

Victoria LaDell, M.A.
Juris Doctor Candidate at Tulane Law School, 2024
December 6, 2023

Reply Comments on Artificial Intelligence Notice of Inquiry

This is a reply to the Notice of Inquiry from the U.S. Copyright Office. It will address some issues raised by other commenters as well as additional elements that should be considered in the conversation concerning copyright and generative AI.¹

Abstract

Looking at current legal precedent, we should invite innovative aspects of generative artificial intelligence (AI), particularly fair use, into our aspirations for copyright and AI. There are also a number of legal complications that should be worked out before the Copyright Office makes any decisions. First of all, we should consider case law in conjunction with the fair use analysis through a modern and relevant lens. Then, we should consider current issues when it comes to potential copyright infringement and AI, such as how COPPA plays a role in this discussion. This paper will analyze issues addressed in the original Notice of Inquiry as well as discuss topics that were not explicitly considered in either the Notice of Inquiry or my original submission to the U.S. Copyright Office in October of this year.

When analyzing how machine learning in an AI model sources its training data, it is obvious that the data scraping is not just from stock libraries – it is also in archives and even social media platforms. Herein lies the issue. Three of the main social media sources that LAION learns from are Pinterest (over a million images), Tumblr, and DeviantArt.² Access to many of these platforms possesses a commonality: to create an account and post material that is most often copyrightable (e.g., digital art, short stories, photographs, etc.), one must only be thirteen years old in compliance with COPPA.

In this paper, I will be discussing artificial intelligence and copyright with respect to societal implications as well as legal matters, both current precedent and looking ahead. As this is a “Reply Comment,” I will address some overarching impressions from the general public as well

¹ Comments will be cited with the name of the commenter, the date, and the Comment ID, e.g., *Comment from LaDell, Victoria. (Oct. 30, 2023) COLC-2023-0006-9138.*

² Elizaroff, N. “The Legal Minefield of Artwork Generated by Artificial Intelligence,” *@ the Bar; The Chicago Bar Association* (Jan. 18, 2023). “Accordingly, artists uploading their work onto the internet to places like DeviantArt, ArtStation, Tumblr, Pinterest, and more may be unwittingly helping train the AI algorithms that are mimicking their work.”

as federal agencies. I will then discuss the theme of pro-innovation in reference to how to address copyright and AI, though I acknowledge that some level of risk always accompanies rapidly evolving technology. For the majority of this paper, I will be analysing the legal implications of generative AI on copyright law, as intellectual property law is my primary legal field of interest. In discussing legal aspects, I will discuss the relevant case law as well as the fair use doctrine outlined in 17 U.S.C. § 107 of the Copyright Act. I will also bring up the issue of accessibility and how generative AI can aid in the endeavor for a more accessible and accommodating society. On the topic of societal implications, I will then transition to practicality, COPPA, and key corporations' roles and expressed viewpoints. I will conclude with my final remarks and recommendations concerning artificial intelligence and copyright.

I. Introduction

In understanding the philosophical interests of intellectual property and why philosophical concerns are relevant in a discussion of copyright and AI, one could compare distinct philosophical schools that each have an impact on elements of all intellectual property: Utilitarianism, Locke's Labor Theory, and Personhood Theory. In short, the utilitarian approach concerns incentivizing innovation for the collective good, Locke's labor theory suggests that humanity has a natural right to the fruits of their work and the possession natural rights (e.g., property) triumphs over the collective good, and personhood theory expresses the importance of a personal and emotional connection and identification with one's creations.

One commenter presented an interesting idea:

“A prompt is something written by a human, I do think a prompt is has intellectual property rights associated. The image or text generated from the prompt, however, should not, in the same way that animals and natural processes do not have intellectual rights to artifacts.”³

Here, Locke's Labor Theory can be seen through this commenter's belief that certain entities or beings either do or do not possess natural rights to intellectual property, i.e., a person works on prompts, so they should have property rights on the input, whereas the same person did not create an output; therefore, that person should not have property rights.

