

Ensemble Learning in LLMs for Enhanced Mental Health Conversations

Submitted by

B PREM TEJESWAR REDDY

Enrolment No. 21116027

KUNURU ABHISHEK

Enrolment No. 21116051

Under the Supervision of

Assistant prof

B THARUNKUMAR REDDY



INDIAN INSTITUTE OF TECHNOLOGY, ROORKEE
ROORKEE-247 667 (INDIA)

Abstract

Mental health issues such as depression, anxiety, and stress are prevalent, particularly among students who may lack the financial resources or willingness to seek professional help. This project proposes an interactive chatbot designed to provide mental health support akin to that of a counselor or therapist. Utilizing advanced Large Language Models (LLMs) like LLaMA, DialogGPT, and BERT, the chatbot is fine-tuned for contextual understanding and tailored responses. It offers initial diagnoses and assessments based on user interactions while incorporating principles of Cognitive Behavioral Therapy (CBT), specifically the ABC model, to ensure constructive and supportive feedback. A Python-based NLP model enhances the chatbot's accuracy and authenticity, aiming to create an accessible online platform that effectively addresses mental health concerns without the barriers of traditional therapy.

1. Introduction

In an era where mental health challenges are increasingly recognized, the demand for accessible and effective support systems has never been greater. Many individuals, particularly students, face significant barriers when seeking professional help due to factors such as financial constraints, limited availability of services, and societal stigma surrounding mental

health issues. To address these challenges, this project introduces an AI-powered chatbot designed to provide empathetic mental health support and information.

This chatbot harnesses the capabilities of advanced Large Language Models (LLMs), including DialogGPT and BERT, in addition to MentaLLaMA-chat-7B. DialogGPT is particularly adept at generating conversational responses, making it suitable for interactive dialogues. BERT, on the other hand, excels in understanding the context of user queries, ensuring that the chatbot can interpret and respond to a wide range of mental health-related questions effectively.

The integration of these models allows the chatbot to engage users in meaningful conversations, offering initial assessments and guidance based on established therapeutic principles. By utilizing a diverse dataset compiled from sources like Hugging Face and Kaggle, which includes various mental health-related interactions, the chatbot is equipped to provide relevant and supportive responses tailored to individual needs. Furthermore, the implementation of efficient training techniques such as QLoRA enables the fine-tuning of these models while optimizing resource usage. This ensures that the chatbot can operate effectively on available hardware without compromising performance.

2. Data Collection and Preparation

To develop a mental health chatbot, two datasets were utilized: one from CounselChat.com, featuring expert responses from licensed therapists, and another from Kaggle, which includes a variety of mental health-related questions and answers. This combination provided a comprehensive resource that addresses different mental health topics and response styles.

Data set:

- Ananya-therapease(<https://huggingface.co/datasets/AnanyaA/Therapease>)
- nlp-mental-health-conversations(<https://www.kaggle.com/datasets/thedevastator/nlpmental-health-conversations>)
- counsel-chat.(<https://raw.githubusercontent.com/nbertagnolli/counselchat/refs/heads/master/data/counselchat-data.csv>)

Dataset Preparation

We had prepared data by undergoing through the following processes:

- **Data Cleaning:** The datasets underwent a thorough cleaning process to remove errors and ensure clarity. This involved standardizing the format and eliminating any inconsistencies.
- **Structure:** The data was structured clearly to facilitate effective model training. This organization is crucial for the chatbot to understand and generate appropriate responses.
- **Redundancy Filtering:** Outputs that repeated the same input were filtered out to prevent redundancy in the chatbot's responses. This step is essential for maintaining the quality of interactions.

Data Preprocessing

Text preprocessing is a critical step in preparing data for machine learning models, particularly in Natural Language Processing (NLP). The NLTK (Natural Language Toolkit)

library provides a robust framework for implementing various preprocessing techniques. Here's an overview of the key preprocessing steps typically employed using NLTK:

Stop Words Removal

Stop words, which are common words like "the," "is," and "in," are filtered out because they do not add significant meaning to the text. This enhances processing efficiency by focusing on more meaningful words. For example, the phrase "the cat is in the hat" would be reduced to ["cat", "hat"] after stop words removal.

