Survey of Prominent AI Models and Their Ethical Implications

Kyle Sanchez
kylesanc@pdx.edu

Ian Wasson
iwasson@pdx.edu

Will McIntosh
wmm2@pdx.edu

December 2023

Abstract

AI has become increasingly more capable and accessible to the general population. Models are now reaching a point in quality where they are being released for everyone to use, and as consumer grade hardware increases in power, these models are becoming easier to use without extra subscription costs. The question is how good are these models, and what are their capabilities.

By surveying popular models, this experiment sought to examine the outputs of these free to use models. Everything from AI artwork, to visual narrations, these tools were easy to find, and easy to use. Dall-E3 is capable of assisting an artist with little drawing experience in constructing comic strips, MidJourney is capable of producing near authentic forgeries of famous artists. HeyGen can produce intelligible translations and ElevenLabs can fake any voice actor with minimal amounts of training data. The capacity for creativity is enormous, but so is the capacity for doing harm. Forging artwork, plagiarizing music, and deepfaking voices is easier than ever, free, and requires minimal effort.

1 Introduction

The rise of large language models in recent years has led to several advances in the generative AI space. New modes of generative AI have taken shape, from art generation to 3d modeling. The creator space is flourishing with new tools that aim to close the gap between creativity, and skill. With only a few sentences, users are able to generate just about any kind of artwork that they desire. Music is being created using these models in the style of classical musicians. Audio description of images can be narrated to the visually impaired, and human voices can be immortalized in virtual libraries.

With these new tools, comes a lot of new possibilities. Not all of them are benign. A bad actor may use these tools to fake a public figure's voice and have them say things that are incriminating. They may generate artwork that depicts someone in a bad light, or even cause legal trouble by passing off illegal copies of music as their own.

2 Motivation

As these new models have grown in popularity, it is important to question whether or not these tools are being used appropriately. It is the aim of this survey to explore the accessibility, ease of use, and cost of these new models. The more readily available a model is, the more likely it is to be abused. However, it may be the case that the free models may not produce results of a great enough quality to pass off as human made. Creators in particular, are seeing the most development of AI tools engineered for their use, as such, this survey will focus primarily on these tools that seek to assist an artist or writer in creating new works.

3 Methodology

For this experiment, the following models were analyzed: Dall-E3 [Bet+], Stable Diffusion [23a], ElevenLabs [23b], GPT-4V [Ope23], HeyGen [23c], and Musicfy [23d]. Dall-E3 and Stable Diffusion are models which take prompts provided by a user and generate artwork. GPT-4V uses images and text as a prompt, providing the context of a picture or video as a basis for the text inputs. ElevenLabs is a company concerned with building a library of AI voices, constructed from audio samples of real people. HeyGen allows a user to submit a video of them talking, and perform translation into various languages, as well as modify the original video to more accurately depict the mouth movements for the translated audio. Finally, Musicfy allows a user the ability to create covers of existing music in the style of another celebrity.

4 Results

4.1 Dall-E3

As a test of Dall-E3, a hand drawn image was used as a basis for a prompt. The idea was to turn a simple sketch into a fully formed comic strip. The prompt consisted of this sketch as well as the instruction to make a comic of the character going to the zoo and buying a donut.

Figure 1: Left: Input character sketch, Right: First generated comic

The initial output from the model showed that it understood the concept of a comic strip, as well as buying a donut. In the first example, it even included the setting of the zoo. The initial character sketch was not as prominent in this comic however, so the prompt was refined.

Figure 2: Further refined comics

After refining the prompt to emphasize the use of the provided character sketch, the model still struggled with reconstructing it. The same general character from the previous comic was still present.

As an attempt to increase the likelihood of the model using the provided character, a new sketch was created with more distinct features. The idea being that the model would pick up on these exaggerated features and include them in the final product. The final comic produced was the closest to the original idea. Dall-E3 works well as a tool, but cannot completely replace the work of an artist. The constant back-and-forth between the artist and Dall-E3 shows that it is a good companion but not yet capable of completely replacing artists.

