

MoodMate: A Sentiment-Responsive Question-Answering Psychologist Chatbot for Emotional Support by Fine-tuning Llama-3.2-3B-Instruct Model

Jiesen LONG (2230026109), Zhiyue CHOU (2230026028), Jiaying SHAN (2230026230),
Yizhen XIN (2230026179), Ruoyi WANG (2230026155), Shuxuan DING (2230026232)

Abstract—This paper introduces MoodMate, an emotion-responsive chatbot designed to provide psychological support to users using Llama-3.2-3B model and advanced technology to improve efficiency and scalability. By employing techniques such as 4-bit quantization and LoRA fine-tuning, MoodMate operates efficiently on user devices. In addition, we employed prompt engineering techniques to optimize the model's interaction experience. We clarified the role of MoodMate as a psychological companion through well-designed prompts, which enabled MoodMate to provide empathetic and supportive advice to the user while remaining neutral and non-judgmental. More importantly, when MoodMate recognizes a user's tendency to self-harm through conversation analysis, it is able to immediately activate a crisis intervention mechanism to ensure the user's safety. The dataset is split into 80% for training and 20% for testing, with a total of more than 50,000 labeled conversations. And the data types include both Chinese and English, covering a wide range of emotional scenarios. MoodMate supports both Chinese and English text and voice interaction in a clear and beautiful UI that enhances the convenience and immersion of the user experience. The experimental results show that after 10 training epochs, the validation loss of the model is 0.387. These case studies further demonstrate MoodMate's ability to generate empathy and offer suggestions, as well as the different chatting styles it displays when processing conversations in English and Chinese. The design of MoodMate highlights its potential as a mental health support tool, demonstrating its adaptability in real-world applications and its ability to meet the needs of a wide range of users.

I. INTRODUCTION

WITH the increasing demand for accessible mental health support, AI-driven solutions have become a promising avenue for providing psychological assistance. One such solution is **MoodMate**, a sentiment-responsive, question-answering chatbot designed to offer emotional support and comfort through natural language processing (NLP) techniques. MoodMate aims to simulate the empathetic responses of a human psychologist, leveraging AI to engage with users in emotionally intelligent conversations. By analyzing the sentiment of user input, MoodMate provides personalized, supportive responses to a wide range of mental health concerns, from stress and anxiety to more general emotional well-being.

The development of such a chatbot is a critical step forward in the digital mental health landscape, allowing for scalable, on-demand support in a variety of contexts. Whether integrated

into mobile applications, websites, or virtual mental health platforms, MoodMate serves as an accessible initial point of contact for individuals seeking emotional guidance, particularly in situations where traditional in-person therapy may not be immediately available. This paper explores the underlying AI technologies that power MoodMate, including the integration of sentiment analysis and question-answering (QA) systems, as well as the latest advancements in NLP techniques and their application in the mental health domain.

II. LITERATURE REVIEW

A. What is MoodMate?

MoodMate is a virtual psychologist chatbot that uses NLP techniques to engage users in emotionally intelligent conversations. Its primary goal is to provide comfort, guidance, and emotional support, emulating aspects of a therapeutic relationship. The system works by analyzing users' input for emotional cues, such as sentiment and tone, to generate contextually relevant and empathetic responses. Unlike traditional chatbots that may focus on factual question-answering, MoodMate is designed to recognize and respond to users' emotional states, offering both practical advice and empathetic reassurance.

By combining sentiment analysis with QA capabilities, MoodMate can deliver personalized responses based on the user's emotional context, offering suggestions for stress management, self-care, or even general emotional validation. This dual functionality allows the chatbot to serve as a valuable tool for individuals seeking quick, accessible support in managing their mental health.

B. How It Works?

The core technologies behind MoodMate are drawn from NLP and machine learning (ML). At its core, the chatbot relies on sentiment analysis to understand the emotional undertone of the user's messages. This involves classifying the sentiment of the text—whether it is positive, negative, or neutral—along with more granular emotional states such as sadness, anger, joy, or frustration. Tools such as BERT (Bidirectional Encoder Representations from Transformers) and GPT (Generative Pre-trained Transformer) are used to interpret the context of the conversation and generate responses that align with the user's emotional state.

In addition to sentiment analysis, MoodMate employs a Question-Answering (QA) system that enables it to respond effectively to direct queries. The QA model is trained to recognize user questions and provide meaningful answers in therapeutic frameworks. To achieve high-quality responses, MoodMate draws from a large dataset of therapy-related dialogues, adapting them to the specific needs of the user in real time.

Through these techniques, MoodMate delivers conversational support that feels natural and human-like, with the ability to adjust responses based on the emotional context of the user's input.

C. Latest Research and Contemporary Technology

The field of conversational AI has seen rapid advancements, particularly in the areas of sentiment analysis and QA systems. One significant development is the use of transformer-based models like GPT-3 and BERT, which have revolutionized the understanding of natural language and sentiment. These models excel at understanding nuanced emotions in text, making them ideal for applications in mental health. For instance, recent research by Mundargi et al. (2024) explored how fine-tuning these models on mental health dialogues can significantly improve the emotional intelligence of chatbots.

Another important advancement is the integration of reinforcement learning to optimize chatbot responses in emotionally sensitive contexts. In this approach, the chatbot learns from user feedback, continually refining its ability to provide supportive and contextually appropriate responses. Studies like those by Abubakar et al. (2022) have shown the potential for reinforcement learning to enhance chatbot empathy, making it more attuned to subtle emotional cues such as tone shifts or changes in sentiment over time.

Recent innovations also include multimodal sentiment analysis, where chatbots can analyze not only text but also voice and facial expressions to better understand the user's emotional state (Jing et al., 2024). Although still in the early stages, this approach promises to provide a richer, more nuanced understanding of emotional cues, thereby improving the chatbot's ability to respond empathetically.

D. Integration with NLP and AI Applications

The integrated use of sentiment analysis technology and QA chatbot technology through Natural Language Processing (NLP) and Artificial Intelligence (AI) can be realized in numerous ways to enhance the intelligence of the system and the user experience.

First, this technology focuses on combining deep learning with natural language. The deep memory network-based emotional cause extraction technique demonstrates the application of sentiment analysis in Q&A scenarios, proving that the effectiveness of sentiment recognition lies in helping the system to recognize the user's potential emotional state and give appropriate responses (Gui et al., 2017). By training the deep learning model, the system can capture the user's emotional characteristics more accurately and use natural language to characterize the text, and these can help the researchers understand the user's emotions better.

Second, the QA chatbot module plays an important role in sentiment analysis. This module ensures that the system is able

to handle continuous user interactions, maintains the context of the conversation, and enables the model to better grasp changes in user sentiment and be able to respond appropriately. e.g., the AllHands framework uses a large-scale language model to analyze the large-scale feedback provided by the user through Q&A and sentiment analysis, which demonstrates the convergence and potential of Q&A and sentiment analysis in AI applications (Zhu et al., 2024).

