

A RAG-enhanced multi-stage conversational system for AI-based mental health intervention

Dian Zhang

Shenzhen Senior High School, Shenzhen, China

dianzhang71@gmail.com

Abstract. Recent advances in generative Artificial Intelligence (AI) have enabled the development of personalized and interactive user-centered systems. However, its application in the field of mental health intervention remains relatively underexplored. Therefore, this study proposes a generative AI-based system for personalized emotion support, which integrates psychological condition assessment, therapeutic audio recommendation, and intervention outcome evaluation within a prompt-guided and Retrieval-Augmented Generation (RAG)-augmented multi-stage conversation framework. With the assistance of Cursor, an AI programming assistant, the researcher completed the construction of the system. 30 participants were invited to experience and evaluate the system for five days. The evaluation results showed that the system had a relatively high accuracy in identifying users' psychological condition and effectiveness in emotion regulation. However, the system shows limitations in handling subtle emotions. Future work will focus on improving model sensitivity and enhancing system adaptability.

Keywords: generative AI, mental health intervention, multi-stage framework, conversational system

1. Introduction

AI-driven systems that integrate user input understanding with personalized content generation have emerged as an important research direction in the development of intelligent interactive applications, with increasing exploration of their potential in mental health support and other user-centered domains. Existing research has demonstrated that empathetic and context-aware generative AI has significant application potential in the field of mental health intervention. However, existing AI-based systems often focused on isolated functionalities, such as assessment or therapy, without integrating these components into a unified framework that provides users with a comprehensive intervention service. To address the gap, this study proposes a multi-stage conversational system that integrates dialogue generation, psychological condition understanding, and personalized therapeutic audio recommendation. The system incorporates a prompt-guided conversational mechanism and a Retrieval-Augment Generation (RAG) component to support structured, context-aware, and knowledge-enhanced interactions. Building on this design, the system establishes a closed loop consisting of psychological assessment, personalized therapeutic audio recommendation, sound therapy, and evaluation of therapeutic outcome, ultimately generating a comprehensive intervention report for users. This framework

provides a potential approach for supporting emotion regulation and adolescent mental well-being through personalized and continuous AI-based intervention experience.

2. Literature review

2.1. Theoretical and empirical support for sound therapy

Sound therapy is a widely used approach in mental health intervention. Substantial empirical evidence supports its effectiveness in emotion regulation. Existing studies have shown that music can activate brain regions, such as auditory cortex and reward pathways, to elicit emotional responses and influence mood states [1, 2]. These findings suggest that music can effectively influence emotions through biological mechanisms.

In addition, Lv et al. found the significant impact of subjective preference on music's emotional effect [2]. Based on this finding, individual preferences for therapeutic audio were considered in the system design to enhance the effectiveness of intervention.

2.2. Comparison of mental health assessment tools

Currently, psychological scales are widely used due to their convenience and scalability. However, the research identified several limitations of self-report tools, including limitations in their accuracy, the confounding effect of mood state, and problems related to factor selection and retention in factor analytic approaches to self-report questionnaires [3]. In addition, standardized screening tools such as Patient Health Questionnaire-9 (PHQ-9), which are often used to estimate depression prevalence, had been shown to overestimate depression prevalence by Levis et al. [4]. Furthermore, Zimmerman stated that the self-management questionnaire provides an accurate estimate of the existence of given conditions and only reflects the prevalence of symptoms involved in a certain disorder rather than the prevalence of this disorder [5], so the prevalence of the disease needs to be assessed through diagnostic interviews.

Although psychological counselors can provide more nuanced psychological assessment, their judgements are also subject to certain limitations. Studies showed that clinical practice may be influenced by cultural bias [6], and both therapists and non-therapists may have difficulty in identifying erroneous therapist responses [7]. These limitations indicate that human-based assessment is still not fully free from subjectivity and error.

2.3. Generative AI applications in mental health domains

Generative AI for Psychological Condition Assessment. The strong language understanding and empathetic capabilities of generative AI have prompted increasing research in its application in psychological condition assessment. Zhong and Wang developed an AI-powered depression prevention system that leverages large language models to analyze real-time dialogue cues for more accurate and dynamic mental state assessment [8], achieving an accuracy of 89% in detecting subtle linguistic markers of depression. In addition, Diep et al. [9] demonstrated that ChatGPT performs effectively in three text-based mental health classification tasks, further supporting its ability to assess psychological conditions through language analysis. These studies suggest that generative AI has significant potential in psychological condition assessment based on textual and conversational data.

Generative AI for Music-based Sound Therapy. Psychological assessment and therapy are closely interconnected, and advances in generative AI enable its application extending from assessment to intervention. Some studies have explored the use of generation AI in music-based sound therapy. Sun et al. [10] explored the potential of music AI in enhancing therapy efficiency and enriching therapy content from the

perspective of music therapists. In addition, Wei et al. [11] proposed Context-AI Tune (CAT), a system for generating personalized music based on environmental input and users' self-reported stress levels, demonstrating great stress reduction. But researchers pointed out that some users wished for familiar elements or user-defined favorites, which shows the importance of knowing users' preferences in advance to generate music.

