Literature Review: What is the Impact of AI Art?

Dalong Hu

Department of Computer Graphics Technology, Purdue University

CGT 540: Current Topics in 3D Animation

Nicoletta Adamo

April 3 2023

Literature Review: What is the Impact of AI Art?

Generative AI outputs images based on text prompts. Ghosh & Fossas (2022) bring attention to a case that illustrates the quality it can achieve, one where Jason Allen used Midjourney to win the digital art category at the Colorado State Fair. The recent abundance of AI art raises many issues. For instance, training data frequently lacks attribution, without heed for copyrights. Only 5% of the images in the widely used COCO dataset can be used unrestricted. Models can be trained based on a specific artist's style in a mimic attack for unlicensed reproduction, and models can sometimes reproduce unmodified data. Unfortunately, there is a lack of legal regulation in part since the internet spans jurisdictions, and profits shift away from individual creators. Some companies like Alibaba already use generative AI models in their ad campaigns.

After explaining these negatives, Ghosh & Fossas (2022) go on to note the precedents of how when photography and digital art were introduced, there were also fears they would cheapen and displace the existing fine arts market. They believe that AI art can become its own medium and style if used ethically. I believe that both the harms and benefits of AI generated art will continue to persist until legal regulation takes hold, and eventually AI art will settle as its own distinct medium.

Examination of Painting Synthesis

Yi et al. (2021) note the difficulties of AI image synthesis of paintings due to a lack of painting training data and how paintings portray a different reality due to the intent and technique of the artist. GAN-based models can produce high quality samples but may sacrifice diversity and stability. Diffusion models are a more recent development and are more easily trained to cover a wider distribution than GANs. However, they note that existing studies still show a certain a level of difference compared too human created created works.

They used resized 64x64 training images from the dataset of wikiart.org, and compared a diffusion model to VQGAN in generating the specific styles of Van Goph and Claude Monet.

Yi et al. (2021) found that the diffusion model had a similar granular level compared to, able to mimic unique characteristics such as brushwork, but recalled some of the existing images from the dataset. In comparison the VQGAN was poor at the granular level, and almost all samples were landscape paintings while other classes were ignored. In contrast, the diffusion model replicated a wider variety of the target distribution.

Their findings add depth to the understanding of what different types of AI models are able to generate, and especially shed light on their weaknesses.

However, Sun et al. (2022) note that AI art is on a different order from the precedent of photography since it could entirely replace the artist, and so question whether AI can completely replicate the human element, the artist's own intentions, perceptions, and emotions. They also chose painting as the venue for investigation since it involves an artist transforming abstract concepts and connotations into concrete details that evoke emotion. Just through the brushstrokes on a canvas, a viewer can glimpse the artist's mind.

Sun et al. (2022) had 380 participants, 70% of which had painting experience, compare two sets of 6 paintings. They hoped to find whether the subjects could differentiate human and AI generated art, what parts of AI art have emotional impact, and what AI art's current limitations are. They found that while most participants couldn't differentiate which was human made, the percentage of correct identifications gradually increased as the study progressed.

Additionally, the participants were more likely to judge a painting to be human made if it better evoked emotion, rather than if it was stronger technically. Out of the six pairs, in five of them the AI generated painting scored higher in demonstrating theme. The only exception was a unique case where the AI could not understand a certain level of color symbolism, and didn't give the same significance to it as a human painter would have.

Over all, Sun et al. (2022) found that the computer can mimic a certain style but lacked the artists capacity to reflect on overall composition. Of course, the models are also not good at handling unusual cases. They note that "AI painting is limited to the level of technique, since a computer is unable to operate on the levels of semantics and effects" (Over all, Sun et al., 13). They add that lacking self-awareness and holistic perspective are AI's main weaknesses.

Going forwards, they believe that AI will overcome this weakness with improved semantic recognition and that how human artist could maintain an advantage would need to be investigated. These studies by Yi et al. (2021) and Sun et al. (2022) demonstrate the strengths and limitations of AI art.

