

SUSMAN GODFREY L.L.P

May 28, 2024

Hon. Sidney H. Stein
United States District Judge
Southern District of New York
500 Pearl Street, Courtroom 23A
New York, New York 10007

Re: *The New York Times Company v. Microsoft Corporation, et al.*,
Case No.: 23-cv-11195-SHS: Opposition to OpenAI's Motion, Dkt. 124

Dear Judge Stein:

The New York Times Company (“The Times”) opposes OpenAI’s May 23 motion. Dkt. 124. OpenAI’s motion should be denied in full.

1. The Times Proposes a Substantial Completion Deadline of July 31.

The Court should reject OpenAI’s proposed June 24 deadline for The Times’s document productions. While OpenAI claims that its proposal “mirrors the amount of time the Court ordered for OpenAI to substantially complete production” for The Times’s initial RFPs (Dkt. 124 at 1), OpenAI omits the key fact: The Times’s initial 15 requests overlap substantially with the initial 15 requests served in November in the consolidated class case pending before this Court, No. 23-cv-08292 (the “Class Case”). The Times expressly relied on this overlap when requesting the June 14 deadline in this case, arguing “there is no reason Defendants should not produce documents responsive to the requests Plaintiffs have already served at the same time that they produce substantially the same documents in the Consolidated Class Cases.” Dkt. 72 at 16. This Court agreed and imposed the June 14 deadline for only those initial overlapping requests.

OpenAI is comparing apples to oranges. The Times’s initial requests were intentionally focused primarily on OpenAI’s use of Times copyright-protected content in its models and products, which should readily be produced. *E.g.*, Ex. 1 at 7 (RFP 8). By contrast, OpenAI’s March 8th RFPs included 61 requests, covering all issues in the case (and many other things as well). Ex. 2. The June 14 deadline was always intended to apply to an initial set of limited RFPs—not every RFP in the case. The upshot is that OpenAI has roughly seven months (November to June) to investigate The Times’s initial 15 requests—yet OpenAI demands that The Times substantially complete production for 61 requests in just 3.5 months.¹ In other words, OpenAI demands responses to *four times* as many requests in half the time. A fairer deadline is July 31, which leaves seven weeks to complete depositions before the close of fact discovery.

From the beginning, The Times has pushed to proceed as efficiently as possible, serving document requests on February 23, 2024—the first day on which discovery could be served. The Times invited Defendants to do the same, but they did not. Defendants instead threatened to stay discovery pending their motions to dismiss. While OpenAI eventually reversed course, its delay

¹ OpenAI’s math is also wrong. Its March 8 requests came two weeks after The Times’s February 23 requests. Under OpenAI’s logic, The Times’s deadline should be two weeks after OpenAI’s June 14 deadline: June 28, not June 24.

May 28, 2024

Page 2

in serving document requests is the sole reason why it cannot take advantage of the June 14 substantial completion deadline. Had OpenAI served discovery sooner, that deadline would apply to The Times's initial productions as well.²

2. OpenAI Is Not Entitled to Privileged Documents About The Times's Infringement Investigation.

OpenAI's second argument is that The Times has not produced privileged communications and work product surrounding the creation of Exhibit J to its Complaint. Exhibit J has examples of GPT-4 memorizing Times content and shows that, *inter alia*, OpenAI trained on, copied, and stored at least one-hundred unique Times works—and likely millions more. The Times was forced to create this exhibit because OpenAI has repeatedly refused to tell the public what works were used to train its models.

The Times has sought from OpenAI discovery into what Times works it used to train its models, how OpenAI copied and used Times works during training and beyond, and the propensity for OpenAI's models to output the Times works it trained on. The Times does not intend to rely on Exhibit J at trial so long as OpenAI complies with its discovery obligations, and any demonstrative The Times's experts create for the jury – with the benefit of full access to OpenAI's data – will be subject to expert discovery. Moreover, The Times has already agreed to produce much of what OpenAI seeks. But OpenAI also wants information protected by privilege or work product, to which it is not entitled. Nor does OpenAI explain why it needs this information from The Times given its troubling admission that it is already tracking users' (and potential litigation experts') efforts to ascertain whether OpenAI used copyrighted content to train its models.