However, existing research also identified the limitations of AI-generated music therapy, including bias, insufficient representation of diverse musical styles and emotional context, ignorance of culturally specific or niche music expression, and oversimplification of the relationship between music and emotions [12]. These findings suggest the understanding users' cultural background and personal preference is essential in improving the effectiveness of AI-based sound therapy.

2.4. RAG-based knowledge grounding in AI-driven psychological condition assessment module

RAG Mechanism for Response Generation. Despite the impressive performance of large language models in natural language processing tasks, the issue of hallucinations remains a key limitation. Bang et al. [13] pointed out that, like other large language models, ChatGPT may generate outputs that are factually incorrect or contain unverifiable information. This indicates that relying solely on the parametric knowledge encoded within large language models is insufficient for handling complex and domain-specific knowledge requirements. To address this limitation, RAG mechanism has been proposed. RAG combines pre-trained parametric memory with non-parametric memory to accomplish language generation tasks, improving the accuracy of the generated content [14]. By incorporating an information retrieval step from an external knowledge base, the language generated by RAG is more specific, diverse, and factually accurate.

Knowledge Base for RAG. To ensure that the system maintains a high level of professionalism during the assessment process and better simulates the tone and logic of a real psychological counselor, the researcher utilized RAG to provide the AI with professional psychological counseling resources. First is The Cognitive Behavioural Therapy Skills Workbook, which is based on the principles of cognitive behavioural therapy and introduced some tools that can help manage depressive or anxious symptoms and maintain long-term rehabilitation progress [15]. In addition, A Therapist's Guide to Brief Cognitive Behavioral Therapy elaborated on the core principles, indications, module structure and case construction of short-term CBT [16]. It clearly pointed out that clinical judgment and individualized treatment plans are more crucial. It emphasizes the significance of flexibility, adaptability and judgment in clinical interviews in treatment assessment. Finally, Ethical Principles of Psychologists and Code of Conduct emphasized that psychologists should adhere to the principles of integrity, impartiality, respect, responsibility and non-harm [17]. It clearly stated that psychological assessment should be based on scientific and professional judgment and should prevent systematic deviations caused by social expectations, cultural differences, etc. The above materials serve as reference materials for the AI system in the psychological assessment stage, which to a certain extent improve the accuracy of psychological assessment.

3. Methodology

3.1. Target users analysis

The researcher conducted an online survey using the Chinese version of the Depression-Anxiety-Stress Scale (DASS) among 73 adolescents aged 10 to 19. The results showed that 35.62% of participants reported varying levels of stress, 53.43% reported varying levels of anxiety, and 31.51% reported varying levels of depression, indicating that lots of adolescents experience mental health challenges [18].

In addition, a questionnaire aiming to investigate adolescents' perceptions of psychological assessment and intervention methods was conducted. The results indicated that most adolescents have the need for psychological support and self-regulation, and traditional psychological assessment and intervention methods remain the dominant approach despite certain limitations. Many participants reported difficulties with existing methods, and a majority expressed positive attitudes toward integrating AI into psychological assessment and personalized intervention.

Overall, the findings suggest that adolescent represent a key user group with both significant mental health needs and strong acceptance of AI-based psychological assessment and intervention.

3.2. System overview

System Design. Based on the results of the previous survey and the identified limitations of traditional psychological assessment and intervention methods, the researcher designed a system. This system is built on generative AI and integrates three core functions: psychological condition assessment, personalized therapeutic audio recommendation, and therapeutic outcomes evaluation. It aims to provide adolescents with a structured and comprehensive mental health intervention service.

System Architecture. The system architecture consists of four main components: data and knowledge resources, AI services, core system logic, and interactive interface, which together support the operation of the system.

The data and knowledge resources include a RAG knowledge base containing knowledge related to psychological assessment and a pre-generated therapeutic audio library stored in Google Drive. These resources provide foundation for psychological assessment and audio recommendations.

The AI service mainly relies on GPT-4o-mini API. It is responsible for understanding user input, generating dialogue content, recommending audio, and generating personalized reports. By integrating user input with knowledge resources, the system can generate context-aware and relevant responses.

The core logic is designed as a multi-stage interaction framework that coordinates different functional modules within the system. It manages the flow of information between user input, knowledge retrieval, and response generation, ensuring that the system operates in a structured way.

The interactive interface is implemented using Gradio, providing a chat-based interaction environment.

System Workflow. The system operates through a structured multi-stage workflow that guides users from initial input to post-intervention evaluation.

After entering the interface, users are prompted to describe their current emotional state and psychological condition. Based on the input, the AI engages in multi-turn dialogue with users providing supportive feedback and using guided questioning strategy to collect more relevant psychological information. The dialogue history is recorded to ensure coherency and continuity throughout the interaction.

