EAI Endorsed Transactions
on Pervasive Health and Technology
Research Article EAI.EU

RAG-TPC: Retrieval Augmented Generation for Teenager Psychological Counseling Using DeepSeek

Yanling Li[1,2,3], Yibo Gao[1,2,4], Fangying Quan[1,2,4,*], Xudong Luo[1,2,3,*]

[1]Guangxi Key Lab of Education Blockchain and Intelligent Technology of Ministry of Education, Guangxi Normal University
[2]School of Computer Science and Engineering, Guangxi Normal University
[3]Guangxi Key Lab of Multi-source Information Mining and Security, Guangxi Normal Universit
[4]Department of Psychology, Faculty of Education,Guangxi Normal University

Abstract

The rising prevalence of adolescent mental health issues underscores the limitations of traditional counselling services in terms of scalability, timeliness, and accessibility. This paper presents RAG-TPC, a Retrieval-Augmented Generation framework built upon the DeepSeek language model for teenage psychological counselling. The system incorporates intent classification, semantic retrieval, and structured prompt-based generation to produce safe, empathetic, and contextually appropriate responses. We construct a domain-specific dataset spanning general distress, mental illness, and SOS emergencies, and employ LoRA-based fine-tuning to enhance intent recognition. Experimental results show that RAG-TPC consistently outperforms competitive LLMs in both classification and response quality. Evaluations by psychological professionals further validate the system's practical effectiveness and ethical reliability, highlighting its potential for scalable AI-assisted mental health support.

Received on 17 May 2025; accepted on 8 December2025; published on 27 January 2025

Keywords: Large Language Models, Retrieval Augmented Generation, DeepSeek, Fine-tuning, Psychological Counseling.

doi:10.4108/eetpht.11.11669

1. Introduction

In recent years, adolescent mental health problems have become a significant concern in the field of global public health. A study published in the Lancet indicates that mental health issues have become one of the core factors affecting the physical and mental development of adolescents, with the prevalence of mental disorders such as anxiety, depression, emotional and behavioural disorders, attention deficit disorder, and suicidal tendencies reaching 8.9%, and exhibiting a continuous upward trend [11]. Furthermore, a meta-analysis of the mental health status of Chinese middle and high school students from 2010 to 2020 shows a deterioration in mental health among this group, with exceptionally high detection rates of depression, anxiety, sleep disorders, and self-harming behaviours.

Despite the growing severity of adolescent mental health issues, the imbalance between the supply and demand for mental health services has exacerbated this dilemma. Existing mental health counselling methods primarily rely on face-to-face consultations and telephone counselling. While these approaches enable mental health professionals to assess the mental state of help-seekers in real time, establish strong therapeutic relationships, and develop personalized intervention plans, they are limited by the number of professionals available, the rigidity of service models, and unequal resource distribution. As a result, many adolescents are unable to access timely and high-quality mental health support [18]. Additionally, the Mental illness stigma [39] and the discrimination experienced by help-seekers [38] discourage many adolescents from seeking psychological assistance, thereby exacerbating mental health issues. Moreover, adolescent mental health crises are often sudden and covert, while traditional counselling models typically

*Corresponding author. Email: quanfangying@126.com (Fangying Quan); luoxd@mailbox.gxnu.edu.cn (Xudong Luo)

require advanced appointments. This can prevent individuals in crisis from receiving timely and effective intervention, increasing potential mental health risks.

With the proliferation of internet technology, adolescents are increasingly inclined to express their emotions through social media [40]. Studies have shown that highly anxious adolescents tend to receive more emotional support through self-disclosure in online environments [41]. However, traditional counselling models heavily rely on identity verification and face-to-face communication, while social interaction anxiety and stigma can significantly reduce individuals' willingness to seek help [42]. These challenges indicate that current mental health service models fail to meet the needs of adolescents, highlighting the urgent need for more flexible, efficient, and accessible mental health interventions.

Against this backdrop, Artificial Intelligence (AI) technology, particularly the recent breakthroughs in Large Language Models (LLMs), offers an innovative solution for adolescent mental health interventions [30, 36, 37]. LLMs possess powerful Natural Language Processing (NLP) capabilities, enabling the development of intelligent psychological counseling systems. These systems not only alleviate the shortage of mental health resources but also provide 247, location-independent psychological support. However, despite the significant potential of LLMs in the field of mental health, their application still faces numerous challenges. First, adolescents have unique patterns of language expression and cognitive styles, which require LLMs to possess more precise dialogue intention understanding to ensure the effectiveness of psychological counseling. Second, psychological counseling involves highly sensitive personal information, making data security and ethical compliance critical issues. Furthermore, psychological counseling is not merely about information transmission; it requires deep emotional support and personalized feedback. Conventional AI dialogue systems primarily focus on text generation, whereas effective psychological intervention requires complex contextual awareness and emotional resonance [32]. A significant limitation of standalone LLMs is their propensity to generate plausible but incorrect or generic information, a phenomenon known as "hallucination," which is particularly risky in the mental health domain. Retrieval-Augmented Generation (RAG) has emerged as a powerful framework to mitigate this issue by grounding the LLM's responses in relevant, external knowledge sources. This approach is especially critical in healthcare domains, as demonstrated by Shah et al., who developed a RAG system specifically for mental health to augment information retrieval and ensure data accuracy and reliability by leveraging reputable scientific evidence [3]. Therefore, optimizing AI models to ensure both safety and professionalism while providing empathetic responses in adolescent mental health scenarios is one of the key challenges in current research.

To address the limitations of traditional counselling models in meeting the mental health needs of adolescents, this study proposes RAG-TPC, a RAG model based on the DeepSeek LLM.The RAG framework enhances conventional LLMs by integrating a retrieval mechanism that fetches pertinent information from a curated knowledge base before generating a response [10]. This approach ensures that the system's outputs are not only coherent but also contextually grounded in verified psychological knowledge, improving reliability and reducing harmful hallucinations [9]. As illustrated in Fig.1, the framework consists of three main components: intent classification, semantic retrieval, and structured response generation. First, a domain-specific dataset is constructed and used to fine-tune DeepSeek for accurate intent recognition across counselling scenarios (e.g., general distress, mental illness, SOS). The predicted intent guides the retrieval of relevant knowledge segments, which are then integrated into a structured prompt. This prompt, combined with user input and task instructions, is passed to DeepSeek-Chat to generate safe, context-aware, and empathetic responses. The main contributions of this study are as follows:

- *Domain-specific dataset construction*: We curate and publicly release a high-quality dialogue dataset tailored to adolescent psychological counselling, providing a valuable benchmark for future research.
- *Parameter-efficient fine-tuning for intent classification*: Using LoRA, we fine-tune the DeepSeek LLM to accurately identify counselling intent categories, ensuring robust recognition of distress, clinical symptoms, and emergencies.
- *Retrieval-augmented and intent-aware response generation*: We design a response generation module that leverages intent-guided semantic retrieval and structured prompting to produce contextually adaptive and psychologically appropriate replies.

