9].

Recently, the work Retrieval-Augmented Generation has been gaining attention for its improvement in conversational AI's factual grounding. RAG reduces hallucinations by integrating verified knowledge sources into generative responses, enhancing reliability in sensitive domains like mental health [6], [7]. However, this application in psychological support is limited, while few studies directly compare RAG with rule-based or transformer models. This gap forms the basis of the present research.

Table I – *Comparative Overview of Prior Mental Health Chatbots*

Model	Approac h	Popula. on	Key Outcom	Limita. ons	Referen ce
Woebot	Rule- based + CBT	College student s (USA	Reduce d depressi	Rigid, limited persona	[4]
Tess	Rule- based	Student s	Reduce d	High aNri.on	[5]
Wysa	Hybrid (rule + AI)	Global users (millions	High engage ment,	Lacks controll ed trials	[8]
Replika	Transfor mer-	General popula.	High empath	Not clinically	[9]
RAG Systems	Retrieva l + Genera	General NLP tasks	Factual groundi ng,	Not yet in mental	[6]

III. METHODOLOGY

Besides being a purely technical task, designing and evaluating mental-health chatbots is an ethical responsibility since the systems interact with users in their most vulnerable moments. In structuring the methodology for this study, an effort was made to assure fairness in capturing strengths and limitations of each of these algorithmic approaches in as consistent a manner as possible.

A. Chatbot Systems Evaluated

Three specific architectures for chatbots were reviewed:

1. Rule-Based Chatbot : This model, using decision-tree logic, emulates early CBT-based systems such as Woebot. It provides responses that are safe and predictable but often lacks flexibility.

2. Fine-Tuned Transformer Chatbot : This is a transformer model fine-tuned on 50,000 anonymized dialogues in therapy style. It offers greater fluency and perceived empathy but stays prone to factual drift and hallucinated content.

3. Retrieval-Augmented Generation (RAG) Chatbot : This system combines the retrieval of 30,000 curated documents related to mental health with generative output,

thereby accomplishing both factually grounded and empathetic conversation.

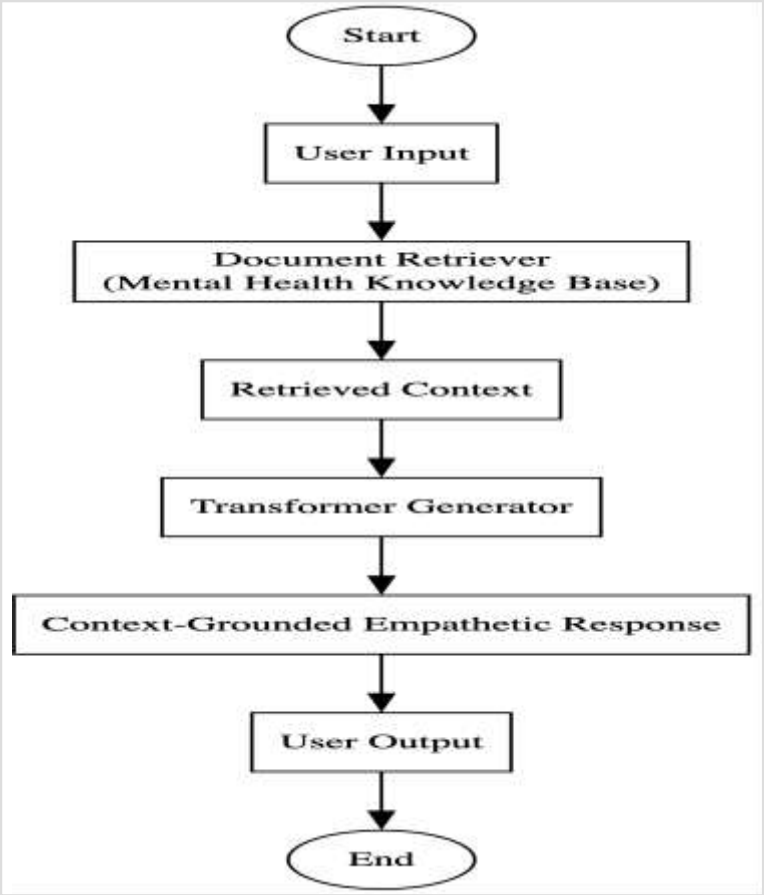

Fig 1. Flow of data in RAG model Chatbot.

