November 24, 2025

In re OpenAI, Inc., *Copyright Infringement Litigation*, 25-md-3143 (SHS) (OTW)
This Document Relates To: All Actions

Dear Magistrate Judge Wang:

Pursuant to this Court's Order at Dkt. 751, Class and News Plaintiffs submit this letter brief outlining their proposed scope for additional Microsoft discovery following Judge Stein's ruling at Dkt. 707. Plaintiffs seek targeted additional discovery into less than [redacted]% of Microsoft's "implementations" of GPT-4o and GPT-4o Mini. Microsoft refuses to provide this discovery and instead offers only limited discovery on a single implementation (Refreshed Copilot). Because the additional discovery Plaintiffs seek is relevant and proportional, the Court should compel Microsoft to provide it.

I. Introduction

Since May 2024, Plaintiffs have been diligently seeking discovery into the GPT-4o series of models and Microsoft's products incorporating these models. When Microsoft refused to provide this discovery, Plaintiffs promptly sought relief from this Court. *See, e.g., Authors Guild* Dkt. 356 at 3 (Class Plaintiffs' March 31, 2025 motion to compel finances for Microsoft's products, including products that incorporate GPT-4o and GPT-4o); *Times* 487 (The Times's March 31, 2025 motion to compel discovery into Refreshed Copilot).

Judge Stein's October 27, 2025 Order rejected Microsoft's reasons for refusing to provide this discovery and made clear this MDL includes GPT-4o and GPT-4o Mini. Dkt. 707 at 6 ("At the May 22 hearing, the Court was clear that the Consolidated Class Action Complaint could include any models that were already a part of this action, which were the models identified in Judge Wang's Authors Guild discovery ruling. ***Microsoft's suggestion that it can evade litigation about some of those models because Judge Wang's ruling was based on a discovery response with which Microsoft disagreed is unavailing."*** (emphasis added)). Judge Stein also denied Microsoft's motion to strike allegations about products that incorporate GPT-4o and GPT-4o Mini "such as the Microsoft Office suite of products and Microsoft Copilot." *Id.*

Microsoft subsequently disclosed to Plaintiffs that it uses GPT-4o and GPT-4o Mini in [redacted]"[1] Because each implementation represents a separate use of Plaintiffs' copyrighted works (and therefore separate instances of infringement), all implementations are relevant. Yet Plaintiffs do not seek detailed discovery into anywhere near all [redacted] implementations of these models and have instead requested narrow discovery into just 10 of them, or [redacted]%, including the product Microsoft calls "Refreshed" Copilot. Microsoft, by contrast, seeks to limit any additional discovery to Refreshed Copilot, ignoring all other implementations.

Plaintiffs now ask the Court to order Microsoft to provide the following targeted discovery: (1) deposition testimony and limited document and data productions about Refreshed Copilot, (2) limited document and data productions for Copilot Search, an "implementation" of GPT-4o Mini that Microsoft uses to provide news content to users, (3) limited document and data productions for eight other Microsoft "implementations" that incorporate GPT-4o and GPT-4o

[1] Exhibit 14 is the list of [redacted] implementations that Microsoft provided to Plaintiffs. During the parties' November 17, 2025 conferral, Microsoft's counsel was unable to answer questions about this list and the implementations.

Mini (including products specifically marketed at writing) like [redacted], (4) Microsoft's copies of any training data used for GPT-4o or 4o-Mini, (5) supplemental interrogatory responses describing how Microsoft developed and uses GPT-4o and 4o-Mini and (6) financial information for all Microsoft products that incorporate the earlier-in-time GPT models, like GPT-3, GPT-3.5, and GPT-4. Plaintiffs made this proposal on November 19, following the November 17 call. Ex. 13 (email from A. Frawley). Microsoft never responded.

II. Argument

1. "Refreshed" Copilot

Microsoft offered Plaintiffs the following limited discovery on Refreshed Copilot:

1. Two witnesses for deposition, one to discuss how the product works and one to discuss data (output logs and any existing metrics related thereto), but limited to the existing deposition hours caps;
2. Source code for inspection and a small number of additional "go and get" documents sufficient to show how the product works; and
3. Financials related to the product.

