

Victoria LaDell, M.A.

RESPONSE TO NOTICE OF INQUIRY

[original footnotes included]

NOTE: This document has been written before the Executive Order issued by President Biden today, October 30, 2023, merely hours ago, and some edits have been included to reflect the executive order (which is yet to be uploaded to the Federal Register, hence the lack of proper citation); however, this document is not fully up to date as this Executive Order was issued less than twenty-four hours preceding the deadline for submission. It is worth noting that the Executive Order appears to be largely in line with my beliefs on ethical and legal concerns of generative artificial intelligence.

General Questions

The Office has several general questions about generative AI in addition to the specific topics listed below. Commenters are encouraged to raise any positions or views that are not elicited by the more detailed questions further below.

1. As described above, generative AI systems have the ability to produce material that would be copyrightable if it were created by a human author. What are your views on the potential benefits and risks of this technology? How is the use of this technology currently affecting or likely to affect creators, copyright owners, technology developers, researchers, and the public?

As for the benefits of this technology from a creative standpoint, my views are largely concerned with “deepfakes.” There are many benefits to artificial intelligence, but the use of generative-AI to create deepfakes which can undermine the democratic values that our laws are based on.

A paramount benefit of generative AI is the furtherance of innovation and expression.

The use of AI technology, particularly generative AI, will transform various aspects of our understanding of the world; thus, it will affect all of the aforementioned sectors of society. There is a concern over what is to be considered “real art” when it comes to generative AI and a fear that ‘steal jobs’ across various sectors that are affected by copyright.

2. Does the increasing use or distribution of AI-generated material raise any unique issues for your sector or industry as compared to other copyright stakeholders?

I will address this question in the capacity of a legal researcher, educator, and ethicist.

Researcher: The use of artificial intelligence within academic and scholarly research can be highly beneficial when used correctly.

Educator: Students benefit from understanding how to use generative AI in an ethical manner, as AI is and will continue to be a ubiquitous part of our society.

Ethicist: As mentioned, deepfakes and use for malicious purposes can undermine democratic values and intentions.

3. Please identify any papers or studies that you believe are relevant to this Notice. These may address, for example, the economic effects of generative AI on the creative industries or how different licensing regimes do or could operate to remunerate copyright owners and/or creators for the use of their works in training AI models. The Office requests that commenters provide a hyperlink to the identified papers.

Karnouskos, Stamatis. "Artificial intelligence in digital media: The era of deepfakes." IEEE Transactions on Technology and Society 1.3 (2020): 138-147.

(hyperlink:https://www.researchgate.net/profile/Stamatis-Karnouskos/publication/342795647_Artificial_Intelligence_in_Digital_Media_The_Era_of_Deepfakes/links/5f2dc189458515b7290d312f/Artificial-Intelligence-in-Digital-Media-The-Era-of-Deepfakes.pdf)

4. Are there any statutory or regulatory approaches that have been adopted or are under consideration in other countries that relate to copyright and AI that should be considered or avoided in the United States?¹ How important a factor is international consistency in this area across borders?

[**NOTE:** This has not been properly edited since President Biden's Executive Order concerning AI was issued on October 30, 2023.]

European Union: Digital Single Market Directive

¹ For example, several jurisdictions have adopted copyright exceptions for text and data mining that could permit use of copyrighted material to train AI systems. Separately, the European Parliament passed its version of the Artificial Intelligence Act on June 14, 2023, which includes a requirement that providers of generative AI systems publish "a sufficiently detailed summary of the use of training data protected under copyright law." See Artificial Intelligence Act, amend. 399, art. 28b(4)(c), EUR. PARL. DOC. P9_TA (2023)0236 (2023), https://www.europarl.europa.eu/doceo/document/TA-9-2023-0236_EN.html.

Articles 3 and 4 of the Directive imply there is an opt-out ability for “rightholders” of a work with certain exceptions. Article 3(3) holds that:

“Rightholders shall be allowed to apply measures to ensure the security and integrity of the networks and databases where the works or other subject matter are hosted. Such measures shall not go beyond what is necessary to achieve that objective.”

What is “necessary” is not expressly addressed in this section and appears up to interpretation.

Article 4(3) states: “[W]orks and other subject matter referred to in that paragraph has not been expressly reserved by their rightholders in an appropriate manner, such as machine-readable means in the case of content made publicly available online,” indicating what could be considered as an opt-out ability as an express right.

However, there seem to be exceptions to this opting out. Though vague, there seem to exist exceptions in data mining for “research, innovation, education and preservation of cultural heritage, digital technologies.”