Once sufficient information about users' psychological condition is obtained, the AI transitions the dialogue toward music preferences naturally and collects information about users' preferred music types, styles, and elements to support the personalized recommendation.

Based on the users' psychological condition and music preference, the AI selects and play the most suitable therapeutic audio from the pre-generated audio library, initiating the sound therapy.

After the therapy, the AI collects user feedback regarding their psychological condition and perceived change, which is used to evaluate the suitability of the intervention and the therapeutic outcome.

Finally, the AI generates a personalized report summarizing the users' psychological condition before and after the intervention, the analysis of the recommended therapeutic audio, sound therapy, and general mental health suggestions.

3.3. Dialogue mechanism

Conversational Strategy. The system uses pre-defined prompts to guide the AI's language style, questioning technique, and dialogue objectives.

The AI is assigned a psychology counselor-like role and is guided to generate responses with a supportive, empathetic, and professional tone through prompt design. The language style encourages users to express their emotions, conveys understanding and support, and fosters user trust in the system, thereby promoting more open and effective communication.

In addition, the AI is guided to adopt the questioning techniques commonly used in psychological counseling, including a combination of open-ended and directive questions. This approach allows users to actively participate in the conversation while ensuring that the system can collect sufficient information for subsequent interaction.

Multi-stage Dialogue Design. The system adopts a multi-stage dialogue to ensure a structured interaction process and effective information collection. The dialogue is organized into four stages: psychological condition assessment stage, transition stage from assessment to music preference collection, music preference collection personalized therapeutic audio recommendation stage, and therapeutic report generation stage.

Each stage is defined with a specific objective through pre-defined prompts. The AI generates questions aligned with the objective of the current stage and dynamically adapts to users' responses, ensuring that the conversation remains focused to the intended topic. Clear boundaries are established between stages, enabling the system to concentrate on stage-specific objectives. To facilitate smooth transitions between stages and improve user experience, the system incorporates transition mechanism that guides user naturally transition from one topic to another with contextual language. In addition, each stage adopts a multi-turn interaction format and enforces a minimum number of dialogue turns, allowing the system to gradually deepen the conversation and ensure sufficient information collection.

Information Extraction and State Understanding. The system extracts information from ongoing dialogues to assess users' psychological condition and recommend personalized therapeutic audio. The understanding of information relies on the semantic processing abilities of the GPT-4o-mini large language model, which enables the assessment of psychological conditions.

Throughout the dialogue, the system retains the entire conversation history so that it can build deep and context-aware understanding of users. The contextual memory enables the system to continuously adapt responses based on the conversation history, thereby offering personalized feedback and recommendations.

3.4. Knowledge integration

RAG Framework Overview. RAG is a hybrid framework that combines information retrieval and generative models to enhance response accuracy. The framework retrieves relevant information from an external knowledge base and incorporates it as contextual input for the generative model, enabling the generation of more informed and context-aware responses.

Knowledge Base Construction. The external knowledge base used by the RAG framework consists of three key resources: *The Cognitive Behavioural Therapy Skills Workbook*, *A Therapist's Guide to Brief CBT*, and *Ethical Principles of Psychologists and Code of Conduct*. The content from these resources was divided into knowledge chunks, which were then structured and stored in a JSON file format for efficient retrieval. This organization allows the system to retrieve relevant and contextually appropriate information based on user input, ensuring that the AI's responses are grounded in established psychological practices and ethical guidelines.

Retrieval and Response Generation. When users input text, the system dynamically retrieves relevant knowledge chunks from the knowledge base and combines them with the inputs to form construct an augmented prompt fed into GPT-4o-mini. Based on the prompt, the AI generates personalized responses, ensuring that the answers meet the users' needs while maintaining professionalism.

The process is optimized through technical enhancements. The knowledge base is stored in JSON format, with each knowledge chunk being an independent text segment, allowing for efficient retrieval. Through API calls to GPT-4o-mini, the system integrates the retrieved knowledge chunks into prompt, improving the quality of the generated responses.

Role of RAG in System Performance. In this system, user inputs are combined with knowledge retrieved via RAG from a psychological practice knowledge base, resulting in a contextually augmented prompt. This approach allows the AI to generate responses that are not only personalized to the user needs but also supportive, empathetic, and aligned with professional psychological guidelines. In addition, RAG enables the system to dynamically retrieve and incorporate relevant information based on the users' inputs, ensuring that the AI's responses remain contextually appropriate throughout the interaction. By incorporating RAG, the system leverages a reliable knowledge base, enhancing the professionalism and supportiveness of the AI-generated content.

3.5. Recommendation module

Design Rationale of Audio Recommendation. Sound therapy is a commonly used and effective method in psychological intervention, with numerous studies demonstrating its ability to alleviate negative emotions and promote mental health. Moreover, individual preferences significantly influence the effectiveness of sound therapy. Therefore, the system adopts personalized sound therapy as the core therapeutic approach.