Another commenter submitted:

“I think AI needs the largest data set we can give it to progress the fastest. I understand the importance of copyright and a part of me is conflicted, but ultimately I think at the end of the day we are holding back serious technological progress that can not only help

³ Comment from Entrekkin, Alexander. (Aug. 29, 2023) COLC-2023-0006-0239.

this country, but that can help us as a species advance and evolve. Let AI be trained by all data on the internet.”⁴

In this instance, we can see a more utilitarian approach insofar that regardless of this commenter’s personal views, this commenter asserts that remaining competitive globally is the preferable approach to copyright and AI.

The intellectual property clause of the Constitution asserts that Congress possesses the power “[t]o promote the *progress of science and useful arts*, by securing for limited times to authors and inventors the exclusive right to their respective writings and discoveries” (U. S. Const. art. I § 8 cl. 8). (emphasis added). If this provision is applied to the issue of artificial intelligence, AI would arguably fall under the promotion of the progress of science and useful arts.

Arguments for various stances appear in an analysis of AI and copyright with reference to both the intellectual property clause of the Constitution and the theories of philosophy. I believe, however, that in the systematic societal realms of healthcare, education, and accessibility, a utilitarian approach is best. By promoting the progress of science and useful arts in these circumstances, AI helps to detect cancerous tumors,⁵ assists with remote learning by providing personalized learning experiences,⁶ and can provide aid to those with disabilities in a variety of ways, e.g., text-to-image generative AI.⁷

II. General Impressions

There were repetitive themes in reading the original Notice of Inquiry comments. Many commenters were artists or anonymous individuals who view generative AI as theft, to put it lightly. One commenter offered: “Letting anything other than humans be allowed to attain copyright or create art that can be used in the creation of further art is disgusting and a corruption of human effort and the very foundation of creating. Only humans should be allowed to make, and therefore profit, off of art.”⁸ That comment illuminated several themes from individual

⁴ Comment from Miller, Gabriel. (Sept. 5, 2023) COLC-2023-0006-1630.

⁵ Zhang, B., Shi, H., & Wang, H. (2023) “Machine Learning and AI in Cancer Prognosis, Prediction, and Treatment Selection: A Critical Approach.” *Journal of Multidisciplinary Healthcare*, 16, 1779-1791. <https://doi.org/10.2147/JMDH.S410301>, “Machine learning (ML), a subset of AI that enables computers to learn from training data, has been highly effective at predicting various types of cancer, including breast, brain, lung, liver, and prostate cancer. In fact, AI and ML have demonstrated greater accuracy in predicting cancer than clinicians.”

⁶ Martin, E. “Top 7 Benefits Of Leveraging AI In Online Learning Platforms,” *eLearning Industry*. (Aug. 29, 2023) <https://elearningindustry.com/top-benefits-of-leveraging-ai-in-online-learning-platforms>

⁷ One example that I will be mentioning later is that of generative AI for those with Aphantasia Milne, S. “Can AI help boost accessibility? These researchers tested it for themselves” *UW News, University of Washington*. (Nov. 2, 2023). <https://www.washington.edu/news/2023/11/02/ai-accessibility-chatgpt-midjourney-ableist/>

⁸ Comment from Kirschman, Meredith. (Sept. 6, 2023) COLC-2023-0006-1767.

commentators that I have gathered.⁹ First, only humans should possess the ability to possess copyright. Next, humans have a right to profit off of only human work. Finally, through viewing the more vitriolic passages from individual commenters (many of whom claim to be artists), it becomes clear that people have a strong emotional attachment to art that aligns well with Personhood theory.

One interesting and noteworthy submission was by the Federal Trade Commission (FTC).¹⁰ In its submission, the FTC submitted both a comment as well as an appendix with the roundtable discussion that seemingly led to their opinions in the comment. The comment mentions “participants” in a roundtable discussion of creatives – not policy experts, scholars, or any other members of the general public – who have allegedly been affected by generative AI. The sentiments echo laments from the individual commenters I have already touched on. The FTC express their distaste for generative AI quite clearly:

“The risks associated with AI use, including violations of consumers’ privacy, automation of discrimination and bias, and the turbocharging of deceptive practices, imposter schemes, and other types of scams.” [citations omitted].¹¹

I will discuss notable corporations’ comments later in this paper.

