PROJECT PROPOSAL – TEAM 04

Team Members:
Jothiswaran Arumugam - JA23L
Venu Madhuri Devineni - VD23C
Sainath Erukulla - SE23L

RESEARCH QUESTIONS AND MOTIVATION

1.What is the typical structure of conversations between developers and ChatGPT? How many turns does it take on average to reach a conclusion?

To better understand the model's effectiveness and efficiency, it may be useful to provide an answer to the question regarding the typical number of turns required to reach a resolution in chats with ChatGPT. Developers and users may find this information valuable in a range of contexts, such as chatbots, customer support, content creation, and any scenario requiring human-AI interaction. Knowing how many rounds are often needed can help developers create conversational systems that are more effective, and users will have better expectations about the complexity and length of interactions.

2. How precise is ChatGPT at providing responses to developer's inquiries about various programming languages?

The goal of employing ChatGPT and other AI to answer programming-related questions is to offer developers a handy and easily available source of knowledge and support. It may be used for a variety of jobs, including:

- Quickly seeking up a programming language's or library's syntax or use.
- Obtaining advice on recommended procedures and coding standards.
- Debugging or troubleshooting code bugs.
- Investigating new frameworks or programming languages.
- Locating answers to typical coding issues or difficulties.

3.How ChatGPT expedites the solution-finding process for developers which leading to time savings in their quest for answers?

The motivation of the statement is to highlight how ChatGPT helps developers identify answers more rapidly, making it a useful tool for them. By emphasising the time-saving benefit, the statement intends to emphasise the useful advantages and productivity enhancements that developers may have while utilising ChatGPT in their work.

METHODOLOGY

To begin with, we will import the GitHub dataset into a **Pandas** DataFrame (Pandas library) and explore it to learn more about its structure and content. Using **data filteration** techniques, we will narrow the dataset to concentrate on the problem with a certain state.

For the first query, our goal was to compute the **average prompt count** in order to ascertain the normal number of turns in dialogues between developers and ChatGPT for addressed issues. To depict the distribution of prompt numbers, we shall use **visualisation** techniques.

For the second query, we will determine accuracy by comparing ChatGPT's replies with the actual data to **calculate precision**. We can presume that your dataset's **'RepoLanguage'** field refers to the programming language. To make your point more obvious, see the precise outcome. An increased precision rating denotes more accuracy in giving the right answers.

For the third question, we'll concentrate on the **time-related variables** to examine how long different phases of pull requests take to complete. The time difference will then be calculated and shown using the "**matplotlib**" and "**seaborn**" libraries to show trends. The next step is to examine the temporal patterns to find any trends, outliers, or prospective regions where timesaving is obvious. We will try to use machine learning models to anticipate how long particular steps will take based on various parameters, which can help us find the causes of time delays or savings.

RESOURCES:

The datasets are publicly available on GitHub(https://github.com/NAIST-SE/DevGPT/). We shall implement the code in **Python** programming language with its respective libraries on a Windows 11, 16 GB RAM and 512 GB SSD system and MacOS Big Sur Memory 8 GB RAM.

WORK PLAN:

Task	Deadline	Task Assigned to
Question 1	October 19th	JA23L
Question 2	October 23rd	VD23C
Question 3	October 27th	SE23L
Status report video presentation and report	October 31st (Actual deadline – November 05th)	JA23L, VD23C, SE23L
Final project report, code, and artifacts in GitHub and Canvas	November 24th (Actual deadline – November 30th)	JA23L, VD23C, SE23L
In-class project presentations	Nov 28, Nov 30, Dec 5, Dec 7	JA23L, VD23C, SE23L