

**UNITED STATES DISTRICT COURT
SOUTHERN DISTRICT OF NEW YORK**

THE NEW YORK TIMES COMPANY,

Plaintiff,

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.

Civil Action No. 1:23-cv-11195-SHS

Hon. Sidney H. Stein

DAILY NEWS, LP; CHICAGO TRIBUNE
COMPANY, LLC; ORLANDO SENTINEL
COMMUNICATIONS COMPANY, LLC; SUN-
SENTINEL COMPANY, LLC; SAN JOSE
MERCURY-NEWS, LLC; DP MEDIA
NETWORK, LLC; ORB PUBLISHING, LLC;
AND NORTHWEST PUBLICATIONS, LLC,

Plaintiffs,

v.

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

Defendants.

Civil Action No. 1:24-cv-03285-SHS

Hon. Sidney H. Stein

**MICROSOFT’S JOINDER BRIEF IN SUPPORT OF
THE OPENAI DEFENDANTS’ MOTION TO CONSOLIDATE**

Pursuant to Federal Rule of Civil Procedure 42, Microsoft submits this supplementary brief to join in the request for relief by the OpenAI Defendants in their Motion to Consolidate The New York Times Company (“The Times”) case with a related case—*Daily News LP, et al. v. Microsoft*

Corp., et al. ECF No. 1 (1:24-cv-03285-SHS, filed Apr. 30, 2024)—for discovery purposes and pre-trial motion practice. ECF. No. 142. Specifically, in the interests of fundamental fairness and judicial economy, Microsoft joins in the OpenAI Defendants’ request to consolidate *The Times* case with the *Daily News* case, but only so long as these matters proceed on a separate, and later schedule than that which the parties’ 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”).¹

In addition to the reasons for consolidation presented by the OpenAI Defendants, *The Times* case and the *Daily News* case should be consolidated because both cases involve nearly identical allegations relating to the **same new technology**—a technology that is **not at issue** in the Consolidated Class Actions. Specifically, with respect to Microsoft, both *The Times* and the *Daily News* cases (the “Newspaper Cases”) accuse Microsoft’s innovative Copilot Chat (formerly Bing Chat) generative search technology. Referred to in the Newspaper Cases complaints as “retrieval augmented generation,” Copilot Chat is a combination of both large language models and the Bing search engine, and is employed in several distinct products. Search engines are not at issue at all in the Consolidated Class Action. Moreover, the Newspaper Cases accuse the **outputs** of the generative search technology – another theory that is not at issue at all in the Consolidated Class Actions. This separate new technology underpins each of plaintiffs’ counts in the Newspaper Cases, as opposed to the Consolidated Class Actions where only the training—or **inputs to** rather than **outputs from**—large language models (“LLM(s)”) is at issue. And, as described and shown below, on a count-by-count basis, The Times’ and the Daily News’ factual and legal allegations

¹ Microsoft previewed this position for the Court in its Conditional Opposition to The New York Times Company’s Motion for Leave to File First Amended Complaint. *See* ECF No. 129, at 6-10.

concerning Microsoft's generative search technology are virtually identical.

Lastly, and as explained in Microsoft's Conditional Opposition to The Times' Motion for Leave to File its First Amended Complaint (ECF No. 129), the different technologies and concordant counts at issue between this case and the *Daily News*, as compared to the Consolidated Class Actions, necessitate that any consolidation between both cases must also involve setting a new schedule that follows behind that of the Consolidated Class Actions. This is so because the different technology at issue in the Newspaper Cases implicates additional, broader fact and expert discovery work that simply cannot be completed on the timeline presently set in the Consolidated Class Actions. ECF No. 112 (Scheduling Order). Additionally, given the early stage of the *Daily News* case, where responses to the Complaint have just been filed (ECF Nos. 76 & 80), the defendants will need time to conduct necessary discovery specific to *eight* new newspaper plaintiffs. Discovery commenced in the *Daily News* case on June 14, 2024.

Accordingly, Microsoft respectfully joins in the OpenAI Defendants' request that the Court consolidate both *The Times* and the *Daily News* cases and set a case schedule for the consolidated newspaper cases with at least six months of fact discovery.