Microsoft's offer is deficient for four reasons.

First, Microsoft should confirm that the two witnesses will be 30(b)(6) witnesses, and the depositions should not count against existing hours caps. Because Microsoft improperly instructed its 30(b)(6) witness on Microsoft's data metrics and dashboards not to answer questions about Refreshed Copilot during his September 30 deposition, News Plaintiffs must re-do that deposition for Refreshed Copilot. Microsoft chose that approach knowing full well that Judge Stein could rule against them and that the witness would need to be re-deposed. *See* Ex. 1 (S. Manivannan 30(b)(1) Deposition Tr. at 19:25-20:19 (Judge Wang addressing Microsoft's instruction to its witness not to answer questions about Refreshed Copilot: "if you want to deal with a later ruling that does allow the discovery and then you're having to haul the witness back in to answer these questions, if these questions are indeed only a few questions. That's your decision to make … if it turns out that such questions are allowed or would be allowed, you would have to either move or agree that you would bring this witness back"), 20:22-21:3 (counsel for Microsoft refusing to allow Mr. Manivannan answer questions regarding Refreshed Copilot and representing that "[Microsoft is] prepared to bring the witness back if Judge Stein decides that new Copilot is within the scope of the case."). Plaintiffs should not be taxed for these do-over depositions.

Second, for the "go and get" documents, Microsoft should confirm that those documents will include: (i) documents sufficient to show any classifiers that Microsoft uses to track how people are using Refreshed Copilot, including to access news content, (ii) metrics data showing how people are using Refreshed Copilot, including data comparing the click-through rates for Refreshed Copilot as compared to traditional Bing Search, and (iii) documents marketing or advertising Refreshed Copilot's ability to assist with drafting novels, stories, books, news, or other writings. These three discrete sets of documents are minimally burdensome and are relevant to Plaintiffs' rebuttal to Microsoft's fair use analysis, including the first and fourth fair use factors with respect to whether Copilot uses copied content in a transformative way and is harming the market for Plaintiffs' original copyrighted works.

Third, the "log data" portion of Microsoft's offer (i.e., prompts, outputs, and retrieval augmented generation data) includes an improper limitation: Microsoft is "only offering to collect and produce information that is already reasonably accessible and not to develop anything new or run any new reports that do not already exist." Ex. 2 (November 17 email from A. Hurst) at 1. That limitation is improper because even if Microsoft does not currently have copies of the specific data reports that Plaintiffs requested, Microsoft has the underlying data available and can readily search and query that data to obtain the information that Plaintiffs seek.[2] For example, at least one of Microsoft's witnesses testified that his team [redacted]. Ex. 4 (E. Abib (30)(b)(1) Rough Dep. Tr.) at 49.

Fourth, Microsoft should produce documents it already collected for the agreed custodians and search terms but then withheld on the ground that the document related to Refreshed Copilot. For avoidance of doubt, Plaintiffs are not requesting that Microsoft add any additional custodians or search terms. This request should place minimal burden on Microsoft because Plaintiffs seek the production of documents that Microsoft has: (i) already collected, (ii) hit on the search term "Copilot," and (iii) have not yet been produced. Indeed, Plaintiffs understand this may be a very limited set of documents based on Microsoft's representation to this Court that, of the existing custodians, "[o]nly 2 people overlap with BingChat." MDL Dkt. 274 at 89.

2. Copilot Search

Plaintiffs seek limited discovery into Copilot Search, which is a generative AI Microsoft chatbot, like Bing Chat and Refreshed Copilot. Microsoft launched Copilot Search in April 2025, describing it as the "next evolution of search." Ex. 5 at 1. It is relevant because it is powered by [redacted] *See* Ex. 6 (Ribas Tr.) 22:2-12 (deposition testimony confirming that Copilot Search is [redacted]. Judge Stein expressly denied Microsoft's request "to exclude Microsoft products that incorporated GPT-4o and GPT-4o Mini," paving the way for Plaintiffs to obtain "discovery into these [] Microsoft products," which include Copilot Search. Dkt. 707 at 6.

