Generative Artificial Intelligence: Calling for a New Legal Front

By: Frank Bruno, Ilaria Maggioni, Esteban Monge-Morera, and Simone Charles
Client Alert
1.16.24

New York Times Copyright Suit and Key Facts

On December 27, 2023, The New York Times Company ("NYT") filed a Complaint in the Southern District of New York against Microsoft Corporation ("Microsoft") and several OpenAI entities ("OpenAI") for what the NYT argues is the "unlawful use of the Times's work to create artificial intelligence products that compete with [The Times and] threatens The Times's ability to provide [its] service." [1] The Complaint alleges that OpenAI's "generative artificial intelligence" ("GenAI") tool produces textual output that copies the NYT's articles, prose, and style, including pulling *verbatim* swaths of text. The Complaint further states that OpenAI relies on large-language models ("LLMs") and other large datasets to train its AI products and tools. In doing so, the Complaint alleges that Microsoft and OpenAI infringe on the NYT's right to exclude others from using its work by copying its articles and commentary word-for-word.[2]

At issue in this case is whether an AI product that relies on large data sets, is unlawfully copying copyright-protected work when the output is substantially similar to the underlying data, or whether in doing so it is merely ingesting data to transform it into something new.

The Complaint alleges that once the NYT discovered that OpenAI was using its content to train its AI models and tools, it attempted to enter negotiations "in accordance with its history of working productively with large technology platforms to permit the use of its content in new digital products (including the news products developed by Google, Meta, and Apple)."[3] The NYT's concern, among other things, is to secure fair compensation for its original content and to be a key player in the responsible development of GenAI technology and its use in the news industry.[4]

What is the Issue with GenAI?

AI is not a new concept in computing and technology industries and research programs. However, AI systems have garnered increasing media attention in the last year and a half partly due to the rapid growth of GenAI products and their capabilities. The issue has prompted President Biden to issue the Executive Order on AI on October 30, 2023, to begin the process of setting standards for the "safe and secure" development and use of AI.[5] Key to the order, among other things, is that it calls on the Director of the United States Patent and Trademark Office ("USPTO") and the United States Copyright Office ("USCO") to provide the President with a set of recommendations that address the scope of protection for works produced using AI and works used to train AI models and systems.[6]

GenAI models can now produce output like text, images, video, and audio—that emulates the identity, tone, style, and even diction of human speech—which, if produced by a human, would constitute copyrightable material. GenAI tools produce these outputs in part by "learning" statistical patterns in large data sets that often include copyrighted works.[7] However, there is considerable disagreement about whether or when the use of copyrighted material to train AI models and tools is infringing (in both generative and non-generative systems). On the one hand, some proponents argue that AI models are merely ingesting data or text to obtain inferences about the input material that the machine then uses to accomplish some other task.[8] On the other hand, other proponents maintain that when GenAI outputs are found to be substantially similar to the underlying copyrighted material used in the training dataset, one could argue copyright infringement.[9]

What is the Issue in the NYT case?

The key issue in the NYT case, according to the Complaint, is that the outputs produced by OpenAI's GenAI tools are substantially similar to those of the NYT archive of articles and commentary—and those outputs are arguably competing against the NYT as a source of news.

While OpenAI has not yet filed its Answer to the Complaint, it is anticipated that one of the possible defenses that will be asserted to support OpenAI's use of copyright protected material used to train its AI models and tools will be that they are not copying at all—arguing that the taking is simply ingesting the material so that its tool "learns" how to create something new. Additionally, OpenAI is also expected to claim a defense of "transformative fair use."

Possible Defense: Transformative Fair Use

Fair Use as a defense to copyright infringement is a statutory defense grounded in the enumerated factors explained below. "Transformative fair use" is a doctrine developed by US federal courts to interpret the first listed factor of the fair use factors and their interplay in articulating and applying the defense to particular circumstances.

