OXFORD

Journal of Intellectual Property Law & Practice, 2025, Vol. 20, No. 8

DOI:10.1093/jiplp/jpaf046
Advance access publication date 4 August 2025
Editorial

AI inputs, fair use and the US Copyright Office Report

The USA has yet to produce determinative caselaw on whether inputting works to compile a generative Artificial Intelligence (AI) system's training data is a fair use. Judicial rulings, however, may soon emerge, as many of the multiple pending cases are reaching the stage of a judgment on the merits of the copyright owners' infringement claims. In addition, the US Copyright Office recently issued *Part 3 Generative AI Training* of a report requested by Congress on *Copyright and Artificial Intelligence*, in which the Office extensively and rigorously examined the application of copyright law to the copying of protected works to assemble data to train generative models.

The Copyright Office primarily focused its analysis of copyright infringement liability for massive copying into training data on fair use, having concluded that entering works into AI systems made copies in prima facie violation of the exclusive rights in the works: 'Creating and deploying a generative AI system using copyright-protected material involves multiple acts that, absent a license or other defense, may infringe one or more rights'. The Office also reserved questions of liability regarding the *outputs* of AI systems for a future Part 4 (whose completion and disclosure remain uncertain, given Register of Copyrights Shira Perlmutter's firing, closely following her dissemination of a prepublication version of Part 3).

The Report acknowledged that 'The stakes are high, and the consequences are often described in existential terms. Some warn that requiring AI companies to license copyrighted works would throttle a transformative technology, because it is not practically possible to obtain licenses for the volume and diversity of content necessary to power cutting-edge systems. Others fear that unlicensed training will corrode the creative ecosystem, with artists' entire bodies of works used against their will to produce content that competes with them in the marketplace. The public interest requires striking an effective balance, allowing technological innovation to flourish while maintaining a thriving creative community'. Accordingly, eschewing general pronouncements and grand policy stances, the Report carefully considered the four statutory fair use factors, emphasizing the importance of the facts underlying particular inputs. Its cautious conclusion: 'Various uses of copyrighted works in AI training are likely to be transformative. The extent to which they are fair, however, will depend on what works were used, from what source, for what purpose, and with what controls on the outputs—all of which can affect the market. When a model is deployed for purposes such as analysis or research—the types of uses that are critical to international competitiveness—the outputs are unlikely to substitute for expressive works used in training. But making commercial use of vast troves of copyrighted works to produce expressive content that competes with them in existing markets, especially where this is accomplished through illegal access, goes beyond established fair use boundaries'.

The Report is available on the Copyright Office website[1]; it is not my purpose here to summarize its entire fair use analysis. Rather, I wish to examine how the Office addressed the issue of amassing training data in order to produce newly-generated works that compete with, but do not substantially copy from, the ingurgitated works. The fourth fair use factor—often dubbed the 'single most important factor'—instructs courts to inquire into the 'effect of the use upon the potential market for or value of the copyrighted work'. The doctrine generally excludes copying that takes from the author's work in order to produce a competing substitute, thereby usurping the market for the work. In the AI context, however, the effect of copying into training data may principally impact the market for or value of the artist's present and future *oeuvre* in general, rather than the market for any specific copyrighted *work* (ie the works copied into the system).

Under Berne Convention Article 9(2), and WIPO Copyright Treaty (WCT) Article 10, the 'three-step test' authorizes member states to create exceptions and limitations to the reproduction right; the third step requires that the exception or limitation 'not unreasonably prejudice the legitimate interests of the *author*'. An AI output that competes with an artist's *oeuvre* in general or with her future work may not supplant the market for any particular copied work, but that output may indeed 'unreasonably prejudice the legitimate interests of the author' in making a living and continuing her creative activities. The relevant interests in the Berne Convention and the WCT focus on the author rather than on particular works and might provide an avenue for objection to copying works into AI systems for the purpose of emulating (without specifically copying) an author's creative output.

The Berne Convention, however, is not self-executing in the USA, and Congress has declared that 'the provisions of the Berne Convention … shall not be enforceable in any action brought pursuant to the provisions of the Berne Convention itself'. (Berne Convention Implementation Act of 1988, Pub. L. No. 100-568, 102 Stat. 2853, sec. 3(a)(2).) Nonetheless, the disparity between unreasonable prejudice to authors and deleterious effects on the markets for the copied works may be narrower than first appeared. The Copyright Office Report shows why.