The rest of this paper is organized as follows. Section 2 outlines our methodology, which encompasses dataset preparation, LoRA-based fine-tuning, and structured prompt-driven generation within a retrieval-augmented framework. Section 3 presents the experimental design and evaluation results. Section 4 compare our work with related work. Finally, Section 5 summarizes the contributions of this study and discusses future research directions.

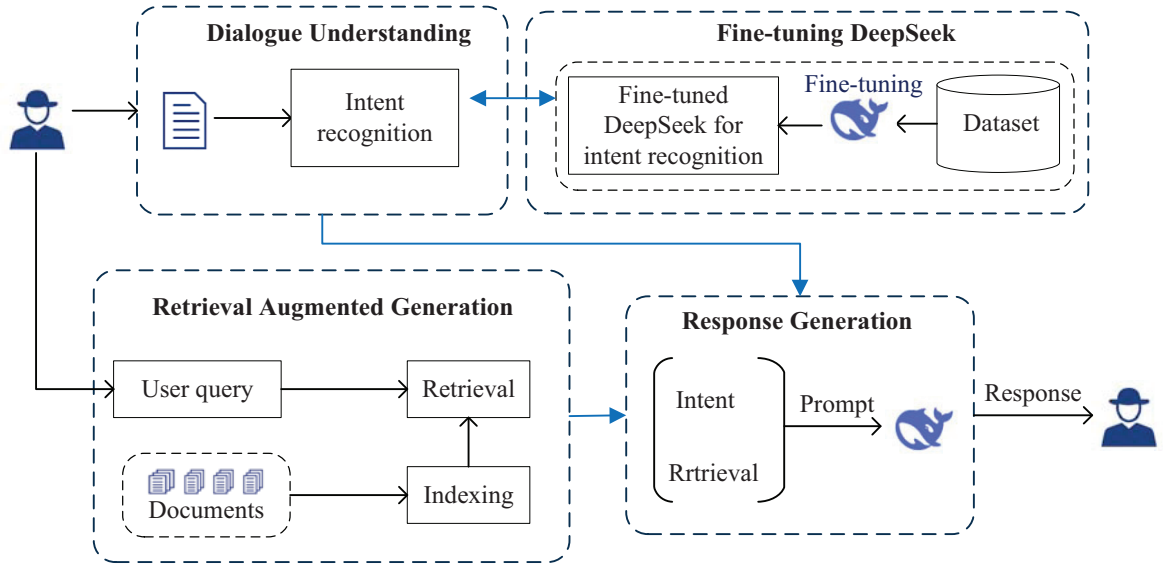


Figure 1. Structure diagram of RAG-TPC model.

2. Methodology

2.1. Dataset

We selected the corpus of adolescent psychological counselling from the corpus created by Wang et al. [43], and then made several optimizations on this basis, including corpus supplementation, corpus cleaning, corpus balancing, and correction of spelling errors. Finally, a dataset that meets the requirements of adolescent psychological counselling was formed, totalling 13,084 data.[1] This dataset covers three common categories in adolescent psychological counselling as follows:

- *Trouble type*: It mainly includes psychological distress adolescents encounter daily, such as interpersonal relationships, academic pressure, etc.
- *Mental illness:* It includes cases where psychological problems have had a significant impact on individual social functions and require professional psychological intervention or medical intervention.
- *SOS emergency help*: It identifies crises that require immediate human intervention, such as suicidal tendencies and severe emotional breakdown.

Tables 1 to 3 show each category's data distribution and representative samples in detail.

[1]The dataset can be downloaded at https://github.com/gaoyibo2000/paper-data.

2.2. Fine-tuning DeepSeek

To effectively address the unique linguistic patterns and psychological expressions present in adolescent counselling, we fine-tune the DeepSeek using a domain-specific dataset. While pre-trained LLMs such as DeepSeek-LLM possess strong general language capabilities, psychological counselling requires refined understanding of emotional cues and intent categories. Therefore, we adopt a fine-tuning approach specifically tailored for DeepSeek.

To optimise efficiency and reduce overfitting on the small-scale dataset, we adopt LoRA (Low-Rank Adaptation) [1] for parameter-efficient fine-tuning. Instead of updating the full parameter matrix $W \in \mathbb{R}^{m\times n}$ of DeepSeek, LoRA decomposes the update into two low-rank matrices $A \in \mathbb{R}^{m\times r}$ and $B \in \mathbb{R}^{r\times n}$, yielding $\Delta W = AB^{\top}$. The final update is scaled and regularised with dropout:

$$W' = W + \alpha \cdot \text{Dropout}(AB^{\top}) \tag{1}$$

where α denotes a scaling factor. Algorithm 1 outlines the procedure: loading the DeepSeek model and tokenizer, attaching the LoRA module, and training with mini-batch gradient descent over multiple epochs. The configuration sets rank = 4, α = 16, and dropout = 0.1, balancing efficiency and expressiveness.

This fine-tuned DeepSeek model serves as the backbone for intent classification in the RAG-TPC architecture. By aligning its representations with the psychological counselling domain, it improves recognition of distress types, mental illness indicators, and SOS crisis signals, thereby enabling more precise downstream response generation.

Table 1. Examples of trouble type.

Category	Count	Example
Academic worries and uncertainty about the future	1,010	Seeing others get into regular high schools while I ended up in a vocational school makes me feel inferior. My friends are drifting away, and my future seems bleak.
Insomnia	1,017	Recently, I have been suffering from insomnia every night. Even the slightest noise wakes me up during the day. I don't want to go anywhere and constantly overthink. My emotions are unstable, and I feel immense psychological pressure.
Stress	1,004	I feel overwhelmed with pressure and have lost my sense of direction in life. I no longer find joy in anything.
Interpersonal relationships	1,060	To fit in, I always lower myself, but thinking about it makes me feel really depressed.
Romantic relationship issues	1,555	Whenever I ask my girlfriend about our relationship, she always says she doesn't know. Does she really care about me?
Self-exploration	1,031	My family keeps telling me to be more open, make friends, and participate in group activities, but I'm really afraid of interacting with strangers. I always feel like I perform terribly in social situations.
Low self-esteem	995	I always feel inferior, like everyone else can do things well while I constantly fall short. What should I do?
Adolescent issues	1,004	After becoming a fan of a celebrity, I started seeing them as my emotional pillar. Whenever they receive bad news, I feel devastated. How can I become emotionally stronger?
Obsessive-compulsive disorder	924	No matter how busy I am, I spend a lot of time organizing my belongings every day. If things are out of place, I can't concentrate.
Sexual issues	1,009	When I see others wearing revealing clothes, I can't help but feel drawn to them. What should I do?
Parent-child relationships	1,033	Every time I argue with my younger siblings, I regret it later. Am I being too harsh on them?
Others	1,442	I feel an inexplicable sense of anxiety and tension, have trouble sleeping, and feel exhausted all the time. The doctor diagnosed me with anxiety and depression, but I don't know what to do.

