RUTTENBERG IP LAW
A PROFESSIONAL CORPORATION

January 29, 2026

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

cc: All Counsel of Record (via ECF)

Re: In re: OpenAI, Inc., Copyright Infringement Litigation, 1:25-md-03143-SHS-OTW
This Document Relates To: Ziff Davis, Inc. et al. v. OpenAI, Inc. et al., Case No. 25-cv-04315-SHS-OTW

Dear Judge Wang:

Enclosed, the parties (Ziff Davis and OpenAI) provide an update regarding the issues raised in Ziff Davis's motion to compel (Dkt. 1052). The Court directed the parties "to continue meeting and conferring and provide a status update no later than January 29, 2026." Dkt. 1170 at 2.

Ziff Davis's Position

As explained below, Ziff Davis believes that the parties have made considerable progress, but that further meeting and conferring is necessary. Giving the importance of these issues to Ziff Davis's claims and the fast-approaching discovery deadline, Ziff Davis suggests that the parties continue meeting and conferring and report back to the Court in the forthcoming joint discovery chart or as otherwise directed by the Court.

I. URL Searching

In its Motion, ZD asks that the Court "compel OpenAI to run URL searches across all training data with URL fields" by a date certain. Dkt. 1052 at 2. In its opposition, OpenAI asserts that it already "agreed to run" the searches requested by Ziff Davis. Dkt. 1115 at 2. The parties have made significant progress, but the issue is yet resolved.

Since the January 15 discovery conference, OpenAI has represented that it ran the URL searches against datasets identified in a January 22, 2026 email. OpenAI further represents that it produced those results on January 25, 2026. Ziff Davis and its experts are still evaluating what OpenAI's production contains. Further, Ziff Davis has asked OpenAI to confirm whether there are any additional datasets containing a URL field that could be searched in the same way but has not received confirmation.

OpenAI's statements (below) are incorrect. OpenAI insists that this Court's involvement is no longer necessary, and it contends that Ziff Davis supposedly only provided its domains before the Christmas holiday. Conveniently, OpenAI is pointing to a pared-back list that was intended to expedite the process and moot OpenAI's ever-changing objections. Indeed, Ziff Davis has been

seeking URL searches going back to August and September. OpenAI has only now started to provide these materials, given Ziff Davis's motion and the obligation to report back to the Court.

To continue this momentum, Ziff Davis requests that the Court direct that the parties continue to meet and confer and report back in the forthcoming discovery chart, or as otherwise directed by the Court.

II. N-Gram Searching

Ziff Davis's Motion also asks that the Court "compel OpenAI to run N-Gram searches by a date certain in substantially the same manner that OpenAI ran these searches for other News Plaintiffs."[1] Dkt. 1052 at 3. The parties continue to meet and confer on this point, but absent a firm agreement, Ziff Davis would seek an order compelling OpenAI to run the requested N-Gram searches.

As an update, OpenAI has now asked Ziff Davis to re-format its production of asserted works. While Ziff Davis believes this is unnecessary, Ziff Davis is attempting to comply with OpenAI's request in order to moot the dispute.

OpenAI insists that its "team has not been able to evaluate" Ziff Davis's data "to determine what sort of search may be appropriate or feasible." Without further "feedback" from its engineers, OpenAI's counsel claim that they "cannot tell [Ziff Davis] what we will or will not be able to do." However, OpenAI's counsel has emphasized that they "were able to reach agreements with the other News Plaintiffs."

As with the URL searching (discussed above), OpenAI contends (below) that the Court's intervention is not needed, and blames Ziff Davis for delays. This is not correct—indeed, despite repeated requests from Ziff Davis dating back for months, OpenAI would not commit to any searches. Even now, it will not commit to run these searches.

Ziff Davis requests that Court direct the parties to continue meeting and confer and report back in the forthcoming discovery chart, or as otherwise directed by the Court.

III. Reproduction of Source Code Printouts Provided to Other News Plaintiffs.

In its Motion, Ziff Davis seeks "production of source code…printouts" provided to any other News Plaintiffs. Dkt. 1052 at 1–2. In response, OpenAI argued that it had already ***agreed*** "to cross-produce source code printouts," and that motion practice regarding source code printouts was therefore unnecessary. Dkt. 1115 at 1.

To date, OpenAI has produced certain specific source code printouts that were identified by Ziff Davis. However, despite multiple inquiries, OpenAI has not confirmed whether it has produced any additional source code printouts to other News Plaintiffs beyond those that have been provided to Ziff Davis to date.

