MENTAL HEALTH CHATBOT

-Rushikesh Deotale

-Samved Shah

Abstract

Our project aims to develop a mental health chatbot using the existing large language models fine-tuned on a dataset of mental health-related questions and answers. The chatbot's purpose is to provide empathetic and supportive responses to users seeking assistance with their mental health concerns. The development process encompasses training the chatbot on the custom dataset, implementing a user interface using Tkinter, and evaluating the models' performance in generating contextually relevant and compassionate responses.

Overview

The project report details the creation and evaluation of a mental health chatbot designed to address the increasing need for accessible and stigma-free support. The initial motivation stems from the societal challenges surrounding mental health, where individuals often face barriers to seeking help. The chatbot serves as a conversational agent, offering users a safe space to discuss their mental health concerns and receive supportive responses.

What is the Problem?

The problem addressed by this project is the limited accessibility to mental health support and the associated stigma. Many individuals hesitate to seek help due to societal judgments or lack of confidential and easily accessible platforms. The motivation is to create a virtual space where users can freely express their feelings and receive empathetic responses, thereby promoting mental well-being.

Why is this Problem Interesting?

This project is interesting because it tackles the broader societal issue of mental health support accessibility. The chatbot serves as a virtual companion, providing a confidential and judgment-free environment for users. It is relevant in society as it offers a potential solution to the challenges individuals face in expressing their mental health concerns and seeking support.

Approach:

The proposed approach involves fine-tuning two models, the GPT-2 language model and the Roberta Model, on a dataset comprising mental health-related questions and answers. The models are trained to understand user inputs and generate contextually relevant and supportive responses. The approach leverages natural language processing techniques and machine learning to simulate empathetic conversations.

Rationale Behind the Proposed Approach:

The rationale behind the approach is to create a chatbot that understands and responds to users in a way that simulates human-like empathy. The fine-tuning process ensures that the model captures the nuances of mental health conversations. While there are references to previous work on chatbots, this project distinguishes itself by prioritizing empathy and context-aware responses, making it more suitable for mental health support.

GPT-2 is an autoregressive model, which generates words in a left-to-right manner, this makes GPT2 a favorable choice. GPT-2 is suited for generating contextually relevant responses, GPT-2 is also good at creating a question answering environment.RoBERTa on the other hand works in a different way where the left and right both contexts are used during training purposes.It uses a bidirectional context understanding. The approach used in RoBERT gives a more holistic understanding of the input.

GPT-2 and RoBERTa were chosen as models for testing and training since these had different approaches while training and understanding text data.GPT-2 and RoBERTa have different pre-training objectives and architectures, providing complementary strengths. GPT-2 focuses on autoregressive generation and creative text, while RoBERTa emphasizes bidirectional context understanding and language representation.

Key Components

The key components include a large language model fine-tuned on a mental health dataset, a Tkinter-based user interface, and a response generation mechanism. The chatbot demonstrates effectiveness in generating empathetic responses, but it has limitations, particularly in handling complex emotions. The user interface allows for easy interaction with the chatbot, enhancing the overall user experience.

Limitations:

One notable limitation is the chatbot's potential bias based on the training data and its struggle with complex emotional expressions. Continuous improvement is necessary to address these limitations and enhance the chatbot's overall performance.Due to the scarcity of examples, the chatbot may have difficulty generalizing its knowledge to a broader range of mental health-related questions. Responses may lack diversity and fail to cover a wide spectrum of user queries

Experiment Setup

The experiment involved fine-tuning the GPT-2 and Roberta language model on a custom dataset of mental health-related questions and answers. The dataset was preprocessed to combine questions and answers into input sequences. The fine-tuning process aimed to enhance the model's ability to generate contextually relevant and empathetic responses in mental health conversations. The training setup utilized the Hugging Face Transformers library, and the model was trained on a computing environment with sufficient resources.

Dataset:

The dataset used for fine-tuning the chatbot consisted of a collection of mental health-related FAQs. Basic statistics of the dataset include the total number of questions, answers, and the combined input text. This dataset was carefully curated to ensure diversity in mental health topics, providing a comprehensive training ground for the chatbot. The questions and answers were combined to create a text corpus for training the model.

Implementation:

GPT- 2:

The implementation involved using the Hugging Face Transformers library to load the pre-trained GPT-2 model and tokenizer. The dataset, created by combining questions and answers, was saved to a text file. The TextDataset and DataCollatorForLanguageModeling classes were utilized to tokenize and process the dataset for training. The training process employed a Trainer class, specifying training arguments such as the number of epochs, batch size, and output directory. The model was fine-tuned on the custom dataset in a computing environment suitable for training deep learning models.

