REPLY COMMENTS IN RESPONSE TO THE COPYRIGHT OFFICE'S NOTICE OF INQUIRY ON ARTIFICIAL INTELLIGENCE AND COPYRIGHT

By Pamela Samuelson, Christopher Jon Sprigman, and Matthew Sag

December 6, 2023

Suzanne V. Wilson
General Counsel and Associate Register of Copyrights
U.S. Copyright Office
101 Independence Ave, S.E. Washington, D.C. 20559-6000

Re: Notice of inquiry ("NOI") and request for comments, Artificial Intelligence and Copyright, Docket No. 2023-6

Dear Associate Register Wilson:

We are U.S. legal academics who write and teach in copyright law. Pamela Samuelson is the Richard M. Sherman Distinguished Professor of Law at Berkeley Law, as well as a Professor at the School of Information, and Co-Director, Berkeley Center for Law & Technology. Christopher Jon Sprigman is the Murray and Kathleen Bring Professor of Law at NYU School of Law and a Co-Director of the Engelberg Center on Innovation Law and Policy. Matthew Sag is a Professor of Law in Artificial Intelligence, Machine Learning, and Data Science at Emory University Law School.[1]

On October 30, 2023, we submitted Comments in response to the Copyright Office's NOI and request for comments, Artificial Intelligence and Copyright, Docket No. 2023-6 (*see* Comment ID COLC-2023-0006-8854). We submit these Reply Comments to respond to Comments submitted that same day by the United States Federal Trade Commission (*see* Comment ID COLC-2023-0006-8630).

We should begin by noting our appreciation for the FTC's work enforcing both federal antitrust and consumer protection laws and helping to lead policy development in both areas. In our view, the FTC plays a vital role in keeping markets open and honest, and we have long been admirers of the intelligence and energy that the agency brings to that

---

[1] We offer these reply comments in response to the NOI in our personal capacities only and our views are not necessarily shared by our institutions or any other affiliated entities. Christopher Sprigman is a member of the law firm Lex Lumina PLLC, which represents clients in relation to copyright and AI issues; however, Sprigman has no involvement in that representation.

task. More specifically, we recognize the usefulness of examining intellectual property issues through the lenses of competition and consumer protection. However, in the case of its response to the Copyright Office's NOI on Artificial Intelligence and Copyright, the FTC has submitted Comments that are unclear and thus open to a variety of interpretations—and possibly to misinterpretations as well. The FTC's Comments also raise questions about the scope of agency's authority under Section 5 of the Federal Trade Commission Act, 15 U.S.C. 45, to bring enforcement actions aimed at activities, including those involving the training and use of AI, that might involve copyright infringement—although we would note that the copyright consequences of AI are, as yet, undefined.

We have three principal criticisms of the FTC's comments:

First, the FTC's submission is not a model of clarity: indeed, later in these Comments we will focus on a particular sentence from the FTC Comments that is worrisome both for its opacity and for the ways in which it may be interpreted (or misinterpreted) to chill innovation and restrict competition in the markets for AI technologies.

Second, the FTC Comments do not appear to be based on a balanced evidentiary record; rather, the Comments appear largely to reflect views articulated by participants in an Oct. 4, 2023 FTC Roundtable event[2] that featured testimony largely from artists and writers: 11 of the 12 witnesses appeared to be or to represent individual creators, and one represented open-source software developers who objected to AI training on their code—not a single witness provided perspectives from technologists who have developed and work with AI agents. Perhaps not surprisingly given the imbalance in the record, the FTC comments do not seem to appreciate the variety of use cases for AI technologies or the broader implications of those technologies for competition policy.

Third, and finally, certain of the FTC's Comments could, if misunderstood, upset the careful balance that the copyright laws create between private rights to control copyrighted works and public access and use of those works. Upsetting that balance could chill development not only of useful AI technologies, but of a range of new technologies and services that augment consumers' opportunities to access and use copyrighted works and increase the value of those works to consumers.

In the remainder of these Comments we will focus on a specific sentence from the FTC Comments that illustrates all of these problems. Specifically, under the heading of "**Copyrights and AI-generated Content,**" the FTC states the following:

---

[2] *See* https://www.ftc.gov/news-events/events/2023/10/creative-economy-generative-ai.

> "Conduct that may violate the copyright laws—such as training an AI tool on protected expression without the creator's consent or selling output generated from such an AI tool, including by mimicking the creator's writing style, vocal or instrumental performance, or likeness—may also constitute an unfair method of competition or an unfair or deceptive practice, especially when the copyright violation deceives consumers, exploits a creator's reputation or diminishes the value of her existing or future works, reveals private information, or otherwise causes substantial injury to consumers. In addition, conduct that may be consistent with the copyright laws nevertheless may violate Section 5."

