The Revealed Preferences of Pre-authorized Licenses and Their Ethical Implications for Generative Models

Vinith M. Suriyakumar [1] **Peter Menell** [2] **Dylan Hadfield-Menell*** [1] **Ashia Wilson*** [1]

Abstract

This work examines the revealed preferences of creators as reflected in open and quasi-open licensing regimes based on the most commonly used licenses by copyright holders of images in the Creative Commons and copyright holders of code in GitHub code repositories. We discuss the ramifications these preferences and licenses might have absent a determination that training of generative AI and its associated outputs constitute fair use.

1. Introduction

Generative AI models, which create text, images, audiovisual works, and other multimedia resembling human creativity introduce new challenges and prospects for artists and rightsholders. As AI systems become more sophisticated, their ability to autonomously generate works indistinguishable from those created by humans raises critical questions about copyright protection (originality, authorship, direct and indirect liability, defenses, and remedies), the industrial organization of creative activities, and ultimately social justice in a post-generative AI world.

Ongoing lawsuits between generative AI developers (such as OpenAI, Google, Meta, Stable Diffusion) and rightsholders (such as The New York Times, Getty Images, and author classes broadly) have amplified debate in the legal community about the propriety of ingestion of copyright-protected internet ”data” without authorization and the generation of outputs remixing such ”data” (Reisner, 2024). While the legal question of what constitutes fair use and the scope of pre-authorized licenses are central to how these tools will develop, we focus here on the ethical question of what it would mean to respect artist and rightsholder preferences.

*Equal contribution [1]Department of Electrical Engineering and Computer Science, MIT, USA [2]School of Law, UC Berkeley, USA. Correspondence to: Vinith M. Suriyakumar <vinithms@mit.edu>.

Proceedings of the 41^{st} International Conference on Machine Learning, Vienna, Austria. PMLR 235, 2024.

Initial attempts to understand what is ethical and desirable have either surveyed artists and rightholders (Lovato et al., 2024) or speculated about their preferences based on existing academic sources (Jiang et al., 2023; Attard-Frost, 2023; Latikka et al., 2023; Brunder). These surveys and analyses capture a skewed sample of artists/rightsholders and are only able to investigate the *stated* preferences of these artists. Yet, the *stated* preferences of individuals does not always align with their beliefs and actions (De Corte et al., 2021; Craig et al., 2017). Thus, investigating the *revealed* preferences of artists and rightsholders provides a fuller understanding of the ethics of ingestion and generative outputs.

Our work expands the ethical landscape by examining the revealed preferences of creators as reflected in open and quasi-open licensing regimes. We analyze the most commonly used licenses by copyright holders of images in the Creative Commons and copyright holders of code in GitHub code repositories. We discuss the ramifications that these licenses might have on the existing generative AI training models. Finally, we discuss the technical affordances needed from the AI community to meet the artists' and rightsholders' license conditions.

2. Background

Initial analyses of the desires of artists and rightsholders as well as what is considered to be ethical regarding generative models have focused on surveying these individuals or drawing speculation based on existing academic sources. Lovato et al. (2024) through a survey of 459 artists found that the majority felt that AI art is a threat to their own work and that model developers should be required to disclose the art used in their training data. There was less agreement on whether works produced using generative models should be owned by the original artists whose work contributed the output or the user who prompted the model to generate the output. Half of respondents did not feel the need to be compensated. They were, however, concerned with for-profit companies profiting from the outputs of models trained on their art. Overall, this suggests that many artists are comfortable with their art being used in the training of generative models as long as for-profit companies are not benefiting financially from these outputs.

Other explorations of the effects of generative AI models on artists and rightsholders survey the academic literature. Jiang et al. (2023) comment on the potential harms that artists will face due to AI art and how the AI art fits within the current copyright law. They also investigate regulations and technical tools that could help prevent some of these harms. Their analyses are grounded in examples of ongoing lawsuits in the US but do not include surveys or desires directly from artists / rightsholders themselves. They identify economic loss, potential forgery, stereotyping, and the slowing of cultural production and consumption as potential harms. Attard-Frost (2023) comment on the impact of generative AI art on Canadian artists and gaps in the Canadian law to address these harms. They describe similar harms with real world examples as those presented by Jiang et al. (2023).