Source: Directive (EU) 2019/790 of the European Parliament and of the Council of 17 April 2019 on copyright and related rights in the Digital Single Market and amending Directives 96/9/EC and 2001/29/EC (hyperlink:<https://eur-lex.europa.eu/eli/dir/2019/790/oj>)

United Kingdom: AI Regulation Policy Paper

The United Kingdom has an intriguing approach to AI that should be considered. The United Kingdom released a white paper in 2022 that establishes its “Pro-innovation and risk-based” approach to AI in the “AI Regulation Policy Paper.”

(hyperlink:<https://www.gov.uk/government/publications/establishing-a-pro-innovation-approach-to-regulating-ai/establishing-a-pro-innovation-approach-to-regulating-ai-policy-statement>)

5. Is new legislation warranted to address copyright or related issues with generative AI? If so, what should it entail? Specific proposals and legislative text are not necessary, but the Office welcomes any proposals or text for review.

I think there should be legislation addressing the copyright practicality as well as ethics of generative AI within the scope of copyright. One idea is a new manner of communicating whether a registered work possesses aspects of AI. For example, New legislation could require a notice that appears alongside the copyrighted work in any publication or republication of the work. rather than denying copyright registration.

- Perhaps adding a symbol that derives meaning from mathematical principles e.g., \cap (intersection), \in (element of), etc.,

- Rather than denying copyright registration of all AI-assisted works or working meticulously on a case-by-case basis.

The Executive Order issued on October 30, 2023, by President Biden regarding AI appear to be, at face value, a step in the right direction when considering federal legislation, statutes, and law.

Training

If your comment applies only to a specific subset of AI technologies, please make that clear.

As I do not have a background in computer science, I will refrain from answering questions regarding aspects that I lack proper ethos. I will include potentially relevant comments to particular questions. I will indicate where I have “no knowledge” pertaining to particular questions when applicable.

6. What kinds of copyright-protected training materials are used to train AI models, and how are those materials collected and curated?

Materials that are posted or reposted online, often without authorization (i.e., piracy), may be used in AI training models. Some companies are more forthcoming, such as Adobe, Getty Images and Google, versus corporations such as Meta.

6.1. How or where do developers of AI models acquire the materials or datasets that their models are trained on? To what extent is training material first collected by third-party entities (such as academic researchers or private companies)?

Google’s Bard is a good example of a general manner in which models can be trained.

In its overview, Google states that Bard is “pre-trained on a variety of data from publicly available sources.” One could interpret this as using forums as well as all information available through Google or other locations online.

Source: “An overview of Bard: an early experiment with generative AI”
(hyperlink:<https://ai.google/static/documents/google-about-bard.pdf>)

Opposed to a GPT-4 model, Google employs LaMDA, “Language Model for Dialogue Applications,” for its search engine, natural language processing, and other AI-related aspects.

6.2. To what extent are copyrighted works licensed from copyright owners for use as training materials? To your knowledge, what licensing models are currently being offered and used?

No knowledge.

6.3. To what extent is non-copyrighted material (such as public domain works) used for AI training? Alternatively, to what extent is training material created or commissioned by developers of AI models?

Comment on public domain works in generative AI training: If AI models rely solely on works presently in the public domain, that would seem to underscore the drive for creativity. All creative works are derivative in one manner or another and are a reflection of culture and society. If only public domain works are used for generative AI training, then the outputs would not be pursuant to creative endeavors or the ability to express ideas

6.4. Are some or all training materials retained by developers of AI models after training is complete, and for what purpose(s)? Please describe any relevant storage and retention practices.

No knowledge.

7. To the extent that it informs your views, please briefly describe your personal knowledge of the process by which AI models are trained. The Office is particularly interested in:

Personally, I think a large quantity is imperative and should continue, with proper safeguards, to operate as it presently does as long as it follows legal and moral principles.

7.1. How are training materials used and/or reproduced when training an AI model? Please include your understanding of the nature and duration of any reproduction of works that occur during the training process, as well as your views on the extent to which these activities implicate the exclusive rights of copyright owners.

My understanding of generative AI training is the process involving dataset inputs, deep neural networks, and the resulting outputs. To my knowledge, a general GPT 3.5 model uses 96 attention layers to produce an output. As I am aware, generative AI models are trained more on pattern analysis rather than training with unauthorized materials outright. I personally believe that, in regard to datasets, large quantities are important to render higher quality and more accurate outputs.

7.2. How are inferences gained from the training process stored or represented within an AI model?

No knowledge.