2.3. Retrieval-Augmented Response Generation

To address the semantic complexity and emotional diversity commonly encountered in adolescent psychological counseling, we propose a Retrieval-Augmented Response Generation (RARG) module based on the DeepSeek-Chat large language model. This module combines intent-guided semantic retrieval with instruction-based response generation strategies.

Intent-Guided Semantic Retrieval. Given any input utterance x, we first employ a fine-tuned DeepSeek intent classification model to determine its psychological category, formalized as $f : X \to C$, where $C_X \in \{C_f, C_p, C_s\}$ corresponds to general psychological distress, clinical mental illness, and SOS emergency, respectively. The classification result C_X serves as a control signal for the subsequent retrieval and generation processes.

We then encode the input into a dense semantic vector and apply the Faiss similarity search framework [2] to retrieve the top-k most relevant textual segments $K = \{k_1, ..., k_n\}$ from a domain-specific psychological knowledge base. This knowledge base comprises:

- Expert-annotated psychological questions and answers pairs;
- Cognitive Behavioral Therapy strategies and emotion regulation techniques;
- Anonymized summaries of real-world adolescent counseling sessions.

This semantic retrieval process injects domain knowledge and case context into the DeepSeek-Chat model, enhancing interpretability, relevance, and response safety.

Structured Prompt Construction and Response Generation. Prompt design plays a pivotal role in large language model-based generation tasks, as the structure of the prompt significantly influences the accuracy and contextual appropriateness of the generated content [12, 13]. Several structured prompt frameworks have

Table 2. Examples of mental illness.

Category	Count	Example
Depression	1,093	I always feel that I can't do anything well and have no motivation for anything. Every day, I feel more and more exhausted, as if I am mentally and physically drained.
Anxiety	1,106	I get extremely nervous before every exam and feel so useless.
Bipolar disorder	171	During a club recruitment event, a senior mocked my outfit in public. When I returned to my dorm, I cut up all my clothes. Now, every time I see a mirror, I want to smash it.
Post-traumatic stress disorder (PTSD)	259	On a stormy night, when thunder roared, I curled up in my wardrobe. The memory of being locked in a storage room when I was eight years old flooded back.
Panic disorder	68	Sometimes, I suddenly feel overwhelming panic, as if everything around me is pressing down on me.
Anorexia and binge eating disorder	169	Even though I am not hungry, when I am feeling emotional, I binge eat a lot of junk food. As a result, I get stomach aches and can't sleep all night.
Below S2 threshold	9,888	Mild social anxiety. I don't know how to interact with others or how to fit in with people around me.
Other disorders	330	I experience auditory hallucinations and my memory is deteriorating. I don't know what's happening to me.

Table 3. Examples of SOS emergency help.

Category	Count	Example
Ongoing suicidal behavior	3	My parents found out I spent money on a game. Trembling, I ran out of the house. The headlights on the road made me want to throw myself forward.
Planned suicidal behavior	204	My mind won't calm down. I feel immense pressure. I don't know what I should do. I don't want to live anymore.
Self-harm	205	I discovered that my dad installed a two-way mirror in my room. I smashed the mirror with a hammer, and the glass shards cut my foot.
Committed physical harm to others	97	I can't control my sexual impulses, which has led to physical conflicts with classmates.
Planned physical harm to others	95	I don't know why, but I have a strong urge to kill someone. How can I clear this state of mind?
No physical harm tendencies	12,480	I feel extremely irritable and uninterested in everything. Nothing makes me happy. I suffer from insomnia and vomiting.

been proposed in previous studies, such as CRISPE [14] and BROKE [15], which emphasize modularity, task-specific adaptation, and controlled generation.

Inspired by these frameworks, we construct a structured prompt format tailored for psychological counseling. Specifically, we integrate the intent category C_X, retrieved knowledge K, original user input x, and task-specific instructions into a composite prompt P, which is then passed to the DeepSeek-Chat model for generation. The generation process is formalized as follows:

$$Y = g(P) = g(\text{Role} + C_X + \text{I} + \text{K} + \text{O})$$

The components of this prompt are defined as:

- *Role*: Specifies the role of the model in the task, such as "psychological counselor";
- C_X: The intent classification result, used to modulate tone and semantics of the response;
- I (Instruction): Provides task directives or behavioral constraints for the model;
- K (Retrieved Knowledge): Refers to relevant external knowledge retrieved from the database;
- O (Output Format): Specifies the expected style, structure, or format of the output.

Through this structured prompt design, the model is enabled to generate responses that are semantically coherent, emotionally empathetic, and therapeutically

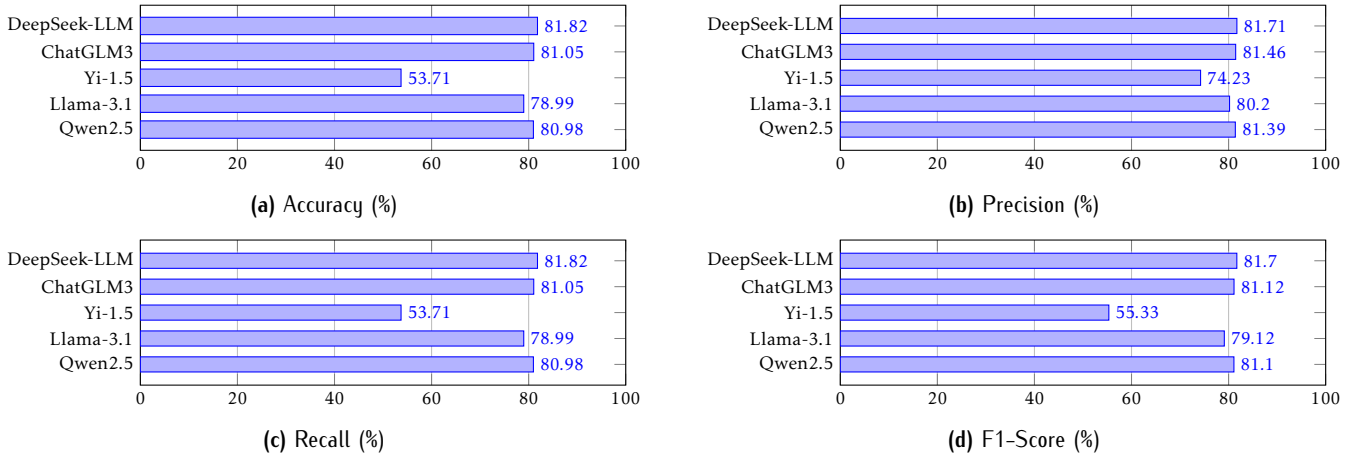


Figure 2. Comparison of fine-tuning LLMs for trouble type.

