Before the U.S. Copyright Office

Docket No. 2023-6

**Comments of the Library Copyright Alliance on the Inquiry Concerning
Artificial Intelligence and Copyright**

The Library Copyright Alliance ("LCA") consists of two major U.S. library associations: the American Library Association and the Association of Research Libraries. These associations represent over 100,000 libraries in the United States employing more than 300,000 librarians and other personnel. These two associations cooperate in LCA to address copyright issues that affect libraries and their users.

LCA participated in one of the Copyright Office's Public Listening Sessions regarding the copyright issues raised by generative artificial intelligence ("AI"), and also joined the Office's educational webinars. LCA welcomes the opportunity to provide comments in response to the Office's August 30, 2023 notice of inquiry ("NOI").

The recent emergence of generative AI systems has focused significant public attention on the intersection of copyright and AI. LCA developed the following principles for copyright and AI that we believe should steer policy discussions in this area. Based on these principles, LCA will respond to some of the NOI's questions.

I. Library Copyright Alliance Principles for Copyright and Artificial Intelligence

The existing U.S. Copyright Act, as applied and interpreted by the Copyright Office and the courts, is fully capable at this time to address the intersection of copyright and AI without amendment.

- Based on well-established precedent, the ingestion of copyrighted works to create large language models or other AI training databases generally is a fair use.

  - Because tens—if not hundreds—of millions of works are ingested to create an LLM, the contribution of any one work to the operation of the LLM is *de minimis*; accordingly, remuneration for ingestion is neither appropriate nor feasible.

  - Further, copyright owners can employ technical means such as the Robots Exclusion Protocol to prevent their works from being used to train AIs.

- If an AI produces a work that is substantially similar in protected expression to a work that was ingested by the AI, that new work infringes the copyright in the original work.

  - If the original work was registered prior to the infringement, the copyright owner of the original work can bring a copyright infringement action for statutory damages against the AI provider and the user who prompted the AI to produce the substantially similar work.

- Applying traditional principles of human authorship, a work that is generated by an AI might be copyrightable if the prompts provided by the user sufficiently controlled the AI such that the resulting work as a whole constituted an original work of human authorship.

AI has the potential to disrupt many professions, not just individual creators. The response to this disruption (*e.g.*, support for worker retraining through institutions such as community colleges and public libraries) should be developed on an economy-wide basis, and copyright law should not be treated as a means for addressing these broader societal challenges.

AI also has the potential to serve as a powerful tool in the hands of artists, enabling them to express their creativity in new and efficient ways, thereby furthering the objectives of the copyright system.

II. Responses to Specific NOI Questions

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?

Generative AI has the potential to dramatically democratize teaching, learning, research, and creativity.[1] As such, it could greatly assist libraries in achieving their mission to enhance learning and ensure access to information for all. Generative AI, however, may have a disruptive impact on the economy. Libraries are important community institutions that can play a critical role in the retraining of workers adversely affected by generative AI. They can also build on their historic roles supporting literacy in general and digital literacy in particular to assist users in understanding generative AI, how it might be used by others, and how they might use it.

5. Is new legislation warranted to address copyright or related issues with generative AI?

New legislation is not needed to address the copyright issues related to generative AI. The Copyright Act is sufficiently broad and flexible to enable courts to address the many different

---

[1] Alan Inouye, *Do your homework on the good of AI. Then legislate it*, The Hill, Sept. 23, 2023, https://thehill.com/opinion/congress-blog/4220232-do-your-homework-on-the-good-of-ai-then-legislate-it/.

copyright issues that may arise in the generative AI context. Since the Copyright Act of 1790, Congress has entrusted the courts to figure out how to apply a minimalist statutory framework to the myriad different factual circumstances that may arise. The Act does not define many of its critical terms, including "original work of authorship" and "expression." The Act does not set forth how one shows that an exclusive right has been violated; the phrase "substantial similarity" does not even appear in the Act. Nor does the Act establish standards for secondary liability. Courts, not Congress, first created the fair use right; and even when Congress finally codified fair use more than a century after its judicial creation, it provided four terse factors that have left much to the judicial imagination.

