UNITED STATES DISTRICT COURT
SOUTHERN DISTRICT OF NEW YORK

IN RE: OPENAI, INC., COPYRIGHT INFRINGEMENT LITIGATION This document relates to: 23-cv-8292 23-cv-10211 24-cv-84 25-cv-3297 25-cv-3482 25-cv-3483	25-md-3143 (SHS) (OTW) Hon. Sidney H. Stein Hon. Ona T. Wang. **ORAL ARGUMENT REQUESTED**

[PROPOSED] REPLY IN SUPPORT OF OPENAI DEFENDANTS' OBJECTION TO DISCOVERY ORDER AT ECF NO. 846 PURSUANT TO FEDERAL RULE OF CIVIL PROCEDURE 72(A)

TABLE OF CONTENTS

TABLE OF AUTHORITIES

Page

CASES

STATUTE AND RULES

OTHER AUTHORITY

I. INTRODUCTION

Since class plaintiffs first asked about OpenAI's reasons for deleting Books1 and Books2 in January 2025, OpenAI has consistently rebuffed plaintiffs' inquiries by asserting attorney-client privilege. Plaintiffs do not seriously dispute that OpenAI's reasons for deletion are privileged. *See* OpenAI Br. 9-11. Instead, plaintiffs defend Magistrate Judge Wang's conclusion that OpenAI intentionally waived that privilege. ECF 846 ("Order"). But nothing in plaintiffs' opposition salvages the Order's misconstruction of the record or its failure to apply controlling precedent.

Plaintiffs first contend that OpenAI waived privilege by partially disclosing one of its privileged reasons for deletion. Plaintiffs' only evidence is a solitary phrase repeated in two off-topic discovery letters suggesting that OpenAI deleted the datasets "due to their non-use." In context, those later-retracted statements disclose no privileged information. And they certainly do not clear the high bar for intentional waiver, meaning that Federal Rule of Evidence 502's protections against inadvertent waiver apply.

Plaintiffs next contend that OpenAI waived privilege by changing its position on privilege. But OpenAI has always insisted that all its reasons are privileged. And because OpenAI has both produced recovered copies of Books1 and Books2 and denied an intent to "rely on any non-privileged reason for the removal of the … datasets in th[is] litigation," ECF 504 at 2, there is no risk whatsoever of a selective presentation of the evidence. In any event, plaintiffs fail to identify a single case holding that a party waived privilege through litigation conduct that did not amount to a violation of the federal or local rules—and Judge Wang found no rules violation here.

Finally, plaintiffs reiterate Judge Wang's striking assertion that OpenAI waived privilege merely by denying allegations of willfulness. Plaintiffs suggest that OpenAI's privileged reasons may be "probative" of OpenAI's state of mind during the alleged infringement. ECF 995 ("Opp.") 1, 16-17, 25. But even supposing that is true, controlling precedent is crystal clear: the fact that

"privileged information may be in some sense *relevant*" to a lawsuit is not enough for waiver. *In re County of Erie* (*Erie II*), 546 F.3d 222, 229 (2d Cir. 2008).

This Court should set aside the Order and return this case's focus to the copyright claims at issue.

II. ARGUMENT

A. OpenAI Did Not Waive Privilege

1. OpenAI Did Not Waive Privilege by Partial Disclosure

a. OpenAI has not waived privilege by partial disclosure for the simple reason that it has *never* disclosed *any* of its privileged reasons for deleting the Books1 and Books2 datasets. OpenAI has consistently maintained that all its reasons for deleting the datasets are privileged. *See, e.g.*, OpenAI Br. 6-8, 12 (collecting examples).

Class plaintiffs' counterargument myopically seizes on stray remarks in two off-topic discovery letters. Opp. 11-12. In context, however, the letters' use of the phrase "due to their non-use" refers only to the *fact* that the datasets were not in use when OpenAI deleted them. *See* OpenAI Br. 13. As plaintiffs admit, the letters addressed efforts to "secure production of the datasets" themselves, not OpenAI's reasons for deleting them. Opp. 12. There is no reason to think that in addressing that discrete discovery question, OpenAI intended to disclose its sensitive and confidential deliberations. Instead, when OpenAI said it deleted the datasets "due to" non-use, OpenAI expressed only that non-use "provided the occasion for discussions with in-house counsel about whether OpenAI should remove the datasets." OpenAI Br. 13.