Mindful of Judge Stein's guidance that any new discovery must be proportional, *id.*, News Plaintiffs have prioritized Copilot Search, among the many other products and implementations that Microsoft recently disclosed as powered by GPT-4o and GPT-4o Mini. Discovery into Copilot Search is warranted because this product can be used to access news content, which makes it relevant to fair use factors one and four as well as News Plaintiffs' harm allegations. For example, for factor four "the key question" is whether the copying "would substantially diminish the market for the original work." *Kadrey v. Meta Platforms, Inc.*, 788 F. Supp. 3d 1026, 1038 (N.D. Cal. 2025). Copilot Search (like Bing Chat and Refreshed Copilot) delivers news content directly to users, thus diminishing the need for people to visit news sites. Marketing materials expressly note that Copilot Search provides "summaries" of search results, which makes it "perfect for when you want more than just a list of links [to websites]." Ex. 7 at 3. By contrast, Microsoft describes its "Traditional Search" product as better for "finding websites" *Id.*

[2] To avoid any delay or burden associated with producing the log data for Refreshed Copilot, Plaintiffs as a compromise asked Microsoft to search for and produce documents, reports, metrics, and other analyses of how Refreshed Copilot is used. Ex. 3 (November 5, 2025, email from J. Maisel). Plaintiffs have also offered to discuss treating the output log data for Bing Chat/"original" Copilot as representative of the output log data for Refreshed Copilot in view of the relatively advanced stage of that discovery. (Although Microsoft has yet to produce this data to Plaintiffs, it has promised to do so by December 4). Microsoft declined this proposal. *Id.*

News Plaintiffs seek limited document productions and data for Copilot Search. Microsoft should produce: (1) documents sufficient to show how people are using Copilot Search, including any metrics or classifiers that reveal the extent to which the product is used to access news content, and (2) data sufficient to compare click-through rates for Copilot Search with click-through rates for traditional search—i.e., data on whether users are more or less likely to click on links (and visit sites) in Copilot Search as compared with traditional search. News Plaintiffs are not seeking any additional custodians or search terms for Copilot Search.

3. Additional Implementations of GPT-4o and GPT-4o Mini

In addition to Refreshed Copilot and Copilot Search, Plaintiffs also seek discovery on a small number of Microsoft's [redacted] implementations of GPT-4o and 4o-Mini: [redacted] and (9) any other implementations that Microsoft is aware of that are marketed towards or used to draft novels, stories, books, or news content. *See, e.g.*, Ex. 8. Plaintiffs selected these implementations among the [redacted] that Microsoft disclosed because these implementations appear most related to books and news content. These implementations are relevant to Plaintiffs' direct infringement claims against Microsoft because each represents a separate infringing use of Plaintiffs' works. *See, e.g.*, CCAC ¶ 5 (explaining that models including GPT-4o and 4o-Mini power ChatGPT, which Defendants have "followed with a suite of other commercial offerings"). These implementations are also relevant to both statutory and disgorgement damages.

For each implementation, Plaintiffs seek a limited set of discovery, including (i) financials, (ii) documents sufficient to show any classifiers that Microsoft uses to track how people are using the implementation, (iii) documents marketing or advertising the implementation's ability to assist with drafting novels, stories, books, news content, or other writings, (iv) any documents that Microsoft already collected that hit on the name of the implementation that have not yet been produced, and (v) a 30(b)(6) witness on these implementations that does not count against Plaintiffs' existing deposition time. These *requests*, limited to a *fraction of Microsoft's implementations*, are minimally burdensome and proportional to the needs of the case.

4. Training Data for GPT-4o and GPT-4o Mini

Microsoft should produce the training data in its 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. Plaintiffs understand this request should not impose any additional burden on Microsoft [redacted] Plaintiffs now ask that Microsoft confirm this representation in writing and further confirm whether GPT-4o and GPT-4o mini were trained using Bing Index data. If these two models were trained using Bing Index data, Microsoft should produce that data.

