**International Journal of Advances in Applied Sciences (IJAAS)**
Vol. 14, No. 3, September 2025, pp. 849~858
ISSN: 2252-8814, DOI: 10.11591/ijaas.v14.i3.pp849-858


# Large language models and retrieval-augmented generation-based chatbot for adolescent mental health

**Andi Riansyah[1], Imam Much Ibnu Subroto[1], Intan Nur'aini[1], Ratna Supradewi[2], Suyanto Suyanto[3]**
[1]Department of Informatics Engineering, Faculty of Industrial Technology, Universitas Islam Sultan Agung, Semarang, Indonesia
[2]Department of Psychology, Faculty of Psychology, Universitas Islam Sultan Agung, Semarang, Indonesia
[3]Faculty of Nursing, Universitas Islam Sultan Agung, Semarang, Indonesia

**Article Info**

***Article history:***

Received Sep 12, 2024
Revised Apr 17, 2025
Accepted Jun 10, 2025

***Keywords:***

Artificial intelligence
Chatbot
Large language model
Mental health services
Retrieval-augmented generation
Youth counseling

**ABSTRACT**

Access to fast and efficient information is crucial in today's digital era, especially for teenagers in obtaining mental health services. The manual method used by Youth Information and Counselling Centre (PIK R) to provide mental health information requires significant time and effort. This research presents an AI-based solution by developing a chatbot system using retrieval-augmented generation (RAG) and large language models (LLM). This chatbot is designed to provide accurate and effective mental health information for teenagers throughout the day. An analysis of a dataset consisting of articles on teenage mental health and data from the Alodokter website was used as the basis for the development of this chatbot. The research results show that the chatbot is capable of providing relevant and accurate information, with evaluations using the recall-oriented understudy for gisting evaluation (ROUGE) score method yielding an average of ROUGE-1 with a precision of 87.8%, recall of 83.0%, and F1-measure of 84.0%; ROUGE-2 with a precision of 82.8%, recall of 76.8%, and F1-measure of 78.2%; and ROUGE-L with a precision of 88.0%, recall of 82.6%, and F1-measure of 83.4%. These findings indicate the potential use of chatbots as an effective tool to support the mental health of adolescents.



***Corresponding Author:***

Andi Riansyah
Department of Informatics Engineering, Faculty of Industrial Technology, Universitas Islam Sultan Agung
St. Raya Kaligawe Km. 4, Semarang 50112, Indonesia
Email: andi@unissula.ac.id

## 1. INTRODUCTION

The access to information for teenagers has been revolutionized because of the fast-paced development of communication technology. In the current age of the digital world, chatbots have become a tool that is essential for 24/7 support and enabling natural language interaction [1], [2]. Though past research has shown that conversational agents can provide effective solutions in various settings, little is known regarding how they can be leveraged to support mental health, particularly for teenagers [3].

There is evidence that the mental disorders prevalence among adolescents is increasing, as indicated by an increasing number of studies attesting to widespread emotional and behavioral disorders like depression and anxiety [4]–[6]. The conventional services, as offered by the Youth Information and Counselling Centre (PIK R), lack timely and precise mental health information [7]. In addition, while other research has assessed the usefulness of chatbots in conveying information, such research frequently falls short in terms of providing the intense analysis required for complicated or big questions, such as those involving multiple-choice exams or measurable data retrieval [8]–[11]. A review of existing literature indicates that while past research has employed chatbots to optimize the accessibility of information, a number of important matters have not been sufficiently addressed.

For one, existing solutions provide generalized responses that cannot fulfill the intricate needs that accompany adolescent mental health. Secondly, there is minimal discourse on the methods of overcoming challenges associated with information retrieval and contextual comprehension. Recent advancements in large language models (LLMs) and retrieval-augmented generation (RAG) have shown potential in addressing these gaps by enabling more contextually relevant and data-informed responses [12], [13]. Comparative research on the use of these technologies more specifically for offering mental support for teenagers has yet to be conducted, however. The gaps are filled in our research through the introduction of a novel chatbot system that combines RAG and LLM paradigms. With the use of this in-depth methodology, the chatbot can not only recognize the nuance of user queries but also derive and incorporate accurate and related information from a carefully filtered pool of user-created content that is retrieved from reliable websites and academic papers related to mental health.

In this paper, we test the performance of advanced LLMs, namely, generative pre-trained transformer (GPT)-4 and GPT-3.5 Turbo, on tasks like yes-or-no questions, multiple-choice answering, and numerical information extraction [14], [15]. By comparing these models directly to expert-annotated data and evaluating performance against metrics like recall-oriented understudy for gisting evaluation (ROUGE) scores, this research builds on the efforts of earlier contributions. This research provides several significant enhancements. Firstly, we propose a unifying framework that brings together the accurate information retrieval capabilities of RAG and the natural language comprehension capabilities of LLM. Such a fusion has not been exhaustively explored in the domain of adolescent mental health. Secondly, we give a detailed evaluation of the performance of the system, listing its strengths and limitations while discussing the implications of our findings. While the combined strategy succeeds in yielding accurate and contextually appropriate information, based on our analysis, challenges remain, particularly for complex linguistic constructions and complicated questions. Last but not least, the implications of our study suggest that this chatbot system can significantly enhance the quality and availability of mental health services for adolescents [16]–[18]. While previous studies have explored the use of chatbots in various information retrieval contexts, they did not explicitly address their influence on providing nuanced, contextually rich mental health support specifically tailored for adolescents. This research seeks to bridge this gap by combining accurate information retrieval and natural language understanding to enhance adolescent mental health support [19], [20].

