UNITED STATES DISTRICT COURT
SOUTHERN DISTRICT OF NEW YORK

IN RE:

OPENAI, INC. COPYRIGHT INFRINGEMENT LITIGATION

This Document Relates To
All Putative Class Actions

23-cv-8292
23-cv-10211
24-cv-84
25-cv-3291
25-cv-3297
25-cv-3482
25-cv-3483

25-md-3143 (SHS) (OTW)

OPINION & ORDER

SIDNEY H. STEIN, U.S. District Judge.

OpenAI and Microsoft have separately moved to strike portions of the Consolidated Class Action Complaint entitled *David Baldacci et al. v. Open AI Inc. et al.*, No. 25-md-3143, on the basis that it includes new products and new causes of action in violation of the Court's May 22, 2025 order (Dkt. No. 60 at 2) that the consolidated complaint "is to include only the same products and causes of action that have already been asserted in the pending putative class actions" that were filed prior to consolidation. (Dkt. Nos. 333 ("OpenAI Mot."), 327 ("Microsoft Mot.").)

OpenAI asks that the Court strike allegations related to certain GPT models and that the Court strike what OpenAI characterizes as a new "download claim" for copyright infringement.

OpenAI is correct that the Court limited the models to be included in the Consolidated Class Action Complaint to GPT-3, GPT-3.5, GPT-3.5 Turbo, GPT-4, GPT-4 Turbo, GPT-4o, and GPT-4o Mini. Therefore, the Consolidated Class Action Complaint's allegations regarding any additional model will be stricken. OpenAI's request to strike allegations related to the so-called "download claim" is denied because the claim for relief and factual allegations underlying a download theory of infringement were present in the underlying class action complaints before consolidation.

Microsoft makes three requests: (1) that plaintiffs be required to clarify that this putative class action implicates only *OpenAI*-trained models, not models trained by *Microsoft*; (2) that plaintiffs be required to clarify that this action is limited to books and does not include other copyrightable works such as magazines, films, or songs, and (3) that allegations regarding models before GPT-3 and after GPT-4 Turbo be stricken as to Microsoft.

Plaintiffs agree with Microsoft that this action is limited to OpenAI-trained models and to books. Microsoft's request that allegations regarding models before GPT-3 and after GPT-4 Turbo be stricken as to Microsoft is granted as to allegations involving GPT-4V, GPT-4.5, GPT-5, and "derivatives" and "successors" of any model but is denied as to GPT-1, GPT-2, GPT-4o, and GPT-4o Mini, including any Microsoft products incorporating GPT-4o and GPT-4o Mini.

I. BACKGROUND

The Court assumes familiarity with the basic allegations and procedural history of this action.

On May 22, 2025, the Court ordered class plaintiffs to file a consolidated class action complaint. At that hearing, the Court directed that the forthcoming complaint should "include simply the products and causes of action that ha[d] already been asserted" in the prior pre-consolidation class action complaints and should contain no new models or causes of action. (May 22, 2025 Hearing Tr. at 37:16–19.) Specifically, the consolidated class action complaint would be limited to GPT-3, GPT-3.5, GPT-3.5 Turbo, GPT-4, GPT-4 Turbo, GPT-4o, and GPT-4o Mini. These models had been identified by Magistrate Judge Ona Wang in her discovery ruling in one of the underlying class actions, which the Court referenced at the May 22 conference. (*See id.* at 45:13–16.) That discovery ruling itself incorporated a list of models identified in one of OpenAI's interrogatory responses in the underlying action entitled *Authors Guild et al. v. OpenAI Inc. et al.*, No. 23-cv-8292 (the "*Authors Guild* action"). (*See* Dkt. No. 335-1 (OpenAI Interrogatory Response) at 7–8; *Authors Guild v. OpenAI Inc.*, No. 23-cv-8292, Dkt. No. 293 (Discovery Ruling) at 2.)[1]