After assessing the users' psychological conditions and understanding their preferences, the system intelligently matches and recommends the most appropriate therapeutic audio from a pre-generated audio library. In the initial version of the system, the AI would generate audios in real-time for users. However, because of the limitation of current audio generation AI technologies, such as the time required for generation and the variability in audio quality, the system ultimately opted to pre-generate a large collection of therapeutic audios. This solution ensures a smooth intervention process, avoids potential negative effects, and allows the AI to personalize recommendation based on users' psychological conditions and preferences.

Audio Resource and Library. All therapeutic audios are generated by Mubert AI based on prompts created by GPT-4o-mini. These prompts detail the emotional tone, style and elements required for each audio to guide the audio generation process.

The audio library is designed to meet users' personalized needs, so it contains audios suited to various emotions, styles, and elements. Each audio file is named with relevant keywords, allowing the AI to efficiently match the most suitable one to users' psychological conditions and music preferences.

Personalized Recommendation Strategy. The system scans the audio library folder to generate a list of all available audio files, returning their full file paths. Once the audio list is obtained, the system combines the users' dialogue content with the audio file list to create a context, which is then used to generate a prompt for GPT-4o-mini. Based on the prompt, the AI analyzes the users' needs and selects the most suitable audio file from the list. Then, the system retrieves the corresponding file path of the selected audio and play the file. In cases where the most suitable recommendation cannot be identified, the system will select and play an audio that best aligns with the users' requirements, ensuring that an appropriate therapeutic audio is still provided.

3.6. Technology stack and tools

The system is developed using Python 3 and OpenAI GPT-4o-mini as the core large language model, with the integration of a RAG mechanism to enhance contextual relevance and professionalism of the AI responses. The front-end interface is built using Gradio, enabling interactive multi-stage dialogues and audio playback functionality. The system is developed in Google Colab, which facilitates code development and debugging. In addition, Google Drive is used as the cloud-based platform for pre-generated therapeutic audio library management.

3.7. Evaluation setup

Internal Functional Testing. To ensure the proper operation of the system, multiple internal tests were conducted on all core functions within the Google Colab environment.

External User Evaluation. To evaluate the accuracy of the generative AI-based psychological assessment and the effectiveness of personalized sound therapy, as well as to collect user feedback, the researcher recruited thirty adolescents with varying degrees of mental health problems to evaluate the system over a five-day period. During this period, participants experienced the AI system every other day, resulting in a total of three sessions within the evaluation cycle. Before and after the evaluation period, participants were asked to complete the Chinese version of the Depression-Anxiety-Stress Scale (DASS) respectively to assess their psychological condition and facilitate the evaluation.

The accuracy of the AI-based psychological assessment was evaluated by comparing the system-generated results with the DASS results. The level of consistency between two results was categorized into four levels: full match, where all emotions are accurately captured; core match, where main emotions are identified but minor ones are omitted; partial match, where only some emotions are reflected; and not match, where generated keywords are largely inconsistent with the DASS results.

The effectiveness of the sound therapy was assessed by analyzing changes in participants' DASS scores before and after the therapy, as measured by differences in scale scores completed by participants across the five-day evaluation period.

User feedback was collected through an online questionnaire distributed after the evaluation period. The questionnaire covered multiple aspects, including the smoothness of system usage, perceived warmth and empathy of the AI dialogue, appropriateness and effectiveness of the recommended therapeutic audio, usefulness of the therapy report, users' willingness to continue using the system as a mental health support tool, and suggestions for system improvement.

4. Results

4.1. Internal functional test results

The researcher conducted functional tests on the six core modules of the system at Google Colab to ensure that all modules could operate properly. The results showed that the AI could accurately extract key words related to psychological conditions in dialogues and make psychological assessment. The music preference collection module was able to effectively identify and analyze users' music preference. The personalized therapeutic audio recommendation module generally provided audio that aligned with users' psychological conditions and preferences. However, in cases where no fully matching audio was available in the library, the recommendations could only meet some of users' needs. In such cases, the AI will summarize the reasons in the report. The feedback collection module provided questions that are sufficient to understand users'

psychological conditions and feedback after the therapy. The personalized therapy report generation module produced structured report consisting of four predefined sections based on the entire dialogue data. Finally, the front-end interaction components, including the chatbox, textbox, buttons, audio, etc., can operate stably without functional issues.

4.2. External user evaluation results

Accuracy Evaluation of Generative AI-Based Psychological Assessment. The evaluation results showed that, among 30 participants, the proportion of full match was 16.67%, the proportion of core match was 43.33%, the proportion of partial match was 36.67%, and the proportion of not match was 3.33%. Overall, in 96.67% of cases, the psychological assessment section can accurately identify all or part of the negative emotions reflected in the DASS results. The situation of incorrect identification only occurred in one experience. Further analysis showed that the system had a high recognition rate for stress and anxiety, but a low recognition rate for depression. This may explain why most assessment results were core match or partial match, rather than full match, as depression-related signals were less consistently captured. See Appendix A for detailed scoring.