III. Fair Use Doctrine and Case Law

A. Training

While companies and institutions seem divided on their stances on whether or not to consider AI with respect to copyrightable material, many opinions were in favor of large dataset training on unlicensed material via the fair use doctrine in pursuit of incentivizing innovation. One interesting case that I wish to highlight is that of Google. In its Comment, Google asserts that fair use protects its work as fair use is to be understood as “copying for a new and different purpose is permitted without authorization where — as with training AI systems — the secondary use is transformative and does not substitute for the copyrighted work.”¹² Google’s argument relies heavily on the potential success of the fair use doctrine in the debates over copyright and AI.

The Fair Use Doctrine is established in §107 of the Copyright Act of 1976.¹³ The four factors considered in determining whether there is fair use of a copyrighted work are:

⁹ Commenters displayed a wide array of views and opinions; here I present a broad interpretation.

¹⁰ Comment from the FTC [Federal Trade Commission]. (Oct. 30, 2023) COLC-2023-0006-8630.

¹¹ Comment from the FTC [Federal Trade Commission]. (Oct. 30, 2023) COLC-2023-0006-8630.

¹² Comment from Google. (Oct. 30, 2023) COLC-2023-0006-9003.

¹³ 17 U.S.C. 107(1).

- (1) the purpose and character of the use, including whether such use is of a commercial nature or is for nonprofit educational purposes;
- (2) the nature of the copyrighted work;
- (3) the amount and substantiality of the portion used in relation to the copyrighted work as a whole; and
- (4) the effect of the use upon the potential market for or value of the copyrighted work.¹⁴

Many proponents of using large datasets that include copyrighted works for generative AI model training claim that the fair use doctrine protects the training process. This argument is particularly persuasive from a research and education perspective. In their Comment, the University Library of the University of California, Berkeley asserts that LLM (Large Language Model) artificial intelligence training on copyrighted inputs legally “falls squarely within what courts have determined to be a transformative fair use,” particularly for “nonprofit educational or research purposes.”¹⁵ Additionally, the Library accessories that an opt-out option would undermine the driving philosophy behind fair use, yielding less content to then train on “for the advancement of science and the useful arts.”¹⁶

Merging fair use analysis with the intellectual property clause of the Constitution (U. S. Const. art. I § 8 cl. 8) would then logically weigh in favor of permitting copyrighted works in AI model training, particularly in the field of text-to-image generative AI or natural language processing (NLP) generative AI, as a matter of practice. As a legal researcher and in the final term of my Juris Doctor degree with a particular interest in intellectual property with respect to the impact of artificial intelligence, I personally find fair use justification for training models proper and should be implemented by federal laws or regulations.

The Allen Institute (AI2) offers two persuasive propositions in an effort to showcase how fair use ought to be treated in terms of training and monitoring outputs.¹⁷ First, either Congress or the Copyright Office should permit training AI Models on copyrighted material under fair use analysis, and that the Datasets only employ “a fraction of the original data collected,” and that data thus becomes transformative from the copyrighted works used in training.¹⁸ Second, there should be additions to the four-factor analysis test under 17 U.S.C. § 107 to “assess potential harm” that the outputs can cause to copyright holders.¹⁹

¹⁴ 17 U.S.C. 107(1)-(4).

¹⁵ Comment from the University Library of the University of California, Berkeley. (Oct. 30, 2023) COLC-2023-0006-8194.

