

**Artificial Intelligence and Copyright Protection:
Exploring the Boundaries of Law, Technology, and Ethics**

Nicholas Mangan

Department of Philosophy, Whatcom Community College

PHIL 301: Professional Ethics

Professor Jim Drinkwine

June 8, 2023

In this paper, we will examine generative artificial intelligence, specifically the ethical and legal implications of using existing art to train artificial intelligence to produce more art. Taking a consequentialist perspective, we will use the outcomes of past legislation and jurisprudence on copyright to argue that art generated by an AI trained using public data (copyrighted or not) ought to belong to the public domain, as any other approach represents a deviation from the ethical purpose of copyright.

In his three-part series of appearances on NPR's *On the Media*, noted technical columnist and author Cory Doctorow explains what he calls "Enshittification," a theory of monopolistic process akin to predatory pricing with which Doctorow explains the decline of user experience in many popular platforms. Platforms initially operate at a loss, offering desirable services for free to attract users. They then make changes to benefit business customers (i.e. advertisers) at the expense of user experience. Finally, they impose unfavorable terms on business customers to make money for themselves. At each change, the group getting the short end of the stick is prevented from leaving by the destruction of competing services which occurred during the initial phase while the company was operating at a loss (Doctorow, 2023b; Doctorow, 2023c). This is relevant to the debate over AI ethics because it explains the perennial abuses of tech platforms through simple economic incentives rather than resorting to "tech exceptionalism," a concept which Doctorow explains in this context in his Medium article *Twiddler*. In brief, tech exceptionalism is the idea that the internet represents a fundamental paradigm shift with respect to all that came before it, and that all unique qualities of tech companies (good and bad) are inherent to the technology rather than being the results of the same social and economic forces which drive the rest of the world (Doctorow, 2023a). Doctorow argues against this here and in his *On the Media* appearance, contending that most if not all of the uniquely monopolistic

and abusive practices in the tech world arise from lack of regulation and the speed at which the digital landscape changes (Doctorow, 2023a; Doctorow, 2023d). Doctorow's stance against tech exceptionalism when criticizing the actions of tech companies makes for a useful lens to understand the (il)legalities of AI training – at the very least it cautions us to attempt to apply existing standards of ethics regarding intellectual property before resorting to novel ethical formations constructed around an over-hyped image of new technology. The behavior of these platforms, and their ability to alter and block regulation to suit their needs (Doctorow, 2023c), speaks to the futility of banning or restricting AI entirely – such a move would inevitably serve only to enrich those companies powerful enough to circumvent it.

A program is an example of generative artificial intelligence if, given a collection of training data belonging to some category (the precise definitions of “data” and “category” here left deliberately ambiguous), that program is designed to produce new data, which belong in the category but do not appear in the training data set, in response to a prompt (Karpathy et al., 2016). Presently, prominent examples of generative AI include Stable Diffusion, which generates images (Rombach et al., 2023), and ChatGPT, which generates text (OpenAI, 2022). The former has been trained on images paired with descriptions of their content, and produces new images in response to descriptions prompted by the user (Rombach et al., 2023). The latter has been trained on text and functions as a “chatbot,” maintaining one side of a conversation with its prompter (OpenAI, 2022). Of note here is the data used to train these models was procured by scraping the internet (Rombach et al., 2023; OpenAI, 2022), and thus was used for this purpose without the explicit permission of its authors.

At the heart of any copyright debate is a potential tension between consequentialism and deontology emerging from two not-always-aligned purposes of copyright as a legal construct.