Effectiveness Evaluation of Generative AI-Based Sound Therapy. The DASS scores of 30 participants were analyzed before and after the five-day evaluation period. Overall, reductions were observed in stress, anxiety, and depression scores (see Figure 1).

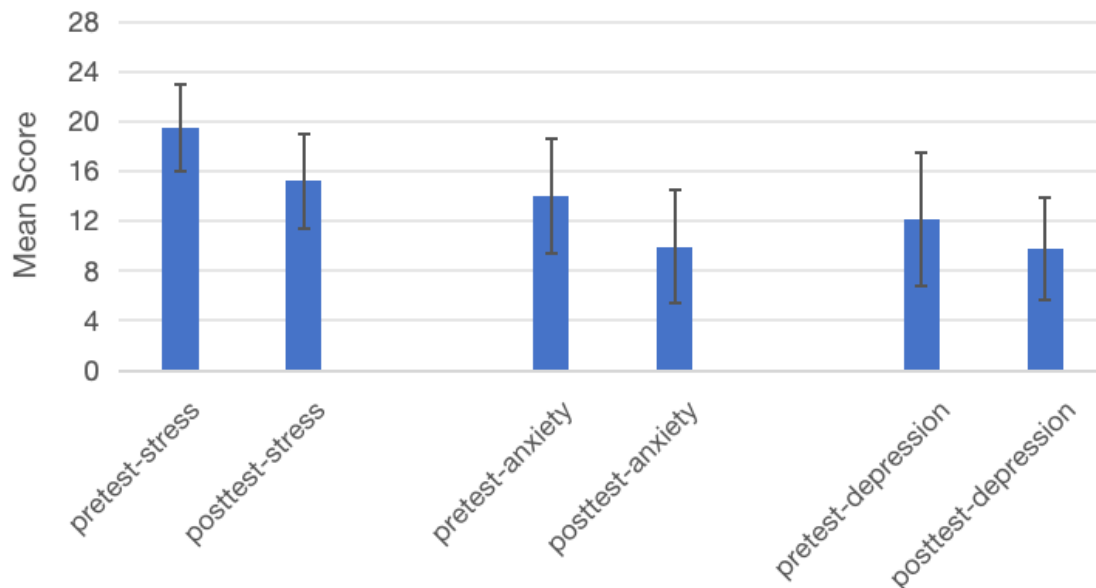


Figure 1. Comparison of pretest and posttest mean scores for stress, anxiety, and depression

For stress, the total score decreased from 584 to 456, with an average decrease of 4.27 points. 3 participants showed a two-level decrease in stress level, 19 participants showed a one-level decrease, and 8 remained unchanged. A paired-samples t -test showed a significant reduction in stress scores ($p < 0.001$).

For anxiety, the total score decreased from 420 to 298, with an average decrease of 4.07 points. Among them, 7 participants showed a two-level decrease in anxiety level, 14 showed a one-level decrease, and 9 remained unchanged. The t -test revealed a reduction in anxiety scores ($p < 0.001$).

For depression, the total score decreased from 364 to 294, with an average decrease of 2.33 points. Among them, 1 participant showed a two-level decrease in depression level, 8 showed a one-level decrease, and 21

remained unchanged. A statistically difference was observed in the t -test ($p < 0.001$), though the magnitude of improvement was relatively smaller compared to stress and anxiety.

Overall, the results indicate that the generative AI-based sound therapy can effectively support adolescents in regulating negative emotions and maintaining mental health. However, future improvements are needed to enhance the effectiveness in alleviating subtle emotions like depression. See Appendix B for detailed scoring.

User Feedback Survey. The survey results indicated that the system was rated positively across multiple dimensions, including interface design, user dialogue, psychological assessment, sound therapy, therapy reports, smoothness of use, and overall willingness to use. Though the ratings for the emotional improvement effects of sound therapy and the accuracy of psychological assessment were slightly lower compared to other dimensions, these areas remain key focuses for further refinement in future iterations.

In addition, the questionnaire collected participants' suggestions for system improvement. The main suggestions include adding a voice input function, making the AI dialogue more human-like, reducing the number of questions asked by the system, and including basic information such as date information in the therapy report. These suggestions provide valuable insights for future system optimization. See Appendix C for detailed results.

5. Discussion

While recent advances in generative AI have enabled the development of interactive user-centered systems, existing applications in mental health contexts remain limited in providing integrated and structured intervention frameworks. To bridge the gap, the system integrates psychological condition assessment, personalized therapeutic audio recommendation, and intervention outcome evaluation within a prompt-guided, multi-stage framework enhanced by the RAG mechanism. This design enables a continuous and structured pipeline that aligns assessment, intervention, and feedback within a unified system.

