KEKER VAN NEST & PETERS LATHAM & WATKINS LLP MORRISON FOERSTER

August 4, 2025

VIA ECF

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

cc: *All Counsel of Record (via ECF)*

Re: ***OpenAI's Opposition to Class Plaintiffs' Motion Regarding Privilege Over Reasons for Removal of Books1 and Books2 Datasets***
In re OpenAI Inc. Copyright Infringement Litigation, No. 1:25-md-03143-SHS-OTW
This Document Relates to the following Class Cases: Case No. 1:23-cv-08292, Case No. 1:23-cv-10211, Case No. 1:24-cv-00084, Case No. 1:25-cv-03291, Case No. 1:25-cv-03482, Case No. 1:25-cv-03483.

Dear Judge Wang:

Plaintiffs' motion is not a genuine effort to obtain discovery about the Books1 and Books2 datasets because OpenAI has already provided all relevant discovery. Rather, it is an attempt to paint OpenAI in a negative light and obtain an unwarranted tactical litigation advantage. But the relevant facts about Books1 and Books2 are not in dispute. No ChatGPT model currently available has been trained on books from Library Genesis ("LibGen"), and GPT-3.5—released in 2022 and no longer used for ChatGPT by early 2023—was the last model trained on such data. OpenAI removed the Books1 and Books2 datasets in July 2022—a full year *before* any plaintiff filed litigation challenging OpenAI's training on books—and has since recovered and produced those datasets to Plaintiffs. Plaintiffs now seek discovery *not* into the *content* of those datasets or when they were used (since they have all that information), but into the *reasons* for the removal. Yet, as OpenAI has consistently explained, those reasons are privileged because they were based on legal advice. As OpenAI has informed Plaintiffs, there are *no* non-privileged reasons for the removal, and OpenAI does not intend to rely on the advice of counsel either affirmatively or in rebuttal. Plaintiffs have all the facts they need regarding the Books1 and Books2 datasets.

Plaintiffs nonetheless seek an extraordinary order compelling production of OpenAI's privileged communications—or, alternatively, precluding OpenAI from "contradicting statements Plaintiffs

may make" regarding the reasons for removal or OpenAI's "states of mind." *See* ECF 413 at 2, 4 (emphasis added). There is no legal basis for such relief. OpenAI has not waived privilege and does not intend to rely on advice of counsel. It is not wielding privilege as a sword and shield. Plaintiffs' invocation of the crime-fraud exception is equally meritless. That exception is sparingly applied and requires a substantial showing that Plaintiffs come nowhere close to establishing. Plaintiffs' theory would have this Court address—on a discovery motion—the question at the heart of the case: whether training large language models on copyrighted works was fair use. That will be addressed at summary judgment or trial. Indeed, the court in *Kadrey v. Meta* rejected crime-fraud and the argument that downloading from shadow libraries—the conduct Plaintiffs claim as the predicate "crime"—was categorically unlawful. To the contrary, that court held on summary judgment that reproduction and use of training data from so-called shadow libraries ***was*** fair use. And another court suggested that the better practice is to dispose of such data once the company ceases training on it.

Argument

As OpenAI has previously explained (ECF 381, 1:23-cv-08292)—and Plaintiffs do not meaningfully challenge—"confidential communications between [OpenAI] and [its] counsel" regarding the decision to remove Books1 and Books2 were "made for the purpose of obtaining or providing legal assistance," and therefore are privileged. *See In re Cnty of Erie*, 473 F.3d 413, 418 (2d Cir. 2007). Plaintiffs argue that OpenAI has waived privilege or that the crime-fraud exception applies. Both arguments are wrong.

A. OpenAI has not waived the attorney-client privilege.

Plaintiffs claim OpenAI has waived privilege based on the false premise that OpenAI is using privilege "as a shield and a sword." ECF 413 at 4. But OpenAI is *not* relying on the advice of counsel either affirmatively or in rebuttal. *See* ECF 413-11 at 1.[1] Because OpenAI is not "assert[ing] a claim or defense that [it] intends to prove by use of the privileged materials," and "the privileged materials [are not] indispensable to [OpenAI's] claims or defenses," an "at-issue waiver [has not] occur[ed]." *GLD3, LLC v. Albra*, 2024 WL 4471672, at *7 (S.D.N.Y. Oct. 11, 2024).[2] No sword is being drawn; only the shield of privilege is being maintained.

OpenAI has not placed privileged communications "at issue." OpenAI has revised the 2024 letters that stated (in passing) that removal was "due to" non-use. *See* ECF 188. Those letters were intended to inform Plaintiffs that OpenAI had ceased using the datasets (and then removed them) long before any litigation. They did not—nor were intended to—disclose the *reasons* for the removal, which OpenAI has always maintained are privileged. Contrary to Plaintiffs'

[1] OpenAI has not yet answered the Consolidated Amended Complaint. However, as OpenAI has told Plaintiffs, it does not intend to assert any affirmative defense that puts its state of mind at issue.