I. The Factual Allegations and Legal Claims Concerning Microsoft's Newly Accused Generative Search Technology Are Virtually Identical in the Newspaper Cases.

Federal Rule of Civil Procedure 42 permits consolidation where "actions before the court involve a common question of law or fact[.]" FED. R. CIV. P. 42(a). In the event this is so, as is the case here, "the court may: ... consolidate the actions; or issue any other orders to avoid unnecessary cost or delay." *Id.*, 42(a)(2)-(3). In addition to the reasons for consolidation set forth in the OpenAI Defendants' Motion, consolidation is warranted here as both *The Times* and the *Daily News* cases concern overlapping (if not completely identical) allegations about Microsoft (and the OpenAI Defendants') generative search technology. Specifically, with respect to

Microsoft, these cases concern Microsoft’s generative AI products (such as Bing Copilot) that consist of underlying generative AI models as well as, and in combination with, innovative search engine tools that allow a user to enter any prompt. Microsoft’s generative search technology is unique to the Newspaper Cases and does not feature in the Consolidated Class Actions.

For example, both *The Times* and the *Daily News* allege that Microsoft “[i]n collaboration with OpenAI” “commercialized OpenAI’s GPT-based technology, and combined it with its own Bing search index” to formulate the newly accused generative search technology. *The Times* Compl.,² ¶ 72; *Daily News* Compl.,³ ¶ 69 (describing same). According to both sets of plaintiffs, the result of this collaboration is a generative search product “with the ability to generate natural language summaries of search result contents” that allegedly includes the copyrighted works at issue spanning both cases. *The Times* Compl., ¶ 72 (further describing how “search results purport to answer user queries directly and may include extensive paraphrases and direct quotes”); *Daily News* Compl., ¶ 69 (same). In addition, *The Times* and *Daily News* both describe how “search responses include links to source materials” which allegedly entails that “users have less need to navigate to those sources because their expressive content is already included in the narrative result.” *The Times* Compl., ¶ 109; *Daily News* Compl., ¶ 115 (same).

The Times and *Daily News* further elaborate on how this newly accused generative search technology works in their complimentary allegations:

<i>The Times</i> Complaint (¶ 81)	<i>Daily News</i> Complaint (¶ 78)
Once trained, [large language models] [“]LLMs[“] may be provided with information specific to a use case or subject matter in order to “ground” their outputs. For example, an LLM may be asked to generate a text output based on specific external data, such as a	Once trained, [large language models] [“]LLMs[“] may be provided with information specific to a use case or subject matter in order to “ground” their outputs through retrieval augmented generation. For example, an LLM may be asked to generate a text output based

² ECF No. 1, 1:23-cv-11195-SHS.

³ ECF No. 1, 1:24-cv-03285-SHS.

<p>document, provided as context. Using this method, Defendants’ synthetic search applications: (1) receive an input, such as a question; (2) retrieve relevant documents related to the input prior to generating a response; (3) combine the original input with the retrieved documents in order to provide context; and (4) provide the combined data to an LLM, which generates a natural-language response.¹¹ As shown below, search results generated in this way may extensively copy or closely paraphrase works that the models themselves may not have memorized.</p>	<p>on specific external data, such as a document, provided as context. Using this method, Defendants’ synthetic search applications: (1) receive an input, such as a question; (2) retrieve relevant documents related to the input prior to generating a response; (3) combine the original input with the retrieved documents in order to provide context; and (4) provide the combined data to an LLM, which generates a natural-language response. As shown below, search results generated in this way may extensively copy or closely paraphrase works that the models themselves may not have memorized.</p>
---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

Beyond these examples, *The Times*’ and *Daily News*’ allegations overlap entirely with respect to this newly accused technology, including technological and legal issues surrounding: generative (search) outputs and the particular relationship between generative AI and corresponding generative search engine technology. *See generally The Times Compl.; Daily News Compl.*

Indeed, *The Times*’ and *Daily News*’ shared technological focus, distinct from the Consolidated Class Actions, is reflected in multiple directly corresponding counts in their Complaints. As set forth in the chart below, and further illustrated in the attached Exhibit A (comparing the allegations in the Complaints, paragraph by paragraph), the newspaper plaintiffs’ claims and allegations are substantively (and in many cases textually) identical:

