

UNITED STATES DISTRICT COURT
SOUTHERN DISTRICT OF NEW YORK

THE NEW YORK TIMES COMPANY,

Plaintiffs,

v.

MICROSOFT CORPORATION, OPENAI, INC.,
OPENAI LP, OPENAI GP, LLC, OPENAI, LLC,
OPENAI OPCO LLC, OPENAI GLOBAL LLC,
OAI CORPORATION, LLC, and OPENAI
HOLDINGS, LLC,

Defendants.

Case No. 1:23-cv-11195-SHS

**MICROSOFT’S CONDITIONAL
OPPOSITION TO THE NEW
YORK TIMES COMPANY’S
MOTION FOR LEAVE TO FILE
FIRST AMENDED COMPLAINT**

Under the current case schedule, substantial completion of document production will occur in a few days (June 14) and fact discovery will close on September 17. This case is presently on the schedule that Microsoft and the OpenAI Defendants stipulated to in *Authors Guild et al. v. OpenAI, Inc. et al.* (No. 1:23-cv-08292), *Alter et al. v. OpenAI, Inc. et al.* (No. 1:23-cv-10211), and *Basbanes et al. v. OpenAI, Inc. et al.* (No. 1:23-cv-00084) (the “Consolidated Class Actions”), but objected to in this case. Even before this proposed amendment, The Times’ claims were significantly broader than those in the Consolidated Class Actions. With the further expansion of the case proposed in The Times’ amendment, it will be impossible to complete fact and expert discovery under the current case schedule. But an efficient solution to this issue exists: aligning the schedule of this case to whatever schedule the Court sets in the New York Daily News case, as the two matters are nearly identical.

Microsoft does not dispute that The Times has generally met the liberal standard for amendment pursuant to Federal Rule of Civil Procedure 15, and that it made its request for leave to amend in accordance with the Scheduling Order governing this case. As Microsoft

communicated to The Times during the meet and confer process, Microsoft is not specifically opposed to The Times’ First Amended Complaint (“FAC”) but requests that any leave to amend is accompanied by an adjustment of the case schedule to account for the fact that the proposed amendment will more than triple the burden associated with data analysis in this case by increasing the number of works at issue from 3,000,000 works to 10,512,015 works. The schedule can be modified for maximum efficiency and judicial economy by matching it to whatever schedule the Court sets in the existing case before this Court brought by the New York Daily News and seven other newspaper publishers. *Daily News LP et al. v. Microsoft Corp. et al.* ECF No. 1 (1:24-cv-03285-SHS, filed Apr. 30, 2024). The *Daily News* case features the same expansive counts, technology, and products that are at issue here and not at issue in the Consolidated Class Actions. In fact, the Court has already related the *Daily News* case and this case on the public docket. Given the overlapping scope of this matter and the *Daily News* matter, it would make much more sense to align The Times’ case with the other newspaper plaintiff cases so that they all can – and should – efficiently proceed together on a later schedule. The Consolidated Class Actions can then proceed on their current schedule without loss of judicial efficiency because the newspaper plaintiff cases contain broader and materially different claims.

Permitting The Times to file its FAC—which “add[s] approximately 7 million additional works to the suit”—without any adjustment to the case schedule would severely prejudice Microsoft’s ability to fully and fairly defend itself in this matter. ECF No. 118-1 at 1. Even without the proposed amendment, managing The Times’ case on the schedule of the Consolidated Class Actions is extraordinarily challenging. Unlike the class cases, this case presently involves: (1) additional and distinct causes of action, directed towards not only the data purportedly used to train OpenAI’s large language models (“LLMs”) but also the resultant outputs, and (2) significant

additional technology (*e.g.*, generative search) utilized in Microsoft Generative AI product offerings that are not at issue in the Consolidated Class Actions. The additional technology at issue means that the parties will have to take discovery not just about using copyrighted material to train large language models, but also will need to prepare a case that addresses the interaction of two *distinct* technologies—generative AI models *and* search engines. Discovery requests served by The Times lay bare the amount of work that these claims require, over and above those in the Consolidated Class Actions, including production and analysis of massive amounts of data that reflect the interaction of these two technologies that produce search outputs. For each alleged copyrighted work, this data must be analyzed to ascertain whether any part of it is mentioned in response to user prompts directed to the Bing Copilot generative search technology. Now The Times wants to bring in over 7,000,000 additional works authored over a period of nearly 100 years to that analytical project. Simply put: the scope of The Times’ case is already orders of magnitude larger than the Consolidated Class Actions, even before The Times’ proposal to add more than 7,000,000 *additional* works through its FAC. In addition to unduly prejudicing Microsoft, there are real technological limitations that may prevent compliance with the current schedule as is, much less as proposed.