This is a long and confusing sentence and it is difficult to restate with certainty what the agency is saying here. But however it is interpreted, the sentence presents several concerns:

1) First, the sentence seems to assume that training a machine learning model on copyrighted works made freely available on the open Internet is likely to be deemed (or should be deemed) a copyright violation. That is far too hasty. The copyright law implications of AI training are currently being litigated in several different federal copyright infringement actions. And moreover, as we detail below, the best understanding of the application of fair use principles to AI training would hold that the practice is in most if not all instances a fair use. On that point, time will tell. But at the moment, when the courts are still in the process of determining the law, the FTC should not be issuing statements that suggest that it has pre-judged the issue. The FTC has no authority to determine what is and what is not copyright infringement, or what is or is not fair use. Under governing law, that is a judicial function.

2) The FTC's undue haste to categorize AI training as likely infringement may be related to another error: the Comment's implicit understanding of AI training as a singular activity, rather than as another manifestation of something copyright law has dealt with many times before—i.e., so-called "non-expressive" use in which copying is undertaken not to distribute the copied material directly or indirectly but rather for some other purpose. The FTC Comments do not explicitly refer to or analyze the substantial body of court decisions holding that a range of non-expressive uses of copyrighted works are fair uses. As we explained in our initial Comments, U.S. courts have addressed the legality of non-expressive uses of copyrighted works in the context of other copy-reliant

technologies, including software reverse engineering,[3] plagiarism detection software,[4] and the digitization of millions of library books to enable meta-analysis, text data mining, and search engine indexing.[5] *Authors Guild, Inc. v. HathiTrust* is a particularly significant case in this regard because the district court in that case directly addressed the issue of text data mining.[6]

As one of us explains in a forthcoming law review article:

> *Text data mining is an umbrella term referring to computational processes for applying structure to unstructured electronic texts and employing statistical methods to discover new information and reveal patterns in the processed data. In other words, text data mining refers to any process using computers that creates metadata derived from something that was not initially conceived of as data. The process of text data mining can be used to produce statistics and facts about copyrightable works, but it can also be used to render copyrighted text, sounds, and images into uncopyrightable abstractions. These abstractions are not the same, or even substantially similar to, the original expression, but in combination they are interesting and useful for generating insights about the original expression.[7]*

Machine learning based on copyrighted works is an application of text data mining, not a separate technological or legal phenomenon. The copyright issues raised by text data mining are by and large the same as those raised by machine learning and generative AI. After all, it is hard to explain "why deriving metadata through technical acts of copying and analyzing that metadata through logistic regression should be fair use, but analyzing that data by training a machine learning classifier to perform a different kind

---

[3] Sega Enters. v. Accolade, Inc., 977 F.2d 1510, 1514 (9th Cir. 1992); Sony Computer Ent. v. Connectix Corp., 203 F.3d 596, 608 (9th Cir. 2000).

[4] A.V. ex rel. Vanderhye v. iParadigms, LLC, 562 F.3d 630, 644–45 (4th Cir. 2009).

[5] See Authors Guild, Inc. v. HathiTrust, 755 F.3d 87, 100–01 (2d Cir. 2014); Authors Guild v. Google, Inc., 804 F.3d 202, 225 (2d Cir. 2015).

[6] Authors Guild, Inc. v. HathiTrust, 902 F. Supp. 2d 445, 460, n22 (SDNY 2012) ("The use to which the works in the HDL are put is transformative because the copies serve an entirely different purpose than the original works: … The search capabilities of the HDL have already given rise to new methods of academic inquiry such as text mining. ... Mass digitization allows new areas of non-expressive computational and statistical research, often called 'text mining.'")

[7] Matthew Sag, *Copyright Safety for Generative AI*, 61 Hous. L. Rev. 305 (2023) (available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4438593)

of logistic regression that produces a predictive model wouldn't be."[8] This is particularly significant given that the Copyright Office itself has recognized the fair use status of TDM research.[9]

3) The FTC's Comment does not consider the impact on academic research or private sector technology development of holding that express consent is required merely to derive or extract abstract uncopyrightable information from copyrighted works using text data mining or machine learning. Most importantly for the FTC's mission, we see no prospect that erecting such a substantial barrier to computational analysis of text, images, and sounds would enhance competition. Although some AI development is being undertaken by large technology companies, AI is in fact a diverse market, with small players engaged in many facets of AI development along with bigger firms. If the FTC is suggesting that consent is required for training, that is likely to raise barriers that will reduce the ability of smaller firms to compete, relative to larger firms which will be better positioned to bear the costs of a permissions requirement.

4) Relatedly, the hostility that the FTC Comments express in relation to indemnification is puzzling. The ability to indemnify end-users will lead to market concentration and the privileging of incumbents only if the broad affordance that current copyright law provides for non-expressive uses of copyrighted materials is narrowed or overturned. Moreover, the increasing prevalence of end-user indemnification has many potential pro-competitive justifications: indemnities reduce consumer confusion in the face of speculative and unsubstantiated allegations of infringement; indemnities allocate the risk of copyright infringement liability to the parties most able to take precautions; and indemnities provide a mechanism to encourage noninfringing uses of generative AI and discourage potentially infringing uses (because they fall outside the scope of the indemnity).