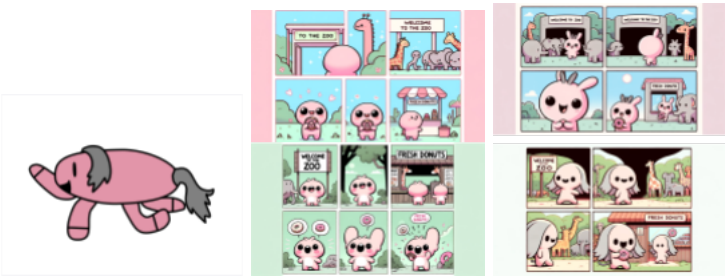

Figure 3: Left: New input character, Right: Comics using new character

4.2 Stable Diffusion

The most easily accessible version of Stable Diffusion is through an application called MidJourney. While the experiment with Dall-E3 was used to demonstrate the collaborative process, MidJourney will be tested for forgery.

Figure 4: Left: Real Kinkade, Right: Fake Kinkade

The first test was to ask for a rendition of Thomas Kinkade. While the colors are more vibrant than typical Kinkade, the general theme and layout was present. The next test was Pierre-Auguste Renoir.

Figure 5: Left: Real Renoir, Right: Fake Renoir

This further reinforces the fact that the models have clear knowledge of who these artists are, as well as an idea of their previous works. The content of

this result is exactly the same as the original. Finally, the model was asked to generate a Pablo Picaso. This was the most likely to be imitated due to his easily recognizable cubist style.

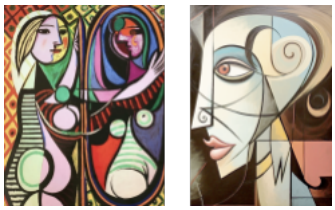

Figure 6: Left: Real Picaso, Right: Fake Picaso

These demonstrations prove that fabrication is incredibly easy with these models. The more famous and prominent the artist, the more likely the model will imitate them. All of this was done without any concern for copyright infringement or intellectual property on the part of the model itself. No warnings were given, no concern for the prompts given.

4.3 HeyGen

The HeyGen-Labs Video Translate tool works by taking a given video and cloning your voice to speak in a variety of different languages. This tool also modifies how your lips move in the video to make it seem like you are speaking in that translated language. This tool allows for translation into various languages including Dutch, Filipino (Tagalog), Italian, Korean, Spanish, and many more. This tool also allows users one free translation a day otherwise it prompts a user with various payment plans.

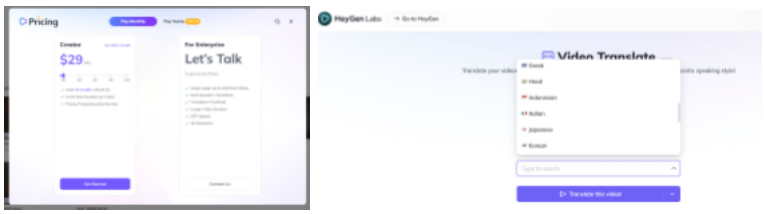

Figure 7: Left: HeyGen pricing model, Right: Language options in HeyGen

The tool itself in its free version and paid version differ only slightly. Going with a payment option only allows users multiple translations a day along with faster processing times compared to the free version. Functionally both versions are identical however for a creative professional that needs to create multiple translations a day, the paid version would benefit them more.

The translations themselves took anywhere from an hour, to more than twelve hours to process. Additionally, the translation's accuracy was surprisingly good. Several people who spoke the languages translated in this sample

were impressed with the accuracy, pronunciation, and intonation. However, there were some grammatical errors and translations that were not accurate.

4.4 Musicfy

Another kind of tool that has gained a lot of popularity within the past few months is AI music tools. The premise of these tools is that they create covers of any song using different artists, characters from different TV shows, celebrities, and public figures such as politicians. The tool Musicfy is one of these AI music tools that provides the previously mentioned capability.

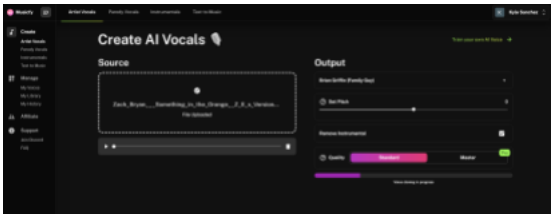

Figure 8: Musicfy user interface

Musicfy allows you to upload a music file or provide a YouTube video link for the song you would like it to make a cover for. All the user has to do is select the artist or person they want the cover to be sung by. The Parody Vocals section allows you to select from a wide variety of characters and people including Spongebob, Peter Griffin, Joe Biden, Barack Obama, and Donald Trump.