Microsoft therefore respectfully requests that if the Court grants The Times’ Motion, the Court also adjust the current discovery schedule and align it with whatever schedule gets set in the *Daily News* case in order to ameliorate significant and undue prejudice to Microsoft’s preparation and defense of this case.

I. FACTUAL BACKGROUND

The Times filed this case on December 27, 2023, and pled seven counts against Microsoft, including *inter alia*: direct, vicarious, and contributory copyright infringement, as well as

violations of the Digital Millenium Copyright Act, trademark dilution, and common law unfair competition. ECF No. 1. In its initial Complaint, The Times’ infringement allegations describe how “the training datasets contain[] millions of copies of Times Works to *train* the GPT models on Microsoft’s supercomputing platform[.]” *Id.*, ¶ 161 (emphasis added).¹ But The Times’ Complaint also contains accusations against Microsoft concerning its supposed “*disseminati[on]*” of “generative *output[s]* containing copies and derivatives of Times Works through the Bing Chat offerings[.]” *Id.*, ¶ 164 (emphasis added).

The parties held a Rule 26(f) Conference to establish a schedule—and associated discovery parameters to address The Times’ allegations—and submitted a report to the Court on March 8, 2024, in which the parties’ provided competing scheduling proposals and justifications for their positions. ECF No. 72. The parties did not agree that this case should be set on the same schedule as the class actions. As explained in the Rule 26(f) Report, Microsoft and OpenAI stipulated to an agreed class schedule in the Consolidated Class Actions “as part of a broader case administration agreement (which included, *inter alia*, stipulations as to the sequencing of class certification and summary judgment).” *Id.*, ¶ 40. Microsoft and OpenAI agreed upon that course of action because “the Consolidated Class Actions”—as opposed to The Times’ case—“are narrowly focused on the use of works during the process for *training* [OpenAI’s] Gen[erative AI] models.” *Id.*, ¶ 41 (emphasis in original). By contrast, The Times’ case concerns Microsoft and OpenAI’s Generative AI products’ generation of *outputs* in response to user prompts, requiring analysis of a large pool of user data to assess whether there were in fact any infringing outputs. Additionally, The Times’

¹ Until The Times filed this Motion, it was not readily apparent that The Times was alleging that Microsoft’s purported infringement spanned such a massive number of works, as opposed to The Times simply maintaining a portfolio of “over 3 million registered, copyrighted works[.]” ECF No. 1, ¶ 14. The Times now seeks to explicitly multiply this case more than three times in the number of works. ECF No. 118-1 at 1.

allegations concern not just generative AI outputs in general, but the outputs of the new Bing Copilot search engine. This product includes generative search technology that goes far beyond the question of training large language models, or even their outputs. Rather, such technology implicates the interaction between generative AI and search engines and requires not only an understanding of both technologies but also the detailed interaction between them. *See generally*, ECF No. 1; *see also* ECF No. 59 at 16 (The Times, in opposition to a non-party’s motion to intervene, describing how the instant case involves more than training and includes “additional forms of direct copyright infringement, including based on how [Microsoft’s and OpenAI’s] artificial intelligence products generate infringing output in response to user prompts.”).

Since the Rule 26(f) report was filed, several key developments have occurred. On April 30th, the Daily News filed a similar suit against Microsoft and OpenAI. ECF No. 1 (1:24-cv-03285-SHS). The Daily News’ claims overlap significantly with The Times’ case (as opposed to the Consolidated Class Actions) and likewise pertain to outputs and corresponding generative technology—such as generative search functionality. *See e.g., id.*, ¶¶ 69, 78, 212.

On May 3rd, the Court entered a Scheduling Order in this case that largely adopted the schedule governing the Consolidated Class Actions. ECF No. 112. Shortly thereafter, on May 16th, The Times reached out to request Microsoft and OpenAI’s consent to file the FAC. Ex. A (Email from Z. Savage (dated 05.16.24)). The Times represented that “the sole change would be to add additional works” and that “[t]he amendment would not change any claims or theories that The Times is asserting[.]” *Id.* When Microsoft sought to clarify The Times’ request, The Times informed Microsoft that it expected to “add on the order of 7 million works” to the FAC, “going back to 1926.” Ex. B at 2 (Email from I. Crosby to A. Hurst (dated 05.20.24)). The Times would

not clarify whether it was claiming that all such works were the subject of output-based claims. *See id.* The instant Motion followed. ECF No. 118.