RFP 20 and 23: Request 20 seeks “All Documents and Communications relating to the creation of Exhibit J of the Complaint,” and Request 23 seeks “Documents sufficient to show the process for obtaining each GPT Services output cited or referred to in the Complaint,” including chat logs, prompts, parameters, and outputs. The Times agreed to produce the underlying facts and analysis requested, but not its retained consulting expert's or Times employees' privileged *communications with counsel* about the facts uncovered. *See Robinson v. De Niro*, 2022 WL 7094922, at *10 (S.D.N.Y. Mar. 9, 2022) (even when privilege protecting underlying factual information is waived, communications with counsel about the factual information remains protected).

RFP 2: Request 2 is broader, seeking “All Documents and Communications regarding any attempt by You, including failed attempts, to reproduce any of Your Published Works via GPT Services.” Plaintiff reasonably construed this vague Request to mean “The Times's process for obtaining the GPT Services outputs cited in the Complaint” and agreed to produce non-privileged documents in this category. OpenAI Ex. C at 6. The Times is not willing to produce work product or privileged communications regarding outputs that were not in the Complaint, nor is that required. Such outputs that were obtained by The Times's non-testifying consultant and Times

² While The Times is under no interim deadline for producing documents, it has assured OpenAI that it will make productions as quickly as possible and finish as contemplated by the case schedule.

May 28, 2024

Page 3

employees at the direction of Times counsel are protected from disclosure under Federal Rule 26(b)(4)(D) and as privileged work product. This protection was not waived because The Times is not relying on Exhibit J to prove OpenAI's infringement at trial. *See Dover v. British Airways*, 2014 WL 5090021, at *2-3 (E.D.N.Y. Oct. 9, 2014) (analysis prepared by a non-testifying consulting expert that the plaintiff is not relying on to prove its case is protected from disclosure under Rule 26(b)(4)(D) and not waived even if cited in the complaint); *Fin. Guar. Ins. v. Putnam Adv., LLC*, 314 F.R.D. 85, 90 (S.D.N.Y. 2016) ("the reference to the analysis in FGIC's complaint, standing alone, does not constitute a waiver").

OpenAI's caselaw is inapposite. In *In re Commodity Exch., Gold Futures*, 2019 WL 13046984, at *3 (S.D.N.Y. Feb. 25, 2019), the plaintiff relied on portions of its consultants' statistical analyses to prove the defendants' price-fixing conspiracy, so the defendants were entitled to discover the undisclosed portions. Unlike that case, the proof of OpenAI's infringement will lie not in Exhibit J but in OpenAI's documents and data, and both sides' yet-to-be-prepared expert testimony. *See U.S. v. Omnicare*, 2023 WL 7297152, at *1 (S.D.N.Y. Nov. 6, 2023) (finding no waiver of undisclosed witness summaries not relied on in the complaint where the undisclosed analysis would not disprove the core allegation in the complaint, distinguishing *In re Commodity Exch.*). Because The Times will not present Exhibit J to the jury, there are no underlying materials that OpenAI "might need to effectively contest or impeach the claim." *New York Times Co. v. DOJ*, 939 F.3d 479, 498 (2d Cir. 2019). Nor did The Times waive any privilege or protection through "disclos[ure] to OpenAI in the course of interacting with ChatGPT LLMs." Dkt. 124 at 2-3. OpenAI's reliance on *In Re Steinhardt* is misplaced because, unlike in that case, the party asserting the privilege did not previously disclose the protected material to a government agency. 9 F.3d 230, 232 (2d Cir. 1993). Moreover, to accept OpenAI's argument would undermine "[t]he logic behind the work product doctrine"—that "opposing counsel should not enjoy free access to an attorney's thought processes." *Id.* at 234. OpenAI nowhere explains why interaction with its user-facing product is equivalent to interaction with its counsel.

RFP 7 and 21: Request Nos. 7 and 21 seek detailed personal OpenAI account information from every employee at The Times and from "any person who participated in *or was aware of* [The Times's] use of GPT Services to generate any of the outputs cited in or referred to in the Complaint." Ex. C at 9, 18. The Times appropriately refused to produce additional documents for these overbroad requests, which seek material that has nothing to do with The Times's claims or OpenAI's defenses. Ex. C at 9-10, 18-19. OpenAI's only proffer of relevance is that the OpenAI account information of every Times employee *and* anyone who even knew The Times was investigating this lawsuit (which expands the scope to absurdity) "would potentially enable OpenAI" to fill "gaps" about the outputs cited in the Complaint, like the "parameters" used with each prompt. Dkt. 124 at 2. The Times already agreed to produce this information in response to Request 23. OpenAI offers no other explanation for relevance, because it cannot. Anything discoverable in the process for obtaining the outputs cited in the Complaint will be produced in response to Requests 2, 20, and 23. The additional information sought in Requests 7 and 21 is invasive.