The Consolidated Class Action Complaint, filed in June 2025, asserts three claims for relief against OpenAI and Microsoft: copyright infringement, vicarious copyright

[1] GPT-3 was not included in the OpenAI interrogatory response referenced by the Judge Wang, but the parties agree that GPT-3 was included in the underlying class action complaints and is therefore properly part of the Consolidated Class Action Complaint. (*See* Microsoft Mot. at 14; OpenAI Mot. at 6 n.15; Dkt. No. 456 ("Opp. to OpenAI Mot.") at 23 n.10.)

infringement, and contributory copyright infringement. (Dkt. No. 183 ("CCAC") ¶¶ 308–27.) The Consolidated Class Action Complaint includes allegations about the models identified by the Court at the May 22 hearing—GPT-3, GPT-3.5, GPT-3.5 Turbo, GPT-4, GPT-4 Turbo, GPT-4o, and GPT-4o Mini—but it also includes allegations regarding "GPT-4V," "GPT-4.5 (also known as Orion)," "the as-yet unreleased GPT-5 model," and all "derivatives" and "successors" of the models identified in the Consolidated Class Action Complaint. (*Id.* ¶ 5.)

II. ANALYSIS REGARDING OPENAI'S MOTION

First, OpenAI has moved to strike allegations related to GPT-4V, GPT-4.5, GPT-5, and "derivatives" and "successors" of any model.

Federal Rule of Civil Procedure 12(f) permits a court to "strike from a pleading . . . any redundant, immaterial, impertinent, or scandalous matter." Motions to strike are generally disfavored, *see Lipsky v. Corn United Corp.*, 551 F.2d 887, 893 (2d Cir. 1976), but district courts routinely strike material "where the court granted leave to amend for a limited purpose and the plaintiff filed an amended complaint exceeding the scope of the permission granted," *Palm Beach Strategic Income, LP v. Salzman*, 457 F. App'x 40, 43 (2d Cir. 2012) (summary order) (collecting cases). *See, e.g.*, *Miles v. City of New York*, No. 14-cv-9302, 2018 WL 3708657, at *5 (S.D.N.Y. Aug. 3, 2018); *Cummings v. City of New York*, No. 19-cv-7723, 2021 WL 1163654, at *3 (S.D.N.Y. Mar. 26, 2021).

The Court was pellucid when it directed that the consolidated complaint was to be limited to "simply the products and causes of action that ha[d] already been asserted" in the class action complaints that had been filed prior to the Court directing the parties to file a Consolidated Class Action Complaint. (May 22, 2025 Hearing Tr. 37:16–19, 39:24–25, 40:2.) To remove any doubt, the Court ordered that only the specific models identified in Judge Wang's discovery ruling in the underlying *Authors Guild* action—GPT-3, GPT-3.5, GPT-3.5 Turbo, GPT-4, GPT-4 Turbo, GPT-4o, and GPT-4o Mini—would be permitted in the consolidated class complaint, which the Court noted "should clear up what the consolidated amended complaint can have." (*Id.* 45:13–21.)

Despite this specific limit on the models that could be included in the Consolidated Class Action Complaint, plaintiffs included allegations concerning GPT-4V, GPT-4.5, GPT-5, and "derivatives" and "successors" of any model in the Consolidated Class Action Complaint. These allegations exceed the scope of the Court's order. Therefore, allegations regarding these models will be stricken from the Consolidated Class Action Complaint in paragraphs 5, 81 to 90, 96, 127, and 161. *See Palm Beach Strategic Income*, 457 F. App'x at 43.

