Bourguiba Team Project - AI Saturdays
Medical Frequently Asked Question Model Project

This report provides an update on the progress of our Medical Frequently Asked Questions (FAQ) Model project. As of the 15th of December, the project has advanced through the initial stages of project proposal submission, data collection, model development, and is currently in the model deployment phase.

Project Proposal Submission:
The project was initiated with the submission of a comprehensive project proposal outlining the objectives and necessary activities of the team in building the Medical Q&A Model. The proposal received approval, paving the way for subsequent stages of the project.

Data Collection:
Data collection was a crucial phase in ensuring the model's robustness and effectiveness. We sourced data from various platforms, including Hugging Face and Kaggle. This diverse dataset comprises medical question and answer pairs.

Data Validation and Cleaning:
We did data validation and cleaning on the data set we collected to ensure usability. The cleaning steps included the removal of irrelevant information from question-and-answer pairs, splitting data columns into separate question column and answer columns, and compiling the refined dataset into a structured format. The organized data is now stored in a designated Google Drive folder for accessibility and collaboration.

Current Stage: Model Development:
The project has now progressed to the model development stage. Key components of this phase include:

1. Requirements**:**
   - Python 3.x
   - Pytorch
   - Hugging Face's Transformers library

2. Download Pre-trained Model:
   Obtain the pre-trained GPT-2 model weights.

3. Install Dependencies
 Install the necessary dependencies using the following command:
   pip install torch

4. Exploration of Different Models:
   - Classical Models: Initial experimentation with traditional machine learning models to establish a baseline performance.
   - Transformer Models: Utilisation of transformer-based architectures, leveraging their capabilities in capturing complex relationships within sequences. We considered GPT-2, XLNet
   - Language Models: Exploration of language models for natural language understanding tasks. We considered BERT
   - Retrieval-Augmented Generation Model: Uses pre-trained models and a knowledge base like a CSV file to generate answers to questions. We considered the Facebook RAG Token model, Open AI GPT 3.5 with Langchain, LLAMA2 with Langchain

   5. Generate Text with a Fine-tuned Model

Next Steps:
1. Model Evaluation:
   - Conduct thorough evaluations of each model to assess their performance based on predefined metrics.
   - Identify strengths and weaknesses of different models to inform the selection of the most suitable architecture.
- Initial evaluation was done. More evaluations still need to be done by a medical practitioner who needs to review the performance in order not to release harmful content.

2. Further Fine-Tuning:
   - Commence a further fine-tuning process on the best model to further improve performance. This will be done with consideration of available time.

3. Model Deployment on Streamlit or Gradio
   We were unable to deploy this yet but will be added.

4. Model/ Product Validation

<u>Conclusion</u>:

We learned to experiment with different models and methods for building question answering systems.

While we plan to create a demo using the model, we also believe that before a model like this can be released, its output would need to be reviewed by medical professionals. Also, with more resources, advanced models like GPT3.5 or GPT4 can be used for the same task.