Implementation of Medical FAQ Chatbot

Proposed by: Cohort -8 Bourguiba Team

Introduction

This project has the potential to bridge the information gap in healthcare and promote better health practices in Africa. It will make healthcare information available and accessible 24/7 for individuals to make informed decisions, improve healthcare outcomes and contribute to public health improvement in the continent.

Problem Description

The public needs a better way to access accurate medical information. The medical field is vast and complex, and it can be difficult for the public to keep up with the latest information. This can lead to confusion and misinformation, which can have negative consequences for patients and their health.

In addition, doctors need a way to attend to a large amount of patients, which overburdens them. This usually leads to some patients having to wait for a very long time before they get the attention of the doctors. Which also reduces the quality of care.

This problem is becoming increasingly acute as the population ages and people live longer with chronic diseases. There is a growing need for a system that can provide the public with accurate and up-to-date medical information in a timely manner, and that can help reduce the doctors burden of attending to a large amount of patients.

Proposed Method

The proposed model for solving the problem of misinformation, overburdened doctors and delayed patients is to develop a chatbot that addresses this issues. The end product for this chatbot would go through the following ML basic standards of data

sourcing, data cleaning and preparation, machine learning model development, and model deployment, which would be further explained.

i) Data Sourcing

The first step is to source the data that will be used to train and deploy the chatbot system. This data could come from a variety of sources, such as: Kaggle data set, Google's dataset search platform, also scraping from Health related websites if necessary.

ii) Data cleaning and preparation

Once the data has been sourced, it will need to be cleaned and prepared for use in the machine learning model. This would involve natural language processing as we would be making use of mostly textual data.

iii) ML Model

Once the data has been cleaned and prepared, it can be used to train our selected models. The specific type of model we would be using are large language models that would be fine-tuned with our data set. Our selections would be language models and also classical models, which we would carry accuracy tests on to determine the best model for our use case.

These models will be further trained or fine-tuned on the data that was sourced and prepared in the previous step. Once the model is trained, it will be able to perform the tasks that it was designed to do, such as providing medical information.

iv) Model Deployment

Once the model has been trained, it would be deployed so that it can be used by the public and doctors. Our considerations for deployment are streamlit, Gradio and huggingface. We selected these options due to their ease of use and possible implementation of the chatbot into these platforms.

Proposed Split

The various subteams are

1. Data sourcing and verification sub team – Ayodeji Akande, Caleb Balogun
2. Data cleaning and preparation sub team – Monsurat Ariyo, Martins Joseph
3. Model subteam – Bala Abduljalil, Ayodeji Akande
4. Model deployment subteam – Caleb Balogun, Monsurat Ariyo,
5. End product verification subteam – Martins Joseph, Bala Abduljalil,

Conclusion

The proposed AI-powered system has the potential to be a valuable tool for the public and doctors. By providing accurate and up-to-date medical information, combating medical misinformation, and helping doctors attend to patients information needs, the system can help to improve the quality of care that patients receive.

Possible Extensions

One possible extension of the project is to make the model able to receive images of physical symptoms and also audio input. This would allow the system to be used for a wider range of tasks, such as:

1. Diagnosing diseases based on images or audio recordings of symptoms.
2. Providing personalized medical advice based on the patient's individual needs and circumstances.
3. Monitoring patients' health remotely.

Another possible extension of the project is to make the system multilingual. This would make the system accessible to a wider range of people, and would help to reduce the barriers to healthcare for people who speak languages other than English.

References

Cobs, G. (2023). Using LangChain To Create Large Language Model (LLM) Applications Via HuggingFace. Medium. https://cobusgreyling.medium.com/langchain-creating-large-language-model-llm-applications-via-huggingface-192423883a74

Langchain, inc. (2023). Langchain documentation. https://docs.langchain.com/docs/

Meta (2023). LlaMa. Introducing llama 2. https://ai.meta.com/llama/

Murtuza, K. (2023). Using LLaMA 2.0, FAISS and LangChain for Question Answering on Your Own Data. Medium. https://medium.com/@murtuza753/using-llama-2-0-faiss-and-langchain-for-question-answering-on-your-own-data-682241488476