Assignment 3 Reflection: GPT, The Worst Mathematician Ever

For this assignment, I built a Python bot using the OpenAI API that repeatedly multiplies a number by itself for a chosen number of iterations. Instead of letting Python perform the math, I asked GPT-4.1-mini to generate each multiplication result and then compared it with the correct answer calculated in Python. I treated Python as a reliable "ground truth" to check whether the AI was right or wrong. Whenever the model made a mistake, the bot responded with a humorous self-deprecating message, giving it a personality that becomes more frustrated over time.

Since I am still relatively new to Python and generative AI, it took me some time at the beginning to understand how the API worked and how to structure the loop properly. After testing the bot with different numbers, I noticed that GPT does not always fail immediately. With smaller numbers, it often produces correct answers. However, as the numbers grow larger through repeated squaring, the model becomes more likely to make mistakes.

The main reason for these failures is that GPT is not actually performing calculations like a calculator. It is a language model that predicts the most likely sequence of tokens based on patterns from its training data. While this works well for generating text, it is less reliable for precise mathematical operations, especially with very large numbers. I also observed that once the model makes one incorrect calculation, the error can affect the following steps since each iteration depends on the previous result.

Overall, this assignment helped me better understand both the strengths and limitations of generative AI. Even though these models can sound confident, they are not always accurate, which is why verification is important when precision matters. Building this bot gave me a more realistic perspective on when AI can be useful and when traditional programming is the more dependable choice.