## 2. METHOD

### 2.1. Research design

This research begins with planning and preparation, where the problem and the need for a mental health chatbot for teenagers are identified. A literature review is conducted to understand technologies such as LLMs and RAG, as well as to establish evaluation metrics such as accuracy, relevance, and response speed. Next, mental health data is collected from trusted sources, processed, and labeled to facilitate its use in the model. Once the data is ready, the chatbot system is developed by integrating an LLM as the core and RAG to retrieve relevant information from a knowledge database. The model is then trained using the prepared data and initially tested to ensure its functionality. The system is evaluated by measuring its performance using the established metrics, and teenagers are involved as users to provide feedback. The evaluation results are analyzed to identify shortcomings and areas for improvement. Based on the evaluation results, the system is refined through iterations, such as updating the model and knowledge database, and then retested to ensure improved performance. After finalization, the chatbot is implemented and launched on platforms easily accessible to teenagers, such as apps or websites. Periodic monitoring is conducted to ensure the system remains relevant and accurate. The research results are reported and published in journals or scientific conferences, while the chatbot system is disseminated to the public and relevant stakeholders. Long-term evaluation is also carried out to monitor the chatbot's impact on teenagers' mental health, such as increased awareness or reduced stress levels, and periodic updates are performed to maintain the system's quality. Through this process, the research is expected to produce an effective and beneficial chatbot for teenagers in the context of mental health.

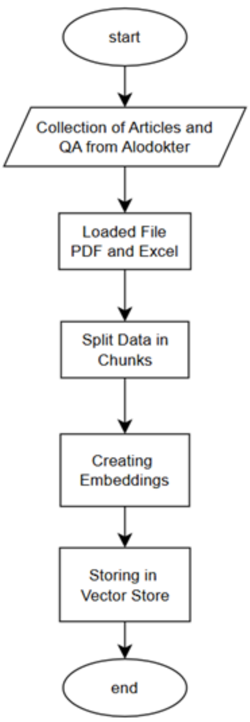
### 2.2. Research procedures

In this research, the dataset used was obtained from two main sources: articles related to mental health and data from the Alodokter website. The mental health articles were collected from various scientific publications and other trusted sources, containing in-depth information on various aspects of mental health, such as anxiety disorders, depression, and coping techniques [21]. Meanwhile, data from Alodokter consists of frequently asked questions and answers by users of the site regarding mental health topics [22]. The dataset from Alodokter provides valuable insights into common mental health issues faced by the community, as well as the medical responses offered by experts on the platform [21], [22]. The combination of these two sources provides a strong foundation for further analysis, especially in developing a chatbot system capable of delivering relevant and accurate mental health information and support.

The data processing, as shown in Figure 1, involves chunking, which is dividing the text into smaller, relevant pieces to facilitate embedding. Chunking is done to ensure that the embedded information remains structured and easily accessible by the model. An effective chunking strategy can also significantly enhance the speed and accuracy of the mode.

After chunking, the processed data is transformed into vectors using the OllamaEmbedding model. This transformation of text into vectors is crucial in the data preprocessing stage so that the collected data can be processed and understood by the LLM model [23], [24]. This is because the development of the chatbot involves integrating LLM and RAG. LLM is used to understand the context of the questions, while RAG assists in retrieving relevant information from the embedded data. This system is designed to provide accurate and relevant answers based on the available data.

In Figure 2, the stages of system modeling depicted in the flowchart above are divided into several phases, which include embedding the user's input using the same model as the embeddings during data collection [12]. This is followed by a similarity search for retrieving or extracting the relevant context between the input data and the data in the database, using the create_retrieval_chain from the library. From the relevant contexts that have been retrieved, they will be generated with llama3, and the results from the llama3 generation will be presented to the user [13].



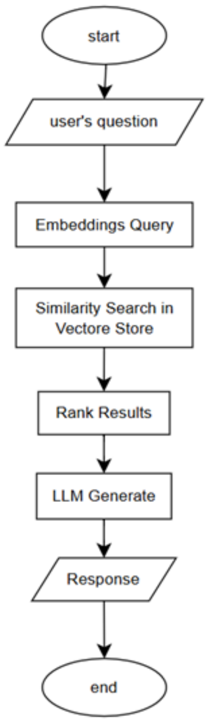




Figure 1. Data collection flowchart Figure 2. System modeling flowchart

### 2.3. Developments of chatbots

The chatbot system integrates RAG architecture and LLM. RAG consists of a retriever and a generator [25], where the retriever fetches relevant documents and the generator uses LLM to produce answers. The OllamaEmbedding model is used to convert text data into vectors utilized by the retriever. LLM is employed to understand the context and generate answers based on the retrieved data [13].

