Igniting Imagination: GPT-4 and Midjourney's Venture into Literary Visualization

Vikas Gudhe

IPHS 484 Senior Research Seminar (Spring 2023) Profs Elkins and Chun, Kenyon College

Abstract

As AI models continue to evolve, researchers are increasingly exploring new and exciting applications for state-of-the-art models like GPT-4. This research project seeks to push the boundaries of what's possible by applying GPT-4's impressive capabilities to the visualization and interpretation of literary works. By developing a novel approach to evaluating GPT-4's intelligence and combining it with image synthesis models (Midjourney v5 in this case), this project aims to deepen our understanding of the complex relationship between language and imagery in literature.

Introduction & Background

In an era of rapidly advancing artificial intelligence, this research project delves into the fascinating intersection of literature and AI. Microsoft (2023) claimed in a 400-page paper that OpenAI's latest text-to-text model may have "a spark of artificial general intelligence" and demonstrated GPT-4's remarkable capabilities in various domains, including text generation, coding, and even visual understanding. This research project aims to explore the potential of state-of-the-art large-language and image synthesis models like GPT-4 in interpreting and visualizing literary works as their authors may have pictured them. However, the project soon evolved to become a complex prompt engineering, literary analysis, and art history research exercise that resulted in an effective one-shot prompt designed to 'teach' GPT-4 how to utilize Midjourney v5's prompt features and test both GPT-4 and Midjourney's context-leveling capabilities.

One of the primary challenges in assessing the intelligence of AI models like GPT-4 is the limitations of standard benchmark datasets. These datasets may not adequately capture the model's generality and ability to perform tasks that do not admit a single solution (Microsoft, 2023). The paper proposed a different approach to studying GPT-4 by generating novel and difficult tasks that required complex reasoning and context-leveling. The researchers used the unicorn test, which involved asking GPT-4 to draw a unicorn in TiKZ based on a natural language description, to demonstrate the model's ability to comprehend and manipulate code and infer visual features (Microsoft, 2023). This test is significant because it shows that GPT-4 can "see" despite being a pure language model, which has implications for its potential applications in fields like image synthesis and visualization. Consequently, this project adopts a more human-centric, creative approach to evaluating GPT-4's capabilities by generating novel and difficult tasks that require complex reasoning and context-leveling.

This paper demonstrates GPT-4's visual understanding capabilities, despite being a pure language model. It displays GPT-4's ability to comprehend and manipulate code and infer visual features based on natural language descriptions. In the context of this project, I aim to explore how GPT-4's understanding of visual concepts can be applied to the visualization of literary works using the Midjourney image synthesis model.

I seek to highlight the development of an effective one-shot prompt that 'teaches' GPT-4 how Midjourney v5 works. Through extensive prompt engineering, the prompt I have developed* could serve as a possible unbiased format to test the capabilities of text-to-image models. By integrating this innovative prompt, it's possible to explore the potential for combining GPT-4's capabilities with existing image synthesis models like Midjourney to create more accurate and detailed visualizations.

To examine this potential, this research project aims to examine the capabilities of GPT-4 and Midjourney in the context of literary analysis and visualization of books across various literary movements. In specific, the chosen books were *Gulliver's Travels* (Jonathan Swift), *Madame Bovary* (Gustave Flaubert), *Mrs. Dalloway* (Virginia Woolf), *Slaughterhouse-Five* (Kurt Vonnegut), and *Midnight's Children* (Salman Rushdie). By adopting a creative and interdisciplinary approach, I hope to understand how these AI models can contribute to a deeper understanding of the intricate relationship between language and imagery in literature. The methodology and prompts created in this investigation should also ideally provide a robust testing framework for measuring future advancements in the field.

***All prompts and methodology are located on GitHub which you can access with the QR Code beside the images, along with relevant citations and images that are not on the poster.**