Second, OpenAI has moved to strike what it characterizes as a new "download claim" for copyright infringement. OpenAI does not appear to contest that the

Consolidated Class Action Complaint includes no new factual allegations regarding OpenAI's download and reproduction of plaintiffs' books and asserts only three claims for relief—copyright infringement, vicarious copyright infringement, and contributory copyright infringement—that were asserted in the class complaints filed prior to consolidation. OpenAI contends, however, that the Consolidated Class Action Complaint adds new claims because it does not link its factual allegations regarding OpenAI downloading plaintiffs' books to OpenAI's training of its LLMs, while the prior class complaints allegedly uniformly connected allegations about OpenAI downloading plaintiffs' books to OpenAI's use of the books to train its LLMs (so-called "training-based claims"). (OpenAI Mot. at 2, 4, 12.)

It is well settled that "a complaint need not pin plaintiff's claim for relief to a precise legal theory," *Skinner v. Switzer*, 562 U.S. 521, 530 (2011), and that "the essence of a cause of action is found in the facts alleged and proven by the plaintiff, not the particular legal theories articulated" in the complaint, *Oneida Indian Nation of N.Y. v. County of Oneida*, 617 F.3d 114, 139 (2d Cir. 2010). Indeed, "[f]ederal pleading is by statement of claim upon which relief may be based rather than by legal theory," and plaintiffs "need only plead facts showing entitlement to relief and are not required to specify the legal theory upon which the claim is based." *Baker v. Latham Sparrowbush Assocs.*, 808 F. Supp 981, 989 (S.D.N.Y. 1992) (citation omitted); *see Davis v. Rumsey Hall Sch., Inc.*, No. 20-cv-1822, 2023 WL 6379305, at *9 (D. Conn. Sep. 29, 2023) ("Plaintiff need not broadcast its strategy at the pleading stage, and need not spell out each legal theory on which he might rely to prove his claims, when varying theories can be supported by his factual allegations.")

Here, the prior class complaints asserted a cause of action for copyright infringement and alleged that OpenAI impermissibly downloaded and reproduced plaintiffs' books.[2] The fact that many of the allegations in the prior class complaints suggested that the ultimate purpose of the reproduction was to train OpenAI's LLMs is

[2] *See, e.g.*, *Authors Guild v. OpenAI Inc.*, No. 23-cv-8292, Dkt. No.69 ¶ 100 ("OpenAI has admitted that 'training' LLMs 'require[s] large amounts of data,' and that 'analyzing large corpora' of data 'necessarily involves first making copies of the data to be analyzed." (alteration in original)); *id.* ¶ 415 ("Defendants infringed . . . by, among other things, reproducing the works . . . used to train their artificial intelligence models."); *Alter v. OpenAI Inc.*, No. 23-cv-10211, Dkt. No. 26 ¶ 87 ("OpenAI . . . chose to copy a massive corpus of copyrighted books right from the internet, almost certainly from illegal sources . . . without even paying for an initial copy."); *Chabon v. OpenAI, Inc.*, No. 25-cv-3291, Dkt. No. 11 ¶ 39 (alleging that OpenAI's "practice necessarily leads OpenAI to capture, download, and copy copyrighted written works"); *id.* ¶ 70 ("Whether Defendants violated the copyrights of Plaintiffs and the Class when they downloaded and copied Plaintiff's and the Class's copyrighted books.").

not dispositive. A complaint need only provide defendants with fair notice of the claims alleged against them to comply with Federal Rule of Civil Procedure 8, *see Greicus v. Liz Claiborne, Inc.*, No. 00-cv-9518, 2002 WL 244598, at *3 (S.D.N.Y. Feb. 20, 2002), and "[f]actual allegations alone are what matters," *Albert v. Carovano*, 851 F.2d 561, 571 n.3 (2d Cir. 1988). The prior complaints' factual allegations adequately put OpenAI on notice of plaintiffs' claims against it for copyright infringement based on its alleged download and reproduction of plaintiffs' books. Therefore, the Consolidated Class Action Complaint's claim of infringement based on OpenAI's alleged download and reproduction of plaintiffs' books is not a new claim in violation of the Court's May 22 order. The Court's May 22 order not to add new claims in the Consolidated Class Action Complaint did not lock class plaintiffs into the legal theories intimated in the prior class complaints. Therefore, OpenAI's motion to strike the so-called "download claim" from the Consolidated Class Action Complaint is denied.