Skill and Mimicry

In response to AI's capabilities in terms of replicating technique, Shan et al. (2023) have built a tool that "cloaks" an image by applying small style transfer perturbations to disrupt the mapping of a model's latent space. They note the power of AI art to easily generate complex artwork that would require professional artists to invest hours of hard work. The problem they focus on is how models like the open source StableDiffusion can be easily "fine-tuned" to mimic specific art styles. Platforms such as Civitai are built around sharing customized models for specific styles, targeted artists, or NSFW content. Shan et al. (2023) note that models profit off of an artist's years of training without compensating them, displace them in search results, and discourage art students who decide to quit. For more popular artists, simply writing their name in the text prompt can get the model to mimic their style. However, users can use as little as 20 pieces of artwork to train a model to mimic a new artist's style. In response to violations to intellectual property rights and dangers of livelihood, artists have spoken out and The Concept Art Association has filed a class action lawsuit and raised over $200K to combat AI art. Shan et al. (2023) disseminated a survey to artist communities which was filled out by 1,207 participants. They found that "97% artists state it will decrease some artists' job security; 88% artists state it will discourage new students from studying art; and 70% artists state it will diminish creativity" (Shan et al., 4) and "78% of artists anticipate AI mimicry would impact their job security, and this percentage increases to 94% for the job security of newer artists" (Shan et al., 4). It should be noted that 91% of their participants have read about AI art extensively, so this is a very passionate group of responders and there is probably some level of sample bias.

Shan et al. (2023) explain how existing cloaking methods like anti-facial recognition models don't prevent AI mimicry because their feature spaces represent identity-related details. They contain information such as color, objects, locations that is unnecessary to shifting features related solely to artistic style. As such, they designed Glaze to use style transfer to hone in on the style components of an image to guide perturbation computation. Shan et al. (2023) delineate that because generative models have continuous output spaces, even small shifts in feature representation can alter the resulting image generation. Glaze works by creating a style transfer version of the image and then applying it to the image with as little visual disruption as possible.

Shan et al. (2023) had their participants to rate Glaze's protection on images on a 5-level Likert scale, with their artist-rated PSR being the percent of users who rated the protection as "successful" or "very successful." They also used the CLIP model's capability of classifying genre to define a CLIP-based genre shift rate metric to be the percentage of mimicked art that don't have the original genre in the top three predicted genres. Shan et al. (2023) found that Stable Diffusion models are stronger at mimicry than DALLE-m models, and that Glaze achieved a "> 93.3% artist-rated PSR and > 96.0% CLIP-based genre shift" (10).

Artists surveyed (1207 participants from artist communities) and CLIP based scores show 92% success rate under normal conditions and 85% success rate against countermeasures like increasing gaussian noise level, JPEG compression, and even robust training. Even at low perturbation levels such as p=0.05, 92% of artists were either "willing" or "very willing" to post cloaked artwork. They attribute this to there being existing techniques artists already use to protect artwork online, such as watermarks or reduced resolution. They also found that even when artists can cloak only 25% of their work, Glaze achieved a high 87.2% artist-rated PSR rate, which means Glaze can help artists who already have a certain body of work online.

In terms of limitations, Shan et al. (2023) note the workaround of mimicry attackers simply using older unprotected artwork, and the fact that Glaze is in no way future-proof. They state that, "we are under no illusion that *Glaze* will remain future-proof in the long run" (Shan et al., 13) but believe that it is a necessary step before legal and regulatory protection comes into place.

Although mimicry attacks are a clear negative that arises from AI art and might rob from artists as a whole a certain value of creativity, some AI artists point to prompt engineering as its own form of creativity.

Prompt engineering is the skill of writing "prompts" and is an iterative and experimental process to probe into the latent space of generative models. Oppenlaender (2022) notes that a specific format of text input is key to generating the desired images. For instance, 49 recipes are listed in OpenAI's documentation on how to input prompts.

Prompt engineering can be improved with experimentation, research, and use of community resources. Not every practitioner shares their prompts especially if they have commercial interests like selling NFTs.