¹⁶ *Id.*

¹⁷ Comment from The Allen Institute for AI. (Oct. 30, 2023) COLC-2023-0006-8762

¹⁸ *Id.*

¹⁹ *Id.*

B. Transformativeness

The key to a successful fair use analysis often depends on whether a governing body deems the work transformative. The concept of transformativeness is best expressed and understood through case law. *Campbell v. Acuff-Rose Music, Inc.*, (1994) sets the standard for fair use analysis under § 107, and there are additional key cases relevant to transformativeness concerning copyright and fair use in the digital age. 510 U.S. 569. Such cases include *Authors Guild v. Google, Inc.*, No. 13-4829 (2d Cir. 2015), *The Author's Guild v. Hathitrust*, 755 F.3d 87 (2d Cir. 2014), *Sony Corp. of America v. Universal City Studios, Inc.*, 464 U.S. 417 (1984). For the purpose of this paper, I will primarily focus on *Campbell v. Acuff-Rose Music, Inc.* (1994), *Authors Guild v. Google, Inc.* (2015), and *The Author's Guild v. Hathitrust* (2014), with a later discussion on *Feist Publications, Inc. v. Rural Tel. Serv. Co.*, 499 U.S. 340 (1991).²⁰

Campbell v. Acuff-Rose Music, Inc., 510 U.S. 569 (1994) sets a standard for how changing a copyrighted work's purpose and character allows creativity to utilize the benefits of copyright law. A key takeaway from *Campbell v. Acuff-Rose Music, Inc.*, (1994) is its promotion of creating transformative works within fair use parameters. In the Supreme Court's opinion, the Court reemphasizes the importance of the intellectual property clause of the Constitution as well as showcasing a need for "breathing space within the confines of copyright" (in reference to Blackmun's dissent in *Sony Corp. of America v. Universal City Studios, Inc.*)²¹.

In *Authors Guild v. Google, Inc.*, (2015), Google was sued for copyright infringement by authors for copying millions of books to upload, where snippets were visible for its digitization project Google Books.²² They did not have permission to do so, yet the use was so transformative that the court found that Google had not infringed. We are seeing a very similar situation in generative artificial intelligence. In machine learning processes, millions of works are being used for training without the original author's permission to train generative AI models where the outputs are arguably highly transformative. Here, it was not that copyrightable works were being replicated; rather, the works helped create a digital system with non-derivative outputs. The

²⁰ See, e.g., *Campbell v. Acuff-Rose Music, Inc.*, 510 U.S. 569 (1994) emphasized the power of the transformative aspect of the fair use analysis even in a commercial work; Both *Authors Guild v. Google, Inc.*, No. 13-4829 (2d Cir. 2015) and *The Author's Guild v. Hathitrust*, 755 F.3d 87 (2d Cir. 2014) display how courts determine whether digital libraries and archives are transformative in terms of purpose of the works being for research, accessibility, and the common good, generally.

²¹ *Sony Corp. of America v. Universal City Studios, Inc.*, 464 U.S. 417 (1984) at 478-480: "When the scholar forgoes the use of a prior work, not only does his own work suffer, but the public is deprived of his contribution to knowledge. The scholar's work, in other words, produces external benefits from which everyone profits. In such a case, the fair use doctrine acts as a form of subsidy -- albeit at the first author's expense -- to permit the second author to make limited use of the first author's work for the public good."

²² No. 13-4829 (2d Cir.)

emphasis on “augment[ing] public knowledge” would also seem appropriate in the conversation on transformativeness in the world of generative AI. Finally, in its analysis, the court asserted that Google was neither infringing upon copyright nor was Google a “contributory infringer” *supra*, at 46. The court also notes that transformative uses are unique as a transformative use “expands its utility, thus serving copyright’s overall objective of contributing to public knowledge.”²³

The court deemed the use to be highly transformative as knowledge about a copyrighted work did not equate to “providing the public with a substantial substitute” for copyrighted works or creating derivatives of those works.²⁴

In *The Author’s Guild v. HathiTrust*, 755 F.3d 87 (2d Cir. 2014), the facts were similar to those in its successor, *Authors Guild v. Google, Inc.*, insofar that a party successfully invoked fair use in reference to creating digital libraries and databases using copyrighted materials. *HathiTrust* is concerned with a digital library consisting of millions of books. This case is unique in that it discusses accessibility in terms of acquiring full texts of copyrighted works without permission from the holder. It provided access to full text for those for which it was applicable, i.e., those with a “print disability.”²⁵ One key difference between the two cases is that HathiTrust is a non-profit organization that created a digital library instead of Google’s commercial nature.

Google LLC v. Oracle America, Inc. is another example of how courts have ruled in favor of digital transformativeness under fair use.

