Kluwer Copyright Blog

Is Generative AI Fair Use of Copyright Works? NYT v. OpenAI

Mira T. Sundara Rajan (UC Davis Law School) · Thursday, February 29th, 2024

The case recently brought against OpenAI by the New York Times is the latest in a series of legal actions involving AI in the United States, and mirrored in other countries –notably, the UK. In order to train their technologies, should AI companies be allowed to use works under copyright protection without consent? The lawsuits brought by the owners of such works, including artworks in the case of image-generators and journalism in the NYT case, claim that this should not be allowed. Such uses, they argue, constitute copyright infringement.

Photo by Jakayla Toney on Unsplash

Fair Use Precedent? Google Books and Transformative Use

The past two decades have seen a wealth of technological developments, but generative AI is qualitatively different from everything that has come before. Rather than focusing on the reproduction and dissemination of existing materials, the goal of AI is to rework them to create something new. In this regard, an important precedent lies in the history of US litigation involving Google Books. Over the course of a decade, Google copied large volumes of books and made them available online, both through excerpts, known as "snippets", and as entire publications. As in the present context, the initial concern of copyright holders was that their consent had not been acquired by Google prior to scanning their works.

Judge Denny Chin initially found Google liable for failing to secure the consent of copyright owners before scanning their books. But he eventually reversed his own position. In 2013, after a decade of litigation, accompanied by a counterpoint of shifts in the book publishing industry driven by rapid technological change, Judge Chin ultimately found that Google's scanning of the books amounted to fair use of those works. As such, it was permissible under United States copyright law.

The key finding in Google Books was that Google's actions were "transformative." In other words, Google did not merely copy the books; it made use of them to create a new and valuable product, in the form of the Google Books service, and one that, according to the court, did not compete with the existing market for books. Instead, Google Books was found to support the marketing of books by giving them increased public exposure. Since there was no appreciable harm to the copyright owners, according to the Court – quite the contrary –it was clearly acceptable under the terms of United States copyright law.

Given this background, it should come as no surprise that OpenAI now claims fair use. Nevertheless, its ability to make this case successfully is far from self-evident.

Clarifying fair use: the role and limits of "transformative use"

Under United States law, as elsewhere, eligible works are protected automatically upon their creation by copyright law. To be eligible, they must be original works of human authorship which are recorded, or "fixed," in a tangible medium. Once these threshold requirements have been met, and as long as the copyright term continues to be in force, any substantial use of the copyright work– for example, copying large parts of it to use in the creation of another work – requires the consent of its author.

Notwithstanding this framework, not every use of a work during the term of copyright is restricted. Certain insubstantial or minor uses – a single line quoted from a book, for example – are allowed. Further, specific criteria go on to outline the possibility of more significant permitted uses under U.S. copyright law. These criteria are found in section 107 of the Copyright Act.

Copyright laws throughout the world incorporate features that allow for circumstances in which works can be used in spite of copyright restrictions – for example, fair dealing in the UK and Canada, free use in Germany, and limitations and exceptions in European and international copyright regulation. However, the U.S. doctrine of fair use is also unique in certain respects. The key to understanding it lies in the language of section 107, which sets out the criteria. It begins with a non-exhaustive list of examples of permitted fair uses, such as criticism and research. This list is followed by the famous "four factor test" – a description of the four factors which must be considered when assessing "whether…any particular case is a fair use". This is the great strength of the fair use doctrine: it is said to be flexible, making it responsive to technological change. This responsiveness is why other countries have become interested in U.S.-style fair use: South Korea adopted it in 2011, and it has been considered for adoption by Australia. However, this very flexibility can also make fair use difficult to apply in practice. Indeed, it has been called "the most troublesome in the whole law of copyright."

The modern understanding of transformative use, which lies at the heart of fair use, originally emerged from a 1990 article by Judge Pierre Leval. The idea arises from factor one, "the purpose and character of the use, including whether such use is of a commercial nature or is for nonprofit educational purposes," which Judge Leval calls the "soul of fair use." "Transformative use" arguably brings a new dimension to fair use, infusing the doctrine as a whole. In a sense,

transformative use has helped further to update United States copyright law for technology.

NYT v. OpenAI: The Defendant's Position

In the NYT case, Open AI's reliance on the doctrine of transformative use, particularly as it has been recognized in the Google Books precedent, is logical. OpenAI's arguments focus on the capacity of generative AI to "transform" the works used in training into a new form – its "generative" capacity, which neither aims at, nor leads primarily to, the creation of exact or "substantially similar" copies of the original works. Crucially, the use of NYT works without consent has not been contested by Open AI. Instead of arguing against the allegations of prima facie copyright infringement made by the Times, OpenAI is simply arguing that any such infringements are justified under the doctrine of fair use.

Notably, it is for this reason that Sam Altman, the CEO of OpenAI, has directly addressed only one of the specific types of allegations of copying made by the New York Times. This involves the making of exact reproductions of NYT articles, which OpenAI argues is a "bug" that the developers intend to resolve – that the purpose of OpenAI is not the exact reproduction of training materials, but the generation of new texts arising out of the information absorbed through training. In other words, the company claims that its products are intended to generate, and do generate, "transformed" works, which will ultimately not bear any significant resemblance to the works that were originally copied. Accordingly, OpenAI has suggested that the exact copying of NYT works detailed in the complaint was facilitated by detailed prompts that would ordinarily violate OpenAI's terms of use. All of this is meant to support the idea of the "transformativeness" of OpenAI's technology.

NYT v. OpenAI: The Plaintiff's Case and Why it Should Succeed

However, it is by no means a foregone conclusion that OpenAI will succeed by asserting transformative use. On the contrary, there are powerful arguments against a finding of fair use in the NYT case.

