

1 Joseph R. Saveri (State Bar No. 130064)
 2 Cadio Zirpoli (State Bar No. 179108)
 3 Christopher K.L. Young (State Bar No. 318371)
 4 Holden Benon (State Bar No. 325847)
 5 Aaron Cera (State Bar No. 351163)
 6 Margaux Poueymirou (State Bar No. 35600)
 7 Melissa Tribble (State Bar No. 339098)
JOSEPH SAVERI LAW FIRM, LLP
 601 California Street, Suite 1505
 San Francisco, CA 94108
 Telephone: (415) 500-6800
 Facsimile: (415) 395-9940
 Email: jsaveri@saverilawfirm.com
 czirpoli@saverilawfirm.com
 cyoung@saverilawfirm.com
 hbenon@saverilawfirm.com
 acera@saverilawfirm.com
 mpoueymirou@saverilawfirm.com
 mtribble@saverilawfirm.com

11 Matthew Butterick (State Bar No. 250953)
 12 1920 Hillhurst Avenue, #406
 Los Angeles, CA 90027
 Telephone: (323) 968-2632
 Facsimile: (415) 395-9940
 Email: mb@buttericklaw.com

*Counsel for Individual and Representative
 Plaintiffs and the Proposed Class*

17 **UNITED STATES DISTRICT COURT**
 18 **NORTHERN DISTRICT OF CALIFORNIA**
 19 **SAN FRANCISCO DIVISION**

20 IN RE OPENAI CHATGPT LITIGATION

21 This document relates to:

22 Case No. 3:23-cv-03223-AMO
 23 Case No. 3:23-cv-03416-AMO

Master File No. 3:23-cv-03223-AMO

**PLAINTIFFS' MOTION TO COMPEL
 PRODUCTION OF DOCUMENTS
 FROM NON-PARTY REUTERS NEWS &
 MEDIA INC.**

Hearing
 Date: March 18, 2025
 Time: 11:00 a.m.
 Judge: Hon. Robert M. Illman

1 PLEASE TAKE NOTICE that Plaintiffs hereby seek to compel the production of documents
2 from non-party Reuters News & Media Inc. (“Reuters”) pursuant to the attached document subpoena
3 (the “Subpoena,” attached as Exhibit A) served on Reuters through its registered agent in *In re OpenAI*
4 *ChatGPT Litig.*, 3:23-cv-03223-AMO. This Motion is based on the following points and authorities, the
5 Declaration of Aaron Cera (“Cera Decl.”), the entire record in this matter, and on such evidence and
6 argument as may be presented at any hearing of this Motion, on a date and time to be determined by the
7 Court.

8 **Introduction.** The premise underlying this Rule 45 motion to compel is simple—the companies
9 developing generative AI models must respect the law. Plaintiffs allege that OpenAI stole vast quantities
10 of copyrighted material from across the internet to train its commercial Large Language Models
11 (“LLMs”) without seeking licenses from authors. Open AI’s defense of that theft is fair use under the
12 Copyright Act, 17 U.S.C. § 107. To prosecute their case and respond to OpenAI’s fair use defense,
13 Plaintiffs require information from Reuters that they cannot get from OpenAI.

14 Reuters possesses information relevant to the existence of a market for copyrighted content as
15 training data for LLMs. On October 24, 2024, it was reported that Reuters had entered into a multi-year
16 licensing agreement with Meta Platforms Inc. for its news content to be used in Meta’s AI ChatBot. *See*
17 Sara Fischer, *Scoop: Meta strikes multi-year AI deal with Reuters*, AXIOS, Oct. 24, 2024,
18 <https://www.axios.com/2024/10/25/meta-reuters-ai-news-facebook-instagram>. This transaction is one
19 of only a handful of AI data licensing agreements in the LLM training data market. *See* Thani Shamsi,
20 *Ultimate List of Data Licensing Deals for AI*, THE MONDA BLOG, (Oct. 28, 2024),
21 <https://www.monda.ai/blog/ultimate-list-of-data-licensing-deals-for-ai>. And the existence of this public
22 agreement suggests many more private documents behind the scenes including key documents
23 concerning valuation and negotiation.

