Visual Art in the Age of AI

Roshnica Gurung
William & Mary

Visual Art in the Age of AI

Roshnica Gurung

College of William and Mary

Research Mentor: Christopher Boyle

Business, Computer Science & IT Pathway

Tidewater Community College

Table of Contents:

ABSTRACT

Artists and researchers have been deeply interested in using AI programs that generate art for quite some time now. As a result, there have been many advancements in making AI more accessible and easier to use for the public. This is because AI is not just for business anymore. Nowadays an individual without a college degree with even the slightest interest in artcan go on a website like Stable Diffusion and create an artistic image using a text prompt in a quick couple minutes. The only limit is your imagination- and your internet's stability. This accessibility was a huge reason behind the popularity of generative AI programs.

In this paper, I look into how AI is currently being used in art generation, some of its shortcomings, along with some risk mitigation strategies that could help lessen the impact of possible privacy attacks. Later on I go into dataset models and how the programs that use them have failed to protect individuals' privacy. On top of that, I also investigate the mistreatment of artists in the hands of large AI companies, providing real life examples. In summary, this work investigates the problems plaguing the generative AI industry and the individuals involved in it, and attempts to find solutions that will minimize the negative impact of it as much as possible.

KEYWORDS:

artificial intelligence, generative AI, visual art generation, privacy risks

I. Background

A. What is Generative AI?

Generative AI is a type of artificial intelligence that is designed to generate content based on existing data. They can consume copious amounts of data, including text, imagery, and synthetic data, which is artificially produced. Generative Artificial Intelligence uses machine-learning algorithms to create new content from the data it receives on the user's end. This AI has been growing in popularity, especially with organizations making their versions of it openly sourceable. When a program becomes "open-source" that means it is available for use to the public and can be used by developers to create their own versions of the program. Many generative AI programs use large language models as data to train on.

B. What is a large language model?

An LLM (Large Language Model) is an artificial intelligence system that uses deep-learning algorithms to process natural language. They are trained using the DL (Deep Learning) technique which uses artificial neural networks to make predictions on a dataset. These neural networks imitate the way a human brain would function in the way it recognizes complex patterns in different datasets and gains new information from it, eventually growing big enough to make predictions on its own about topics pertaining to that data. LLMs can be used for text summarization, translation to other languages, and even answering questions a user may

have. It studies the relationship between different words and groups them in pairs depending on their relationships.

The combination of deep learning with the vast amount of data in an LLM holds both positive and malicious possibilities.

II. Current Uses of AI in Visual Art

Recently there has been a surge in tools such as text-to-image generators thanks to the regular releases of new open-source programs. These programs have skyrocketed to fame, especially ones like LensaAI, due to them going viral on social media. If you have been active on apps like TikTok, Instagram, and Twitter since December 2022, then you have probably come across multiple posts of cartoon images of the same person in varying art styles. Those posts were made using a paid subscription AI app called Lensa. Its “Magic Avatars” feature takes in a user’s selfie and creates a collection of artistic images with their face based on the prompt the user picks. Some examples of it are fantasy, anime, pop, and even fairy princesses. It quickly climbed the App Store charts with people turning it into a trend. People were sharing their reworked selfies with friends and strangers online, commenting on whether they thought the AI had done a good job capturing their faces in the images.

Although Lensa, initially released in October 2019, is a child of PrismaLabs, it uses Stable Diffusion, a latent text-to-image converter released on August 22, 2022, by Stability AI Ltd. Another popular text-to-image generator is Midjourney which was released by Midjourney Inc. on March 14, 2022, a couple of months before Stable Diffusion.

Both Stable Diffusion and Midjourney use LAION 5B, which is a large-scale dataset consisting of images and captions that were taken from the internet. LAION 5B contains over 5.85 billion image-text pairs that were filtered through the CLIP neural network, which connects texts and images based on their relationship. For example, an image of a girl studying in a classroom could be connected to the word “education”. LAION 5B is used by many artificial

intelligence programs because of the large amount of data it covers. 2.3 billion of the image-text pairs contain the English language with 2.2 billion from 100+ other languages, and the final 1 billion not falling under a language assignment, i.e. names.

