Evaluating the Planning Capabilities of Large Language Models: A Case Study in Study Plan Generation

The point of the study

This study explores how Large Language Models (LLMs)—such as GPT-4 (the model behind ChatGPT)—can be used for planning, with study planning as a real-world test case.

Research suggests that while LLMs can generate plausible-sounding plans, they often struggle to create truly realistic and effective ones. Automated Planning, a long-standing area of AI, is traditionally handled by specialized systems known as planners (such as Fast Downward). However, these require experts to manually define the problem domain, possible actions, and desired outcomes—a complex and rigid process.

LLMs, on the other hand, are flexible and can generate human-like responses. This study aims to explore whether LLMs alone, LLMs combined with traditional planners, or LLMs optimized through prompt engineering can create better, more adaptable study plans.

What you'll be doing

You'll interact with the system (Athena) by providing a study-related prompt. The system will then generate three study plans—each created using a different method:

1. A basic LLM-generated plan
2. A plan generated using an LLM and a traditional planner
3. A plan generated by an LLM optimized using prompt engineering

To keep the study unbiased, these plans have been assigned neutral names—you won't know which method produced which plan. Your task is simply to review them and evaluate their quality based on different factors.

Thank you so much for participating! Your insights will help us better understand how AI can be leveraged for planning. I hope you find the study both interesting and enjoyable! 😊

Things to Keep in Mind

- I've kept the design very simple to ensure it doesn't affect your perception of the plans or the system. Kindly evaluate the plans itself, not the design.
- You can adjust the generated plans by prompting the continuously prompting the system.
- If you encounter any issues, kindly refresh, or contact me at sln2000@hw.ac.uk