2.1. Stated *vs.* Revealed Preferences

All of the studies discussed are based on either analyses from the authors of current lawsuits or surveys from artists themselves. The surveys are the closest to better understanding the proper technical tools and ethical considerations to help address their concerns. Yet there are a couple of issues with these surveys that can be improved upon to better understand the necessary technical affordances and ethical considerations. First, most of these surveys are biased towards artists with an interest in generative models and digital artists. The sample is also biased towards artists that are not renowned or fully established because successful artists can be more difficult to contact or reluctant to participate in such surveys. Thus, there is an important subset of artists whose opinions are underrepresented in these analyses.

The second issue is that the stated preferences of artists and rightsholders can be misleading. Various studies show individuals often derive positive utility from expressing opinions that reflect social responsibility and generosity (De Corte et al., 2021; Taylor & Brown, 1994), especially when those opinions are not binding (Kahneman & Knetsch, 1992). Artists might want to signal to the public that they support sharing of creative works notwithstanding their personal interest in profiting from such works. The controversy over file-sharing during the early 2000 period illustrated this phenomenon. Many successful recording artists were reluctant to express their opposition out of concern that it would alienate fans or be seen as aligning with record labels (see e.g. Menell (2013; 2002); Nimmer & Menell (2001) describing the recording industry's backroom legislative effort to extinguish recording artist ability to terminate copyright transfers). Yet most successful songwriters and recording artists lent their names to briefs filed in opposition to the file-sharing companies. (See Brief of Amici Curiae National Academy of Recording Arts & Sciences, et al., in Support of Petitioners, Metro-Goldwyn-Mayer Studios, Inc., et. al., v. Grokster, Ltd., 2005 (comprising over 20,000 members including including Jimmy Buffet, Sheryl Crow, Don Henley, Billy Joel, Alanis Morisette, Stevie Nicks, Bonnie Raitt, Bruce Springsteen, and Trisha Yearwood) (National Academy of Recording Arts & Sciences, 2005); Brief of the American Federation of Musicians of the United States and Canada, et al., in Support of Petitioners, Metro-Goldwyn-Mayer Studios, Inc., et. al., v. Grokster, Ltd., 2005 (comprising over 300,000 musicians and performers) (American Federation of Musicians of the United States and Canada, 2005). Such strategic motivations could well skew survey results toward a "copyleft" perspective on generative AI.

License	Attribution Required	Remixing Allowed	Commercial Use Allowed
CC BY	✓	✓	✓
CC BY-NC	✓	✓	✗
CC BY-SA	✓	✓	✓
CC BY-NC-SA	✓	✓	✗
CC BY-ND	✓	✗	✓
CC BY-NC-ND	✓	✗	✗
CC0	✗	✓	✓

Table 1. Creative Commons licenses and their requirements.

In our study, we aim to address these two issues of sample bias and hypothetical bias for studying the desires of artists and rightsholders regarding ethical generative AI. We do so by studying *revealed* preferences data across multiple domains with millions of rightsholders. In particular, we study the data generated by pre-authorized licensing in both the open source image community (i.e. Creative Commons) and the open source code community (i.e. Github).

3. Analysis

We analyze the breakdown of licenses used in two "open-source" licensing regimes: images (Creative Commons) and code (GitHub). In studying the revealed preferences of creators in these two different regimes we highlight similarities and differences between the two communities that will affect generative models for images vs. code.

3.1. Images (Creative Commons)

The Creative Commons (CC) is an international non-profit organization established in 2001 with the mission of enabling easier and more ethical use of copyrighted works. The CC organization generally reflects an open philosophy, although it offers users a range of pre-authorized licensing options. Table 1 presents the six CC licenses. The salient features of the licenses for generative models are: waiver of rights (CC0), attribution (BY), authorization for editing (or remixing) (preparation of derivative works) (ND), authorization for commercial use (NC), and requirement to share alike (SA). Licensors may waive all rights or pre-authorize usage with one or more reservations of rights.

