

October 17, 2023

To the U.S. Copyright Office in response to Notice of Inquiry and request for comments on artificial intelligence and copyright.

Comments by David Newhoff, creator and copyright advocate; colleague and member of Copyright Alliance; writer of the pro-copyright blog *The Illusion of More*; author of *Who Invented Oscar Wilde?: The Photograph at the Center of Modern American Copyright*; and co-founder of RightsClick, a software suite used by creators to register works and manage copyright portfolios.

Please accept my responses to the following selected questions:

8.1. In light of the Supreme Court's recent decisions in *Google v. Oracle America* and *Andy Warhol Foundation v. Goldsmith*, 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, raise different considerations under the first fair use factor?

In my view, neither case is helpful to a putative AI-developer defendant regarding the first factor question being asked. Under *Warhol*, there is no colorable defense that the purpose of AI training is to achieve “critical bearing” on the works used, and it is difficult to imagine how most, if any, developers would make such a claim. In *Oracle*, the reimplementation of APIs for the development of new computer programs is highly distinguishable from, for instance, copying a billion images in their entirety to “train” a machine to generate images. Further the Court cautioned that *Oracle* is narrowly tailored to computer programs as the copyrightable works in question.

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? 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?

This is one example in which generative AI can upend copyright doctrine. Even where a use may involve millions of works (e.g., *Google Books*), the fourth factor considers potential harm to the market for the *works*, whereas generative AI—if it does not produce market substitutes—primarily represents potential harm to *authors* and future authorship. While it is beyond the scope of copyright to protect creative jobs against technological changes per se, the consideration in the context of “training” should be expansive and doctrinal—namely that a potential threat to “authorship” cannot, by definition, “promote the progress” of “authorship.” Therefore, the fair use defense should be unavailable to the developer in this context. Where the AI does produce market substitutes, case law should be sufficient, including the fair use factor four inquiry.

10.3. Should Congress consider establishing a compulsory licensing regime? 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. With the development of streaming platforms, compulsory licensing devastated the livelihoods of songwriters. AI development is still nascent, and we cannot predict how the market will change in the future. Legislation of this nature is likely to be short-sighted and may lock in regimes that fail to serve authors.

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. Both parties should be obligated to collect and maintain records to foster transparency. That said, considering the lack of good faith shown by the tech sector in complying with regimes like the DMCA, provisions of this nature should contain actual penalties for failure to comply. Any proposal based upon a civil liability shield in exchange for compliance should be dismissed as a non-starter and a waste of time.

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?

The threshold for copyrightability must still be a “modicum of originality” contributed by the human, whether the tool used is a camera, keyboard, paintbrush, or, perhaps, an AI-based application. It is certainly possible for a human using a generative AI to be the “author” of the work, though, I believe, not the developer of the AI model or system. Merely selecting the materials used to “train” an AI cannot be considered “authorship,” let alone “authorship” of what may be tens of millions of outputs. Further, because it appears that most models require the ingestion of tens of millions of works in order to “learn,” this volume of collection by means of, for example, internet scraping is too indiscriminate to be considered “selection.”

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?

My answer to the first two questions is No, and, therefore, the third question is moot. In general, it is not desirable to attach copyright rights to AI-generated material any more than it is desirable to vest civil rights in robots. As stated above, if individual humans use certain AI-based tools to create works of expression, the use of these tools should not automatically disqualify the entire work from copyright protection.

But we must be cautious about vesting copyright rights in enterprise-scale, corporate production of works by, for instance, an AI-developer/producer. Beyond posing a threat to the careers of creative professionals (and to the cultural value of creative work), at a certain point, the application of copyright law itself may become irrelevant and/or unconstitutional. For instance, if generative AI were to foster an oligopoly of developer/producers, it is conceivable that copyright enforcement would become meaningless. Imagine the chaos (or futility) arising from a claim that AI-Developer Alpha allegedly infringed the work of AI-Developer Beta. Such a scenario raises difficult questions of standing and, as noted below, may frustrate the substantial similarity inquiry to the point of irrelevance.

Meanwhile, if these potential outcomes result in shrinking the population of working authors and the diversity of works, this would be anathema to the constitutional purpose to “promote progress.”

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”? If so, how?

While patent protection for AI systems may promote the “useful arts,” copyright protection for AI-generated works does not inherently promote “science” (a.k.a.) new creative expression. Again, if generative AI is likely to reduce the number of working “authors” in the U.S., this is offensive to the Progress Clause in Article I. That the American Framers could only conceive of human “authors” is not just a technicality of history. Whether in 1787 or today, law, like art, is a human construct that serves no purpose beyond human experience. As I have stated a few times here, it is contradictory to believe that one can promote “authorship” while obviating the role of “authors.” That the Progress Clause could be interpreted to encompass this result defies textual, doctrinal, and historical reason.

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?

Yes, and it’s happening right now. Users of generative AI are producing famous—and famously protected—reproductions and derivative visual works of, for instance, Marvel characters. While copyright owner Disney may not elect to take legal action against, for instance, parties sharing these outputs on social media, there is nothing about the use of AI in these examples that militates against finding that the copies and derivatives are infringing. In fact, it seems certain that if these outputs were to be used commercially, the inevitable litigation would be short work for the court.