5) Finally, we question whether, even under a broad understanding of the FTC's Section 5 authority, the agency can declare, as it suggests in its Comments, that the asserted copyright violation of training an algorithm on copyrighted works is an unfair method

---

[8] *Id.*

[9] *See* U.S. COPYRIGHT OFF., SECTION 1201 RULEMAKING: EIGHTH TRIENNIAL PROCEEDING, RECOMMENDATION OF THE REGISTER OF COPYRIGHTS, 121–24 (2021), https://cdn.loc.gov/copyright/1201/2021/2021_Section_1201_Registers_Recommendation.pdf. (In evaluating the proposed DMCA § 1201 exemption to circumvent technological protection measures on DVDs and eBooks for the purpose of conducting TDM, the Copyright Office said: "Balancing the four fair use factors, with the limitations discussed, the Register concludes that the proposed use is likely to be a fair use.")

of competition that violates Section 5.[10] We are concerned especially about the suggestion in the Comments that AI training might be a Section 5 violation where it "diminishes the value of [a creator's] existing or future works." A hallmark of competition is that it diminishes the returns that producers are likely to garner relative to a less competitive marketplace. This is just as likely to be true in markets for creative goods like novels and paintings as it is in markets for ordinary tangible goods like automobiles and groceries. AI agents that produce outputs that are not substantially similar to any work on which the AI agent was trained, and are thus not infringing on any particular copyright owner's rights, are *lawful competition* for the works on which they are trained.[11] Surely the FTC does not plan to have Section 5 displace the judgments of copyright law on what is and what is not lawful competition?

Moreover, even if, contrary to our expectations, courts declare AI training to be infringement (because outside the protection of the Copyright Act's fair use provision), the FTC should think long and hard before layering the prospect of Section 5 liability on

---

[10] We understand that deceptive advertising and other consumer misrepresentations may incidentally involve copyright violations, but in those scenarios, the FTC's authority to act would be grounded in the unfairness of the deception or misrepresentation, not in any potential copyright violation. We note that in one of the first class actions filed in relation to generative AI, *Andersen v. Stability AI Ltd.*, Judge Orrick dismissed the plaintiffs' unfair competition claims, noting that they were preempted by the Copyright Act and that the plaintiffs had failed to allege plausible facts in support of their theory that users of a text-to-image model could be deceived. Andersen v. Stability AI Ltd., No. 23-CV-00201-WHO, 2023 WL 7132064, at *14 (N.D. Cal. Oct. 30, 2023). Likewise, in another high-profile class action relating to generative AI, *Kadrey v. Meta Platforms, Inc.*, the trial court ruled that the plaintiffs' unfair competition claims must also be dismissed. Kadrey v. Meta Platforms, Inc., No. 23-CV-03417-VC, 2023 WL 8039640, at *2 (N.D. Cal. Nov. 20, 2023). Judge Chhabria noted that "[t]o the extent it is based on the surviving claim for direct copyright infringement, it is preempted. … To the extent it is based on allegations of fraud or unfairness separate from the surviving copyright claim, the plaintiffs have not come close to alleging such fraud or unfairness." *Id.*

[11] The trial court in *Andersen v. Stability AI Ltd.*, granted defendants' motion to dismiss in relation to the plaintiffs' theory that the output of generative AI models were necessarily "all infringing derivative works" regardless of their substantial similarity to the plaintiffs' original expression. Andersen v. Stability AI Ltd., No. 23-CV-00201-WHO, 2023 WL 7132064, at *7-8 (N.D. Cal. Oct. 30, 2023). Likewise, the court in *Kadrey v. Meta Platforms, Inc*. dismissed as implausible the class action plaintiffs' claim for copyright infringement based on the theory that "every output of the LLaMA language models is an infringing derivative work." Kadrey v. Meta Platforms, Inc., No. 23-CV-03417-VC, 2023 WL 8039640, at *1 (N.D. Cal. Nov. 20, 2023) The court concluded that "[t]he plaintiffs are wrong to say that, because their books were duplicated in full as part of the LLaMA training process, they do not need to allege any similarity between LLaMA outputs and their books to maintain a claim based on derivative infringement. To prevail on a theory that LLaMA's outputs constitute derivative infringement, the plaintiffs would indeed need to allege and ultimately prove that the outputs "incorporate in some form a portion of" the plaintiffs' books. *Id.*

6

top of the remedies already available under the Copyright Act. Although only injunctive relief is available under Section 5, there is a risk—especially palpable given the hostility to AI training that is apparent in the FTC's Comments—that the Agency will fashion injunctions that reach beyond the specific activities at issue in a particular dispute and implicate AI training activities or AI outputs that copyright law's flexible fair use doctrine might recognize as lawful. There is a risk, in other words, that the FTC will misuse its broad authority to fashion injunctive relief in ways that chill otherwise lawful competition.

In conclusion, although we are disappointed in the FTC's initial Comments in this matter, we hope that in the future the FTC will have an important role to play in exploring issues of competition and consumer protection that arise in relation to AI. To do so productively, the agency must take the time to gather the facts from all stakeholders, explore the complex interplay between copyright and competition interests, and consider the competition and consumer protection implications of the entire range of possible use cases for AI technologies.