The brevity of the Act's core provisions has enabled courts to adapt the application of the Act to rapidly evolving technology. Indeed, courts have been applying the Act to AI for more than a decade. *Kelly v. Arriba Soft Corp.*, 336 F.3d 811 (9th Cir. 2003), *Perfect 10, Inc., v. Amazon.com, Inc.*, 508 F.3d 1146 (9th Cir. 2007), *A.V. ex rel. Vanderhye v. iParadigms, LLC*, 562 F.3d 630 (4th Cir. 2009), *Authors Guild, Inc. v. HathiTrust*, 755 F.3d 87 (2d Cir. 2014), and *Authors Guild v. Google, Inc.*, 804 F.3d 202 (2d Cir. 2015), all involved various forms of artificial intelligence.[2] To be sure, these cases did not concern *generative* AI, but the courts now hearing the pending challenges to ingestion for training generative AI models are perfectly capable of applying these precedents to the cases before them. In particular, the fair use conclusion concerning the ingestion of training data should not change just because the AI systems in dispute produce new text in response to user prompts instead of compilations of snippets of existing text.[3] Of course, the courts considering these cases may rule differently. Regardless, the most appropriate forum at this time for resolving the lawfulness of ingestion of training data is the judiciary, which can weigh all the nuances of a particular case, including the issues raised in question 8 of the NOI.[4]

If the courts apply fair use so narrowly that the development of generative AI is constrained, Congress should then consider appropriate legislation to enable the advancement of this technology. As revealed in the Copyright Office's educational webinar concerning international developments, other countries with sophisticated technology sectors have enacted provisions in their copyright laws that facilitate generative AI. The United States must not place itself at a competitive disadvantage relative to other countries by hindering the evolution of this important

---

[2] On the basis of these cases, the Copyright Office concluded that assembling a corpus of works in order to text and data mine the corpus for research purposes is a fair use. U.S. Copyright Office, *Section 1201 Rulemaking: Eighth Triennial Proceeding to Determine Exemptions to the Prohibition on Circumvention, Recommendation of the Register of Copyrights*, Oct. 2021, at 107-17, https://cdn.loc.gov/copyright/1201/2021/2021_Section_1201_Registers_Recommendation.pdf

[3] Because ingested works are "deconstructed" in the training process, the ingestion of any particular work would not harm the market for that work. AI providers use mitigation techniques such as de-duplication to prevent the possibility of unintended anomalous outputs.

[4] The Copyright Office in the report it produces after this inquiry should not attempt to answer these questions or even summarize the views of commenters. The Copyright Office shouldn't be giving advisory opinions on issues in active litigation.

tool. But at this point, we should let the courts apply fair use principles to the many different fact patterns presented in the cases centering on ingestion.

Similarly, the courts are fully capable of addressing whether the outputs of a generative AI system are infringing, and determining which parties are direct infringers and which are secondary infringers. As the law regarding infringement is virtually entirely judge-made, the courts are once again the most appropriate forum for adjusting existing tests for infringement, if necessary, in the generative AI output context.

Likewise, the courts can assess whether the Copyright Office has correctly applied the human authorship requirement to the issue of copyrightability. We have confidence that in most cases the Copyright Office will draw the line in the right place, but when it doesn't, the courts can step in and redraw the line.

Finally, with respect to voices and likeness, LCA objects to deepfakes that mislead users and harm creators. Although LCA is concerned about the potential abuse of an individual's name, image, and likeness, this is an issue outside of the scope of copyright law, and thus outside the expertise of the Copyright Office.[5] LCA would oppose legislation that purported to restrict the copying of "styles." This could chill creativity because artists in all fields imitate the styles of other artists. LCA looks forward to comments that identify the gaps, if any, in the Lanham Act, the Federal Trade Commission Act, and state rights of publicity and unfair competition law that need to be addressed by federal legislation in a manner that respects the First Amendment.

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)?

Copyright owners already have the technical means to opt out of the use of their works for training materials. If they place their works behind technological protections, the circumvention of those technological protections for the purpose of ingestion into training datasets would violate 17 U.S.C. § 1201.[6] Copyright owners can also employ robot exclusion protocols, which many AI providers respect.[7] Both Google and Microsoft recently fine-tuned their technology for responding to bot exclusion headers such that a webmaster can allow the crawling of her site for

---

[5] We note that members of the Senate Intellectual Property Subcommittee have circulated a discussion draft of the No Fakes Act of 2023. This is not the appropriate forum for discussion of the discussion draft.

