

**MORRISON FOERSTER**

425 MARKET STREET  
SAN FRANCISCO  
CALIFORNIA 94105-2482  
  
TELEPHONE: 415.268.7000  
FACSIMILE: 415.268.7522  
  
WWW.MOFO.COM

MORRISON FOERSTER LLP  
AUSTIN, BERLIN, BOSTON, BRUSSELS, DENVER, HONG  
KONG, LONDON, LOS ANGELES, MIAMI, NEW YORK, PALO  
ALTO, SAN DIEGO, SAN FRANCISCO, SHANGHAI,  
SINGAPORE, TOKYO, WASHINGTON, D.C.

September 27, 2024

*Sent via ECF*

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

Writer's Direct Contact  
+1 (415) 268-6066  
JGratz@mofocom

RE: *The New York Times Co. v. Microsoft Corp., et al.*, No. 1:23-cv-11195  
*Daily News, L P, et al. v. Microsoft Corp., et al.*, No. 1:24-cv-3285

Dear Judge Wang:

We write on behalf of all parties in the above-entitled actions to provide an update as to the status of the outstanding disputes raised at the Status Conference on September 12, 2024 in accordance with the Court's subsequent Order on September 13, 2024 (*The New York Times*, ECF No. 243). OpenAI has leave of The New York Times Company ("The Times"), the Daily News Publishers (collectively with The Times, "Plaintiffs"), and Defendant Microsoft to file this letter jointly on behalf of all parties to this action.

**I. Plaintiffs' Letter Motion to Compel (*Times*, ECF 128) re: GPT, GPT-2, and GPT-3**

The parties met and conferred on the scope of OpenAI's productions related to GPT, GPT-2, and GPT-3 on Friday, September 20, 2024 and exchanged further correspondence on September 23, 25, and 26. The parties have made progress with respect to this dispute and anticipate continuing to confer in the coming days. At present, OpenAI has agreed to produce or otherwise make available the following: (1) training data for GPT; (2) training data and models for GPT-2 and GPT-3; and (3) non-custodial documents related to GPT-2 and GPT-3, as it has for other models. Specifically, OpenAI plans to produce the following non-custodial documents related to GPT-2 and GPT-3: the model cards, README files, research papers, and announcements. The Times has agreed to defer further requests for non-custodial documents related solely to GPT pending The Times's inspection of the GPT training data but continues to seek custodial data that is related to these models and hits on search terms from appropriate custodians.

The Times has proposed search terms intended to capture documents concerning GPT, GPT-2, and GPT-3, and while The Times understands that the exact terms are still being negotiated, The Times contends that OpenAI should commit to running terms that are intended to capture documents concerning these models. The Times also contends that OpenAI should produce the same kinds of non-custodial documents for GPT-2 and GPT-3 that OpenAI agreed to produce for

## MORRISON FOERSTER

September 27, 2024

Page 2

the later-in-time models, including documents “intending to cover the entire training process.” Dkt. 136-1 at 1.

OpenAI received The Times’s most recent search-term proposal, which included search terms related to GPT, GPT-2, and GPT-3, on September 25, 2024. OpenAI is evaluating that proposal, as explained in more detail below. OpenAI is not withholding documents that are otherwise discoverable and responsive solely on the basis that they relate to the models discussed herein.

The parties anticipate continuing to confer regarding this issue.

### **II. Plaintiffs’ Letter Motion to Compel (*Times*, ECF 141) re: RFP 6 (Negotiations with Third Parties)**

The parties met and conferred on the scope of OpenAI’s productions of third party agreements and related negotiations documents and communications on September 20, 2024. OpenAI provided a written proposal memorializing its position at the conferral on September 24, 2024. On September 25, 2024, The Times responded to and sought clarification about some portions of OpenAI’s proposal.

OpenAI has agreed to produce some communications between OpenAI and third parties with whom OpenAI had a term sheet or an executed data access agreement. The parties are continuing to negotiate the scope of this production. The Times contends that the production should include not only communications relating to the financial terms of deals, but also the agreements’ scope, releases/covenants not to sue for past use, and term and termination provisions.