III. ANALYSIS REGARDING MICROSOFT'S MOTION

First, Microsoft requests that plaintiffs be required to clarify that this action implicates only *OpenAI*-trained models, not models trained by *Microsoft*. (Microsoft Mot. at 12–13.) There is no dispute among the parties that this action does not include "Microsoft-trained LLMs" (*see* Dkt. No. 453 ("Opp. to Microsoft Mot.") at 4), so no clarification is needed and Microsoft's request is dismissed.

Second, Microsoft requests that plaintiffs be required to clarify that this action is limited to books rather than other copyrightable works. (Microsoft Mot. at 17–18.) No such clarification is necessary because, reading the Consolidated Class Action Complaint as a whole, it is clear that this action is limited to books (*see* CCAC ¶¶ 2–3, 6, 9, 13, 14 ("Plaintiffs seek to represent a Class of *book* copyright holders . . ." (emphasis added)), 67–68, 74, 97, 116, 119–20, 171–292) and plaintiffs concede that "this case is about books" (Opp. to Microsoft Mot. at 9). Microsoft's request is dismissed.

Finally, Microsoft requests that the Consolidated Class Action Complaint's allegations regarding all GPT models other than GPT-3, GPT-3.5, GPT-3.5 Turbo, GPT-4, and GPT-4 Turbo—*i.e.*, allegations concerning GPT-1, GPT-2, GPT-4o, GPT-4o Mini, GPT-4V, GPT-4.5, GPT-5, and "derivatives" and "successors" of any model—be stricken as to Microsoft.

There is no claim against Microsoft based on GPT-1 and GPT-2 given the fact (and plaintiffs' agreement) that the Consolidated Class Action Complaint "does not allege that Microsoft was involved with [GPT-1 and GPT-2]." (Opp to Microsoft Mot. at 7.) Microsoft's motion to strike allegations against it based on GPT-1 and GPT-2 is therefore denied.

As explained above, the Court will strike allegations regarding GPT-4V, GPT-4.5, GPT-5, and "derivatives" and "successors" of any model from paragraphs 5, 81 to 90, 96, 127, and 161 of the Consolidated Class Action Complaint.

That leaves only allegations involving Microsoft based on GPT-4o and GPT-4o Mini. Microsoft contends that, while OpenAI conceded that GPT-4o and GPT-4o Mini were part of this action prior to the consolidation of the class actions, Microsoft never acknowledged that these models were within the scope of this litigation as to Microsoft, so Microsoft should not be held to OpenAI's concession. (*See* Microsoft Mot. at 14 n.8; Dkt. No. 496 ("Microsoft Reply") at 5 n.3.) Microsoft slices the baloney too thin. 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. The Court expressly stated on May 22 which models are within the scope of this action, and those models include GPT-4o and GPT-4o Mini. Microsoft's motion to strike allegations related to GPT-4o and GPT-4o Mini is denied. Microsoft's further request that the Court expressly limit the scope of the MDL to exclude Microsoft products that incorporated GPT-4o and GPT-4o Mini—such as the Microsoft Office suite of products and Microsoft Copilot—is also denied. The Court notes, however, that discovery into these or other Microsoft products, like all discovery, is subject to the proportionality standard of Federal Rule of Civil Procedure 26(b)(1).

IV. Conclusion

In sum, the allegations in the Consolidated Class Action Complaint regarding GPT-4V, GPT-4.5, GPT-5, and "derivatives" and "successors" of any model in paragraphs 5, 81 to 90, 96, 127, and 161 are stricken. The motions to strike are otherwise denied or dismissed.

Dated: New York, New York
October 27, 2025

SO ORDERED:

Sidney H. Stein, U.S.D.J.