

This Document is prepared to demonstrate the relation between prompt, enhanced text and image that has been generated. In this document general results and assumptions are discussed and examined. Examples will be provided, and finally, general findings and insights will be presented by me.

Model Used For Generating the Enhance Text: GPT-3.5 Turbo (OpenAI API)

Models Used For Generating the image: stable-diffusion-v1-5 , stable-diffusion-2-1 base (Hugging Face Diffusion Library)

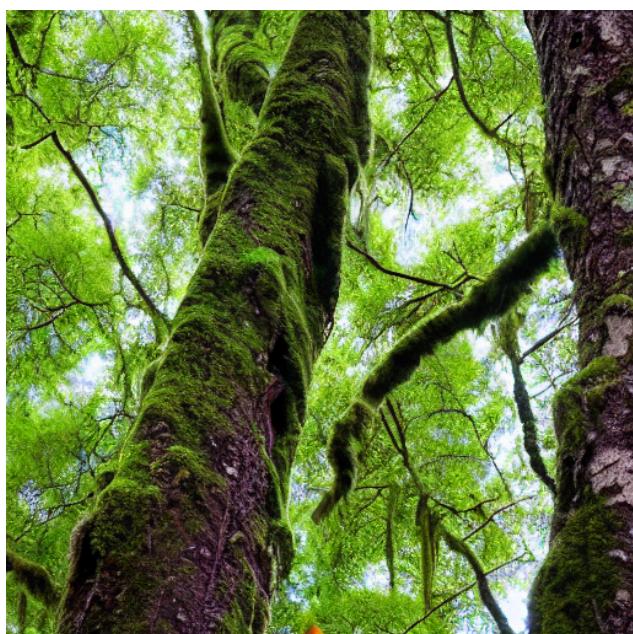
Two different Diffusion models used to compare the images generated from the same enhanced text. The models are run in Google Colab to generate faster outputs via T4 GPU

Examples

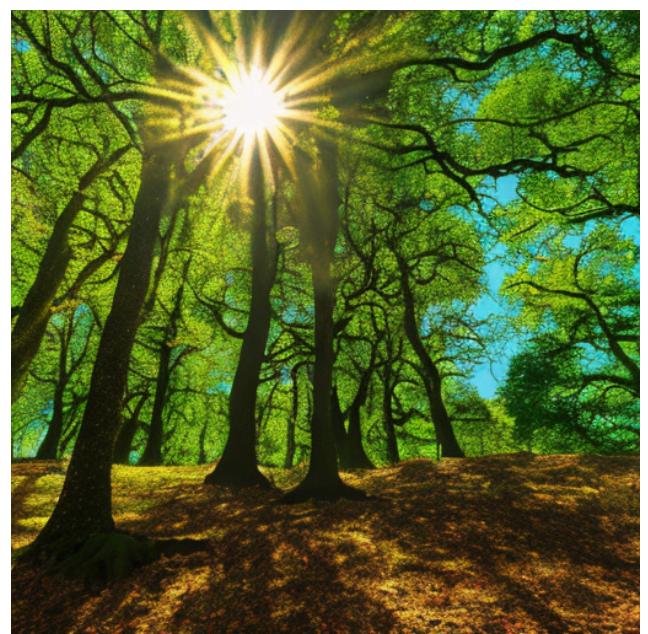
1.Tree

User Input: A tree in the forest

Enhanced text for 20 years Later :In the dappled sunlight of the forest, a majestic oak tree stands tall, its gnarled branches reaching towards the sky. Vibrant emerald leaves flutter in the gentle breeze, casting intricate shadows on the forest floor. Moss blankets the ancient bark, a tapestry of life woven over decades.



