Self-Reflection

Assignment 3 CART 498
Aalok Sud
40214911

The mathematical comparison bot demonstrates several critical failure points when comparing calculations between Python and OpenAI's output, particularly when dealing with large numbers and iterative multiplication.

The most prominent failure pattern emerges around the 6th iteration when using base numbers like 2, 3, or 4. For example, with base number 2 and 15 iterations, the bot starts failing predictably at iteration 6, where:

- Python calculation: 18446744073709551616
- GPT calculation: 18446744065119617024

The failures can be categorized into several types:

1. **Precision Loss**: Starting from iteration 6, OpenAI's API begins to lose precision with large numbers. This is evident in how the GPT calculations start deviating from Python's exact calculations, initially by small margins but growing exponentially with each iteration.
2. **Complete Breakdown**: By iteration 11, the bot shows signs of complete numerical breakdown, where GPT starts returning nonsensical responses or fails to maintain proper numeric format. For example, in iteration 11, while Python calculates an exact value, GPT returns a heavily rounded number with many trailing zeros.

Why?
Most Likely because of Language Model’s prediction system rather than actually calculating.

We can see what the model is good at, i.e. word generation when it comes to comparing two numbers.

The Code:
The code has three functions, the python function that calculates the numbers using a for loop; the openAi gpt calculator which also does the calculations and the gpt bot that compares both results and generates a text output.