The Current State of Protection Against AI for Visual Artists

Audrey Xu *

Crystal Springs Uplands School, Hillsborough, CA, USA

*axu26@crystal.org

Abstract. Visual artists find themselves scattered as generative AI continues to make its way into business and mundane life. Though the technology behind has the capacity to aid artists, their current usage conflicts with artist careers. In addition, behind popular AI models used today are mountains of stolen work and data used to train those very models. As a result, there have been recent research efforts into designing new technologies that help artists defend their work from being used by AI without permission. Moreover, there are multiple ongoing legal cases where the outcome could lead to more regulation of this newfound technology. However, in spite of these recent efforts, artists still continue to struggle because available protection methods remain esoteric and complicated. In this paper, multiple forms of protection for artists are discussed, listing various advantages and why they ultimately still fail. A survey was then conducted to garner artist opinions on what is available now. Furthermore, they were asked what they wanted to be changed in the future in order for artists and AI to co-exist without opposing each other.

Keywords: Generative AI, artist, protection

1 Introduction

AI-generated art has swept the art world by storm after its release to the wider public. While the novelty of this technology is exciting, eventual consequences are sure to hit the artistic world [1]. Already, AI-generated art faces reasonable backlash from artists, and day by day, it only grows more advanced. Some artists have improved their productivity through the aid of generative AI, as the images it creates are technically more advanced and quickly made. Other people entirely remove human labor from the equation, relying solely on AI for the work. Because companies are focused on the output, more and more artists are being replaced in favor of their AI counterparts [2]. Hence, in today's system, the existence of AI creates competition between artists, as those who are able to output works faster are more valued while those who require more time are shunned.

Moreover, the very creation of generative AI harms artists, in the way it gathers data. Most, if not all, models scrape the internet for data and use artists' works without their permission [3][4]. This certainly raises ethical issues [5][6]. However, data scraping is

K. Chulsoo et al. (eds.), *Proceedings of the 2025 4th International Conference on Art Design and Digital Technology (ADDT 2025)*, Advances in Computer Science Research 123,
https://doi.org/10.2991/978-94-6463-815-8_16

not considered illegal, thus it is difficult to argue that this is copyright infringement. With artists losing lawsuits, larger AI companies grow stronger and increasingly untouchable by individuals [7]. It is becoming clear that with the current laws in place, while AI images are not protected by copyright, artists aren't either.

Without copyright protection, artists must turn to community resources in order to protect their work. Despite resources like Glaze [8] and Nightshade [9] available to the public, these tools are not streamlined, especially for artists who likely have less experience with technology in comparison to researchers. In essence, there is a major gap between artists' understanding of generative AI and perceived understanding by researchers. While the intentions behind anti-AI tools are to aid artists, many artists find the process needlessly complicated and are left without simpler ways to protect.

This paper will start from related works and a brief summary of the generative AI involved in the art world. Then a few artist protection methods will be discussed with advantages and weaknesses. To study how artists utilize those protection methods, we sent a survey to broad artist forums and online communities. The survey results and conclusion will be represented towards the end of this paper.

2 Related Works

A similar concept to study the dynamic between AI and artists has been executed, notably through the works "Understanding Generative AI in Art" [2] and "Foregrounding Artist Opinions" [10]. The first paper focuses on how artists interact with generative AI, as well as their views on its usage in a professional setting. Many artists surveyed were rather accepting of it, viewing it as a collaborator, though they brought up the ethical error of AI replacing artists' work, as well as the regurgitation of pre-existing content through AI. The survey also highlighted the gap in knowledge between the creators of AI models and its users, concluding that this technology needs to consider those unfamiliar with it.

The latter paper contains questions more focused on ownership and usage rights. The majority of artists-- though not overwhelmingly--believed that AI would harm artists and did not view it as a positive development. However, almost thirty percent of survey takers were neutral on whether or not AI is a positive development. To add on, few believed that AI work trained on a distinct recognizable style should be considered and owned by the artist whose style it is replicating. General consensus was that artists were concerned about who was profiting off of AI, though many were not concerned with receiving profit themselves.