The development of the chatbot involves the integration of the OllamaEmbedding model, RAG, and LLM. The system architecture diagram can be seen in Figure 3, which illustrates how data is processed and how the system components interact. The workflow diagram for data retrieval and processing is presented in Figure 1. The system is designed to provide relevant answers based on the embedded data [22].

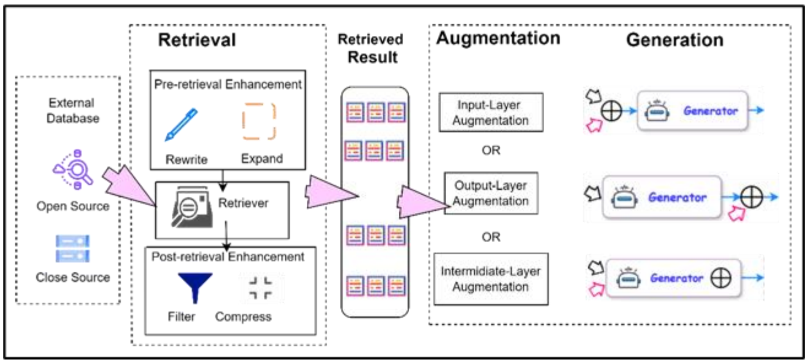



Figure 3. RAG architecture

Based on Figure 3, the RAG architecture has three main components:

– Retriever: this component is responsible for retrieving relevant documents from the knowledge source based on the given query. Sources can be either open source or closed source [26]. A retriever functions to measure the relevance between a query and documents for effective information retrieval. Pre-retrieval and post-retrieval enhancement techniques are used to improve the accuracy and relevance of retrieval results.
– Augmentation: this component processes the retrieved documents to extract important information. Augmentation is applied at three stages: input layer (combining documents with the original query), output layer (combining retrieval and generation results), and intermediate layer (integrating results in the internal generator layer) [26]. Intermediate-layer augmentation offers the potential for performance improvement, but it requires deeper access to the LLM.
– Generation: this component is a language model that produces final answers based on the input and the augmented retrieval results. Augmentation can be performed at three stages (input, output, or intermediate), depending on the system's needs [26].

Figure 1 shows how retrieval and generation techniques work together to produce relevant answers. Augmentation at various layers ensures accurate and contextually appropriate answers. Figure 2 illustrates the application of RAG in chatbots. Figure 4 illustrates the architecture of the RAG system applied to a chatbot for mental health information services. The process begins with input in the form of a document or article (the red file symbol on the left), which is then divided into smaller pieces through a chunking process [22]. These pieces are processed by the retriever component (represented by the orange chip symbol), which functions to search for and retrieve relevant information from a vector-based database (blue bucket symbol).

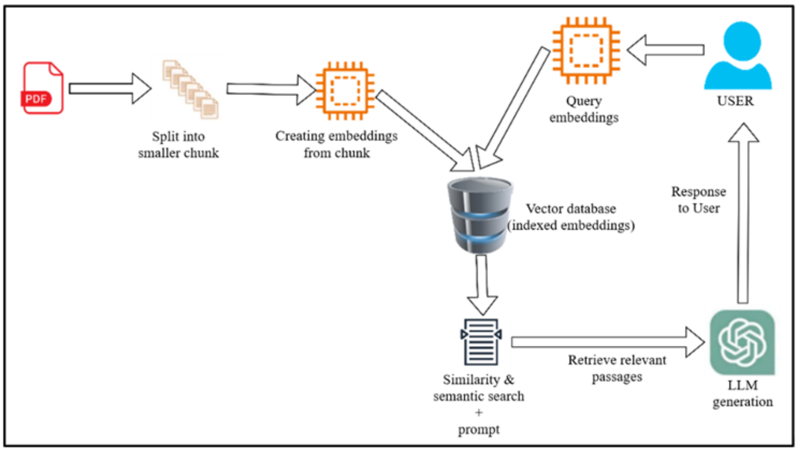



Figure 4. Implementation of RAG and LLM in chatbots

The retriever conducts a search based on the semantic similarity between the user's query (the person symbol in the upper right) and the processed documents [27]. Relevant information is sent to the generator (another chip symbol on the right), which uses OpenAI's LLM to generate answers based on augmented search results. The final result is a response provided back to the user, tailored to the relevant context. Overall, this diagram illustrates how the combination of similarity search and prompting within the RAG system enables the chatbot to provide more accurate and informative responses, assisting users in obtaining mental health information quickly and accurately.

### 2.4. Evaluation

The evaluation of LLMs is an important component in contemporary artificial intelligence research because it helps to understand their operational capabilities and shows the way for future advancements [28]. The chatbot is tested by providing various questions to assess the system's ability to provide accurate answers. The evaluation was conducted using ROUGE metrics, namely ROUGE-1, ROUGE-2, and ROUGE-L, to measure the quality and relevance of the answers [29]. This test aims to ensure that the chatbot can provide effective mental health information that meets the needs of adolescents [30]–[32].