<i>The Times</i> Complaint	<i>The Daily News</i> Complaint
<p>Count I: Copyright Infringement (17 U.S.C. § 501) – Alleging direct infringement by Microsoft and OpenAI resulting from <u>“disseminating generative output[s]”</u> containing copies and derivatives of Times Works. <i>The Times</i> Compl. ¶¶ 158-168</p>	<p>Count I: Copyright Infringement (17 U.S.C. § 501) – Alleging direct infringement by Microsoft and OpenAI resulting from <u>“disseminating generative output[s]”</u> containing copies and derivatives of Times Works. <i>Daily News</i> Compl. ¶¶ 190-201</p>
<p>Count IV: Contributory Copyright Infringement – Alleging Microsoft materially contributed to end-users’ infringement by co-developing LLM models capable of <u>“distributing unlicensed copies of Times</u></p>	<p>Count IV: Contributory Copyright Infringement – Alleging Microsoft materially contributed to end-users’ infringement by co-developing LLM models capable of <u>“distributing unlicensed copies of</u></p>

<i>The Times Complaint</i>	<i>The Daily News Complaint</i>
<u>Works to end-users” as well as having “ground[ed] output in Times Works”</u> , and Microsoft knew of same because of its extensive development and troubleshooting. <i>The Times</i> Compl. ¶¶ 178-180	<u>Publishers’ Works to end-users” as well as having “ground[ed] output in Publishers’ Works”</u> , and Microsoft knew of same because of its extensive development and troubleshooting. <i>Daily News</i> Compl. ¶¶ 211-214
Count V: Digital Millennium Copyright Act – Removal of Copyright Management Information (17 U.S.C. § 1202) – Alleging Microsoft and OpenAI “<u>removed The Times’ copyright-management information</u>” in developing GPT models and their <u>resultant “generati[ve] outputs”</u> . <i>The Times</i> Compl. ¶¶ 181-191	Count V: Digital Millennium Copyright Act – Removal of Copyright Management Information (17 U.S.C. § 1202) – Alleging Microsoft and OpenAI “<u>removed the Publishers’ copyright-management information</u>” in developing GPT models and their <u>resultant “generati[ve] outputs”</u> . <i>Daily News</i> Compl. ¶¶ 215-225
Count VII: Trademark Dilution (15 U.S.C. § 1125(c)) – Alleging Microsoft and OpenAI’s generative AI products make “<u>unauthorized use of The Times’ distinctive, famous trademarks</u> on lower quality and inaccurate <u>“outputs generated by Defendants’ GPT-based products.”</u>” <i>The Times</i> Compl. ¶¶ 198-204	Count VII: Trademark Dilution (15 U.S.C. § 1125(c)) – Alleging Microsoft and OpenAI’s generative AI products make “<u>unauthorized use” of the Publishers’ distinctive, famous trademarks</u> on lower quality and inaccurate <u>“outputs generated by Defendants’ GPT-based products.”</u>” <i>Daily News</i> Compl. ¶¶ 233-249

In sum, The Times’ and Daily News’ cases include virtually identical legal claims and allege virtually the same facts based on the same new generative search technology to support such claims. *See* Ex. A. Under these circumstances, consolidation of these cases together is warranted and appropriate.

II. The Consolidated Newspaper Cases Should Be Set On A Schedule That Provides At Least Six Months Of Fact Discovery Commencing In June 2024.