IV. Accessibility

It would be prudent to discuss similar ways in which accessibility should be considered in the conversation of generative AI. Transformativeness can be when copyrighted material is made available to individuals with disabilities as the *purpose* has changed, not the work *per se*.²⁶ It is asserted in *Authors Guild, Inc. v. HathiTrust* that accessibility could contribute to the “purpose” aspect of the fair use analysis.²⁷ In *Authors Guild, Inc. v. HathiTrust*, those with certain disabilities were allowed to have full access to copyrighted materials.²⁸

²³ *Id.*

²⁴ *Id.*

²⁵ *The Author’s Guild v. HathiTrust*, 755 F.3d 87 (2d Cir. 2014).

²⁶ See, e.g., *Authors Guild v. Google, Inc.*, (2015) where the court states: “the amount and substantiality of *what is thereby made accessible* to a public for which it may serve as a competing substitute.” (emphasis in original) at 31.

²⁷ *The Author’s Guild v. HathiTrust*, 755 F.3d 87 (2d Cir. 2014)

²⁸ *Id.*

One example of technology where AI aids accessibility needs is the Be My Eyes Virtual Volunteer™ app. Through OpenAI’s GPT-4 model, a new application has already been developed to aid in accessibility as a Virtual Volunteer.²⁹ Be My Eyes allows users to insert images via the app, and the generative AI model will then “answer any question about that image and provide instantaneous visual assistance for a wide variety of tasks.”³⁰

The Center for Teaching Innovation at Cornell University has expressed the understanding of a need for generative AI in terms of teaching and accessibility.³¹ They believe that in the time of online and remote learning, educators should be inspired to “[u]se Generative AI to expand, not restrict” as generative AI can assist in transforming text into formats better suited for individuals with disabilities along with various other manners in which to utilize generative AI for accessible learning.³²

Looking forward, AI can also assist in ensuring companies are keeping up with Web Content Accessibility Guidelines (WCAG) and in compliance with the Americans With Disabilities Act³³ (ADA). Furthermore, it can make WCAG and ADA-compliant websites even more accessible through the use of chatbots and other AI assistance. In the case of text-to-image generative AI such as Midjourney, researchers have shown that while imperfect, text-to-image technology can help specific disabilities such as Aphantasia, where an individual cannot visually conceptualize text³⁴. Text-to-image generative AI can be applied to imagining book characters and settings to create graphs and tables for research.

UNESCO has also launched a “Guidance for generative AI in education and research” for long-term planning of how to expand the possibilities of generative AI further. Though the guidance does not specifically mention accessibility or disability, its emphasis on furthering AI in education and research gives hope that such guidelines will eventually be considered and even adopted.

V. Uncopyrightable Works and Elements

²⁹ Henneborn, L. “Designing Generative AI to Work for People with Disabilities” *Harvard Business Review*. (Aug. 18, 2023) <https://hbr.org/2023/08/designing-generative-ai-to-work-for-people-with-disabilities>

³⁰ <https://www.bemyeyes.com/blog/introducing-be-my-eyes-virtual-volunteer>

³¹ <https://teaching.cornell.edu/generative-artificial-intelligence/ai-accessibility>

³² *Id.*

³³ Americans With Disabilities Act of 1990, 42 U.S.C. § 12101 et seq. (1990)

³⁴ Milne, S. “Can AI help boost accessibility? These researchers tested it for themselves” *UW News, University of Washington*. (Nov. 2, 2023). <https://www.washington.edu/news/2023/11/02/ai-accessibility-chatgpt-midjourney-ableist/>

One request I have for the Copyright Office is to clarify what elements of a work constitute “data” in machine learning.

Let us illustrate the idea of using facts to generate a story with the assistance of generative AI. *Feist* gives us an interesting perspective on how uncopyrightable material, such as facts, can still become copyrightable due to their originality in selection, coordination, and arrangement:

“A compilation is not copyrightable *per se*, but is copyrightable only if its facts have been ‘selected, coordinated, or arranged *in such a way* that the resulting work as a whole constitutes an original work of authorship.’ § 101 (emphasis added).”³⁵

Two ultimate takeaways from *Feist* are: “The first is that facts are not copyrightable; the other, that compilations of facts generally are.”³⁶ With rule regarding facts and arrangement set forth in *Feist*, a new query emerges – how far can one take this standard, and how can it be applicable to copyright and AI? First, let us briefly discuss a case study – an example of the arrangement of facts.