RoBERTa :

For the other model we preprocessed the data in the similar way as for the GPT-2 model. We changed the tokenizer and used Hugging Face inbuilt tokenizer RoBERTa.The training process employed a Trainer class, specifying training arguments such as the number of epochs, batch size, and output directory. The model was fine-tuned on the custom dataset in a computing environment suitable for training deep learning models.

Model Architecture:

RoBERTa:

The model is a variant of the BERT architecture which is based on bidirectional encoder representations of transformers. It is a transformer-based language model that uses self-attention mechanisms to generate output words in a sentence. The difference between BERT and RoBERTa is that the training size of the latter was much larger , making it more effective. Also RoBERTa uses a dynamic masking technique during training which helps the model learn more robustly.

The model uses transformers along with attention mechanisms which considers text as whole compared to using a sequence of text either from left to right or right to left giving it's nature of bidirectional representation. The transformer includes two mechanisms, an encoder mechanism followed by a decoder mechanism.The encoder is used to understand the inputs and decoder is used to generate the output sentences.The self-attention mechanism allows the model to consider the entire input sequence bidirectionally, capturing relationships between words regardless of their position in the sequence. The attention mechanism enables RoBERTa to understand context and dependencies between words in both directions.

For fine tuning we used the pretrained weights of the model, followed by modifying the input sizes along with training masks. The data was loaded using the TextDataset library and training was done on our dataset.

GPT-2

The model architecture utilized for fine-tuning was GPT-2, a pre-trained language model known for its capabilities in natural language processing tasks. The GPT-2 architecture is based on a transformer neural network, employing attention mechanisms to capture contextual information.

GPT-2 consists of a stacked layer of multiple transformer models along with feed-forward mechanisms.The model uses the input tokenized sentence along with a vector which compasses the sentence context using attention mechanisms. This vector is used for understanding the context of the sentences and the words which are important for the entire context of the input. These are passed for the decoder to generate the output sequence. GPT-2 first uses the self attention block for focusing on context words, the idea of the attention is that some words and phrases capture more meaning needing more attention. These are then passed to a feedforward network which is a complex network made of input layers accepting attention vectors. The hidden layers capture the correlations between each data point. The output is then tokenized back into a sentence. GPT-2 uses the input sequence to generate a new word, which is then again passed to it as new input. This means that the input sequences sequentially produces the output one after the other and doesn’t exactly reads the text as whole at once like the BERT architecture.

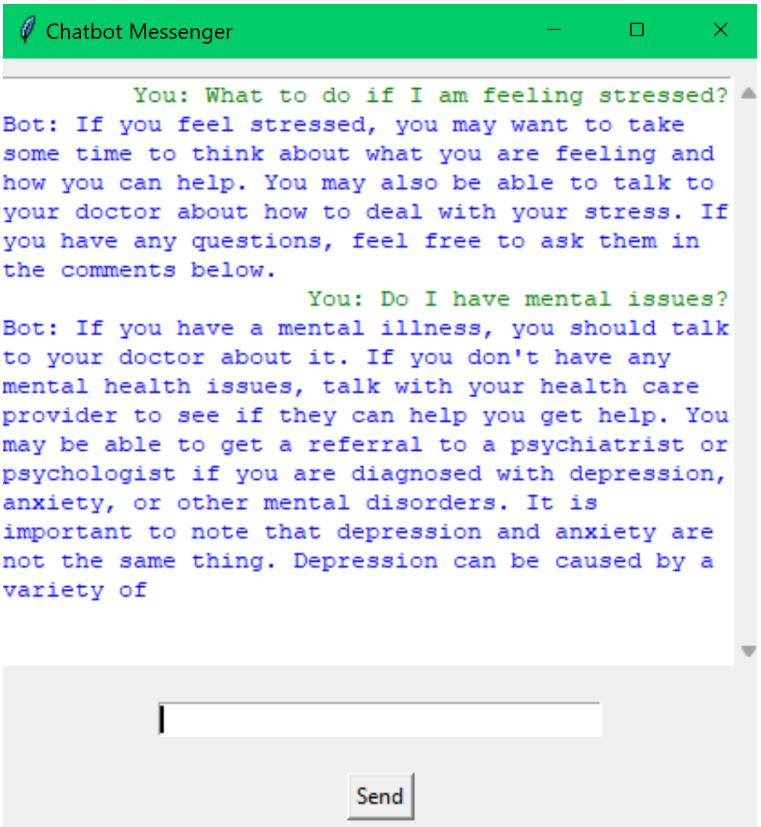
The fine-tuning process adapted the pre-trained GPT-2 model to the mental health domain, enhancing its ability to generate coherent and empathetic responses in conversations.

