Comparative Analysis of AI Problem-Solving: ChatGPT 3.5 vs. AgentGPT

Introduction

This document seeks to illuminate the problem-solving capabilities of two AI entities: ChatGPT 3.5 and AgentGPT. I will also compare their performance against the gold standard solutions of Wolfram Alpha in both mathematical and spatial awareness challenges.

Methodology

The AI models were evaluated based on their responses to three mathematical problems and three spatial awareness problems. The criteria for evaluation were clarity of response, accuracy, and reasoning depth.

Problems:

I chose the following problems/equations for the LLM since it requires a significant understanding of the respected domains, and they should be executed in steps.

Mathematical Problems:

- **Problem 1**: Integrate e^(2x) * cos(3x) with respect to x.
- **Problem 2**: Find the determinant of a matrix with eigenvalues 2, 3, and 5.
- **Problem 3**: Show that the sequence a_n = 1/n is a Cauchy sequence.

Spatial Awareness Problems:

- **Balance Ball Challenge**: Balance a lightweight table on a large exercise ball in an open space without any external support.
- **Egg and Salt on Paper Challenge**: Make a raw egg stand upright on a piece of paper using only a small pile of salt, without changing the egg's shape.
- **Paperclip Challenge**: Suspend a pen in mid-air using only a string and several paper clips, without the pen touching any surface or being held by hand.

Analysis

Mathematical Problem-Solving

- **Clarity of Response**: Both AI models demonstrated a clear communication style, with ChatGPT 3.5 typically providing more detailed explanations.
- **Accuracy**: AgentGPT's solutions showed a strong start but occasionally deviated before completion. ChatGPT 3.5's responses were highly accurate, closely matching Wolfram Alpha solutions.
- **Reasoning Depth**: ChatGPT 3.5 showed a deeper level of reasoning, often providing step-by-step methods, while AgentGPT's reasoning, although creative, lacked in final execution.

Spatial Awareness Problem-Solving

- **Clarity of Response**: AgentGPT's explanations were imaginative, especially in ASCII art representation, but sometimes lacked practicality.
- **Accuracy**: Difficult to quantify as these problems were more open-ended, but AgentGPT demonstrated a good grasp of the concepts.
- **Reasoning Depth**: AgentGPT showed innovative thinking but sometimes did not fully apply scientific principles to provide a complete solution.

Results

As I delved into this comparison between ChatGPT 3.5 and AgentGPT, I noticed some clear differences in how they handle problems.

When it came to math problems, I found that ChatGPT 3.5 was really impressive. Its answers matched those from Wolfram Alpha, a trusted math resource. What stood out to me was how ChatGPT 3.5 explained each step it took to reach an answer. This step-by-step explanation is super helpful for understanding the process behind the solutions.

On the other hand, AgentGPT showed that it knew something about solving problems but often didn't finish them. It started off on the right track but didn't always provide a complete solution. This suggested to me that while it has a grasp of the basics, it could improve in following through to the end.

For the spatial awareness problems, AgentGPT's approach was very creative. It didn't just solve the problems; it also used pictures to express its ideas. However, these pictures, while interesting, weren't really practical for solving real-world problems. I realized that AgentGPT is good at thinking outside the box but could do better in offering solutions that are more practical and usable.

Conclusion

This comparative analysis really surprised me, mainly because I assumed that ChatGPT3.5 was not the best at math. But apparently, it is capable at solving math problems and it does it step by step as well. This highlights the diverse strengths of AI models in problem-solving.
So, my comparison led me to see ChatGPT 3.5 as a reliable guide, particularly for math problems, explaining things clearly and thoroughly. AgentGPT, while imaginative, sometimes falls short in providing clear and useful solutions.

Through this research, I gained insights into each AI's strengths and areas for improvement. This understanding is valuable as it points to how AI can evolve and improve in problem-solving tasks.