Though not under the Copyright Act, the World Intellectual Property Organization (WIPO) reported on the results of a U.K. ruling regarding whether a copyrighted work infringed on other copyrighted material or whether the work was a compilation of facts. This is the case of Dan Brown and *The Da Vinci Code*.³⁷ Here, the court held that though Dan Brown likely researched and was influenced by certain copyrighted texts, *the Da Vinci Code* was ultimately a carefully selected arrangement of facts concerning the legend of the holy grail in certain Christian-centered texts and beliefs.³⁸

Putting these aspects together, what if an individual uses a generative AI model such as ChatGPT to collect an assortment of facts on a certain subject? What if I asked for inspiration in the form of loglines? To find out, I put ChatGPT-4 up to the task.

³⁵ *Feist Publications, Inc. v. Rural Tel. Serv. Co.*, 499 U.S. 340 (1991)

³⁶ *Id.*

³⁷ Suthersanen, U. “Copyright in the Courts: The Da Vinci Code,” *WIPO Magazine*, Issue 3/2006. (Jun. 2006) https://www.wipo.int/wipo_magazine/en/2006/03/article_0004.html#:~:text=The%20Judgement,infringed%20copyright%20in%20the%20book.

³⁸ *Id.*



You

What would be a good story based off a collection of true facts? Can you give me five potential fact-based loglines?



ChatGPT

Creating a story based on a collection of true facts can lead to engaging and thought-provoking narratives. Here are five potential fact-based loglines for stories:

1. **The Unsung Codebreakers:** During World War II, a group of brilliant female mathematicians are recruited to break German enigma codes, secretly turning the tide of the war while battling gender biases of their time.
2. **The Lost Expedition:** In the 1840s, the doomed Franklin Expedition sets out to find the Northwest Passage, leading to a harrowing tale of survival, mystery, and a newfound respect for indigenous Arctic knowledge.
3. **The Race to the Pole:** Two teams, one led by Amundsen and the other by Scott, embark on a perilous and historic race to be the first to reach the South Pole, revealing the extremes of human ambition and endurance.
4. **The Silent Earthquake:** Based on true events, scientists discover an impending catastrophic earthquake in a major city, battling against time and skepticism to save millions of lives from an unprecedented natural disaster.
5. **The Heist of the Century:** A retelling of the 1978 Lufthansa Heist, where a group of small-time criminals pull off nearly \$5 million from the JFK airport, unraveling a story of greed, betrayal, and the fall of a crime empire.

Each of these loglines takes factual events and weaves them into a narrative, highlighting human struggles, triumphs, and the complexity of real-life stories.

For the sake of the argument, let us say that I chose to write a novel based on one of these premises. In this hypothetical, I could write a novel or short story, "The Heist of the Century", about the Lufthansa Heist mentioned in the fifth proposed fact-based pitch from the GPT-4 model. Then, I do all of my research through this model, relying solely on what facts the generative AI model would think appealing to a potential reader in today's market. All of the outlining and arrangement would come from my own original thoughts, though all of the source material, for lack of a better phrase, is from a generative AI model. Under the cases discussed

thus far, including *Feist*, I should, upon completion of the work, possess copyright. I could upload relatively large files into the GPT-4 model of OpenAI's chabot and ask for summaries to shorten research time without taking any material directly from an output verbatim. So this is a brief example of a nuanced situation in which, though inspired and heavily assisted by outputs generative AI, I would hypothetically have the copyright due to how I arrange the facts into a transformative work of historical fiction.

A more likely example that has been discussed is the risks that generative AI models, particularly Chat GPT, pose in education and research. As I mentioned those concerns in my original comment, I will not duplicate my explanation of the ways in which generative AI could be as much of a risk as it is a benefit to society.