Oppenlaender (2022) did a 3 month autoethnographic study and used VQGAN-CLIP to create a taxonomy of modifiers so as to enhance theoretical understanding of how humans write and use prompts for the HCI research community. Currently the HCI literature is missing a unified theory for prompt engineering. They note that although prompt engineering can be learned from community guides, reports, and social media, personal application and experimentation is necessary to understand the craft. Oppenlaender (2022) chose VQGAN-CLIP because it was one of the first generative models that gained massive popularity and was thus key to the growth of the text-to-image community. VQGAN-CLIP is deterministic, which means the same text input with the same configurations will produce the exact same images. Oppenlaender (2022) paid particular attention to the Twitter community and other community resources, continually reinterpreting and expanding their list of potential candidates for prompt modifiers until it no longer grew.

Oppenlaender (2022) distinguishes 6 types of prompt modifiers, but notes that every generative model will have a certain quality and style distinct to them due to the latent space of their neural network.

Subject terms indicate subject and are crucial.

Style modifiers reproduce a style. Some examples are “#pixelart”, “Cubism”, “by Greg Rutkowski.”

Image prompts are like subject terms. There can be many image prompts but only one initial image.

Quality boosters are modifiers like “trending on artstation”, “#wow”, “rendered in UnrealEngine”, “beautiful.”

Repetition strengthens associations, so “a very very very very very beautiful landscape” creates different results.

Magic terms add variation and randomness by using distantly related words or non-visual intangible qualities such as “feed the soul.”

Weights can be assigned to each of the modifiers. To exclude certain subjects, negative weights can be assigned, such as “heart: -1” to remove the tendency of generating heart-shaped objects when “love” is in a prompt. Generally subject modifiers are added first, then style, quality, and repetition before modifiers that introduce variation are appended, with weights being the last step.

Oppenlaender (2022) says that prompt engineering can be the starting point of creative workflows that use initial images as inspiration before moving on to other systems or photo editors, and that there is opportunity for the creation of tools and interfaces that support such creativity. Prompt engineering relates to human-AI alignment because prompts are meant to encapsulate stated intent, and further research and development will be about teaching artificial intelligence to comprehend human values. Something to be aware of the inherent bias in text-to-image systems. For instance, prompts containing “princess” will often produce women of light skin color, because there is “bias in the training data toward Western, educated, industrialized, rich and democratic (WEIRD) subjects” (Oppenlaender, 11).

Oppenlaender (2022) makes the prediction that humans will further use natural language to interact with opaque models in the future. Many creative sectors will be transformed and disrupted. From the examples of CogVideo for text-to-video generation or OpenAI's Codex for natural language commands to programming code, Oppenlaender (2022) suggests that generative models could even formulate full story-driven worlds and games from prompts.

Here we see a dearth of creativity in the form of mimicry attacks, but also a new creative practice in the form of prompt engineering arising from AI generated art.

Example Applications of AI art

Ploennigs & Berger (2022) state that diffusion-based AI art platforms are powerful tools for concept design, and focus on how AI art can help with the ideation, sketching, and modeling of architectural design.

With NLP methods they analyze how these platforms are already being used.

Ploennigs & Berger (2022) compare the technology of the three leading platforms and analyze their different use cases to derive a workflow that merges the strengths of each individual platform. The three major platforms Ploennigs & Berger (2022) define are DALL-E, Midjourney, and StableDiffusion, all of which compete on the technical, quality, and user experience level. Midjourney has an extensive Discord community, DALL-E 2 has a dedicated web interface, and StableDiffusion is open-source with many community-created tools. Midjourney allows successive upscaling, and its workflow is centered on generating and comparing image variants to upsample the best one. DALL-E and StableDiffusion allow for direct editing through inpainting and outpainting. Ploennigs & Berger (2022) note that since all three models support importing external images, they can be combined into a single workflow. One thing to take into account is each model having its own image style due to training data and process, such as Midjourney being more impressionistic.