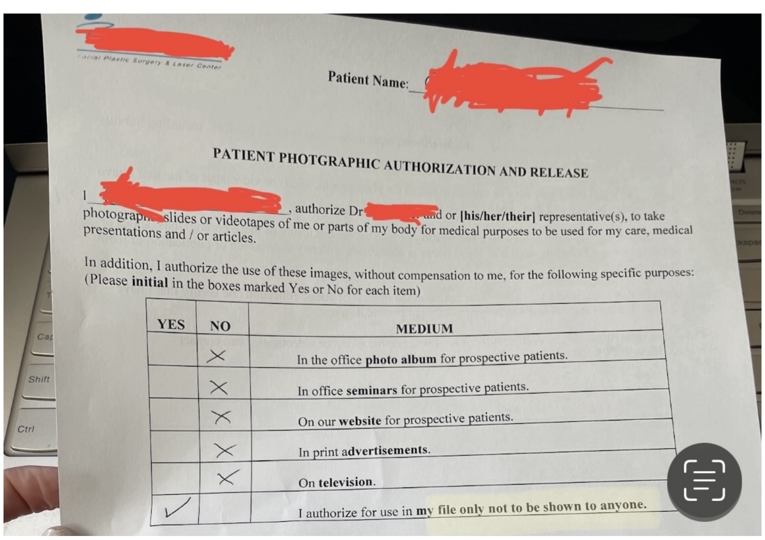
With the large size of the LAION 5B dataset, it comes with a myriad of privacy and security issues which range anywhere from issues with the dataset to stealing work from artists.

III. Current Problems Caused by AI in Visual Art

A. LAION 5B Issues

In September 2022, an artist named Lapine found out that her medical photos from 2013 were referenced in the LAION 5B dataset. Lapine used the website “Have I Been Trained” to see if LAION 5B had used any data with her face on it. “Have I Been Trained” is a website that searches the LAION 5B and LAION 400M image datasets to look for an image or text given by the user. Artists have been using this website to see if their work has been used as training data without their consent. “Have I Been Trained” works in collaboration with LAION, so if people do find their work there, they are able to flag it for removal.

Lapine uploaded a recent selfie of hers and reverse image searched it to find a picture of her that was taken by her doctor back in 2013. Her tweet stated that she had signed a consent form for her doctor, choosing the “YES” option for “I authorize for use in **my file only not to be shown to anyone.**”. In another tweet, Lapine says, “The other part is- this probably has happened to other patients. What would be a good protocol for a patient who finds part of their private medical record in a public dataset.” (Lapine, 2022)



[redacted]
Facial Plastic Surgery & Laser Center

Patient Name: [redacted]

PATIENT PHOTGRAPHIC AUTHORIZATION AND RELEASE

I [redacted], authorize Dr [redacted] and or [his/her/their] representative(s), to take photograph[redacted] slides or videotapes of me or parts of my body for medical purposes to be used for my care, medical presentations and / or articles.

In addition, I authorize the use of these images, without compensation to me, for the following specific purposes: (Please **initial** in the boxes marked Yes or No for each item)

YES	NO	MEDIUM
	X	In the office **photo album** for prospective patients.
	X	In office **seminars** for prospective patients.
	X	On our **website** for prospective patients.
	X	In print **advertisements**.
	X	On **television**.
✓		I authorize for use in **my file only not to be shown to anyone.**

Picture of Lapine’s “Patient Photographic Authorization and Release” Form
Credit: @LapineDeLaTerre on Twitter

B. Collection and deletion of data

1. Biased data

Systemic racism is ingrained in the core of American society and affects almost every part of our lives here. Since a lot of the data used to train AI reflects current societal biases towards particular groups of people, this can result in insensitive content being produced by AI. In this way, negative stereotypes and biases can be inadvertently perpetuated by AI.Besides being extremely inappropriate, it can also negatively impact marginalized communities if used in processes like hiring and grading. For example, if an AI-powered system was used to make decisions about hiring people, it may favor a person who comes from a wealthier background compared to somebody who comes from a poorer one. The AI-system fails to consider the social factors that connect socioeconomic status and literacy rates. This can cause financially disadvantaged individuals to be treated even more unfairly. This generalization puts groups of people such as queer individuals, people of color, disabled people, and first-generation students in major risk for discrimination.

2. Incorrect/sensitive data

Because of the vast size of the datasets, it is hard to verify each unit of data as safe and trustworthy. The bigger the dataset, the more privacy concerns arise. If the dataset is big, it's hard to personally verify each and every instance of data as true.