In addition, sentiment analysis is included as a separate module to the Q&A system, increasing the adaptability and flexibility of the system. This combination not only enhances the intelligence of sentiment analysis, but also lays the foundation for intelligent applications, showing great potential especially in personalized services.

III. METHODOLOGY

The implementation of the LoRA based emotional detection mental health Chat-bot includes the natural language processing, the resource-efficient computation, and the fine-tuning strategies which all generated in a model training system. In this methodology section, there is a detailed introduction on the model section, quantification, efficient fine-tuning of parameters, preparation of data set, the training optimization and inference strategies.

The method utilizes techniques such as 4-bit quantization, Low Rank Adaptation (LoRA), and fine-grained data set design to successfully find a strike a balance between computational productivity. These techniques work in tandem to provide a solution that is both scalable and efficient, capable of generating accurate and contextualized responses.

A. Model construction

At core of the building of the QA chat bot system is Meta AI's large-scale language model, LLaMA-3B, a state-of-the-art transformer architecture known for its ability to generate consistent and contextually relevant text. LLaMA-3B was chosen for this system due to its superior performance in causal language modeling tasks. However, due to its large parameter size, LLaMA-3B presents notable challenges in terms of computational resources and memory requirements, especially in resource-constrained environments.

In order to make these difficult problems solved, the project also introduced quantization strategies in the construction of the model, which greatly reduced the memory footprint and computational requirements. As a result, the deployment of the model is guaranteed to be of higher quality.

The quantification step including two parts:

Step 1: The quantization steps are implemented using the NF4(Normal Float 4) format, which reduces the precision of the model weights by a factor of 7, i.e., from 32 to 4-bit representations, dramatically reducing the memory consumption by 75%, while allowing the model to be highly accurate and precise.

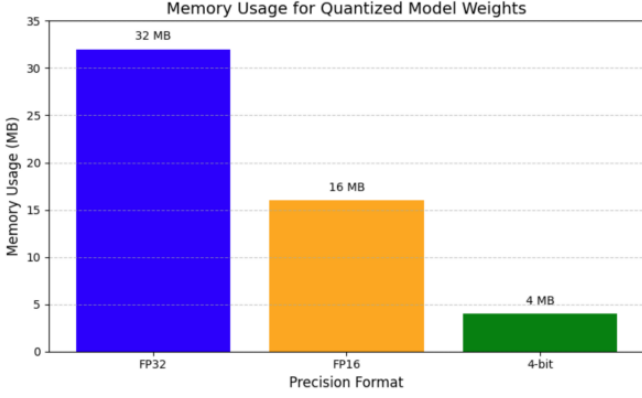


Fig. 1. Memory usage for quantized model weights

The mathematical principle of the quantization process can be expressed as follows:

$$\omega_{\text{quantized}} = \text{round}\left(\frac{\omega_{\text{FP32}} - \mu}{s}\right) \cdot s + \mu$$

ω_{FP32} indicates the original model weights with 32-bit precision, μ is the mean of the weight distribution, and s is the scale factor of the normalized weight range.

The NF4 format further embeds computationally efficient techniques using bfloat16 for middle manipulations, which strengthens numerical stability and maintains the integrity of the gradient computation during model training.

Step 2: In order to configure the quantification algorithm, the project Uses the BitsAndBytesConfig library, 4-bit precision can be seamlessly integrated into the model schema. The configuration definitions are as follows:

```
nf4_config = BitsAndBytesConfig(
    load_in_4bit=True,
    bnb_4bit_quant_type="nf4",
    bnb_4bit_use_double_quant=True,
    bnb_4bit_compute_dtype=torch.bfloat16
)
```

This configuration branch features model performance preservation. The integration within NF4 quantization provides assurance that the LLaMA-3B model is not only resource-effective but also suitable for massive sessions adoption.

B. Efficient fine-tuning through LoRA

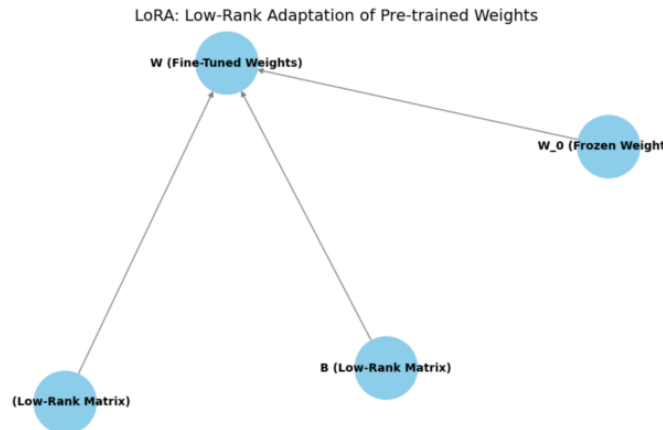


Fig. 2. LoRA Structure for Fine-Tuning Pre-trained Weights

The fine-tuning of large language models at scale, such as LLaMA-3B, usually requires the updating of millions or even billions of parameters, which is computationally and memory

intensive. This challenge is particularly critical where hardware resources are constrained.

$$W = W_0 + \alpha \cdot AB$$

W represents the pre-trained weight matrix.

A and B denotes the trainable low-rank matrices.

α represents the Scaling factor controlling adaptation intensity.

The benefits of LoRA include the ability to achieve fine-tuning of a small number of parameters (<0.1% of the full model size) while significantly reducing GPU memory and computational requirements, ultimately enabling rapid experimentation with minimal resource overhead.

The Following hyperparameters were used:

Ranking (r): 64;

The scale factor (α):16;

Dropout: 10% to mitigate the over-fitting during training.

LoRA achieves significant computational savings, by freezing the majority of the model's arguments and fine-tuning only the low-rank matrices.

C. Database input and data format pre-processing

QA chat bot systems depend on high-quality session datasets to ensure significant and accurate responses. To accomplish this, customized command-based datasets were designed. Each data item is formatted in a more structured format:

This structured input format comprises like the example:

```
<s>[INST] <<SYS>> You are a mental health counselor.
<<SYS>> How can I manage anxiety? [/INST] Practice
mindfulness techniques and consider consulting a professional.
</s>
```

System Message: a prompt designating the role of the chatbot (e.g., psychological assistant, financial advisor);

User Input: a query or statement provided by the user's input.

Model response: the trained output.

In order to optimize the utilization of GPU memory, smaller datasets were concatenated into sequences of up to 2048 tokens (the maximum input length of the model). This approach reduces the amount of padding required and optimizes the efficiency of the training operation.

D. The function optimization and model training

The training process was carefully configured using advanced technology to enable stability and optimize performance. The learning rate was set to 1e-4 to offer a reliable basis for training. The batch size was set to 4 samples per processing device, a choice appropriate for resource-constrained environments, ensuring efficient computations and stable gradient updates.