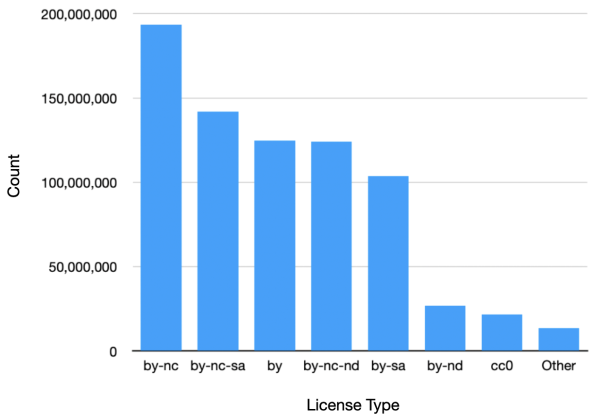

Figure 1. Breakdown of CC licensed used by images on Openverse.org.

There are currently over 2.5 billion works across the internet that use CC licenses. These span text, audio, and images. Many of these works are scraped by Common Crawl (Patel & Patel, 2020), a commonly used tool for obtaining internet-based data, and subsequently used as training data for most generative models. For our analysis, we will focus on CC-licensed images curated by Openverse (`openverse.org`). Openverse sources over 700 million CC-licensed images from open APIs (e.g. Flickr) and Common Crawl. From this database of images sourced by Openverse and engineers at Openverse we calculate the breakdown of the CC licenses used (Figure 1).

This breakdown elucidates multiple findings that are pertinent to generative model training and outputs. First, more than 90% of theses licenses require attribution. The CC does not have strict guidelines for what is considered sufficient attribution. Rather, they request reasonable attribution (rea), such as retaining a copyright notice or adding the hyperlink / URI associated with the copyrighted material. Additionally, the majority of image CC licenses do not allow commercial uses or remixing of the copyright images. Overall, it is clear that while owners of copyrighted works using CC licenses are often considered to be "copyleft," the majority nonetheless require attribution and impose restrictions on how the content is used and for what purposes. It is more accurate to characterize these licenses as promotional: you may use these images for non-commercial purposes so long as you provide reasonable attribution and you may not use these images for commercial purposes without express authorization.

3.2. Open-Source Code (GitHub)

GitHub is a cloud-based platform that allows developers to store Git repositories for their code. It is the most popular platform for open-sourcing code throughout the software community. We present the most commonly used open source licenses and their differences in Table 2. Similar to CC-licensed images, we focus on the most salient permissions and conditions of the licenses: whether license and copyright notices are required, if modification of the code is allowed, and whether commercial use is allowed. Notably, all of the licenses allow commercial use. An additional requirement that is present in pre-authorized licenses for code is the "same license" condition. This condition requires that all derivative works must use the same pre-authorized license as the copyrighted works on which they are based. This is salient for code generation because it means absent a fair use determination, any code outputs that use code with these licenses must also use the same license.

License	Copyright Notice	Modification Allowed	Commercial Use Allowed	Same License
MIT	✓	✓	✓	✗
GPLv2	✓	✓	✓	✓
Apache	✓	✓	✓	✗
GPLv3	✓	✓	✓	✓
BSD 3-clause	✓	✓	✓	✗
BSD 2-clause	✓	✓	✓	✗
LGPLv3	✓	✓	✓	✓
AGPLv3	✓	✓	✓	✗

Table 2. Open Source Software licenses and their requirements.

As of May 2024, over 420 million repositories are stored on GitHub. Lacking direct API access, we focus on an analysis of license usage conducted by GitHub in 2015 (Balter). The breakdown presented in Figure 2 indicates a much different set of preferences than the CC-licensed breakdown. The MIT and Apache licenses have similar permissions and conditions with regard to those that will have the most impact on generative model training and outputs. One striking difference with these two most commonly used licenses compared to the most commonly used Creative Commons licenses is that attribution is not required. A copyright notice is required by both of these licenses but that does not necessarily mean that attribution of the exact data used to generate an output is required. Especially since all of the licenses require copyright notices, model developers could simply append these notices to all generated code outputs. While attribution is not required, the same license condition present in GPL licenses (accounting for approximately 20% of repository) requires this technical tool to identify whether outputs have been generated using copyrighted works that have this "same license" requirement.