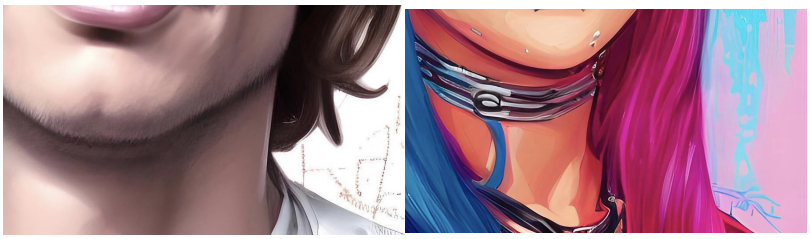
Another concern is LLMs getting sensitive data such as names, addresses, and other identifying information about a person and compromising their privacy. An example of that

could be a selfie you post on your Instagram account with your family ending up on a dataset without your or your family's consent. This could be added to the model's algorithm and your and your family's faces could be used to generate new content without you knowing it.

C. Art theft

Art theft is a serious legal and ethical issue that has been a growing concern for many artists with the rise of AI use in creating visual art. The thing about these concerns is that they're not rooted in hypotheticals- art theft is already happening. With companies like Stability AI and Midjourney, making a piece using a text prompt has never been easier, with many online trends skyrocketing their user growth.

There is currently a lawsuit against Stability AI, Midjourney, and Deviant Art, which has made its own AI art generator called DreamUp. The lawsuit was made by three artists - Sarah Andersen, Kelly McKernan, and Karla Ortiz. They claim that these organizations trained their AI on images collected from the internet, also called "training images", without the consent of their original artists. This is to support their other claim that "AI Image Generators are 21st-century collage tools that violate the rights of millions of artists". They say that Stability AI grabbed more than 5 billion images from websites such as Getty Images, Shopify, Tumblr, and Flickr to use as training data without any licenses. The lawsuit, called "Andersen v. Stability AI Ltd, U.S. District Court for the Northern District of California, No. 3:23-cv-00201." has garnered the attention of hundreds of thousands of people, including artists who are worried about having their art stolen by these organizations with no promise of compensation, and rightfully so.



Screenshots of blended signatures from images generated by Lensa

Credit: @LaurynIpsum on Twitter

On top of the lack of compensation and consent to use their artwork, artists have also reported not getting credited for their work. Lauryn Ipsum, an illustrator, went on Twitter to point out that “mangled remains of an artist’s signature is still visible” on the images generated by Lensa. With a quick look at Lensa artworks, one can find fragments of different artists’ signatures combined together, usually on the bottom right of the AI-generated image. This confirms that datasets like LAION 5B, used by Lensa, Stable Diffusion, and Midjourney, take art from artists without their knowledge with no recognition.

IV. Other Problems Caused by AI

A. New forms of Privacy Attacks

Before diving into Prompt Injection Attacks, one of the newest privacy attacks in the AI scene, we need to know what prompt-based learning is. Prompt-based learning, also known as prompt fine-tuning, is a strategy that allows programs to adjust themselves and their knowledge according to the prompt given to them by the user. This strategy relies heavily on the quality of the prompt provided in order to provide good output.

Prompt Injection Attacks are a fairly new addition to the world of privacy attacks. While injection attacks have existed for a while, prompt injection attacks were recently popularized thanks to programs like ChatGPT and Bing's AI chatbot. An injection attack happens when an attacker supplies untrustworthy input to a program which causes it to execute differently than it should. There are many different types of injection attacks like code and os command injection attacks, but we'll mostly be focusing on prompt-based injection attacks for now.

In a prompt injection attack, the attacker inserts untrusted text as part of the prompt. From there the attacker can change the output to virtually anything. This technique is often used to manipulate users into revealing sensitive information about themselves, including but not limited to full name, address, bank account number, etc. The reason why this specific attack is so dangerous is because it can be customized to the person interacting with the program. This equips attackers with a new tool that allows them to shapeshift into the type of person the user would trust.

V. Privacy Risk-Mitigation Strategies

While there isn't one end-all be-all solution to protecting user privacy, there are different risk-mitigation strategies that are available for use.