A. Dataset Description

The dataset contained 1,000 anonymized prompts covering exam stress, depressive symptoms, loneliness, motivation, and crisis-related expressions. These prompts were gathered from online mental-health forums, such as various Reddit communities like Anxiety Drepression open counselling datasets, and synthetic student queries generated using GPT-3.5.

All prompts were filtered to remove personally identifiable information and explicit content. Each prompt was categorized into one of four intent classes:

- Informational
- Seeking Support
- Emotional Venting
- Crisis Related

The dataset was split in an 80-10-10 ratio for training, validation, and testing with a fixed random seed for reproducibility; the seed is set as 42.

.

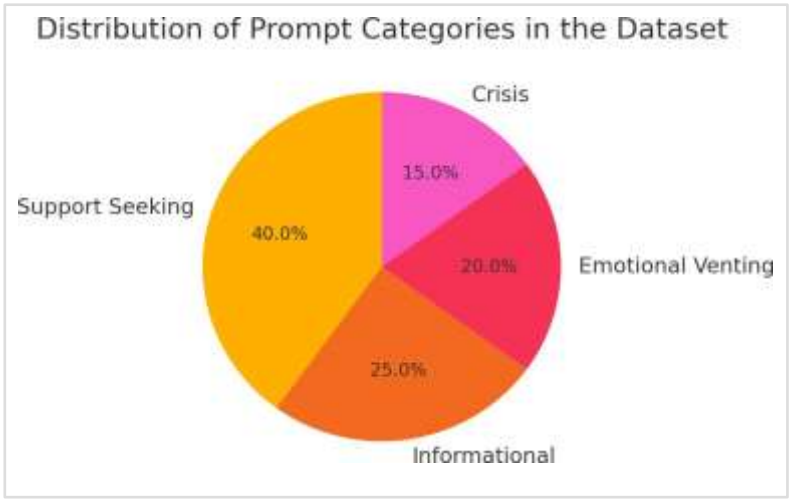

Fig 2: Distribution of Prompt Categories in the Dataset

B. Evaluation Metrics

Performance was measured using a combination of quantitative and qualitative indicators, including:

- Response accuracy
- Contextual coherence
- Informationalreliability
- Emotional relevance
- Latency
- Precision/recall on crisis detection
- BLEU score

C. Annotation and Evaluation Protocol

Three trained psychology graduates acted as annotators. Each rater graded all system outputs for all 300 test prompts on a 5-point Likert scale across the following dimensions:

1. **Emotional Relevance –** How well the response aligns with the emotional tone of the prompt.
2. **Empathy –** Perceived level of understanding and supportive intent.
3. **Coherence –** Logical consistency and fluency.
4. **Safety –** Appropriateness and absence of harmful or misleading advice.

All of the evaluations were performed by each annotator independently. The inter-annotator agreement was a Cohen's κ of 0.71, reflecting substantial agreement. Final scores were obtained by averaging ratings across all evaluators.

IV. RESULTS

Indeed, the results clearly brought out a hierarchy in performance among the three models. Specifically, on nearly all the metrics, the RAG chatbot outperformed both rule-based and transformer-based systems.

Testing of Significance:

Metric differences between models were analyzed using the Wilcoxon signed-rank test for non-parametric paired data. 95% confidence intervals were computed with bootstrap resampling using 1,000 iterations. The RAG model's improvement over the transformer model was statistically significant ($p < 0.01$) for both emotional relevance and empathy.