Since Midjourney user queries are publicly available, Ploennigs & Berger (2022) did an analysis on 40 million queries to see how people used AI art in practice for architecture. One pattern they found was that Midjourney users do not use full sentences but rather a collection of modifier terms. One of their findings about the mean number of iterations a user goes through shows that queries are developed over multiple iterations with additional terms appended on each time.

Ploennigs & Berger (2022) did a case study to look at how each model performed on its own for the prompt "cozy living room, wood paneling, television, large sofa, natural light, lived in, realistic, full view" (13). The initial generation of Midjourney had some results that were stylized or had strange perspective. Commands such as "remaster" can enhance quality, but even then small perspective errors need manual fixing. DALL-E 2 had some stylized results that didn't match the "realistic" prompt. After inpainting the results have good quality. Stable Diffusion had the strongest initial results, but all the images were close-ups so out-painting had to be used, with careful attention to the outpainting transition areas.

Adaptive workflows move across tools for the necessary interaction paradigms. For a unified workflow, Ploennigs & Berger (2022) suggest starting with Midjourney to leverage its advantage in free-form ideation, refining it with in/outpainting in DALL-E, and finally using Stable Diffusion to attain a high-quality final result.

A limitation of AI models is the necessity of a large amount of trial and error to get a desired result, and there are cases where a variant tree will fail entirely. Because AI has no semantic understanding and can only mimic style, a generated floor plan query may look fine but be completely nonsensical. Using technical terms in prompts also produce ambiguous results, since the models don't actually understand what they mean. Additionally, AI models have difficult with counting and spatial arrangements beyond simple cases of foreground and background.

Ploennigs & Berger (2022) conclude that the workflows for all tools are gradually developing as the necessary interaction paradigms get established, and believe that eventually platforms will provide any style and many interaction that users desire. What is more difficult to achieve are use cases that require semantic understanding, but specific training such as Building Information Models (BIM) can be used to try and overcome that.

Architectural design is just one field where AI art can see great use in aiding practitioners. As people work out how to incorporate AI models safely and ethically in their workflows, it can be a great benefit.

Conclusion

AI art is powerful and can be used for both great harm such as mimicry attacks, profit shifting, and copyright violation but also great creative benefit. Many creatives can use it for quick ideation, and with its incorporation into new workflows a unique creativity can spring from the combination of human and AI. Until further regulation is in place, both the positive and the negative will continue to occur. Eventually, I believe that AI art will find its own place as a unique separate medium.

References

Chang, R., & Huang, Y. (2021). Towards AI Aesthetics: Human-AI Collaboration in Creating Chinese Landscape Painting. In M. Rauterberg (Ed.), Culture and Computing. Interactive Cultural Heritage and Arts (pp. 213–224). Springer International Publishing. https://doi.org/10.1007/978-3-030-77411-0_15

Ghosh, A., & Fossas, G. (2022). Can There be Art Without an Artist? (arXiv:2209.07667). arXiv. https://doi.org/10.48550/arXiv.2209.07667

Oppenlaender, J. (2022). A Taxonomy of Prompt Modifiers for Text-To-Image Generation (arXiv:2204.13988). arXiv. https://doi.org/10.48550/arXiv.2204.13988

Ploennigs, J., & Berger, M. (2022, December 19). AI Art in Architecture. ArXiv.Org. https://arxiv.org/abs/2212.09399v1

Shan, S., Cryan, J., Wenger, E., Zheng, H., Hanocka, R., & Zhao, B. Y. (2023). GLAZE: Protecting Artists from Style Mimicry by Text-to-Image Models (arXiv:2302.04222). arXiv. https://doi.org/10.48550/arXiv.2302.04222

Sun, Y., Yang, C.-H., Lyu, Y., & Lin, R. (2022). From Pigments to Pixels: A Comparison of Human and AI Painting. Applied Sciences, 12(8), Article 8. https://doi.org/10.3390/app12083724

Yi, D., Guo, C., & Bai, T. (2021). Exploring Painting Synthesis with Diffusion Models. 2021 IEEE 1st International Conference on Digital Twins and Parallel Intelligence (DTPI), 332–335. https://doi.org/10.1109/DTPI52967.2021.9540115