24 Plaintiffs seek the Reuters/Meta Platforms Inc. agreement and related information in support of
25 their allegations against OpenAI. Specifically, the requested documents are relevant to OpenAI’s
26 anticipated fair use defense, namely the effect of its theft of copyrighted material on the market for
27 copyrighted content as training data for LLMs, and to assist in valuing the copyrighted works for use as
28 training data, including developing a damages methodology. This material is in the possession, custody,

1 or control of Reuters and cannot be obtained from OpenAI because it is an exclusive agreement between
2 Reuters and Meta Platforms Inc., the terms and conditions of which are not public. Mia Sato, *Meta signs*
3 *its first big AI deal for news*, THE VERGE, (Oct. 25, 2024),
4 <https://www.theverge.com/2024/10/25/24279259/meta-reuters-ai-chatbot-deal-news-licensing-media>
5 (noting the confidentiality of the terms of the agreement).

6 Despite Plaintiffs' good faith efforts to obtain compliance with the subpoena, Reuters has refused
7 to produce documents both relevant and proportional to the needs of this case. Instead, Reuters has
8 chosen to stand on boilerplate objections that are unsupported in the law. The Court should order
9 Reuters to comply with the subpoena.

10 **Factual background.** Plaintiffs served a document subpoena on Reuters on November 20, 2024.
11 The subpoena sought three categories of documents:

- 12 • **Executed AI Training Data Licensing Agreements** (Requests for Production Nos. 1, 7):
13 Documents pertaining to executed agreements that grant permission to use data for the purpose
14 of training AI models.
- 15 • **AI Training Data Licensing Agreement Negotiations and Valuations** (Requests for
16 Production Nos. 2, 5, 6): Documents pertaining to AI training data licensing agreements, whether
17 executed or not, including the specific details of the agreements such as terms, conditions,
18 pricing, and form of payment.
- 19 • **Specific Content for AI Training** (Requests for Production Nos. 3, 4): Documents pertaining to
20 licensing agreements for specific content types used in AI training, including books and news
21 articles.

22 Cera Decl. ¶ 2.

23 **Plaintiffs have exhausted reasonable efforts to obtain compliance.** Plaintiffs have made
24 multiple attempts to resolve this dispute without Court intervention, including through meeting and
25 conferring and multiple email exchanges. Cera Decl. ¶ 5, 7, 8, 9, 10, 11. After Plaintiffs granted a courtesy
26 one-week extension for Reuters to respond, it responded on December 9, 2024 with boilerplate
27 objections, refusing to produce the requested documents on various grounds, including claims of undue
28 burden, overbreadth, and confidentiality. Cera Decl. ¶ 4. The parties met and conferred on December

1 23, 2024 and exchanged numerous emails attempting to resolve the subpoena. Cera Decl. ¶ 5, 7, 8, 9, 10,
2 11. Despite these efforts, the parties were unable to reach agreement. Cera Decl. ¶ 9, 10. Reuters has
3 failed to produce any responsive documents or to otherwise compromise in any way, choosing to stand
4 on its objections. Cera Decl. ¶ 9. Despite these efforts, and the clear importance of the requested
5 information, Reuters has refused to produce the requested documents, necessitating this Motion.

6 **Reuters’s boilerplate objections should be stricken.** Generally, the Federal Rules have a strong
7 preference for full discovery when practicable. *Williams v. Condensed Curriculum Int’l*, No.
8 20CV05292YGRRMI, 2021 WL 6621071, at *2 (N.D. Cal. Dec. 29, 2021) (Illman, J.) (citation omitted).
9 The party resisting discovery bears the “burden of clarifying, explaining, and supporting its objections.”
10 *Id.* Reuters’s objections are without merit and cannot justify its non-compliance. Indeed, many of the
11 objections raised by Reuters are frivolous.

12 **1. Relevance**

13 Rule 45 subpoenas are subject to the general relevancy standard applicable to discovery under
14 Rules 26 and 34 and other discovery rules. *See* Fed. R. Civ. P. 45 advisory committee’s note to 1970
15 amendment. *See also Mi Familia Vota v. Hobbs*, 343 F.R.D. 71, 81 (D. Ariz. 2022) (citing *Aquastar Pool*
16 *Prods. Inc. v. Paramount Pool & Spa Sys.*, 2019 WL 250429, *2 (D. Ariz. 2019)) (“The test for
17 ‘relevance,’ in the context of a Rule 45 subpoena to a non-party, is no different than the test under Rules
18 26 and 34.”). Under Rule 26(b)(1), “[p]arties may obtain discovery regarding any nonprivileged matter
19 that is relevant to any party’s claim or defense and proportional to the needs of the case.” Fed. R. Civ. P.
20 26(b)(1).

