Project Proposal

Sunny Zhou

The Task

For this project, I would like to choose to task 2: a RAG conversational assistant.

The Purpose

At some point, I would like to have a personal website showcasing my projects, portfolio, and resume. To elevate this website, it would be interesting to have a chat assistant to guide visitors through the work and answer questions about the work.

Data

Instead of some occupational domain, such as legal or scientific, I plan to make it about my projects. This will be sourced from my GitHub. This variety of this data will be challenging to deal with. It be mostly consists of code samples (some with comments, some without), reports, and some config/system files. Some repos will need some cleaning and polishing of the README.md files to provide context for the rest of the content. It will be challenging to tying all of these different documents to each other, without crossing repos.

Knowledge Base

The data chunking to build the knowledge base will include cutting up all of the code and reports located in my public GitHub. From an initial search of the vector searching tools listed in the guidelines, ElasticSearch seems the most attractive, for its speed and existing support community. However, something like Milvus would be interesting to use given the high-dimensional embeddings, which allow me to capture information tied less to the document. The final decision will require more research.

Retriever and Generator

Like the previous considerations, I would like to start with ElasticSearch for the retriever, but may pivot. For the generator, I found people on forums using Mistral to match author styles for generation, which would fit my need to match my talking habits.

Goal and Evaluation

It is obvious that this task is ambitious especially considering the other work I have, so setting a realistic goal – one that I would be happy to reach – is important. I would be happy if this assistant would be able to generate and return accurate work chunks with explanations relevant to the users' queries or questions. Any other goals spelled out below, such as the tone and some hallucination concerns, will be cherries on top.

The evaluation for this assistant would come with a ground-truth (i.e. did it retrieve the intended document), which accuracy would work fine for, and a more open-ended metric (e.g. how well did it answer a clarifying question); the latter being more difficult to answer and requiring human evaluation.

Foreseeable Challenges: Security

There are many challenges that I need to solve before I would be comfortable having this assistant represent me independently. Some of these challenges, listed here and above, I do not plan to solve for this project, but are still interesting to think about.

In particular, I have a big concern regarding the credibility of this assistant. Due to the nature of the task and dataset, the main challenge will be making sure the assistant not hallucinating. For this to be a useful additional to a personal website, fake knowledge or stretched truths would give the wrong impression of me. For example, if the assistant oversells my abilities to employers or states false problematic political opinions posing as me, this could lead to obvious complications. Additionally, if this assistant starts sharing personal information (e.g. Personal Tokens) that could cause security risks for me. This will require a lot of tweaking and testing that might span longer than the allotted time.