[2] *Arista* offers no support for Plaintiffs' position. That unpublished decision involved a post-discovery motion *in limine* where the defendant affirmatively invoked a good faith defense. Here, OpenAI has not put its state of mind at issue and has disclaimed reliance on advice of counsel.

assertion, OpenAI has *never* represented to the Court that the *fact* that the datasets were not in use at the time of their removal means that OpenAI does not claim privilege over the *reasons* for removal. *See* ECF 413-9 at 70:2–3, 70:9–12 ("just because there may be an element of this issue that is not privileged, doesn't mean that every aspect of the issue is not privileged.")[3]

Because there has been no waiver, there is no basis to compel production of privileged communications or to bar OpenAI from responding to "statements Plaintiffs *may* make" regarding the reasons for the removal. Regardless, evidentiary rulings are reserved for the trial judge. Ex. A (5/27/2025 Hr'g Tr., Afternoon Session) at 41:16–19. And Plaintiffs cite no authority entitling a party to flood the jury with rank speculation when an opponent properly asserts privilege.

B. The crime-fraud exception does not apply.

Nor have Plaintiffs come close to satisfying their heavy burden to establish the crime-fraud exception. There has been no crime or fraud; only fair use. Plaintiffs identify no evidence that OpenAI's privileged communications about the reasons for removing the datasets were made "in furtherance of contemplated or ongoing criminal or fraudulent conduct" and were "*intended* in some way to facilitate or to conceal the criminal activity." *In re Grand Jury Subpoenas Dated Sept. 13, 2023*, 128 F. 4th 127, 141–42 (2d Cir. 2025).

Whether the use of copyrighted works to train large language models qualifies as fair use is central to this case, to be addressed at summary judgment or trial, not on a discovery motion. Plaintiffs' mere invocation of "crime-fraud" does not justify piercing the privilege. Courts "should exercise considerable caution" when asked to pierce the privilege at this early stage, particularly where the claim is weak and touches on the merits. *In re Omnicom Grp., Inc. Sec. Litig*., 233 F.R.D. 400, 407 (S.D.N.Y. 2006).

Plaintiffs cite *Kadrey v. Meta*, but Judge Chhabria's rulings support *preserving* privilege. He rejected a crime-fraud exception during discovery (Ex. B (2/27/25 Hr'g Tr.) at 16:18–18:1) and held at summary judgment that the torrenting of books from a shadow library is fair use: "Because Meta's ultimate use of the plaintiffs' books was transformative, so too was Meta's downloading of those books" from LibGen. *See Kadrey v. Meta Platforms, Inc.*, 2025 WL 1752484, at *12 (N.D. Cal. June 25, 2025). The same logic applies here. If the initial download was lawful, later removing those datasets cannot be a crime, and communications about the removal cannot further or conceal a nonexistent crime.

Plaintiffs' spoliation theory fares no better. It wrongly assumes that OpenAI had a duty to preserve simply because its 2019 PTO submission acknowledged "substantial uncertainty" around fair use in AI training. But legal uncertainty does not trigger a preservation duty; only reasonably foreseeable litigation does. *In re Keurig Antitrust Litig.*, 341 F.R.D. 474, 532 (S.D.N.Y. 2022). Here, the datasets were removed a full year before any lawsuit was filed

[3] Plaintiffs' claimed inability to "investigat[e] OpenAI's non-infringement claims" because they "cannot inspect several versions of the[] datasets" is false, ECF 413 at 4—OpenAI has located and produced the datasets used to train GPT-3. Regardless, it is irrelevant to *waiver*.

challenging OpenAI's training on books and before OpenAI had even decided to launch ChatGPT. *See* ECF 424-1 (July 31, 2025 Kwon Decl.) ¶¶ 7–8. No duty to preserve existed. The involvement of outside counsel does not suggest otherwise. Responsibly consulting with counsel to assess risk or ensure compliance does not trigger a preservation obligation. *See Cruz v. G-Star Inc.*, 2019 WL 4805765, at *10 (S.D.N.Y. Sept. 30, 2019).

Even if Plaintiffs could conjure some novel preservation duty, their theory still fails. Removal (and recovery) of datasets is not a crime. In *Bartz v. Anthropic*, the court suggested that books downloaded from LibGen should be removed once no longer in use. 2025 WL 1741691, at *16 (N.D. Cal. June 23, 2025). Ironically, Plaintiffs' own counsel argued in that case that "hoard[ing]" such data was itself infringement. *See* ECF 275, *Bartz v. Anthropic*, Case 3:24-cv-05417-WHA (N.D. Cal., July 28, 2025). Yet here, they claim removal is a "crime." That contradiction underscores the incoherence of their theory: they cannot argue that it is unlawful to *retain* LibGen data in one case, only to argue that it is unlawful to *remove* it in another. The Court should reject this manufactured Catch-22 and deny Plaintiffs' motion.

Respectfully,

KEKER, VAN NEST & PETERS LLP[4]	LATHAM & WATKINS LLP	MORRISON & FOERSTER LLP
/s/ R. James Slaughter	*/s/ Herman H. Yue*	*/s/ Rose S. Lee*

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