CEN5035 - Project Proposal

Group 7

Members: Zhixi (Zack) Lin, Linwei Jiang, Yuki Hanyan Zhang

Research questions and motivation

(1) If developers were to rerun their prompts with ChatGPT now and/or with different settings, would they obtain the same results? (From Challenge #9)

- The motivation behind answering this question is to determine whether ChatGPT will be able to provide a better response if the question asker is unsatisfied with the initial response. For instance, if a developer asks ChatGPT to provide them with an algorithm that will resolve their issues, but the algorithm that ChatGPT provided does not resolve the issue. Does the developer have to come up with a new prompt to get a new result? Or can they input the same prompt repeatedly with/without a different setting until a better solution is found. By answering this research question, developers will be able to save a lot of time from either having to come up with a new query every time for a different answer, or run the same query multiple times for the same incorrect response.

(2) Will it be typically difficult for users to specify a setting verbally for ChatGPT, such that ChatGPT will be able to return a code that satisfies all the described criterias?

- As ChatGPT is able to recognize different settings that the developer requested for their code, and potentially generate code accordingly, it would allow developers to customize the code that could potentially suit their own application perfectly. Thus, through this prompt we would be able to evaluate the feasibility of such customization of generated code, and how far will ChatGPT be able to withstand different settings and not fail.

(3) Provided that ChatGPT has generated a snippet of code with/without a specific tuning/setting, will ChatGPT be able to convert that to a different language and the code will still perform correctly as well with similar efficiency.

- This research question wants to analyze whether users including a more specific setting when asking a question will result in different efficiencies for different programming languages. If adding settings or leaving out some settings will result in lower efficiency of certain programming languages, that means more training will be required for generating code for those specific programming languages. The motivation of this research question is to allow programmers who use ChatGPT to generate code for all programming

languages to receive an equally efficient result without relying on whether specific settings are given or not.

Methodology

- Attempt to research online and see if we are able to find existing, recent data that are relevant to our topic and have been parameterized to suit our needs in evaluating the 3 research questions. This will mostly comprise a dataset from the Mining Challenge site.

- Attempt to use ChatGPT API calls to write programs that are able to generate data points for the above evaluation hopefully as automatically as possible, but a lot of manual evaluation would be necessary for this process. These data points will be from running the same prompts from the DevGPT dataset on the Mining Challenge site and any other dataset we decide to use on ChatGPT again and recording the responses. We will also add different settings onto some prompts.
    - For research question 1: We will feed the same prompts that other developers used to ChatGPT and record responses down to compare later. We will also incorporate some key words into some of these prompts to add different settings to the prompts. For evaluation on whether the same results are obtained, we will either attempt to manually evaluate a small part of the resulting dataset, or we will see if we can discover a way to automate it.
    - For research question 2: Attempt to find (in DevGPT) or generate (in ChatGPT) codes with prompts that consist of none all the way up to 20 different customization settings that range from implementation detail, logics, algorithm used, library used, etc.. Then evaluate these results and see if the code generated by ChatGPT is reasonable, logical, and matches with the settings.
    - For research question 3: Prompt ChatGPT with coding problems, with settings or without settings, and ask it to generate results using different commonly used programming languages, store these results to create our own dataset. Lastly, we want to evaluate the level of efficiency of the generated code snippet in different programming languages for the same prompt i.e. same coding problem.

- Regarding the prompts we can feed into ChatGPT to retrieve the dataset, we may want to use the coding problems in Leetcode. Since the level of difficulty of these problems

are divided between easy, medium, and hard, we can also try to find out whether changing the level of difficulty can affect the conclusion of our research problems.

Work plan

1. Data Gathering and Topic Extension stage (10/12 - 10/22)
2. Resolve Question 1, Compose the paper along the way. (10/22 - 11/12)
3. Resolve Question 2 with 3, Compose the paper along the way. (11/13 - 11/23)
4. Finalize the Paper.

* For now we will be doing each stage as a team and together. However, after stage 1, if it appear that it may be easier to divide each question among us, then we will assign the questions as follows:

Question 1 - Linwei Jiang

Question 2 - Zhixi (Zack) Lin

Question 3 - Yuki Hanyan Zhang