II. ARGUMENT

In accordance with “Rule 15(a)(2) of the Federal Rules of Civil Procedure, ‘[a] court should freely give leave [to amend] when justice so requires.’” *Holmes v. Grubman*, 568 F.3d 329, 334 (2d Cir. 2009) (quoting *McCarthy v. Dun & Bradstreet Corp.*, 482 F.3d 184, 200-01 (2d Cir. 2007)). However, there is a limit to this principle: the “[C]ourt has discretion to deny leave for good reason, including futility, bad faith, undue delay, or undue prejudice to the opposing party.” *McCarthy*, 482 F.3d at 200.

It is the significant risk of undue prejudice that is implicated here. As explained above, Microsoft agrees that – in a vacuum – The Times is entitled to amend its Complaint; however, without a schedule modification to address the numerous additional works, and the implications of the massive addition to the already distinct and expansive claims and corresponding technologies implicated in this case (as compared to the Consolidated Class Actions), the proposed amendment will unduly and significantly prejudice Microsoft, preventing it from fully and fairly defending itself in this case.

The Times’ case was fundamentally different from the Consolidated Class Actions from the outset. As Microsoft pointed out in the Rule 26(f) Report, the Consolidated Class Actions’ copyright claims focus exclusively on infringement of plaintiffs’ works based on their purported inclusion in the data sets used to train the LLM models. By contrast, The Times’ case, just like the newly-filed *Daily News* case, also involves copyrighted works allegedly present in the resultant generative outputs, which implicates significant additional technology, as well as Microsoft’s and OpenAI’s commercial offerings, which are not at issue in the Consolidated Class Actions. As

described in The Times' Complaint, the products potentially at issue here include: Microsoft's Azure AI platform, Microsoft's Bing Copilot (formerly Bing Chat), and the ChatGPT Browse with Bing plug-in. ECF No. 1 ¶¶ 102, 118; *see also* ECF. No. 118-3 ¶¶ 102, 118 (same).

Discovery exchanged to date spotlights the larger scope of The Times' case compared to the Consolidated Class Actions. For example, The Times has served discovery on Microsoft specifically about its generative search technology (not at issue in the Consolidated Class Actions), including a request to inspect the same. *See, e.g.*, Ex. C at 8 (RFP No. 11, seeking "[d]ocuments concerning Defendants' use of ... [generative search] in Generative AI Products and Services... including "how the results from [generative search] search products are similar to or different from standard search results..."); *id.* (RFP No. 12, seeking "[d]ocuments concerning the features or capabilities of [generative search]."); *see also* Ex. D at 2 (Req. for Inspection ¶ 5, seeking "Copilot branded Generative AI Products and Services, including APIs, plug-ins, Prometheus and other features offered through the Copilot branded platform, including software components available in source code form and the associated source code repositories and documentation.").

Further, The Times has served numerous requests for production on Microsoft that seek to inspect large collections of data regarding user outputs and post-training activity. *See, e.g.*, Ex. C at 6 (RFP No. 1, seeking "[d]ocuments concerning the use of Times Content for ... every Text Generation AI Model ... for any post-training activity[.]"); *id.* (RFP No. 3 seeking, "[d]ocuments concerning copyrighted material in the Text Generation AI Models...or any Generative AI Products and Services[.]"); *see also id.*, at 7 (RFP No. 8 seeking, "[d]ocuments sufficient to show to what extent and in what form Times Content resides in, is stored in, is copied by, or otherwise exists within the ... the Text Generation AI Models."). The Times' interrogatory requests to Microsoft are to the same effect. *See, e.g.*, Ex. E at 5 (ROG No. 4, requesting Microsoft "[i]dentify

all individuals previously or presently employed by Defendants with knowledge about ... Your efforts to limit ... *outputs* containing copyrighted content.”) (emphasis added); *id.*, at 5-6 (ROG No. 7, requesting Microsoft “[i]dentify all Text Generation AI Models, including versions and release dates, and for each model, specify the existence, custodian, location, and general description of technical documentation and documentation...related to that model, and whether that model was ... otherwise affected by Times content.”); *see also* Ex. D at 2 (Req. for Inspection ¶ 6, seeking “[q]uery, session, and chat logs related to Times Content reflecting user sessions on Your Generative AI Products and Services and Bing (‘Session Data’), including, for each session, user queries paired with responses to those queries.”).