Despite the similar concept, the survey outlined in this paper is geared towards methods for artist protection, not artists' beliefs on AI (though that is still relevant). Before delving into the survey, this paper will also go over various methods of protection that currently exist, along with their benefits and inconveniences for artists.

3 Generative AI and Models

When discussing AI today, most refer to generative AI, which uses models to create outputs, whether that be text or image. Most of these AIs create outputs based on text prompts. In order to function, AI models are trained on samples consisting of massive amounts of data. Within these models, there are many methods of development, notably diffusion models and generative adversarial networks (GANs) [11][12].

As shown in Figure 1, diffusion models use images which have noise continuously added to them. As noise is added, the model tracks each step until it is able to work its way backwards and create an image similar to the original. Following this, it is able to create images through prompting. Out of other models, it is most suited to work with text-to-image generation. Diffusion is one of the most popular models today, as the images it generates are clearer and have less error, though they also take longer to generate. Some well-known diffusion models include Stable Diffusion [13], Midjourney [14], and DALL-E [15].

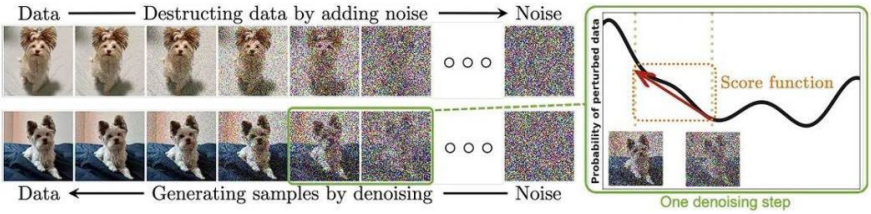

Fig. 1. Diffusion Models. *Source: Yang, Ling, et al. "Diffusion models: A comprehensive survey of methods and applications." ACM Computing Surveys 56.4 (2023): 1–39.*

Generative adversarial networks utilize two different neural networks: a generator and a discriminator, shown in Figure 2. The first generates images based on a given dataset, while the discriminator attempts to differentiate between the generated images and real images. When the generator is able to fool the discriminator, it results in a discriminator loss. If the discriminator is able to detect which of the images are generated, it results in a generator loss. Each loss causes the network to continue to improve itself. Thus, the two are in a constant feedback loop, constantly improving each other, with the end goal being a generator that can fool the discriminator and the human eye completely. However, despite it being a relatively quick model, it is unstable and uncommon compared to its diffusion counterpart.

Despite the differences in their methodology, one major commonality is the use of scrapers when gathering training data. Data scraping uses a program to extract information from websites, storing it in a database. The next session will provide a few methods intended to prevent or protect images to be scrapped by the AI models.

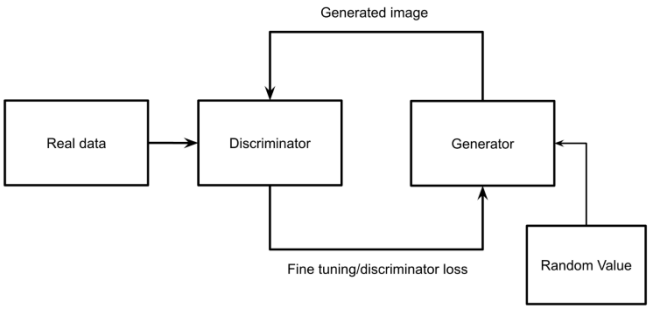

Fig. 2. Generative Adversarial Network Diagram.

4 Protection Methods

4.1 Copyright

Examining AI under copyright law is currently difficult, as it is largely uncharted territory. Most notably, the question of whether or not AI is considered fair use is in play. Within the US, the act of data mining and scraping from the internet is not considered illegal; it is considered fair use. Moreover, each case of generative AI usage has different circumstances and may be judged differently on being a transformative work [16]. Whether or not a work is transformative depends on a number of factors, including if it is commercial or non-commercial. Because generative AI aims to compete with the same market its data is from, the right to fair use is questioned [17]. AI usage must be looked at on a case-by-case basis in order to find whether or not they are negatively affecting the market using the artists' own data.