For all of the reasons described, The Newspaper Cases—unlike the Consolidated Class Actions—involve newly accused generative search technology (*e.g.*, Bing Copilot) and the newspapers’ copyrighted works allegedly present in that technology’s resultant outputs (*i.e.*, as a function of a user’s prompt). The Newspaper Cases also center on the distinctive relationship between Microsoft’s generative AI offerings and search engine technology, as well as how

generative search is different from “standard” search. This technological interaction is likewise not at issue in the Consolidated Class Actions. Instead, the Consolidated Class Actions’ claims strictly concern infringement of plaintiffs’ copyrighted works based strictly on their hypothetical inclusion in LLM model **data training sets**. Indeed, *The Times* and *Daily News* has already made it clear from discovery served in the Newspaper Cases that it is “concern[ed]” with “the use of Times Content for ... every Text Generation AI Model ... for **any post-training activity**[.]” as well as “[d]ocuments concerning the features or capabilities of [generative search].” ECF No. 130-3 at 6, 8 (*The Times*’ RFP Nos. 1, 12); Ex. B at 7, 9 (*Daily News*’ RFP Nos. 1, 12). In total, thus far *The Times* and *Daily News* have served at least five requests for production each, and *The Times* has served two interrogatories directly related to generative search technology that is beyond the scope of the Consolidated Class Actions. Accordingly, if the Court is inclined to consolidate *The Times* and *Daily News* cases, a modified, later schedule for the consolidated matter (as opposed to what is presently set in *The Times* case) is necessary.

Specifically, the additional counts in *The Times*’ and *Daily News*’ Complaints—based on newly accused and different technology not present in the Consolidated Class Actions—renders the scope of both fact and expert discovery far different. *See* ECF No. 129 (*Microsoft’s Conditional Opposition to The Times’ Motion for Leave to File First Amended Complaint*). Microsoft will have to conduct a more extensive investigation, respond to a broader set of discovery, and locate additional witnesses for the generative search technology at issue.

As just one example of the added burden and complexity, in addition to conducting discovery on plaintiffs’ copyrighted material purportedly used to **train** the LLMs, Plaintiffs in these cases seek different and additional discovery regarding the **functionality and operation** of Microsoft’s generative search products. This process will likely involve collecting and analyzing

troves of data (by both machines and technologists) to determine whether any portion of plaintiffs' copyrighted works were referenced following users' prompts of Microsoft's Bing Copilot. And, this gargantuan task is not just a function of expense; rather, there are technological limitations to Microsoft's systems' ability to run queries on such data and the immense amount of time such queries will necessarily take.

Similarly, Plaintiffs in *The Times* and *Daily News* cases seek different discovery regarding the interaction between the newly accused, distinct technologies at issue, including the following Microsoft products: Microsoft's Azure AI platform and Microsoft's Bing Copilot (formerly Bing Chat).⁴ *The Times* Compl. ¶¶ 102, 118; *see also Daily News* Compl. ¶¶ 65, 93, 96. While these products appear in the Consolidated Class Actions' Complaints, their full functionality and operation does not. As a result, the scope of fact discovery will differ significantly across these cases.

Expert discovery will be substantially impacted as well by the broader scope of the Newspaper Cases. Microsoft's technical and expert witnesses will need to analyze Microsoft's generative AI search products' outputs data. Only in this way can a full picture be formed of users' interaction with Microsoft's generative search technology, and the vast amount of substantial non-infringing uses. Much the same, expert analysis is needed on the technical interplay between the consumer-facing generative search products and the internal workings of the LLM and generative AI models.

Finally, while *The Times* and *Daily News* cases raise identical legal and factual issues, any

⁴ For example, *The Times* and *Daily News* have already requested to inspect: "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." ECF No. 130-4 at 2 (*The Times*' Req. for Inspection ¶ 5); *see also* Ex. C at 2 (same for *Daily News*).

consolidated schedule would need to account for Microsoft and the OpenAI Defendants' need to conduct discovery specific to the eight new enterprise *Daily News* plaintiffs—discovery which has not yet begun as the defendants have only recently responded to the Complaint and the parties have just held a Rule 26(f) conference, and have not yet submitted a Rule 26(f) Report, because this case is still in its infancy. This only further amplifies the need for setting these two (potentially consolidated) cases on a later schedule to allow the parties sufficient time to fully and fairly prepare their legal cases.

* * *

For all of these reasons, Microsoft joins in the OpenAI Defendants' request to consolidate *The Times* case with the *Daily News* matter, and respectfully requests that the Court set the schedule for the consolidated cases so that there is at least six months of fact discovery commencing in June 2024 so that Microsoft has the opportunity to fully and fairly defend itself against these claims.

Dated: June 14, 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
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*