Consequentialism, specifically its sub-field utilitarianism, evaluates actions based on their results according to some metric (Ferrell et al., 2021, p. 180). Deontology, conversely, views ethics from the perspective of the inalienable rights of individuals and the intentions of their actions (Ferrell et al., 2021, p. 180). In his article on the moral justification of copyright, Tom G. Palmer contrasts the deontological conception of copyright as deriving from “the natural law right to the fruit of one’s labor” (Palmer, 1990, p. 819) with its utilitarian justification as a driver of innovation and creation (Palmer, 1990, p. 849). Within this consequentialist framework, he identifies a secondary conflict arising from the question of whether copyright is (or can be) optimized to incentivize creation sufficiently to offset the limitations it places on creating derivative works (Palmer, 1990, p. 849). If we provisionally accept the premise that training generative AI on a creative work is a form of copying, the debate over its permissibility becomes this conflict over the ethics of copying and copyright in general – by no means a solved problem, but one outside the scope of this paper. If, on the other hand, we reject that premise and claim that AI is producing wholly new works, then these conflicts have no bearing at all. To evaluate AI on these terms it is therefore necessary to invent a “sliding scale of originality;” in which AI art is neither wholly novel nor wholly derived from the AI’s training data.

It falls to a business considering AI art to determine where on this scale their use of it lies. As discussed, the extremes yield fairly clear-cut answers: If AI art is wholly original, then copyright of training data need not be considered. If it is wholly derivative, then one ought to behave as though one has the same obligations to the creators of training data as one does to the authors of original works of which one wishes to create a derivative. In practice, owing to the impossibility of securing permission from millions of anonymous contributors to an AI’s data set, this latter option amounts to total abstention from the use of any AI model trained on data one does not own the rights to. It is questionable whether taking a position anywhere but one of these extremes is ethically or legally

feasible – legally, at least, whether a work infringes on copyright is black and white. Similarly, if we conclude that AI training data creators are entitled to some arbitrarily small portion of full compensation for their work, inversely proportional to the degree of “originality” we impute to the AI, what precise portion we decide on is irrelevant – if it is non-zero, the aforementioned number and anonymity of contributors makes it impossible to deliver. The deontological perspective is therefore unhelpful here. However, while the impossibility of mass compensation may seem to also obviate the creativity-incentivizing aspect of copyright’s utilitarian justification, it only does so for one side of the utilitarian calculus on copyright. From a utilitarian standpoint, if the goal of copyright is to maximize creative output, it must balance its usefulness for rewarding original artists with its chilling effect on the creation of new art and expression in general (Palmer, 1990, p. 849). A hypothetical copyright regime which allowed the ownership of entire art styles, for example, would reward a select few stylistic pioneers with protection while suppressing countless more artists who would have otherwise made art in an existing style. In the special case of AI trained in such a way that its training data’s originators cannot be compensated, copyright’s utility is best optimized simply by getting out of the way – not preventing the creation of AI art with impossible compensation requirements, but not protecting it either. Were AI art to be protected without granting any rights to its (anonymous, innumerable) trainers, it would in effect be stealing their own art from them. In brief, if we concede that any amount of credit for AI art is due to the artists who produced its training data, then the only utilitarian copyright regime which allows it to exist at all is one which completely ignores it.

It is difficult to have a productive conversation on how different or similar a given AI is to a human being. In the field of artificial intelligence, there exists a social phenomenon in which the popular conception of “intelligence” tends to change to exclude anything machines become capable of (McCorduck, 2004, p. 204). In her wider argument against human-exceptionalist discounting of AI