In the EU, the GDPR (General Data Protection Regulation) system allows artists to license their work and give or deny consent for its use [18]. However, the management of such a licensing and consent system would be inefficient and impractical, to say the very least. With the sheer scale of AI models, it is virtually impossible to contact and manage consent from every contributor. This leaves companies without a coherent and ethical system of obtaining all of their needed data.

Overall, many artists are morally opposed to their works being scraped by AI models. Companies are profit-focused, leading them to take offers from AI companies if the conditions create more profit [19]. While artists and large companies are in opposition, their motivations differ, again lending to the idea that copyright is not an efficient route to take. Thus, in the law, different ways must be found to protect artists.

4.2 Watermarking and Metadata

There are two types of watermarking, visible and invisible watermarks. Visible watermarking refers to a logo or image superimposed on top, while invisible watermarks are

embedded within the host data and have no effect on the visual characteristics of the image. There are many applications for digital watermarking, however, they can be attacked using simple image modifications, such as rotation, cropping, and other minor changes [20]. Therefore, digital watermarks need to be built in a way that maximizes robustness, that is, they must be able to withstand all sorts of edits. Because of this, watermarks, even ones within an image's host data, are difficult to fully rely on.

In a similar vein, a string or tag in an image's metadata could also be used to signify whether or not the owner indicates consent for training in the future. As it is relatively simple to modify an image's metadata using text, this method is accessible towards artists as well. The issue lies in the plethora of untagged images. It is highly likely that most datasets will continue to use images that neither prohibit nor allow use in AI models, as not using them would mean a drop in model efficiency [21]. With the many inactive artists whose images are still on the internet and used for training, it is not certain how effective this method will be either. Though metadata tagging seems promising, lack of knowledge or exposure surrounding it may also prevent it from taking off or becoming widely prevalent, similar to the issues with the following method.

4.3 Third-Party Protection

Through the aid of other organizations, an available option to artists is protection from a third party. An example of this would be through the sites "ArtStation" [22] or "Cara" [23], which are art portfolio sites. They offer a "NoAI" tag, placing a HTML tag within the work's metadata that prevents data scraping tools from using it [22]. On the surface, this seemingly could prevent all unauthorized use of art in AI, but there has not been any recorded proof of this having any effect. To add on, ArtStation still allows the distribution of AI generated images on their site, propagating the use of datasets stolen from countless artists. Also, this "NoAI" tag feature is automatically disabled, leaving artists to figure out its existence.

Besides the above methods, there is software made specially for the purpose of making it impossible to train certain images and "poisoning" AI datasets. Glaze and Nightshade by UChicago are both free-to-use resources which allow artists to protect their works through adversarial attacks. Glaze makes the work virtually indistinguishable to the human eye, yet is able to confuse the AI into not recognizing the subject. This is achieved by identifying and isolating elements of a work specific to its style, then cloaking it. As most text-to-image models are based around style mimicry, this prevents the process [8]. Additionally, Nightshade is a program which uses prompt-specific poisoning to hinder an AI model's ability to generate material for certain prompts [9]. Despite this, the process is not streamlined for artists who are possibly unfamiliar with this area of technology, making it relatively unpopular.

4.4 Withdrawal and Machine Unlearning

With certain methods, data can be removed and forgotten by a machine, called machine unlearning. Largely, this has been theorized for the purpose of correcting models which are biased [24] or for security. In this context, however, it can be useful for removing

large portions of copyrighted artwork, through comparisons of AI-generated images and copyrighted non-AI-generated art. This can be done through several methods, such as training non-copyrighted works of a similar nature or removing the influence of copyrighted works on the model [25]. Despite this, the main goal of unlearning is to retain as much model usability as possible, which would still mimic and regurgitate artists' styles [26]. The same issue of contacting all artists whose works are in a model's dataset would also be relevant here, as creative works created recently are technically protected by copyright by default (though this is not the same as registering ownership). To continue, it has been shown that despite a concept being erased from a model, it's still possible to generate something directly related to that concept [27]. Machine unlearning technology is still incredibly finicky as of now, with many people receiving mixed results.