Stable Diffusion-v1-5



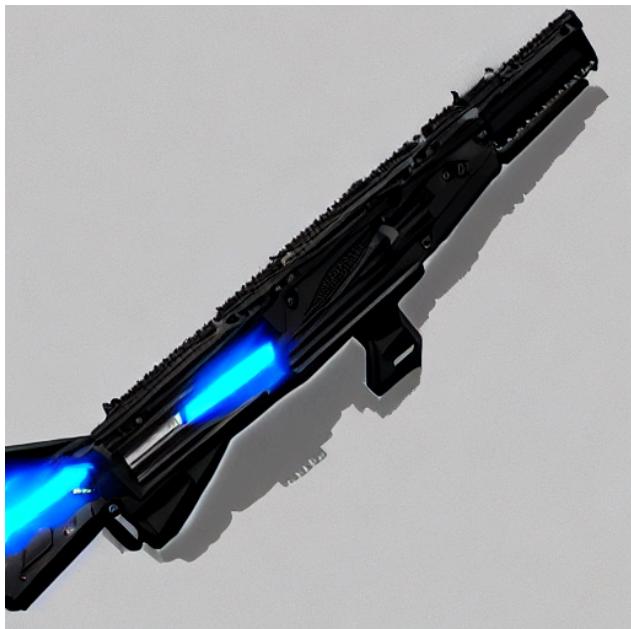
Stable Diffusion-2-1-base

2. Weapon

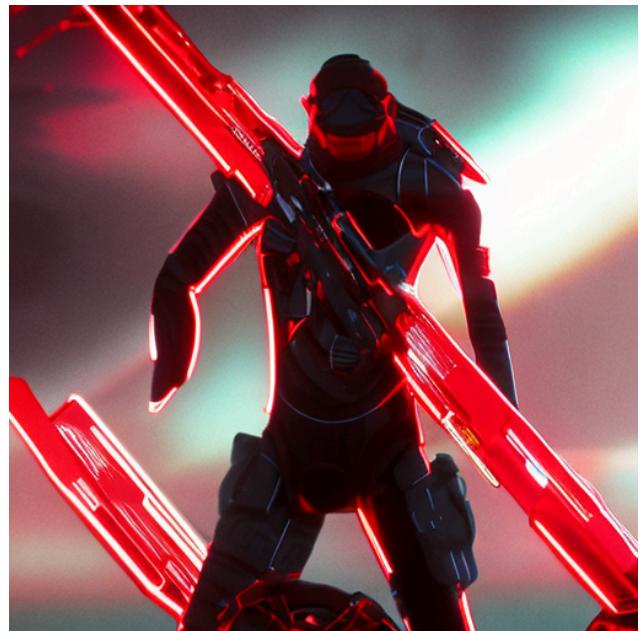
User Input: Weapon

Enhanced Text for 20 Years Later: In 20 years, weapons may feature sleek, ergonomic designs with advanced holographic displays and integrated smart technology. The exterior could be constructed from lightweight, durable materials, adorned with intricate geometric patterns and glowing accents. A touch of elegance fused with deadly efficiency.

Images Generated:



Stable Diffusion-v1-5



Stable Diffusion-2-1-base

3. New York City

User Input: New York City at Night

Enhanced Text for 20 Years Later from Now: In the year 2041, New York City at night glows like a neon dream. Skyscrapers reach toward the starlit sky, adorned with holographic billboards that paint the streets in a kaleidoscope of colors. The air hums with the buzz of hovercars, casting dynamic shadows on the glistening sidewalks below.

Images Generated:



Stable Diffusion-v1-5



Stable Diffusion-2-1-base

4. Human On Mars Surface

User Input: A human On Mars Surface

Enhanced Text For 20 Years Later: A lone human stands on the rust-red surface of Mars, bathed in the soft glow of two setting suns. Their silver suit reflects the warm hues of the rocky terrain, casting long shadows. A gentle breeze stirs the dust, creating a mesmerizing dance of light and shadow against the backdrop of a distant, barren landscape.