7.3. Is it possible for an AI model to “unlearn” inferences it gained from training on a particular piece of training material? If so, is it economically feasible? In addition to retraining a model, are there other ways to “unlearn” inferences from training?

From my understanding, it would be a generally fruitless endeavor to attempt to have a model “unlearn” inferences in both a practical manner as well as an economic matter. Then, we must consider what is more important: innovation or individual intellectual property rights. It is a difficult question.

7.4. Absent access to the underlying dataset, is it possible to identify whether an AI model was trained on a particular piece of training material?

No knowledge.

8. Under what circumstances would the unauthorized use of copyrighted works to train AI models constitute fair use? Please discuss any case law you believe relevant to this question.

There should be an education exception in terms of fair use and the unauthorized world to train AI models. In teaching individuals on how to use or train artificial intelligence, there should be a fair use exception for education, as mentioned in the footnote to Question 8.3.

In addition to the two cases mentioned in question 8.1, *Authors Guild v. Google, Inc.*, 804 F.3d 202 (2015) is highly relevant to the issue of fair use and AI training models. The explanation for the reasoning, in this case, is clearly identified in the Fair Use Index’s description (<https://www.copyright.gov/fair-use/summaries/authorsguild-google-2dcir2015.pdf>).

8.1. In light of the Supreme Court’s recent decisions in *Google v. Oracle America*² and *Andy Warhol Foundation v. Goldsmith*,³ how should the “purpose and character” of the use of copyrighted works to train an AI model be evaluated? What is the relevant use to be analyzed? Do different stages of training, such as pre-training and fine-tuning,⁴ raise different considerations under the first fair use factor?

² 141 S. Ct. 1183 (2021).

³ 143 S. Ct. 1258 (2023).

⁴ See Pre-training, Fine-tuning, and Foundation Models, GenLaw: Glossary (June 1, 2023), <https://genlaw.github.io/glossary.html> (explaining that pre-training is a relatively slow and expensive process that “results in a general-purpose or foundation model” whereas fine-tuning “adapts a pretrained model checkpoint to perform a desired task using additional data”).

In discussing these cases, the first factor of the fair use analysis concerning the “purpose and character” factor, which can also be considered as the transformative factor of derivative works, is called into question, not merely the fourth fair use factor, which is more interested in commercial impact. Both should be considered in evaluating the “purpose and character” of copyrighted works that train AI models.

I do not possess adequate technological knowledge to fully comment on the first use factor in reference to pre-training and fine-tuning. With that said, I believe that it is not an inherent infringement on copyright in training as “purpose and character” if AI-generated work is a reflection of broad cultural trends e.g., minimalism, digital contemporary realism, etc.

I will disclose that I strongly disagree with the Supreme Court’s ruling against the Andy Warhol Foundation in favor of Goldsmith. The Court did not properly apply the fair use doctrine established in 17 U.S.C. § 107 of the Copyright Act of 1976, nor did they properly analyze Warhol’s work using the four fair use factors.

8.2. How should the analysis apply to entities that collect and distribute copyrighted material for training but may not themselves engage in the training?

All parties and entities that contribute to the collection of data and AI model training should be held to the same standards. Issuing accountability to all involved parties, in my view, is the best option legally and ethically.

8.3. The use of copyrighted materials in a training dataset or to train generative AI models may be done for noncommercial or research purposes.⁵ How should the fair use analysis apply if AI models or datasets are later adapted for use of a commercial nature?⁶ Does it make a difference if funding for these noncommercial or research uses is provided by for-profit developers of AI systems?

To properly answer this question, I will include the four fair use factors set forth in 17 U.S.C. § 107:

- (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;

⁵ For example, the generative AI model, Stable Diffusion, was reportedly developed in part by researchers at the Ludwig Maximilian University of Munich but is used by the for-profit company Stability AI. See Kenrick Cai, Startup Behind AI Image Generator Stable Diffusion Is In Talks To Raise At A Valuation Up To \$1 Billion, *Forbes* (Sept. 7, 2022), <https://www.forbes.com/sites/kenrickcai/2022/09/07/stability-ai-funding-round-1-billion-valuation-stable-diffusion-text-to-image/?sh=31e11f5a24d6>.

⁶ 17 U.S.C. 107(1).

(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.

