ASSOCIATED PRESS

200 Liberty Street
New York, New York 10281
T 212.621.1500

www.ap.org

October 27, 2023

United States Copyright Office
Docket No. 2023-6
Artificial Intelligence and Copyright Study
Notice of inquiry and request for comments
Submitted via: regulations.gov

Re: Comments by The Associated Press on Artificial Intelligence and Copyright Study

Dear U.S. Copyright Office,

I write on behalf of The Associated Press ("AP") in response to the August 30, 2023, Notice of Inquiry ("NOI"), issued as part of the Copyright's Office's ongoing AI Initiative and its study of the intersection between copyright law and emerging artificial intelligence.

The AP is an independent global news organization dedicated to factual reporting. Since its founding in 1846, AP, whose members include thousands of U.S. newspapers and broadcasters, has been breaking news and covering the world's biggest stories. AP news reaches more than four billion people every day, published by news organizations on their websites around the world. AP is thus uniquely positioned to comment on the impact that the development of artificial intelligence will have on publishers and the news industry.

In short, we believe there are great opportunities to be gained from emerging AI technologies. AP sees a world in which AI can bolster newsrooms, providing journalists with the tools they need to enhance newsgathering capabilities and allowing news publishers to explore new ways of delivering news and reaching audiences, all consistent with the AP's traditional mission to inform the world. However, we also see tremendous risks, particularly if AI is allowed to develop in a way that threatens the integrity of the information ecosystem.

One concern is that frequent "hallucinations" by generative AI systems, or even abuse of those same systems by malignant actors, will lead to the spread of mis- and dis-information and all the evils associated therewith.

Another concern, and perhaps more relevant to this particular comment request, is that AI systems would be allowed to develop in a way that does not respect the intellectual property rights of those upon whose content the systems are trained.

Indeed, it is a guiding principle for the AP that ***intellectual property rights must be respected, and content creators fairly compensated, for use of their works in developing AI***. AI systems are trained upon massive datasets of copyright protected works, including, we understand, proprietary news articles published by media organizations on their websites. AI developers must be required to obtain authorization from the copyright holders of those works in order to use them in AI training. To allow otherwise would be copyright infringement, and would impose severe economic harm on those who create expressive works.

***

We appreciate the Copyright Office's substantial efforts to examine the applicability of copyright laws to emerging AI technologies. This comment letter focuses mainly on the first policy issue identified by the Copyright Office in its Notice of Inquiry—*i.e.*, "(1) the use of copyrighted works to train AI models." The NOI states, "there is disagreement about whether or when the use of copyrighted works to develop datasets for training AI models . . . is infringing," and it asks, "whether permission by and/or compensation for copyright owners is or should be required when their works are included."

The answer is simple: yes.

Copying protected works for AI training without permission is prima facie infringement.

Copyright law requires that those exploiting copyrighted works do so only with the copyright owner's authorization or under some other recognized defense.

In this case, AI developers use copyrighted works in order to train their AI models. The developers are thus using the underlying works to create a new product of value for their own—and their customers'—exploitation. Consider, for example, a generative AI chatbot that trains upon the copyrighted works of others in order to respond to user requests for a summary of recent news. It is not difficult to imagine that these new AI systems will be of *immense* commercial value, perhaps even replacing user inquiries for the underlying copyright-protected works themselves.

When this training of AI models is done without permission from the owners of the underlying works, it is *prima facie* copyright infringement. It would *also* be copyright infringement if any of the "output" of that AI model includes more than *de minimis* amounts of the expressive content of any of the underlying works used for training.

Training AI models with copyright-protected works is not fair use.

Proponents of the fair use argument for AI training will argue that training on the works of others is a necessary step towards creating transformative new systems. But this argument overstates the nature of change being wrought to existing works. It also ignores the severe economic harm that will be rendered upon content creators if this type of activity is blessed by copyright law.

Several legal cases are working through the fair use questions and will ultimately render a verdict. We are confident courts will interpret existing laws to require that using copyrighted works to train AI systems requires a license, and is not fair use. If they find otherwise, however, it will be important for the Copyright Office to issue guidance and regulations that protect a copyright holder's ability to control the use of their works.

A couple of points worth noting about the fair use analysis:

- *First*, an AI developers' wholesale copying of content to train its AI models supplants the owner's licensable training data, rather than "transforming" it. AI training adds nothing new to the original works; it merely uses them to create new works that supersede or

supplant them. Because one purpose of news publishers' content is to license it to AI developers for model training, the use by AI developers "share[s] the objectives" of news publisher's content. *See Andy Warhol Found. for the Visual Arts, Inc.* v. *Goldsmith*, 143 S.Ct. 1258, 1279 (2023).