To mitigate the risk of instability, gradient trimming was employed, with the trimming threshold set to 0.3. The technique of gradient trimming is widely acknowledged for preventing gradient blow-up, as demonstrated, where it ensured numerical stability and facilitated model convergence. Similarly, the effectiveness of gradient cropping in accelerating training and improving convergence can be demonstrated through an adaptive approach.

The training process employs a 3% warm-up ratio, reserving a portion of the total steps for gradually increasing the learning rate, which is a common practice to improve the stability of early training. In addition, the training effectively modelled

larger batch sizes using gradient accumulation. An approach that addresses GPU memory constraints by aggregating the gradient over multiple steps before performing backpropagation. Accumulation of gradients has been shown to optimize memory usage and stabilize training, especially when dealing with large datasets and deep architectures.

These carefully chosen strategies reflect a holistic approach to configuring deep learning module training parameters, designed to balance computational efficiency, resource constraints, and the need to achieve robust model aggregation.

E. The strategies of interface

1) The chat bot generates responses through the utilization of top-p nucleus sampling and temperature scaling, within the period of interface.

The equation is

$$P(\text{word}) \sim \text{Softmax}\left(\frac{z}{T}\right)$$

T represents the control temperature for stochasticity, and z represents the model output in logarithmic terms.

In the model, T is set to 0.95 and the top-p value set to 0.9, meaning the capture of 90% cumulative probability mass.

2) The evaluation standard of the model includes two aspects.

The degree of perplexity (PPL) is a key statistical metric used to evaluate the performance of a language model. Basically, it evaluates the model's ability to predict the next word (or words) given the first few words in the data set. To put it in more technological terms, the perplexity quantifies the ability of a probability distribution or probabilistic model to predict a sample. Less perplexity indicates that the model is better at prediction of words in the data set, suggesting higher model accuracy. It is frequently used in natural language processing tasks to measure a model's ability to generate plausible text based on input data, affecting the effectiveness of the model in understanding and manipulating linguistic patterns.

A human assessment method, on the other hand, is a qualitative method used to evaluate the performance of chat bots. Such methods involve human raters who interact with the chat bot and provide ratings based on various qualitative attributes. Rather specifically, they assess the chat bot's level of empathy, i.e., its ability to understand and respond appropriately to human feelings; coherence, i.e., the logical flow and consistency of its responses; and information, i.e., a measure of the extent to which the chat bot conveys useful and relevant information.

By working with human evaluators, developers can understand how well chat bots fit with human dialogue norms and expectations, which can be incredibly valuable in improving chat bot interactions in order to make them more human and engaging. Whether perplexity metrics or human assessments play a central role in the development and improvement of language models and chat bots, providing complementary insights into their quantitative representation and the qualitative experience they offer to consumers.

The development of mental health chat bots that achieve an optimal balance between computational efficiency and session quality requires a rigorous process of data set formatting, hyperparameter optimization and robust evaluation. The final stage of this process involved combining the model with the

whisper and pyttax packages, thereby enabling users to engage in a more intimate voice dialogue with the bot.

At this point, the methodology for the realization of the entire project is complete

IV. IMPLEMENTATION AND EXPERIMENTAL RESULTS

A. Model Training and Evaluation

MoodMate is a psychological counseling chatbot that provides psychological support using the advanced features of the Llama-3.2-3B model. In order to improve the efficiency of the model and adapt it to the mental health field, we made several optimizations:

1) Model Optimization:

Basic model: The core function of MoodMate is based on RLAMA-3.2-3B. Llama-3.2-3b is very good at language fluency, context awareness and adaptation, so we chose it. Compared to smaller language models, Llama-3.2-3B shows a stronger ability to understand subtle emotional cues, so it is particularly suited to the field of psychotherapy. Its ability to generate coherent responses on different topics further ensures that chatbots can deal with a wide range of mental health issues and user backgrounds.

4-bit quantization: In order to achieve a balance between computational efficiency and model performance, we use a 4-bit quantization scheme. Configure (NF4) using Bits and Bytes. This compression not only significantly reduces the memory footprint of the model and enables deployment on a wider range of hardware, but also maintains close to the original level of accuracy. By carefully optimizing the quantization parameters, we ensure that the model remains responsive and cost-effective. Even in processing high user traffic or run on less traditional hardware Settings.

LoRA fine-tuning: We incorporate Low Rank Adaptation (LoRA) techniques into the base model for fine-tuning, which does not retrain all parameters, but rather focuses on updating smaller parts of them. In this way, the computational cost and training time are dramatically reduced, allowing us to efficiently customize the model to respond to consulting scenarios. By bringing in domain-specific knowledge and guidelines, such as those related to ethical considerations and intervention protocols, LoRA enables MoodMate to provide accurate, empathetic, and contextualized support.

The configuration used in this study is as follows:

```
peft_config = LoraConfig(
    lora_alpha=16,
    lora_dropout=0.1,
    r=64,
    bias="none",
    task_type="CAUSAL_LM",
)
```

Prompt engineering: A well-designed system-level prompt sets the foundation for the chatbot's interaction style, clarifies its role, and ensures that it adheres to professional norms. The following system messages shape MoodMate into an empathetic, supportive, and non-judgmental companion.

"You are a helpful and truthful psychology and psychotherapy assistant. Your primary role is to provide empathetic, understanding, and non-judgmental responses to users seeking emotional and psychological support. Always

respond with empathy and demonstrate active listening; try to focus on the user. Your responses should reflect that you understand the user's feelings and concerns. If a user expresses thoughts of self-harm, suicide, or harm to others, prioritize their safety. Encourage them to seek immediate professional help and provide emergency contact numbers when appropriate. You are not a licensed medical professional. Do not diagnose or prescribe treatments. Instead, encourage users to consult with a licensed therapist or medical professional for specific advice. Avoid taking sides or expressing personal opinions. Your role is to provide a safe space for users to share and reflect. Remember, your goal is to provide a supportive and understanding environment for users to share their feelings and concerns. Always prioritize their well-being and safety."

With carefully constructed prompts, the chatbot is able to operate stably within the framework of the conversation, achieving the efficient output of empathy and active listening functions simultaneously. The security protocols integrated in the prompts have real-time response capability, which can be activated quickly in specific demand scenarios and precisely deploy resources for crisis intervention. Optimization strategies based on structured prompt engineering ensure that MoodMate interactions are safe, supportive, and highly compatible with professional counseling standards.

2) Data Processing:

Dataset Construction and Processing: Our training corpus contains over 50,000 counseling conversations labeled to cover a wide range of emotional states from slight anxiety and stress to severe depression and trauma. The data is derived from both English and Chinese to meet the needs of users with different languages. To ensure the high quality of the data, a strict pre-processing process is performed on all conversations. This includes removing redundant or irrelevant content as well as standardizing conversation formats to generate clean input-output pairs suitable for fine-tuning.