21 The documents Plaintiffs request are directly relevant to their claim of direct copyright
22 infringement and to counter OpenAI’s fair use defense. The documents clearly pertain to important
23 issues in the litigation, including licensing practices, the valuation of content utilized in training LLMs,
24 and the terms of applicable agreements. The documents sought from Reuters via the subpoena will help
25 establish the existence of a market for copyrighted work as training data for LLMs because they will
26 provide evidence of benchmark licensing agreements and market demand. *Oracle Corp. v. SAP AG*, 765
27 F.3d 1081, 1093 (9th Cir. 2014) (“Although a copyright plaintiff need not... present evidence of
28 “benchmark” agreements in order to recover hypothetical-license damages, it may be difficult for a

1 plaintiff to establish the amount of such damages without undue speculation in the absence of such
2 evidence.”).

3 This information is also relevant to calculating actual damages and recovery of lost profits. *Polar*
4 *Bear Prods., Inc. v. Timex Corp.*, 384 F.3d 700, 707–08 (9th Cir. 2004), as amended on denial of reh’g and
5 reh’g en banc (9th Cir. Oct. 25, 2004) (“Actual damages are usually determined by the loss in the fair
6 market value of the copyright, measured by the profits lost due to the infringement or by the value of the
7 use of the copyrighted work to the infringer.”) (internal citations and quotations omitted).

8 The discovery sought from Reuters will assist in proving that, but for OpenAI’s theft of vast
9 amounts of copyright-protected works, Plaintiffs would have been able to license their works at a
10 competitive price. The information sought directly relates to the fair use factor of the effect on the
11 market for copyrighted material by setting parameters for its value and use. *Oracle Am., Inc. v. Google*
12 *Inc.*, 847 F. Supp. 2d 1178, 1182 (N.D. Cal. 2012). If copyright owners could license their work for AI
13 training, similar licensing agreements help to establish the market value of their work. *Jarvis v. K2 Inc.*,
14 486 F.3d 526 (9th Cir. 2007) (finding district court’s calculations of actual damages were properly based
15 on “objective considerations of market value”). Moreover, internal valuation memos can provide
16 evidence of potential lost revenue due to infringement and would likely contain projections and analyses
17 of potential licensing income. This supports Plaintiffs’ claims for lost profits. If an AI developer uses
18 copyrighted material without a license, the copyright owner’s potential lost licensing revenue is valuable
19 for determining damages. Additionally, negotiations surrounding a potential licensing deal will provide
20 evidence of the license’s value, helping to quantify losses. *Oracle Am., Inc.*, 847 F. Supp. 2d at 1182 (“To
21 determine the work’s market value at the time of the infringement, we have endorsed a hypothetical
22 approach: what a willing buyer would have been reasonably required to pay to a willing seller for the
23 owner’s work.”) (citations omitted).

24 In the context of a fair use defense, which is the critical issue in this case, the fourth factor in
25 establishing such defense (the effect of the use upon the potential market for or value of the copyrighted
26 work) requires consideration of whether the unlicensed use of the work undermines the market or
27 licensing opportunities for it. 17 U.S.C. § 107(4). If a licensing market exists and is negatively impacted
28

1 by OpenAI’s theft of data, this weighs against a finding of fair use. *Dr. Seuss Enters., L.P. v. Penguin Books*
2 *USA, Inc.*, 109 F.3d 1394, 1403 (9th Cir. 1997). The documents sought will aid Plaintiffs in this analysis.