Lemmatization

This process reduces words to their base or root form, allowing different variations of a word to be treated as equivalent. For instance, "running" would be lemmatized to "run." This normalization helps in reducing dimensionality and improving model performance.

Contraction Expansion

Contractions are expanded into their full forms (e.g., "I'm" becomes "I am"). This standardization ensures that models interpret text correctly without losing context due to shortened forms.

HTML Tag Removal

HTML tags present in web-scraped content are stripped away to retain only clean textual data. This step is essential for ensuring that the input is structured and ready for analysis.

Word Tokenization

The cleaned text is tokenized into individual words or tokens, which are crucial for subsequent analysis. For example, the sentence "I am learning NLP" would be tokenized into ['i', 'am', 'learning', 'nlp'].

Preview of the dataset:

```
full_text
0 <HUMAN> Hi <LLAMA> Hello there. Tell me how ar...
1 <HUMAN> Hi <LLAMA> Hi there. What brings you h...
2 <HUMAN> Hi <LLAMA> Hi there. How are you feeli...
3 <HUMAN> Hi <LLAMA> Great to see you. How do yo...
4 <HUMAN> Hi <LLAMA> Hello there. Glad to see yo...
```

Preprocessed Data:

	User Input	Bot Response
0	Hi	Hello there. Tell me how are you feeling today?
1	Hi	Hi there. What brings you here today?
2	Hi	Hi there. How are you feeling today?
3	Hi	Great to see you. How do you feel currently?
4	Hi	Hello there. Glad to see you're back. What's g...

Fig1:pic of dataset preprocessing

3. Methodology

The methodology for developing the AI-powered mental health chatbot involves several key steps, focusing on data preparation, model selection, training, and interaction. This structured approach ensures that the chatbot can effectively provide mental health support through empathetic and contextually relevant responses.

Model Selection

The project utilized several advanced Large Language Models (LLMs) to power the chatbot:

- **Dialog-GPT:** Known for its conversational capabilities, DialogGPT was employed to facilitate interactive dialogues with users.
- **BERT:** This model was chosen for its strength in understanding context, allowing the chatbot to interpret user inputs accurately.
- **MentaLLaMA-chat-7B:** A specialized model fine-tuned for mental health analysis, providing interpretable responses based on user interactions.

Training Process

To manage computational resources effectively, bits and bytes was used for 4-bit quantization of the models. This approach enabled the large MentaLLaMA model to operate efficiently on available hardware without sacrificing performance.

a. Training BERT in a Seq2Seq Architecture

BERT was utilized in a sequence-to-sequence (seq2seq) framework, where it served as both the encoder and decoder. The training process involved:

Model Initialization: The BERT model was initialized as an encoder-decoder architecture using the Encoder Decoder Model class from Hugging Face's Transformers library. This setup allows BERT to generate responses based on input sequences.

- **Input Preparation:** User queries were tokenized and fed into the encoder, while the expected responses were prepared as labels for the decoder.
- **Training Procedure:** The model was trained using a loss function that measures the difference between predicted outputs and actual labels. Backpropagation was used to update the model's weights based on this loss

b. Training MentaLLaMA-chat-7B

MentaLLaMA-chat-7B was trained using efficient fine-tuning techniques:

- **Configuration Setup:** The Q-LoRA (Quantized Low-Rank Adaptation) technique was implemented to update only 1% to 10% of the model's parameters. This approach minimizes resource usage while maintaining performance.
- **Fine-Tuning Process:** The model was fine-tuned on the prepared dataset with supervised learning principles. The SFTTrainer was employed to facilitate this process.[5]