Balancing the Four Factors

The secret lies within the doctrine of transformative use itself. As originally explained by Judge Leval, "[t]he existence of any identifiable transformative objective does not, however, guarantee success in claiming fair use. The transformative justification must overcome factors favoring the copyright owner."

Here, in profound contrast to the situation involving Google Books, generative AI is creating products that are competing directly with works created by the New York Times – and with those of the other, human writers and artists who are suing AI companies. This consideration is not only directly relevant to the fair use doctrine, since it impacts the assessment of the fourth factor ('impact on the potential market'), but it is also among the most serious concerns raised by generative AI. It should profoundly disturb not only authors, artists, and publishers, but also the general public.

Creative activities have always needed a viable social structure to finance them. When these social structures become dysfunctional, culture and knowledge suffer, and the creators of works have to struggle in inhumane and unproductive conditions. Finding ways around this challenge in the age of AI has become important, and copyright may or may not have the answers. Regardless, it should always be remembered that AI has already grown to become a multi-billion dollar industry, and that, whatever the social benefit of their innovations, AI companies enjoy direct, staggering financial gains. At whose expense have these gains been secured?

Accurate Attribution & the Proliferation of False Information

Intriguingly, the NYT complaint goes on to raise a second area of concern: the accurate attribution of information. The complaint points to two problems: first, that "these tools…wrongly attribute false information to The Times" and, secondly, that, "[b]y design, the training process does not preserve any copyright-management information, and the outputs of Defendants' GPT models removed any copyright notices, titles, and identifying information" from the articles.

There is no general right of attribution under United States copyright law, which only recognizes a right of attribution for artists. Attribution for artists under s. 106A of the U.S. Copyright Act is an extremely limited right, and fair uses of artworks are explicitly made exempt from attribution requirements. However, a moral right of authors to be attributed for their works is recognized outside the United States, and, in some cases, this right also enables authors to protest the false attribution of works to them. A corresponding moral right of integrity may also be invoked against the circulation of false information that affects the integrity of their creations, or their authorial reputations.

However, U.S. copyright law does prohibit the removal of digital rights management information – a practical stand-in for attribution in the technological context, as noted by the U.S. Copyright Office in its 2017 report on moral rights.

Along with a number of other points of principle raised in the complaint, this element draws attention to a broader picture: the potentially uncontrollable spread of false and unverifiable

information in the AI environment. AI can generate vast amounts of information and falsify it in new ways. At this stage, AI chatbots are even known to "hallucinate" false information – creating problems for users of the technology by generating everything from false narratives to made up citations for legal cases, with potentially dramatic practical consequences.

It is in this environment that works of New York Times journalism, like other works of human authorship, will have to compete – not only for money, but also for attention, legitimacy, and human connection. Nothing less than truth and reality are at stake.

Given the prominence of the fair use defense in this case, there is a sense that much is balanced on the edge of a knife. Despite Google Books, the courts continue to have choices. The latest Supreme Court decision on fair use is the 2023 case of *Goldsmith v. Warhol,* where the Andy Warhol Foundation was found liable for Warhol's unlicensed use of a photograph originally taken by Linda Goldsmith. Many commentators argued at the time that Warhol's image was a transformative use, yet the Court disagreed and decided that this was not a fair use. While a full discussion of the decision is beyond the scope of this post, it should be noted that the majority of the Supreme Court did not consider this use of Goldsmith's work to be "transformative", instead pointing out that it was used on a magazine cover, much as Goldsmith's own works were used, and represented direct competition with her in a clearly commercial context.

As noted by William F. Patry, decisions on fair use remain unpredictable and, to an extent not always sufficiently acknowledged, fact-specific. Above all, they arguably reflect a broader *zeitgeist* surrounding copyright. Court decisions fluctuate according to the social mood of the times.

Conclusion: Time for a new approach to copyright law?

Ultimately, these considerations may point to the unfitness of copyright law in its current form to meet the tremendous challenges posed by AI. Will this ruling manage to address the potential for social turmoil inherent in generative AI? Will it encourage the transformation of this tool into a humane and creative instrument for human expression? Or are we simply expecting too much from copyright law? Faced with such heavy demands, will the architecture of copyright prove to be more fragile or more resilient?

_____

To make sure you do not miss out on regular updates from the Kluwer Copyright Blog, please subscribe here.

Kluwer IP Law

The **2022 Future Ready Lawyer survey** showed that 79% of lawyers think that the importance of legal technology will increase for next year. With Kluwer IP Law you can navigate the

increasingly global practice of IP law with specialized, local and cross-border information and tools from every preferred location. Are you, as an IP professional, ready for the future?
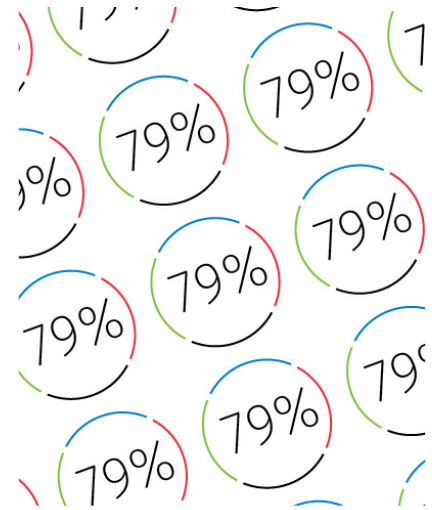
Learn how **Kluwer IP Law** can support you.

This entry was posted on Thursday, February 29th, 2024 at 8:11 am and is filed under Artificial Intelligence (AI), Authorship, Fair Use, Infringement, Text and Data Mining (TDM), USA You can follow any responses to this entry through the Comments (RSS) feed. You can leave a response, or trackback from your own site.