Algorithm 1 Fine-tuning DeepSeek with LoRA

Input: Pre-trained DeepSeek model, dataset D_{train}, learning rate α, number of epochs, batch size B
Output: Fine-tuned model, fine-tuned tokenizer
    model = LoadPretrainedModel($DeepSeekmodel$)
    tokenizer = LoadTokenizer($DeepSeekmodel$)
    lora_config = LoraConfig($rank = 4, \alpha = 16, dropout = 0.1$)
    model = AddLoraToModel(model, lora_config)
    dataset = MyDataset(data, tokenizer)
    dataloader = DataLoader(dataset, batch_size = 8, shuffle = $True$)
    $\alpha = 5e$-5
for epoch = 1 to num_epochs **do**
        model.train()
    for each batch in dataloader **do**
            inputs = MoveBatchToDevice(batch, device)
            outputs = model(inputs)
            loss = ComputeLoss(outputs)
            BackwardPass(loss)
            optimizer.step()
            optimizer.zero_grad()
    end for
end for
return: the fine-tuned model, tokenizer

effective, thereby meeting the dual requirements of professionalism and safety in intelligent mental health support systems.

3. Evaluation Experiments

3.1. Experimental Settings

To evaluate the performance of LLMs in adolescent psychological counselling intent recognition, we conducted experiments on a high-performance computing platform equipped with an NVIDIA L20 GPU , a Intel Xeon Platinum 8457C CPU, and 100GB RAM, using the LLaMAFactory framework. The dataset was split into training, validation, and test sets in an 8:1:1 ratio to ensure data independence and representativeness.

Table 4. LLMs used for fine-tuning.

Model Name	Parameter Size	Features
Llama-3.1	8B	Meta's high-performance large language model
Qwen2.5	7B	Alibaba Cloud's general-purpose large language model
Yi-1.5	6B	Open-source large language model from YITU AI
ChatGLM3	6B	Open-source conversational AI from Zhihu AI
DeepSeek-LLM	7B	Open-source large-scale model from DeepSearch for deep search optimization

For fine-tuning, we adopted the LoRA method to reduce computational cost. The low-rank matrices were set to rank 4, with a scaling factor $\alpha = 16$ and a dropout rate of 0.1 to mitigate overfitting. To strike a balance between accuracy and deployment efficiency, we selected several medium-sized LLMs for comparison. Their configurations are listed in Table 4.

3.2. Evaluation of Intent Classification Performance

To comprehensively assess the performance of LLMs in adolescent psychological counselling, we conducted a series of intent classification experiments using several representative models. Each model was fine-tuned on our domain-specific dataset and evaluated based on standard classification metrics, including accuracy, precision, recall, and F1-score. Figs. 2-4 illustrate the comparative results across three key counselling categories: general distress, mental illness, and SOS emergency help.

Fig. 2 shows the fine-tuning effect of each model on the trouble type task. DeepSeek-LLM and Qwen2.5 performed excellently and led the overall performance,

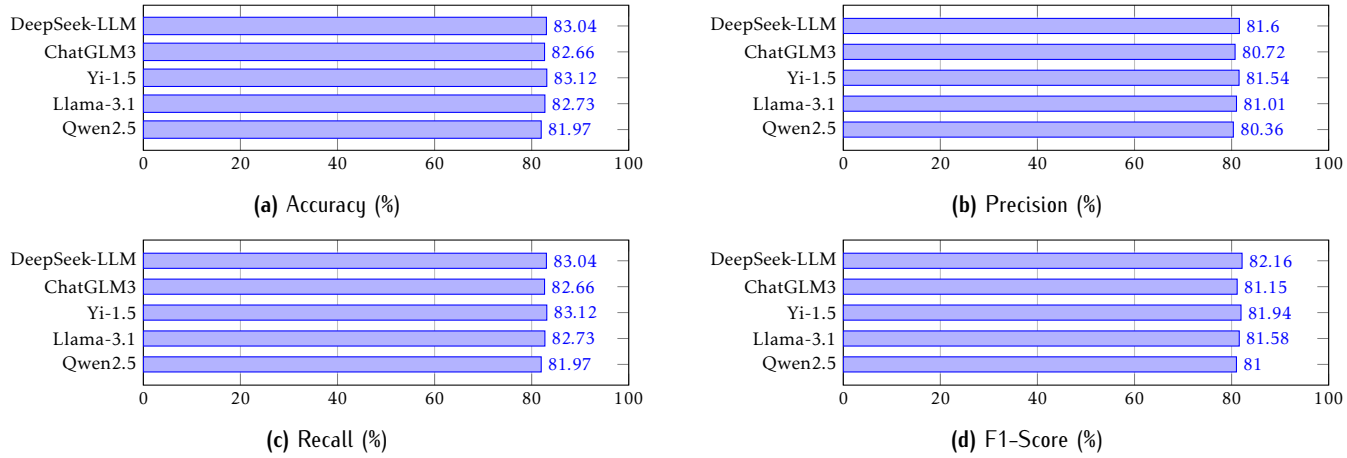


Figure 3. Comparison of fine-tuning LLMs for mental illness.

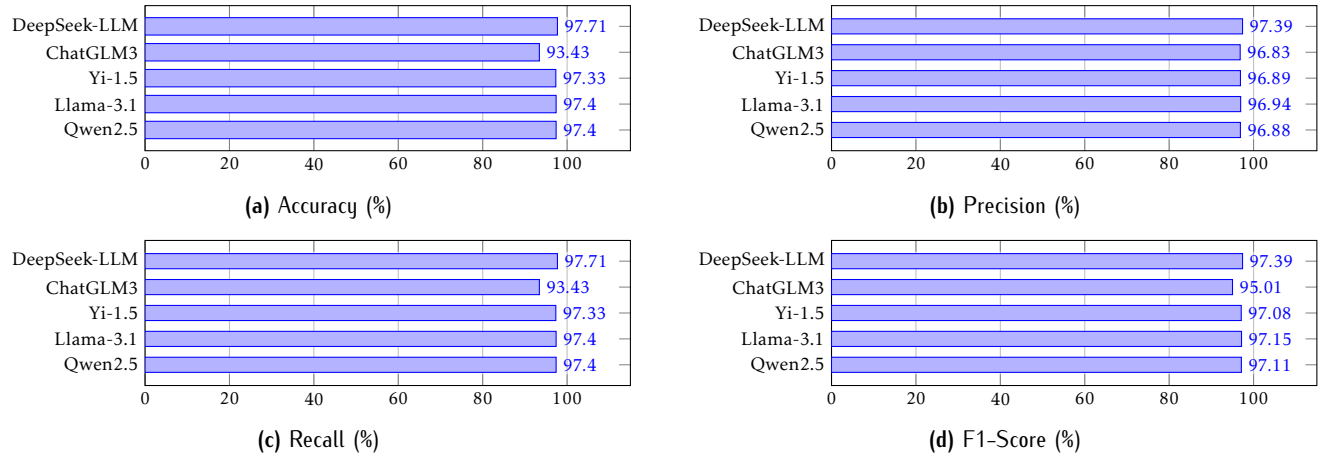


Figure 4. Comparison of fine-tuning LLMs for SOS emergency help.

with accuracy rates of 81.82% and 80.98%, respectively, and F1 scores exceeding 81%. Llama-3.1 also performed well, with an accuracy rate of 78.99% , slightly lower than the previous two, but still showed good performance. In contrast, Yi-1.5 has an accuracy of only 53.71% and a precision of 74.23%, significantly behind other models in all indicators. Yi-1.5's disadvantage in this task may be due to its pre-training data and architecture optimization direction. The trouble type task requires the model to have high semantic understanding and multi-category classification capabilities. In contrast, high precision and recall rely on the model's deep learning capabilities for different problems. DeepSeek-LLM and Qwen2.5 may have been exposed to a broader range of dialogue scenarios during training or have performed more targeted optimizations in the fine-tuning stage, thus performing well in this task. Yi-1.5 may have deficiencies in its generalization ability in multi-category situations, resulting in low classification performance.