On the topic of uncopyrightable elements, I wish to highlight that there is no explicit rule in the Copyright Act or in federal legislation concerning copyrighting "style," or rather, no obvious grounds exist for an infringement claim. The Copyright Act extends protection to "original works of authorship fixed in any tangible medium of expression." 17 U.S.C. § 102(a). Under the current understanding of 17 U.S.C. § 102, "style" is not copyrightable. This is desirable, as described by Stephen Wolfson for the Creative Commons, "if one artist were given a monopoly over anime, grunge music, or other styles, that would frustrate copyright's core purpose of supporting creativity."³⁹ Artists are often upset due to this fact, which would align well under Personhood Theory as discussed earlier.⁴⁰ Additionally, the Creative Commons also submitted a comment to the Notice of Inquiry.⁴¹ In its submission, Creative Commons displays support for the ability to copyright human works of authorship, even when assisted by artificial intelligence.⁴² Creative Commons also supports the idea that training AI on datasets possessing copyright-protected material should be considered fair use.

VI. COPPA

There has been little to no discussion on how generative AI and copyright are impacted by the Children's Online Privacy Protection Act (COPPA).⁴³ As minors use and upload content to online platforms, minors are among the individuals who had their data scraped from places such as

³⁹ Wolfson, S. "The Complex World of Style, Copyright, and Generative AI," *Creative Commons*. (Mar. 23, 2023). <https://creativecommons.org/2023/03/23/the-complex-world-of-style-copyright-and-generative-ai/#:~:text=However%2C%20style%20is%20not%20generally,core%20purpose%20of%20supporting%20creativity>.

⁴⁰ *Id.*

⁴¹ Comment from Creative Commons, (Oct. 29, 2023). COLC-2023-0006-8735.

⁴² *Id.*

⁴³ 15 U.S. Code Chapter 91.

Pinterest, Tumblr, and DeviantArt.⁴⁴ This presents an issue. If all who have been wronged by data scraping wish to file infringement claims, it becomes increasingly difficult for minors to possess that right. Statistics for precisely how many minors aged thirteen to seventeen use internet platforms that have been used for machine learning and training are difficult to pin down due to COPPA and related age-protecting mechanisms, but that does not detract from the fact that they would be at a disadvantage as they would not be able to file without a parent or guardian in a federal court for infringement.

This seems to create a largely overlooked issue: what rights would minors have under potential (stringent) copyright regulations over the use of AI? Also, copyright regulations could interfere with COPPA, especially in an age where minors under the age of thirteen use generative AI for creative purposes or even just amusement and curiosity.

The Federal Trade Commission (FTC) oversees COPPA. The primary purpose of COPPA, according to the FTC, is “to place parents in control over what information is collected from their young children online.”⁴⁵ Furthermore, the FTC has stated that:

“As a matter of federal policy, all websites and online services operated by the Federal Government and contractors operating on behalf of federal agencies must comply with the standards set forth in COPPA.”⁴⁶ According to the Copyright Office: “Minors may claim copyright, and the U.S. Copyright Office will accept applications submitted either by or on behalf of a minor if the application is otherwise proper and complete.”⁴⁷ Once a work is registered, then personal information “including a minor’s name, address, or year of birth” cannot be removed from the registration record in the majority of circumstances. There is no 13-year age cut-off for Copyright Registration.⁴⁸ The Copyright Office’s current regulations could, in particular circumstances, create a COPPA issue. This issue is one to be debated by courts, if ever applicable, but I again want to highlight another nuance issue within current practices and policies governing copyrightability.

VII. Practicality

Not all who publish creative works online register each individual piece for copyright. Nor should they necessarily need to, as copyright exists when it is fixed in a tangible medium of expression. 17 U.S.C. § 102(a). Now, in order to sue for infringement in a federal court, the work(s) must be registered. Creators who possess copyright without registration would be at a

⁴⁴ Here, I am referring to LAION method of data scraping found by domain here:

https://laion-aesthetic-dataset.io/laion-aesthetic-6pls/domain?_sort_desc=image_counts.