In addition, The Times contends that OpenAI should produce (1) internal communications between OpenAI employees about the above deals, and (2) internal and external communications related to *potential* deals that did not ultimately materialize into a signed agreement, including but not limited to communications relating to OpenAI’s negotiations with The Times.

OpenAI has agreed to produce the relevant agreements and term sheets by October 11, 2024 to, in its view, allow third parties adequate time to seek protective orders if they so wish. OpenAI has further agreed to produce related communications with third parties after October 11, 2024, in accordance with its standard document review procedures. OpenAI has also agreed to produce third party communications concerning the scope of the agreements and requested that The Times articulate the relevance of releases/covenants not to sue for past use, and term and termination provisions. With respect to “potential” deals, OpenAI will confer with The Times about how to define “potential” deals in order to evaluate the scope and feasibility of a possible production. OpenAI has further requested that The Times withdraw or narrow its request for OpenAI’s internal communications about the agreements due to the burden of reviewing irrelevant and privileged material.

The parties anticipate continuing to confer regarding this issue.

## **MORRISON FOERSTER**

September 27, 2024

Page 3

### **III. Plaintiffs' Letter Motion to Compel (*Times*, ECF 141) re: RFP 11 (Documents concerning OpenAI's For-Profit Entity Formation)**

The parties met and conferred on the scope of OpenAI's productions related to the creation of OpenAI's for-profit entity on September 20, 2024 and exchanged further correspondence on September 23, 25, and 26. The Times has asked OpenAI follow-up questions related to OpenAI's positions with respect to this dispute.

The Times's position is that, absent a specific set of documents responsive to this RFP that OpenAI identifies as irrelevant (e.g., documents about the tax consequences of forming the new entity), OpenAI should produce all responsive and non-privileged documents about the creation of OpenAI's for-profit entity, including documents about how the creation of the for-profit entity related to OpenAI's overall business strategy (e.g., strategies for securing investments). The Times also contends that OpenAI should produce documents concerning this transition that relate to Microsoft, including communications with Microsoft.

OpenAI has agreed to produce documents related to its decision to form a for-profit entity that address or discuss the products and services at issue in this case—that is, OpenAI's GPT products and OpenAI's GPT large language models. That agreement necessarily includes documents related to Microsoft that also touch on both OpenAI's decision to form a for-profit entity and OpenAI's GPT products and GPT large language models. OpenAI has asked The Times for an explanation of why it believes Microsoft-related documents that do not fall within what OpenAI has agreed to produce are nevertheless relevant. OpenAI awaits The Times's response. It is OpenAI's understanding that, to date, The Times has not narrowed its request with respect to this dispute.

The parties anticipate continuing to confer regarding this issue.

### **IV. Plaintiffs' Letter Motion to Compel (*Times*, ECF 228) re: Search Terms**

OpenAI provided The Times with a search-term counterproposal and hit report on September 19, 2024. The parties met and conferred on the issue on September 20, 2024. The Times provided OpenAI with a follow-up proposal on September 25, 2024. Under the parties' agreed upon, but not yet filed, ESI protocol, OpenAI has until October 2, 2024 to provide The Times with hit counts and a new counterproposal to the extent it does not accept in full the terms in The Times's counterproposal.

### **V. Plaintiffs' Letter Motion to Compel (*Times*, ECF 221) re: Confidentiality Designations**

The parties have made progress toward resolving this dispute. Microsoft agreed to undertake to identify and re-designate particular categories of documents in its existing production where, in Microsoft's view, there are not likely to be competitive, strategic, or sensitive technical concerns,

**MORRISON FOERSTER**

September 27, 2024

Page 4

and to implement similar considerations on a going-forward basis. The parties are continuing to meet and confer regarding the timing and scope of Microsoft's re-designations.

**VI. Model Inspection Protocol**

Plaintiffs and OpenAI have reached an impasse on a dispute concerning the model inspection protocol, and set forth their respective positions below.

