

# Retrieval Augmented Generation with MedQA

AI Club

September 16<sup>th</sup>, 2024

## Abstract

Your task is to use Retrieval Augmented Generation to enable LLMs to answer medical questions with accurate information sourced from textbooks. The MedQA dataset[1] is a collection of questions and answers, as well as textbooks from which reliable information can be sourced

## 1 Introduction

Retrieval Augmented Generation (RAG) is a useful technique for grounding LLM answers in reality. A retrieval model can be used to extract documents related to a question. These can then be passed to the LLM to reduce hallucinations and the reliance on data in their training set. Medicine is a domain where RAG can be really useful, as data that LLMs were originally trained on can be misleading, inaccurate, or incomplete for many medical topics, especially with more detailed questions.

## 2 MedQA Dataset

The data is available at <https://github.com/jind11/MedQA>. You only need to use the US section of this dataset. The data is sourced from medical exams, and each question has multiple choices, of which one is correct. Use the available textbooks as the set of documents for retrieval. You can split these textbooks into further sections if you feel they are too large.

You don't need to train the model on this dataset, you can use it for quick testing and evaluation. You do need to use the textbooks from this dataset.

## 3 Instructions

- You must use a hybrid search method with variable alpha to retrieve possible answer sources for a question
- You can use pretrained embedding models for the retriever, as well as pretrained LLMs for the model.
- Create a basic UI using streamlit (or another framework if you're more familiar with it) to use as a chatbot.

Look into langchain, vector databases for storing embeddings, sparse and dense retrievers.

### 3.1 Bonus

- Evaluate your model on the US portion of the dataset (Use accuracy as the metric).
- Accept PDFs in your UI to add more information for the retriever.
- Try and prevent hallucinations - see if your model can refuse to answer questions if the relevant data source is removed.
- Compare the performance of the LLM with and without RAG.

## References

- [1] Di Jin et al. "What Disease does this Patient Have? A Large-scale Open Domain Question Answering Dataset from Medical Exams". In: *arXiv preprint arXiv:2009.13081* (2020).