UNITED STATES DISTRICT COURT
SOUTHERN DISTRICT OF NEW YORK

THE NEW YORK TIMES COMPANY, Plaintiff, -against- MICROSOFT CORPORATION, OPENAI, INC., et al., Defendants.	:	23-cv-11195 (SHS) (OTW) **OPINION & ORDER**

ONA T. WANG, United States Magistrate Judge:

I. BACKGROUND

The New York Times (the "Times" or "Plaintiff") brought this action alleging, *inter alia*, that Defendants unlawfully used Plaintiff's copyrighted works to train Defendants' large-language models ("LLMs"). Defendant OpenAI, Inc. ("Defendant") seeks to compel[1] production of: (1) the Times's use of nonparties' generative artificial intelligence ("Gen AI") tools; (2) the Times's creation and use of its own Gen AI products; and (3) the Times's position regarding Gen AI (e.g., positions expressed outside of litigation, knowledge about the training of third-party Gen AI tools using the Time's works). (ECF 236). Defendant asserts that this outstanding discovery is relevant to their fair use defense. (ECF 236). Plaintiff asserts that the disputed discovery concerning Plaintiff's interactions with their own and nonparties' Gen AI tools are

[1] Plaintiff has already provided or agreed to produce: (1) documents regarding the Times's use of the Defendants' Gen AI tools in reporting or presentation of content, and documents regarding the Times's trainings about Defendants' Gen AI products; (2) documents relating to the Times's A.I. Initiatives program; and (3) nonprivileged documents and communications with third parties about the Defendants' use of Times content in their Gen AI products and this litigation and whether to license Times works to OpenAI. (ECF 238).

neither relevant nor proportional to the needs of the case. (ECF 238). Because Defendant has not demonstrated the relevance of the information sought, their motion to compel is **DENIED**.

II. LEGAL STANDARD

Federal Rule of Civil Procedure 26(b)(1) permits discovery of "any nonprivileged matter that is relevant to any party's claim or defense and proportional to the needs of the case." The party moving to compel, here OpenAI, "bears the initial burden of demonstrating relevance and proportionality." *See Winfield v. City of New York*, No. 15-CV-5236 (LTS) (KHP), 2018 WL 840085, at *3 (S.D.N.Y. Feb. 12, 2018). "Motions to compel and motions to quash a subpoena are both entrusted to the sound discretion of the court." *Howard v. City of New York*, No. 12-CV-933 (JMF), 2013 WL 174210, at *1 (S.D.N.Y. Jan. 16, 2013).

III. DISCUSSION

The Copyright Act (the "Act") allows for certain "fair" uses of copyrighted works and sets out four non-exclusive factors for courts to consider in determining whether a particular use is "fair":

> I. the purpose and character of the use, including whether such use is of a commercial nature or is for nonprofit educational purposes;
> II. the nature of the copyrighted work;
> III. the amount and substantiality of the portion used in relation to the copyrighted work as a whole; and
> IV. the effect of the use upon the potential market for or value of the copyrighted work.

Hachette Book Group, Inc. v. Internet Archive, 115 F.4th 163, 178-79 (2d Cir. 2024). Each of these factors requires scrutiny of a defendant's purported use of the copyrighted work(s), and whether that defendant's use may constitute "fair use" under the Act. The factors do not require a court to examine statements or comments a copyright holder may have made about a

defendant's general industry, whether the copyright holder has used tools in the defendant's general industry, whether the copyright holder has admitted that other uses of its copyrights may or may not constitute fair use, or whether the copyright holder has entered into business relationships with other entities in the defendant's industry.

Defendant argues that the discovery they seek is relevant to "the Times's own claim that the mere existence of this technology is a threat to its business model and the enterprise of journalism." (*See* ECF 236, at 2). However, the "statement" referenced by Defendant is not a claim or defense; it is a heading in the Amended Complaint: "GenAI Products Threaten High-Quality Journalism," which precedes paragraphs 47 through 54. (ECF 170, at 14). This section discusses the Times's protection of its own journalistic content, the limited content available to search engines, and prior discussions with Defendant to "explore the possibility of an amicable resolution," which apparently were unsuccessful. (ECF 170 ¶ 54). There is no wholesale indictment of Gen AI tools, nor is there any suggestion that the Times allows third parties unfettered, unpaid access to its copyrighted journalistic content.[2] The AC is tightly focused on Defendant's particular Gen AI products and their alleged use of the Times's copyrighted content.