There is nothing "incoherent" about the proposition that "an occasion for a decision is not a reason for the decision." *Contra* Opp. 11 & n.4. When a judge decides to hire a law clerk, it is in some sense "due to" the clerk's submission of an application during the judge's hiring period. But the mere submission of the application is not the reason the clerk gets the job—standout

grades, experience, and interview skills are what distinguish hopeful applicants from successful law clerks. In just the same way, "non-use" of the datasets set the stage for their deletion, even though "non-use" was not a *reason* for deletion.

Plaintiffs fault OpenAI for supposedly failing to "clarify" what "non-use" meant before Judge Wang. Opp. 11. True, parties may not raise new issues "for the first time in objections to a magistrate's discovery order." Opp. 10 (citation omitted). But whether OpenAI's reference to "non-use" waived privilege is not a new issue. As Judge Wang recognized, OpenAI argued that the phrase "due to non-use" did not "describe a 'reason' for the deletion of [the datasets]." Order 15 n.14. Regardless, parties may—and, in fact, *must*—challenge the magistrate judge's reasoning in their Rule 72 objections. Second Circuit precedent requires parties to address the "specific legal conclusion[s] in the [magistrate judge's] report and recommendation"—warmed over versions of "previously filed papers or arguments" are "[in]adequate." *Miller v. Brightstar Asia, Ltd.*, 43 F.4th 112, 120-21 (2d Cir. 2022) (citation omitted). Here, OpenAI's objections explain why Judge Wang erred in reading the letters to disclose a privileged reason for deletion. *See* Order 16.

Plaintiffs themselves raise a new argument, asserting for the first time that OpenAI *was* still using Books1 and Books2 to train GPT models when it deleted the datasets in 2022. Opp. 12. In fact, the newly unearthed messages plaintiffs rely on confirm that by mid-2022, OpenAI "d[idn't] really need the datasets anymore" and had stopped using them for "any aspect of embedding research." Opp. Ex. 2 at 4, 7. At most, plaintiffs' exhibits confirm that OpenAI used an extremely large mix of datasets called "[redacted]"—containing Books1 and Books2, plus a host of other datasets like Common Crawl—to train the GPT-3 model in 2021 and then continued to reference [redacted] in "[redacted]" on GPT-3 in 2022. Opp. Ex 2 at 6 (emphasis added). Nothing in plaintiffs' exhibits disproves that OpenAI stopped training new models using

the datasets in 2021. OpenAI Br. 5. Nor do the exhibits show that OpenAI somehow waived privilege.

b. Even if OpenAI *had* disclosed a privileged reason for deletion, Federal Rule of Evidence 502 would foreclose waiver based on partial disclosure. Plaintiffs do not dispute that Rule 502 applies here, instead objecting that OpenAI failed to mention Rule 502 in its briefing before Judge Wang. Opp. 15. But Judge Wang cited Rule 502 in her articulation of the waiver standard, Order 13-14, and OpenAI is free to challenge Judge Wang's misapplication of that standard, *supra* at 3.

Rule 502 bars waiver here. To start, waiver is unwarranted under Rule 502(b) because any disclosure was inadvertent and OpenAI took "reasonable steps to rectify the error" by filing revised versions of the discovery letters omitting the "due to … non-use" phrasing. OpenAI Br. 15-16 (citation omitted). Plaintiffs say that OpenAI waited too long to file the revised letters. Opp. 15-16. But plaintiffs themselves identify the obvious reason for the wait: Between May and June 2025, the parties were "litigat[ing] … whether the letters waived privilege." Opp. 15-16. And it was not until the parties' positions had crystalized, including at a May 27 hearing before Judge Wang, that revision became necessary. In any event, "inordinate delay" justifies "waiver" under Rule 502(b) and there is no suggestion of any such foot-dragging here. *Valentin v. Bank of N.Y. Mellon Corp.*, 2011 WL 1466122, at *3 (S.D.N.Y. Apr. 14, 2011) (citation omitted).