4. Potential Impact on Generative AI

We briefly discuss some of the potential impacts based on these revealed preferences.

State of the art and commonly used generative models for image generation such as DALLE (Ramesh et al., 2021), Stable Diffusion (Rombach et al., 2022), and Midjourney (mid)

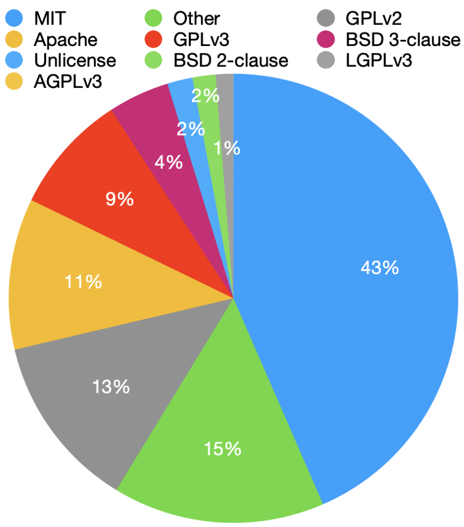

Figure 2. Breakdown of OSS licenses used by repos on GitHub.

were trained on images gathered from Common Crawl, which includes many CC-licensed images captured in the Openverse database. As reflected in Figure 1, many of the CC-licensed images do not allow commercial use. What is currently unclear is how these individual licenses affect the overall use of a generative model. For example, does the use of these CC-licensed images with an NC license in training mean that these models are not allowed to be distributed for commercial use? If so, absent a determination that such training and associated outputs constituted fair use, OpenAI and StabilityAI are already in violation of these licenses due to the commercial nature of their activities. In contrast with the impact of CC on image generation models, the majority of licenses used for GitHub repos which have been cited as a primary source for training code generation models such as GPT-4 (Chen et al., 2021) and CoPilot (Gershgorn, 2021) allow commercial use. Thus, there is no impact on the monetization of GPT-4 and CoPilot for these purposes.

4.1. Attribution

Across both regimes, it is clear that attribution is an important technical affordance, whether to satisfy the CC license conditions, to understand what copyright notices would need to be included for generated code, or to ensure that generated code has the same license in cases where the outputs are based on code that has a GPL license. Similarly, many images use a `CC ND` license which prohibits generation of derivative works. How this applies to image generation models is dependent on how we view the mechanism by which outputs are created. If we view the mechanism as simply an interpolation of the training data, that interpretation would bar most outputs that are similar to images restricted by a `CC ND` license.

Given the prevalence of `CC ND` licenses, what constitutes "reasonable attribution" for generative AI ingestion and outputs? Clearly though there is an important need for rigorous attribution methods that can identify the most relevant training data for an output and potentially attach the hyperlinks / URIs for those images as is described by Creative Commons for "reasonable attribution."

Finally, we turn to the current state of affairs for scalable and accurate attribution. A drawback of many proposed methods is that they require retraining (Park et al., 2023; Ilyas et al., 2022; Feldman & Zhang, 2020; Jain et al., 2023), which is a challenge for large scale models. Instead, one of the most commonly proposed scalable approximations is the influence function (IF), which was introduced in Hampel (1974) in the setting of robust statistics. A notable feature of influence functions is that these methods do not require retraining, which is a key advantage in recent large scale settings, and has sparked interest in developing methods principled in IF for explaining models in a black-box fashion. (Koh & Liang, 2017; Schioppa et al., 2022; Grosse et al., 2023; Choe et al., 2024). While the IF is practically appealing, there has been limited work on understanding when it fails to return accurate attributions. Koh et al. (2019) present empirical results backed by some theoretical guarantees (albeit under slightly restrictive assumptions) to help better understand the differences between the IF and the output of a true attribution method. Their key findings indicate that these approximations are not always accurate, and in fact tend to underestimate the true attribution. This raises an important questions concerning the feasibility of data-attribution methods, and what role data-attribution could play in helping to satisfy these license requirements, especially if it turns out we are unable provide accurate attributions with existing or future algorithms?