5 Survey and Results

The survey first aims to acquire the profile of the artist, such as the medium they work in and their online presence or lack thereof. It then asks whether or not artists use any currently available protection on their art and why or why not. The survey was run over the period of a month, and received 39 responses at the time of this data analysis.

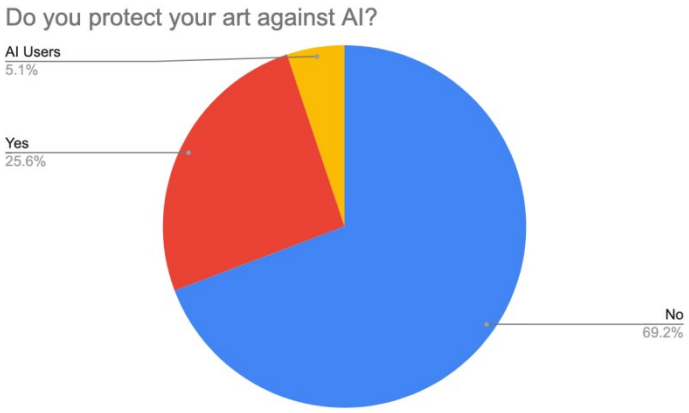

Fig. 3. Percentage of Artists Using AI Protection.

Figure 3. shows the percentage of people who used available protective resources, 25.6%, compared to the percentage who did not, 69.2%. Alongside this, there were two artists using AI already, accounting for the remaining 5.1%.

Figure 4. represents the protection methods chosen by artists who answered "Yes" on the previous figure. Some artists used multiple methods, however there were a total of 11 respondents. The majority of respondents used Glaze. Despite this, only 3 people utilized Nightshade, created by the same group as Glaze. Two artists rely on Cara, which is a portfolio sharing platform for artists that prohibits the sharing of AI-

generated images.[16] Similar to Artstation, Cara uses the "NoAI" tag that notifies AI scrapers to not scrape tagged images, but this system is built on the honor code, meaning that AI scrapers are able to use tagged images regardless. At last, a number of artists also chose to not upload their work, preventing any AI scrapers from reaching it in the first place.

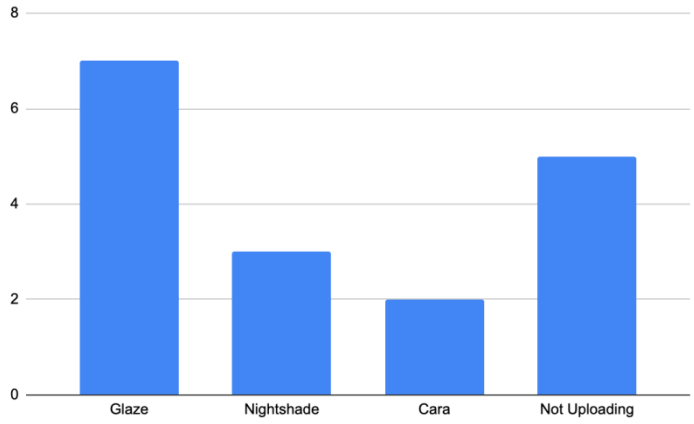

Fig. 4. AI Protection Tool Usage.

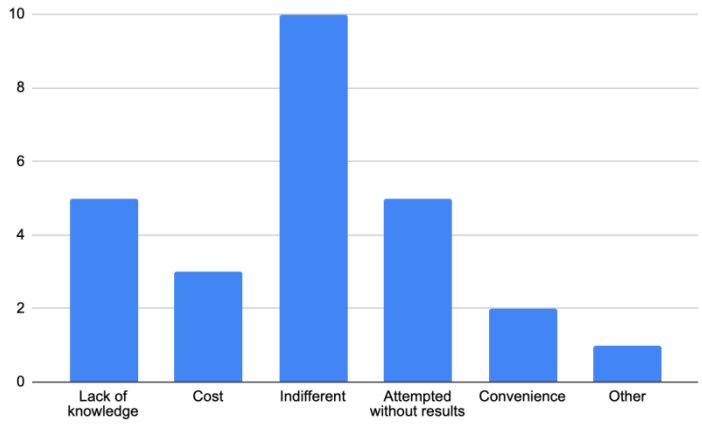

Fig. 5. Reasons for AI Protection Tool Non-Usage.