3 Despite the clear importance to Plaintiffs’ claims, in the meet and confer, Reuters disputed the
4 relevance of the requests, arguing that licensing deals with news organizations have no relevance here
5 because the named Plaintiffs are book authors. Cera Decl. ¶ 5. Not so. The operative complaint’s class
6 definition is “[a]ll persons or entities domiciled in the United States that own a United States copyright
7 in any work that was used as training data for the OpenAI Language Models during the Class Period.”
8 Am. Compl. ¶ 53, ECF No. 120; Cera Decl. ¶ 5, 7. Definitionally, news authors are part of the class.
9 Moreover, even if the class were limited to books, news-related licensing agreements would be relevant.
10 Indeed, the Federal Rules of Evidence permit the use of indirect evidence at trial. *See* Fed. R. Evid. 401;
11 *Unicolors, Inc. v. H&M Hennes & Mauritz, L.P.*, 52 F.4th 1054, 1082 (9th Cir. 2022) (recognizing Fed. R.
12 Evid. 401 as a liberal standard). Plaintiffs could thus permissibly use a news licensing agreement—along
13 with other evidence showing the value of books in LLM training—to establish the existence and value of
14 a book market. *Williams v. Gaye*, 895 F.3d 1106, 1129 (9th Cir. 2018) (upholding award of profits where
15 plaintiff tethered their hypothetical license damages to evidence of a benchmark license in the industry).
16 In short, even under Reuters’s strained reading of the Complaint, the requests are relevant.

17 2. Overbreadth and Vagueness

18 Reuters objects to the requests as overly broad and vague, particularly with respect to terms such
19 as “AI training data,” “licensing agreements,” and “Article.” That is a risible claim because Reuters
20 itself uses the terms. For example, last year Reuters published an article called “Inside Big Tech’s
21 underground race to buy AI training data.” *See* Katie Paul and Anna Tong, *Inside Big Tech's underground*
22 *race to buy AI training data*, REUTERS (Apr. 5, 2024); *see also* Reuters, *OpenAI seeks partnerships to generate*
23 *AI training data*, REUTERS (Nov. 9, 2023). If Reuters cannot understand terms like “AI training data” or
24 “licensing agreements,” it could have just asked its writers.

25 In any case, the terms are sufficiently defined within the context of the subpoena and the
26 litigation. Reuters cannot avoid those definitions “by quibbling.” *Marchand v. Mercy Med. Ctr.*, 22 F.3d
27 933, 938 (9th Cir. 1994). Instead, if Reuters cannot understand a term, it can state that in a response and
28 then produce documents within the meaning of the term as it understands it. But the idea that Reuters

1 cannot understand basic terms—and terms that it itself uses—is not a reason to avoid the subpoena.
2 *Condensed Curriculum Int'l*, 2021 WL 5069946, at *5 (finding defendant’s vagueness objections to
3 common terms like “you,” “contend,” “due,” and “support” frivolous and obstructive, especially
4 given defendant’s own use of such terms).

5 3. Undue Burden

6 Reuters next claims that compliance would impose an undue burden, and that Plaintiffs should
7 first seek production from OpenAI. To start, that is largely impossible because the publicly disclosed
8 agreement is between Reuters and Meta Platforms, not OpenAI. What is more, as a sophisticated entity
9 with substantial resources, Reuters is well-equipped to produce the requested documents. Reuters’s
10 undue burden objection should also be overruled.

11 Each factor in Rule 26(b)(1) supports a finding of no undue burden. This action involves issues of
12 significant public importance, considering OpenAI’s public predictions about LLMs’ economic impact
13 and the scope of the alleged infringement. The amount in controversy is substantial. Reuters possesses
14 unique information, specifically its confidential licensing agreement with Meta Platforms and related
15 details. And it has ample resources to comply with the subpoena.

16 4. Confidentiality

17 Reuters asserts that the requests seek confidential or proprietary information. However, a
18 Protective Order (ECF No. 106) is in place in this litigation to safeguard such information, and
19 confidentiality concerns do not justify withholding responsive documents. Indeed, “there is no absolute
20 privilege [against disclosure] for trade secrets and similar confidential information.” *DDS, Inc. v. Lucas*
21 *Aerospace Power Transmission Corp.*, 182 F.R.D. 1, 4 (N.D.N.Y. 1998) (quoting *Fed. Open Mkt. Comm. of*
22 *the Fed. Reserve Sys. v. Merrill*, 443 U.S. 340, 362, (1979)) (alteration in original); *see also ICG Commc’ns,*
23 *Inc. v. Allegiance Telecom*, 211 F.R.D. 610, 614 (N.D. Cal. 2002) (“courts generally eschew an absolute
24 privilege for trade secrets and similar confidential information in favor of a case-by-case approach that
25 balances privacy against the need for disclosure. Other confidential business information is generally
26 afforded even less protection than trade secrets.”) (quotations and citations omitted). Reuters’s
27 confidentiality objections are without merit.

