

Before the
United States Copyright Office
Washington, DC

In re

Artificial Intelligence and Copyright

Docket No. 2023-6, COLC-2023-0006

**REPLY COMMENTS OF
THE COMPUTER & COMMUNICATIONS INDUSTRY ASSOCIATION (CCIA)**

In response to the notice of inquiry (“NOI”) and request for comments published by the U.S. Copyright Office (“the Office”) in the Federal Register at 88 Fed. Reg. 59942 (Aug. 30, 2023), and extended at 88 Fed. Reg. 65205 (Sept. 21, 2023) and 88 Fed. Reg. 78393 (Nov. 15, 2023), the Computer & Communications Industry Association (“CCIA”)¹ submits the following reply comments. CCIA appreciates the opportunity to provide additional input on these important issues following our participation in the Office’s listening sessions on copyright and AI this spring, as well as first-round comments.

I. Introduction

As stated in our initial comments, existing U.S. copyright law is capable of addressing issues related to artificial intelligence and serves to promote creative activity in AI technology. The technology-neutral nature of the Copyright Act is sufficient to address present issues regarding AI and copyright. CCIA members have a significant interest in ensuring that the development and use of AI technology continues to be promoted, rather than suppressed, by the U.S. copyright system.

¹ CCIA is an international, not-for-profit trade association representing a broad cross section of communications and technology firms. For more than 50 years, CCIA has promoted open markets, open systems, and open networks. CCIA members employ more than 1.6 million workers, invest more than \$100 billion in research and development, and contribute trillions of dollars in productivity to the global economy. A list of CCIA members is available at <https://www.cciagnet.org/members>.

II. The FTC's Submission Misunderstands Copyright

The Federal Trade Commission (“FTC” or “the Commission”), which has no jurisdiction over copyright, submitted a comment in this docket. Almost half of the Commission’s filing is devoted to the FTC’s interest in AI generally, and is not germane to the NOI. When the FTC finally does address copyright and AI-generated content (in a section barely a page long), it claims that such content may unfairly harm creators’ ability to compete, and may deceive consumers when they think a work has been created by a particular musician or other artist but it has been generated by someone else using an AI tool. The Commission offers neither analysis nor solutions—nor does it respond to any of the questions the Office asked about this issue in the NOI.

More problematic are the FTC’s unsupported statements regarding training. The FTC assumes, without evidence, that generative AI will cause harm to the creators whose works were included in training databases. The FTC further assumes that the scraping of training data from publicly available websites should trigger liability. The discussion reflects no awareness of the policies underlying copyright: that copyright is a monopoly granted by the government to authors only for the purpose of providing them with an economic incentive to create works for public benefit; and that this monopoly contains important limitations to ensure that the public receives that benefit. Among these limitations is fair use, which the Supreme Court recognized in *Google v. Oracle* is essential to promoting competition. Rather than understanding the pro-competitive nature of fair use, the FTC implies that it encourages unfair competition and that “the evolution of the doctrine could influence the competitive dynamics.”

While the FTC appears skeptical of the position that the unauthorized ingestion of training data can be lawful, it also is troubled by the ability of “large technology firms . . . to

obtain exclusive licenses to copyrighted (or otherwise proprietary) training data, potentially further entrenching the market power of these dominant firms.” In the FTC’s view, if AI firms collect training data without authorization, they are engaging in unfair competition, but if they license the training data, they are entrenching their market power. The FTC also ignores that licensing training data is likely to benefit content industry behemoths, further encouraging consolidation and oligopolistic behavior in that industry: only large copyright owners have the ability to offer large-scale licenses, giving them a further leg up on individual creators. These dynamics are already apparent in other digital licensing markets.² Fair use is the obvious way out of this conundrum, yet the FTC stresses that “conduct that may be consistent with the copyright laws nevertheless may violate Section 5.” This is certainly possible—as deceptive conduct is rarely infringing—but the observation does not play a role in this inquiry.