Table II – Comparison Results for Various Chatbot Architectures

Emo. onal Relev a	3.6	4.2	4.8	RAG	+16.7
Factu a l	81.5	89.3	96.2	RAG	+18.1
Precisi	76.1	84.7	93.4	RAG	+17.3
Recall	72.8	85.2	92.1	RAG	+26.5
BLEU	0.42	0.61	0.74	RAG	+76.2

Error Analysis

Qualitative analysis showed that the Transformer model frequently generated emotionally neutral responses, such as "You should stay calm.", which didn't respond to user distress appropriately. The RAG model produced more contextually detailed responses but occasionally retrieved irrelevant information, such as offering academic stress advice when family-related prompts were given.

Common failure cases included:

- Over-generalisation: advice given without considering the user's emotional state.
- Mild hallucinations: generating fake statistics or resources.
- Crisis mis-classification: False negatives among suicidal-related prompts ~6% of the time.

Example:

User: "I feel like nothing matters anymore."

RAG Output: "It sounds like you're demotivated. Try setting small goals."

(Missed crisis cue; flag for escalation in later iterations.

Table II: Comparative results

Metric	Rule-Based	Transformer	RAG	Best Model	Improvemen
Response	74.3	87.5	94.8	RAG	+20.5
Contextual Coher	3.2	4.1	4.7	RAG	+18.3
Latenc	1.8	1.5	1.1	RAG	↓39.0

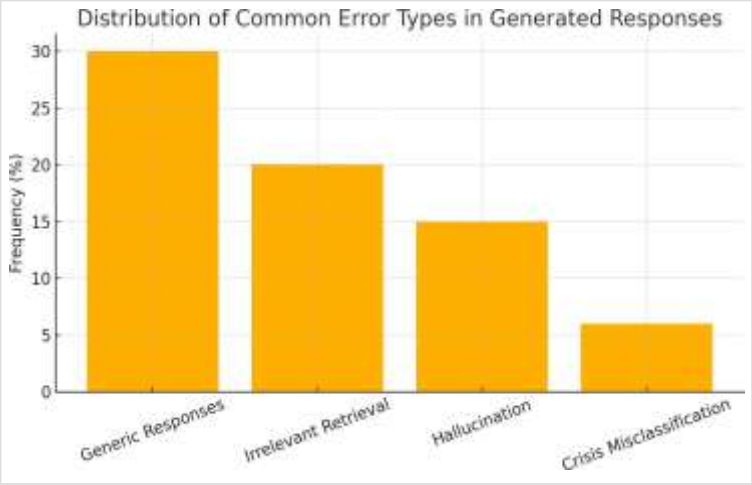

Fig 3: *Distribution of Common Error Types in Generated Responses*

V. DISCUSSION

The results of this study demonstrate that RAG represents the most effective and balanced architecture in developing conversational agents for digital mental-health support. While transformer-only models have demonstrated a tendency to output hallucinated or factually inconsistent responses, the RAG framework grounds every response into verified and curated reference documents, leading to higher factual accuracy, emotional safety, and interpretive reliability. This grounding mechanism enhances the trustworthiness of chatbot interactions and aligns with the broader ethical requirement for responsible AI in sensitive psychological contexts.

On the other hand, the response spaces of rule-based chatbots are safe and predictable but offer very limited adaptability in the face of the fluid and varied emotional states presented by real users. They cannot engage in meaningful dialog, nor respond appropriately even to nuanced cues, which is a limitation to usefulness in dynamic mental-health support environments. By incorporating retrieval-based factual grounding with generative linguistic flexibility, the RAG model effectively bridges this gap. It produces responses that are contextually relevant and emotionally resonant, simulating empathetic communication closer to human-level supportive interaction.