Even if the discovery letters had *intentionally* disclosed privileged information, as plaintiffs suggest, *see* Opp. 16-17, Rule 502(a) would still forbid any finding of waiver. For intentional disclosures, Rule 502(a) compels waiver over additional privileged information concerning the same "subject matter" only if fairness requires it. Fed. R. Evid. 502 note. There is no risk of any unfairness here. Plaintiffs complain that OpenAI has prejudiced them by wielding privilege to

"foreclose any further inquiry" into OpenAI's reasons for deletion which, plaintiffs say, might be useful in proving that OpenAI acted willfully. Opp. 16. But Rule 502(a) is a tool to avoid inequities, not a cudgel for plaintiffs to use to extract useful information. Lack of access to privileged communications cannot possibly count as prejudice in and of itself—otherwise there would be prejudice in *every* partial-disclosure case. Instead, plaintiffs must specifically explain why further disclosure is warranted to avoid a "selective and misleading presentation of the evidence" to the factfinder. Fed. R. Evid. 502(a) note. Plaintiffs cannot do that here because OpenAI has disclaimed any intent to rely on evidence about its reasons for deleting the datasets. *See* OpenAI Br. 16.

2. OpenAI Did Not Waive Privilege by Changing Positions

Throughout this litigation, OpenAI's consistently asserted that its *reasons* for deletion are privileged, even though the *fact* of nonuse is not. OpenAI Br. 17 (citing record evidence). Plaintiffs nevertheless make the remarkable assertion that OpenAI has "changed its privilege claims … at least five times." Opp. 5-7, 18. That revisionist account depends on ignoring the critical distinction between facts and reasons.

a. Presumably appreciating that the fact/reason distinction renders OpenAI's statements on privilege entirely consistent, plaintiffs attack the distinction itself. In plaintiffs' view, OpenAI has waffled on whether "non-use" is a *fact* or a *reason* because OpenAI's counsel once described "nonuse of the data set to be [the] cause of the deletions." Opp. 19 (emphasis altered) (quoting ECF 109 at 70:1-6). Not so. As discussed, the fact that OpenAI "d[idn't] really need the datasets anymore" after it stopped using them, Opp. Ex. 2 at 5, was a "cause" of deletion in the sense that it provided the occasion for OpenAI's discussions with in-house counsel. OpenAI thus hewed to its longstanding position when it told Judge Wang that OpenAI was "not asserting privilege" over the fact of "nonuse." ECF 109 at 70:4-6.

Plaintiffs next protest that OpenAI has not "consistently allowed for discovery into the fact of non-use." Opp. 13-14, 19. Plaintiffs point to assertions of privilege at Michael Trinh's July 25 deposition. As the transcript shows, however, counsel invoked privilege only to guard against plaintiffs' efforts to elicit disclosure of OpenAI's privileged reasons. ECF 413-1 at 107:6-17. Counsel can hardly be faulted for instructing the witness on privilege given plaintiffs' aggressive efforts to gin up purported privilege waivers.

b. Even if OpenAI had changed positions, any shift would not meet the standard for waiver. Waivers based on litigation conduct are reserved for "willful[l]" or "cavalier" violations of the federal or local rules—not small changes in a party's position. *In re Honeywell Int'l, Inc. Sec. Litig.*, 230 F.R.D. 293, 299 (S.D.N.Y. 2003) (citation omitted). In *Honeywell*, for example, a party violated Rule 26 by seeking to "reengineer privilege logs to align their privilege assertions with their legal arguments." *Id.* at 299. OpenAI has not done anything like that. At most, OpenAI imprecisely delineated the metes and bounds of a specific privilege over a specific topic.

