Comprehensive Report on Sed-Puzzle Task

Task Overview

The task involved working with sed-puzzles, a type of transformation puzzle that requires applying a sequence of sed commands to achieve a desired output. The details regarding the individual tasks undertaken are documented in their respective files: task1, task2, task3, and task4.

Development of the Generator

A significant portion of the time spent on this project was dedicated to the development of the puzzle generator. The generator is now fully functional and capable of simulating any puzzle found on the sed-puzzle website. A novel approach was employed to make the generator highly efficient and adaptable to various puzzle formats. This work represents a major achievement, as it allows for extensive testing and exploration of different puzzle configurations with ease.

Challenges in AI Model Testing

Manual Testing Limitations

One of the major hurdles encountered was the need for manually entering prompts during testing. This made testing ChatGPT and Claude AI quite limited, as the process was time-consuming and could not be efficiently scaled. Due to this constraint, only a limited number of test cases could be evaluated for these models.

Large-Scale Testing with Gemini

In contrast, testing for the Gemini AI model was conducted through its API, which allowed for large-scale dataset evaluations. This approach enabled extensive benchmarking and provided valuable insights into the model's performance across different puzzle scenarios.

Issues with LLaMA 2

Setting up the LLaMA 2 model on my system was another time-intensive task. While the setup process was eventually completed, the model's execution speed proved to be extremely slow. This limitation hindered meaningful experimentation, and as a result, only a few prompts were run before deciding to move on to more efficient alternatives.

Discrepancies in the Validator

During the project, some discrepancies were discovered in the validator provided by the original repository.

Additional Materials

As part of this report, a paper presentation slide has been attached, summarizing the key aspects of the project paper given as part of the task.

Final Thoughts

Despite the challenges faced, this project was an incredibly engaging and rewarding experience. The development of the generator, testing across multiple AI models, and uncovering discrepancies in existing tools all contributed to a deep learning experience. Above all, I thoroughly enjoyed working on this task and look forward to further advancements in this domain.