Plaintiffs' Position:

The parties have a dispute relating to The Times's and the Daily News Publishers' (together, "Plaintiffs") requests to inspect OpenAI's GPT large language models ("LLMs") under Rule 34. Plaintiffs' discovery into these models bear on significant issues in the consolidated cases, including the extent to which unauthorized copies of Plaintiffs' copyrighted works are encoded within the parameters of the GPT LLMs themselves and the propensity of the GPT LLMs to disseminate or publicly display those works in outputs.<sup>1</sup>

Plaintiffs have made a good faith estimate of the volume of prompting of the models (measured in the form of retail "credits") for this aspect of discovery. Because OpenAI has infringed millions of Plaintiffs' copyrighted works, Plaintiffs' request is proportional to the needs of the case and Plaintiffs' investigation into OpenAI's infringement. Rather than provide Plaintiffs' the access to the models they are entitled to under the Federal Rules, OpenAI seeks to hide its infringement by professing an undue – yet unquantified – "expense." OpenAI has refused to state what its actual costs would be, and instead improperly focuses on what it charges its customers for retail services as part of its (for profit) business.

Although Plaintiffs and OpenAI are otherwise in agreement as to the protocol under which OpenAI will make its GPT LLMs available to Plaintiffs, OpenAI insists upon a cost-shifting paradigm wherein it provides an initial cap of \$15,000 of retail credits for each Plaintiff to spend on prompting OpenAI's GPT LLMs through an application programming interface, and thereafter, for any additional credits they may need, Plaintiffs will have to pay for half the retail cost (that is, pay OpenAI commercial rates for its service in order to get discovery). OpenAI's proposal is improper for at least three reasons. First, Plaintiffs should not bear the cost of finding out how much of their copyrighted content OpenAI copied within the models. *Oppenheimer Fund, Inc. v. Sanders*, 437 U.S. 340, 358 (1978) ("the presumption is that the responding party must bear the expense of complying with discovery requests"). Second, OpenAI has failed to articulate any undue burden or expense associated with Plaintiffs' discovery to justify this cost-shifting proposal beyond vague references to compute costs. OpenAI has refused to provide any specific information quantifying the alleged compute costs or clarify whether any of those costs

---

<sup>1</sup> Although related, these are different issues for discovery than the issue as to whether unauthorized copies of Plaintiffs' works are contained in OpenAI's training datasets. Those datasets have not yet been produced and the parties are still negotiating a protocol that will govern Plaintiffs' inspection of OpenAI's training datasets.

**MORRISON FOERSTER**

September 27, 2024

Page 5

go to its co-defendant Microsoft. Third, OpenAI should not profit from Plaintiffs' discovery by charging retail pricing for access to its models.

Plaintiffs have requested that OpenAI provide an initial \$400,000 in retail credits to each Plaintiff, with the option for Plaintiffs to negotiate with OpenAI for any additional credits. Plaintiffs' \$400,000 retail credit estimate is based on OpenAI's public retail pricing information for the GPT LLMs, the millions of Plaintiffs' works at issue, and Plaintiffs' pre-suit testing of GPT-4 as reflected in Exhibit J to both Complaints. OpenAI's de minimis counteroffer of \$15,000 does not reflect a good faith effort to find a solution to the very real issue of OpenAI's expansive infringement.

OpenAI's Position:

The parties' dispute concerning model inspection does not concern whether the newspaper Plaintiffs will have access to models. They already do. The dispute centers on whether that access is provided with a blank check, as the newspapers want, or with reasonable guardrails to promote efficient use of resources, as OpenAI suggests. The proportionality standards of Rule 26 mandate the latter.

The newspaper Plaintiffs—the New York Times and the eight regional newspapers—are requesting hundreds of thousands of dollars of credits to run an arbitrary and unsubstantiated – and likely unnecessary – number of searches on OpenAI's models, all at OpenAI's expense. This is not a reasonable or proportional request, especially given that the models are not required to determine what data was used to train the models. OpenAI has already separately produced the training data. Plaintiffs instead request unfettered access to the models to run an inordinate number of queries of unknown nature and purpose. The demand that OpenAI cover the entirety of the expense of this unbounded and unsubstantiated fishing expedition through the models should be rejected.