capabilities, AI philosopher Pamela McCorduck writes that attempts to prove that AI has or lacks human intelligence “assume without question that we know and agree upon the nature of intelligence”(McCorduck, 2004, p. 203). The position McCorduck finds prevalent among AI’s skeptics and detractors hearkens to the tech exceptionalism Doctorow warns us against – rather than demonstrating that humans are uniquely intelligent, the dogmatic approach to AI criticism seems to hold that machines are uniquely unintelligent, and that anything a machine becomes capable of ceases to be construed as a sign of intelligence. This kind of goalpost-moving is present on both sides of the debate on AI training data. In a joint symposium held in 2020 by the U.S. Copyright Office and the World Intellectual Property Organization, Rutgers computer science professor Ahmed Elgammal uses an explanation of the function of generative adversarial networks (a type of AI often used in image generation) to outline a viewpoint in which, implicitly, AI art is the original work of the person who prompted the AI to produce it. He explains that a generative adversarial network is really two separate neural networks, one (the discriminator) trained to distinguish AI art from human and another (the generator) trained to produce art that will fool the discriminator into classifying it as human art. Crucially, the generator (which is ultimately responsible for the AI art output of these models) never has access to any of the training data (human art). Rather, it initially produces random visual noise and is guided towards more art-like output by the feedback of the discriminator, which does have access to the training data (United States Copyright Office et al., 2020, pp. 41–42). This description is corroborated by a paper by Goodfellow et al. (2014) which introduced the concept of generative adversarial networks. This disconnection between the part of the AI which “learns” and the part of the AI which has access to the training data is evidence that generative AI of this type does not simply regurgitate its training data or collage pieces of it into a fundamentally unoriginal gestalt. In addition to prompting, users also curate the output of generative AI, choosing which of many output images to keep or discard. In the same symposium, Rob Kasunic, an officer of the US copyright office, treats this

mechanistic production of new content in response to minimal user input as analogous to photography and digital image manipulation – the use of a tool to automate what remains a fundamentally human-created art piece (United States Copyright Office et al., 2020, pp. 55–56, 59). However, he cautions that, at least in the eyes of the law, works lacking this human input lack copyright protection, using as an example paintings created and photographs taken by animals (United States Copyright Office et al., 2020, pp. 60–62).

From this legal standpoint we find a precedent for what we already established as the most utilitarian option for AI art trained on mass-gathered data: a complete lack of both protection and restriction. By classifying generative AIs among the non-human entities not entitled to copyright protection, we can eke out an imperfect but minimally unjust approach to AI. Unfortunately, artists whose work has been used to train AI will receive no compensation – but the output of the models it trains will be free to the public to remix and recreate forever. This satisfies at least one purpose of copyright: to incentivize the creation of new art.

To justify taking this approach over mandating the use of exclusively paid training data (the seemingly ideal solution), we turn once more to Doctorow. In his Medium article, he explains the hazard of compliance costs in the context of regulations. He describes the case of a European Union copyright protection mandate requiring platforms that host user content to scan for illegally uploaded copyrighted material. Despite being costly to all platforms, this mandate was supported by the likes of YouTube and Facebook because they could bear the costs and smaller platforms couldn't – it reduced their competition (Doctorow, 2023a). Similarly, we might expect that limiting AI to companies with the resources to hire people to produce training data would have monopolistic results – already-large companies that could afford to train on larger data sets would improve their AIs faster, leaving poorer

competitors unable to catch up. Ultimately, Doctorow's "Enshittification" would ensue – a single AI company, having dominated the market, would use its power to insist on lower and lower prices for new training data – leading to a situation where artists' work is still stolen for AI training, but for the benefit of a single company rather than for everyone as in the no-copyright model.

Overall, this is a very frustrating topic. Through a combination of practicality and precedent, we have come to the conclusion that the least awful way to produce AI art is the one in which no one gets paid – training data is taken freely, and output is public domain. This is a reasonably fair solution guaranteed to leave a bad taste in the mouth of everyone involved. The reason copyright is a useful thing to have in a society is that it rewards artists for creating art, thus encouraging them to create more of it. If an organization can profit from AI art without compensating the artists whose art was used to train the AI, and can do so in such a way that those artists believe they are being ripped off, then this incentive structure is compromised regardless of whether said artists actually have any claim on the AI art produced. On the other hand, the fact that AI art will still be made even if it cannot be copyrighted is self-evident. Per Brittain (2023), AI art currently cannot be copyrighted in the United States – and that has not stopped companies from churning out reams of the stuff. Therefore, if copyright is to fulfill its utilitarian purpose of art promotion, it must allow, but not protect, AI art.