As to whether the AI developer may be liable for these copies and derivatives, it seems straightforward to find that if, for instance, “Daredevil” was input for training and “Daredevil” was later output by the system, then the developer may be liable for both direct and secondary copyright infringement. Direct copyright infringement in violation of §106(1) occurs during input, and secondary infringement arises due to the developer’s failure to prevent the infringing work from being output.

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?

Combined with 24 below.

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 doctrines of “substantial similarity” and “access” may both be challenged by generative AI. First, if a given system is prompted to produce common or popular probability outcomes, it may generate thousands or millions of similar works, all of which are potentially non-infringing under the doctrine of “independent creation.” For example, the AI user who is unfamiliar with the work of Karla Ortiz may inadvertently produce a work that is “substantially similar” to one of Ortiz’s images, but the copy may be said to be “independently created” by that individual. The difficulty is novel because “independent creation” is historically anomalous whereas AI could make it rampant.

This also goes to the question of “access” and liability for that “access.” The AI developer undoubtably has “access” to any work fed into its model, but it seems unlikely the user of the AI can be said to have had “access” in this context. Then, because of the increased likelihood of multiple, “independently created” similar works, proving “access” may be moot with regard to the liability of the individual user of the 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?

Assuming the user of the AI knowingly made the infringing work, he/she is liable for direct infringement. Secondary liability for any developer may arise if the allegedly infringing work is a copy, or is “substantially similar,” to a whole work that was fed into the model or subsequent application.

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

I fail to see why “open source” would alter the consideration of an alleged infringement.

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?

As in pre-AI considerations, §1202(b) denies the AI developer an “innocent infringer” defense.

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?

A requirement to identify AI-generated material likely addresses topics outside the scope of copyright law. Presumably, the public would be best served by labels used to mitigate fraud and other forms of misinformation, and the complications arising from that intent are best left to the FTC, Congress, and agencies other than the Copyright Office. Even where AI may be used to create forgeries, this is already criminal conduct, and copyright plays little or no role. But unless a creative work is illegally presented as the work of a named artist, there is no compelling interest per se in notifying the public that a work, or part of a work, was generated by AI.

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?

Again, this not a copyright matter, but a federal ROP may address some of the already rampant, unethical uses of AI where the potential harm to both the infringed party and the public is significant. The 25 state ROP laws do not address, for instance, the potential harm caused by AI's capacity to generate "in the style of" works, especially in the commercial market.

If ROP law is expanded, it should 1) apply to all persons, not just celebrities; 2) anticipate and remedy AI-enabled harms stemming from misappropriation of likeness for purposes other than commercial advertising; and 3) not restrict expressive uses of AI-generated likeness for purposes (e.g., biographical films) that fall within the scope of protected speech.

In a world in which all media travels the globe instantly, a federal ROP statute would seem to be the only sensible framework in which to address the myriad potential harms. As for preemption, this may raise the cost of potential litigation by eliminating the option of a state filing, but further study into this as a matter of civil procedure is required.

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?

This concern could be addressed in a new, federal ROP statute while leaving undisturbed the doctrine that copyright does not protect "style." While an amendment to the Copyright Act akin to VARA could be written to encompass a narrow protection for "style," the intent seems better suited to ROP. Additionally, it may be easier and more effective to write a new law for the purpose of federal ROP than to amend the Copyright Act, especially when the U.S. is not a moral rights jurisdiction vis-à-vis copyright. Finally, it should be noted that, if protection of this nature were enforceable, it may create new licensing opportunities for artists and prospective commercial users.

As to application of the law, again, forgery is covered by the criminal code, but the most likely harm would seem to be commercial use of "in the style of" works in a manner that may implicate

the artist's reputation and/or deny her a commission or a licensing opportunity. Still, if such a right were to be established, exceptions would be required so that what we might call "reminiscent of" is distinguishable from "in the style of." This is a highly subjective consideration that may draw lessons from "substantial similarity" doctrine, even if the new right does not sound in copyright. For instance, prompting an AI for a work "in the style of [named artist]" may be analogized to the principle of "access."

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

In regard to disclaiming AI-generated material in a registration application, the current guidelines are likely to confuse applicants and overburden examiners who have neither the resources, nor necessarily the expertise, to engage in assessments normally left to the courts. Although it is understandable that the Office wishes to preserve the "human authorship" doctrine, asking an applicant *how* a work was made is a significant shift that should not be taken lightly.

Applicants are going to submit myriad statements which are either unfounded in law, or which beg for what amounts to a "substantial similarity" test on the part of the examiner. This may strain Office resources and potentially cost applicants additional fees. Instead, if the Office were to add a checkbox at the Certification stage of the application, asking whether the deposit copy(ies) contain any AI-generated material, the applicant will be given the opportunity to make a truthful statement subject to §506(e) while leaving the question of separating the AI material from the human authorship to the courts, as it should be.