**Copyright Office**
**Docket No. 2023-6**
**Artificial Intelligence and Copyright – Notice of inquiry and request for comments**

**COMMENTS OF DIGITAL CONTENT NEXT**

*Filed with www.regulations.gov*

Founded in 2001, Digital Content Next (DCN) is the only trade organization wholly dedicated to serving the unique and diverse needs of high-quality digital content companies that manage trusted, direct relationships with consumers and marketers. DCN's members[1] are some of the most trusted and well-respected publishing brands that, together, have an unduplicated audience of 259 million unique visitors and reach 95 percent of the U.S. online population.

Consumers visit these publishing brands because they value trustworthy information, which has gone through a rigorous editorial process. Indeed, publishers are held to account for the veracity and trustworthiness of their content by audiences who have high expectations. Given the early challenges with hallucinations[2] in AI outputs and the potential for spreading misinformation at scale, our hope is that AI companies will partner with premium publishers who have vast experience in producing trustworthy content.

Against this backdrop, DCN appreciates the opportunity to provide the perspective of the online publisher community to the Copyright Office.

**Copyright Law Applies**

Copyright law was built to promote and protect the creation of original works. Copyright holders have the exclusive right to distribute and monetize their copyrighted works as they see fit for a defined period of time. These protections incentivize the creative industries by allowing them to reap the fruits of their labors and enabling them to reinvest into new content creation. The benefits to our society are nearly impossible to quantify as the varied kinds of copyrighted material enrich our lives daily: music, literature, film and television, visual art, journalism, and other original works provide inspiration, education, and personal and societal transformation. The Founding Fathers included copyright in the Constitution (Article I, section 8, clause 8) because they recognized the value of original works.

Today, publishers use the internet as a key distribution and monetization means. Our members make their copyrighted content available to consumers through a wide range of means,

[1] *See* https://digitalcontentnext.org/membership/members/ for a listing of current DCN members.
[2] https://www.techtarget.com/whatis/feature/Model-collapse-explained-How-synthetic-training-data-breaks-AI

including on websites and apps that are supported by various methods for monetization. The primary revenues for publishers are advertising and, increasingly, subscriptions. Just because copyrighted content is widely available on the internet does not make it free for the taking nor extinguish its copyrighted status.

Generative artificial intelligence systems and large language models, like most technological developments, present both risks and opportunities. They are merely tools, however, and DCN does not believe that under current conditions existing copyright law is not up to the task of promoting the use of those technologies to create copyrighted works and to the task of preventing and remedying any misuse of them that infringes on the rights of the owners of the protected works. Unfortunately, in many cases, these AI systems and models have violated, and will continue to violate, copyright law by scraping and copying, and generating output based on, protected content available on the internet without the consent of the original copyright holder and, in many cases, their knowledge. These actions undermine the Constitutional mandate "to promote the Progress of Science and the useful Arts, by securing for a limited Times to Authors and Inventors the exclusive Right to their respective Writings and Discoveries."

**Fair Use**

Unlicensed use of copyrighted works constitutes infringement unless it falls within a specific statutory exception or constitutes a "fair use." Some entities have claimed that the unlicensed use of copyrighted works to train generative AI systems is per se a fair use. That is incorrect. In determining whether such unlicensed use constitutes a fair use, a court would apply four factors under a fact-specific, case by case analysis. The fair use analysis is a balancing test with no single factor inherently dispositive. The four factors are as follows:

1. **The purpose and character of the use.** Under this factor, a court examines how the party claiming fair use is using the copyrighted work. Commercial uses of AI would be less likely to be found fair as compared to nonprofit educational and non-commercial uses, regardless of whether the user itself is non-commercial. Thus, AI companies that are using copyrighted works to train systems for their own profit are less likely to be found engaging in fair use. This does not mean, however, that all nonprofit education and non-commercial uses would be deemed fair. Indeed, one AI company that started as a not-for-profit is now reportedly raising capital at an $80 billion valuation[3].

   The critical inquiry for the first factor, per the Supreme Court's recent opinion in *Warhol v. Goldsmith,* "focuses on whether an allegedly infringing use has a further purpose or different character, which is a matter of degree, and the degree of difference must be weighed against other considerations, like commercialism." Without a sufficient degree of "further purpose or different character," there is no fair use, and the commercial nature of the use only militates against fair use.

[3] https://www.nytimes.com/2023/10/20/technology/openai-artifical-intelligence-value.html

**2. The nature of the copyrighted work.** Under this factor, a court examines the degree to which the work that the AI system used for training relates to copyright law's objective of promoting original expression. Use of an expressive work is less likely to be deemed fair use than use of materials that merely list facts and data with no contextualization.

**3. The amount of copyrighted material used in relation to the copyrighted material as a whole.** Under this factor, the court looks at both the size and significance of the portion of the copyrighted work that the AI system used. To the extent that AI systems have admittedly copied large swaths of copyrighted material – in many circumstances, 100 percent of the consumer-facing material may be copied – it would weigh against a finding that the use is fair. Even use of a small portion, however, can be found unfair to the extent that it is at the heart of the original work or a unique element.