The final section of the FTC’s submission contains a summary of a roundtable the Commission convened on the “Creative Economy and Generative AI” (“the Roundtable”) — despite the FTC’s decision not to invite any representatives from AI firms to the Roundtable. The FTC states that it appended the transcript of the Roundtable “to assist the Copyright Office in assessing the harms generative AI systems pose to creative professionals.” The FTC’s summary of the Roundtable refers to “great power imbalance,” with “[c]ertain types of creative professionals” being “particularly vulnerable to exploitation” — “especially those who lack awareness or leverage in negotiating contract terms or the resources to enforce their rights.” While this issue could benefit from further attention, it is outside the scope of this inquiry.

² See, e.g., Meredith Rose, *Streaming in the Dark: Competitive Dysfunction within the Music Streaming Ecosystem*, Berkeley J. Ent. & Sports L. (forthcoming 2024), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4586800; Xiyin Tang, *Privatizing Copyright*, 121 Mich. L. Rev. 753 (2023).

III. Text and Data Mining Is Consistent with the Berne Convention

Several comments wrongly claimed that there was only limited support internationally for copyright exceptions for AI, and that those exceptions were out of sync with the Berne Convention. For example, the Copyright Alliance attempts to minimize the international acceptance of copyright exceptions relating to artificial intelligence by stating that “only the European Union, Japan, Singapore, and the United Kingdom have AI exceptions within their copyright laws.”

However, the European Union includes 27 countries and represents 15% of global gross domestic product as well as a substantial share of the world’s creative production. Its member states, such as Germany, France, the Netherlands, and Sweden, are among the world’s most technologically sophisticated, along with the other countries that have adopted AI exceptions: Japan, Singapore, and the United Kingdom. The Copyright Alliance also neglects to mention that the Israel Ministry of Justice recently issued an opinion letter stating that its fair use provision generally permits ingestion of training material for machine learning.³ Further, they imply that Hong Kong, South Korea, and Canada have rejected copyright exceptions for AI, when in fact their consultation processes are still underway.

Finally, both the Copyright Alliance and the National Music Publishers’ Association (“NMPA”) assert that the text and data mining (“TDM”) exceptions in the EU, Japan, Singapore, and the UK do not comply with the Berne three-step test. NMPA claims these TDM exceptions are “[s]weeping carveouts” that “go significantly beyond ‘special cases’” (the first Berne step). To the contrary, these exceptions typically are extremely limited, with clear definitions of text

³ Jonathan Band, *Israel Ministry of Justice Issues Opinion Supporting the Use of Copyrighted Works for Machine Learning*, Disruptive Competition Project (Jan. 19, 2023), <https://www.project-disco.org/intellectual-property/011823-israel-ministry-of-justice-issues-opinion-supporting-the-use-of-copyrighted-works-for-machine-learning/>.

and data mining, and restrictions on who can engage in it, for what purpose, and who has access to the assembled data.

For its part, the Copyright Alliance points to the “quickly developing licensing markets for use of copyrighted works by AI developers” as evidence that AI exceptions would conflict with the normal exploitation of works (the second Berne step). But a non-expressive use for training an AI model could hardly be called a normal exploitation of a work, and those “quickly developing licensing markets” are often licenses for access to works that are behind paywalls and thus currently are not being ingested into training datasets. Indeed, as noted above, such licenses are likely to worsen the problems of monopolization: only highly consolidated industries, or dominant market participants, are likely to be able to offer licenses that have any impact given the massive quantities of data necessary to train large language models.⁴ Further, treating the licensing markets for AI use of copyrighted works as evidence that such a use is normal exploitation is circular. If AI ingestion of lawfully acquired copyrighted works does not require a license, then there is no conflict with the normal exploitation of works.