In its analysis of the fourth fair use factor, the Report considers 'market dilution' as an aspect, broader than 'lost sales', of the effect of copying works into training data. Recognizing concerns that factor four should address only the impact on the market for the copied works, lest a broader assessment of market harm 'stifle innovation and creativity in AI development', the Report responds:

> While we acknowledge this is uncharted territory, in the Office's view, the fourth factor should not be read so narrowly. The

[1] See https://www.copyright.gov/ai/Copyright-and-Artificial-Intelligence-Part-3-Generative-AI-Training-Report-Pre-Publication-Version.pdf.

> statute on its face encompasses any "effect" upon the potential market. The speed and scale at which AI systems generate content pose a serious risk of diluting markets for works of the same kind as in their training data. That means more competition for sales of an author's works and more difficulty for audiences in finding them. If thousands of AI-generated romance novels are put on the market, fewer of the human-authored romance novels that the AI was trained on are likely to be sold. Royalty pools can also be diluted. [Phonogram producer] UMG noted that "[a]s AI-generated music becomes increasingly easy to create, it saturates this already dense marketplace, competing unfairly with genuine human artistry, distorting digital platform algorithms and driving 'cheap content oversupply' - generic content diluting human creators' royalties."

The Report here emphasizes that the effect of the copying impacts currently extant works by putting them in competition with AI-generated outputs. The copyrighted works serve as fodder for new productions for the same markets (romance novels; personalized streaming content) in which the copied works operate. Note that the Report does not claim that the competing AI outputs copy the inputted works; rather, the Report examines the economic consequences of copying the inputs that enable the competing outputs. It is important to retain the focus on the inputs, because absent the predicate copying, there is no infringement action against flooding the market with independently generated romance novels. If 1000 human Barbara Cartland wannabe's publish their (not copied) contributions to the genre, those entries will compete with Cartland, but competition without copying does not infringe copyright (even if the wannabe's read many Cartland novels to learn the tropes of the genre). The question with which the Report grapples is whether fair use excuses predicate copying that does not comprise the end product if the effect of the copying may be to substitute for the source works. The Report's broad understanding of 'potential market for or value of the copied work' underpinned its comprehensive assessment of market harm.

Perhaps more boldly, the Report brought style copying within the ambit of cognizable market harm:

> Market harm can also stem from AI models' generation of material stylistically similar to works in their training data. As the Office noted in Part 1 of this Report, many commenters raised concerns about AI outputs that imitate a creator's style, which copyright does not protect as a separate element. Even when the output is not substantially similar to a specific underlying work, stylistic imitation made possible by its use in training may impact the creator's market. In the words of the Writers Guild of America, because AI systems can be prompted to imitate a writer's style, applying fair use would force writers "to compete with AI-generated scripts trained on their work, without their authorization, and without fair compensation." This threat is more acute because of the technology's ability to produce works so similar in style "that the average person cannot discern a difference in the marketplace[,] … creat[ing] direct competition with the creators whose works have been used to train the model."

The Report calls to mind an arguably abusive employment practice, in which the employer orders an employee to train her successor, then promptly fires her. Doubly deleterious, the substitution effect here results from drafting the author into enabling the AI program to generate outputs that not only occupy the same slice of the market (eg romance novels), but also target the author's own works. Again, the Report does not assert that copying style alone supports an infringement action, even though that copying may compete with the works that inspired the emulations. Rather, the predicate copying of expression (indeed, entire works) preconditions the generation of targeted substitutes. Seen in that light, copying in order to imitate an individual author's style bears an even more direct effect on the market for the author's copied works than copying to enable new entrants into a more broadly defined market (eg romance novels). This analysis suggests (and the Report so indicates in the context of its analysis of the third fair use factor) that AI developers might well consider implementing 'guardrails' to avoid generating the outputs that are most likely to substitute for individual authors' creations.

Jane C. Ginsburg*

*Morton L. Janklow Professor of Literary and Artistic Property Law Columbia University School of Law. Email: jcg2@columbia.edu

Journal of Intellectual Property Law & Practice, 2025, **20**, 521–522
https://doi.org/10.1093/jiplp/jpaf046
Editorial