In reference to the question, the first factor is satisfied if the initial purpose and character of the use are educational and non-commercial. If the purpose and character of the use is for research or non-commercial purposes. The second and third factors do not require an analysis from an educational standpoint here. Finally, the effect on the potential market is hypothetical unless agreed otherwise. Thus, there is no present commercial risk if the

It should not make a difference if the AI training models are funded by for-profit developers. According to the U.S. Copyright Office Fair Use Index's homepage (<https://www.copyright.gov/fair-use/>), the purpose for fair use that it "promotes freedom of expression" by using copyrighted works without express authorization.

Ethically, this issue could be up for debate; however, education and research should be paramount when compared to the fear or potential commercial use for which there can be ramifications later; however, at present, there should not be a problem with for-profit developers funding noncommercial research uses.

Also, a copyrighted work has a shelf life, so to say, in reference to the duration of the copyright. So, hypothetically, if the developers eventually use works that were initially part of non-commercial training, then some works may be in the public domain by that point, which would render the argument moot.

8.4. What quantity of training materials do developers of generative AI models use for training? Does the volume of material used to train an AI model affect the fair use analysis? If so, how?

Though unsure of the quantity of materials used in training, the volume of the materials should satisfy fair use analysis by allowing and encouraging more derivative works to be created. A good example of this fair use analysis is found in the Supreme Court's ruling in *Google v. Oracle America* (citation in footnote no. 2 provided by question 8.1) concerning API and transformativeness.

8.5. Under the fourth factor of the fair use analysis, how should the effect on the potential market for or value of a copyrighted work used to train an AI model be measured?⁷ Should the inquiry be whether the outputs of the AI system incorporating the model compete with a particular copyrighted work, the body of works of the same author, or the market for that general class of works?

⁷ Id. at 107(4).

Particular works used without a license whatsoever are fully unauthorized, taking the *Warhol* case a step further (where a license was provided to a party by Goldsmith originally). If an author can prove copyright, then legal remedies should be made available.

In most circumstances, however, the standards set forth in *Google v. Oracle America* and *Authors Guild v. Google, Inc.* should be adhered to.

9. Should copyright owners have to affirmatively consent (opt in) to the use of their works for training materials, or should they be provided with the means to object (opt out)?

In short, no. This may be an option, but affirmative consent devalues and undermines a democratic society with thriving creative industries because when can that consent be taken back, if ever? Compulsory affirmative consent would create a legal slippery slope and would likely result in a large amount of case law quickly. Requiring affirmative consent goes against democratic principles, and one could potentially make an argument that affirmative consent is unconstitutional in this circumstance.

9.1. Should consent of the copyright owner be required for all uses of copyrighted works to train AI models or only commercial uses?⁸

No, if consent is required in a scenario, then it should be limited to commercial use. An exception may be subscription services such as Midjourney.

9.2. If an “opt out” approach were adopted, how would that process work for a copyright owner who objected to the use of their works for training? Are there technical tools that might facilitate this process, such as a technical flag or metadata indicating that an automated service should not collect and store a work for AI training uses?⁹

If an “opt-out” approach were adopted, then perhaps the copyright holder’s work might be banned, in a sense, from appearing on search engines that use AI algorithms – particularly in

⁸ For example, the European Union’s Directive on Copyright in the Digital Single Market provides for two copyright exceptions or limitations for text and data mining (which may be used in the training of generative AI systems): one for purposes of scientific research and one for any other purpose. The latter is available only to the extent that rightsholders have not expressly reserved their rights to the use of their works in text and data mining. See Directive 2019/790 of the European Parliament and of the Council of 17 April 2019 on copyright and related rights in the Digital Single Market and amending Directives 96/9/EC and 2001/29/EC, 2019 O.J. (L 130), <https://eur-lex.europa.eu/eli/dir/2019/790/oj>.

⁹ For example, some AI companies have reportedly started to allow copyright owners to tag their works as not available for AI training. See Emilia David, Now you can block OpenAI’s web crawler, *The Verge* (Aug. 7, 2023), <https://www.theverge.com/2023/8/7/23823046/openai-data-scrape-block-ai>; Melissa Heikkilä, Artists can now opt out of the next version of Stable Diffusion, *MIT Tech. Review* (Dec. 16, 2022), <https://www.technologyreview.com/2022/12/16/1065247/artists-can-now-opt-out-of-the-next-version-of-stable-diffusion/>.

advertisements. I cannot speak to the current technological aspects of this question as I lack the expertise.

9.3. What legal, technical, or practical obstacles are there to establishing or using such a process? Given the volume of works used in training, is it feasible to get consent in advance from copyright owners?