#### 2.4.1. ROUGE-1

ROUGE-1 is an evaluation model that serves to measure similarity based on unigrams (single words) between the output text of a model and reference text. This method is useful for evaluating how many important words in the reference appear in the model's output. ROUGE-1 is calculated as the ratio of the number of unigrams that overlap with the reference to the total unigrams in the reference or model output, depending on whether precision, recall, or F-measure is being calculated. Here is the calculation formula:

$$Precision = \frac{number\ of\ overlapping\ unigrams}{total\ unigrams\ in\ the\ generated\ summary} \tag{1}$$

$$Recall = \frac{number\ of\ overlapping\ unigrams}{total\ unigrams\ in\ the\ reference\ summary} \tag{2}$$

$$F - Measure = 2 \times \frac{precision \times recall}{precision \times recall} \tag{3}$$

#### 2.4.2. ROUGE-2

ROUGE-2 is an evaluation model that functions to measure similarity based on bigrams (two consecutive words) between the model's output and reference texts. This method is used to evaluate more complex word sequences, providing deeper insights into the cohesion and context captured by the model. ROUGE-2 is calculated as the ratio of the number of bigrams that overlap with the reference to the total number of bigrams in the reference or the model output. Here is the calculation formula:

$$Precision = \frac{number\ of\ overlapping\ bigrams}{total\ bigrams\ in\ the\ generated\ summary} \tag{4}$$

$$Recall = \frac{number\ of\ overlapping\ bigrams}{total\ bigrams\ in\ the\ reference\ summary} \tag{5}$$

$$F - Measure = 2 \times \frac{precision \times recall}{precision \times recall} \tag{6}$$

#### 2.4.3. ROUGE-L

ROUGE-L is an evaluation model that serves to measure similarity based on the longest common subsequence (LCS) between the model output and the reference text. This model is used to evaluate the similarity of long sequences between the model's output and the reference, allowing for the assessment of how well the generated sequence reflects the expected sequence in the reference. Here is the calculation formula:

$$Precision = \frac{Longest\ Common\ Subsquent\ (LCS)}{total\ words\ in\ the\ generated\ summary} \tag{7}$$

$$Recall = \frac{Longest\ Common\ Subsquent\ (LCS)}{total\ words\ in\ the\ reference\ summary} \tag{8}$$

$$F - Measure = 2 \times \frac{precision \times recall}{precision \times recall} \tag{9}$$

## 3. RESULTS AND DISCUSSION

In this section, the evaluation results of the chatbot system designed to assist teenagers in obtaining mental health information will be analyzed in detail. The main objective of this evaluation is to assess the accuracy, relevance, and effectiveness of the chatbot in providing appropriate solutions for the situations faced by users. The evaluation results aim to provide a clear picture of the system's performance and its contribution to supporting the mental health of adolescents.

The evaluation method used is the measurement of ROUGE scores, which is applied to assess the extent to which the answers provided by the chatbot align with the predetermined reference answers. This assessment includes ROUGE-1, ROUGE-2, and ROUGE-L, each of which provides insights into the precision, recall, and F-measure of the system. In addition, the user interface (UI) aspect is also evaluated to ensure that the system is not only effective in terms of content but also easy to use for end users.

To evaluate the chatbot's ability to respond to various types of user inputs, a set of five representative questions was selected. These questions cover both general knowledge and specific user scenarios related to adolescent mental health: i) What is depression? ii) What is emotional mental disorder? iii) What is the development of emotional mental health like between adolescents living in the "Putri Aisyiah" orphanage and those living at home in Sukorejo Village, Kendal Regency? iv) I am an 18-year-old freshman, and I often forget things and can't focus at all. Is there a way for me to regain my concentration? and v) I often laugh excessively and sometimes cry uncontrollably; what symptoms could that be?

Table 1 shows the results of the chatbot performance evaluation based on the ROUGE score, which measures the quality of the chatbot's responses by comparing them to reference answers. This evaluation involves three main metrics: ROUGE-1, ROUGE-2, and ROUGE-L. Each metric is measured from three aspects, namely precision, recall, and F1-measure.

- ROUGE-1 measures the similarity based on unigrams between the chatbot's responses and the reference. The results show an average precision of 87.8%, recall of 83%, and F1-measure of 84%. This indicates that the chatbot is capable of recognizing a majority of the relevant keywords that align with the reference answers, with a good balance between precision and coverage.
- ROUGE-2 measures the suitability based on bigrams. The average precision for this metric reaches 82.8%, recall is at 76.8%, and the F1-measure is 78.2%, which is slightly lower than ROUGE-1. This indicates that when the chatbot is asked to understand the relationships between words, its performance slightly declines.
- ROUGE-L measures the similarity based on the longest matching word sequence between the chatbot's response and the reference. The average precision is 88%, recall is 82.6%, and the F1-measure reaches 83.4%.