[6] Exemptions granted in the Section 1201 triennial rulemaking permit circumvention of technological protections in order to engage in text and data mining for research purposes.

[7] Ignoring a robot exclusion protocol would not constitute a violation of Section 1201. As discussed below in response to question 34, compliance with robot exclusion headers could be addressed by voluntary codes of best practices.

search purposes but not AI training purposes.[8] The situation invites education for creators, who might not yet understand that they can use tools and what those tools are.[9]

At the same time, browse-wrap[10] license terms on websites that prohibit the ingestion of material should not be enforceable. With such licenses, there are serious questions whether there was sufficient manifestation of assent to form a binding contract.[11] Moreover, strong arguments can be made that browse-wrap licenses inconsistent with the fair use right are preempted either by the Constitution[12] or 17 U.S.C. § 301(a).[13] Notwithstanding these arguments, there is authority suggesting that contractual prohibitions on copying otherwise permitted by fair use may be enforceable.[14] Congressional intervention may be necessary to make clear that exceptions and limitations provided in the Copyright Act prevail over license terms inconsistent with those exceptions.[15]

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

This question correctly recognizes that collective licensing is relevant only if the copyright owners' consent is required to train generative AI models, that is, only if ingestion for training purposes is not a fair use. Because we believe it generally is a fair use, as discussed above,

---

[8] *See* Danielle Romain, *An update on web publisher controls*, Sept. 28, 2023, https://blog.google/technology/ai/an-update-on-web-publisher-controls/; *Announcing new options for webmasters to control usage of their content in Bing Chat,* Sept. 22, 2023, https://blogs.bing.com/webmaster/september-2023/Announcing-new-options-for-webmasters-to-control-usage-of-their-content-in-Bing-Chat.

[9] This education could be performed by the Copyright Office, artist rights associations, and of course libraries.

[10] In a browse-wrap license, a website's terms of service declare that by using the website, the user agrees to the website's terms of service. The adverse effects go beyond preventing the training for AI (*e.g.*, research and other scholarly uses).

[11] *See Nguyen v. Barnes & Noble*, 763 F. 3d 1171 (9th Cir. 2014).

[12] *See Vault Corp. v. Quaid Software Ltd*, 847 F.2d 255 (5th Cir. 1988).

[13] *See ML Genius v. Google*; *Data General Corp. v. Grumman Systems Support Corp.*, 36 F.3d 1147 (1st Cir. 1994); 1 Nimmer on Copyright § 1.01[B][1][a][i].

[14] *See Meyer v. Uber Technologies, Inc.*, 868 F.3d 66 (2d Cir. 2017); *Bowers v. Baystate Techs., Inc.*, 320 F.3d 1317 (Fed. Cir. 2003), *cert. denied*, 539 U.S. 928 (2003); *Davidson & Assoc. v. Jung*, 422 F.3d 630 (8th Cir. 2005); *ProCD, Inc. v. Zeidenberg*, 86 F.3d 1447 (7th Cir. 1996).

[15] For legislative protections in other countries on the overriding of exceptions by contracts, *see* Jonathan Band, *Protecting User Rights Against Contractual Override*, AU Washington College of Law Digital Commons (2023), *available at* https://digitalcommons.wcl.american.edu/research/97/. It should be noted that Article 7(1) of the EU Directive on Copyright in the Digital Single Market prohibits enforcement of contractual restrictions on the text and data mining exception in Article 3 of the Directive. *See also* Katherine Klosek, *Protecting Library Rights: Consideration for Congress*, 84 Ass'n of College and Research Libraries News (2023), https://crln.acrl.org/index.php/crlnews/article/view/26045/33975.

legislation establishing extended collective licensing or some other compulsory licensing regime is unnecessary.