It would seem unfeasible to get consent in advance from every copyright owner. Practically, the original author may be unavailable or unable to give consent for one reason or another (e.g., death, physically unavailable, etc.). Furthermore, how informed would this consent be? Though agreeing to lengthy, esoteric terms and conditions may be a legal way to consent, it would be up for debate whether that consent was informed consent. If it is not informed consent, it should not be considered consent, ethically.

9.4. If an objection is not honored, what remedies should be available? Are existing remedies for infringement appropriate or should there be a separate cause of action?

Remedies that are readily available should be employed for AI infringement claims. There should not be separate AI remedies from other remedies. As technology changes, more and more changes would be required, resulting in a plethora of remedies for the different technologies. Intellectual property infringement causes of actions and potential remedies should remain consistent; However, current remedies could be expanded upon to reflect emerging technology and issues as needed.

9.5. In cases where the human creator does not own the copyright—for example, because they have assigned it or because the work was made for hire— should they have a right to object to an AI model being trained on their work? If so, how would such a system work?

No.

10. If copyright owners' consent is required to train generative AI models, how can or should licenses be obtained?

This could be a condition or factor upon registration for a copyright through the Copyright Office.

10.1. Is direct voluntary licensing feasible in some or all creative sectors?

Voluntary licencing in creative sectors such as visual arts, literature, audiovisual works, and other creative works should be contractually considered among all relevant parties and should not be regarded as an administrative concern.

10.2. Is a voluntary collective licensing scheme a feasible or desirable approach?¹⁰ Are there existing collective management organizations that are well-suited to provide those licenses, and are there legal or other impediments that would prevent those organizations from performing this role? Should Congress consider statutory or other changes, such as an antitrust exception, to facilitate negotiation of collective licenses?

Yes, antitrust concerns and exceptions should always be considered from a legal standpoint. Pursuant to the Sherman Antitrust Act of 1890 and the goals of democratic laws, circumstances that raise antitrust issues should be deemed illegal in all circumstances. There is already a concern when it comes to a monopoly of power in the world of AI, so an economic monopoly should be prevented at all costs.

10.3. Should Congress consider establishing a compulsory licensing regime?¹¹ If so, what should such a regime look like? What activities should the license cover, what works would be subject to the license, and would copyright owners have the ability to opt out? How should royalty rates and terms be set, allocated, reported and distributed?

No. this ability should be limited to patents necessary for the emergent welfare of society. A compulsory licensing regime should not be relevant to other areas of intellectual property law, such as copyright.

10.4. Is an extended collective licensing scheme¹² a feasible or desirable approach?

¹⁰ Collective licensing is one alternative to a direct licensing regime, in which copyright owners negotiate and enter into private agreements on an individual basis. Under a collective licensing arrangement, rights are aggregated and administered by a management organization. The management organization negotiates the terms of use and distributes payment to participating copyright owners. See WIPO, WIPO Good Practice Toolkit for CMOs at 6 (2021), <https://www.wipo.int/publications/en/details.jsp?id=4561>.

¹¹ 50 A compulsory or “statutory” license allows for certain uses of a copyrighted work “without the consent of the copyright owner provided that the person adhered to the provisions of the license, most notably paying a statutorily established royalty to the copyright owner.” Music Licensing Reform: Hearing Before the Subcomm. on Intell. Prop. of the S. Comm. on the Judiciary, 109th Cong. (2005) (statement of Marybeth Peters, Register of Copyrights), <http://copyright.gov/docs/regstat071205.html>.

¹² 51 “An Extended Collective Licensing scheme is one where a relevant licensing body, subject to certain safeguards, is authori[z]ed to license specified copyright works on behalf of all rights holders in its sector (including non-members), and not just members who have given specific permission for it to act.” Extended Collective Licensing (ECL) scheme definition, LexisNexis Glossary (2023), <https://www.lexisnexis.co.uk/legal/glossary/extended-collective-licensing-ecl-scheme>; see also Letter from Karyn A. Temple, Acting Register of Copyrights, U.S. Copyright Office, to Rep. Robert Goodlatte, Chair, and Rep. John Conyers, Ranking Member, H. Comm. on the Judiciary (Sept. 29, 2017), <https://www.copyright.gov/policy/massdigitization/house-letter.pdf>; Letter from Karyn A. Temple, Acting Register of Copyrights, U.S. Copyright Office, to Sen. Charles Grassley, Chair, and Sen. Dianne Feinstein, Ranking Member, S. Comm. on the Judiciary (Sept. 29, 2017), <https://www.copyright.gov/policy/massdigitization/senate-letter.pdf>.

