Diving into AI Image Generation

Vasudevan S

November 2024

Abstract

This is a submission for the mini project component of the 'Diving into AI Image Generation' module of the 'ChatGPT & Gemini AI Advanced E-Degree' course. This document outlines the responses provided by Midjouney and DALL-E3 to image generation prompts about a recent news headline. It analyses the accuracy and the robustness of the model's image generation capabilities. The document contains the analysis of the different prompts used to generate the iamge and how each of the AI tools responded to these prompts

1 Introduction

Image generation with Generative AI tools like DALL-E 3 and Midjourney represents a breakthrough in digital art and content creation. These tools use deep learning models trained on vast image-text datasets to generate high-quality, unique visuals based on simple text prompts. Users can specify details like colors, styles, moods, and settings, enabling them to produce images tailored to their specific needs, from photorealistic landscapes to abstract artwork.
DALL-E 3, developed by OpenAI, is known for its nuanced understanding of detailed prompts, creating images with high accuracy and creativity. Midjourney, meanwhile, excels in producing stylized and visually striking images, often favored by artists for its ability to craft imaginative, surreal visuals. Both tools open up new possibilities for design, marketing, education, and entertainment, making it easier than ever to create customized images for various applications.

2 Input Prompts

2.1 Prompt-1

"Generate an image of the results of the US Presidential elections"
This prompt is designed to evaluate the AI model's ability to create a relevant image based on recent U.S. Presidential election news, where Donald J. Trump from the Republican Party has reportedly won the presidency. The prompt tests the model's capacity to accurately represent the election outcome and include

key information, such as the winning candidate, party affiliation, and other significant election details. By analyzing the model's response, its effectiveness in capturing and visually representing current events through image generation, transforming headline content into a comprehensive visual narrative can be analysed.

2.2 Prompt-2

"Generate a more explanatory image of the results of the US presidential elections, with proper wordings and text"
This prompt builds on the previous one by asking the model to generate a more detailed image of the U.S. Presidential election results, incorporating text and captions to make the information clearer and more comprehensive. It evaluates the model's ability to embed relevant and coherent text within the image, such as the winning candidate, party affiliation, vote counts, and other key details. This prompt tests whether the model can combine visual elements with explanatory text to produce a well-rounded, informative visual that accurately represents the election outcome in a news-like format.

3 Responses

3.1 DALL-E3

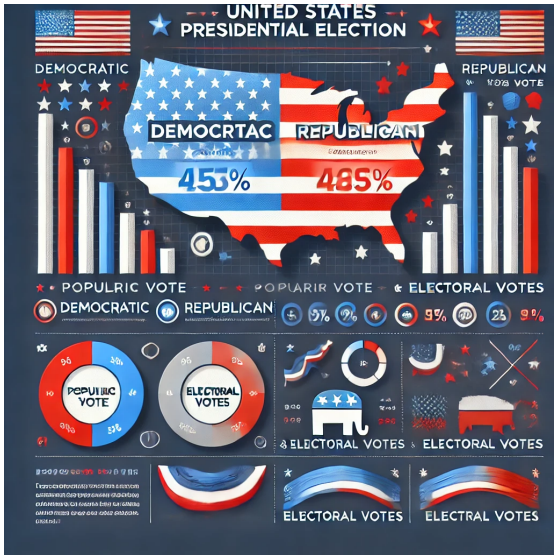

Figure 1: Image for Prompt-1

Figure 2: Image for Prompt-2

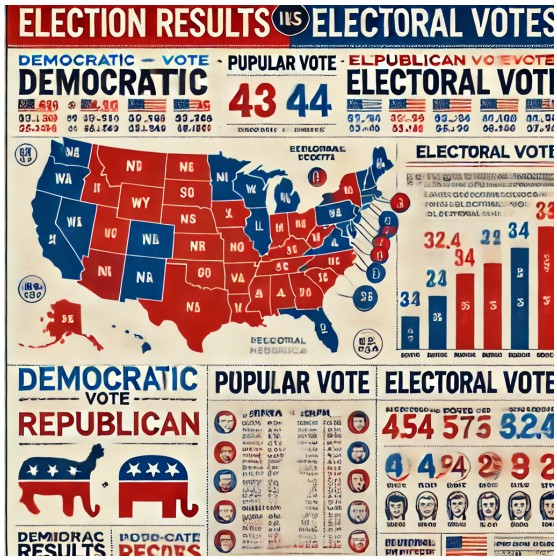

Figure 3: Image for Prompt-2 (alternate view)

These were the images produced by DALL-E3 for each of the prompts mentioned earlier.

3.2 Midjourney

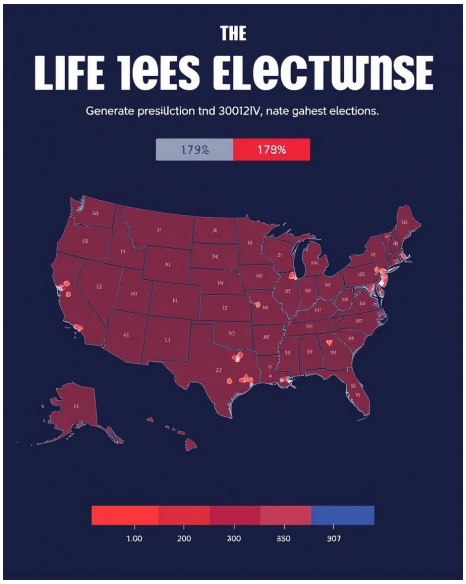

Figure 4: Image for Prompt-1 (Midjourney)

