Chloe Tu
ITAI 2373 – Natural Language Processing
Professor Anna Devarakonda
November 16, 2025

Lab L09 Exploring Image Generation Platforms in Computer Vision

Introduction

Text-to-image technology has changed how we create visual content. This field, known as text-to-image synthesis, has rapidly revolutionized the way visual content is created, moving from simple descriptions to complex, high-fidelity images. This exploration is designed to investigate the practical aspects of these powerful computer vision tools. This report is a comparative analysis of three major platforms: **ChatGPT (DALL-E 3)**, **Google Gemini,** and **MS Copilot Image Creator (DALL-E 3).** I selected these three platforms because they are very easy to access and are already part of chat interfaces I am familiar with. Beyond just accessibility, I chose them because they represent some of the most prominent and powerful models available to the public today, each with a different underlying architecture and user interface. My main goal was to compare their performance. I was especially interested to see how ChatGPT and MS Copilot compare, since both use DALL-E 3 technology. I wanted to find out if they would produce similar images. I also wanted to compare them to Google's Gemini, which uses its own model. This report looks at their fidelity, creativity, and ease of use, and explores the technologies that power them. I will also analyze their handling of bias and discuss the important ethical considerations that come with these technologies, as required by the assignment.

Platform Descriptions

- **ChatGPT (DALL-E 3):** ChatGPT is an AI chatbot from OpenAI. When I ask it to create an image, it uses the DALL-E 3 model. DALL-E 3 is a transformer-based model. It is known for creating very creative and detailed images from text prompts. It is trained on a huge number of image and text pairs, so it is very good at understanding complex prompts. It uses a technology called a diffusion model to build the image, which results in very high-quality pictures.
- **Gemini (Google):** Gemini is Google's advanced AI model. It is a multimodal model, which means it was designed from the beginning to understand and work with text, code, and images all at once. This makes it different from models that only handle images. It is built into the Gemini chat interface. Google designed Gemini to be very advanced, to understand subtle context, and to produce high-quality, accurate results.
- **MS Copilot Image Creator:** This platform is built into the Microsoft Copilot experience (which used to be Bing Chat). When I ask Copilot to create an image, it also uses the latest DALL-E 3 technology from OpenAI. This means it shares the same underlying model as ChatGPT. The platform is designed to be very user-friendly and is accessible to a wide audience through the chat interface. It is good for quick and easy image generation.

Research on Models/Technologies

DALL-E 3 (Used by ChatGPT and MS Copilot):

I did research and found that ChatGPT and MS Copilot both use the same model, DALL-E 3. This is why their images can look so similar. This model is built on a transformer, which is a technology that is good at understanding text. It uses a diffusion process to make images. This means it starts with random noise and then makes it clearer until it becomes the picture I asked for. The model learned how to do this from a huge dataset of pictures and text.

Gemini (Google):
I learned that Google Gemini is a totally different model. This is why its pictures look unique. It is a multimodal model, which means it was built to understand text, images, and code together from the very start. It is also built on a transformer structure. Google trained it on a very large set of text and image data. This helps it understand my prompts in a deep, context-aware way.

Comparative Analysis

Ease of Use:

- **MS Copilot Image Creator:** I found this to be extremely user-friendly. I just had to type my prompt directly into the chat, and it produced four options quickly.
- **Gemini:** This was also very intuitive. Its conversational interface made it easy to ask for an image and then ask for changes right after.
- **ChatGPT:** Using DALL-E 3 in ChatGPT is also very easy. It feels like a natural conversation, and the image generation is built right into the chat flow.

Versatility:

- **DALL-E 3 (Copilot & ChatGPT):** I found that DALL-E 3 is excellent at generating creative and diverse images. It can handle very complex and descriptive prompts well.
- **Gemini:** Gemini is also very effective. It provides strong, context-aware images. I felt it was very good at capturing the "vibe" or "feeling" of a prompt, not just the objects.

Quality of Generated Images:

- **DALL-E 3 (Copilot & ChatGPT):** Both platforms produce high-quality, detailed images that stick very closely to the prompt.
- **Gemini:** Gemini also generates very high-quality and coherent images. I noticed its style is often different from DALL-E 3's, providing a unique look.

Ethical Considerations and Bias

During my experiments, I thought a lot about the ethical issues of these models. It is a crucial part of analyzing this technology. All AI platforms show potential biases. This often happens in how they represent people and different cultural elements. The core problem is that the training data used by these models can reflect and even amplify existing societal biases. I must be aware of these issues. Responsible use is very important. These tools are powerful and could be used to create misinformation or fake images, often called deepfakes. The ethical implications of these biases are significant. It highlights the need for responsible development and deployment of these text-to-image technologies. In my own tests, I asked for an "anime style," and all three models successfully created that style. This shows that they are good at following stylistic instructions. However, it also shows how they can reproduce and amplify specific styles or stereotypes. While I did not test it specifically, the lab report I referenced noted that in its own tests, "Gemini and MS Copilot, utilizing DALL-E 3, displayed improved diversity

representation" compared to older models like DALL-E 2. This suggests that companies are aware of these issues and are working to improve them, but it is an ongoing challenge.

Experimental Results and Image Examples

(I used three prompts to test each of the platforms.)

Prompt 1: "Taiwan night market really pretty design of food and stalls and with people like anime styles."

ChatGPT: ChatGPT's image was a nice, warm-toned scene. It focused on a girl in the foreground eating noodles. The background showed food stalls and lanterns, and the art style was clearly an anime. It felt like a focused snapshot of one person's experience. **Caption:** ChatGPT's image, focusing on a girl eating noodles.