**4. The effect of the use on the current and potential markets**. Under this factor said by the Supreme Court to be the most important factor, the court examines the extent to which the use of the work by the AI system undermines the existing or future market for that work. The AI use would be less likely to be found fair to the extent that the AI system's output competes with the original work. Many AI companies are using these systems to create or power consumer-facing services which threaten to displace traditional publishers. A student researching a specific topic might turn to ChatGPT instead of reading a publisher's news articles on that topic, thus depriving, the publisher of the opportunity to show advertising or charge a subscription or other fee. Similarly, a consumer looking to purchase a new household appliance might ask an AI system for its recommendations, which are sourced from several publishers' well-researched articles on the best and least expensive appliances, instead of reading the publisher's reviews. Publishers lose the opportunity to attract audiences, sell advertising or subscriptions, and to realize any e-commerce revenue. Moreover, there are long-established licensing markets for publishers' content for a plethora of online uses and there are burgeoning markets to license copyrighted works as training data for AI tools. For example, several AI developers such as OpenAI and Adobe have entered into licensing agreements with different types of content owners to use their content as training data. Using copyrighted works without such licenses would eviscerate both the current and potential markets for licensing such works. Thus, any use that either serves as a potential substitute or competitor for the underlying work or supplants a market for the copyright owner to license such works for training AI tools would weigh against a finding of fair use. Finally, this appropriation of value may make it impossible for publishers to continue to create, develop, and publish new articles and other materials, which is surely not in the public interest.

Several courts have found that certain uses of copyrighted works by search engines, as traditionally implemented, qualify under "fair use." The case-by-case nature of the fair use

analysis, however, means that the findings in such cases are not dispositive of the analysis in any given AI case especially since there are some important differences between the search engines at issue in those cases and AI systems. Traditionally, search engines display no more content than is required to alert a user as to where to go to find more and link the user to the original source.

The search engine cases do nothing to help generative AI against charges of copyright infringement. Generative AI returns paragraphs and even entire articles copied or extracted from the content and expression of copyrighted publisher works. Even where links are provided, there is often no reason for the user to use them – they may even serve as signaling that a response is correct and not a "hallucination".

In sum, search engines may provide limited benefits to the creators of the original works. But indiscriminate, unlicensed ingestion of millions of copyrighted works in their entirety to create competing commercial works does not.

Some AI companies have taken steps recently to address some of these concerns by giving content creators greater control over whether their content is scraped by bots, although these methods will not remove previously scraped content.[4] While this may be helpful with respect to future scraping, such steps are not sufficient to eliminate all concerns. They are an opt-out system, rather than truly voluntary. Copyright holders would need to identify and take proactive steps for the ever-expanding number of AI systems, many of which may take different approaches, and would need to implement those approaches for each copyrighted work in what may be extremely large catalogs. There is also no guarantee that all AI systems will adopt and honor such measures. And, of course, such a system runs contrary to the basic philosophy of copyright law that the copyright owner controls who may or may not use its copyrighted work through affirmative licensing.

Commitments by developers of commercial and public-facing AI systems to disclose the works they use for training would also help copyright holders ensure that their works have not been used inappropriately.

To the extent that AI system developers wish to obtain more certainty regarding their use of copyrighted works for training than potentially provided by the fair use doctrine, the best course, as always, is to license the works. Not only would that minimize litigation risk, it would also likely improve results as the process of licensing would result in higher quality and more tailored sets of content upon which to train. The additional virtue of a license agreement is that the parties can agree on the scope and nature of disclosure and attribution, as well as provide a remedy for harms to the publisher caused by AI hallucinations. Finally, licensing would help

---

[4] https://blogs.bing.com/webmaster/september-2023/Announcing-new-options-for-webmasters-to-control-usage-of-their-content-in-Bing-Chat
https://blog.google/technology/ai/an-update-on-web-publisher-controls/

maintain the incentive for publishers to continue creating quality new content, which provides significant value for society and for continued training of AI systems.

LLM providers have claimed that licensing is impractical and a licensing requirement would bring AI to a standstill. This is clearly incorrect. There are a number of successful private models for licensing content in large amounts. Examples include: photographs (e.g., Shutterstock, Getty Images); musical compositions (ASCAP, BMI, SESAC, GMR); Print (the Copyright Clearance Center); Merlin (digital music services); Monotype (fonts) and many others in the U.S. and around the world. In addition, LLMs already pay for other inputs such as computing time and engineering costs. A blanket "fair use" designation would amount to a subsidy for AI systems and LLMs at the expense of copyright holders.

**AI Litigation**

As of the date of this filing, more than 10 lawsuits have been filed this year against AI system companies, alleging violations of the Copyright Act, the Digital Millennium Copyright Act, the Computer Fraud and Abuse Act and various state statutory and common law provisions. We are concerned that it could take years for these cases to be resolved and that the resolutions may not reach precedential determinations of issues of general applicability. During this time, AI companies continue to crawl the web and train their systems on copyrighted works in infringing ways and to the point that it may be difficult to pull back. Courts may therefore benefit from Copyright Office guidance on how to apply copyright law, including the fair use doctrine, to use of copyrighted works to train AI systems.

**Conclusion**

We urge the Copyright Office to make clear that use of copyrighted works to train AI is not per se a fair use, and to explain how each of the fair use factors would apply to various AI training scenarios. Doing so would help promote beneficial uses of AI while minimizing copyright infringement and help courts resolve AI-related copyright disputes correctly and faster.

We appreciate the opportunity to comment in these proceedings and look forward to working with you to ensure the Copyright Act remains relevant going forward.

Digital Content Next
October 30, 2023