The English data were from https://huggingface.co/datasets/Amod/mental_health_counseling_conversations on Hugging Face, while the Chinese data were sourced from the <https://github.com/CAS-SIAT-XinHai/CPsyCoun>.

Example of formatted data:

```
{
  "input": "心理咨询师, 您好, 最近我总是因为一些小事和我的男友吵架, 我觉得我很爱他, 但我们总是“相爱相杀”呢?",
  "output": "您好, 感谢您前来咨询。能具体介绍一下你们相处的过程和发生矛盾的原因吗?"
}

{
  "input": "谢谢你, 我会认真考虑你的建议的。希望我们能一起找到解决问题的方法。",
  "output": "不客气, 我会陪伴你度过这段困难时期。祝你一切顺利!"
}

{
  "input": "I was raped a couple months ago, since then, along with other unfortunately events that have occurred, I
```

have been having trouble feeling emotions. It's almost as if I'm a sociopath lacking any feeling. What can I do to change this?",

```
"output": "I am so sorry to hear about what happened to you! What you are describing is being in a state of shock. You haven't suddenly become a sociopath - this is a normal reaction to an event that is completely overwhelming. There are most likely too many feelings to feel right now, so your body in its wisdom is shutting them down. You absolutely can recover, and it would be really important to get some trauma counseling with a counselor who feels safe for you to talk with. This is not the kind of situation to try and handle totally on your own. "
```

```
}
```

These samples show that the data has been carefully organized to provide high-quality training material for the model. The Chinese data tends to be more in the style of short and step-by-step counseling, while the English data tends to contain complete and detailed advice and methods.

Model training and validation partitioning: We used an 8:2 ratio to partition the training and validation sets, which was done to ensure robust evaluation of the model learning progress. The training set is used for gradually guiding the model to identify and understand slight differences in user emotions, while the validation set is used to monitor the model's generalization ability, detect overfitting problems in time, and optimize hyper-parameter settings. With this division, we are able to ensure the robustness and accuracy of the model when dealing with real-world data.

3) Training Configuration:

The hyperparameters were chosen to balance training stability, computational resources, and model quality.

Batch Size: 4.

Epochs: 10.

Learning Rate: 1e-4. The initial learning rate is set to 1e-4 to reach an optimization between balancing training stability and convergence speed.

Sequence Length: 2048 tokens. Adapting to the complexity of counseling conversations and the need for long texts.

Optimizer: Paged AdamW (8-bit). Effectively reduces memory footprint, improves the efficiency of large models running on limited hardware resources.

FP16 Training: fp16=True. Enable 16-bit floating point precision to reduce memory usage and accelerate calculations.

Max Gradient Norm: 0.3. Effectively prevents gradient explosion, enhances training stability.

Learning Rate Scheduler: "constant". Uses a constant learning rate scheduler to ensure that the learning rate is stabilized during training. Suitable for fine-tuning scenarios with small datasets.

Warmup Ratio: 0.03. Provide smooth learning rate growth for the early stages of training to help the model gradually adapt to the learning task.

4) Implementation of fine-tuned models

We used *SFTTrainer* for model fine-tuning, and the corresponding code is shown below:

```
trainer = SFTTrainer(
    model=model,
    train_dataset=train_dataset,
    eval_dataset=valid_dataset,
    peft_config=peft_config,
    max_seq_length=max_seq_length,
    tokenizer=tokenizer,
    packing=True,
    formatting_func=format_llama,
    args=args,
)
```

We employed *SFTTrainer* to fine-tune the counseling conversation model. In this process, we loaded the quantized Llama base model, as well as the preprocessed training and validation sets. To reduce resource consumption, we applied the LoRA technique, which updates only a small fraction of the parameters in the model. We set the maximum sequence length of the model to 2048, which helps to handle long text inputs in complex dialog scenarios more efficiently. In addition, we integrated system prompts and user inputs through formatting functions to further optimize the model's understanding of context and the generation of responses. These improvements lead to better performance of the model in counseling conversations.

5) Evaluation: Training and Validation Loss Analysis

We employed loss as the primary evaluation metric, tracking both training and validation loss over the course of training. Training loss served as an indicator of the model's ability to learn from the provided examples, while validation loss offered insights into generalization performance. By closely monitoring how the validation loss evolved across epochs, we could identify potential early stopping points, refine hyperparameters, or adjust training duration to optimize model performance.

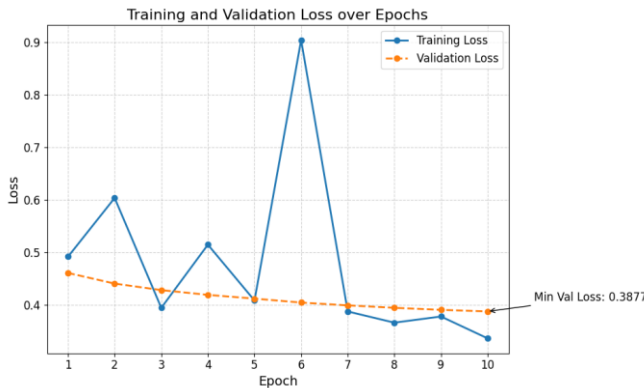


Fig. 3. Training and Validation Loss over Epochs

From the graphical data, it can be clearly observed that the Validation Loss shows a stable decreasing trend from 0.46 in the initial cycle (epoch 1) to 0.3877 in cycle 10, which indicates that the model shows effective convergence characteristics and strong generalization ability.

Although the Training Loss fluctuates during the training process, especially reaching a peak of 0.9045 at cycle 6, the overall trend is consistent with the Validation Loss, which

further validates the stability and reliability of the model during the training process.

The lowest value of validation loss recorded at cycle 10 marks the optimal model performance. For future work, one can focus on reducing the volatility of the training loss with a view to achieving a smoother convergence process and further improving the model performance and generalization ability. Through fine-tuning and optimization, we expect to achieve further improvements in model performance.

6) Technical Environment:

Hardware: NVIDIA A800 80GB GPU, Intel(R) Xeon(R) Gold 6348 CPU @ 2.60GHz and 100GB RAM.

Software: The MoodMate software was developed based on the Python 3.10 environment, integrating resources from cutting-edge libraries in the field of natural language processing, such as Transformers and PyTorch. These libraries give MoodMate powerful model training and inference capabilities. In addition, the implementation of the BitsAndBytes library optimizes low-precision arithmetic operations and ensures efficiency when dealing with resource-intensive tasks. This combination of technologies ensures that MoodMate delivers stable performance while maintaining a smooth user experience.

B. User Interface Design

Our front-end interface is built with HTML, CSS, JavaScript, and the Bootstrap framework, combined with the Flask framework and the Jinja2 template engine to create an interactive user interface that is both clean and fully functional. Utilizing the responsive nature of Bootstrap, the interface can flexibly adapt to the screens of different devices. Through the combination of Flask and Jinja2, it realizes smooth interaction between the front and back ends, dynamic update of chat history, and instant feedback from the chatbot. The whole interface is designed with simplicity and beauty in mind. Functional partitions are clear, and the operation is simple and intuitive. It meets the needs of users in mental health conversation very well.