Plaintiffs contend that a "majority" of courts in this district strictly find waiver based on *any* violation of the federal or local rules, with a "minority" instead asking whether a violation was willful. Opp. 17-18. Though "some courts" follow each approach, *In re Honeywell Int'l.*, 230 F.R.D. at 299, plaintiffs offer no citation for their claim that a "majority" of courts have adopted the strict position. And a strict rule is hard to square with Second Circuit precedent looking to whether a rules violation was "cavalier" or "flagran[t]." *See In re DG Acquisition Corp.*, 151 F.3d 75, 84 (2d Cir. 1998); OpenAI Br. 18 n.3 (collecting authorities). A strict rule would also improperly chill privileged speech and harshly punish privilege holders (clients) whenever their lawyers *inadvertently* fail to comply with the rules.

In all events, because even plaintiffs' preferred rule does not compel a finding of waiver here, the Court need not pick a side in this debate now. Magistrate Judge Wang did not find that OpenAI violated any federal or local rule. And plaintiffs fail to identify a single case in which a court has imposed a privilege waiver based on litigation conduct that did not amount to a rules violation. *See* Opp. 17-20. *S.E.C. v. Yorkville Advisors, LLC* found waiver based on a "fail[ure] to produce in a timely manner a privilege log that *complied with the applicable rules*." 300 F.R.D. 152, 167 (S.D.N.Y. 2014) (emphasis added). Likewise, *In re DiDi GlobalS Inc. Securities Litigation* found waiver where a party placed documents on a privilege log "without conducting a privilege review," which, in the court's view, "fl[ew] in the face of [Federal Rule of Civil Procedure] 26 and Local Civil Rule 26.2." 2025 WL 1735412, at *7 (S.D.N.Y. June 23, 2025); *see also id.* *7 & n.58 (noting that "minor defects" in privilege assertions do not justify waiver). As for *Digital-Vending Services International, LLC v. University of Phoenix Inc.*, the misconduct at issue was so egregious that the court felt compelled to "attac[h]" to its order a copy of Virginia's "Principles of Professionalism for Virginia Lawyers." 2010 WL 11450510, at *1 n.2 (E.D. Va. Apr. 22, 2010). OpenAI's actions are worlds away from that kind of conduct.

3. OpenAI Did Not Waive Privilege by Putting Privileged Communications at Issue

a. Magistrate Judge Wang held that OpenAI waived privilege simply by denying plaintiffs' allegations that OpenAI *willfully* infringed their copyrights. That holding ignores clear Second Circuit precedent limiting at-issue waivers to circumstances in which parties seek to "*rely* on privileged advice from … counsel to make [a] claim or defense." *Erie II*, 546 F.3d at 229. And as plaintiffs concede, OpenAI has *not* sought to rely on privileged advice in support of any claim or defense. Opp. 22.

Plaintiffs instead suggest the standard for at-issue waiver is satisfied here because OpenAI's reasons for deleting the datasets might "bea[r] on whether OpenAI had knowledge that its [alleged] infringing conduct was wrongful," Opp. 21, or shed light on "the lawfulness and propriety of [OpenAI's] conduct," Opp. at 25. But the Second Circuit rejected exactly that approach in *Erie II* when it held that a party does not put "protected information at issue by making it relevant to the case." 546 F.3d at 229 (citation omitted). Controlling precedent could not be clearer: The fact that "privileged information may be in some sense *relevant*" to a lawsuit is not enough, nor is a "mere indication of a claim or defense." *Id.* After all, attorney-client privilege routinely "renders relevant information undiscoverable." *United States v. Mejia*, 655 F.3d 126, 132 (2d Cir. 2011) (citation omitted).

To be sure, parties can put privileged communications at issue without expressly relying on legal advice. An affirmative "good faith defense" waives privilege where the defendant had "the advice of counsel *on the very issue* on which [he] asserts good faith …." *Scott v. Chipotle Mexican Grill, Inc*., 67 F. Supp. 3d 607, 618 (S.D.N.Y. 2014) (emphasis added); *accord* OpenAI Br. 23-24 (collecting cases). But as plaintiffs reluctantly admit, OpenAI has not asserted an affirmative good faith defense in this case. Opp. 25 n.6. Nor is an assertion of good faith a necessary implication of OpenAI's denial of willfulness. *See* OpenAI Br. 21-22. In any case, the privileged communications here do not concern the *very conduct* at issue—*i.e.*, OpenAI's alleged infringing *use* of plaintiffs' copyrights in training the GPT models. Instead, OpenAI has simply asserted privilege over its reasons for deleting datasets *years after* the alleged infringing conduct in this case. *See supra* at pp. 3-4; OpenAI Br. 25.

