CuringBot Project Proposal

Rui Ji, Johnny Yang, Prithvik Gowda

1 Problem and Motivation

Currently, mental health has been a striking problem for adults, especially teenagers and young adults. The lack of access to therapy could cause destructive consequences to those unprivileged. We intend to build a language model that appropriately responds to the situation provided by users, leveraging domain knowledge carried in training data, and delivers the first step of support.

2 Related Work

2.1 Llama-2-Chat:

A fine-tuned model based on Llama-2 that has specifically been used in conversational scenarios and responds to user input as a prompt. The model has sizes 7B, 13B, and 70B and is generally one of the most powerful open sources under each specific size. The model has gone through a training period and is fine-tuned by RLHF to succeed good ability in chatting.

2.2 Falcon-Instruct:

Falcon-Instruct is another famous open-sourced LLM fine-tuned by the Falcon basic model. Right now, Falcon have released 7B, 40B for instruction (relatively lower ability in chat but still fairly good), and 180B for chat recently.

2.3 Vicuna-V1.5:

Another chatable9 generative model fine-tuned from Llama 2 using conversations derived from ShareGPT.com, another robust model using a transformer model and can perform agent-like chatbot abilities..

3 Hypothesis

By fine-tuning base language models such as Llama, GPT-2, ChatGLM, etc with data specific in the mental health counseling field, we could achieve quality responses toward users' inquiries about their mental health problems. To a specific measure, we hypothesize that the quality of response will surpass responses from powerful general language models such as GPT-3.

4 Approach

We begin by locating proper datasets that could be used to train domain knowledge about mental health into general base language models.
https://huggingface.co/datasets/Amod/mental_health_counseling_conversations?row=96

Next, we plan to choose a few base language models, compare the mechanisms/training process behind them, and fine-tune them with the aforementioned dataset.

Once finished, we will develop a scoring mechanism(using GPT-4, for example) that evaluates the quality of these models in the task of responding to mental health problems.

Compare the score of models among different base models and general language models without fine tuning, such as GPT-3.

Find reasoning of the difference, and potentially the advantages and disadvantages of each.