A. Diversity in data

The first strategy is to use smaller, more diverse datasets. This makes the process of verifying the status of the individuals involved much easier. It also helps confirm that consent is given by every individual to have their information in the dataset. Smaller datasets are also good for gathering data involving diverse factors in specific environments. This way we cover instances of problems that may not be as commonly thought of, and in turn, minimizes bias and negative stereotypes that may form if we only looked at data that is popular.

B. Differential Privacy

In addition to using smaller datasets, we can also utilize differential privacy in generative AI. Differential Privacy (DP) is a technique that studies and shares information about the individuals shared in the dataset. DP utilizes noise-adding algorithms that add randomness to data, which provides protection against privacy attacks by de-identifying data and making it untraceable. It plays a very important role in making AI safer for public use because it makes it possible for programs to use and share important information about individuals without disclosing sensitive information like names and family history. It does so by focusing on the different patterns in a dataset and censoring all the private information that doesn't contribute to it. While there are a lot of advantages to using differential privacy for privacy protection, it also

comes with some downsides that we need to consider before crowning it the best way to protect our information.

The noise that DP adds to data minimizes its accuracy, and hence, may provide us with imperfect data. While partially correct data can mess up the final results, it also provides another layer of protection between users and hackers by misleading them with data that may or may not be correct. In most cases, the difference between perfect and imperfect data is not overly stark and only really works as a preventative measure for privacy attacks.

VI. Conclusion

Now that we have examined some of the bigger problems plaguing the AI visual art industry, the million-dollar question awaits: where do we go from here? Should we even implement the changes we talked about in this paper? And if so, how do we do it?

There isn't a strategy that would end all privacy issues at once, but we can focus on the two we talked about for this paper. Through research, we can see that the best way to deal with these issues is to work with the core of the cause, which would be the data being used. The majority of the problems are caused by corrupted, verified, or uniform data, which could be alleviated by using smaller datasets that focus on niche and diverse variables. The secondary strategy that could also work would be implementing differential privacy to make data untraceable back to the individual. After comparing the advantages and drawbacks of each of the strategies, we can conclude that the former is the better choice.

While it is a lot more work than simply using a mega-large dataset such as LAION 5B, there is way less room for error than if we were to simply implement differential privacy strategies to our data models.

References:

Greshake, Abdelnabi, S., Mishra, S., Endres, C., Holz, T., & Fritz, M. (2023). More than you've asked for: A Comprehensive Analysis of Novel Prompt Injection Threats to Application-Integrated Large Language Models. https://doi.org/10.48550/arxiv.2302.12173

Lapine [@LapineDeLaTerre. (2022, September 16). *My face is in the #LAION dataset. In 2013 a doctor photographed my face as part of clinical documentation.* [Image attached]. [Tweet]. Twitter. https://twitter.com/LapineDeLaTerre/status/1570889343845404672

Bender, Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the dangers of stochastic parrots: Can language models be too big? FAccT 2021 - Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency, 610–623. https://doi.org/10.1145/3442188.3445922

Cormode, Jha, S., Kulkarni, T., Li, N., Srivastava, D., & Wang, T. (2018). Privacy at Scale: Local Differential Privacy in Practice. Proceedings of the 2018 International Conference on Management of Data, 1655–1658. https://doi.org/10.1145/3183713.3197390

Schuhmann, Beaumont, R., Vencu, R., Gordon, C., Wightman, R., Cherti, M., Coombes, T., Katta, A., Mullis, C., Wortsman, M., Schramowski, P., Kundurthy, S., Crowson, K., Schmidt, L., Kaczmarczyk, R., & Jitsev, J. (2022). LAION-5B: An open large-scale dataset for training next generation image-text models. https://doi.org/10.48550/arxiv.2210.08402

Philipp Hacker, Andreas Engel, & Theresa List. (2023). Understanding and Regulating ChatGPT, and Other Large Generative AI Models. Verfassungsblog, 2366-7044.

Zohny, McMillan, J., & King, M. (2023). Ethics of generative AI. Journal of Medical Ethics, 49(2), 79–80. https://doi.org/10.1136/jme-2023-108909

Lauryn Ipsum [@LaurynIpsum (2022, December 5). *I'm cropping these for privacy reasons/because I'm not trying to call out any one individual. These are all Lensa portraits* [Image attached]. [Tweet]. Twitter. https://twitter.com/LaurynIpsum/status/1599953586699767808?s=20