5. Conclusion

Our exploration of revealed preferences in open and quasi-open licensing regimes reveals that the legal and ethical landscape surrounding pre-training and output generation is complex and uncertain. Even for those databases and repositories that are to some extent pre-authorized for subsequent use, the case for pretraining without fuller authorization is murky. The overwhelming majority of Creative Commons-licensed images require attribution and a majority do not authorize commercial uses. Open source code licenses are more permissive, but questions remain as to the scope of pre-authorized consent. This analysis not only clarifies some of the ethical considerations of generative AI models but also underscores the necessity for the AI community to develop technical solutions for attribution or alternative training regimes that align with the preferences and conditions set forth by artists, programmers, and rightsholders.

References

Midjourney. https://www.midjourney.com/home. Accessed: 2024-6-3.

Recommended practices for attribution. https://wiki.creativecommons.org/wiki/Recommended_practices_for_attribution. Accessed: 2024-7-22.

American Federation of Musicians of the United States and Canada. Brief of the american federation of musicians of the united states and canada, et al., in support of petitioners, metro-goldwyn-mayer studios, inc., et. al., v. grokster, ltd., no. 04-480, on writ of certiorari to the united states court of appeals for the ninth circuit, 2005.

Attard-Frost, B. Generative ai systems: impacts on artists & creators and related gaps in the artificial intelligence and data act. *Available at SSRN*, 2023.

Balter, B. [no title]. https://github.blog/2015-03-09-open-source-license-usage-on-github-com/. Accessed: 2024-6-10.

Brunder, K. M. " ai art and its implications current and future artists".

Chen, M., Tworek, J., Jun, H., Yuan, Q., de Oliveira Pinto, H. P., Kaplan, J., Edwards, H., Burda, Y., Joseph, N., Brockman, G., Ray, A., Puri, R., Krueger, G., Petrov, M., Khlaaf, H., Sastry, G., Mishkin, P., Chan, B., Gray, S., Ryder, N., Pavlov, M., Power, A., Kaiser, L., Bavarian, M., Winter, C., Tillet, P., Such, F. P., Cummings, D., Plappert, M., Chantzis, F., Barnes, E., Herbert-Voss, A., Guss, W. H., Nichol, A., Paino, A., Tezak, N., Tang, J., Babuschkin, I., Balaji, S., Jain, S., Saunders, W., Hesse, C., Carr, A. N., Leike, J., Achiam, J., Misra, V., Morikawa, E., Radford, A., Knight, M., Brundage, M., Murati, M., Mayer, K., Welinder, P., McGrew, B., Amodei, D., McCandlish, S., Sutskever, I., and Zaremba, W. Evaluating large language models trained on code. July 2021.

Choe, S. K., Ahn, H., Bae, J., Zhao, K., Kang, M., Chung, Y., Pratapa, A., Neiswanger, W., Strubell, E., Mitamura, T., et al. What is your data worth to gpt? llm-scale data valuation with influence functions. *arXiv preprint arXiv:2405.13954*, 2024.

Craig, A. C., Garbarino, E., Heger, S. A., and Slonim, R. Waiting to give: stated and revealed preferences. *Management Science*, 63(11):3672–3690, 2017.

De Corte, K., Cairns, J., and Grieve, R. Stated versus revealed preferences: An approach to reduce bias. *Health economics*, 30(5):1095–1123, 2021.

Feldman, V. and Zhang, C. What neural networks memorize and why: Discovering the long tail via influence estimation. *Advances in Neural Information Processing Systems*, 33:2881–2891, 2020.

Gershgorn, D. GitHub and OpenAI launch a new AI tool that generates its own code. https://www.theverge.com/2021/6/29/22555777/github-openai-ai-tool-autocomplete-code, June 2021. Accessed: 2024-6-3.