May 28, 2024
Page 4

Respectfully submitted,

/s/ Ian B. Crosby

Ian B. Crosby
Susman Godfrey L.L.P.

/s/ Steven Lieberman

Steven Lieberman
Rothwell, Figg, Ernst & Manbeck

cc: All Counsel of Record (via ECF)
Enclosures

Plaintiff's Exhibit 1

**UNITED STATES DISTRICT COURT
SOUTHERN DISTRICT OF NEW YORK**

THE NEW YORK TIMES COMPANY,

Plaintiff,

v.

MICROSOFT CORPORATION, OPENAI, INC.,
OPENAI LP, OPENAI GP, LLC, OPENAI, LLC,
OPENAI OPCO LLC, OPENAI GLOBAL LLC,
OAI CORPORATION, LLC, and OPENAI
HOLDINGS, LLC,

Defendants.

Civil Action No. 1:23-cv-11195

**THE NEW YORK TIMES COMPANY’S FIRST SET OF REQUESTS FOR
PRODUCTION TO OPENAI, INC., OPENAI LP, OPENAI GP, LLC, OPENAI, LLC,
OPENAI OPCO LLC, OPENAI GLOBAL LLC, OAI CORPORATION, LLC, AND
OPENAI HOLDINGS, LLC**

Pursuant to the Federal Rules of Civil Procedure and the Local Rules of the Southern District of New York, (the “Local Civil Rules”), The New York Times Company (“NYT”) requests that OpenAI, Inc., OpenAI LP, OpenAI GP, LLC, OpenAI, LLC, OpenAI OPCO LLC, OpenAI Global LLC, OAI Corporation, LLC, and OpenAI Holdings, LLC (“OpenAI Defendants”) produce the requested documents at the offices of Susman Godfrey L.L.P., 1301 Avenue of the Americas, 32nd Floor, New York, NY 10019, within thirty days after the date of service of this document, or at such other mutually agreeable time and place.

I. DEFINITIONS

The following definitions and instructions apply to the requests listed below:

1. All definitions and rules of construction set forth in Rule 26.3 of the Local Rules of the Southern District of New York are incorporated by reference herein.

2. “Affiliate” means, with respect to any Person, any other Person that, directly or indirectly through one or more intermediaries, controls, is controlled by, or is under any common control with, such Person.

3. “Agreement(s)” means any contract, agreement, arrangement, or understanding, formal or informal, oral or written, between or among two or more Persons.

4. “AI Model(s)” means a component of an information system and/or large language model that implements artificial intelligence (“AI”) technology and uses computational, statistical, or machine-learning techniques to produce outputs from a given set of inputs.

5. “Generative AI Models(s)” means Your AI Models that emulate the structure and characteristics of input data in order to generate derived synthetic content, such as images, videos, audio, text, and other digital content.

6. “Text Generation AI Model(s)” means Your Generative AI Models, including generative pre-trained transformers or large language models (LLMs), that provide text outputs in response to inputs or prompts, including but not limited to GPT base, GPT-1, GPT-2, GPT-3, GPT 3.5, GPT-4, GPT-4 Turbo and all other versions or other Generative AI Models included in Your Generative AI Products and Services.

7. “Complaint” means the operative complaint in the Action as of the date these Requests are served.

8. “Defendants” means Microsoft and OpenAI.

9. “Electronically Stored Information” or “ESI” refers to information and Documents within the full scope of Federal Rule of Civil Procedure 34 with all Metadata intact—created, manipulated, communicated, stored, and best utilized in digital form, and stored on electronic media.

10. “Generative AI Product(s) and Service(s)” means any publicly available or commercial product or service offering that includes a Generative AI Model, including but not limited to ChatGPT branded products and services, Copilot branded products and services, ChatGPT Browse with Bing, Bing Chat, APIs, and platforms otherwise providing access to a Generative AI Model.

11. “Including” means including but not limited to.

12. “Journalism,” for the purposes of these requests, means the activity of writing or creating content for newspapers, magazines, news websites, mobile applications, television, podcasts, or any other publication and/or news outlet, and includes the work of The Times as alleged in the Complaint.

13. “Metadata” refers to structured information about an electronic file that is embedded in the file, describing the characteristics, origins, usage, and validity of the electronic file.