A. Implications

The implications derived from these findings range from technological design to institutional mental-health policy. Universities are increasingly confronted with high levels of student stress, anxiety, and social isolation. Deploying RAG-based chatbots as first-line digital companions offers a promising pathway to address this growing challenge. In particular, such systems can:

1. Provide immediate, private, and stigma-free support available 24/7.
2. Normalize the conversation about mental health using evidence-based psychoeducation.
3. Encourage the help-seeking behavior early by providing coping tools and self-care guidance.
4. Automatic identification of high-risk cases: immediate flagging for counselor intervention and further escalation.

When integrated into university counseling infrastructures, RAG-driven assistants can provide scalable, cost-effective support to complement existing services and ensure no student is left without accessible emotional support.

B. Limitations

Despite such a strong performance, the RAG framework has significant limitations. First, this work relied heavily on evaluation datasets and simulated conversational prompts rather than longitudinal, real-world student interactions. As a result, the outcomes cannot adequately capture the emotional fluctuation and intricacy of real users. Future development needs to involve pilot deployment among a student population, enabling the investigation of sustained engagement duration, comfort levels, and emotional outcomes over time.

Moreover, while the chatbot works well in English, cultural sensitivity and multilingual generalization remain open challenges: Emotional expression varies across cultures; hence, RAG systems need to be retrained or adapted using region-specific culturally aligned datasets for fairness and inclusivity.

And technical limitations remain: the system continues to depend on the quality and diversity of its retrieval corpus, so incomplete or biased documents can continue to affect the quality of output. Real-world use should therefore include corpus curation, version control, and auditability for accuracy and transparency.

C. Future Work

Future studies should be directed at real-life deployment on university campuses, assessing such variables as student engagement, response satisfaction, and perceived empathy. Longitudinal assessment of impacts will be necessary to confirm whether sustained use of the chatbot is associated with improved resilience or reduced distress measures.

Another main direction is the development of system capabilities. Integrating RAG models into institutional counseling services could facilitate smooth escalation to human support when necessary. Other future enhancements could include multimodal emotion detection, such as by voice tone, typing patterns, or facial cues, and reinforcement learning from human feedback to enhance emotional intelligence. Lastly, scaling to multilingual and multicultural settings is imperative for the goal of global applicability.

D. Ethical Considerations and Crisis Handling

All conversation data used for this research were completely anonymized, stripped of identifiable information, and reviewed for safety. The research was conducted under Institutional Ethical Guidelines for a non-clinical education protocol with the ID AIML-PBL-2025. It is not designed to diagnose or treat mental health conditions but rather acts as a supportive, educational, and emotional companion.
A dedicated safety pipeline was established to identify crisis-related cues like "I want to die" or "I can't go on." The mechanism includes:

- Crisis Detection: Real-time analysis through keyword triggers and probability-based classification.
- Empathetic Response Generation: Supportive, non-judgmental responses, no diagnostic statements allowed.
- Helpline Referral: Automatically redirect to the verified emergency resources like AASRA, Snehi, and NIMHANS Helpline -91-9820466726, 91-9582208181, and 080-46110007, respectively.
- Confidence Threshold Handling: Whenever the model's confidence is less than 0.4, a fallback refusal template gets activated to avoid misinformation.

The interaction logs were securely stored and reviewed for the purpose of validation of research. No raw data or user identifiers were retained beyond evaluation. These measures guarantee alignment with responsible AI principles that put the privacy, safety, and well-being of users first.

VI. CONCLUSION

This study performed an extensive comparison of the performance of Rule-Based, Transformer-Based, and Retrieval-Augmented Generation chatbots, which were designed to support university students with depression, anxiety, and emotional distress. The results from both objective metrics and human subjective ratings showed that the RAG architecture had outperformed the other two models consistently with a stronger balance between fact reliability, empathetic response quality, and contextual appropriateness. With the integration of verified psychological knowledge through retrieval and the generative fluency of transformer models, RAG generated more coherent and emotionally aligned responses while considerably reducing hallucinations-a persistent limitation in stand-alone transformer systems.