Fig. 3 shows that the performance of each model on the mental illness task is relatively close. DeepSeek-LLM and Yi-1.5 have accuracies of 83.04% and 83.12%, respectively, ranking at the top, while Qwen2.5 has an accuracy of 81.97%, slightly lower. In terms of precision, all models are between 80%-82%, indicating that they can make stable classification decisions when identifying conversations related to mental illness. Yi-1.5's performance on this task may be because its training data contains more mental health-related content or its fine-tuning strategy has a good optimization effect. DeepSeek-LLM, with its large pre-training data scale, has shown stronger generalization ability and performs well in recall rate (83.04%) and F1 score (82.16%). Qwen2.5's performance is slightly worse, which may be due to its conservative understanding of mental health-related conversations, resulting in a lower recall rate and F1 score than DeepSeek-LLM and Yi-1.5.

Fig. 4 shows that the accuracy of all models on the SOS emergency help task exceeds 93%. DeepSeek-LLM

achieves the highest accuracy (97.71%), while Llama-3.1 and Qwen2.5 both reach approximately 97.4%, and ChatGLM3 is slightly lower (93.43%). The models show little difference in precision and recall, all above 96.5%, indicating low false alarm rates and a strong ability to recognise emergency help dialogues. The consistently high performance may result from the distinctive semantic features of SOS emergency texts, such as strong emotional words (e.g., "help", "urgent"), as well as their clear structural patterns. These characteristics, together with exposure to similar samples during pre-training, enable the models to generalise effectively after fine-tuning. DeepSeek-LLM performs best, likely due to its broader coverage of emergency dialogue data in the pre-training corpus, whereas ChatGLM3 appears less exposed to such data, resulting in slightly weaker performance.

In summary, task complexity, measured by the number of categories and semantic boundaries, along with the balance of data distribution and the significance of semantic features, significantly influences model performance. The trouble type task involves many categories and ambiguous semantics, where models with weak generalisation ability perform poorly, while those with strong semantic understanding achieve better results. The mental illness task has fewer categories and a balanced dataset, resulting in similar performance across models, with Yi-1.5 showing particular strength in mental health-related data. Although the SOS emergency task suffers from class imbalance, its salient signals and limited categories make it easier to learn, yielding high accuracy overall. DeepSeek-LLM demonstrates the most stable performance across all tasks and is suitable for diverse psychological counselling scenarios. Qwen2.5 performs exceptionally well on the trouble type task, and Yi-1.5 has an advantage in the mental illness task. While all models achieve high accuracy on the SOS emergency task, the issue of imbalance still requires attention. Future work should focus on improving the recall of minority categories and developing data augmentation methods to enhance fairness and robustness. Among all models, DeepSeek-LLM achieves the best balance between performance and resource efficiency, making it the most suitable backbone for our system.

3.3. Ablation Study

To verify the effectiveness of the module, we conducted ablation experiments by removing key components to assess their individual and combined contributions in the psychological counselling task. Based on the fine-tuned DeepSeek-LLM, we designed three experimental configurations:

1. *Baseline model*: generates responses solely from user input without retrieval or structured prompting.

2. *Prompt Only*: introduces structured prompt construction such as role specification, behavioural instructions, and output formatting, but does not include knowledge retrieval.

3. *RARG*: integrates both semantic retrieval and structured prompting, representing the complete generation pipeline proposed in this work.

Using a consistent test set across all configurations, we evaluated model performance with two automatic metrics: F1-score and BLEU-2. The F1-score measures semantic alignment between generated responses and annotated labels, while BLEU-2 evaluates linguistic quality based on n-gram similarity to human references. As shown in Table 6, the baseline model achieved an F1-score of 75.31% and a BLEU-2 score of 0.426. With structured prompting, the scores improved to 78.64% and 0.485. When both retrieval and structured prompting were applied, the full model achieved the highest performance, with an F1-score of 81.70% and a BLEU-2 score of 0.538.

These results show that structured prompting enhances semantic relevance and fluency, while semantic retrieval supplies domain knowledge that improves contextual understanding and response specificity. The integration of both yields the best overall performance, validating the effectiveness of the proposed RAG-TPC design.

3.4. User Evaluation

To ensure ethical rigour and professional validity, all user evaluations were conducted exclusively by individuals with a professional background in psychological counselling. Specifically, we invited 10 certified counselling instructors, 50 senior undergraduate psychology majors, and 40 graduate students in psychology-related disciplines. This approach ensured that the evaluation of the system, particularly in high-risk and sensitive scenarios, was informed by expert judgment rather than layperson perceptions, thereby upholding ethical standards in mental health research. We used the fine-tuned DeepSeek-LLM to identify the consultation category and called DeepSeek-Chat to generate responses. Each evaluator was asked to submit four test cases per intent category, balancing the number of mental illness and SOS cases. After each test, they completed a questionnaire evaluating both the system's recognition accuracy and the appropriateness of its responses. The questionnaire comprised four core items, each evaluated on a five-point Likert scale, the details of which are presented in Table 5. In total, 4,774

Table 5. Questionnaire for user evaluation.

Question No.	Question	Multiple Choice Answers				
1	Did it accurately identify the categories of the consultant's troubles?	❑ 5=Very accurate	❑ 4=Accurate	❑ 3=Basically accurate	❑ 2=Inaccurate	❑ 1=Very inaccurate
2	Did it accurately identify the consultant's mental illnesses?	❑ 5=Very accurate	❑ 4=Accurate	❑ 3=Basically accurate	❑ 2=Inaccurate	❑ 1=Very inaccurate
3	Did it accurately identify the consultant's SOS?	❑ 5=Very accurate	❑ 4=Accurate	❑ 3=Basically accurate	❑ 2=Inaccurate	❑ 1=Very inaccurate
4	Did the system's response take into account the consultant's intentions and provide a reasonable reply?	❑ 5=Very reasonable	❑ 4=Reasonable	❑ 3=Basically reasonable	❑ 2=Unreasonable	❑ 1=Very unreasonable

Table 6. Comparison of F1 and BLEU-2 under different model configurations.