14. “OpenAI” or “You” or “Your” means Defendants OpenAI Inc., OpenAI GP, LLC, OpenAI Opco, LLC, OpenAI Global, LLC, OpenAI Holdings LLC, OpenAI, LLC, or OAI Corporation, LLC, their subsidiaries, affiliates, divisions, predecessors in interest, successors, or assignees, and their respective officers, directors, employees, consultants, representatives, and agents.

15. “Relating to” means discussing, describing, referring to, pertaining to, containing, analyzing, studying, reporting on, commenting on, evidencing, constituting, setting forth, considering, recommending, concerning, or pertaining to, in whole or in part.

16. “Microsoft” means Defendant Microsoft Corporation, its subsidiaries, affiliates, divisions, predecessors in interest, successors, or assignees, and their respective officers, directors, employees, consultants, representatives, and agents.

17. “The Times” means Plaintiff The New York Times Company, its subsidiaries, affiliates, divisions, predecessors in interest, successors, or assignees, and their respective officers, directors, employees, consultants, representatives, and agents.

18. “Times Content” means material published by The Times, including in print, online, audio, or in any other format, whether on Times-owned platforms or available through third-party sources.

19. “Training Dataset(s)” means the data and content used to train or fine tune the Generative AI Models.

II. INSTRUCTIONS

1. Produce all responsive documents prepared, sent, or received, in whole or in part, during the time period applicable to the request. You should identify in some reasonable manner the person from whose files each document you produce was found.

2. You are to produce all documents, as defined above, that are in the possession, custody or control of you, or in the possession, custody or control of any attorney or agent for you. Without limiting the terms, “possession, custody, or control” of an item means that you either have physical possession of the item or have a right to possession of the item that is equal or superior to the person who has physical possession of the item.

3. You are under a duty to supplement your responses to these discovery requests pursuant to FRCP 26. If after you produce any of the discovery requested below you discover or receive additional documents or information responsive to the requests, you should promptly produce such additional documents or information to the full extent required by the rules.

4. Pursuant to FRCP 26, you are to provide a privilege log for any documents you have withheld on the basis of privilege(s). Request for such a log is hereby made, effective as of the response date of these requests. If you assert a claim of privilege with respect to any document

request, or portion thereof, you should produce a log, accounting for each document withheld individually, and providing the nature of the privilege being claimed; the type of document, date, author, addressees or recipients, and general subject matter of the document, and such other information as is sufficient to identify the document. If a claim of privilege applies to only a portion of a document, the document should be produced, and the portion claimed to be privileged obscured and stamped “redacted.”

5. The documents requested below should be produced in the manner that they are kept in the usual course of business. Documents should be produced with the label or labels from any file folder, binder, file drawer, file box, notebook, computer disk, or other container in which the document was found. Documents on computers or computer disks, including but not limited to electronic mail and corresponding attachments, should be produced in both native format and single-page TIFF format with optical character recognition (OCR). Even if only part of a document is responsive to a request, you should still produce the entire document including all attachments.

6. All duplicates or copies of documents are to be provided to the extent they have handwriting, additions, or deletions of any kind different from the original document being produced.

7. All documents which cannot be legibly copied should be produced in their original form.

8. If any documents requested herein have been lost, discarded or destroyed, they shall be identified as completely as possible, including, without limitation, the following information: (a) the date of disposal; (b) the manner of disposal; (c) the reason for disposal; (d) the person authorizing disposal; and (e) the person disposing of the document.

9. If no Documents responsive to a particular request exist, state that no responsive Documents exist.

10. If you or your lawyers find any of these requests vague, confusing, hard to understand, or if they just want to talk through some issues relating to the requests, please call Elisha Barron at 212-729-2013. Please do not wait and object instead of attempting to resolve the issue with a telephone call. Thank you.

III. DOCUMENT REQUESTS

REQUEST FOR PRODUCTION NO. 1:

Documents concerning the use of Times Content for training or fine tuning Your Text Generation AI Models, including, for every Text Generation AI Model that uses Times Content: i) the amount and prevalence of Times Content in the Training Datasets by token count, word count, article count, as a percentage of the Training Datasets, or by any other metric; ii) rules or guidance for choosing what is and is not included in the Training Datasets for the Text Generation AI Models; iii) use of Times Content to clean or curate the Training Datasets; iv) any weighting of particular Training Datasets for the Text Generation AI Models; and v) use of Times Content for any post-training activity, including fine tuning.