Works Cited:

Brittain, B. (2023, February 23). AI-created images lose U.S. copyrights in test for new technology.

Reuters. <https://www.reuters.com/legal/ai-created-images-lose-us-copyrights-test-new-technology-2023-02-22/>

Doctorow, C. (2023a, February 28). Twiddler - Cory Doctorow - Medium. *Medium*. Retrieved June 6,

2023, from <https://doctorow.medium.com/twiddler-1b5c9690cce6>

Doctorow, C. (2023b, May 5). *Enshittification Part 1: Why Every Platform Goes Bad | On the Media |*

WNYC Studios (B. Gladstone, Interviewer). WNYC Studios. Retrieved June 5, 2023, from

<https://www.wnycstudios.org/podcasts/otm/segments/enshittification-part-1-where-did-it-all-go-wrong-on-the-media>

Doctorow, C. (2023c, May 12). *Enshittification Part 2: The Mechanisms That Helped Big Digital Go*

Bad | On the Media | WNYC Studios (B. Gladstone, Interviewer). WNYC Studios. Retrieved

June 5, 2023, from <https://www.wnycstudios.org/podcasts/otm/segments/enshittification-part-2-mechanisms-helped-big-digital-go-bad-on-the-media>

Doctorow, C. (2023d, May 19). *Enshittification Part 3: Saving The Internet | On the Media | WNYC Studios* (B. Gladstone, Interviewer). WNYC Studios. Retrieved June 6, 2023, from <https://www.wnycstudios.org/podcasts/otm/segments/enshittification-part-3-saving-internet-on-the-media>

Ferrell, O. C., Fraedrich, J., & Ferrell. (2021). *Business Ethics: Ethical Decision Making and Cases* (13th ed.). Boston: Cengage Learning, Inc.

Goodfellow, I. J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., & Bengio, Y. (2014). Generative Adversarial Networks. *arXiv (Cornell University)*.
<https://doi.org/10.48550/arxiv.1406.2661>

Karpathy, A., Abbeel, P., Brockman, G., Chen, P., Cheung, V., Duan, Y., Goodfellow, I., Kingma, D., Ho, J., Houthoof, R., Salimans, T., Schulman, J., Sutskever, I., & Zaremba, W. (2016, June 16). *Generative models*. OpenAI. Retrieved May 5, 2023, from <https://openai.com/research/generative-models>

McCorduck, P. (2004). *Machines Who Think: A Personal Inquiry Into the History and Prospects of Artificial Intelligence*. A K Peters/CRC Press.

OpenAI. (2022, November 30). *Introducing ChatGPT*. Retrieved June 4, 2023, from <https://openai.com/blog/chatgpt>

Palmer, T. (1990). Are patents and copyrights morally justified? The philosophy of property rights and ideal objects. *Harvard Journal of Law & Public Policy*, 13(3), 817–865.

<http://www.tomgpalmer.com/wp-content/uploads/papers/morallyjustified.pdf>

Rombach, R., Blattmann, A., Lorenz, D., Esser, P., & Ommer, B. (2023, March 28). *High-Resolution Image Synthesis with Latent Diffusion Models - Computer Vision & Learning Group*. Computer

Vision & Learning Group. Retrieved May 30, 2023, from <https://ommer->

[lab.com/research/latent-diffusion-models/](https://ommer-lab.com/research/latent-diffusion-models/)

United States Copyright Office, World Intellectual Property Organization, Strong, M., Rowland, C., Kasunic, R., Ashley, J., Alvarez, K., Smith, R., Levandunsky, W., Gurry, F., Iancu, A., Elgammal, A., Lynch, R., Till, U., Woods, M., Aistars, S., Guadamuz, A., Boog, J., Page, K., . . . Slotin, I. (2020, February 5). *Copyright in the age of artificial intelligence* [Video]. Copyright in the age of artificial intelligence, Washington, District of Columbia, United States of America.

<https://www.copyright.gov/events/artificial-intelligence/transcript.pdf>