The Copyright Alliance also suggests that the “inevitable market harm such exceptions would cause to copyright owners” is proof that these exceptions would unreasonably prejudice the legitimate interests of the right holder (the third Berne step). This market harm is completely speculative as well as again being circular—it presumes that model training requires a license in order to create a legitimate interest. Fair use of copyrighted material may sometimes cause a market harm, but nonetheless does not unreasonably prejudice the copyright owner’s legitimate interests. Moreover, any harm from model training would result from the extraction of non-

⁴ See Rebecca Tushnet, *All of This Has Happened Before and All of This Will Happen Again*, 29 Berkeley Tech. L.J. 1447 (2017) (explaining how Getty Images and similar dominant players attempt to use licensing to entrench monopolistic positions).

protected material from ingested copyrighted works, and accordingly would not be within the legitimate interest of the right holder.

IV. News Industry Claims on Fair Use Are Not Accurate

Contrary to the News Media Alliance’s comments and submitted white paper, AI models demonstrably do not contain the creative expression of news content, both technologically and legally.⁵ Generative AI systems based on large language models (“LLMs”) require the assembly of a dataset which is transformed and analyzed to create the model. An AI language model based on a small amount of data will not have enough information from which to derive the highly complex and nuanced rules, meanings, and contradictions of human language. And a model trained on a narrow category of works—like scientific journals—will not have enough information to accurately discern the patterns of modern, everyday human speech. To build a model that can realistically emulate all facets of human language, the developer needs a collection of data that includes a very large number of examples, reflecting a broad range of speech. However, it does not incorporate that data directly.

Instead, during training, a model evaluates the proximity, order, frequency, and other attributes of portions of words, called tokens, in its training data. In fact, the model itself selects which attributes to use. In this way, training is the discovery of probabilities of relationships between the tokens — ultimately not in any individual text, but in all of the text on which the model is trained. The trained model then comprises a large network of weights that represent these learned relationships. However, the model itself does not contain any of the content or the

⁵ See generally Project DisCo, *Common Misconceptions About Generative AI and Copyright*, Disruptive Competition Project (Aug. 1, 2023), <https://www.project-disco.org/intellectual-property/common-misconceptions-about-generative-ai-and-copyright/>.

expression from the content originally crawled by the bots.⁶

Copyright of course protects the way facts or ideas are expressed, but not the facts and ideas themselves. Leaving facts and ideas unprotected is a constitutional requirement under the First Amendment. Grammar, syntax, and characteristics of human language are also outside of copyright law. The First Amendment also permits the fair use of copyrightable expressive content. Because the LLM does not contain anyone's expression, it does not infringe copyright.

As for the copying necessary to create the dataset from which the LLM is derived, there is clear relevant precedent. Although high-quality generative AI is new, AI itself has been in use for at least two decades; and several courts have found that the copying involved in developing these AI tools is a fair use. These tools include plagiarism detection software, optical character and speech recognition, and search engines for websites and books. Most copyright experts believe that the fair use analysis for generative AI is the same as it is for these other AI tools.⁷

V. Conclusion

CCIA appreciates the opportunity to comment on these important issues and would be happy to provide any additional assistance that might be useful to the Office as it prepares its report.

Respectfully submitted,

Ali Sternburg, Vice President, Information Policy
Josh Landau, Senior Counsel, Innovation Policy
Computer & Communications Industry Association (CCIA)
asternburg@ccianet.org

December 6, 2023

⁶ *C.f.* CCIA Comments In re Artificial Intelligence and Copyright, Docket No. 2023-6, COLC-2023-0006 (filed Oct. 30, 2023), at 15-16 (answering Question 12), <https://ccianet.org/wp-content/uploads/2023/10/CCIA-Comments-to-Copyright-Office-on-AI.pdf>.

⁷ *See, e.g., Artificial Intelligence and Intellectual Property – Part II: Copyright and Artificial Intelligence Before the S. Comm. on the Judiciary, Subcomm. on Intellectual Property* (statement of Matthew Sag, Professor of Law in Artificial Intelligence, Machine Learning, and Data Science, Emory University School of Law), 118th Cong. (2023), https://www.judiciary.senate.gov/imo/media/doc/2023-07-12_pm_-_testimony_-_sag.pdf.