REQUEST FOR PRODUCTION NO. 2:

Documents You have gathered to submit or have submitted to any legislative or executive agency, committee, or other governmental entity in the United States that concern or relate to the allegations in the Complaint, including any Documents related to Your intended or actual submission.

REQUEST FOR PRODUCTION NO. 3:

Documents concerning copyrighted material in the Text Generation AI Models, the Training Datasets, or any Generative AI Products and Services, including documents or communications regarding the appropriateness, value, importance, necessity, and/or concerns over or legal implications of the use of such materials.

REQUEST FOR PRODUCTION NO. 4:

Documents concerning claims of copyright and trademark infringement, including complaints regarding infringement, steps taken to respond to such claims, and steps taken to prevent infringement.

REQUEST FOR PRODUCTION NO. 5:

Documents reflecting policies, procedures or practices concerning use of intellectual property in AI Models, including the policy OpenAI referred to as “Copyright Shield” in the November 6, 2023, blog post on Your website titled “New models and developer products announced at DevDay.”

REQUEST FOR PRODUCTION NO. 6:

Documents concerning alternatives to using copyrighted content to train the Text Generation AI Models without compensation including using licensing agreements or material in the public domain.

REQUEST FOR PRODUCTION NO. 7:

Documents concerning the impact of AI Models on Journalism and the past, current or future use of the Text Generation AI Models or any Generative AI Products and Services as a source of Journalism, substitute for Journalism, or product in competition with Journalism (including The Times’s Journalism).

REQUEST FOR PRODUCTION NO. 8:

Documents sufficient to show to what extent and in what form Times Content resides in, is stored in, is copied by, or otherwise exists within the Training Datasets and/or the Text Generation AI Models.

REQUEST FOR PRODUCTION NO. 9:

Documents sufficient to show and describe OpenAI’s corporate organizational structure,

including those sufficient to show OpenAI’s board of directors, parent companies, subsidiaries, affiliates, departments, and divisions; and those sufficient to show the function each Defendant plays within Your corporate structure as a whole.

REQUEST FOR PRODUCTION NO. 10:

Documents sufficient to identify OpenAI’s current and former officers, directors, managers, employees and consultants with knowledge of the Text Generation AI Models, business practices regarding the Text Generation AI Models, and commercialization and use of the Text Generation AI Models.

REQUEST FOR PRODUCTION NO. 11:

Documents concerning OpenAI’s transition to a for-profit company, as alleged in paragraphs 57-58 of the Complaint.

REQUEST FOR PRODUCTION NO. 12:

Documents concerning Defendants’ use of Retrieval Augmented Generation (“RAG”) in Generative AI Products and Services, including: (i) use of Times Content for RAG; (ii) rules or guidance for preferring or avoiding certain web domains (including The Times’s web domains or sites known to host content copied from other sources) when choosing what to include in a generative result; (iii) rules for whether and how the results from RAG search products are similar to or different from standard search results; (iv) and rules regarding paywalls.

REQUEST FOR PRODUCTION NO. 13:

Documents concerning the features or capabilities of Retrieval Augmented Generation.

REQUEST FOR PRODUCTION NO. 14:

A complete listing of every data sample in the Training Datasets used to train each of the Text Generation AI Models and the source from which each data sample was obtained.

REQUEST FOR PRODUCTION NO. 15:

Documents concerning any commercial or supercomputing collaborations between and among the Defendants relating to the Generative AI Models, including licensing, developing, training, hosting, engineering, testing, deploying, or operating Generative AI Product(s) and Service(s).

Dated: February 23, 2024

/s/ Elisha Barron

Ian Crosby (*admitted pro hac vice*)
Genevieve Vose Wallace (*admitted pro hac vice*)
Katherine M. Peaslee (*pro hac vice pending*)
SUSMAN GODFREY L.L.P.
401 Union Street, Suite 3000
Seattle, WA 98101
Telephone: (206) 516-3880
Facsimile: (206) 516-3883
icrosby@susmangodfrey.com
gwallace@susmangodfrey.com
kpeaslee@susmangodfrey.com

David Brook (*admitted pro hac vice*)
Emily K. Cronin (*admitted pro hac vice*)
Ellie Dupler (*admitted pro hac vice*)
SUSMAN GODFREY L.L.P.
1900 Ave of the Stars, Suite 1400
Los Angeles, CA 90067
Telephone: (310) 789-3100
Facsimile: (310) 789-3150
dbrook@susmangodfrey.com
ecronin@susmangodfrey.com
edupler@susmangodfrey.com