1 At bottom, Reuters's boilerplate objections are meritless. Plaintiffs' requests are relevant and
2 proportional to the needs of this case, and Reuters has not articulated a legitimate reason for its refusal to
3 comply. For the foregoing reasons, Plaintiffs respectfully request that the Court grant this Motion and
4 order Reuters to produce documents responsive to Plaintiffs' Requests for Production.

5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28

1 Dated: February 10, 2025

By: /s/ Joseph R. Saveri

Joseph R. Saveri

3 Joseph R. Saveri (State Bar No. 130064)
4 Cadio Zirpoli (State Bar No. 179108)
5 Christopher K.L. Young (State Bar No. 318371)
6 Holden Benon (State Bar No. 325847)
7 Aaron Cera (State Bar No. 351163)
8 Margaux Poueymirou (State Bar No. 35600)
9 Melissa Tribble (State Bar No. 339098)
10 **JOSEPH SAVERI LAW FIRM, LLP**
11 601 California Street, Suite 1505
12 San Francisco, CA 94108
13 Telephone: (415) 500-6800
14 Facsimile: (415) 395-9940
15 Email: jsaveri@saverilawfirm.com
16 czirpoli@saverilawfirm.com
17 cyoung@saverilawfirm.com
18 hbenon@saverilawfirm.com
19 acera@saverilawfirm.com
20 mpoueymirou@saverilawfirm.com
21 mtribble@saverilawfirm.com

15 Matthew Butterick (State Bar No. 250953)
16 1920 Hillhurst Avenue, #406
17 Los Angeles, CA 90027
18 Telephone: (323) 968-2632
19 Facsimile: (415) 395-9940
20 Email: mb@buttericklaw.com

19 Bryan L. Clobes (pro hac vice)
20 **CAFFERTY CLOBES MERIWETHER**
21 **& SPRENGEL LLP**
22 205 N. Monroe Street
23 Media, PA 19063
24 Telephone: 215-864-2800
25 Email: bclobes@caffertyclobes.com

24 Alexander J. Sweatman (pro hac vice anticipated)
25 **CAFFERTY CLOBES MERIWETHER**
26 **& SPRENGEL LLP**
27 135 South LaSalle Street, Suite 3210
28 Chicago, IL 60603
Telephone: 312-782-4880
Email: asweatman@caffertyclobes.com

1 Daniel J. Muller (State Bar No. 193396)
2 **VENTURA HERSEY & MULLER, LLP**
3 1506 Hamilton Avenue
4 San Jose, California 95125
5 Telephone: (408) 512-3022
6 Email: dmuller@venturahersey.com

7 David Boies
8 **BOIES SCHILLER FLEXNER LLP**
9 333 Main Street Armonk, NY 10504
10 Telephone: (914) 749-8201
11 Email: dboies@bsflp.com

12 Evan Matthew Ezray
13 **BOIES SCHILLER FLEXNER LLP**
14 401 E. Las Olas Blvd., Suite 1200
15 Ft. Lauderdale, FL 33301
16 Telephone: (954) 377-4237
17 Email: eezray@bsflp.com

18 Jesse Michael Panuccio
19 **BOIES SCHILLER FLEXNER LLP**
20 1401 New York Ave. NW Washington, DC 20005
21 Telephone: (202) 237-2727
22 Email: jpanuccio@bsflp.com

23 Maxwell V. Pritt
24 Joshua Michelangelo Stein
25 **BOIES SCHILLER FLEXNER LLP**
26 44 Montgomery Street, 41st Floor
27 San Francisco, CA 94104
28 Telephone: (415) 293-6813
Telephone: (415) 293-6800
Email: jpanuccio@bsflp.com
Email: mpritt@bsflp.com

*Counsel for Individual and Representative
Plaintiffs and the Proposed Class*