

### **Assignment 3: GPT-4o, the Worst Mathematician Ever**

JADEN THOMPSON

CART 498 C - Generative AI

February 5th, 2025

A short reflection (250-350 words) explaining why your bot is failing. If it fails at all!

For assignment 3, the bot seemingly does well enough until it reaches a certain point where the number of digits within the multiplication calculations becomes way too large to properly process. Usually beyond nine-digits of multiplication is when the bot typically fails and sends out the response that it has failed. LLMs such as ChatGPT are somewhat well known within the field of computing to struggle with large number computing and complex computational tasks, their focus within the models tend to operate well with what they are trained to perform such as coding and language processing. While they can still operate and do mathematical problems they were not specifically designed with this task in mind because standard computing algorithms are still the intended method of computational mathematics. Benchmarking LLMs with large-digit multiplication has been a useful method for testing the training data that they use because user input can be so varied while providing completely new test cases. LLMs often struggle to generalize beyond the data they have been trained on, if they have only been trained on nine-digit numbers then extrapolating to ten-digits is beyond their scope. As LLMs continue to evolve over time eventually these issues that they face with computational tasks will become less frequent and their accuracy will improve to compensate for large digit calculations.