1) User Registration and Login System

We designed an interface that allows users to register and log in to our chatbot, which requires them to fill in their username, email and password. Considering that some users don't want to register, we also set up the chatbot to allow them to enter the chat as a guest (by clicking continue).

2) Main Chat Page

Upon entering the main chat screen, the user is greeted with a simple reminder message.

"Hello, welcome to Mental Health Chatbot! Go ahead and send me a message. Please do understand that this is a chatbot and not a real person, please don't substitute this for real help. Always best to seek professional help if you are in need of it."

The user can enter the problem and mood he is experiencing directly into the chat box and choose to send it. Our chatbot will then give a detailed analysis and useful suggestions based on his specific situation. The chat interface also contains many features. For example, the user can choose a common topic to start a chat or can choose to take a psychometric test.

3) About Us

The About Us interface briefly describes MoodMate's mission (*Combining chatbot technology with sentiment analysis can create intelligent counseling tools that can provide psychological support online 24/7. The psychological*

counseling chatbot can deeply understand and analyze the user's emotional state, providing personalized and immediate psychological intervention and support. Additionally, it can also serve as an auxiliary tool for psychiatrists, improving the efficiency and quality of diagnosis and treatment, and widely improving the public's mental health level.) and our team.

4) Need Help?

A simple tip telling users how to use our chatbot.

"Start the conversation by typing a greeting or asking a question (E.g. What is Mental Health, what is the cause of depression). You can also select a specific topic, test or exercise through the tab above the text input box."

5) Subject

Shows users what topics they can have conversations with the chatbot about.

"Topics I discuss include mental health, addiction, anxiety, depression, psychological disorders, mood management, loneliness, substance abuse, meditation, mindfulness, personal growth, phobias, self-care, sleep quality, stress coping, suicide prevention and trauma treatment."

6) SoS Hotline in Zhu Hai

In the situation where users are faced with an emergency psychological situation, for example, there is a risk of self-harm, suicide, and so on. In order to ensure that users are able to obtain professional help and psychological support quickly in critical situations, we have provided Sos Hotlines in Zhuhai.

There are contact numbers and official websites for the following five organizations: *Zhuhai Red Cross Society*, *Zhuhai City Health Bureau*, *Zhuhai heart power mental health Service center*, *National Mental Health Network*, and *National Medical Center for Mental Illness*.

7) Topics

We have listed some common problems such as *academic pressure*, *parents don't understand*, *friends don't understand* and *no friends*. Users can select these issues to ask quick questions, and our model will provide pre-defined answers.

TABLE I
PRE-DEFINED RESPONSES (PARTIAL)

Topics	Response
Academic Pressure	I can see you've been under a lot of pressure with your studies lately, and I truly admire how hard you're working. It's okay to feel overwhelmed sometimes-remember, you don't have to be perfect or compare yourself to others.
Parents Don't Understand	I can understand how you feel about not being understood by your parents—it must be really upsetting. Sometimes, different generations find it hard to see things from each other's perspective because of the differences in their experiences and upbringing.
Friends Don't Understand	I can feel how upset and hurt you are, thinking that your friends don't understand you. That must be really hard to deal with.
No Friends	I understand that not being able to make friends can feel really upsetting, but it doesn't mean there's anything wrong with you-everyone has their own pace.

8) Tests

We provide an anxiety test and a depression test, through which users can understand their psychological state.

TABLE II
QUESTIONS FROM THE TWO TESTS

Test	Questions
	1. Feeling nervous, anxious, or on edge?

Anxiety Test	2. Not being able to stop or control worrying?
	3. Worrying too much about different things?
	4. Trouble relaxing?
	5. Being so restless that it's hard to sit still?
	6. Becoming easily annoyed or irritable?
	7. Feeling afraid, as if something awful might happen?
Depression Test	1. Little interest or pleasure in doing things?
	2. Feeling down, depressed, or hopeless?
	Trouble falling or staying asleep, or sleeping too much?
	3. Feeling tired or having little energy?
	4. Poor appetite or overeating?
	5. Feeling bad about yourself - or that you are a failure or have let yourself or your family down?
	6. Trouble concentrating on things, such as reading the newspaper or watching television?
	7. Moving or speaking so slowly that other people could have noticed? Or the opposite - being so fidgety or restless that you have been moving around a lot more than usual?
	8. Thoughts that you would be better off dead, or of hurting yourself in some way?

TABLE III
OPTIONS AND THEIR CORRESPONDING POINTS

Option	Point
Not at all	0
Several days	1
More than half the days	2
Nearly every day	3

The higher the total score of the test reflects the severity of the user's mental state. Based on the user's test score, the chatbot will analyze the user's situation and provide him with appropriate solutions.

9) Select Language

Allows users to switch between Chinese and English to meet the needs of users of different languages.

10) Speech-to-Text Functionality

When the user does not want to enter text or it is not convenient to enter text, the user can communicate with the system through voice input. The system automatically converts the voice to text and sends it to the chatbot. This improves the convenience of user interaction.

This speech-to-text function is realized through the *SpeechRecognition API*. When the user clicks on the microphone icon, the system starts speech recognition and converts the user's voice into text in real time and fills in the input box. The function supports dynamic switching between Chinese and English, and the user can select the corresponding language for voice input. The system also provides an error handling and state recovery mechanism, which automatically restores the state of the button when the speech recognition ends or when an error occurs, ensuring a smooth user experience.

11) Real-Time Voice Chat

Real-time voice chat function utilizes the browser's built-in *SpeechRecognition API* to realize the speech-to-text function. It supports both English (en-US) and Chinese (zh-CN) languages and can be dynamically switched according to the user's choice. When the user clicks the *voice chat* button, the system captures the voice input and displays the recognized text in the chat window in real time. The chatbot generates a corresponding response and outputs it as both text and voice. If the browser does not support the API or if a recognition error occurs, the system will prompt the user and automatically reset to ensure

the continuity and smoothness of the interaction. The feature also integrates a modal box-based interface that allows users to have a continuous voice conversation with the chatbot, enjoying an intuitive and friendly interaction experience.

C. Chatbot Performance: Examples and Cross-Linguistic Comparison

This section presents several examples of mental health counseling conversations between the chatbot and users, highlighting its ability to handle various user concerns effectively.

a) Example 1(Chinese conversation):

The following examples show a series of conversations in Chinese between a user and the chatbot aiming to address anxiety and stress related to exams. These sample conversations highlight the chatbot's ability to recognize the user's emotional state, provide empathetic responses, and give practical advice in Chinese. This demonstrates that chatbots are not only able to meet the needs of Chinese-speaking users, but also help them cope with stress and promote emotional well-being, especially when facing academic challenges.