The NYT cases is pending in the district court for the Southern District of New York. In the Southern District of New York, the analysis of transformative fair use[10] has traditionally turned on whether the second work is a creative original new work on its own, a necessary step in creating some new work, or rather, an unlawful use of the underlying work. When determining whether a particular use of a work is a fair use, courts must consider the statutory factors as set forth in the U.S. Copyright Act (17 U.S.C. § 107): (1) the purpose and character of the use, including whether such use is of a commercial nature or is for non-profit educational purposes; (2) the nature of the copyrighted work; (3) the amount and substantiality of the portion used in relation to the copyrighted work as a whole; and (4) the effect of the use upon the potential market for or value of the copyrighted work.

The "purpose and character" test is the most important in the analysis of transformative fair use. The analysis asks whether a new work alters the first with new "'expression, meaning, or message.' The more it does so, the more transformative the new work. And 'the more transformative the new work, the less will be the significance of other factors, like commercialism, that may weigh against a finding of fair use.'"[11] While "transformativeness" may weigh considerably in favor of fair use, it will not establish fair use on its own. The remaining factors must still be considered.[12]

What to Expect

Cases on GenAI and the issues raised by this new case in the Southern District of New York are expected to be increasingly debated in different US federal courts by tech companies and copyright holders representing diverse industries, such as photography, film, the writers' guilds among others. It is difficult at this early stage to predict courts' trends on the topics.

In this case, OpenAI is likely to file its Answer in the coming weeks. Our team is following this case closely and will provide updates as they occur.

If you have questions, please contact Frank Bruno, Partner (brunof@whiteandwilliams.com; 215.864.6225); Ilaria Maggioni, Counsel (maggionii@whiteandwilliams.com; 917.747.3693), Esteban Monge-Morera, Associate (monge-morerae@whiteandwilliams.com; 646.766.1350), Simone Charles, Associate (charless@whiteandwilliams.com; 646.837.5772) or another member of the Intellectual Property Group.

[1] Complaint at 2, ¶2, *The New York Times v. Microsoft Corp.,* 1:23-cv-11195 (filed Dec. 27, 2023). Note that Microsoft is OpenAI's largest investor according to OpenAI's website. *See* OpenAI, Our Structure, *https://openai.com/our-structure.*

[2] *See* Complaint at 2, ¶2.

[3] Complaint at 3, ¶7.

[4] *Id.*

[5] Exec. Order No. 14110, 3 C.F.R. § 2(a)(2023).

[6] *Id.* at § 5.2(c)(iii) (providing recommendations are due within 270 days of the order [e.g., July 26, 2024] or 180 days after the Copyright Office publishes its forthcoming AI study, whichever comes later).

[7] U.S. Copyright Office, Artificial Intelligence and Copyright: Supplementary Information, at 16 (Aug. 30, 2023), Artificial Intelligence and Copyright; *see also* Kim Martineau, What is generative AI?, IBM Research Blog (Apr. 20, 2023), https://research.ibm.com/blog/what-is-generative-AI ("At a high level, generative models encode a simplified representation of their training data and draw from it to create a new work that's similar, but not identical, to the original data.").

[8] *See* U.S. Copyright Office, at 17 (defining "machine learning" as a "technique for building AI systems that is characterized by the ability to automatically learn and improve on the basis of data or experience, without relying on explicitly programmed rules. Machine learning involves ingesting and analyzing materials such as quantitative data or text and obtain inferences about qualities of those materials and using those inferences to accomplish a specific task."

[9] *See, e.g.,* Silverman v. OpenAI, Inc., 4:23-cv-3416 (N.D. Cal.); Tremblay v. OpenAI, Inc., 3:23-cv-3223 (N.D. Cal.); Getty Images (US), Inc. V. Stability AI, Inc., 1:23-cv-0135 (D. Del.).

[10] *See, e.g.,* Bill Graham Archives v. Dorling Kindersley Ltd., 448 F.3d 605 (2d Cir. 2006); Authors Guild v. HathiTrust, 755 F.3d 87, 90 (2d Cir. 2014); Andy Warhol Found. for the Visual Arts, Inc. v. Goldsmith, 143 S. Ct. 1258 (2023).

[11] *See* Goldsmith, 143 S. Ct. at 1299 (quoting Campbell v. Acuff-Rose Music, Inc., 510 U. S. 569, 579 (1994)).

[12] *Id.*