Short Reflection

According to the three combinations, (4,6) (12,4) and (3,12) that broke the system. There are many reasons that could contribute to my bot failing. I am using GPT-4o-mini which is said to be generally less reliable than GPT-4o, which is a larger and more powerful model, trained on a more extensive dataset than the mini. According to OpenAI's evaluations, GPT-4o outperforms the GPT-4o-mini on various benchmarks, including math and reasoning tasks. Another potential cause of failure could be due to how Python and GPT-4o-mini handle very large numbers. There might be subtle differences in their internal representations or data types that are causing them to not equal each other. From my test results, I noticed that while the output from gpt-4o and the correct result look identical when printed, the integer converted result from the AI output might be exceeding the maximum value that a regular Python integer can hold which could be why unexpected behaviours or rounding errors happen. Additionally, one of the GPT-4o-mini responses that I received contained a random trailing period at the end of the numbers (e.g. 429981696.) which caused the int() conversion to fail, resulting in an incorrect result. I assume the reason for this is due to how I set the temperature and top_p. As I have set a higher temperature, there is more randomness in GPT-4o'-mini's responses, therefore, it is more likely that either the calculations are random or incorrect. Lastly, with the top_p set to 1.0, although that is the default setting, this allows the model to consider a wider variety of options, which could lead to a potentially less accurate response.