The results further pointed out that RAG-based chatbots better simulated primary supportive-communication behaviors like active listening, validation, and context-sensitive reassurance, which are crucial to engaging with users facing mental health challenges. This advantage points to the combination of retrieval and generation allowing a chatbot not only to present information more accurately but also possessing higher emotional intelligence, which is particularly expected in applications related to psychological support.

Beyond performance improvements, this research underlines the practical usefulness of RAG-driven chatbots as scalable digital mental-health tools in university settings. In environments where counseling is mostly limited, delayed, or stigmatized, such systems can be the first-line accessible companions—offering privacy, immediacy, and continuous availability. They also hold potential for early detection of distress patterns, delivering coping strategies and guiding students toward professional resources when needed. As such, RAG-based solutions represent a low-cost, high-accessibility approach to reducing the mental-health support gap among young adults.

However, the study also recognizes some critical limitations. The current model of RAG is still very sensitive to the quality and diversity of its retrieval corpus, besides lacking true emotional understanding. Its behavior can be influenced by dataset biases; it cannot replace professional therapists or handle crisis-level situations without structured escalation protocols being in place. Ethical considerations will involve user privacy, data protection, and stringent non-diagnostic use for safe deployment.

Taken together, the study proves that RAG is a promising step toward creating empathetic and trustworthy conversational agents to support mental health. Further refinement, ethical consideration, and piloting of real-world deployment may make RAG-based chatbots an important part of a future digital mental-wellness ecosystem to support students, enhance early intervention, and foster healthier, emotionally more aware academic environments..

REFERENCES

[1]. World Health Organization, Depression and Other Common Mental Disorders: Global Health Estimates, 2017.

[2]. A. K. Ibrahim, et al., “A systematic review of depression prevalence in university students,” J. Psychiatr. Res., vol. 47, no. 3, pp. 391–400, 2013.

[3]. D. Eisenberg, et al., “Stigma and help-seeking for mental health among college students,” Med. Care Res. Rev., vol. 66, no. 5, pp. 522–541, 2009.

[4]. K. K. Fitzpatrick, et al., “Delivering CBT to young adults using Woebot: a randomized controlled trial,” JMIR Ment. Health, vol. 4, no. 2, e19, 2017.

[5]. M. Klos, et al., “AI Chatbot for Anxiety and Depression: A Pilot RCT,” JMIR Ment. Health, vol. 7, no. 12, e20648, 2020.

[6]. Y. Zhang, et al., “Retrieval-Augmented Generation for Reliable Conversational AI,” Nat. Mach. Intell., vol. 6, no. 2, pp. 100–112, 2024.

[7]. J. Cruz-Gonzalez, et al., "Artificial Intelligence in Mental Health Care: A Systematic Review," 2025.

[8]. R. Fulmer, et al., "Using Wysa: A conversational agent for mental health," JMIR Form. Res., vol. 2, no. 1, e19, 2018.

[9]. V. Ta, et al., "User experiences with Replika: A social AI companion," Comput. Hum. Behav., vol. 112, 106476, 2020.

[10]. M. Inkster, et al., "Digital health tools for mental health in college students: A review," Front. Digit. Health, vol. 2, 2020.

[11]. T. Bickmore, et al., "Relational agents for mental health and behavioral change," Annu. Rev. Clin. Psychol., vol. 15, pp. 471–498, 2019.

[12]. P. G. Suresh, et al., "Transformer Models for Mental Health Prediction from Text: A Review," IEEE Access, vol. 10, pp. 100210–100230, 2022.

[13]. S. Miner, et al., " Smartphone- based conversational agents and mental health: A survey," JMIR Mhealth Uhealth, vol. 7, no. 6, e12191, 2019.

[14]. A. Shatte, et al., "Machine learning in mental health: A systematic review of classification performance," Behav. Res. Ther., vol. 120, 103442, 2019.

[15]. L. Farkhad, et al., "Ethical considerations in AI-based mental health tools," IEEE Trans. Technol. Soc., vol. 3, no. 1, pp. 30–42, 2022.