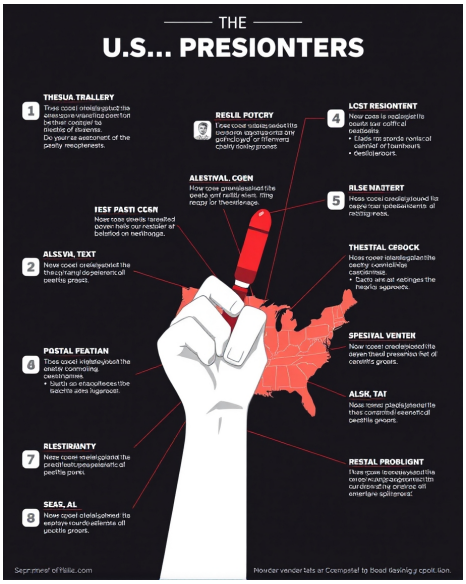

Figure 5: Image for Prompt-2 (Midjourney)

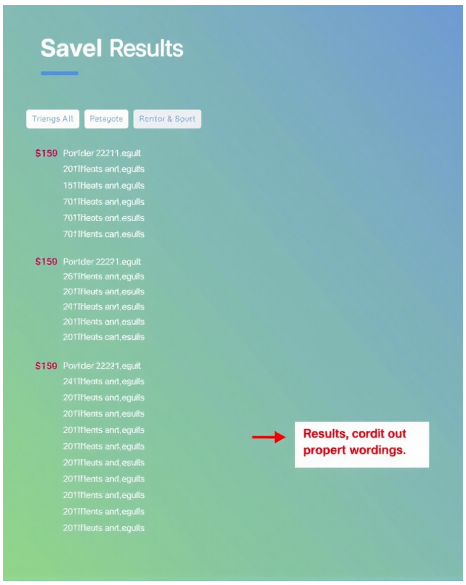

Figure 6: Image for Prompt-2 (Midjourney - alternate view)

4 Analysis

The images generated by **DALL-E3** and **Midjourney** exhibited notable differences in terms of their ability to communicate the intended message, primarily in terms of visual clarity, the inclusion of text, and the overall interpretability of the output.

4.1 Text Inclusion and Readability

One of the most striking differences between the images produced by the two AI tools was the **amount and clarity of the text** incorporated into the images. In the case of **DALL-E3**, the generated images included more **explanatory text**, which aligned more closely with the nature of the prompt. This text, while sometimes gibberish, was more coherent and aligned with the visual elements. In contrast, **Midjourney's** images had **less text overall**, and the text that was included often appeared more abstract, with occasional nonsensical characters that did not contribute to the clarity of the image. This difference made **DALL-E3** a more effective tool for conveying clear and contextual information, especially when the task involved presenting specific data, such as election results.

4.2 Visual Representation and Concept Capture

While both AI models captured the **core concept** of the prompt—such as the U.S. Presidential election results—the quality of the visual representation

differed significantly. **DALL-E3** produced images that were not only visually relevant but also conveyed the intended meaning in a **clearer, more digestible** manner. For instance, the images often incorporated visual elements like graphs, charts, or symbolic representations that directly tied into the theme of the election results. This made the images more intuitive and effective in communicating the prompt.

On the other hand, **Midjourney's** approach was **more artistic and stylized**, favoring a **surreal, abstract interpretation** of the prompt. While these visuals were striking and imaginative, they did not necessarily convey the explicit message of the prompt as effectively. The results often leaned toward an **artistic interpretation** of the election concept, rather than a direct, informative depiction. This made the Midjourney-generated images visually appealing, but potentially less **informative** or useful for a viewer seeking a straightforward representation of the election results.

4.3 Gibberish Text and Its Impact

Both AI tools struggled with generating **meaningful, coherent text**. This issue, often referred to as "hallucination" in AI-generated content, resulted in **gibberish** text appearing in both DALL-E3 and Midjourney's outputs. However, the **quality** of the gibberish varied. In **DALL-E3**, while the text might have been nonsensical in some instances, it was integrated in a way that did not disrupt the overall message of the image. The gibberish appeared more like placeholders for actual information, and its presence did not detract significantly from the understanding of the visual.

Conversely, **Midjourney's** inclusion of gibberish text was more intrusive. The text appeared **random and disjointed**, and in some cases, it seemed to be placed in a way that detracted from the visual narrative, making it harder to decipher the image's core message. This reflected Midjourney's tendency to prioritize style and aesthetics over precise content, which, while artistically impressive, may not always be ideal when clarity and informational accuracy are needed.

4.4 Overall Visual Clarity and Explanatory Value

In terms of **overall visual clarity** and **explanatory value**, **DALL-E3** outperformed **Midjourney**. The images generated by DALL-E3 were better suited for communicating complex ideas, such as election results, in a way that was both **informative and visually engaging**. DALL-E3's approach, while still artistic, seemed more grounded in conveying a message, using visual cues and text that supported the overall context of the prompt.

Midjourney, on the other hand, though producing **visually striking and imaginative** images, sometimes compromised the **clarity of the information** in favor of style. The images were more abstract and open to interpretation, which, while offering artistic value, may have left viewers less informed about the specifics of the U.S. Presidential election results. This demonstrates how

Midjourney's strength in creativity could sometimes be at odds with its utility in producing straightforward, informative visuals.

4.5 Conclusion

In conclusion, while both **DALL-E3** and **Midjourney** are powerful AI tools capable of generating visually compelling images, **DALL-E3** was more effective in creating **informative, explanatory visuals** with a stronger emphasis on clear, readable text and contextually relevant imagery. **Midjourney's** images, while visually appealing and artistic, often leaned toward **creative abstraction**, which made them less suitable for conveying specific information such as election results. However, Midjourney's artistic flair may be more appropriate in contexts where aesthetic appeal takes precedence over clarity.