First, most of these models are publicly available, and Plaintiffs can use the OpenAI accounts they have already established to run these searches. Indeed, that is what Plaintiffs did to prepare the exhibits they attached to their Complaint. But OpenAI, like other AI model providers, charges for searches in order to cover the cost of such searches; the searches require computing power and resources. Plaintiffs, however, would prefer not to pay anything at all and have asked OpenAI to bear the entire financial burden by providing them with \$800,000 of credit—\$400,000 for the New York Times and \$400,000 for the regional newspapers—to run whatever searches Plaintiffs want. Plaintiffs admitted that the \$800,000 is an estimate. When pressed for the details of how that estimate was calculated, Plaintiffs could not offer sufficient detail for OpenAI to replicate or test their math. By their own admission, Plaintiffs are, at this point, uncertain whether they actually need that amount and whether they could inspect the models for

**MORRISON FOERSTER**

September 27, 2024

Page 6

less through more efficient search strategies.<sup>2</sup> OpenAI is willing to accommodate Plaintiffs' discovery needs. But the volume of searches Plaintiffs intend to run will impose costs on OpenAI. OpenAI has therefore proposed a two-step framework to ensure Plaintiffs are efficient and reasonable while getting the discovery they seek and not unduly burdening OpenAI.

Second, OpenAI has offered to provide each set of newspaper plaintiffs (from The Times and Daily News) with an *initial* tranche of \$15,000 of credits to begin their searches, which would allow Plaintiffs to use over a *billion* tokens in inputs<sup>3</sup> This is a starting point; OpenAI has explicitly told Plaintiffs that this is *not* a final "cap." This credit—offered at OpenAI's own expense and without prejudice to additional future credits—would enable Plaintiffs to conduct initial searches, determine the scope of any further searches needed, and investigate how searches can be executed to minimize burden and expense. With the benefit of this real-world experience, Plaintiffs and OpenAI will have actual data—instead of speculative estimates—to determine the scope of additional searches and credits needed.

Third, OpenAI has offered that after Plaintiffs have exhausted their initial tranche of \$15,000 of credits, OpenAI will meet and confer to determine what additional credits are required. At that point, OpenAI would be willing to agree to split the future API credit burden between the parties, i.e., NYT would pay for half of the credits and OpenAI would provide the other half. Plaintiffs' complaints about charging "retail pricing" are thus misleading. If OpenAI was charging retail pricing it would ask for Plaintiffs to pay for all credits; it is only asking for Plaintiffs to pay for some.

OpenAI's two-phased proposal strikes a measured balance to the parties' competing concerns. It provides Plaintiffs with enough resources (paid for by OpenAI) to perfect their search strategy and arrive at a total credit amount that is based on real-world data. It also imposes some discipline on Plaintiffs to ensure their search strategy is targeted and efficient—not unfettered and wasteful—as required by the proportionality guidelines in Rule 26.

**VII. Training Data Inspection Protocol**

The parties are continuing to meet and confer on the training data inspection protocol and have made progress in narrowing or eliminating disputes. Plaintiffs sent their latest proposal to OpenAI on September 25 and the parties are continuing to negotiate.

---

<sup>2</sup> One such strategy is using "batch API" requests—essentially running multiple searches in groups and obtaining results up to 24-hours later—would be more efficient and cost up to 50% less. But Plaintiffs have never indicated if their estimate takes such strategies into account.

<sup>3</sup> For example, the pricing for the accused GPT-4 Turbo model is \$10 for 1 million input tokens and \$30 for 1 million output tokens. See <https://openai.com/api/pricing/>. \$15,000 in credits would allow Plaintiffs to use over a billion tokens in inputs. OpenAI would also provide Plaintiffs the highest tier of API usage, which starts at \$1,000. See <https://platform.openai.com/docs/guides/rate-limits?context=tier-five>.

**MORRISON & FOERSTER**

September 27, 2024  
Page 7

Respectfully submitted,

/s/ Joseph C. Gratz  
Joseph C. Gratz (*pro hac vice*)  
MORRISON & FOERSTER LLP  
425 Market Street  
San Francisco, CA 94105-2482  
JGratz@mofocom  
Telephone: 415.268.7000  
Facsimile: 415.268.7522

***Counsel for OpenAI Defendants***