Depending on safeguards set in place, an extended collective licensing scheme could be a feasible or desirable approach concerning generative AI, but I believe this issue should be discussed in much more depth before it could potentially become a reality.

10.5. Should licensing regimes vary based on the type of work at issue?

Visual art versus deepfakes or issues right of publicity issues should not be given equal weight.

[Portions of this answer have been largely removed since the enactment of President Biden's new Executive Order concerning AI issued on October 30, 2023.]

11. What legal, technical or practical issues might there be with respect to obtaining appropriate licenses for training? Who, if anyone, should be responsible for securing them (for example when the curator of a training dataset, the developer who trains an AI model, and the company employing that model in an AI system are different entities and may have different commercial or noncommercial roles)?

As it would be a burdensome and intricate task, new employee training would be an option. As for who should be responsible, it should be someone with a legal and ethical background as a requirement. There should be mandatory ethics regulations when it comes to all generative AI works, particularly those concerning third-party licensing issues.

12. Is it possible or feasible to identify the degree to which a particular work contributes to a particular output from a generative AI system? Please explain.

Unless entire potential access to AI usage by a human using AI is offered or obtained, then it seems unlikely and unfeasible. Perhaps with an innovation of classifiers to recognize AI patterns or works, this identification may become possible. This may only be relevant to certain natural language processing or machine processing systems that experts in the field could speak to.

13. What would be the economic impacts of a licensing requirement on the development and adoption of generative AI systems?

Simply put, there would be less generative AI works, and worse generative AI works.

14. Please describe any other factors you believe are relevant with respect to potential copyright liability for training AI models.

– Public domain issues

Transparency & Recordkeeping

15. In order to allow copyright owners to determine whether their works have been used, should developers of AI models be required to collect, retain, and disclose records regarding the materials used to train their models? Should creators of training datasets have a similar obligation?

Yes, AI developers should be required to collect, retain, and disclose records regarding the materials used to train their models.

15.1. What level of specificity should be required?

Requirements for created works could include: how it was created, who (human) created it, what methods or models were included, was the work was contracted, who are all relevant parties involved in the creation of the work (e.g., work for hire, joint-authorship, etc.), for what purpose was the work created.

There is an argument that specificity requirements should largely be considered on a case-by-case basis. For registration by an individual author, there should be more leniency in order to promote creativity.

15.2. To whom should disclosures be made?

Disclosures should be made to the Copyright Office or relevant parties. For example, in a work-for-hire, if an individual creates a work with the assistance of AI, then the individual should disclose that information to their employer or commissioner. This type of disclosure could help prevent cases of secondary or vicarious liability on the side of the contracting party. Disclosures should be made in a similar way on joint-author works.

15.3. What obligations, if any, should be placed on developers of AI systems that incorporate models from third parties?

There may be issues of liability, particularly secondary liability or perhaps vicarious liability. It is not my view, however, that the potential for liability issues would constitute grounds for new obligations from an administrative level.

15.4. What would be the cost or other impact of such a recordkeeping system for developers of AI models or systems, creators, consumers, or other relevant parties?

One potential impact could be that recordkeeping systems could create jobs.

16. What obligations, if any, should there be to notify copyright owners that their works have been used to train an AI model?

In pursuant to the purpose of the fair use doctrine, 17 U.S.C. § 107, there should not be an obligation of notice.

17. Outside of copyright law, are there existing U.S. laws that could require developers of AI models or systems to retain or disclose records about the materials they used for training?

Edit: This answer may be addressed by President Biden’s Executive Order concerning AI issued on October 30, 2023.

Generative AI Outputs

If your comment applies only to a particular subset of generative AI technologies, please make that clear.

Copyrightability

18. Under copyright law, are there circumstances when a human using a generative AI system should be considered the “author” of material produced by the system? If so, what factors are relevant to that determination? For example, is selecting what material an AI model is trained on and/or providing an iterative series of text commands or prompts sufficient to claim authorship of the resulting output?

We need to reconsider what it means to be cognizant enough for creativity. A toddler cannot create what high GPT-4 models can create. From a philosophical perspective, one should consider the difference between being human and possessing the capacity for what we consider as aspects of humanity, such as consciousness, cognizance, moral reasoning, empathy, and other traits attributed to humanity. Though we are discussing narrow AI today, general AI, if it comes to fruition, could change our perspective of humanity in legal regard to the term “author.” If only a modicum of creativity is needed for copyrightability in the respect of an author, then AI systems could be authors, arguably.