Main Results:

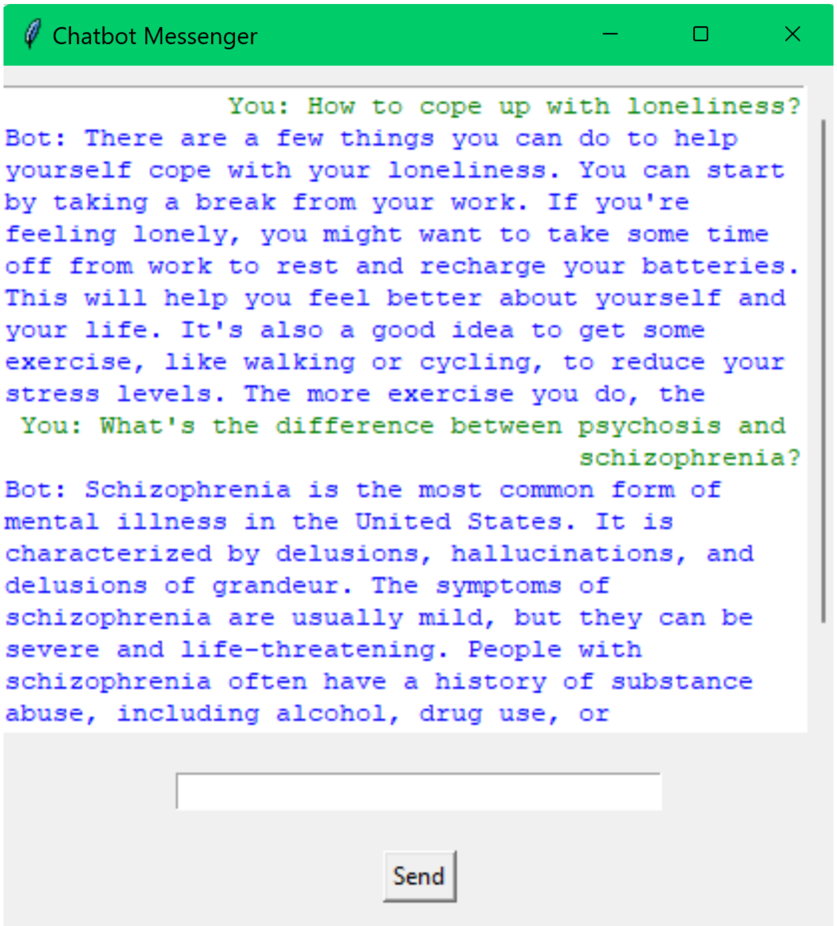
The training accuracies for both models were reported after training epochs. We observed that Roberta was giving highly inaccurate and ineffective results. The result meant that the bidirectional architecture was not enough to generate appropriate responses. One of the reasons for this could be the small corpus of the dataset.Given the very small size of the data samples we couldn't generate responses which are good enough for evaluation.

