VIA ECF

September 9, 2025

Hon. Ona T. Wang
U.S. District Judge for the Southern District of New York
Daniel Patrick Moynihan United States Courthouse
500 Pearl Street
New York, NY 10007

Re: *In re OpenAI, Inc.*, *Copyright Infringement Litigation*, 25-md-3143 (SHS) (OTW)
This Document Relates To: The New York Times, No. 1:23-cv-11195

Dear Magistrate Judge Wang:

Plaintiffs The New York Times, the *Daily News* Plaintiffs, and The Center for Investigative Reporting (collectively, "News Plaintiffs") write to follow up on their motions to compel Microsoft to provide discovery into all versions of Microsoft's Copilot Chatbot. *See* Case No. 25-md-3143 ("*MDL*") Dkt. 372; Case No. 1:23-cv-11195 ("*Times*") Dkt. 487. As addressed in more detail in the prior motions and at oral argument, the latter, "refreshed" versions of Copilot infringe News Plaintiffs' copyrighted content in the same way as the original consumer Copilot product alleged in the Complaints. Microsoft is improperly withholding discovery into the latter versions of Copilot based on its representations about changes to the internal architecture of Copilot, and vague and unsubstantiated assertions of burden.

Following the discussion on the record at the August 12, 2025, discovery conference, and this Court's order directing the parties to continue meeting and conferring on limited discovery, *MDL* Dkt. 459, News Plaintiffs proposed a narrowed set of seven discovery requests into all versions of Microsoft's consumer Copilot product dated October 1, 2024, onwards. These seven requests are for:

1. Training data in Microsoft's possession, custody, or control for the in-scope families of GPT models at issue, which include GPT-1, GPT-2, GPT-3, GPT-3.5, GPT-3.5 Turbo, GPT-4, GPT-4 Turbo, GPT-4o, and GPT-4o mini, to the extent not already provided.[1]

2. Documents, reports, metrics, and other analyses to ascertain, *inter alia,* "regurgitations" of News Plaintiffs' content in output, access and reproduction of News Plaintiffs' content as part of retrieval augmented generation, hallucinations that dilute News Plaintiffs' asserted trademarks, and the extent to which Copilot outputs substitutive news content that harms the market for News Plaintiffs' works.[2]

[1] News Plaintiffs believe this will be minimally burdensome because most of the training data should have already been produced by Microsoft and/or OpenAI.
[2] News Plaintiffs reserve their right to inspect output log data on these issues, but are hopeful that the reports, analyses, and metrics above can obviate the need to do so.

3. Documents sufficient to show user engagement metrics.

4. Documents sufficient to show any differences related to access and reproduction of News Plaintiffs' content in the retrieval augmented generation process as compared to the earlier versions of Copilot.

5. Documents sufficient to show any differences with respect to guardrails, blocklists, or other measures relating to access and reproduction of News Plaintiffs' content as compared to the earlier versions of Copilot.

6. Production of documents improperly withheld as non-responsive on the basis that they pertain to later versions of Copilot.[3]

7. Supplementation of interrogatory responses to encompass the later versions of Copilot.

But Microsoft has refused to even engage with News Plaintiffs' efforts to negotiate a narrower scope of discovery in view of the parties' impasse with respect to the inclusion of GPT-4o and GPT-4o mini in the case, which News Plaintiffs understand power the latter versions of Copilot (and are already in this case via OpenAI). *See* Ex. A (meet and confer correspondence). Microsoft's refusal to produce ***any*** documents concerning GPT-4o and GPT-4o mini—and therefore most if not all discovery into the latter versions of Copilot—has frustrated any further meet and confer efforts to identify tailored areas of discovery into Copilot that get to the core issues in these cases. Microsoft has also failed to substantiate its burden argument with respect to any of these seven narrowly tailored requests. In fact, Microsoft did not articulate any burden with respect to requests 2-5 above. And Microsoft refused discovery into request 1 on the false premise that "GPT-4o and GPT-4o mini are not at issue in this case," and on requests 6 and 7 based on the bare assertion that they are "catch-all" requests that require "restarting discovery" and that Ex. A at 1, 4.

Because the parties have been unable to resolve this dispute after more than six months of meet and confer efforts, News Plaintiffs respectfully request that this Court address and rule on the scope of discovery into all versions of Copilot—including GPT-4o and GPT-4o mini—either at the upcoming September 25, 2025 discovery conference or as part of a confidential settlement conference.

September 9, 2025

Respectfully Submitted,

/s/ Steven Lieberman
Steven Lieberman
Rothwell, Figg, Ernst & Manbeck P.C.
Counsel for The New York Times and *Daily News* Plaintiffs

[3] News Plaintiffs understand this may be a very limited set of documents based on Microsoft's representation to this Court that "[o]nly 2 people overlap with BingChat." *MDL* Dkt. 274 at 89.

/s/ Davida Brook
Davida Brook
Susman Godfrey LLP
Counsel for The New York Times

/s/ Stephen Stich Match
Stephen Stich Match
Loevy & Loevy
Counsel for The Center for Investigative Reporting

cc: All Counsel of Record (via ECF)