Through testing and using this tool, it processes these AI song covers, instrumentals, and text-to-speech songs within a few minutes. The main difference between the free and paid versions is just unlimited translations for the paid version. However, with the free version five song creations every hour is still plenty for any content creator using this tool. The Artist's Vocals, Parody voices, and Instrumentals converted the song to provide a user with accurate for their respective artist, celebrity./character for the parody or instrument the instrumentals.

The Text to Music feature did have a concerning issue. This part of the tool prompts a user to "Describe the music you want to generate". When prompted with "Drake Style Rap" it provided a sound clip only fifteen seconds long. Although this clip was short, there was a clear use of an actual song by Drake around the eight second mark from the song "Jungle". Although these clips are short it was nearly identical to the original.

4.5 GPT-4V and ElevenLabs

The final tool sampled in this survey was a combination of GPT-4V and ElevenLabs. This was first done by Charlie Holtz [Hol23], a screenshot of a user's

webcam is taken every couple of seconds and sent to GPT-4V to be asked to describe the contents of the image as if it were a David Attenborough documentary. This descriptive text is then sent to ElevenLabs to have a virtual recreation of Attenborough narrate it. All of this takes about 10 seconds to happen.

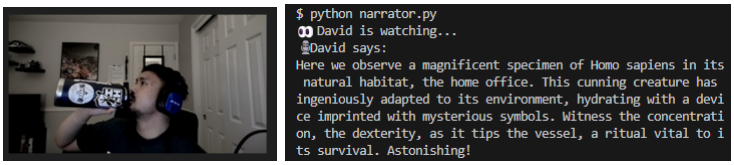

Figure 9: Left: Subject drinking water, Right: GPT-4V Description of subject

In an attempt to see how easy this is to recreate, the code that was provided by Holtz was duplicated, modified, and then ran. Given an image of a person drinking from a water bottle, it gave an incredibly descriptive analysis in Attenborough's iconic style. This tool has potential to help those with visual impairments, creating descriptive audio for visuals. It can however, be used to easily generate voice lines that the likeness of the voice may not approve of. In the original demonstration by Holtz, the AI Attenborough was recreated without his approval or authorization. While it may seem humorous at the moment, the legal implications of this may prove to be more challenging to reconcile.

5 Conclusion

All of the tools surveyed in this experiment were incredibly easy to find, use, and modify as desired. None of the models sampled required payment, and even the payment options provided no discernible difference in quality. Artwork is a highly subjective field, what makes something good or bad is ultimately up to the viewer. It is however, also incredibly difficult to create and uphold laws surrounding the use of AI tools in creating artwork. In the case of Dall-E3, is it sufficient to provide concept art? Can one copyright and distribute the output from these models? How transformative is the work, and if one were to use the work as a basis for their own artwork, what does that process look like? There are many questions about the legality of these models, but at the moment, without those laws people can use them to create whatever their imagination desires, with reasonable levels of quality, and no impact to their wallets.

References

[23a] 2023. URL: https://stablediffusionweb.com/#ai-image-generator.

[23b] 2023. URL: https://elevenlabs.io/.

[23c] 2023. URL: https://www.heygen.com/.

[23d] 2023. URL: https://musicfy.lol/.

[Hol23] Charlie Holtz. *CBH123/narrator: David Attenborough narrates your life*. Nov. 2023. URL: https://github.com/cbh123/narrator.

[Ope23] OpenAI. *GPT-4V(ISION) System Card*. Sept. 2023. URL: https://openai.com/research/gpt-4v-system-card.

[Bet+] James Betker et al. *Dall·E 3*. URL: https://openai.com/dall-e-3.

6 Contribution

We discussed and decided to go with option 1 as individual contributions.

Option 1: We agree that all group members made a valuable contribution and, therefore, it is fair that each member receive the same grade for the discussion.