19. Are any revisions to the Copyright Act necessary to clarify the human authorship requirement or to provide additional standards to determine when content including AI-generated material is subject to copyright protection?

Yes, the Copyright Act should possess a straightforward and clear definition of “author” in reference to being human rather than relying on standard statutory interpretations. Also, there needs to be clarification on the scope of “de minimus” in this seemingly formulaic consideration, or else the system could easily be taken advantage of.

The meaning of “author” should be clarified in 17 U.S.C. § 102 of the Copyright Act.

20. Is legal protection for AI-generated material desirable as a policy matter? Is legal protection for AI-generated material necessary to encourage development of generative AI technologies and systems? Does existing copyright protection for computer code that operates a generative AI system provide sufficient incentives?

Legal protection should be considered desirable in the discussion of policy. I do not hold specific views on other questions put forth here.

20.1. If you believe protection is desirable, should it be a form of copyright or a separate sui generis right? If the latter, in what respects should protection for AI-generated material differ from copyright?

Protection should be a form of copyright law rather than being a separate sui generis right.

21. Does the Copyright Clause in the U.S. Constitution permit copyright protection for AI-generated material? Would such protection “promote the progress of science and useful arts”?¹³ If so, how?

Article I, Section 8, Clause 8 of the United States Constitution could permit some form of protection for AI-generated materials if it were interpreted insofar that it is a “living document.” Our democracy of law depends on the ability to change with the needs of the general public as well as on an individual basis. It can be argued that the Copyright Clause should respect and reflect this philosophy.

Infringement

¹³ 52 U.S. Const. art. I, sec. 8, cl. 8.

22. Can AI-generated outputs implicate the exclusive rights of preexisting copyrighted works, such as the right of reproduction or the derivative work right? If so, in what circumstances?

I struggle to see a clear instance that an AI-generated work would intrinsically trigger issues regarding exclusive rights, given my opinions stated thus far.

23. Is the substantial similarity test adequate to address claims of infringement based on outputs from a generative AI system, or is some other standard appropriate or necessary?

It is my view that the substantial similarity test is no longer adequate to address infringement claims in reference to generative AI. The issue is not whether a particular system is more adequate than the substantial similarities test; rather, the issue is whether there is an existing standard capable of addressing generative AI creations.

24. How can copyright owners prove the element of copying (such as by demonstrating access to a copyrighted work) if the developer of the AI model does not maintain or make available records of what training material it used? Are existing civil discovery rules sufficient to address this situation?

The first question posed here seems rather rhetorical. If the records do not exist, then they cannot be accessed. There is an issue of accessibility to this question. Federal Rules of Civil Procedure (FRCP) can provide some enlightenment on the topic of civil discovery in regard to electronic discovery rules. FRCP 26, in addition to FRCP 33, could be considered. I cannot attest to whether they are sufficient given there are advantages and disadvantages of utilizing civil discovery rules in matters of AI.

25. If AI-generated material is found to infringe a copyrighted work, who should be directly or secondarily liable—the developer of a generative AI model, the developer of the system incorporating that model, end users of the system, or other parties?

The developers of the AI model in standard circumstances should not possess liability the developers are most often unaware of whether models have infringed on specific copyrighted works due to how most AI models are trained – this would still require a case-by-case analysis. Prompt users, or “end users of a system,” could possess direct liability if they possessed knowledge of the infringement. There could be direct or secondary liability in cases such as works for hire.

25.1. Do “open-source” AI models raise unique considerations with respect to infringement based on their outputs?¹⁴

If there is proper disclosure of relevant aspects of “open-source” AI models by the appropriate parties, then unique considerations may not be necessary.

26. If a generative AI system is trained on copyrighted works containing copyright management information, how does 17 U.S.C. 1202(b) apply to the treatment of that information in outputs of the system?

The outputs of AI-generated works should be considered under the same standards set forth in 17 U.S.C. 1202(b) regarding copyright management alteration.

27. Please describe any other issues that you believe policymakers should consider with respect to potential copyright liability based on AI-generated output.

There should be discussions on ethics, bias, data scraping, and prompt engineering.

Labeling or Identification

28. Should the law require AI-generated material to be labeled or otherwise publicly identified as being generated by AI? If so, in what context should the requirement apply and how should it work?

Potentially, the law may require such a label. As mentioned earlier, either plain disclosures or mathematical symbols e.g., \cap (intersection), \in (element of), or other symbols, could be added to the © symbol to identify that AI has assisted in a significant capacity.