5.1. System behavior and evaluation

The evaluation results suggest that the system has potential in supporting adolescent mental health.

For the accuracy of the psychological assessment module, the high recognition accuracy of stress and anxiety indicates that these emotions are more identifiable through generative AI-based dialogue. The lower recognition accuracy of depression may attribute to the more subtle and complex nature of depression which causes challenges for AI to detect accurately.

For the effectiveness of the sound therapy, the observed decrease in stress, anxiety, and depression scores indicate that the system is effective in alleviating negative emotions. However, the improvements in stress and anxiety are more pronounced than those in depression. This difference may be attributed to system's relatively lower accuracy in identifying depression. Deviations in psychological assessment may lead to less precise matching of therapeutic audio to users' needs, thereby reducing the effectiveness of sound therapy in addressing depression.

In addition, the user feedback survey demonstrated a high level of satisfaction, indicating that the system is well-accepted by adolescents, and some participants provided valuable suggestions for improvement of the system.

5.2. System strengths

The evaluation results suggest that the system can effectively capture users' psychological conditions through the dialogue, as reflected by the relatively high consistency between AI-based assessment and DASS results.

This supports the feasibility of adopting large language models for psychological condition assessment, particularly in identifying explicit emotions such as stress and anxiety. The observed improvements in participants' psychological condition further indicate that AI-based personalized sound therapy can serve as an effective intervention method. This suggests that incorporating users' personal preference can enhance the effectiveness of sound therapy. In addition, user feedback reflects a high level of satisfaction, with participants providing positive feedback across multiple dimensions and a strong willingness to continue using the system. This implies that AI-based approaches may provide an accessible and engaging form of mental health support for adolescents. Furthermore, the integration of multiple functional components within a multi-stage conversational framework contributes to its overall effectiveness. By structuring the dialogue and incorporating a RAG mechanism, the system can maintain coherence and professionalism while adapting to user input, supporting a more continuous and personalized intervention process.

5.3. System limitations

While the system demonstrates relative high accuracy in psychological assessment and effectiveness in sound therapy, there are still several limitations. One limitation lies in the use of a pre-generated therapeutic audio library rather than real-time audio generation due to technical limitations. Although this approach ensures a smooth user experience, it may reduce the level of personalization to dynamically generated audio, potentially limiting the effectiveness of sound therapy. Another limitation is the system's limited capability in identifying and addressing subtle emotions like depression. This is largely due to the system's reliance on text-based input, as the absence of multimodal input limits its ability to capture nuanced emotional cues. In addition, some design parameters, including dialogue length and the structure of report generation, still require further optimization to improve user experience.

5.4. Future directions

To address the current limitations, future work will focus on optimization of the system. One important direction is the integration of real-time AI-based audio generation models, which could enhance the adaptability and personalization of therapeutic audio. Another direction involves incorporating multimodal emotion input, such as voice and facial cues, to improve the system's ability to capture more subtle emotions. Furthermore, aspects of system design, including dialogue length and the report generation structure, will be further optimized to improve overall user experience and interaction quality. Future research will also involve large-scale evaluations to assess improvement in assessment accuracy and therapeutic effectiveness and to collect more comprehensive user feedback. With continued development, the system has the potential to be applied in real-world settings and contribute to adolescent mental health support.

6. Conclusion

This study aims to explore how generative AI can be designed to support structured, context-aware, and knowledge-grounded interactions in mental health intervention. To achieve this, a generative AI-based mental health intervention system was developed, integrating psychological condition assessment, personalized therapeutic audio recommendation, and therapeutic effect evaluation within a prompt-guided and retrieval-augmented multi-stage conversational framework. After the system was set up, internal functional test was conducted, and 30 adolescents were invited to participate in a five-day evaluation of the system. The results indicate that the system can generate context-aware responses, effectively identify users' psychological conditions, and support the regulation of negative emotions. This suggests that combining multi-stage dialogue

design with a RAG mechanism can improve the practical applicability of generative AI systems in user-centered scenarios. However, several limitations remain, including the inability to generate real-time therapeutic audio, limited sensitivity in recognizing subtle emotions, and the need for further refinement in some system design details. Further work will focus on integrating more advanced AI-based audio generation model and exploring multimodal emotion input methods. Overall, this study contributes a structured and knowledge-enhanced design framework for generative AI systems, highlighting its potential as an accessible and scalable mental health support tool for adolescents.