None of the cases cited by Defendant support the assertion that the discovery sought is relevant to their fair use defense or to the heading in the Amended Complaint. For example,

[2] Nor is any broader discovery warranted based on Defendant's speculative and conclusory assertion that "if the Times knew about multiple third parties using the Times's works to train generative AI tools but did nothing, that would suggest recognition by the Times of the reasons that such training is protected by fair use – e.g. that no workable market exists for licensing the volume of data required; that it offers significant public benefits; and that it stands to achieve purposes distinct from that of its underlying works." [sic]. (ECF 236 at 3) (emphasis added). Moreover, the Times is already producing documents about its knowledge and awareness of Defendant's training. *See supra*, n. 1.

Google v. Oracle does not support a modification of the fourth fair use factor to include discovery about Plaintiff's views on or statements about the "public benefits" of Gen AI in journalism. 593 U.S. 1, 35-36. Rather, the Supreme Court suggested a more nuanced view of the market effects, one that requires consideration of the importance of the "public benefits the copying will likely produce" to "copyright's concern for the creative production of new expression" and a balancing against the potential loss to the copyright owner, "taking into account … the nature of the source of the loss." *Id.* at 35-36 (internal quotations omitted). Discovery regarding the loss to the copyright owner would consist of documents concerning licensing discussions, which the Times has already agreed to produce, (*see, supra* n. 1), and discovery from Defendant on how its use might "kill demand for the original." *C.f. Oracle*, 593 U.S. at 35 ("But a potential loss of revenue is not the whole story. We here must consider not just the amount but also the source of the loss. As we pointed out in *Campbell*, a lethal parody, like a scathing theatre review, may kill demand for the original… Yet this kind of harm, even if directly translated into foregone dollars, is not cognizable under the Copyright Act.") (internal quotations omitted). Similarly, discovery concerning the "public benefits [from] the copying" would be directed to the Defendant and the public benefits of its copying, not whether nonparties' Gen AI tools (which presumably were developed without copying) serve a general public benefit.

The Second Circuit took the same approach in *Am. Geophysical Un. v. Texaco Inc.*, focusing on how Texaco's copying, and its use of those copies, met (or did not meet) the fair use factors. 60 F.3d 913, 927 (2d Cir. 1995) ("Since we are concerned with the claim of fair use in copying the eight individual articles from [the journal] *Catalysis*, the analysis under the fourth

factor must focus on the effect of Texaco's photocopying upon the potential market for or value of these individual articles."). The copyright holder's other use or licensing of their own works to other nonparties was simply not at issue in the fair use determination, and *Google* and *Texaco* do not support a finding of relevance here for the same.

Similarly, *Andy Warhol Foundation for the Visual Arts, Inc. v. Goldsmith* does not stand for the proposition that "the Times's creation, use and positions on [others' Gen AI] generally is directly relevant" to Defendant's fair use defense. (ECF 236 at 1) ("[T]he technology yields transformative and productive benefits for the enterprise of journalism specifically.") (emphasis added). Whether nonparties' Gen AI tools confer benefits on the journalism industry is not relevant to a determination of whether Defendant's acts—i.e., the alleged copying involving Defendant's Gen AI tools—constitute fair use.[3] The fair use factors are concerned with "the copier's use of an original work." *See Andy Warhol Foundation for the Visual Arts, Inc. v. Goldsmith*, 598 U.S. 508, 528 (2023).

IV. CONCLUSION

This case is about whether Defendant trained their LLMs using Plaintiff's copyrighted material, and whether that use constitutes copyright infringement. (ECF 170, ¶¶ 158-168). It is not a referendum on the benefits of Gen AI, on Plaintiff's business practices, or about whether any of Plaintiff's employees use Gen AI at work. The broad scope of document production sought here is simply not relevant to Defendant's purported fair use defense. For example, if a

[3] OpenAI seems to suggest that if the Times's journalists use any form of Gen AI tools in their work, that Gen AI then "benefits" journalism, and if Gen AI tools "benefit" journalism, that "benefit" would be relevant to OpenAI's fair use defense. But the Supreme Court specifically states that a discussion of "public benefits" must relate to the benefits from the copying. *Oracle*, 593 U.S. at 35.

copyright holder sued a video game manufacturer for copyright infringement, the copyright holder might be required to produce documents relating to their interactions with that video game manufacturer, but the video game manufacturer would not be entitled to wide-ranging discovery concerning the copyright holder's employees' gaming history, statements about video games generally, or even their licensing of different content to other video game manufacturers.

Accordingly, because Defendant has failed to demonstrate the relevance of the information sought, Defendant's motion to compel is **DENIED**.

The Clerk of Court is respectfully directed to close ECF 236.

SO ORDERED.

s/ Ona T. Wang

Dated: November 22, 2024
New York, New York

Ona T. Wang
United States Magistrate Judge