Table 1. Results of the ROUGE score evaluation

| Question | ROUGE-1 | | | ROUGE-2 | | | ROUGE-L | | |
|---|---|---|---|---|---|---|---|---|---|
| | Precision (%) | Recall (%) | F1-measure (%) | Precision (%) | Recall (%) | F1-measure (%) | Precision (%) | Recall (%) | F1-measure (%) |
| i | 70 | 89 | 81 | 64 | 78 | 70 | 74 | 90 | 81 |
| ii | 100 | 94 | 97 | 100 | 94 | 97 | 100 | 94 | 97 |
| iii | 72 | 83 | 77 | 59 | 68 | 63 | 70 | 81 | 75 |
| iv | 100 | 96 | 98 | 98 | 94 | 96 | 99 | 95 | 97 |
| v | 97 | 53 | 67 | 93 | 50 | 65 | 97 | 53 | 67 |
| Average | 87.8 | 83 | 84 | 82.8 | 76.8 | 78.2 | 88 | 82.6 | 83.4 |

Our findings indicate that integrating RAG and LLM significantly enhances the contextual understanding and accuracy of chatbot responses, outperforming simpler chatbot systems explored in prior research. This suggests that employing a retrieval-augmented approach effectively supports detailed and context-specific information delivery without negatively impacting response quality. These results align favorably with similar research, indicating enhanced accuracy and relevance when leveraging advanced LLM techniques.

Figure 5 presents an interactive chatbot interface specifically designed to provide mental health information and advice for adolescents. Users can pose various questions, such as issues related to concentration or forgetfulness, and the chatbot responds with practical recommendations and references to relevant research. This user-friendly interface enables teenagers to quickly and conveniently access valuable mental health information.

In this experiment, the user posed questions related to the difficulties in concentration experienced, which often occur among students. The chatbot responds by providing a comprehensive explanation of the possible causes of concentration difficulties, such as lack of sleep, stress, and multitasking habits. In addition, the chatbot provides practical suggestions to address these issues, such as maintaining a sufficient sleep pattern, managing stress, and reducing gadget use before bedtime.

The responses provided by the chatbot in this experiment demonstrate a fairly high level of accuracy and relevance. The chatbot successfully identified the issues faced by the user and provided answers that were relevant to the context of the questions. This indicates that the system can recognize and provide accurate answers related to the mental health issues faced by adolescents, in accordance with the context that has been embedded into the model. Figure 6 shows the conversation page between the user and the bot, where the input questions are outside the context present in the vector store. For every question that is out of context, the bot will confirm that there is no relation between the question and the topic being discussed.

The main objective of this research is to develop and evaluate an automated system that uses advanced language models and retrieval techniques to generate summaries of mental health-related information for adolescents. This research aims to improve the accessibility and accuracy of mental health information provided to young individuals, ensuring that the information is both relevant and easy to understand. The hypothesis of this study is that the use of a RAG approach will enhance the accuracy and comprehensiveness of the generated summaries compared to traditional summarization methods. Specifically, it is hypothesized that integrating retrieval techniques with advanced language models will improve both the precision and recall of the system, leading to summaries that better reflect the key points of mental health content while maintaining relevance and clarity for adolescent users.