[1] Based on the searches conducted for other News Plaintiffs, the parameters involve a threshold of 50 6-gram matches, with a stride of 1 for content that is 200 words and shorter, and a stride of 2 for other content.

If OpenAI confirms that it will cross-produce any source code printouts provided to other News Plaintiffs moving forward (as suggested in response to Ziff Davis's motion), this issue will be resolved.

IV. Reproduction of Training Data Printouts Provided to Other News Plaintiffs.

In its Motion, Ziff Davis also seeks reproduction of training data printouts that were provided to other News Plaintiffs. Dkt. 1052 at 1-2. In its responses, OpenAI insists that such training data is plaintiff-specific, although it does not object to Ziff Davis reviewing such training data at the offices of counsel for other News Plaintiffs.

In the interim, OpenAI has produced other printouts of training data requested by Ziff Davis. Accordingly, Ziff Davis is reviewing those recently produced printouts to assess whether it still requires the training data printouts provided to other News Plaintiffs. To be clear, Ziff Davis does not agree with OpenAI's characterizations below.

With the Court's permission, Ziff Davis suggests that the parties continue to meet and confer and report back in the forthcoming discovery chart, or as otherwise directed by the Court.

V. Reproduction of Training Data Using Databricks/Snowflake

In its Motion, Ziff Davis also seeks production of training data using Databricks or Snowflakes, which (according to cited deposition testimony) are available in the ordinary course to OpenAI engineers, and which would enable more searches of the training data. Dkt. 1052 at 3. In response, OpenAI opposed that request (Dkt. 1115 at 3) and continues in its refusal.

For context, searching the OpenAI training data in its current form is exceedingly slow, particularly given the number of asserted works and the size of the databases. However, the database could be made available with other tools that would exponentially accelerate such searches, such as through Databricks, which is how OpenAI recently made available its output log productions. OpenAI has insisted that it would be expensive and time-consuming to re-produce the training data (or even a subset thereof) in Databricks.

Ziff Davis understands that other News Plaintiffs ultimately overcame these hurdles by having OpenAI conduct its own searches or URLs and N-grams, as requested above. *See*, 24-cv-4872, Dkt. 223 at 2 ("As directed by the Court, the parties met and conferred regarding proposed datasets and threshold values for the proposed 'N-Gram Search' of OpenAI training data. OpenAI will attempt to run an agreed N-gram search on an identified set of training datasets and produce the results"). With the benefit of those search results, Ziff Davis expects that (like other plaintiffs) it could conduct more limited and targeted searches of training data, using the existing architecture. Accordingly, to the extent OpenAI conducts the URL and N-gram searches requested by Ziff Davis (as discussed above), that may moot Ziff Davis's request for reproducing the training data in another, more readily accessible form such as Databricks.[2]

[2] Again, OpenAI's factual assertions below are wrong and perplexing. Ziff Davis's experts inspected the training data room months ago, but progress was exceedingly slow given the limitations of the review tools OpenAI made available. Further, whereas other News Plaintiffs received results of URL searches before conducting inspections, OpenAI refused to provide these materials to Ziff Davis until ***after*** its experts inspected the training data (and source code).

Accordingly, with the Court's permission, Ziff Davis suggests that the parties continue to meet and confer regarding the URL and N-Gram searches, and report back whether those productions obviate this aspect of Ziff Davis's motion.

OpenAI's Position

Ziff Davis's motion to compel, ECF 1052, remains unnecessary and a waste of the Court's resources, as evidenced by the updates OpenAI provides below about each of Ziff Davis's requests.

I. URL Searches

Ziff Davis provided the domain names that it wanted OpenAI to run just before the December holiday period. Before Ziff Davis filed its motion for the January discovery conference, OpenAI had already agreed to run the searches that Ziff Davis requested. OpenAI ran these searches promptly once the relevant engineers returned to the office. And OpenAI produced the results of these searches to Ziff Davis on January 25. Despite Ziff Davis's protestations above, these productions are complete, and no further Court action is needed (or was ever needed).

II. N-gram Searches

As explained in prior briefing, before OpenAI can evaluate what type of n-gram search might be appropriate (let alone actually run any n-gram searches), it needs Ziff Davis to produce the text of its copyrighted works in a format that can be extracted and processed for analysis. OpenAI asked for this data over four months ago, on September 17, 2025. In December, Ziff Davis finally responded by asking OpenAI to confirm that a particular file format would work (which OpenAI promptly confirmed).