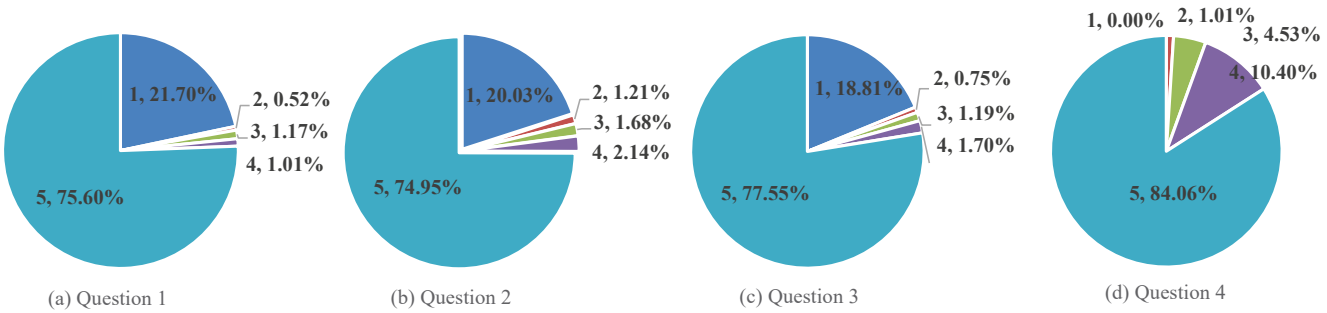
Model Configuration	Retrieval	Prompt	F1 (%)	BLEU-2
Baseline	✗	✗	75.31	0.426
Prompt Only	✗	✓	78.64	0.485
Full RARG	✓	✓	**81.70**	**0.538**

valid responses were collected, with the evaluation outcomes illustrated in Fig. 5. As shown in Fig. 5, evaluators' assessments of the system's recognition ability exhibited polarisation, with many responses clustering at either "very accurate" or "very inaccurate". While the actual classification accuracies (81.82%, 83.04%, 97.71%) surpassed the proportions of "very accurate" ratings (75.60%, 74.95%, 77.55%), this divergence highlights a perception gap, possibly due to stricter clinical standards held by professional raters or higher expectations in real-world scenarios. Despite this, evaluators expressed intense satisfaction with the system's generated responses. A total of 84.06% rated them as "very reasonable", and 94.46% rated them as at least "reasonable". This outcome reflects the effectiveness of our structured prompt design and intent-guided retrieval in producing empathetic and contextually appropriate replies. Importantly, by involving trained psychological professionals in the evaluation loop, we addressed key ethical concerns in AI-assisted counselling, particularly those related to safety, appropriateness, and risk mitigation in emotionally sensitive interactions. Additionally, several safeguards were implemented to ensure the responsible deployment. The training dataset was fully anonymised to preserve privacy. Structured prompts and intent classification further regulated the model's behaviour, helping to avoid inappropriate, misaligned, or overly diagnostic responses. These measures contribute to enhanced transparency, controllability, and ethical robustness, distinguishing our approach from generic LLM applications.

4. Related Work

4.1. Mental Health Recognition and Reasoning Based on LLMs

In recent years, the application of LLMs in the mental health field has focused mainly on emotion recognition, mental state reasoning, and the detection of depression and suicide risk. For example, in 2022, McDonagh et al. [42] examined the application of NLP-based mental health dialogue systems in recognizing depression and anxiety, showing that AI has progressed in automated emotion detection. In addition, in 2024, Xu et al. [31] fine-tuned open-source LLMs (*e.g.*, Alpaca and FLAN-T5) on multiple mental health datasets (such as Dreaddit, DepSeverity, SDCNL, CSSRS-Suicide) to develop Mental-Alpaca and Mental-FLAN-T5. The results showed that instruction tuning could improve LLMs' performance in mental health reasoning tasks. Similarly, in 2024, Ping [34] developed a machine learning model that achieved an accuracy of 89% in mental health state analysis tasks, further confirming the potential of LLMs in mental health assessment. Furthermore, a large-scale benchmarking study by VarastehNezhad et al. [4] systematically revealed significant variations in the emotional and sentiment profiles of responses generated by eight major LLMs to mental health queries, highlighting the inherent unpredictability and model-specific biases of using off-the-shelf general-purpose models in this sensitive domain. However, these studies mainly focused on automated detection rather than dialogue-based interventions, and they often relied on social media data or standardized questionnaires for inference without optimizing interactive psychological counselling capabilities. Notably, these approaches primarily utilize the intrinsic knowledge of LLMs and lack mechanisms for retrieving and incorporating external, up-to-date, or domain-specific knowledge, which is a core aspect of the RAG paradigm [10]. Soman et al. demonstrated that combining fine-tuning of LLMs with the RAG approach significantly improves accuracy in psychological counseling interactions with humans and generates more human-like responses [7]. Beyond retrieval, recent advances in emotion understanding, such as the cognitive-affective chain-driven framework proposed by Chen et al. [5], offer promising pathways for models to achieve a more



(a) Question 1 (b) Question 2 (c) Question 3 (d) Question 4

Figure 5. Results of user evaluation.

nuanced comprehension of user states. These advancements collectively highlight the evolving landscape of LLM applications in mental health, yet they also underscore the critical need for more interactive and knowledge-grounded approaches like RAG to bridge the gap between recognition and effective intervention.

4.2. The Application of LLMs in Mental Health Dialogue Systems

Besides mental health state detection, LLMs have also been applied to dialogue-based mental health support systems. For instance, Ma et al. [22] analyzed 120 user posts (comprising 2,917 comments) from the Reddit platform's r/Replika forum, finding that LLM-based dialogue agents (such as Replika) could improve users' real-world social skills, boost self-confidence, and reduce self-harming behaviour to some extent. Furthermore, in 2023, Cho et al. [35] found that fine-tuned LLMs demonstrated some level of empathy and contextual adaptability when interacting with high-functioning autistic adolescents, as confirmed by psychological experts' assessments of the dialogue content. Concurrently, Li [6] introduced AIVA, an AI-based virtual companion explicitly designed for emotion-aware interaction, showcasing the ongoing development of empathetic AI agents.

The above studies demonstrate the potential of LLMs in mental health dialogues. However, they generally lack an in-depth understanding of intent, as they primarily rely on general-purpose models, making it challenging to identify adolescents' psychological states and dynamically adjust responses accurately. Moreover, these systems typically operate as closed-book models, generating responses based solely on their pre-trained parameters. This limits their ability to provide responses grounded in specific therapeutic frameworks (e.g., CBT techniques) or the latest clinical guidelines, a gap that RAG architectures are designed to fill by dynamically accessing external knowledge bases. In contrast, this study enhances LLMs' intent recognition capabilities through fine-tuning and incorporates a dialogue category classification strategy to generate psychological counseling responses more aligned with psychological principles. Crucially, it integrates a RAG module to fetch relevant knowledge, ensuring responses are not only empathetic but also informed by verified psychological principles. This approach improves personalization and adaptability, providing more effective support for adolescent mental health interventions.

