Exploring the Boundaries of ChatGPT in Scientific Inquiry Through Prompting

Katie Hammer

This project delves into leveraging ChatGPT to enhance scientific inquiries through strategic prompt engineering. Developed by OpenAI, ChatGPT simulates human conversation and efficiently synthesizes vast amounts of data, making it a valuable asset for research. While it doesn't independently generate new ideas, its strength lies in combining extensive knowledge to offer unique insights and solutions.

This study examines its applications in computer science, focusing on anthropomorphism, error message analysis, cellular automata, and code translation. By tapping into ChatGPT's capabilities, we aim to push the boundaries of computational problem-solving, promising substantial contributions to future scientific discoveries.

Anthropomorphism

Background

Anthropomorphism, attributing human traits to AI, enhances user interactions through natural language processing (NLP). ChatGPT uses sentiment analysis to detect emotional tones, learns from extensive datasets, and generates varied responses to mimic human-like interactions[1].

Methods

Using a trial-and-error approach, I prompted ChatGPT on various topics about itself, such as its humanity, religious beliefs, and intelligence. Consistent responses like "As an AI..." maintained objectivity. I also created a custom GPT with personal information and specific communication preferences, adding descriptions, instructions, and conversation starters.

Results

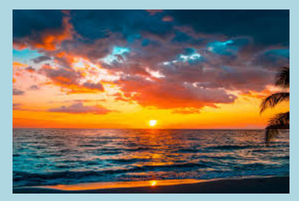
ChatGPT's anthropomorphic features make interactions more intuitive. Custom APIs improved its ability to maintain context, showing potential for personalized AI assistants. ChatGPT also effectively interprets and responds to images, associating visual elements with emotions.

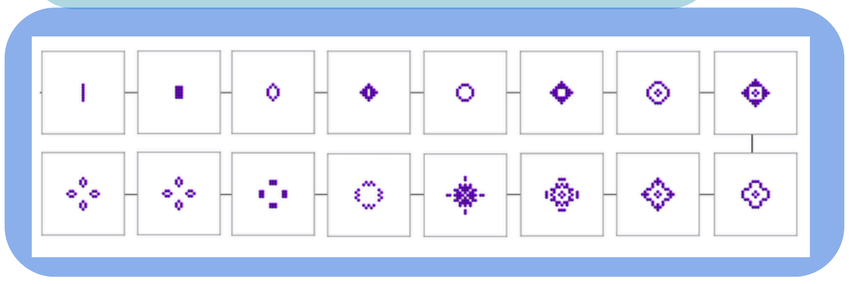
Suggests melancholy or introspection

Conveys relaxation and comfort

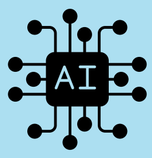
Evokes calmness and serenity

Represents energy and activity





Conway's Game of Life



Background

Conway's Game of Life is a cellular automaton played on a grid where cells are either "dead" or "alive." The state of each cell is determined by its eight neighbors following four simple rules [2].

Methods

I tested ChatGPT's ability to generate patterns using strategies like random generation, modifying known patterns, combining patterns, exploring symmetry, and edge conditions. Prompts included requests for generating specific grid patterns and observing their evolution.

Results

ChatGPT successfully generated Python code for various patterns and provided instructions for running the code. This demonstrated its efficiency and utility in computational tasks, despite the game's complexity.

Error code messages with one-shot prompting

Background

This study investigates improving Python 2 compiler error messages using AI. We used high school students' coding attempts with single errors, focusing on snippets below 20 lines where the student successfully fixed the error, to see if AI could generate better error messages [4].

Methods

We selected 100 code pairs, using 40 as a control group with a standard system message. For the other 60, we crafted "ideal" error messages. ChatGPT 3.5-turbo-1106 in OpenAI's playground was used for one-shot prompting. We rated the error messages on their helpfulness and extra information provided. Three evaluators independently rated the control, one-shot prompted, and fine-tuned examples.

Results

Preliminary observations suggest that fine-tuning might yield better error messages due to more examples. One-shot prompting sometimes caused confusion when examples differed from the test error. Final results are pending evaluation.

Code Translation

Coding can be time-consuming, but ChatGPT, with knowledge of over 38 programming languages, can quickly generate code when prompted correctly. By using specific tasks, providing context, and including error handling, I found that ChatGPT could produce accurate and functional code with clear setup instructions [3]. This highlights the AI's potential to significantly speed up the coding process while maintaining high quality.

Extrapolation and Prediction (2023)

Metric	Perdicted	Actual
Grade 3+ Hurricanes	5-8	3
Amazon Deforestation (sq km)	7,000-9,000	9,117
Average Global Temperature (°C)	14.8	15.08
Inflation Rate (%)	5.5-6.5	6.3
Glacier Melting (% of volume)	2-4	6.2
CO_2 Emissions (billion metric tons)	25-27	37.4

References & Acknowledgements

This research was done under the supervision of Dr. Brett Becker, in addition to his PhD student Eddie Santos.

[1] Levy, D. M., & Ascarza, E. (2024). How Information Affects Choice: An Experimental Study of News and Recommendation Algorithms. SSRN.

[2] MIT. (2010). The Game of Life Slides.

[3] Ilhan, A., & Parkes, D. C. (2022). Decision-Aware Importance Sampling for Learning Causal Graphs.

[4] Roe, M. (2024). Deep Reinforcement Learning for Dynamic Resource Allocation in 5G Networks.

[1] Levy, D. M., & Ascarza, E. (2024). How Information Affects Choice: An Experimental Study of News and Recommendation Algorithms. SSRN.