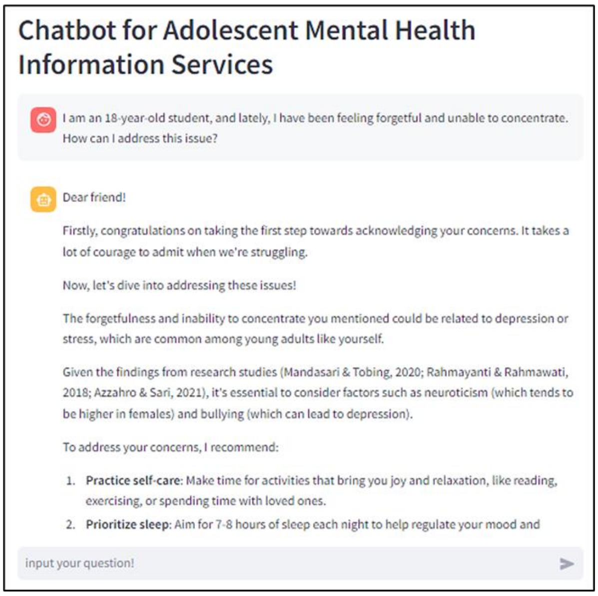



Figure 5. Chatbot conversation page

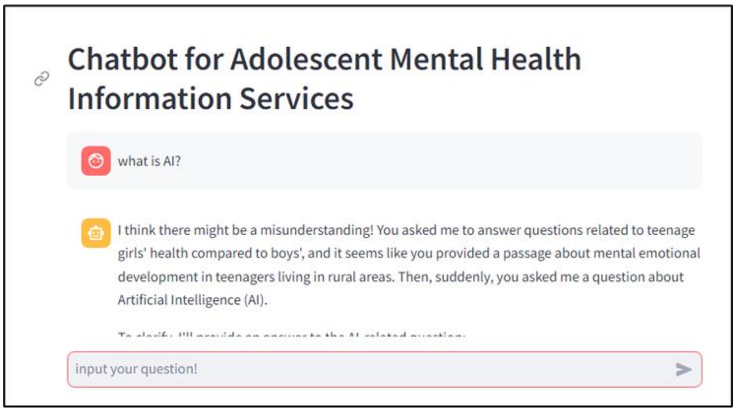



Figure 6. Chatbot conversation out of context

## 4. CONCLUSION

Based on the findings of this study, the chatbot system developed using LLM and RAG methods shows strong potential as a mental health information service for adolescents. Evaluation results confirm a satisfactory performance, with an average ROUGE score of ROUGE-1: precision 87.8%, recall 83.0%, and F1-measure 84.0%; ROUGE-2: precision 82.8%, recall 76.8%, and F1-measure 78.2%; and ROUGE-L:

precision 88.0%, recall 82.6%, and F1-measure 83.4%, indicate the developed chatbot system effectively addresses adolescent mental health information needs. Our findings offer definitive proof that employing RAG and LLM significantly enhances chatbot performance in terms of contextual accuracy and information relevance, rather than being limited by traditional response generation methods. Looking ahead, future research could expand on this approach by incorporating real-time connections to mental health professionals or adding personalized, context-aware recommendations based on individual user profiles. Additionally, exploring adaptations for multiple languages and cultural settings can help increase the system's reach and relevance. Overall, these results underscore the promise of AI-driven solutions like chatbots in enhancing mental health resources for adolescents, laying a foundation for broader applications in digital healthcare and well-being interventions.

## ACKNOWLEDGMENTS

We would first like to thank the psychologists who verified the dataset we collected; without them, this research would not have been possible. In addition, we thank our academic advisor for their invaluable suggestions and assistance during the investigation. This research aims to benefit everyone who is striving to improve mental health care for adolescents in Indonesia.

## FUNDING INFORMATION

This research was funded by Universitas Islam Sultan Agung, Semarang, Indonesia. The authors gratefully acknowledge the financial support provided by the university in conducting and completing this study.

## AUTHOR CONTRIBUTIONS STATEMENT

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration.

| **Name of Author** | **C** | **M** | **So** | **Va** | **Fo** | **I** | **R** | **D** | **O** | **E** | **Vi** | **Su** | **P** | **Fu** |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Andi Riansyah | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Imam Much Ibnu Subroto | ✓ | ✓ | | ✓ | ✓ | ✓ | | | ✓ | ✓ | | | ✓ | |
| Intan Nur'aini | ✓ | ✓ | ✓ | | ✓ | ✓ | ✓ | | ✓ | ✓ | ✓ | | | |
| Ratna Supradewi | ✓ | ✓ | | | ✓ | | | | ✓ | | | | | |
| Suyanto Suyanto | ✓ | ✓ | | | ✓ | | | | ✓ | | | | | |

| | | |
|---|---|---|
| C : **C**onceptualization | I : **I**nvestigation | Vi : **Vi**sualization |
| M : **M**ethodology | R : **R**esources | Su : **Su**pervision |
| So : **So**ftware | D : **D**ata Curation | P : **P**roject administration |
| Va : **Va**lidation | O : Writing - **O**riginal Draft | Fu : **Fu**nding acquisition |
| Fo : **Fo**rmal analysis | E : Writing - Review & **E**diting | |

## CONFLICT OF INTEREST STATEMENT

The authors declare that there are no known conflicts of interest, either financial or non-financial, that could have influenced the results of this research. This study was conducted independently, and all findings have been presented objectively based on the data and analysis carried out by the research team.

## DATA AVAILABILITY

The data supporting the findings of this study were obtained from various scientific articles related to adolescent mental health. The data were processed and anonymized prior to use in the development of the chatbot system. Due to copyright restrictions and privacy considerations, the data are not publicly available but may be obtained from the corresponding author upon reasonable request.

## REFERENCES

[1] Y. Zhu, J. Liang, and Y. Zhao, "Expert or partner: the matching effect of AI chatbot roles in different service contexts," *Electronic Commerce Research and Applications*, vol. 71, p. 101496, May 2025, doi: 10.1016/j.elerap.2025.101496.
[2] M. Skjuve, A. Følstad, K. I. Fostervold, and P. B. Brandtzaeg, "My chatbot companion - a study of human-chatbot relationships," *International Journal of Human-Computer Studies*, vol. 149, p. 102601, May 2021, doi: 10.1016/j.ijhcs.2021.102601.

[3] A. Riansyah, M. Qomaruddin, M. Indriastuti, and M. Sagaf, "Clustering digital transformation of small and medium enterprises (SMEs) using fuzzy K-means method," in *2023 10th International Conference on Electrical Engineering, Computer Science and Informatics*, IEEE, Sep. 2023, pp. 540–544. doi: 10.1109/EECSI59885.2023.10295664.

[4] S. Ito, E. Furukawa, T. Okuhara, H. Okada, and T. Kiuchi, "Leveraging artificial intelligence chatbots for anemia prevention: a comparative study of ChatGPT-3.5, Copilot, and Gemini outputs against google search results," *PEC Innovation*, vol. 6, p. 100390, Jun. 2025, doi: 10.1016/j.pecinn.2025.100390.