5. Supplemental Interrogatory Responses

Plaintiffs seek supplemental interrogatory responses to address the GPT-4o and GPT-4o Mini models that Judge Stein has declared to be in scope. Microsoft has already responded to interrogatories to answer questions about Microsoft's involvement with other large language models. But when Microsoft answered those interrogatories, it refused to provide any information about GPT-4o and GPT-4o Mini, claiming those models were outside the scope of the case. Given Judge Stein's ruling, Microsoft should now supplement those responses to answer the same questions as to GPT-4o, GPT-4o Mini, and Refreshed Copilot. These interrogatories are: Class Interrogatory Nos. 1, 6 (Ex. 9 at 4, 11); NYT Interrogatory Nos. 7-9, 15, 18 (Ex. 10 at 6-10, 17-20, 28-29); *Daily News* Interrogatory Nos. 3-11 (Ex. 11 at 6-31), and CIR Interrogatory Nos. 4 and 8 (Ex. 12 at 3, 5). Providing updated interrogatory responses is not burdensome; in fact, Microsoft has never raised any burden concern as to this aspect of Plaintiffs' request.

6. Financials for products that incorporate OpenAI's LLMs

In addition to the financials for Refreshed Copilot and the implementations listed above in Section II(3), Microsoft should also produce financials for all products that incorporate GPT-3, GPT-3.5, GPT-3.5 Turbo, GPT-4, and GPT-4 Turbo. Microsoft has refused to produce financials for products that are not priced specifically for providing GPT LLM capability.[3] Class Plaintiffs moved to compel this discovery in March. *See Authors Guild* Dkt. 374. As Plaintiffs argued in their prior motion, Microsoft's refusal is based on two false premises: (1) that Plaintiffs must prove a "causal nexus" between Microsoft's infringement and the revenue in question at the discovery stage to be entitled to this discovery and (2) that the profits at issue are indirect profits rather than direct profits. *See id.*[4]

Both arguments are wrong. At the discovery stage, all that Plaintiffs must show is that the financials for Microsoft's products incorporating the earlier models are relevant. *See* FRCP 26(b)(1); *see also Prolo v. Blackmon*, 2022 WL 2189643, at *9 (C.D. Cal. Mar. 25, 2022) (rejecting argument that a plaintiff must prove there is a causal link between "the allegedly improper use of a photograph and Defendants' indirect profits to be entitled to obtain discovery about indirect profits"). And the profits at issue here are direct profits Microsoft earned from products powered by OpenAI's LLMs. *See, e.g.*, *Graham v. Prince*, 2023 WL 5917712, at *5 (S.D.N.Y. Sept. 11, 2023) ("Direct profits arise from the sale of the infringing good."); *see also On Davis v. The Gap, Inc.*, 246 F.3d 152, 160 (2d Cir. 2001), *as amended* (May 15, 2001) (explaining that a publisher would have to produce gross revenues for sales of an anthology containing an infringing poem but not gross revenues for works or genres that do not include the infringing poem).

Microsoft must produce financials for its products that incorporate OpenAI's earlier LLMs and cannot limit its production to only those products priced specifically for providing OpenAI GPT LLM capability (like Bing Chat and CoPilot). Under Microsoft's strained view of relevance, Microsoft is withholding financials for products like Microsoft Office that incorporate OpenAI's LLMs and *were specifically mentioned in Judge Stein's order. See* Dkt. 707 at 6. Microsoft must produce these financials.

---

[3] Class Plaintiffs raise this issue here to clarify that the proposed compromises for financial discovery related to GPT-4o and GPT-4o Mini should not relieve Microsoft of its obligation to provide financial discovery for the other models.

[4] The Court denied the motion without prejudice "pending filing of the proposed consolidated class action complaint." Dkt. 78 at 6.

Respectfully,

/s/ Davida Brook
Davida Brook
Susman Godfrey LLP
Counsel for The New York Times Company
News Plaintiffs' Liaison Counsel

/s/ Justin Nelson
Justin Nelson
Susman Godfrey LLP
Interim Lead Class Counsel

cc: All Counsel of Record (via ECF)