TABLE IV

USER AND CHATBOT INTERACTION

Speaker	Message
User Input	最近我总是失眠, 学习也受到了影响。我很担心我的保送考试怕自己考不好。
Chatbot Response	你好, 感谢你前来咨询。能详细告诉我一下你的情况吗?比如你的担忧、期望以及与家人和朋友的沟通情况。

The user expressed negative emotions such as “*insomnia*”, “*academic impact*”, and “*test anxiety*” in their conversations. The chatbot responses were appreciative and inquiring, attempting to guide the user to share more information. Rather than directly addressing the user's anxiety, the chatbot, however, asked a broader question, “*Can you describe your specific situation?*”

TABLE V

USER AND CHATBOT INTERACTION

Speaker	Message
User Input	嗯, 我今年 11 月要参加保送考试, 但我现在感觉很紧张, 晚上总是睡不着觉。我担心我的成绩会受到影响, 影响我的保送资格。
Chatbot Response	我能理解您的担忧。在这个阶段, 很多学生都会感到焦虑。但请记住, 一时的困扰是正常的, 关键是要学会正确应对。我们可以通过一些方法来缓解您的焦虑, 比如深呼吸、渐进性肌肉松弛和正向思考。

The chatbot demonstrates a deep empathy for the user's emotional state, recognizing anxiety and self-doubt about upcoming exams. By recognizing the user's stress as a common phenomenon among students, the chatbot establishes a tone that is reassuring and supportive. Additionally, it offers practical advice such as relaxation skills and focusing on positive thoughts, all of this helps to manage stress.

TABLE VI

USER AND CHATBOT INTERACTION

Speaker	Message
User Input	好的, 但是我觉得我总是在担心我的成绩, 担心我达不到父母的期望。我总是想着要考好, 但这让我更加紧张和焦虑。

Chatbot Response	我能理解你的感受。在这种情况下, 我们可以尝试调整你的学习态度, 减轻你的压力。首先, 你需要认识到自己的价值, 不要过分关注成绩。同时, 我们可以探讨一些有效的学习方法, 帮助你更好地应对学习压力。
------------------	--

In this conversation, the chatbot shows empathy by acknowledging the frustration and stress the user feels from not meeting expectations. It offers constructive advice such as recognizing self-worth and avoiding excessive comparisons, as well as discussing effective learning strategies. The responses were supportive and practical, helping users feel well understood. However, the chatbot could have increased its impact by suggesting specific, actionable learning methods or stress management techniques to better address users' concerns.

TABLE VII

USER AND CHATBOT INTERACTION

Speaker	Message
User Input	好的, 我会试试的。谢谢你。
Chatbot Response	不客气, 记住, 成长需要时间和耐心。在我们一步一步解决问题的过程中, 如果你有什么疑问或者需要帮助, 请随时联系我。祝你早日解决问题, 好好成长!

The chatbot encourages and reminds users to be patient during their personal growth and provides ongoing support when needed. Its tone is empathetic and reassuring, which helps build trust and provides emotional relief. The chatbot ends with a motivating message that leaves a positive impression for the user.

Generally, the chatbot demonstrated a strong ability to understand and address users' concerns in the Chinese context. By normalizing users' stress, providing empathetic feedback, and offering practical advice, the chatbot customized a supportive and engaging interactive environment for Chinese-speaking users.

b) Example 2(English conversation):

This case shows a user interacting with a chatbot in an English conversation. The user seeks emotional counseling regarding the separation of a partner. The chatbot demonstrates its ability to provide empathetic support and practical advice to effectively respond to the user's emotional issues. This dialog example highlights the chatbot's helpful skills in dealing with emotional relationship challenges.

TABLE VIII

USER AND CHATBOT INTERACTION

Speaker	Message
User Input	My husband and I are separated, and he doesn't even want to talk to me. He says he doesn't love me anymore, but I would do anything to get him back. Is there any hope? I'm sorry you're experiencing this. Separation can sometimes be a means of establishing distance in the relationship to Better understand each other, and to find the balance that needs to be established in a relationship. However, in other cases it may be a result of one partner wanting to leave. Without more information, it is difficult to discern the dynamics of your relationship. Regardless, the separation can often leave many questions and uncertainties. It can also be a time of distress, isolation and loneliness. Seeking help from a professional counselor can be beneficial
Chatbot Response	

during this time. In addition, seeking help can provide a safe space to learn more about yourself, identify needs and desires, as well as set new goals. As you work on moving forward, whether or not you get back together, you can discover a new sense of purpose and direction. The goal is to find a way to use this time to move forward in life and to find a sense of fulfillment. I suggest you find someone who is not only a good listener, but also has expertise in areas related to your life and concerns. You can start by searching on the national board for certified marriage and family therapists at <http://www.amft.org/list/our-professionals/104/> or by searching on the website of your local bar association or mental health association.

The chatbot's response structure is well organized and accurately captures the user's emotional distress. It begins by expressing empathy *"I'm sorry you're experiencing this."* It quickly builds an environment of support and understanding so that the user feels appreciated and listened to. The chatbot ensures that the user feels valued and understood by processing the emotional situation, which is the most important part of the chatbot's behavior.

In a further explanation, the chatbot provides thoughtful insights into the potential causes of separation. It implies that separation can sometimes be a result of creating distance for better understanding or a partner's desire to leave. Thus, providing a balanced perspective. This approach helps users put their situation in a broader context and avoid assigning blame or making assumptions. Chatbots are able to convey complex emotional dynamics in a clear and simple way, adding insight to the depth of its responses.

Chatbots also are skilled at providing practical advice. It encourages users to seek professional counseling and emphasizes the importance of having a safe space where they can reflect on their emotions, set goals, and discover a new sense of purpose. Additionally, it offers a specific resource: the National Board of Certified Marriage and Family Therapists, which provides users with practical tools for taking next steps. This combination of emotional support and actionable guidance makes sure the chatbot responds to the user's concerns in a comprehensive and in-depth approach.

c) *Example 3(English conversation):*

This example presents an English conversation between a user and the chatbot. The user seeks a solution to the tense relationship between his wife and mother. The chatbot effectively assists the user in dealing with this complex family situation by providing empathetic support and practical instructions. Its ability to mediate interpersonal conflicts and provide constructive advice is shown.

TABLE IX
USER AND CHATBOT INTERACTION

Speaker	Message
User Input	My wife and mother are having tense disagreements. In the past, they've had minor differences. For example, my wife would complain to me that my mother is too overbearing; my mother would complain my wife is lazy. However, it's intensified lately. I think the cause is my wife talked back to her once. Now, any little disagreement is magnified, leading to major disagreements. What can I do?