In fact, news publishers have already developed a licensing market for machine learning. For example, in July 2023, the AP announced a licensing deal allowing Open AI to train its AI models on portions of the AP's text archive of news stories.[1] In that respect, news publishers are similar to Lynn Goldsmith, who had licensed her photograph to serve as an artist reference for Andy Warhol; the Warhol Foundation's subsequent unlicensed use was therefore unfair. *Id.* at 1278. Freely copying a work for a use that the creator had otherwise licensed weighs against fair use.

- The end product of AI training can also closely mimic the purpose and character of the underlying copyrighted works, even if the final product does not technically reproduce the expressive elements of those works. Consider again the example of a generative AI system that trains upon copyrighted works to produce a chatbot that can be prompted to provide a summary of recent news. Whether or not that GAI summary reproduces expressive elements from the underlying works in its output, its delivery of news is nonetheless made possible by the initial copying of the underlying works. It then creates an end product with virtually the same purpose—*i.e.*, to inform users of newsworthy events. *See, e.g.*, *Associated Press* v. *Meltwater U.S. Holdings, Inc.*, 931 F. Supp. 2d 537, 553 (S.D.N.Y. 2013) (Meltwater's news-clipping service was not fair use because it "free ride[s] on the costly news gathering and coverage work performed by other organizations," and "avoid[s] paying licensing fees [which] gives it an unwarranted advantage over its competitors who do pay licensing fees.").

- *Second*, a "finding of fair use is more likely when small amounts, or less important passages, are copied than when the copying is extensive, or encompasses the most important parts of the original." *Authors Guild* v. *Google, Inc.*, 804 F.3d 221 (2d Cir. 2015). This factor plainly weighs in favor of news publishers because generative AI models use works in their entirety, and do so on a massive scale.

- *Finally*, unauthorized use of news publisher content by AI generators can cause significant commercial harm to the market for the underlying works, including (i) lost licensing opportunities with the specific AI developers; (ii) undermining licensing opportunities with other customers working with large data sets; (iii) normalizing the behavior, making it more difficult to address from a legal or licensing perspective at later date; and (iv) creation by AI developers of commercial products that compete with news publishers.

[1] *AP, Open AI agree to share select news content and technology in new collaboration*, The Associated Press (July 13, 2023), *available at* https://www.ap.org/press-releases/2023/ap-open-ai-agree-to-share-select-news-content-and-technology-in-new-collaboration.

Under the fourth fair use factor, courts consider "not only the extent of market harm caused by the particular actions of the alleged infringer, but also 'whether unrestricted and widespread conduct of the sort engaged in by the defendant . . . would result in a substantially adverse impact on the potential market' for the original." *Campbell* v. *Acuff-Rose Music, Inc.*, 510 U.S. 569, 590 (citation omitted). Courts consider the "impact on potential licensing revenues for traditional, reasonable, or likely to be developed markets." *Fox News Network, LLC* v. *TVEyes, Inc.*, 883 F.3d 169, 180 (2d Cir. 2018).

As stated, a licensing market already exists for the use of copyrighted works in AI training and development. These AI developers recognize the copyright risk as well as the importance of compensating creators. Thus, permitting free use of publisher content for training "would effectively destroy that broader market" for training data. *Andy Warhol Found. for Visual Arts, Inc.* v. *Goldsmith*, 11 F.4th 26, 50 (2d Cir. 2021), *aff'd*, 143 S. Ct. 1258 (2023).

Moreover, that market is extremely valuable, with "deep-pocketed consumers . . . willing to pay well" for content to train their AI models. *TVEyes*, 883 F.3d at 180. Allowing free use of publisher content would decrease its value and the incentive to create, which is the goal of copyright.

- News publishers would be seriously harmed if the AI model "outputs" result in substantially similar and competing commercial products that take away audiences from the underlying content. The failure to establish an AI legal framework that respects IP rights would have a devastating impact on the media industry, undermining core media business models based on audience, licensing and advertising. Among other things, it would allow AI systems to leverage the value of news content—for free—while creating commercial products designed to compete with the industry. This is against the law and fundamentally unfair, and would impose severe economic harm on those who create expressive works.

It is the last point that is perhaps the most concerning. The news industry already sits in a perilous position, with many news organizations, particularly local news outlets, having suffered significant damage to their business models by the emergence of digital platforms that dominate the advertising market. News content is created at great expense for media companies, and often with great personal sacrifice by individual journalists. To allow AI companies to leverage this valuable content to create competing products that take away readership—and to do so for free, without compensation—would be a bridge too far.

Damage to the news industry would mean news companies are no longer able to produce news. The flow of critical information to the public breaks down if there is no means to support it. This results in an uninformed citizenry and impacts a fundamental pillar of democracy.

***

In sum, the AP believes there are significant benefits to be gained from the development of AI technologies. But it is critical that such systems be developed responsibly and in a way that

respects the intellectual property rights of content creators and ensures creators are fairly compensated for use of their works in training AI.

We appreciate your consideration of this matter.

Respectfully submitted,

The Associated Press

Karen Kaiser, Senior VP & General Counsel
Brian Barrett, Associate General Counsel