In the context of claims directed at an entirely distinct search-based technology at issue in this case and the *Daily News* case, adding millions more copyrighted works into the mix will grind the gears of case management to a halt. In these newspaper cases (as opposed to the Consolidated Class Action), for each alleged copyrighted work at issue, Microsoft has to query user logs in order to respond to fact discovery requests, as described *supra*, as well as assess The Times’ infringement allegations. For 10,512,015 works the effort is extremely time and resource intensive, and practically impossible to complete by the current close of fact discovery. It takes weeks or months to construct, test, and run queries regarding such data logs. This cannot be solved by throwing more resources at the problem, since there are technological limits to the capability of the systems hosting the data. Moreover, once the results are known, they must be analyzed by humans in order to assess the user intents as well as the nature and context of any matching snippets. Additionally, the outputs must be assessed as to source—whether they are the result of a straightforward search engine query or something else. The compressed schedule that was designed solely to address the training fair use issues simply is not suitable to complete such an

intensive analysis of such a large pool of data. It is extremely prejudicial to Microsoft's ability to fully and fairly defend itself, as it is likely not even feasible to conduct such an analysis in a timely fashion.

Further, Microsoft must expend substantial resources to conduct meaningful sampling and statistical expert analyses of *all* outputs—not just any matching ones—in order to address the broader context of consumer usage of generative search technology. Even if the generative search results occasionally spit out a portion of a work authored by The Times, and even if that were an infringement by an end user who acted with likely entirely legitimate purposes, Microsoft is further entitled to prove that the drops in the bucket of matching outputs are dwarfed by all of the other noninfringing uses of the technology. *Sony Corp. of Am. v. Universal City Studios, Inc.*, 464 U.S. 417, 440-42 (1984).

Finally, Defendants must devise some method of assessing the requisites of ownership of the more than 10 million works at issue, many of which are close to 100 years old. By increasing the total to *three times* as many works for the project of assessing ownership, there can be no doubt that The Times has expanded the scope of the case to a level that simply cannot be conducted on the current schedule.

All of this will present undue prejudice to Microsoft's ability to defend itself if the case schedule is not amended or otherwise extended. The Times argues that “[t]he proposed amendment will not affect the schedule that the Court has set for hearing the threshold fair use question because the answer to that question does not turn on the number of works asserted.” ECF No. 118-1 at 1. This statement would be correct if The Times' claims were limited to alleged copyright infringement by LLM training, where fair use arguments are identical to the Consolidated Class Action. But as shown above, this is not the case. And the Court's Scheduling

Order is not limited to the issue of fair use by training; indeed, the Scheduling Order does not even mention or use that term. *See* ECF No. 112. Instead, the current case schedule contemplates the completion of fact and expert discovery, as well as dispositive motion briefing, on all issues across all cases. This is simply impracticable given the scope of this case which requires assessing both the input and the outputs from Microsoft’s Generative AI products in a universe dramatically expanded by over 7,000,000 new works.

* * *

For all these reasons—as a matter of fundamental fairness and to prevent any prejudice—Microsoft respectfully requests that if the Court grants The Times’ Motion, the Court also permits an extension of this case’s discovery schedule to coordinate it with the *Daily News* matter.²

Dated: June 3, 2024

By: /s/ Annette L. Hurst

ORRICK, HERRINGTON & SUTCLIFFE LLP
Annette L. Hurst (admitted *pro hac vice*)
The Orrick Building
405 Howard Street
San Francisco, CA 94105-2669
Telephone: (415) 773-5700
Facsimile: (415) 773-5759
ahurst@orrick.com

Christopher J. Cariello
51 West 52nd Street
New York, NY 10019
Telephone: (212) 506-3778
Facsimile: (212) 506-5151
ccariello@orrick.com

Sheryl Koval Garko

² In its Motion, The Times represents that: “[b]ecause motions to dismiss are pending with regards to certain of those causes of action, The Times proposes that it should be required to file the FAC within twenty-one days after a ruling on those motions.” ECF No. 118-1 at 2. Microsoft is in agreement with The Times’ proposal on this issue.

Laura Brooks Najemy
222 Berkley Street
Boston, MA 02116
Telephone: (617) 880-1800
Facsimile: (617) 880-1801
sgarko@orrick.com
lnajemy@orrick.com

FAEGRE DRINKER BIDDLE & REATH LLP
Jeffrey S. Jacobson
1177 Avenue of the Americas
New York, New York 10036
Telephone: (212) 248-3191
jeffrey.jacobson@faegredrinker.com

Jared B. Briant (admitted *pro hac vice*)
1144 15th Street, Suite 3400
Denver, Colorado 80202
Telephone: (303) 607-3588
jared.briant@faegredrinker.com

*Attorneys for Defendant
Microsoft Corporation*