Stable Diffusion-v1-5



Stable Diffusion-2-1-base

5. An Iphone

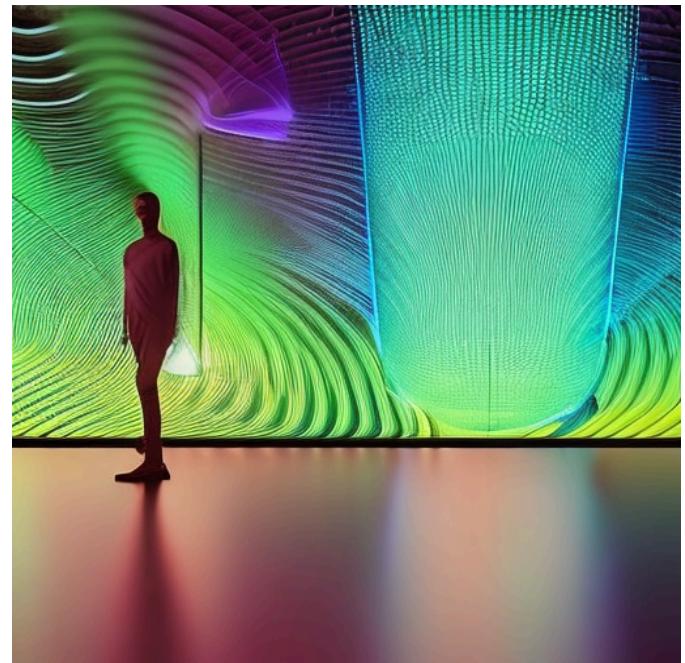
User Input: An Iphone

Enhanced Text For 20 Years Later: In a dimly lit room, a holographic iPhone hovers, emitting a soft glow of iridescent colors. Its sleek surface shimmers with a shifting pattern of ethereal light, while intricate neural interfaces pulse beneath the translucent casing. The atmosphere is charged with a sense of futuristic elegance and technological mystique.

Images Generated:



Stable Diffusion-v1-5



Stable Diffusion-2-1-base

6. Coca-Cola

Enhanced Text For 20 Years Later: The year is 2041. A holographic Coca-Cola logo shimmers above a sleek, transparent bottle filled with effervescent liquid that glows with a neon blue hue. Surrounded by a bustling cityscape, the drink exudes a futuristic aura, reflecting the vibrant lights and energy of a technologically advanced world.

Images Generated:



Stable Diffusion-v1-5



Stable Diffusion-2-1-base

Conclusions

1. It was observed that when the same enhanced text was used, sometimes the generated images closely matched the intended concept, while other times irrelevant and unsatisfying visuals were produced based solely on keywords from the enhanced text.
2. In most cases, the enhanced texts were found to be logically consistent. However, despite detailed descriptions and clear instructions, the resulting images often failed to meet expectations.
3. It was noted that the term "sleek" appeared frequently in the enhanced texts. To prevent this and generate more diverse descriptions, the `temperature` and `top_p` parameters were adjusted. However, as the outputs became too extreme and inconsistent, default values were retained.
4. In the enhanced texts, futuristic concepts like "holographic" and "lightning" were frequently associated with the idea of something 20 years into the future.
5. Even for abstract concepts such as selfishness or loneliness, the generated images often oddly featured holographic displays and neon lighting, regardless of their relevance.
6. The same enhanced texts were also input into ChatGPT-4 Turbo, and image generation was attempted. The results showed that GPT-4 Turbo produced far more detailed and rich descriptions compared to outputs from Stable Diffusion models.

To conclude Everything That Has Been Stated So Far:

1. The system failed to generate meaningful results for abstract concepts (e.g., loneliness, selfishness).
2. System failed to produce clear writings.
3. Although many possibilities could be imagined for an object 20 years from now, the system frequently relied on clichés like neon lights and holograms, which led to a loss of realism in some cases.

4. There is no access to the previously generated images. A data structure (possibly a FILO stack) could have been used to store previous images, prompts, and enhanced texts.
5. By enabling user authentication, each user could have a unique session which may allow them to keep track of their conversations and access their previously generated images.