Nor do we see any policy reason for Congress to enact legislation declaring that ingestion for training purposes is infringing, assuming that such legislation limiting the scope of fair use would be constitutional. It is premature, and completely speculative, to conclude that generative AI poses "a serious risk of market dilution from machine-generated works that can be cheaply mass-produced and inevitably lower the value of human-authored works."[16] And even if generative AI did pose this risk of market dilution, there is no evidence of a causal relationship between the ingestion of specific works of professional authors and the development of AI technologies that could have this dilutive effect. Moreover, generative AI's ability to enhance the creative process for all its users could offset the adverse impacts the technology might have on certain individual professional authors. Further complicating Congress's policy assessment is that the impacts of generative AI on the creative sector must be viewed in the much broader context of the impacts of generative AI on the economy and society at large.

Because of the enormous complexity of determining with any confidence the best path forward in this space for the American people, the copyright discussion relating to generative AI most profitably should focus not on inputs but on outputs: not on creating new collective licensing schemes for ingestion but on preventing the generation of infringing derivative works. Unfortunately, the Copyright Office appears to be paying more attention to the legal issues relating to training than those relating to infringement by outputs; the NOI asks 32 questions about training, and only seven questions about infringement by outputs. Hopefully the report itself will not reflect this imbalance.

Specifically, the report should note that AI providers already employ mitigation measures to inhibit their models from creating infringing works.[17] The report should encourage the

---

[16] Authors Guild, *FAQ on the Authors Guild's Positions and Advocacy Around Generative AI*, https://authorsguild.org/advocacy/artificial-intelligence/faq/. It is reasonable to assume that generative AI will have a disruptive effect on some artists in some sectors, but it is too early to predict which artists in which sectors will be most affected. Some authors appear more concerned with the fairness of technology companies profiting off their hard work than market dilution. William D. Cohan, *AI is learning from stolen intellectual property. It needs to stop.*, The Washington Post, Oct. 19, 2023, https://www.washingtonpost.com/opinions/2023/10/19/ai-large-language-writers-stealing/?utm_medium=email&utm_source=newsletter&wpisrc=nl_popns&utm_campaign=wp_opinions_pm. However, the Supreme Court made clear in *Feist v. Rural Telephone*, 499 U.S. 340, 349 (1991), that while "[i]t may seem unfair that much of the fruit of the compiler's labor may be used by others without compensation," copyright meets its Constitutional objective of promoting "the Progress of Science and useful Arts" precisely by "encourag[ing] others to build freely upon ideas and information conveyed by a work." *Id.*

[17] Peter Henderson, et al., *Foundation Models and Fair Use*, March 29, 2023, https://arxiv.org/pdf/2303.15715.pdf. Generative AI technology can also be used to develop tools that would assist courts in the infringement analysis. *See* Uri Hacohen and Niva Elkin-Koren,

enhancement, updating, and application of these measures. Additionally, although the courts are the most appropriate forum for adjusting existing tests for infringement in the generative AI output context, the Copyright Office can serve as a resource to courts and stakeholders by exploring in depth the complicated and subtle issues relating to the infringing status of outputs and direct and secondary liability.

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

One possible solution to many of the issues identified in the NOI is voluntary measures such as licensing arrangements and codes of best practices. The NOI asks about legislation relating to collective licensing, transparency, recordkeeping, labeling, identification, and likeness. But generative AI is in its infancy, rendering legislation premature. Private ordering in this new space has already begun. The Associated Press has licensed its archive of news stories to OpenAI.[18] The new collective bargaining agreement between the Writers Guild of America and the studios addresses the use of AI in television and film projects.[19] Legislation would interfere with this type of private ordering. Legislation would also preempt the emergence of voluntary codes of best practices relating to generative AI. Issues such as compliance with exclusion headers, transparency, recordkeeping, and labeling probably are better addressed by voluntary codes of best practices that could be updated regularly to respond to technological and commercial developments.

Respectfully submitted,

Jonathan Band
LCA Counsel
jband@policybandwidth.com

October 30, 2023

Copyright Regenerated: Harnessing GenAI to Measure Originality and Copyright Scope, Sept. 6, 2023, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4530717.

[18] Matt O'Brien, *ChatGPT-maker OpenAI signs deal with AP to license news stories*, July 12, 2023, https://apnews.com/article/openai-chatgpt-associated-press-ap-f86f84c5bcc2f3b98074b38521f5f75a.

[19] WGA Contract 2023, https://www.wgacontract2023.org/the-campaign/summary-of-the-2023-wga-mba.