But Ziff Davis did not even *start* producing the necessary data until just days before the motion deadline, and productions continued during the hearing (and *after*). OpenAI promptly reviewed samples of that data and confirmed that Ziff Davis did not export the data in a usable format. Last week, OpenAI shared additional details on how Ziff Davis can fix its productions to make this data usable for purposes of determining the appropriate parameters for any n-gram searches. Until OpenAI receives the corrected data, it cannot analyze the data to determine what sort of n-gram search might be appropriate, and it certainly cannot agree to run any n-gram searches without this data. The parties continue to meet and confer. Again, despite Ziff Davis's complaints above, no Court action is needed at this time because Ziff Davis has yet to even produce usable n-gram search data.

III. Source Code Printouts

Again, this issue was resolved before Ziff Davis filed its motion. Prior to Ziff Davis filing its motion, OpenAI had already agreed to produce the source code pages that Ziff Davis requested. That production was shipped on January 20. No further Court action is needed (or was ever needed).

As for the additional source code printouts provided to *other* Plaintiffs that Ziff Davis now seeks, Ziff Davis still has not followed the Protective Order's requirements for requesting these printouts, *see* ECF 367 at 17-18. If Ziff Davis makes a request for *specific* source code files—*i.e.*, a request demonstrating that Ziff Davis even knows what code files and number of pages it is requesting—OpenAI will consider that request in good faith in the same manner as before. What

OpenAI is *not* willing to do is throw out basic (and heavily negotiated) protections around source code. The Protective Order specifically prohibits Plaintiffs from using print requests as a means of "avoiding" an initial "electronic review" to determine what source code is actually *necessary*. *See* ECF 367 at 17. And the Protective Order also exempts source code from the cross-production requirement for other discovery materials. *See* ECF 367 at 7. These are just two of the many mechanisms built into the Protective Order to help ensure that OpenAI's source code—which is ***incredibly*** sensitive—is reproduced no more than is absolutely necessary.

When Ziff Davis has followed the rules in the Protective Order, and requested source code printouts with specificity, OpenAI has promptly printed it. But if Ziff Davis is asking for a modification of the Protective Order, then it must file a motion to amend that demonstrates that there is good cause why it cannot follow the same rules as everyone else. Since Ziff Davis has not filed that motion, no further Court action is needed.

IV. Training Data Printouts

OpenAI has completed several productions of training data that Ziff Davis identified for export in compliance with the terms of the governing Protective Order, *see* ECF 367 at 17-18. As for the additional training data printouts previously provided to *other* Plaintiffs, OpenAI remains willing to discuss possible cross-production of any *specific* training data exports that it produced to other Plaintiffs. But first, Ziff Davis must explain their relevance to Ziff Davis's claims or responsiveness to Ziff Davis's requests for production. Since Ziff Davis has not done so, and since these printouts are incredibly burdensome to cross-produce (as implicitly acknowledged by their exemption from the cross-production requirement for other discovery materials, *see* ECF 367 at 7), this discovery is not authorized under Rule 26. No Court action is needed.

V. Databricks

As explained in prior briefing, Ziff Davis inexplicably chose not to engage technology consultants or begin its training data inspections until the fact discovery period was nearly over. Then Ziff Davis's team wasted time looking for their content in datasets that contain source code and other irrelevant information (like non-text datasets), rather than web content. Ziff Davis, having lost so much time due to its own mismanagement of inspection capacity, asked OpenAI to build a new, bespoke training data platform for Ziff Davis alone, despite the fact that ***all*** other plaintiffs have successfully conducted training data inspections. OpenAI investigated the feasibility of Ziff Davis's requests prior to the January discovery hearing and confirmed that it would take several months to build such a platform.

Since the hearing, Ziff Davis asked if this engineering work could be dramatically shortened if it agreed to exclude irrelevant datasets (like the ones it had not realized it was scanning). OpenAI has now confirmed that this would ***not*** meaningfully reduce the burden because building a new platform has upfront engineering costs. In the meantime, OpenAI allowed Ziff Davis to leave its search scripts running on the virtual inspection machine (which imposes computing costs and burden on OpenAI), and it has already produced the responsive results.[3] No Court action is needed.

[3] Ziff Davis's assertion that OpenAI refused to provide URL search results before Ziff Davis's inspections is baffling; again, Ziff Davis did not pursue URL searches until late December. OpenAI promptly ran those searches and produced responsive results.

Respectfully submitted,

/s/ *Thomas Gorman*
Thomas E. Gorman
Keker, Van Nest & Peters LLP
Counsel for OpenAI Defendants

/s/ Guy Ruttenberg
Guy Ruttenberg
Ruttenberg IP Law, A Professional Corporation
Counsel for Ziff Davis Plaintiffs