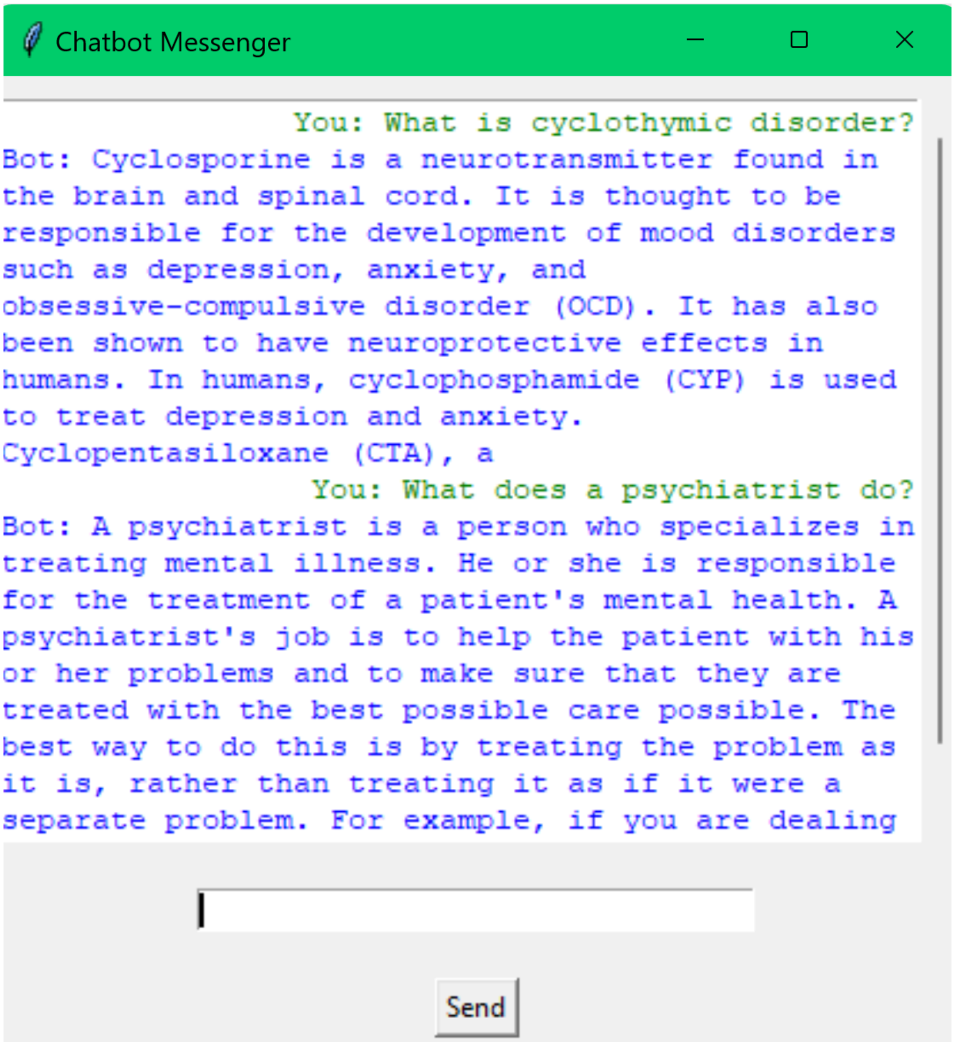
On the other hand when we trained the GPT-2 model, it gave responses which were meaningful, the feedword nature of the GPT-2 architecture enables it to generate responses which are meaningful sentences even if the dataset samples are smaller in nature. GPT-2 is known to perform for generative approach on a smaller dataset since it takes into account the context from previous words and keeps on adding the generated word in the new input. While the BERT architecture based RoBERTa uses a mask modeling approach which masks random words from the input which act as attention masks giving those words more weightage

In conclusion the results underscore the chatbot's effectiveness in simulating empathetic conversations related to mental health. We found value in the chatbot's ability to understand and respond to their queries, creating a supportive virtual space for discussing mental health concerns. The chatbot's success in generating coherent and relevant responses signifies its potential as a helpful tool in mental health support.


Chatbot Messenger
You: What to do if I am feeling stressed?
Bot: If you feel stressed, you may want to take some time to think about what you are feeling and how you can help. You may also be able to talk to your doctor about how to deal with your stress. If you have any questions, feel free to ask them in the comments below.
You: Do I have mental issues?
Bot: If you have a mental illness, you should talk to your doctor about it. If you don't have any mental health issues, talk with your health care provider to see if they can help you get help. You may be able to get a referral to a psychiatrist or psychologist if you are diagnosed with depression, anxiety, or other mental disorders. It is important to note that depression and anxiety are not the same thing. Depression can be caused by a variety of
Send

Additional Screenshots





Supplementary Results:

Parameter choices during the experiments included setting the block size for tokenization, training epochs, and batch size. The block size was chosen to balance computational efficiency and context retention. The number of training epochs was determined based on achieving satisfactory convergence, and the batch size was selected to fit the available computing resources. These choices aimed to optimize the trade-off between model performance and computational efficiency, ensuring a balance in training time and effectiveness.

Conclusion and Future Work:

The project acts as a starting point for creating a mental health chatbot which is good enough to work on a small dataset.The project aims to use large language models which have provided valuable insights into the challenges and potential of leveraging large language models for empathetic conversation in the mental health domain. The project addresses problem of limited mental health support and acts as a virtual space for users to interact with whenever they seek mental health.

The use of GPT-2 and RoBERTa architectures explored different approaches to simulating conversations via a question and answering chat.GPT-2 demonstrated effectiveness in generating contextually relevant and meaningful responses, highlighting its suitability for generative tasks with limited data. On the other hand, RoBERTa, with its bidirectional context understanding, faced challenges in handling the nuances of mental health-related queries, especially in the context of a small dataset.

In addition to the existing data one can look for non-publicly available datasets which are large corpuses of text containing questions and answers. The training on larger data corpus could increase performance of the above models. Also trying out other large language models might be something which can be explored as well other than the models one used here.With a larger data the BERT architecture could perform much better compared to the current performance helping it overcome challenges it faced in handling complex emotional expressions and generating meaningful responses.

Adding a simple user interface using Tkinter library is useful for user testing and easy user interaction. A simple user interface the one shown in the project can be easily amalgamated on different platforms as well with ease.