Figure 5. represents the reasons listed for not using protective resources by artists who selected "No" in Figure 1. Some artists listed multiple reasons, however there were a total of 26 respondents. First of all, some artists do not know where to find protective resources. Others do not have the available income to afford protection. Some stated that they tried to use resources, but they caused technical difficulties: "I tried to use a protecting tool twice (I don't remember the name), but it crashed my computer and the

artworks they put as examples didn't [have] noticeable changes, so I thought it didn't work well." Other artists also saw a worsened image quality and differences within artworks. Another inconvenience is that it is another process in the step, with one respondent saying they would use these resources "if it was automatic on export from [Adobe] Lightroom."

Lastly, most answered in an indifferent fashion, whether that be apathy towards their works being used in AI or a pessimistic outlook on trying to defend themselves from AI, leading to non-action. Survey responses include "It is a futile effort. All information can and is accessed by someone with [the] ability to do so," and "...there's a plethora of content they can scrape already, mine literally won't make a difference." These indicate an idea of helplessness throughout artists dealing with generative AI, that their work is a drop in an ocean of other content. Contrasting this, another (the "other" marked in Figure 5) says "I don't care if my art is used in training datasets."

6 Conclusion and Future Work

Despite the plethora of available protection for artists, the propagation of generative AI overpowers artists. Furthermore, many artists do not use these resources by choice, the most indicated answer in the survey. Part of this is a lack of trust in these resources, there is no evident proof either way. It is also unsurprising that many artists lack knowledge of protective resources; in comparison to AI in advertisements and billboards in daily life, protective resources are mainly spread by word of mouth. In order to improve what is offered now, many stated features along the line of minimal steps, available at no cost or low cost, and proved reliability. Conversely, some believed that instead of trying to push back technologically, protection should come from the law. As AI continues to rapidly progress, artists and researchers must also be prepared to collaborate in kind.

References

1. Kawakami, R. and Venkatagiri, S., 2024, June. The Impact of Generative AI on Artists. In Proceedings of the 16th Conference on Creativity & Cognition (pp. 79-82).
2. Shi, J., Jain, R., Duan, R. and Ramani, K., 2023. Understanding Generative AI in Art: An Interview Study with Artists on G-AI from an HCI Perspective. arXiv preprint arXiv:2310.13149.
3. Choi, E.D., 2023. Protecting Visual Artists from Generative AI: An Interdisciplinary Perspective. In ICML Workshop on Generative AI and Law.
4. Šarčević, T., Karlowicz, A., Mayer, R., Baeza-Yates, R. and Rauber, A., 2024. U Can't Gen This? A Survey of Intellectual Property Protection Methods for Data in Generative AI. arXiv preprint arXiv:2406.15386.
5. Somepalli, G., Singla, V., Goldblum, M., Geiping, J. and Goldstein, T., 2023. Diffusion art or digital forgery? investigating data replication in diffusion models. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 6048-6058).
6. Heikkilä, M., 2022. The Algorithm: AI-generated art raises tricky questions about ethics, copyright, and security. Artificial Intelligence.

