A3 Reflection

In developing my self-deprecating math calculator bot, I encountered several challenges that have led to its failing or, more precisely, its miscalculations. The project's goal was to integrate the OpenAI API so that the bot (embodied as "Gio's Got Math") would perform iterative multiplications using python and then compare it to gpt4's language model, it provided an incorrect answer in a humorous manner. However, despite careful prompt engineering and client integration, there have been several points of failure that are worth discussing.

First, the inherent limitations of language models in handling precise arithmetic calculations have posed significant challenges. While GPT-4o and its variations excel at many language tasks, they were not meant to do precise numerical computations. I found that the model may include additional comments or fail to comply exactly to the required format compared to the python code, especially when the compounding grew exponentially. This produces inconsistent outputs that depart from the expected pattern.

Second, I found the openAI gpt model became worse the higher the iterations got. I believe it is because the model does not have many examples of multiplying big numbers like that. It seems to stop after 5-6 iterations then the numbers start to become off. This tells me models are not to the point of replacing mathematical programs.

Lastly, the challenge of enforcing a strict output format from a generative model proved non-trivial. The model occasionally output more verbose explanations or neglected to apply the intentional error logic as expected. This reflects the difficulty in controlling generative behavior when the task involves checking for inaccuracy. Overall, while the bot sometimes behaves as intended, providing wrong answers in a self-deprecating tone, it fails consistently due to both technical integration issues and the limitations of the language model in strictly controlled arithmetic tasks.