Plaintiffs fail to identify any case in which a court has waived privilege based on a defendant's bare denial of a willfulness allegation. *Oklahoma Firefighters Pension & Retirement*

System v. Musk found an at-issue waiver where defendants intended to present a "good faith defense," yet sought to "ba[r] [the] plaintiff from examining" legal "advice … about th[e] *very conduct*" at issue. 2025 WL 2982003, at *6 (S.D.N.Y. Oct. 23, 2025) (emphasis added) (cited Opp. 24). Tellingly, the court cited *Erie II* for the normal rule that "it [is] not enough to effectuate a waiver that the party ha[s] claimed its actions [are] lawful." *Id.* at *5. That proposition is impossible to square with plaintiffs' suggestion that OpenAI waived privilege through an "intent to rely on its belief about the lawfulness and propriety of its conduct to rebut willfulness." Opp. 25.

Plaintiffs' heavy reliance on *Arista Records LLC v. Lime Group LLC* is similarly unavailing. 2011 WL 1642434 (S.D.N.Y. Apr. 20, 2011) (cited Opp. 22-24). Unlike OpenAI, the copyright defendants in *Arista Records* asserted a good-faith defense. *Id.* at *1-2. And unlike OpenAI, the defendants put legal advice squarely at issue by testifying about their belief in the "legal risk" and "liab[ilities]" involved in operating the LimeWire file-sharing website. *Id.* at *1. In those circumstances, the court deemed waiver appropriate because it would be "unfair" for the defendants to use privilege to stop discovery into the legal advice they received "before infringement." *Id.* at *3 (citation omitted). Conversely, OpenAI has not testified about its belief in legal risk, nor do the communications at issue concern legal advice OpenAI received before the alleged infringement.

b. Plaintiffs attempt to shore up Judge Wang's analysis by offering three reasons why OpenAI's reasons for deleting the datasets concern the "very issue" in dispute here. Opp. 25. None holds water.

First, plaintiffs suggest that OpenAI's reasons for deleting the datasets might be "probative" of OpenAI's state of mind during its earlier alleged infringing use of the datasets to

train GPT models. Opp. 25. Again, *Erie II* forecloses the notion that relevance alone is sufficient for at-issue waiver. 546 F.3d at 229.

Second, plaintiffs speculate that OpenAI's reasons may be especially relevant to OpenAI's state of mind during the alleged infringement if, as they contend, OpenAI *was* "using the datasets at the time of deletion." Opp. 26 (emphasis altered). But at most plaintiffs' exhibits suggest that in 2022 Books1 and Books2 were part of a larger mix of datasets referenced in "[redacted]" on models already trained on the same mix. Opp. Ex 2 at 6. The exhibits do not prove that OpenAI used Books1 or Books2 to train new models after 2021. Nor do the exhibits demonstrate that OpenAI's reasons for deleting the datasets are "critically relevant" or "indispensable" to plaintiffs' copyright claims. *Chin v. Rogoff & Co.*, 2008 WL 2073934, at *5 (S.D.N.Y. May 8, 2008) (citation omitted).

Third, plaintiffs suggest that OpenAI's reasons may be relevant to whether OpenAI spoliated evidence by deleting the datasets. But OpenAI has *already* produced copies of the datasets, undercutting any allegation of spoliation. Regardless, hypothetical relevance to an illusory "forthcoming spoliation motion" hardly demonstrates that OpenAI's reasons are critical to plaintiffs' *copyright* claims. Opp. 26.