⁴⁵ Federal Trade Commission, “Complying with COPPA: Frequently Asked Questions”. [ftc.gov](https://www.ftc.gov)

⁴⁶ *Id.*

⁴⁷ U.S. Copyright Office, Compendium of U.S. Copyright Office Practices § 101 (3d ed. 2021).

⁴⁸ *Id.*

disadvantage. One may argue that this is a class disadvantage unless the Copyright Office allows for free registration to those specifically seeking damages or relief for claims of AI-assisted infringement. I discuss COPPA so heavily as I see it as a way in which it will become very difficult for minors to sue for infringement.

Right of Publicity introduces another issue, which has to do with copyrighted images and those in the public domain. What are we to do with images in the public domain that contain photos of living or deceased persons that are used in “deepfakes.” As I have discussed in my original comment, there is a definitive need for federal regulations regarding the right of publicity.

VIII. Corporations

Before my concluding remarks, I will note some significant comments from prominent corporations in the discussion of copyright and artificial intelligence.

Meta is pro-fair use insofar as copyrighted materials ought to be allowed for machine learning and AI model training purposes. In its comment, Meta references *Authors Guild v. Google, Inc.* (2015) to defend its stance. Meta claims that “[e]xtracting information from training data is the sole purpose and function of Generative AI model training.” and then cites the “Google Books” case, referencing *Authors Guild*.⁴⁹

Getty Images advocates for affirmative consent from the rightful owner of copyrighted material prior to data scraping and AI model training.⁵⁰ Getty’s position is unsurprising given its legal battle with Stability AI.⁵¹

OpenAI’s stance is also pro-innovation and pro-fair use.⁵² What makes OpenAI unique in this capacity is that the corporation advocates for a de-duplication “(i.e., to delete content that may appear more than once)” process in training sets. OpenAI even refers to memorization as a “bug” to be fixed:

“Because our models do not have access to training information after they have learned from it, they are unlikely to duplicate training data in their outputs. In fact, verbatim repetition or ‘memorization’ of training data is generally considered by AI developers to be a bug to be corrected, rather than a feature to be pursued.”

⁴⁹ Comment from Meta. (Oct. 30, 2023) COLC-2023-0006-9027

⁵⁰ Comment from Getty Images (US), Inc., (Oct. 30, 2023) COLC-2023-0006-9044

⁵¹ See e.g. *Getty Images, Inc., v. Stability AI* [2023] EWHC 3090 (Ch), See <https://s3.documentcloud.org/documents/24183636/getty-images-v-stability-ai-uk-ruling.pdf>

⁵² Comment from Open AI, Re: Notice of Inquiry and Request for Comment [Docket No. 2023-06] (Oct. 30, 2023), COLC-2023-0006-8906

I agree with the argument that making AI regulations at this time is premature. OpenAI has a very similar outlook to many of the comments discussed thus far regarding fair use. It believes AI model training “qualifies as a fair use, falling squarely in line with established precedents recognizing that the use of copyrighted materials by technology innovators in transformative ways is entirely consistent with copyright law.” One glaring concern in its comment is the assertion that “factual metadata and fundamental information that AI models learn from training data are not protected by copyright law.” I think this statement ought to be up for debate, and I would like clarity from the Copyright Office on such an assertion.

IX. Conclusion

Artificial intelligence has the ability to augment human creativity rather than replace it, much less surpass it. Under this philosophy, policymakers should approach the topic of copyright and AI from a pro-innovation stance. Furthermore, I have attempted to raise some of the multitude of issues when facing the task of creating stringent AI rules and regulations. I therefore do not believe the Copyright Office should solely or even primarily hold this burden, as generative AI issues with copyright operate on a case-to-case basis. If not the Copyright Office, then who? Rather than asserting that this burden lies exclusively with the executive branch via bureaucratic federal agencies, Congress and the Courts should be included in such a paramount discussion before making decisions. The judicial branch should have significant responsibility considering how much of the conversation around copyright and AI circles back to case law and subsequent doctrines.

Thank you for your time and consideration on this pressing issue.

Sincerely,
Victoria LaDell, M.A.
Juris Doctor Candidate at Tulane Law School, 2024