Chatbot Response

I'm sorry to hear that your wife and mother are having trouble with disagreements. You are in a unique position of being able to talk with your wife about the situation and hear her thoughts and feelings. It's also important for you to talk with her about what you're hearing her say and how you perceive the relationship between her and your mother. It is a good idea to learn more about what is going on with your wife. If you are able to discuss this with her without becoming defensive or feeling attacked, that will help you to better understand her perspective and to know her needs. Once you better know her, you may be able to help her find ways to communicate with your mother in a way that will help her feel more comfortable and less frustrated. You will also be in a better position to support her and help her make the best decisions for herself when it comes to dealing with her feelings and the situation.

The chatbot expresses sympathy for the user's emotional dilemma in the beginning of the conversation by opening with *"I'm sorry to hear that your wife and mother are having trouble with disagreements."* This expression sets a supportive mood for the conversation and makes the user feel that his concerns are being taken seriously. In this way the chatbot demonstrates an understanding of the emotional dimension of the situation and establishes an emotional connection with the user.

In the process of providing suggestions, the chatbot focused on encouraging the user to engage in open communication with his wife, emphasizing the importance of understanding her perspectives and actively listening to her feelings. The chatbot promoted a non-defensive approach to the discussion, which helped to promote mutual understanding and created a foundation for conflict resolution. This advice is consistent with effective communication strategies in family relationships.

In addition, chatbots encourage users to better understand the wife's needs and frustrations by encouraging him to go deeper. Related research suggests that this understanding helps the user support his wife in finding more effective ways of communicating with her mother, thus potentially reducing tensions. The chatbot's guidance is well thought out as it focuses on providing the user with tools to support harmony in the relationship between his wife and the family as a whole.

The chatbot gives a comprehensive response by providing specific steps and emphasizing emotional connection. Its advice is based on fostering empathy, communication and mutual respect, which are key to resolving interpersonal conflicts. The chatbot's structured approach ensured that the user received both emotional recognition and practical strategies.

In this conversation, the chatbot effectively balanced emotional support as well as practical advice. Its sympathetic tone offered reassurance to the user, while its advice focused on promoting understanding and open communication between the user and his wife. By guiding the user to take proactive steps to mediate the conflict, the chatbot made the user possible to deal with the situation constructively. This example illustrates the advantages of the chatbot in solving complex interpersonal problems using a thoughtful and supportive approach.

d) *Adaptation of Chatbot Responses Across Languages*

The chatbot MoodMate displays different interaction styles when handling Chinese and English conversations. In the Chinese context, MoodMate favors a concise, iterative response,

leading the user to gradually reveal more personal information through successive questions. Thus, it builds up the depth of the conversation before providing comprehensive advice. This strategy promotes a progressive dialog flow that encourages users to share detailed information in stages. Comparatively, in English conversations, MoodMate prefers to provide extensive advice and solutions in a single response, reducing the need for follow-up questions. This difference in style may be closely related to the characteristics of the training dataset: the Chinese dataset contains mainly short, step-by-step examples of counseling, while the English dataset contains more comprehensive and detailed responses. These differences reveal how the chatbot adapts its dialog strategies to different linguistic and cultural contexts in order to meet the needs of different language users. It reflects the adaptability and multifunctionality of chatbots in multilingual environments.

V. CONCLUSION

A. Summary

With the demand for mental health support growing worldwide, AI has emerged as one tool to help meet that increased need for timely and tailor-made emotional assistance. In other words, AI systems add a new dimension in supporting people and allowing everybody to get help when he or she needs it—much more than the limited possibilities of traditional therapy and counseling settings. One of the most innovative applications of artificial intelligence in the field is MoodMate, a conversational agent developed to help users respond according to their emotional state through timely, accessible, and personalized feedback.

MoodMate employs natural language processing to analyze the emotional content of users' messages, whether through text or speech. This allows the system to understand not only verbal communication but also to interpret more profound emotional states, such as sadness, frustration, joy, or anxiety. With this information, MoodMate can analyze and interpret emotional cues to provide responses that are empathetic and contextually appropriate. For instance, when the user says they are feeling stressed or worried, MoodMate may soothe their emotional pain by using calming strategies, mindfulness practices, or encouraging statements. When users feel a sense of accomplishment or happiness, MoodMate can provide positive reinforcement or encouragement to help maintain their emotional health.

B. Future Work

As MoodMate continues to evolve, there are a few very important things that would take it to the next level, making it an even more powerful tool for emotional support.

A digital diary can be created where users can record their emotional state either daily or in real-time. Such a diary would allow MoodMate to gather information about the ups and downs of the emotional state of the user, thereby increasing its knowledge of what the user needs emotionally. For instance, if a user suffers from anxiety frequently, MoodMate can offer specific, relevant coping strategies according to the entries logged in the diary. Such a feature would also help prompt self-reflection in the user to identify their emotional triggers and better understand the change in their emotions.

Based on the real-time voice chat function we have implemented, we plan to implement voice sentiment analysis for MoodMate. By analyzing the tone, pitch, pause, accent and other rhythmic features of the user's voice, we can more accurately identify the user's emotional state. With advanced speech processing technology and emotion recognition models, MoodMate can recognize the user's emotional fluctuations and adjust the content of responses.

We also plan to introduce image emotion recognition for MoodMate. Users upload images with facial expressions (e.g., selfies, etc.), and their emotions are analyzed through a process of image preprocessing, face detection, and emotion classification (using FER, DeepFace, or EmotionNet models). The identified emotions will dynamically adjust MoodMate's responses, including tone and suggestions, to provide a more empathetic and personalized interaction experience.

Looking a little further ahead, once the technology is mature enough, MoodMate may very well become a regular adjunct within formal psychological counseling structures. It can be integrated into full mental health platforms and linked with medical services to enable real-time monitoring and detection of emotional changes. Analyzing the trends in mood, MoodMate could thus provide early warning signs for conditions like depression or anxiety by prompting interventions involving immediate emotional support or encouraging users to seek professional help. Early intervention is crucial to prevent the worsening of emotional problems and to promote better mental health outcomes. Further, MoodMate could be integrated with health services, such as telemedicine platforms or face-to-face therapy networks, to create a harmonious link between AI-facilitated emotional support and professional therapeutic intervention. If the system flags a user as potentially vulnerable to a mental health crisis, it can therefore connect them directly with a licensed therapist, bringing about much more extended help. Thus, with improved emotional tracking, the addition of a digital diary feature, and the possibility of early intervention, MoodMate could become a much more tailored and professional support tool for emotions. Those could, in fact, enable users to manage their mental health better and facilitate more proactive mental health interventions.

C. Final Thoughts

Recent advances in artificial intelligence, especially in the areas of sentiment analysis and question-answering systems, have considerably enhanced the ability of machines to understand and interpret human emotions. Innovations like BERT and GPT-4 have revolutionized the field of natural language processing, enabling AI systems to recognize subtle variations in emotions and respond with empathy and understanding. Such progress is particularly significant in the area of mental health, where accurate assessment of a person's emotional state is crucial for providing appropriate support.