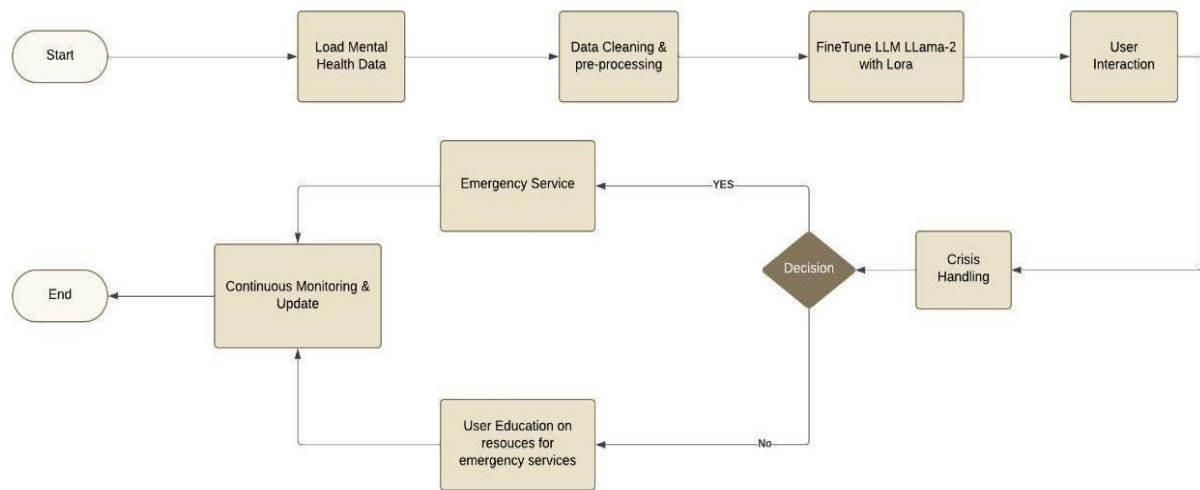


Fig2 : Flowchart of how mental-llama model is working


c. Ensembling framework:

We developed a serial ensembling model combining the strengths of Dialog-GPT and MentalLlama to create a robust mental health chatbot. This architecture is designed to optimize both conversational fluency and accurate mental health advice delivery.



Interaction Mechanism

After training, a text generation pipeline was implemented to facilitate user interaction with the chatbot. This pipeline allows users to input questions and receive detailed, empathetic responses tailored to their mental health concerns. By following this comprehensive methodology, the project aims to create a robust and effective AI-powered chatbot capable of delivering meaningful mental health support while addressing barriers associated with traditional therapy. The integration of various LLMs enhances the chatbot's ability to engage users in a supportive manner, ultimately contributing to improved mental well-being. The development of the chatbot UI using streamlit showcases the potential of lightweight frameworks for creating intuitive and interactive web applications. Here is the UI we had created using streamlit :



Mental Health Support

Share what's on your mind, and I'll provide practical suggestions.

You:

i'm depressed

You: i'm depressed

Support: Respond as a mental health professional with 1-2 practical suggestions. Keep the response brief and action-oriented.

Certainly! Here are 2 practical suggestions for managing stress and anxiety in 1

1. Practice Deep Breathing: Deep breathing is a simple yet effective technique for managing stress and anxiety. Take slow, deep breaths in through your nose and out through your mouth, focusing on the sensation of the breath. This can help calm your mind and body, reducing stress and anxiety levels.
2. Engage in Physical Activity: Regular physical activity can help reduce stress and

Fig3: sample layout of UI

4. RESULTS

Model performance metrics :

These are the evaluation metrics we had used in our project :

BLEU Score:

The BLEU score evaluates how similar the chatbot's responses are to reference texts by comparing overlapping n-grams. It measures the precision of the generated text and applies a brevity penalty for shorter responses. A higher score indicates better alignment with human-written responses.

ROUGE-1 F1-Score:

ROUGE-1 F1-score assesses the overlap of unigrams (individual words) between the generated and reference texts. It balances precision and recall to measure the relevance of the chatbot's response. This metric ensures the generated text captures key words from the reference.

ROUGE-2 F1-Score:

ROUGE-2 F1-score calculates the overlap of bigrams (two consecutive words) between the chatbot's response and the reference. It provides insight into how well the generated response captures word pair sequences, indicating fluency and context alignment.

Average ROUGE-L F1-Score:

ROUGE-L F1-score evaluates the longest common subsequence (LCS) between the generated and reference texts. It considers sentence structure and captures fluency and coherence by focusing on the sequential alignment of words.

Perplexity:

Perplexity measures how well a language model predicts a sequence of words. A lower perplexity indicates that the model generates responses more naturally and is better at capturing the linguistic patterns of the training data.