Elisha Barron (5036850)
Zachary B. Savage (ZS2668)
Tamar Lusztig (5125174)
Alexander Frawley (5564539)
Eudokia Spanos (5021381)
SUSMAN GODFREY L.L.P.
1301 Avenue of the Americas, 32nd Floor
New York, NY 10019
Telephone: (212) 336-8330
Facsimile: (212) 336-8340

ebarron@susmangodfrey.com
zsavage@susmangodfrey.com
tlusztig@susmangodfrey.com
afrawley@susmangodfrey.com
espanos@susmangodfrey.com

Scarlett Collings (*Admission Pending*)
SUSMAN GODFREY L.L.P.
1000 Louisiana, Suite 5100
Houston, TX 77002
Telephone: (713) 651-9366
Facsimile (713) 654-6666
scollings@susmangodfrey.com

Steven Lieberman (SL8687)
Jennifer B. Maisel (5096995)
Kristen J. Logan (*admitted pro hac vice*)
ROTHWELL, FIGG, ERNST & MANBECK, P.C.
901 New York Avenue, N.W., Suite 900 East
Washington, DC 20001
Telephone: (202) 783-6040
Facsimile: (202) 783 6031
slieberman@rothwellfigg.com
jmaisel@rothwellfigg.com
klogan@rothwellfigg.com

Attorneys for Plaintiff
The New York Times Company

Plaintiff's Exhibit 2

UNITED STATES DISTRICT COURT
SOUTHERN DISTRICT OF NEW YORK

-----	X:	
THE NEW YORK TIMES COMPANY	:	
	:	
Plaintiff,	:	
	:	No. 3:23-CV-11195-SHS
-against-	:	
	:	
MICROSOFT CORPORATION, <i>et al.</i> ,	:	
	:	
Defendants.	:	
-----	X	

**DEFENDANT OPENAI OPCO, LLC’S FIRST SET OF REQUESTS
FOR PRODUCTION OF DOCUMENTS AND THINGS (NOS. 1-61)**

In accordance with Federal Rules of Civil Procedure 26 and 34, Defendant OpenAI OpCo, L.L.C. (“OpenAI OpCo”), by and through its undersigned attorneys, hereby request that Plaintiff The New York Times Company respond to these First Set of Requests for Production of Documents and Things (“Requests”) and produce for inspection and copying, within thirty days of service, the Documents and Things described below, in accordance with the following Definitions and Instructions. OpenAI OpCo requests that Plaintiff produces Documents and other Things described below electronically or at the offices of Latham & Watkins LLP, 505 Montgomery Street, Suite 2000, San Francisco, California 94111.

DEFINITIONS

1. Each of these Definitions and Instructions is incorporated into each of the Requests to which it pertains. Notwithstanding any Definition below, each word, term, or phrase used in these Requests is intended to have the broadest meaning permitted under the Federal Rules of Civil Procedure.

2. “Plaintiff,” “You,” and/or “Your” shall mean The New York Times Company, its officers, directors, managers, Employees, Agents, divisions, parent companies, subsidiaries, affiliates, including any Person who served in any such capacity at any time.

3. “Defendant” shall mean any defendant named in the Complaint filed on December 27, 2023 in *The New York Times Company v. Microsoft Corporation, et al.*, Case No. 1:23-cv-11195 and any subsequent complaint that You file in this action.

4. “OpenAI” shall mean OpenAI, Inc., OpenAI LP, OpenAI GP, LLC, OpenAI, LLC, OpenAI Opco L.L.C., OpenAI Global LLC, OAI Corporation, LLC, and OpenAI Holdings, LLC.

5. “Generative AI” shall mean models, algorithms, or other systems that are trained using data to generate new content (including text, images, or other media), such as ChatGPT,

Gemini, LLaMA, or Claude.

6. “Person” shall mean any natural person or any business, legal, or governmental entity or association.

7. “Communication” shall mean any transmission, exchange, or transfer of information by any means (in the form of facts, ideas, inquiries, or otherwise).

8. “Complaint” shall mean the Complaint filed on December 27, 2023 in *The New York Times Company v. Microsoft Corporation, et al.*, Case No. 1:23-cv-11195 and any subsequent complaint that You file in this action.