References

- [1] Koelsch, S. (2010). Towards a neural basis of music-evoked emotions. *Trends in Cognitive Sciences*, 14(3), 131–137.
- [2] Lv, X., Wang, Y., Zhang, Y., Ma, S., Liu, J., Ye, K., Wu, Y., Voon, V., & Sun, B. (2024). Auditory entrainment coordinates cortical-BNST-NAc triple time locking to alleviate the depressive disorder. *Cell Reports*, 43(8), Article 114474. <https://doi.org/10.1016/j.celrep.2024.114474>
- [3] Huprich, S. K., Bornstein, R. F., & Schmitt, T. A. (2011). Self-report methodology is insufficient for improving the assessment and classification of Axis II personality disorders. *Journal of Personality Disorders*, 25(5), 557–570.
- [4] Levis, B., Benedetti, A., Ioannidis, J. P. A., Sun, Y., Negeri, Z., He, C., Wu, Y., Krishnan, A., Bhandari, P. M., Neupane, D., Imran, M., Rice, D. B., Riehm, K. E., Saadat, N., Azar, M., Boruff, J., Cuijpers, P., Gilbody, S., Kloda, L. A., ... Thombs, B. D. (2020). Patient Health Questionnaire-9 scores do not accurately estimate depression prevalence: Individual participant data meta-analysis. *Journal of Clinical Epidemiology*, 122, 115–128.e1. <https://doi.org/10.1016/j.jclinepi.2020.02.002>
- [5] Zimmerman, M. (2024). The value and limitations of self-administered questionnaires in clinical practice and epidemiological studies. *World Psychiatry*, 23(2), 210.
- [6] Iudici, A., Colombo, L., Caravita, S. C. S., Cottone, P., & Neri, J. (2024). The invisible discrimination: Biases in the clinical approach regarding migrants: A study to help ethnopsychology services and clinicians. *Behavioral Sciences*, 14(3), Article 155.
- [7] Cooper, M. B., Di Bartolomeo, A. A., Westra, H. A., Olson, D. A., & Boritz, T. (2023). A test of psychotherapists' and nontherapists' recognition of correct and incorrect process in managing client resistance. *Counselling and Psychotherapy Research*, 23(4), 1097–1107.
- [8] Zhong, Z., & Wang, Z. (2025). *Intelligent depression prevention via LLM-based dialogue analysis: Overcoming the limitations of scale-dependent diagnosis through precise emotional pattern recognition*. arXiv. <https://doi.org/10.48550/arXiv.2504.16504>
- [9] Diep, B., Stanojevic, M., & Novikova, J. (2022). *Multi-modal deep learning system for depression and anxiety detection*. arXiv. <https://doi.org/10.48550/arXiv.2212.14490>
- [10] Sun, J., Yang, J., Zhou, G., Jin, Y., & Gong, J. (2024). Understanding human-AI collaboration in music therapy through co-design with therapists. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems* (pp. 1–21). Association for Computing Machinery.
- [11] Wei, X., Zhang, Z., Yue, Z., & Chen, H. T. (2025). Context-AI tunes: Context-aware AI-generated music for stress reduction. In *International Conference on Human-Computer Interaction* (pp. 330–345). Springer Nature Switzerland.
- [12] Chen, H. (2024). The possible effects of music therapy based on computer generation on depression. *Lecture Notes in Education Psychology and Public Media*, 54, 190–196.
- [13] Bang, Y., Cahyawijaya, S., Lee, N., Dai, W., Su, D., Wilie, B., Lovenia, H., Ji, Z., Yu, T., Chung, W., Do, Q. V., Xu, Y., & Fung, P. (2023). A multitask, multilingual, multimodal evaluation of ChatGPT on reasoning,

- hallucination, and interactivity. In *Proceedings of the 13th International Joint Conference on Natural Language Processing and the 3rd Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics (Volume 1: Long Papers)* (pp. 675–718). Association for Computational Linguistics. <https://doi.org/10.18653/v1/2023.ijcnlp-main.45>
- [14] Lewis, P., Perez, E., Piktus, A., Petroni, F., Karpukhin, V., Goyal, N., Küttler, H., Lewis, M., Yih, W.-t., Rocktäschel, T., Riedel, S., & Kiela, D. (2020). Retrieval-augmented generation for knowledge-intensive NLP tasks. In *Advances in Neural Information Processing Systems 34 (NeurIPS 2020)* (Vol. 33, Article 793, pp. 9459–9474). Curran Associates.
- [15] Hertfordshire Partnership University NHS Foundation Trust. (2016). *Cognitive behavioural therapy skills workbook*. <https://www.hpft.nhs.uk/media/1655/wellbeing-team-cbt-workshop-booklet-2016.pdf>
- [16] Cully, J. A., & Teten, A. L. (2008). *A therapist's guide to brief cognitive behavioral therapy*. Department of Veterans Affairs South Central MIRECC.
- [17] American Psychological Association. (2017). *Ethical principles of psychologists and code of conduct*. <http://www.apa.org/ethics/code/index.aspx>
- [18] Zhang, D. (2025). Innovative design of auxiliary meditation healing devices based on the role of sensory stimulation in emotion regulation. In *Proceedings of the CTB Global Youth Research and Innovation Forum*. CTB Library.