7. Wenzel, K., Kaufman, G. and Dabbish, L., 2023. Beyond Fairness: Alternative Moral Dimensions for Assessing Algorithms and Designing Systems. arXiv preprint arXiv:2312.12559.
8. Shan, S., Cryan, J., Wenger, E., Zheng, H., Hanocka, R. and Zhao, B.Y., 2023. Glaze: Protecting artists from style mimicry by {Text-to-Image} models. In the 32nd USENIX Security Symposium (USENIX Security 23) (pp. 2187-2204).
9. Shan, S., Ding, W., Passananti, J., Wu, S., Zheng, H. and Zhao, B.Y., 2024, May. Nightshade: Prompt-Specific Poisoning Attacks on Text-to-Image Generative Models. In 2024 IEEE Symposium on Security and Privacy (SP) (pp. 212-212). IEEE Computer Society.
10. Lovato, J., Zimmerman, J., Smith, I., Dodds, P. and Karson, J., 2024. Foregrounding Artist Opinions: A Survey Study on Transparency, Ownership, and Fairness in AI Generative Art. arXiv preprint arXiv:2401.15497.
11. Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A. and Bengio, Y., 2014. Generative adversarial nets. Advances in neural information processing systems, 27.
12. Karras, T., Laine, S. and Aila, T., 2019. A style-based generator architecture for generative adversarial networks. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 4401-4410).
13. Nguyen, Loc X., Pyae Sone Aung, Huy Q. Le, Seong-Bae Park, and Choong Seon Hong. "A new chapter for medical image generation: The stable diffusion method." In 2023 International Conference on Information Networking (ICOIN), pp. 483-486. IEEE, 2023.
14. Jaruga-Rozdolska, Anna. "Artificial intelligence as part of future practices in the architect's work: MidJourney generative tool as part of a process of creating an architectural form." Architectus 3 (71 (2022): 95-104.
15. Marcus, Gary, Ernest Davis, and Scott Aaronson. "A very preliminary analysis of DALL-E 2." arXiv preprint arXiv:2204.13807 (2022).
16. Zhong, H., Chang, J., Yang, Z., Wu, T., Mahawaga Arachchige, P.C., Pathmabandu, C. and Xue, M., 2023, April. Copyright protection and accountability of generative ai: Attack, watermarking and attribution. In Companion Proceedings of the ACM Web Conference 2023 (pp. 94-98).
17. Rahman, N. and Santacana, E., 2023. Beyond fair use: Legal risk evaluation for training LLMs on copyrighted text. In ICML Workshop on Generative AI and Law.
18. Zhang, D., Xia, B., Liu, Y., Xu, X., Hoang, T., Xing, Z., Staples, M., Lu, Q. and Zhu, L., 2024, April. Privacy and Copyright Protection in Generative AI: A Lifecycle Perspective. In Proceedings of the IEEE/ACM 3rd International Conference on AI Engineering-Software Engineering for AI (pp. 92-97).
19. de la Durantaye, K., 2023. Garbage In, Garbage Out. Regulating Generative AI Through Copyright Law. Regulating Generative AI Through Copyright Law (August 28, 2023). ZUM, 10, pp.645-660.
20. Lipiński, P., 2011. Watermarking software in practical applications. Bulletin of the Polish Academy of Sciences: Technical Sciences, (1).
21. Ippolito, D. and Yu, Y.W., 2023. DONOTTRAIN: A Metadata Standard for Indicating Consent for Machine Learning. In Workshop on Generative AI and Law.
22. "NOAI Tagging on Projects - Artstation Magazine." ArtStation Magazine, 16 Dec. 2022, magazine.artstation.com/2022/12/noli-tag/.
23. https://blog.cara.app/blog/faq
24. Cherepanova, V., Reich, S., Dooley, S., Souri, H., Dickerson, J., Goldblum, M. and Goldstein, T., 2023, August. A deep dive into dataset imbalance and bias in face identification.

In Proceedings of the 2023 AAAI/ACM Conference on AI, Ethics, and Society (pp. 229-247).
25. Mantri, K., Sasikumar, N., 2024. Developing Methods for Identifying and Removing Copyrighted Content from Generative AI Models. In ICML Workshop on Generative AI and Law.
26. Nguyen, T.T., Huynh, T.T., Nguyen, P.L., Liew, A.W.C., Yin, H. and Nguyen, Q.V.H., 2022. A survey of machine unlearning. arXiv preprint arXiv:2209.02299.
27. Liu, Ken Ziyu. "Machine Unlearning in 2024." Ken Ziyu Liu - Stanford Computer Science, May 2024, ai.stanford.edu/~kzliu/blog/unlearning.