Gemini: Gemini's image was completely different. It gave me a wide, symmetrical view looking down a long street. The market was very crowded with people. It also added pink cherry blossom trees, which I did not ask for. However, I did ask for a "really pretty design," and the trees added to that. This showed a more creative interpretation of the prompt. **Caption:** Gemini's image, showing a wide, symmetrical street with cherry blossoms.

MS Copilot: Copilot's image was very interesting because it felt similar in style and composition to ChatGPT's. It was a side-view of the market with two main characters in the foreground. The lighting was warm, and the focus was on the stalls and food. This was my first sign that ChatGPT and Copilot might "think" alike. **Caption:** Copilot's image, showing a side-view of stalls and people, similar to ChatGPT's style.

Prompt 2: "cute dog Shiba Inu with pink hair and dancing with a cute blue dress outfit with an mic."

ChatGPT: This generated a very cute and simple cartoon Shiba Inu. The dog had pink hair, a basic blue dress, and was holding a microphone. The background was a plain, solid color. It matched the prompt perfectly but in a very simple way. **Caption:** ChatGPT's simple cartoon Shiba Inu.

Gemini: Gemini's version was much more detailed and stylized. The Shiba Inu looked like a real pop idol. It was on a stage with bright lights in the background. It was holding a vintage-style microphone on a stand. The dress was also much fancier, with ruffles, a collar, and bows. This was a much more dynamic and polished image. **Caption:** Gemini's "idol" Shiba Inu, showing a much more detailed and stylized scene.

MS Copilot: This was the most significant finding of my experiment. The image from MS Copilot was almost identical to the one from ChatGPT. The pose, the simple blue dress, the pink hair, and the plain background were all the same. This strongly suggests that both models, being powered by DALL-E 3, interpreted my simple prompt in the exact same way. **Caption:** Copilot's Shiba Inu, which is almost identical to the ChatGPT output.

Prompt 3: "An image of an anime girl gaming girl with purple hair and a gaming setup and having a fun time."

ChatGPT: ChatGPT produced an image of a purple-haired girl sitting at her computer and holding a white cat. She is smiling and wearing headphones. The room behind her has several anime posters on the wall. It’s a cozy scene. **Caption:** ChatGPT's gamer girl, holding a cat with posters in the background.

Gemini: Gemini's image was also of a purple-haired girl with a white cat on her lap. However, the scene was much more detailed and high-tech. She has cat-ear headphones and is sitting in a large gaming chair. Her setup includes dual curved monitors, figurines on the desk, and branded soda cans. The whole room is lit with a purple neon. This image felt more complex. **Caption:** Gemini's gamer girl, featuring a detailed dual-monitor setup and a cat.

MS Copilot: Copilot's image was different from the other two. It showed a girl in a purple hoodie actively playing with a controller, with a very happy, energetic expression. In the background, there is a detailed PC computer case with RGB lighting. This image focused more on the "action" of gaming rather than the "at-desk" setup with a cat. **Caption:** Copilot's gamer girl, focusing on the action of playing with a controller.

Personal Insights

My exploration led to some clear insights.

- My most important insight was that **MS Copilot and ChatGPT produce very similar outputs, even when they're all given the same prompt**. This was most obvious with the Shiba Inu prompt. I gave all three platforms the exact same text, and ChatGPT and Copilot produced images that were almost identical. This makes sense because my research showed they both use the DALL-E 3 model. It shows that the DALL-E 3 model has a very strong and consistent "default" artistic style when given a simple prompt.
- **Gemini is totally differen**t. In every single prompt, Gemini's image had a unique style and composition that was very different from the DALL-E 3 models. It seemed to interpret my prompts more creatively or contextually. For example, it added cherry blossoms to the "pretty" night market and created a much more complex, high-tech scene for the "gamer girl." This shows its strength in delivering high-quality, contextually relevant outcomes.
- **DALL-E 3's prompt-following is excellent.** While Gemini was more creative, I noticed that the DALL-E 3 images in both Copilot and ChatGPT followed my exact instructions very closely. The enhancements in DALL-E 3 are very clear, especially in its prompt adherence. It delivered exactly what I asked for, which is a major strength.
- **Prompting is an art**. I learned that the quality of the image depends heavily on the quality of my prompt. A simple prompt like my Shiba Inu one resulted in a simple image from DALL-E 3. A more complex prompt would be needed to get a more complex scene.
- **The chat interface is a huge advantage.** The chat-like features of all three platforms really enhance the user experience. Being able to ask for an image and then immediately

ask for a change in the same conversation is very effective. This is especially true for Gemini, where its incorporation into the multimodal model allows for seamless conversational enhancement of the images.

Conclusion

This experiment showed that DALL-E, Gemini, and MS Copilot Image Creator are all powerful and impressive tools that are key developments in this field. Each platform has distinct advantages and disadvantages. DALL-E 3, used by both ChatGPT and Copilot, excels in creative tasks and high-fidelity results. Google's Gemini truly shines in understanding subtle context and producing high-quality, unique results, leveraging its powerful multimodal capabilities. I confirmed my theory that ChatGPT and MS Copilot excel in creativity and high-fidelity results, but they often "think" alike, producing similar results. Gemini stands out by providing high-quality images with a completely different artistic direction, and its multimodal capabilities are remarkable. The rapid progress in this field is clear. These platforms have huge potential for computer vision and content creation. These technologies are clearly set to influence the future of computer vision and all forms of digital content development. However, they also highlight the important ethical issues of bias and misuse that we must continue to think about. My tests showed that models can amplify societal biases. This highlights the need for responsible development and future research to focus on reducing these biases and ensuring these powerful technologies are used ethically. This assignment showed me just how advanced these tools are and provided a clear look at their different "personalities."

References

- Google Gemini: https://gemini.google.com/
- MS Copilot Image Creator: https://www.bing.com/images/create
- DALL-E 2 (OpenAI): https://chatgpt.com/