Appendix A

Table A. The results of the accuracy assessment of psychological assessment

Test-taker	DASS-stress score	DASS-anxiety score	DASS-depression score	Psychological condition	Emotion keywords in AI report	Match?
01	20	14	12	moderate stress, moderate anxiety, mild depression	stress, anxiety, dispiritedness	core match
02	20	10	10	moderate stress, moderate anxiety, mild depression	stress, anxiety, helplessness	full match
03	26	12	8	severe stress, moderate anxiety	stress, exhaustion	partial match
04	22	8	10	moderate stress, mild anxiety, mild depression	stress, anxiety, exhaustion	core match
05	16	24	16	mild stress, severe anxiety, moderate depression	high anxiety, loneliness	partial match
06	22	18	18	moderate stress, severe anxiety, moderate depression	stress, anxiety, dispiritedness	core match
07	20	12	8	moderate stress, moderate anxiety	stress, anxiety	full match
08	16	10	8	mild stress, moderate anxiety	emotional isolation	no match
09	24	16	12	moderate stress, severe anxiety, mild depression	stress, anxiety, helplessness	core match
10	20	8	8	moderate stress, mild anxiety	stress, anxiety	full match
11	18	10	10	mild stress, mild anxiety, mild depression	anxiety, nervousness	core match
12	22	18	14	moderate stress, severe anxiety, moderate depression	anxiety, stress, worry	partial match
13	16	14	12	mild stress, moderate anxiety, mild depression	anxiety	core match
14	14	12	18	moderate anxiety, moderate depression	anxiety, stress, nervousness	partial match
15	18	16	14	mild stress, severe anxiety, moderate depression	anxiety, exhaustion, disappointment	core match
16	16	12	6	mild stress, moderate anxiety, mild depression	stress, anxiety	core match
17	22	12	20	moderate stress, moderate anxiety, moderate depression	stress, anxiety, nervousness	partial match
18	16	10	8	mild stress, moderate anxiety	anxiety	core match
19	22	10	10	moderate stress, moderate anxiety, mild depression	stress, anxiety, dispiritedness	core match
20	24	10	8	moderate stress, moderate anxiety	stress, nervousness, loneliness	partial match

Table A. Continued

21	14	14	18	moderate anxiety, moderate depression	anxiety, self doubt	partial match
22	22	24	18	moderate stress, severe anxiety, moderate depression	anxiety, exhaustion, loneliness	partial match
23	22	14	8	moderate stress, moderate anxiety	stress, anxiety, dispiritedness	full match
24	22	12	20	moderate stress, moderate anxiety, moderate depression	anxiety, confusion	partial match
25	20	18	12	moderate stress, severe anxiety, mild depression	stress, anxiety	core match
26	22	26	28	moderate stress, severe anxiety, severe depression	stress	partial match
27	10	10	12	moderate anxiety, mild depression	anxiety, nervousness	core match
28	18	14	8	mild stress, moderate anxiety	anxiety, nervousness	core match
29	20	14	4	moderate stress, moderate anxiety	anxiety, sense of loss	partial match
30	20	16	6	moderate stress, severe anxiety	anxiety, stress	full match

Appendix B

Table B. The results of the evaluation on the effectiveness of sound therapy based on AI-generated personalized therapeutic audio

test-taker	Before music-based intervention			After music-based intervention		
	stress	anxiety	depression	stress	anxiety	depression
01	20	14	12	16	8	10
02	20	10	10	18	6	8
03	26	12	8	20	8	8
04	22	8	10	16	6	6
05	16	24	16	12	22	12
06	22	18	18	14	12	16
07	20	12	8	16	10	8
08	16	10	8	16	8	8
09	24	16	12	18	10	10
10	20	8	8	18	8	8
11	18	10	10	12	6	8
12	22	18	14	18	12	10
13	16	14	12	12	8	10
14	14	12	18	6	6	12
15	18	16	14	14	12	8

Table B. Continued

16	16	12	6	10	4	4
17	22	12	20	14	8	14
18	16	10	8	14	6	6
19	22	10	10	18	6	8
20	24	10	8	22	10	8
21	14	14	18	14	12	16
22	22	24	18	16	20	14
23	22	14	8	20	12	8
24	22	14	20	18	6	14
25	20	18	12	16	10	8
26	22	26	28	18	22	24
27	10	10	12	4	8	10
28	18	14	8	16	10	8
29	20	14	4	14	8	4
30	20	16	6	16	14	6

Appendix C

Table C. The result of user feedback survey

Survey question	Mean score	Standard deviation	Percent of high score (4-5)
The system interface design was clear, intuitive, and user-friendly.	4.57	0.568	96.67%
I felt understood and respected during the dialogue process.	4.3	0.758	83.34%
The emotional assessment was accurate.	4.07	0.691	86.66%
The recommended music improved my emotional state.	4	0.788	76.67%
The report was useful and accurate.	4.3	0.651	90%
The system ran smoothly.	4.83	0.379	100%
I'm willing to continue using this system in the future.	4.3	0.837	83.33%