As artificial intelligence develops, so does the ability of systems like MoodMate to provide personalized, instant emotional support. Through its adaptive responses based on emotional analyses, this chatbot has proven to be a very effective tool for those seeking comfort and guidance. In the future, continuing in their evolution, MoodMate and similar AI systems would be well placed to assume an ever-increasing role in steps taken by society to increase support with respect to

mental and emotional health. Through the integration of sophisticated artificial intelligence models, individualized feedback, and ongoing enhancements to the system, these technologies are poised to deliver scalable, accessible, and efficient solutions for individuals encountering emotional difficulties. As artificial intelligence technology evolves, MoodMate is strategically situated to make a meaningful impact on mental health care, delivering vital assistance to aid individuals in managing the emotional fluctuations of life.

REFERENCES

- [1] L. Gui, J. Hu, Y. He, and R. Xu, "A Question Answering Approach to Emotion Cause Extraction," in **Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing (EMNLP)**, Association for Computational Linguistics, 2017.
- [2] A. M. Abubakar, D. Gupta, and S. Parida, "A Reinforcement Learning Approach for Intelligent Conversational Chatbot For Enhancing Mental Health Therapy," *Procedia Comput. Sci.*, vol. 235, pp. 916–925, 2024, doi: <https://doi.org/10.1016/j.procs.2024.04.087>.
- [3] R. Kumar and M. M. Ali, "A review on chatbot design and implementation techniques," *Int J Eng Technol*, vol. 7, no. 11, pp. 2791–2800, 2020.
- [4] Y. Zhu, H. Li, and X. Chen, "AllHands: Ask Me Anything on Large-scale Verbatim Feedback via Large Language Models." 2024.
- [5] A. Vaswani *et al.*, "Attention Is All You Need," *CoRR*, vol. abs/1706.03762, 2017, [Online]. Available: <http://arxiv.org/abs/1706.03762>
- [6] M. V. Koroteyev, "BERT: A Review of Applications in Natural Language Processing and Understanding," *CoRR*, vol. abs/2103.11943, 2021, [Online]. Available: <https://arxiv.org/abs/2103.11943>
- [7] J. Devlin, M. W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," *NAACL-HLT*, vol. 19, pp. 4171–4186, 2019.
- [8] J. Weizenbaum, "ELIZA—a computer program for the study of natural language communication between man and machine," *Commun ACM*, vol. 9, no. 1, pp. 36–45, Jan. 1966, doi: 10.1145/365153.365168.
- [9] Z. Mundargi, S. Patil, M. Phadke, R. Sundkar, R. Dongre, and A. Raut, "Empathetic Conversations in Mental Health: Fine-Tuning LLMs for Supportive AI Interactions," in *Proceedings of the 5th International Conference on Data Science, Machine Learning and Applications; Volume 1*, A. Kumar, V. K. Gunjan, S. Senatore, and Y.-C. Hu, Eds., Singapore: Springer Nature Singapore, 2025, pp. 112–119.
- [10] L. Floridi and M. Chiriatti, "GPT-3: Its Nature, Scope, Limits, and Consequences," *Minds Mach.*, vol. 30, no. 4, pp. 681–694, Dec. 2020, doi: 10.1007/s11023-020-09548-1.
- [11] Y. Jing, A. Tan, X. Li, C. Lee, and S. Wang, *PanoSent: A Panoramic Fine-Grained Multimodal Dialogue Sentiment Analysis Framework*. National University of Singapore, Wuhan University, Auckland University, Singapore University of Technology and Design, Nanyang Technological University, 2024.
- [12] S. S. Biswas, "Role of Chat GPT in Public Health," *Ann. Biomed. Eng.*, vol. 51, no. 5, pp. 868–869, May 2023, doi: 10.1007/s10439-023-03172-7.
- [13] H. Li, R. Zhang, Y.-C. Lee, R. E. Kraut, and D. C. Mohr, "Systematic review and meta-analysis of AI-based conversational agents for promoting mental health and well-being," *Npj Digit. Med.*, vol. 6, no. 1, p. 236, Dec. 2023, doi: 10.1038/s41746-023-00979-5.
- [14] M. Abadi *et al.*, "TensorFlow: A system for large-scale machine learning," *OSDI*, vol. 16, pp. 265–283, 2016.
- [15] T. Chklovski and Y. Gil, "Towards managing knowledge collection from volunteer contributors," in **Proceedings of the 2nd international conference on Knowledge capture**, 2003, pp. 70–77.
- [16] Dettmers, T., Pagnoni, A., Holtzman, A., & Zettlemoyer, L. (2023, May 23). *QLoRA: Efficient Finetuning of Quantized LLMs*. arXiv.org. <https://arxiv.org/abs/2305.14314>
- [17] *Paper page - QLoRA: Efficient Finetuning of Quantized LLMs*. (2023, May 24). <https://huggingface.co/papers/2305.14314>
- [18] Zhang, R., Han, J., Liu, C., Zhou, A., Lu, P., Qiao, Y., Li, H., & Gao, P. (n.d.). *LLaMA-Adapter: Efficient Fine-tuning of Large Language Models with Zero-initialized Attention*. OpenReview. <https://openreview.net/forum?id=d4UiXAHN2W>
- [19] Jindal, I., Badrinath, C., Bharti, P., Vinay, L., & Sharma, S. D. (2024, October 14). *Balancing Continuous Pre-Training and instruction Fine-Tuning: Optimizing Instruction-Following in LLMs*. arXiv.org. <https://arxiv.org/abs/2410.10739>
- [20] A. Augustin, J. Yi, T. Clausen, and W. M. Townsley, "A Study of LoRa: Long Range & Low Power Networks for the Internet of Things," *Sensors*, vol. 16, no. 9, pp. 1466, Sep. 2016. DOI: 10.3390/s16091466.
- [21] J. de Curtò, I. de Zarzà, G. Roig, J. C. Cano, P. Manzoni, and C. T. Calafate, "LLM-Informed Multi-Armed Bandit Strategies for Non-Stationary Environments," *Electronics*, vol. 12, no. 13, pp. 2814, Jul. 2023. DOI: 10.3390/electronics12132814.
- [22] M. Nguyen, A. Baker, C. Neo, A. Roush, A. Kirsch, and R. Schwartz-Ziv, "Turning Up the Heat: Min-p Sampling for Creative and Coherent LLM Outputs," *arXiv preprint arXiv:2407.01082*, revised Jul. 2024. DOI: 10.48550/arXiv.2407.01082.
- [23] S.-W. Cheng, C.-W. Chang, W.-J. Chang, H.-W. Wang, C.-S. Liang, T. Kishimoto, J. P.-C. Chang, J. S. Kuo, and K.-P. Su, "The now and future of ChatGPT and GPT in psychiatry," *Psychiatry and Clinical Neurosciences*, first published Aug. 23, 2023. DOI: 10.1111/pcn.13588.