28.1. Who should be responsible for identifying a work as AI-generated?

In short, humans should not be the sole, holders of this tremendous responsibility. I do not know who does, but humans are too fallible to identify AI-generated works, especially as the works become better as technology and AI users each improve. The Andy Warhol Foundation no longer authenticates works for this reason.

¹⁴ 53 Some AI models are released by their developers for download and use by members of the general public. Such so-called “open-source” models may restrict how those models can be used through the terms of a licensing agreement. See, e.g., Llama 2 Community License Agreement, Meta AI (July 18, 2023), <https://ai.meta.com/llama/license/> (requiring users of Llama 2 AI model to include an attribution notice and excluding use in services with greater than 700 million monthly active users).

AI itself could potentially identify AI-generated works, but that does not seem an appealing or foolproof system.

28.2. Are there technical or practical barriers to labeling or identification requirements?

The most practical barrier is whether or not it is feasible to create a system or method to correctly identify all AI-generated or AI-assisted works.

28.3. If a notification or labeling requirement is adopted, what should be the consequences of the failure to label a particular work or the removal of a label?

The consequences should not be dissimilar to current practices such as notice, takedown, fines, and so forth.

29. What tools exist or are in development to identify AI-generated material, including by standard-setting bodies? How accurate are these tools? What are their limitations?

No knowledge.

Additional Questions About Issues Related to Copyright

30. What legal rights, if any, currently apply to AI-generated material that features the name or likeness, including vocal likeness, of a particular person?

As this is a right of publicity issue, it is necessary to address here that the right of publicity is not governed by federal law; rather, it is governed by state law (as is acknowledged in the following question). . . .

Note: I cannot provide an adequate answers to this question since the enactment of President Biden's new Executive Order concerning AI issued on October 30, 2023.

31. Should Congress establish a new federal right, similar to state law rights of publicity, that would apply to AI-generated material? If so, should it preempt state laws or set a ceiling or floor for state law protections? What should be the contours of such a right?

There should be federal statutes for “deepfakes” that are intended to undermine the value and power of democracy. AI should not be allowed to influence elections or any other matter of the democratic process. This should especially apply to the name, image, and likeness of politicians in spreading false information that could be dangerous.

Note: Congress should follow the ideas set forth by President Biden’s new Executive Order concerning AI issued on October 30, 2023.

32. Are there or should there be protections against an AI system generating outputs that imitate the artistic style of a human creator (such as an AI system producing visual works “in the style of” a specific artist)? Who should be eligible for such protection? What form should it take?

Generally speaking, there should not be protections against imitations of style. Style is not intellectual property in and of itself. If copyright explicitly does not protect fact or idea, then style should not be protected as long as there is no false assertion of ownership (e.g., People should be allowed to create images in the styles of Picasso, Magritte, Warhol, etc., as long as they do not claim it as an original work of the human creator its style is imitating).

This type of protection, if any such protection should arise, should not be to protect human creators whose style is so well-known that it would be easy for a reasonable person, in this situation, to distinguish an original versus a generated version.

If “style” is an *idea*, *concept*, or *method* of creation, then it should not be copyrightable therefore, should not possess protection. To quote the Copyright Office itself:

“Copyright does not protect ideas, concepts, systems, or methods of doing something. You may express your ideas in writing or drawings and claim copyright in your description, but be aware that copyright will not protect the idea itself as revealed in your written or artistic work.”

(hyperlink:<https://www.copyright.gov/help/faq/faq-protect.html#:~:text=Copyright%20does%20not%20protect%20ideas,your%20written%20or%20artistic%20work>)

Note: This is subject to change with President Biden’s new Executive Order concerning AI issued on October 30, 2023.

33. With respect to sound recordings, how does section 114(b) of the Copyright Act relate to state law, such as state right of publicity laws?¹⁵ Does this issue require legislative attention in the context of generative AI?

Edit: I cannot appropriately answer this question until a more detailed reading of President Biden’s new Executive Order concerning AI issued on October 30, 2023.

¹⁵ Under 17 U.S.C. 114(b), the reproduction and derivative work rights for sound recordings “do not extend to the making or duplication of another sound recording that consists entirely of an independent fixation of other sounds, even though such sounds imitate or simulate those in the copyrighted sound recording.”

34. Please identify any issues not mentioned above that the Copyright Office should consider in conducting this study.

- Prompt engineering
- AI prompt engineering
- Ethics
- De minimus
- De-duplication versus unlearning

Conclusion: Just as the law, my views are constantly evolving with new technology and cultural shifts. The views expressed here are those of the deadline of this inquiry.