[5] H. Wu, X. Yin, K. Miao, J. Yin, and H. Zhao, "Cardiorespiratory fitness in adolescents with mental sub-health in China," *Annales Médico-psychologiques, revue psychiatrique*, vol. 179, no. 1, pp. 33–38, Jan. 2021, doi: 10.1016/j.amp.2020.08.002.

[6] T. A. Matthews, H. Shao, M. Forster, and I. Kim, "Associations of adverse childhood experiences with depression and anxiety among children in the United States: racial and ethnic disparities in mental health," *Journal of Affective Disorders*, vol. 362, pp. 645–651, Oct. 2024, doi: 10.1016/j.jad.2024.07.121.

[7] A. Marwanto, A. Riansyah, and M. K. Anwar, "An improvement of apple leaf diseases detection using convolutional neural network methods based on mobile systems," in *2024 11th International Conference on Electrical Engineering, Computer Science and Informatics*, IEEE, Sep. 2024, pp. 440–447. doi: 10.1109/EECSI63442.2024.10776345.

[8] E. Kusumaningtyas, E. Laurentino, and A. Barakbah, "Responsive chatbot using named entity recognition and cosine similarity," in *Proceedings of the 5th International Conference on Applied Science and Technology on Engineering Science*, SCITEPRESS - Science and Technology Publications, 2022, pp. 239–245. doi: 10.5220/0011756300003575.

[9] D. Gao *et al.*, "FashionGPT: LLM instruction fine-tuning with multiple LoRA-adapter fusion," *Knowledge-Based Systems*, vol. 299, p. 112043, Sep. 2024, doi: 10.1016/j.knosys.2024.112043.

[10] G. Jiang, Z. Ma, L. Zhang, and J. Chen, "EPlus-LLM: a large language model-based computing platform for automated building energy modeling," *Applied Energy*, vol. 367, p. 123431, Aug. 2024, doi: 10.1016/j.apenergy.2024.123431.

[11] Z. Pang, Y. Luan, J. Chen, and T. Li, "ParInfoGPT: an LLM-based two-stage framework for reliability assessment of rotating machine under partial information," *Reliability Engineering & System Safety*, vol. 250, p. 110312, Oct. 2024, doi: 10.1016/j.ress.2024.110312.

[12] L. Shi, M. Kazda, B. Sears, N. Shropshire, and R. Puri, "Ask-EDA: a design assistant empowered by LLM, hybrid RAG and abbreviation De-hallucination," *arXiv*, Jun. 2024, [Online]. Available: http://arxiv.org/abs/2406.06575

[13] J. Roberts, M. Baker, and J. Andrew, "Artificial intelligence and qualitative research: the promise and perils of large language model (LLM) 'assistance,'" *Critical Perspectives on Accounting*, vol. 99, p. 102722, Mar. 2024, doi: 10.1016/j.cpa.2024.102722.

[14] S.-C. Chien *et al.*, "Using large language model (LLM) to identify high-burden informal caregivers in long-term care," *Computer Methods and Programs in Biomedicine*, vol. 255, p. 108329, Oct. 2024, doi: 10.1016/j.cmpb.2024.108329.

[15] M. Stadler, M. Bannert, and M. Sailer, "Cognitive ease at a cost: LLMs reduce mental effort but compromise depth in student scientific inquiry," *Computers in Human Behavior*, vol. 160, p. 108386, Nov. 2024, doi: 10.1016/j.chb.2024.108386.

[16] M. H. S. Lesmana and M.-H. Chung, "Psychometric properties of the peer mental health stigmatization scale in Indonesian adolescents," *SSM - Mental Health*, vol. 7, p. 100400, Jun. 2025, doi: 10.1016/j.ssmmh.2025.100400.

[17] S. Wecht, M. Hendrixson, and A. Radović, "A mixed method investigation of parent-adolescent communication about mental health," *Journal of Adolescent Health*, vol. 75, no. 6, pp. 904–911, Dec. 2024, doi: 10.1016/j.jadohealth.2024.07.012.

[18] E. Jen, H.-Y. Chan, and H. N. Cheung, "Addressing adolescent social and emotional concerns: insights from loneliness, burnout, and preferred conversation topics in Asian and UK contexts post-pandemic," *Acta Psychologica*, vol. 247, p. 104326, Jul. 2024, doi: 10.1016/j.actpsy.2024.104326.

[19] M. Theberath *et al.*, "Effects of COVID-19 pandemic on mental health of children and adolescents: a systematic review of survey studies," *SAGE Open Medicine*, vol. 10, Jan. 2022, doi: 10.1177/20503121221086712.

[20] K. Q. Lin, M. C. C.-Breen, C. Nowell, A. F. Jorm, and M. B. H. Yap, "Parenting strategies to support adolescent mental health during a pandemic: a Delphi consensus study," *Mental Health & Prevention*, vol. 36, p. 200363, Dec. 2024, doi: 10.1016/j.mhp.2024.200363.