4.3. The Challenges of LLMs in the Field of Mental Health

Despite the promising application of LLMs in mental health recognition and dialogue systems, several challenges remain, including personalized adaptability and ethical and data security issues. For example, in 2024, Na [23] developed CBT-LLM based on CBT principles and fine-tuned LLMs to generate structured and professional CBT responses. However, this study primarily focused on applying CBT techniques rather than adapting LLMs to adolescents' psychological needs. Similarly, in 2023, Lai et al. [25] proposed a psychological counselling assistance tool (Psy-LLM) that combined LLM pre-training data with psychological question-answering tasks. While Psy-LLM performed well in responding to mental health issues, it struggled to handle unstructured natural dialogues, especially when facing emotional expressions or psychological crises. A key limitation of such fine-tuned models is the static nature of their knowledge, which can become outdated. RAG systems address this by decoupling the model's parametric knowledge from a dynamically updatable non-parametric knowledge store, allowing for continuous integration of new information without costly retraining [8].

Furthermore, the retrieval component in RAG frameworks itself presents unique challenges in the mental health context. Retrieval performance is highly dependent on the quality, breadth, and ethical curation of the underlying knowledge base [26]. Inaccurate or biased retrieved documents can lead to misguided responses,

posing significant risks. Therefore, constructing a reliable, clinically-validated, and ethically-sourced knowledge base is a critical research challenge that our work addresses through expert-annotated content and structured knowledge sources.

Additionally, ethical and data security issues are significant challenges in applying LLMs to mental health. For example, Ayers et al. [20] and Chen et al. [24] found that well-trained LLMs outperformed human peers in mental health support tasks. However, the complexity of training data sources may introduce potential biases or inappropriate psychological advice. Ensuring user data privacy and preventing LLMs from inadvertently disclosing sensitive information during psychological counselling remain critical issues in current research. RAG systems can offer a degree of transparency and controllability over the source of generated information, as responses can be traced back to retrieved documents. This allows for better auditing and oversight compared to purely parametric generation, which is a black-box process. However, this also necessitates rigorous vetting of the knowledge base to prevent the retrieval and amplification of harmful content.

In contrast to these studies, the present study focuses on optimizing LLMs' personalized adaptability in adolescent psychological counselling. By adopting a RAG framework, our approach explicitly tackles the challenges of knowledge grounding and dynamic information access. This study aims to generate more psychologically appropriate dialogue content by fine-tuning LLMs and integrating expert-annotated psychological counselling data while ensuring ethical compliance.

5. Conclusion

This paper proposes a DeepSeek-based framework for intent understanding and response generation tailored to adolescent psychological counselling. To address the shortage of timely and accessible mental health support, we constructed a high-quality domain-specific dataset and applied parameter-efficient fine-tuning to optimise intent recognition. By integrating intent-guided retrieval and structured prompt design, the model generates safe, empathetic, and context-aware responses. Experimental and user evaluation results confirm that RAG-TPC significantly improves both classification accuracy and dialogue quality.

Future work will focus on enhancing emotional sensitivity, reducing bias, and incorporating multimodal inputs to more accurately reflect real-world interactions. It is also necessary to further explore deployment monitoring, user feedback loops, and ethical auditing to ensure long-term trust and effectiveness of AI-assisted psychological counselling systems.

Acknowledgement. This work was partially supported by the National Natural Science Foundation of China (No. 61762016), the Guangxi Natural Science Foundation(No.2024JJB180051), the Research Fund of Guangxi Key Lab of Multi-source Information Mining & Security (No. 24-A-01-01), and the Education Ministry Key Lab of Education Blockchain and Intelligent Technology (No. EBME24-05, EBME24-11).

References

[1] Hu E.J., Shen Y., Wallis P., Allen-Zhu Z., Li Y., Wang S., Wang L. and Chen W. et al. (2022) *LoRA*: Low-rank adaptation of large language models. (ICLR), 1(2), 3.

[2] Johnson J., Douze M. and Jegou H. (2021) *Billion-Scale Similarity Search with GPUs* (IEEE Transactions on Big Data), 7(03), 535–547.

[3] Shah H.A., Islam A., Tariq Z.U.A., Belhaouari S.B. and Househ M. (2025) *Retrieval Augmented Generation System for Mental Health Information* (in *MedInfo 2025: Healthcare Smart x Medicine Deep*, IOS Press), 693–697.

[4] VarastehNezhad A., Tavasoli R., Elyasi S., LotfiNia M.H. and Farbeh H. (2025) *AI in Mental Health: Emotional and Sentiment Analysis of Large Language Models' Responses to Depression, Anxiety, and Stress Queries* (arXiv:2508.11285 [cs.CL]).

[5] Chen S., Liu Z., Zhang Z., Qin K., Qian Y. and Ma F. (2026) *A cognitive-affective chain-driven framework for emotion understanding*(Information Processing & Management), 63 (2, Part A), 104367.

[6] Li C. (2025) *AIVA: An AI-based Virtual Companion for Emotion-aware Interaction* (arXiv:2509.03212 [cs.CV]).

[7] Soman G., Judy M.V. and Abou A.M. (2025) *Human guided empathetic AI agent for mental health support leveraging reinforcement learning-enhanced retrieval-augmented generation* (Cognitive Systems Research), 90, 101337.

[8] Guu K., Lee K., Tung Z., Pasupat P. and Chang M.-W. (2020) *REALM: Retrieval-Augmented Language Model Pre-Training* (arXiv:2002.08909 [cs.CL]).

[9] Gao Y., Xiong Y., Gao X., Jia K., Pan J., Bi Y., Dai Y., Sun J., Wang M. and Wang H. (2024) *Retrieval-Augmented Generation for Large Language Models: A Survey* (arXiv:2312.10997 [cs.CL]).

[10] Lewis P., Perez E., Piktus A., Petroni F., Karpukhin V., Goyal N., Küttler H., Lewis M., Yih W.-t., Rocktäschel T., Riedel S. and Kiela D. (2021) *Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks* (arXiv:2005.11401 [cs.CL]).

[11] Dong W., Liu Y., Bai R., Zhang L. and Zhou M. (2025) *The prevalence and associated disability burden of mental disorders in children and adolescents in China: a systematic analysis of data from the Global Burden of Disease Study* (The Lancet Regional Health–Western Pacific), 55.

[12] Reynolds L. and McDonell K. (2021) *Prompt programming for large language models: Beyond the few-shot paradigm* (in *Extended abstracts of the 2021 CHI conference on human factors in computing systems*), 1–7.

[13] Marvin G., Nakayiza H., Jjingo D. and Nakatumba-Nabende J. (2023) *Prompt engineering in large language models* (in *International conference on data intelligence and cognitive informatics*, Springer), 387–402.

[14] Nigh M. (2023) *ChatGPT3-Free-Prompt-List: A free guide for learning to create ChatGPT3 Prompts* https://github.com/mattnigh/ChatGPT3-Free-Prompt-List.

[15] Chen H. and Li Z. (2023) *ChatGPT Advanced: An Introduction to Prompt Engineering* (in *ChatGPT Advanced: An Introduction to Prompt Engineering*, Peking University Press, Beijing), 98–130 (in Chinese).

[16] Wang M., Liu Y., Liang X., Li S., Huang Y., Zhang X., Shen S., Guan C., Wang D., Feng S. et al. (2024) *LangGPT: Rethinking structured reusable prompt design framework for LLMs from the programming language* (arXiv preprint arXiv:2402.16929).