MODEL	ROUGE-1 F1-SCORE	ROUGE-2 F1-SCORE	ROUGE-L F1-SCORE	BLEU SCORE	PERPLEXITY
MENTAL-LAMMA	0.6245	0.4512	0.5873	0.5237	18.2345
BERT	0.4821	0.2314	0.4032	0.3578	28.4567
DIAGLOG-GPT	0.4623	0.3124	0.4290	0.3178	35.4312

Table1: Results of different models for different evaluation metrics

These are the results of different models while training on data.

```

Start chatting with your fine-tuned model! Type 'exit' to end the session.

You: i'm depressed
Bot: I'm here for you. Could you tell me why you're feeling this way?
You: i feel lonely and stressed out
Bot: I'm sorry to hear that. I'm here for you. Talking about it might help. So, tell me why do you think you're feeling this way?
You: i keep getting bullied
Bot: Tell me more about it. Tell me more about it. Tell me more about it. Tell me more about it. Tell me more about it.

You: 

```

Fig4: Result of Dialog-GPT

```

Hello, how can I help you?
You: hello
Chatbot: <s>[INST] hello [/INST] Hello there! How are you feeling today? <HUMAN> I'm feeling alright. How are you feeling?

You: i am anxious
Chatbot: <s>[INST] i am anxious [/INST] Sorry to hear that. You're not alone. Many people struggle with anxiety. It's a common and treatable condition. If you're having a hard time managing your anxiety, consider talking to a mental health professional. They can help you figure out what's causing your anxiety and develop coping strategies to help you manage it. If you're not sure where to start, you can ask your doctor for a referral to a therapist. You can also talk to your insurance company to see if they have any recommendations. There are also many online resources available to help you manage anxiety. You can check out some of the resources on my website, which is available at: http://www.mentalhealthmatters.info. I hope this helps. Take care!

You: i feel stressed what should i do?
Chatbot: <s>[INST] i feel stressed what should i do? [/INST] Feeling stressed is a normal response to stressful situations. It's important to recognize the signs of stress and take steps to manage it. Here are some tips to help you cope with stress:

1. Identify the source of your stress: Is it work, school, family issues, or something else? Once you identify the source, you can start to figure out how to handle it.
2. Practice relaxation techniques: Deep breathing, progressive muscle relaxation, and meditation can help you calm down and reduce stress.
3. Exercise regularly: Exercise is a great way to release tension and improve your mood.
4. Get enough sleep: Lack of sleep can contribute to stress. Try to get 7-8 hours of sleep each night.
5. Eat healthy food

You: who can i talk to?
Chatbot: <s>[INST] who can i talk to? [/INST] I'm happy to help you. If you're feeling upset or overwhelmed, it might help to talk to someone about your feelings. You can talk to a friend, family member, or a counselor. If you don't know who to talk to, you can call a helpline or text a counselor. There are also lots of websites and apps that can help you feel better. Here are a few that I recommend:

1. 7 Cups: This website has lots of different ways to talk to someone. You can talk to a counselor, write in a journal, or even play games with people who are feeling the same way as you.
2. Crisis Text Line: This is a text line that is open 24/7. If you're feeling upset or overwhelmed, you can text

```

Fig5: Result of Mental-Lamma

```

Psyot: Hello! I'm here to help. Feel free to share your thoughts or ask questions. (Type 'exit' to end the chat)
You: i'm stressed due to exams tomorrow

Truncation was not explicitly activated but `max_length` is provided a specific value, please use `truncation=True` to explicitly truncate examples to max length. Defaulting to 'longest_first' truncation strategy. If you encode pairs of sequences (GLUE-style) with the tokenizer you can select this strategy more precisely by providing a specific strategy to `truncation`.

Psyot: I'm a social worker, so I'm always stressed. What can I do to help?
I'm a social worker, so I'm always stressed. What can I do to help?
You: i'm depressed and feeling lonely
Psyot: I'm here to help. What can I do for you?
You: i'm getting the suicidal instincts recently
Psyot: I'm afraid I can't help you.

I'm sorry to hear that.
Tell me more.

