SE Proj 1d1

Team: Anmol Koul, Arnav Mejari, Om Kumar Singh, Shweta Patki

Reflection Document

1.   What are the pain points in using LLMs?

i) Generalization : Having used 3 LLMs (Chat GPT, Gemini & Perplexity) in building the requirement document for our project, we have observed that the LLMs lack good generalization. This is sometimes countered by multiple careful prompting, however it is not guaranteed and the users may need to begin a new chat.

ii) Summarization : We have observed during our that the LLMs do not summarize the requirement content consistently, bringing some noise or irrelevant data which degrades the quality of the answer.

iii) Memorization: We have faced problems with LLM's extracting required data from documents, especially when the number of documents becomes high. Furthermore, the limit of attaching documents(say 10) in each prompt causes further issues with the LLM sometimes forgetting some batches of document information.

2.   Any surprises? Eg different conclusions from LLMs?

i) After giving the same prompts to multiple LLM's for comparison, we noticed that the answers (usecases) changed significantly with each LLM. In some cases, the answer from one of the LLM would fit in with the others. However, there were instances where the answers do not correctly depict what was asked from the LLM itself.

ii) Depending on the LLM, the answers were either too technical or too casual with some LLMs (we noticed this with Perplexity AI).

3.   What worked best?

i) Multiple careful prompts provided with examples seems to help in creating good answers.

ii) Surprisingly, writing certain keywords in ALL CAPS instructed the LLM to prioritize and provide satisfactory answers based on those keywords.

iii) Merging answers from multiple LLMs to get a better overall answer.

4.   What worked worst?

i) Utilizing zero shot prompting provided the worst generalized answers.

ii) Using vague prompts rather than clear prompts.

5.   What pre-post processing was useful for structuring the import prompts, then summarizing the output?

Using capitalized text, a clear file structure, and asking for short, simple answers in a regulated format was useful to be able to understand and summarize the output.

6.   Did you find any best/worst prompting strategies?

i)   From our experience, we found that careful prompting with a few examples provided the best results for our requirements. Zero shot prompting, on the other hand, provided answers which looked fine at first glance, but were not sufficient when we dug deeper into the subject.

ii)   We found that speaking as if to a colleague rather than a friend (with more technical terms and older internet terminology) was useful in getting clearer and more centralized answers.

iii)   Different strategies work for different LLMs, and we can guess the sources of the data based on what works. Gemini works with some disparagement, ChatGPT works fine if spoken to as a 'character' like in a book, and Perplexity works best with lots of technical terms.