KEKER
VAN NEST
& PETERS   LATHAM&WATKINS LLP   IIIORRISON FOERSTER

November 22, 2024

Hon. Ona T. Wang
Daniel Patrick Moynihan
United States Courthouse
500 Pearl St.
New York, NY 10007-1312

cc: All Counsel of Record Line (via ECF)

Re: OpenAI's Response to News Plaintiffs' supplemental letter (ECF 328)[1] in *The New York Times Co. v. Microsoft Corp.*, Case No. 23-cv-11195; *Daily News, LP v. Microsoft Corp.*, Case No. 24-cv-3285

Dear Judge Wang:

OpenAI writes in response to the News Plaintiffs' supplemental letter regarding training data. Contrary to Plaintiffs' insinuations, the facts are much more banal and the consequences too. OpenAI did not delete any evidence. What happened? *Plaintiffs* requested a configuration change to one of several machines that OpenAI has provided to search training datasets. Implementing Plaintiffs' requested change, however, resulted in removing the folder structure and some file names on one hard drive—a drive that was supposed to be used as a temporary cache for storing *OpenAI* data, but evidently was also used by Plaintiffs to save *some* of their search results (apparently without any backups). In any event, there is no reason to think that any files were actually lost, and Plaintiffs could re-run the searches to recreate the files with just a couple days of computing time.

As background, OpenAI first made training data available for inspection in June, but Plaintiffs delayed their review until October. Once they began, Plaintiffs triggered a series of technical issues due to their own errors. As a direct result of Plaintiffs' self-inflicted wounds, OpenAI has been forced to pour enormous resources into supporting Plaintiffs' inspection, much more than should be necessary. Meanwhile, Plaintiffs have tried to parlay their own difficulties to request relief that extends well beyond fixing the inspection set-up they keep breaking: namely, they seem to want an order compelling OpenAI to respond to nearly 500 million requests for admission. Fortunately, a more reasonable approach exists. The Court should direct Plaintiffs to collaborate with OpenAI to develop a plan for reasonable, targeted searches to be executed either by Plaintiffs or OpenAI. OpenAI has already invited Plaintiffs to collaborate on such searches. They just need to say "Yes."

Technical Issues and Mischaracterizations. As outlined in the November 2 letter (ECF 305 at 2-4), Plaintiffs' inspection efforts began with them repeatedly running flawed code that overwhelmed and crashed the file system. To address Plaintiffs' self-imposed injuries, OpenAI

---

[1] All record citations herein are to material in Case No. 23-cv-11195.

Page 2

invested significant resources, including days of engineer time as well as hardware and compute resources to get Plaintiffs back on track. OpenAI also offered to run at least some of Plaintiffs' searches for them and asked Plaintiffs to make a comprehensive proposal. Despite OpenAI's support, Plaintiffs returned to their inefficient "boil-the-ocean" searches, demanding ever-increasing hardware performance. When Plaintiffs requested a configuration change that they thought would accelerate their work, OpenAI accommodated their request, despite concerns that the proposed change would yield no speed improvements and might even hinder performance. It is now clear that Plaintiffs did not fully understand the change that they requested—a change that, upon implementation, inadvertently removed file-system information (such as folder structures and some file names) from a hard drive set up to cache OpenAI's data. Plaintiffs apparently also saved some of their own working files there (without making backups) and lost that organizational metadata from that drive. However, contrary to Plaintiffs' insinuations, there is no reason to think that the contents of any files were lost.

The core obstacle here is not technical; it is Plaintiffs' unwillingness to collaborate. Rather than address this minor setback jointly, Plaintiffs have exaggerated the nature and scope of the issue, portraying it as some catastrophic failure that irrevocably destroyed evidence. Not so. The enterprise-grade server that was affected (one of several that OpenAI has provided) was not their primary workstation. Plaintiffs have left this machine largely idle since it was set up, and in total, used that server for less than 3 days' worth of computing time. In other words, by the time the parties appear before the Court on December 3, Plaintiffs will have had enough compute time to re-run their search script five times over.

Plaintiffs' Delays and OpenAI's Efforts to Cooperate. OpenAI has offered to take over Plaintiffs' searches, provided Plaintiffs supply clear and reasonable proposals. On October 18, OpenAI asked Plaintiffs to explain what they thought OpenAI could do to help. Plaintiffs delayed over two weeks to provide an overbroad list of URL-based keywords and almost a month to outline a proposed text-search protocol. In response, OpenAI has worked diligently with its engineering team to evaluate these proposals. OpenAI's counsel has also held numerous meetings with the internal engineering team to go over project expectations and technical challenges, and testing has already begun.

Far from "refusing" to file a joint submission, OpenAI suggested using this time to collaborate on a comprehensive plan to finish Plaintiffs' searches—a solution that could resolve not only the present dispute but also prevent future technical issues. Plaintiffs failed to respond and filed their supplemental letter brief the next day.

Path Forward. OpenAI remains committed to resolving these issues efficiently and equitably. Provided Plaintiffs engage in meaningful collaboration, OpenAI stands ready to conduct reasonable searches on Plaintiffs' behalf. Plaintiffs' focus, however, must shift from manufacturing conflicts to achieving solutions. To this end, OpenAI respectfully requests that the Court direct the parties to meet and confer on search proposals. This approach balances Plaintiffs' discovery needs with the need to avoid unnecessary disputes and discovery burdens.

2833826

Page 3

Respectfully,

KEKER, VAN NEST & PETERS LLP[2]	LATHAM & WATKINS LLP	MORRISON & FOERSTER LLP
/s/ Thomas E. Gorman Thomas E. Gorman	*/s/ Elana Nightingale Dawson* Elana Nightingale Dawson	*/s/ Carolyn Homer* Carolyn Homer

---

[2] All parties whose electronic signatures are included herein have consented to the filing of this document.