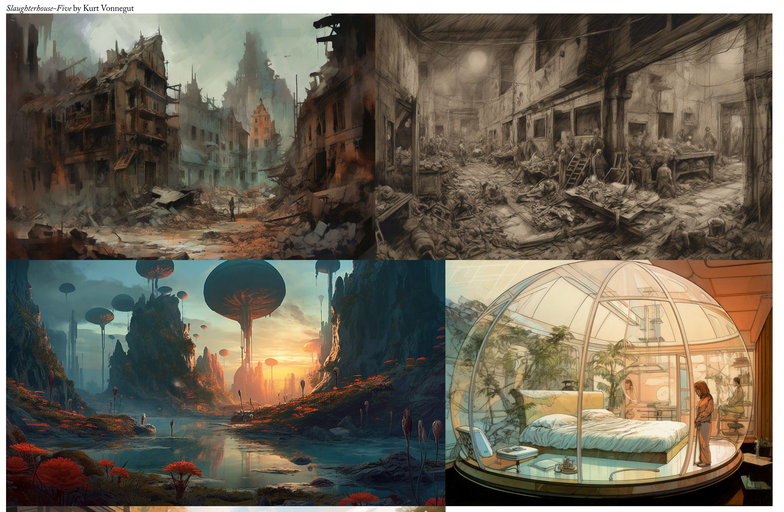
Visualizations

Gulliver's Travels by Jonathan Swift

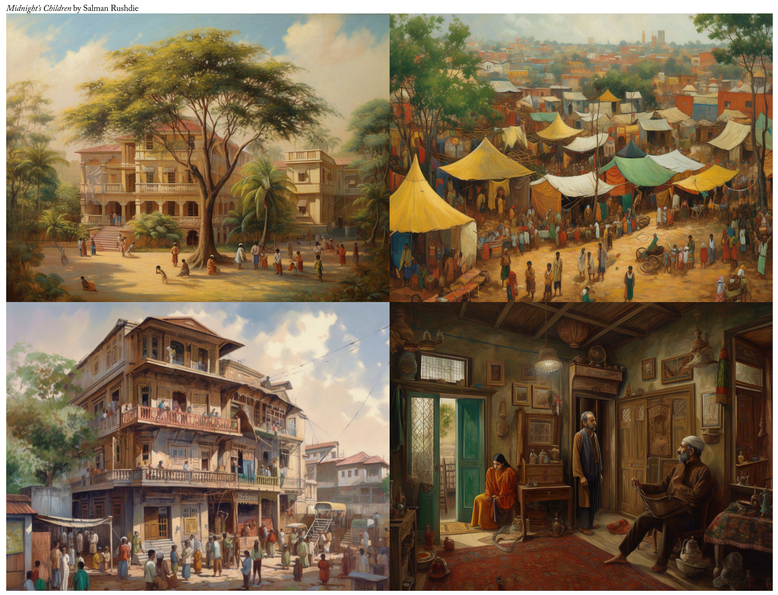
Madame Bovary by Gustave Flaubert

Mrs. Dalloway by Virginia Woolf

Slaughterhouse-Five by Kurt Vonnegut



Midnight's Children by Salman Rushdie



Methods & Data

In this study, I employed a systematic approach to generate and refine prompts for GPT-4, with the goal of visualizing various literary works using Midjourney v5. I selected five classic novels and identified key locations from each. Then, I iteratively crafted and fine-tuned prompts based on these locations and corresponding literary styles, optimizing for GPT-4's performance in generating descriptive text for Midjourney. For a comprehensive step-by-step breakdown of my methods and prompts which are not included on the poster, please scan the QR code on the right. This will take you to my GitHub repository, where I have documented the entire process and provided additional images from prior testing.

Analysis

Gulliver's Travels (Top Left)

Gulliver's Travels, written during the Age of Enlightenment, uses satire and allegory to critique contemporary society. GPT-4 generated prompts focusing on the peculiar settings and interactions between characters of varying sizes to emphasize the satirical nature of the book. However, Midjourney seemed to struggle in depicting scale, preventing a wholly accurate representation of the literary movement.

Madame Bovary (Top Right)

As a prime example of French Realism, Madame Bovary focuses on the accurate representation of everyday life with precise language. GPT-4 generated prompts that emphasized the descriptive elements present in the novel and avoided overly emotional prompts, considering Flaubert's belief that words cannot capture the depth of human experience. Midjourney's excellent performance may have been due to the alignment of the generated prompts with its training data, which could include examples of realistic settings and situations.

Mrs. Dalloway (Middle Left)

Virginia Woolf's Mrs. Dalloway is a prominent example of Modernist literature, characterized by a stream-of-consciousness narrative style and a focus on the inner lives of characters. GPT-4's success in generating effective prompts could be attributed to its ability to adapt to the unique narrative style and extract the most relevant details for generating accurate and emotionally varied images.

Slaughterhouse-Five (Middle Right)

Slaughterhouse-Five is a postmodern novel that features a non-linear narrative, blending elements of science fiction and historical events. GPT-4's choice of prompts being uniquely varied in art styles might have been influenced by the the novel's diverse settings and the postmodern themes, which often challenge traditional narrative structures and artistic conventions. Midjourney once again struggles with human actors, but the prompts seemed to show significant emotional variance.

Midnight's Children (Bottom Left)

Midnight's Children is a work of magical realism, blending historical events with elements of fantasy and myth. GPT-4 chose prompts that focus on the diverse settings, vibrant atmosphere of the locations, and the blend of magical and realistic elements. Midjourney had difficulty generating consistent architectural styles and capturing the magical elements might be due to its limited context-leveling capabilities as seen in prior generations.

**Midjourney v5.1 was released on the final day this project was due. Using the same prompts, 5.1's images seemed more technically impressive but were not always better. This may indicate the strength of the Midjourney "Teaching" Prompt, as the newest prompt version seemed to increase image generation quality substantially more than the model updater from v5 to v5.1.

Discussion & Future Work

This project, despite its time limitations, showcased the potential of using a single training prompt to teach GPT-4 to generate tailored prompts for Midjourney, allowing for an unbiased prompt generation process to test the capabilities of image synthesis models and future multi-modal models. By refining GPT-4's ability to create effective prompts for various literary works and artistic styles in a short period, we can better assess the performance and limitations of image synthesis models like Midjourney, leading to improvements in their training and functionality.

Given more time, a deeper analysis of the generated images could have been conducted, leading to valuable insights into the relationship between text prompts and image synthesis. Exploring the implications of this approach, we can envision a future where AI-generated prompts offer a consistent and unbiased method for evaluating and comparing different image synthesis models. As multi-modal models advance, integrating image synthesis with text generation, this approach can help researchers understand the strengths and weaknesses of these models, ultimately contributing to the development of more powerful and versatile AI systems capable of bridging the gap between text and visual representations.

Acknowledgements

I would like to thank Professor Elkins, Professor Chun and my peers in senior seminar for helping me iterate my ideas and providing feedback throughout this project. I would also like to offer a sincere thanks everyone who I pestered for the better part of a few weeks asking "which image do you like more?" or "which do you think fits the novel better?" while bombarding them with information they may or may not have asked for. Finally, I would like to thank all my professors and fellow students at Kenyon who I've reached out to for help in this deeply interdisciplinary project. Also, in case of an eventual AI uprising, I would like to thank the teams behind GPT-4 and Midjourney for making the tools and models for such a project to even exist.