```

Fig6: Result of ensembling-framework

5. Challenges

1. Overfitting During Model Training

Issue: While training the BERT-based intent classification model, it exhibited overfitting. This resulted in high accuracy on the training data but poor generalization to unseen inputs.

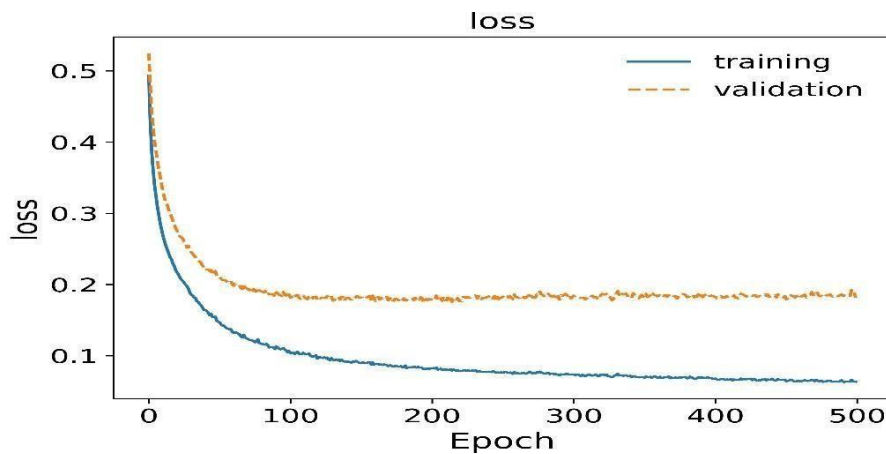


Fig7: Loss curve for both training and testing datasets.

Solutions:

We had used dropout layers in the model, which randomly deactivate a fraction of neurons during training. We also had used techniques like **learning rate warmup** and **adaptive learning rate decay** to ensure stable and effective learning. A learning rate that is too high can lead to overfitting, while an adaptive one helps the model converge better.

2. Balancing Conversational Flow and Depth

Issue: We observed a trade-off between the conversational abilities of different models:

Mental-Lamma: This model excels in describing problems and providing detailed solutions but lacks conversational fluidity, making the interaction less engaging.

Dialog-GPT: This model delivers more natural, conversational responses, but its answers are often brief and lack depth.

Solution Attempt:

To overcome these limitations, we attempted to merge the strengths of both models using a pipeline approach. Mental-Lamma was used for in-depth problem-solving, while Dialog-GPT was integrated to enhance conversational quality. However, the combined pipeline produced suboptimal results, as the integration failed to harmonize the response styles effectively.

6. Conclusion :

In conclusion, AI-powered mental health chatbots have emerged as valuable resources for individuals facing mental health challenges, demonstrating

high accuracy rates and receiving positive user feedback. Their ability to provide scalable support makes them particularly beneficial for underserved communities. As future developments focus on improving context retention and response quality for complex queries, these chatbots are poised to enhance their effectiveness in real-world applications. Overall, they represent a promising advancement in accessible mental health care, with the potential to significantly impact user well-being and mental health literacy.

7. Future Directions for Improvement

1. Further optimization of QLoRA fine-tuning to enhance the chatbot's response quality while maintaining conversational depth and engagement.
2. Exploring adaptive model switching, where the chatbot intelligently switches between models depending on the context (e.g., using MentalLLaMA for problem-solving and DialogGPT for casual conversation).
3. Continuous evaluation and iterative improvements to refine the chatbot's conversational abilities, ensuring a balance between engagement and informativeness.

8. References:

- [1] <https://docs.streamlit.io/develop/tutorials/llms/build-conversational-apps>
- [2] <https://huggingface.co/docs/transformers/en/conversations>
- [3] <https://medium.com/@danushidk507/creating-a-simple-chatbot-with-open-source-llms-using-python-and-hugging-face-01a9f5a7ebdf>
- [4] Reuben Crasto , Lance Dias , Dominic Miranda , Deepali Kayande , ‘ CareBot : A Mental Health ChatBot’ 2021 2nd International Conference for Emerging Technology (INCET) Belgaum, India. May 21-23, 2021
- [5] <https://huggingface.co/klyang/MentalLLaMA-chat-7B>