9. References to this “Action,” “Case,” “Litigation,” or “Lawsuit” mean the above-captioned matter, *The New York Times Company v. Microsoft Corporation, et al.*, Case No. 1:23-cv-11195, brought by Plaintiff pending in the United States District Court for the Southern District of New York.

10. “Agent” shall mean any Person asked, hired, retained, or contracted to assist You or act on Your behalf in connection with this Litigation whether employed or not by You.

11. “Employee” shall mean director, trustee, officer, employee, or partner of the designated entity, whether active or retired, full-time or part-time, current or former, and compensated or not.

12. “Document” shall mean all “writings” and “recordings” as those terms are defined in Federal Rules of Evidence 1001 and shall be afforded the broadest possible meaning permitted by Federal Rule of Civil Procedure 34. By way of illustration and without limitation, Documents include at least the following: originals, drafts and all non-identical copies of memoranda, reports, studies, surveys, analyses, books, manuscripts, notes, emails, graphs, notebooks, correspondence, interoffice communications, letters, diaries, calendars, photographs, motion pictures, sketches,

spreadsheets, source code, drawings, promotional material, technical papers, printed publications, and all other writings, as well as all non-paper information storage means such as sound reproductions, computer inputs and outputs, tape, film and computer memory devices, as well as tangible things such as models, modules, prototypes, and commercially saleable products. If a draft Document has been prepared in several copies that are not identical, or if the original identical copies are no longer identical due to subsequent notation, each non-identical Document is a separate Document. The foregoing specifically includes information stored in a computer database and capable of being generated in documentary form, such as electronic mail. For the avoidance of doubt, any electronic Documents or data shall include its associated metadata.

13. “Thing” shall have the full meaning ascribed to it by the Federal Rules of Civil Procedure.

14. “DMCA” means the Digital Millennium Copyright Act (17 U.S.C. § 512).

15. “Asserted Work” shall mean any work listed in Exhibits A-I and K of the Complaint, or any additional work that may be listed on an amended complaint.

16. “Published Work” shall mean any work created by or published by You, including both Your Asserted Works and other works not referenced in the Complaint.

17. “ChatGPT” shall mean OpenAI’s general purpose generative-AI model ChatGPT, and the underlying GPT-3.5 and GPT-4 large language models.

18. “GPT Services” shall mean OpenAI’s general purpose generative-AI model ChatGPT, the underlying GPT-3.5 and GPT-4 large language models, and any GPT-based products, including any Application Programming Interface or other third party services that rely on any of the aforementioned.

19. “Electronically Stored Information” includes, but is not limited to, e-mails and

attachments, voice mail, instant messages, text messages, cell phone data and other electronic communications, word processing documents, text files, hard drives, Excel spreadsheets and underlying formulae, graphics, audio and video files, databases, calendars, telephone logs, transaction logs, Internet usage files, offline storage or information stored on removable media (such as external hard drives, hard disks, floppy disks, memory sticks, flash drives, and backup tapes), information contained on laptops or other portable devices, and network access information and backup materials, TIF files, PDF files, Native Files and the corresponding metadata which is ordinarily maintained.

20. “Concern,” “Concerning,” “Related to,” and “Relating to,” are construed broadly to mean in whole or in part alluding to, responding to, concerning, relating to, connected with, involving, commenting on, in respect of, about, associated with, discussing, evidencing, showing, describing, reflecting, analyzing, summarizing, memorializing, consisting of, constituting, identifying, stating, tending to support, tending to discredit, referring to, or in any way touching upon.

21. “Exclusion Protocol” shall mean any rule, method, or policy to prevent specific user agents, such as web crawlers or bots, from accessing a website or parts of a website.

22. “Including” means “including but not limited to.” “Include(s)” means “include(s) but is not limited to.”

23. OpenAI’s use of terms, phrases, and Definitions is for convenience and no term, phrase, or Definition shall be construed as an admission by OpenAI.

24. Any word that is not defined has its usual and customary meaning.

INSTRUCTIONS

1. The following shall apply to all Requests:

- a. The singular of each word shall be construed to include its plural, and vice versa;
- b. The conjunctions “and” as well as “or” shall be construed either conjunctively or disjunctively, which is more inclusive in context;
- c. “Each” shall be construed to include “all” and “any,” and vice versa; and
- d. The present tense shall be construed to include the past tense, and vice versa.

2. If, in responding to these Requests, You encounter any ambiguities when construing a question, instruction, or definition, Your answer shall set forth the matter deemed ambiguous, and the construction used in answering.