Grosse, R., Bae, J., Anil, C., Elhage, N., Tamkin, A., Tajdini, A., Steiner, B., Li, D., Durmus, E., Perez, E., et al. Studying large language model generalization with influence functions. *arXiv preprint arXiv:2308.03296*, 2023.

Hampel, F. R. The influence curve and its role in robust estimation. *Journal of the american statistical association*, 69(346):383–393, 1974.

Ilyas, A., Park, S. M., Engstrom, L., Leclerc, G., and Madry, A. Datamodels: Predicting predictions from training data. *arXiv preprint arXiv:2202.00622*, 2022.

Jain, S., Salman, H., Khaddaj, A., Wong, E., Park, S. M., and Madry, A. A data-based perspective on transfer learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 3613–3622, 2023.

Jiang, H. H., Brown, L., Cheng, J., Khan, M., Gupta, A., Workman, D., Hanna, A., Flowers, J., and Gebru, T. Ai art and its impact on artists. In *Proceedings of the 2023 AAAI/ACM Conference on AI, Ethics, and Society*, pp. 363–374, 2023.

Kahneman, D. and Knetsch, J. L. Valuing public goods: the purchase of moral satisfaction. *Journal of environmental economics and management*, 22(1):57–70, 1992.

Koh, P. W. and Liang, P. Understanding black-box predictions via influence functions. In *International conference on machine learning*, pp. 1885–1894. PMLR, 2017.

Koh, P. W. W., Ang, K.-S., Teo, H., and Liang, P. S. On the accuracy of influence functions for measuring group effects. *Advances in neural information processing systems*, 32, 2019.

Latikka, R., Bergdahl, J., Savela, N., and Oksanen, A. Ai as an artist? a two-wave survey study on attitudes toward using artificial intelligence in art. *Poetics*, 101:101839, 2023.

Lovato, J., Zimmerman, J., Smith, I., Dodds, P., and Karson, J. Foregrounding artist opinions: A survey study on transparency, ownership, and fairness in ai generative art. *arXiv preprint arXiv:2401.15497*, 2024.

Menell, P. S. Envisioning copyright law's digital future. *NYL Sch. L. Rev.*, 46:63, 2002.

Menell, P. S. American copyright life: Reflections on re-equilibrating copyright for the internet age, this. *J. Copyright Soc'y USA*, 61:235, 2013.

National Academy of Recording Arts & Sciences. Brief of amici curiae national academy of recording arts & sciences, et al., in support of petitioners, metro-goldwyn-mayer studios, inc., et. al., v. grokster, ltd., no. 04-480, on writ of certiorari to the united states court of appeals for the ninth circuit, 2005.

Nimmer, D. and Menell, P. S. Sound recordings, works for hire, and the termination-of-transfers time bomb. *J. Copyright Soc'y USA*, 49:387, 2001.

Park, S. M., Georgiev, K., Ilyas, A., Leclerc, G., and Madry, A. Trak: Attributing model behavior at scale. *arXiv preprint arXiv:2303.14186*, 2023.

Patel, J. M. and Patel, J. M. Introduction to common crawl datasets. *Getting structured data from the internet: running web crawlers/scrapers on a big data production scale*, pp. 277–324, 2020.

Ramesh, A., Pavlov, M., Goh, G., Gray, S., Voss, C., Radford, A., Chen, M., and Sutskever, I. Zero-shot text-to-image generation. In *International conference on machine learning*, pp. 8821–8831. Pmlr, 2021.

Reisner, A. Generative AI is challenging a 234-Year-Old law. *The Atlantic*, February 2024.

Rombach, R., Blattmann, A., Lorenz, D., Esser, P., and Ommer, B. High-resolution image synthesis with latent diffusion models. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 10684–10695, 2022.

Schioppa, A., Zablotskaia, P., Vilar, D., and Sokolov, A. Scaling up influence functions. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 36, pp. 8179–8186, 2022.

Taylor, S. E. and Brown, J. D. Positive illusions and well-being revisited: separating fact from fiction. 1994.