At bottom, plaintiffs' waiver theory remains premised on the idea that defendants waive privilege simply by denying allegations of willfulness. That unprecedented rule would give skillful pleaders the ability to render privilege a nullity by making allegations about an adversary's knowledge. OpenAI Br. 23. Plaintiffs offer little reason to ignore the obvious implications of their position, instead suggesting that privilege has not "been rendered a 'nullity' in the 14 years since *Arista Records*." Opp. 24. But past practice is cold comfort because plaintiffs' proposed rule goes far beyond any previous decision. The Court should not follow plaintiffs down that path.

B. The Crime-Fraud Exception Does Not Apply

Finally, plaintiffs half-heartedly assert that Magistrate Judge Wang erred by declining to permit them access to OpenAI's privileged information under the crime-fraud exception. Opp. 26-27. In plaintiffs' view, the exception applies because OpenAI communicated with counsel about deleting the datasets in furtherance of criminal copyright infringement. Opp. 27; *see* 18 U.S.C. § 2319. But as Judge Wang held after "*in camera* review" of a "subset" of OpenAI's documents, the crime-fraud exception does not apply here because OpenAI's attorney-client communications were not themselves "in furtherance of" any "crime or fraud." Order 25-26 & n.24. Communications between OpenAI and counsel about deleting the datasets "necessarily must have taken place" after OpenAI's alleged *use of* plaintiffs' copyrighted works. Order 26. Like Judge Wang, other courts have had no difficulty rejecting plaintiffs' arguments. *See* Discovery Order at 2-3, *Kadrey v. Meta Platforms, Inc.*, No. 23-cv-3417 (N.D. Cal. Sept. 24, 2025), Dkt. No. 627.

Plaintiffs observe that efforts to "conceal" criminal conduct, too, fall within the crime-fraud exception. Opp. 27. True, but irrelevant. Plaintiffs offer zero evidence that OpenAI "*intended* … to conceal" criminal conduct, *In re Grand Jury Subpoenas Dated September 13, 2023*, 128 F.4th 127, 141-42 (2d Cir. 2025) (citation omitted), and any such suggestion is belied by the fact that OpenAI ultimately *produced recovered copies of the very datasets that plaintiffs say prove criminal activity*. *See* OpenAI Br. 1. Plaintiffs also accuse OpenAI of spoliating evidence by deleting the datasets and claim that "spoliation is sufficient" for the crime-fraud exception. Opp. 27. But even if spoliation (which is not a crime or fraud) suffices—a proposition the Second Circuit has never embraced—plaintiffs say nothing to challenge Judge Wang's finding that they "fail[ed] to allege sufficient factual basis" for this theory. Order 27.

III. CONCLUSION

The Court should grant OpenAI's objections and set aside the Order at ECF No. 846.

Respectfully submitted,

Dated: December 23, 2025

By: */s/ LISA S. BLATT*

LISA S. BLATT
WILLIAMS & CONNOLLY LLP
680 Maine Avenue SW
Washington, DC 20024
(202) 434-5000
lblatt@wc.com

Dated: December 23, 2025

By: */s/ R. JAMES SLAUGHTER*

R. JAMES SLAUGHTER
KEKER, VAN NEST & PETERS LLP
633 Battery Street
San Francisco, CA 94111
(415) 391-5400
rslaughter@keker.com

Dated: December 23, 2025

By: */s/ MARGARET GRAHAM*

MARGARET GRAHAM
LATHAM & WATKINS LLP
1271 Avenue of the Americas
New York, NY 10020
(212) 906-1200
margaret.graham@lw.com

Dated: December 23, 2025

By: */s/ ROSE LEE*

ROSE LEE
MORRISON & FOERSTER LLP
707 Wilshire Boulevard
Los Angeles, CA 90017
(213) 892-5355
roselee@mofo.com

Attorneys for Defendants
OPENAI, INC., OPENAI GP, LLC, OPENAI, LLC, OPENAI OPCO LLC, OPENAI GLOBAL LLC, OAI CORPORATION, LLC, OPENAI HOLDINGS, LLC

* All parties whose electronic signatures are included herein have consented to the filing of this document, as contemplated by Rule 8.5(b) of the Court's ECF Rules and Instructions.