[21] A. D'Arqom *et al.*, "Mental health and decision-making participation of adolescent orphans: a pilot study," *Jurnal Keperawatan Padjadjaran*, vol. 11, no. 3, pp. 176–183, Dec. 2023, doi: 10.24198/jkp.v11i3.2358.

[22] S. E.-Lacko, P. F. Gilabert, and M. Knapp, "The double disadvantage faced by adolescents from low socioeconomic backgrounds with mental health problems affects earnings up to mid-life," *Social Science & Medicine*, vol. 362, p. 117385, Dec. 2024, doi: 10.1016/j.socscimed.2024.117385.

[23] L. Wu *et al.*, "A framework enabling LLMs into regulatory environment for transparency and trustworthiness and its application to drug labeling document," *Regulatory Toxicology and Pharmacology*, vol. 149, p. 105613, May 2024, doi: 10.1016/j.yrtph.2024.105613.

[24] D. Barman, Z. Guo, and O. Conlan, "The dark side of language models: exploring the potential of LLMs in multimedia disinformation generation and dissemination," *Machine Learning with Applications*, vol. 16, p. 100545, Jun. 2024, doi: 10.1016/j.mlwa.2024.100545.

[25] Y. Gao *et al.*, "Retrieval-augmented generation for large language models: a survey," *arXiv*, Mar. 2024, [Online]. Available: http://arxiv.org/abs/2312.10997

[26] W. Fan *et al.*, "A survey on rag meeting LLMs: towards retrieval-augmented large language models," in *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, New York, NY, USA: ACM, Aug. 2024, pp. 6491–6501. doi: 10.1145/3637528.3671470.

[27] K. Forth and A. Borrmann, "Semantic enrichment for BIM-based building energy performance simulations using semantic textual similarity and fine-tuning multilingual LLM," *Journal of Building Engineering*, vol. 95, p. 110312, Oct. 2024, doi: 10.1016/j.jobe.2024.110312.

[28] Y. Liu *et al.*, "Are LLMs good at structured outputs? a benchmark for evaluating structured output capabilities in LLMs," *Information Processing & Management*, vol. 61, no. 5, p. 103809, Sep. 2024, doi: 10.1016/j.ipm.2024.103809.

[29] A. Riansyah, S. Mulyono, and M. Roichani, "Applying fuzzy proportional integral derivative on internet of things for figs greenhouse," *IAES International Journal of Artificial Intelligence*, vol. 10, no. 3, pp. 536–544, Sep. 2021, doi: 10.11591/ijai.v10.i3.pp536-544.

[30] J. J. G. Torres *et al.*, "Automated question-answer generation for evaluating rag-based chatbots," in *Proceedings of the First Workshop on Patient-Oriented Language Processing (CL4Health) @ LREC-COLING 2024*, ELRA and ICCL, 2024, pp. 204–214. [Online]. Available: https://aclanthology.org/2024.cl4health-1.25.pdf

[31] J. Lu *et al.*, "Evaluation of large language models (LLMs) on the mastery of knowledge and skills in the heating, ventilation and air conditioning (HVAC) industry," *Energy and Built Environment*, Mar. 2024, doi: 10.1016/j.enbenv.2024.03.010.

[32] Z. Rasool *et al.*, "Evaluating LLMs on document-based QA: exact answer selection and numerical extraction using cogtale dataset," *Natural Language Processing Journal*, vol. 8, p. 100083, Sep. 2024, doi: 10.1016/j.nlp.2024.100083.

## BIOGRAPHIES OF AUTHORS

**Andi Riansyah** is a computer scientist with a bachelor of computer science degree from Universitas Islam Sultan Agung and a master of information systems degree from Universitas Diponegoro. He possesses over 7 years of research experience and has acquired a profound understanding of computer science, making noteworthy contributions to scientific advancement. He can be contacted at email: andi@unissula.ac.id.

**Imam Much Ibnu Subroto** received the Ph.D. degree from Universiti Teknologi Malaysia in 2015. He is currently a Senior Lecturer with the Department of Informatics Engineering, Universitas Islam Sultan Agung. He is the founder of SINTA and the founder of GARUDA. His research interests include computer science, artificial intelligence, machine learning, data mining, and education technology. He can be contacted at email: imam@unissula.ac.id.

**Intan Nur'aini** primary focus lies in computer science, particularly in the field of data engineering and artificial intelligence, where he has completed numerous courses and research projects. Aside from his academic pursuits. She can be contacted at email: intannura@std.unissula.ac.id.

**Ratna Supradewi** is a psychology expert. She received a Bachelor of Psychology and a Master of Psychology from Gajah Mada University. She can be contacted at email: supradewi@unissula.ac.id.

**Suyanto Suyanto** is a nursing lecturer. He received a Bachelor of Nursing degree (S.Kep.), a Master of Nursing degree (M.Kep.), and a Medical-Surgical Nursing Specialist certification. (Sp.Kep.MB.). He can be contacted at email: suyanto@unissula.ac.id.