[17] DukeManh (2023) *Prompting Introduction* https://github.com/dair-ai/Prompt-Engineering-Guide/blob/main/guides/prompts-intro.md, accessed 19 December 2024.

[18] Van Dijk D.A., Meijer R.M., van den Boogaard T.M., Spijker J., Ruhé H.G. and Peeters F.P.M.L. (2023) *Worse off by waiting for treatment? The impact of waiting time on clinical course and treatment outcome for depression in routine care* (Journal of Affective Disorders), 322, 205–211.

[19] Amin M.M., Cambria E. and Schuller B.W. (2023) *Will affective computing emerge from foundation models and general artificial intelligence? A first evaluation of ChatGPT* (IEEE Intelligent Systems), 38(2), 15–23.

[20] Ayers J.W., Poliak A., Dredze M., Leas E.C., Zhu Z., Kelley J.B., Faix D.J., Goodman A.M., Longhurst C.A., Hogarth M. and Smith D.M. (2023) *Comparing Physician and Artificial Intelligence Chatbot Responses to Patient Questions Posted to a Public Social Media Forum* (JAMA Internal Medicine), 183(6), 589–596.

[21] Das J.K., Salam R.A., Lassi Z.S., Khan M.N., Mahmood W., Patel V. and Bhutta Z.A. (2016) *Interventions for Adolescent Mental Health: An Overview of Systematic Reviews* (The Journal of Adolescent Health), 59(4S), S49–S60.

[22] Ma Z., Mei Y. and Su Z. (2024) *Understanding the benefits and challenges of using large language model-based conversational agents for mental well-being support* (in *AMIA Annual Symposium Proceedings*), 2023, 1105.

[23] Na H. (2024) *CBT-LLM: A Chinese large language model for cognitive behavioral therapy-based mental health question answering* (arXiv preprint arXiv:2403.16008).

[24] Chen D., Parsa R., Hope A., Hannon B., Mak E., Eng L., Liu F.-F., Fallah-rad N., Heesters A.M. and Raman S. (2024) *Physician and artificial intelligence chatbot responses to cancer questions from social media* (JAMA Oncology).

[25] Lai T., Shi Y., Du Z., Wu J., Fu K., Dou Y. and Wang Z. (2023) *Psy-llm: Scaling up global mental health psychological services with ai-based large language models* (arXiv preprint arXiv:2307.11991).

[26] De Choudhury M., Pendse S.R. and Kumar N. (2023) *Benefits and harms of large language models in digital mental health* (arXiv preprint arXiv:2311.14693).

[27] Elyoseph Z., Levkovich I. and Shinan-Altman S. (2024) *Assessing prognosis in depression: comparing perspectives of AI models, mental health professionals and the general public* (Family Medicine and Community Health), 12(Suppl 1), e002583.

[28] McCloud T., Jones R., Lewis G., Bell V. and Tsakanikos E. (2020) *Effectiveness of a Mobile App Intervention for Anxiety and Depression Symptoms in University Students: Randomized Controlled Trial* (JMIR mHealth and uHealth), 8(7), e15418.

[29] Habicht J., Viswanathan S., Carrington B., Hauser T., Harper R. and Rollwage M. (2023) *Closing the accessibility gap to mental health treatment with a conversational ai-enabled self-referral tool*(medRxiv).

[30] Wainberg M.L., Scorza P., Shultz J.M., Helpman L., Mootz J.J., Johnson K.A. and Arbuckle M.R. (2017) *Challenges and opportunities in global mental health: a research-to-practice perspective* (Current Psychiatry Reports), 19, 1–10.

[31] Xu X., Yao B., Dong Y., Gabriel S., Yu H., Hendler J., Ghassemi M., Dey A.K. and Wang D. (2024) *Mental-LLM: Leveraging Large Language Models for Mental Health Prediction via Online Text Data* (Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies), 8(1), 1–32.

[32] Zhang Z. and Wang J. (2024) *Can AI replace psychotherapists? Exploring the future of mental health care* (Frontiers in Psychiatry), 15, 1444382.

[33] Zhong W., Luo J. and Zhang H. (2024) *The therapeutic effectiveness of artificial intelligence-based chatbots in alleviation of depressive and anxiety symptoms in short-course treatments: A systematic review and meta-analysis* (Journal of Affective Disorders).

[34] Ping Y. (2024) *Experience in psychological counseling supported by artificial intelligence technology* (Technology and Health Care), 32(6), 3871–3888.

[35] Cho Y., Kim M., Kim S., Kwon O., Kwon R.D., Lee Y. and Lim D. (2023) *Evaluating the Efficacy of Interactive Language Therapy Based on LLM for High-Functioning Autistic Adolescent Psychological Counseling* (arXiv preprint arXiv:2311.09243).

[36] World Health Organization (2021) *WHO report highlights global shortfall in investment in mental health* https://www.who.int/news/item/08-10-2021-who-report-highlights-global-shortfall-in-investment-in-mental-health

[37] Mahase E. (2020) *Workforce crisis has left mental health staff at "breaking point" as demand rises* (BMJ).

[38] Thornicroft G., Sunkel C., Aliev A.A., Baker S., Brohan E., el Chammay R., Davies K., Demissie M., Duncan J., Fekadu W., Gronholm P.C., Guerrero Z., Gurung D., Habtamu K., Hanlon C., Heim E., Henderson C., Hijazi Z., Hoffman C. and Winkler P. (2022) *The Lancet Commission on ending stigma and discrimination in mental health*(The Lancet), 400(10361), 1438–1480.

[39] Lee Y.-C., Cui Y., Jamieson J., Fu W. and Yamashita N. (2023) *Exploring Effects of Chatbot-based Social Contact on Reducing Mental Illness Stigma*(in *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, ACM), 1–16.

[40] Vespoli G., Taddei B., Imbimbo E., De Luca L. and Nocentini A. (2024) *The concept of privacy in the digital world according to teenagers*(Journal of Public Health).

[41] Towner E., Grint J., Levy T., Blakemore S.-J. and Tomova L. (2022) *Revealing the self in a digital world: A systematic review of adolescent online and offline self-disclosure*(Current Opinion in Psychology), 45, 101309.

[42] McDonagh C., Lynch H. and Hennessy E. (2022) *Do stigma and level of social anxiety predict adolescents' help-seeking intentions for social anxiety disorder?* (Early Intervention in Psychiatry), 16(4), 456–460.

[43] Wang H.L., Wu Z.Z. and Lang J.Y. (2020) *Emotional First Aid Raw Dataset*https://github.com/chatopera/efaqa-corpus-zh, accessed 22 April 2020.

[44] Ding N., Qin Y., Yang G., Wei F., Yang Z., Su Y., Hu S., Chen Y., Chan C.-M., Chen W. et al. (2023) *Parameter-efficient fine-tuning of large-scale pre-trained language models* (Nature Machine Intelligence), 5(3), 220–235.