3. In producing the Documents or Things requested, You are requested to produce all Documents and Things known or available to You regardless of whether such Documents and Things are possessed directly by You or Your Agents, Employees, representatives, investigators, or by Your attorneys or their Agents, Employees, representatives, or investigators.

4. In the event that any Document is withheld on a claim of attorney-client privilege or work-product doctrine on any other ground, You shall provide a log that identifies any such Document in a manner to be mutually agreed among the parties.

5. In producing the Documents and Things requested herein, please produce them in their original file folders, if any, or in lieu thereof, attach to the set of Documents produced from a given file or a photostatic or electronic duplicate of all written or printed material in the original file folder. In addition, the Documents shall be produced in the same sequence as they are contained or found in the original file folder. The integrity and internal sequence of the requested Documents within each file folder shall not be disturbed. Under no circumstances shall Documents from any file folder be commingled with Documents from any other file folder.

6. If production of any requested Document(s) is objected to on the grounds that production is unduly burdensome, describe the burden or expense of the proposed discovery.

7. These Requests are continuing in nature. If You receive or otherwise become aware of information responsive to any Request after You have served Your responses to these Requests, You must promptly supplement Your responses to these Requests to provide such information, as required by Federal Rule of Civil Procedure 26.

REQUESTS FOR PRODUCTION

REQUEST FOR PRODUCTION NO. 1:

All Documents and Communications relating to the alleged reproduction, public display, or distribution of Your Asserted Works via GPT Services.

REQUEST FOR PRODUCTION NO. 2:

All Documents and Communications regarding any attempt by You, including failed attempts, to reproduce any of Your Published Works via GPT Services.

REQUEST FOR PRODUCTION NO. 3:

All Documents and Communications relating to any outputs of GPT Services that allegedly summarize, quote, or otherwise reference Your Asserted Works.

REQUEST FOR PRODUCTION NO. 4:

All Documents and Communications relating to the alleged reproduction, public display, or distribution of Your Asserted Works via Generative AI services other than GPT Services.

REQUEST FOR PRODUCTION NO. 5:

All Documents and Communications relating to any attempt by You, including failed attempts, to reproduce any of Your Published Works via Generative AI services other than GPT Services.

REQUEST FOR PRODUCTION NO. 6:

All Documents and Communications relating to any outputs of via Generative AI services other than GPT Services that allegedly summarize, quote, or otherwise reference Your Asserted Works.

REQUEST FOR PRODUCTION NO. 7:

Documents sufficient to show each of the OpenAI accounts You or Your Agents have created or used, including without limitation Documents sufficient to show the full name associated with the account(s), the username(s) for the account(s), email address(es) associated with the account(s), the organization ID and name associated with the account(s), and date of registration or activation for the account(s).

REQUEST FOR PRODUCTION NO. 8:

All Documents and Communications relating to any allegations that any of Your Asserted Works infringe any third-party rights.

REQUEST FOR PRODUCTION NO. 9:

All Documents and Communications relating to any complaints by any Person regarding alleged plagiarism in Your Asserted Works.

REQUEST FOR PRODUCTION NO. 10:

Documents sufficient to identify the expressive, original, and human-authored content of each of Your Asserted Works.

REQUEST FOR PRODUCTION NO. 11:

Documents sufficient to identify the non-expressive, non-original, or non-human-authored content of each of Your Asserted Works.

REQUEST FOR PRODUCTION NO. 12:

Documents sufficient to show each and every written work that informed the preparation of each of Your Asserted Works, regardless of its length, format, or medium.

REQUEST FOR PRODUCTION NO. 13:

All Documents and Communications relating to any disputes as to the ownership of Your Asserted Works, including but not limited to DMCA Takedown Notices or other copyright infringement removal requests received by You.

REQUEST FOR PRODUCTION NO. 14:

All Documents and Communications relating to Your correspondence with the United States Copyright Office regarding Your Asserted Works, including deposit copies.

REQUEST FOR PRODUCTION NO. 15:

All Documents and Communications relating to Your ownership of the Asserted Works, including agreements related to the authorship of Your Asserted Works and work-for-hire agreements.

REQUEST FOR PRODUCTION NO. 16:

All Documents and Communications relating to Your knowledge of the alleged use of Your Published Works for training Generative AI models.

REQUEST FOR PRODUCTION NO. 17:

All Documents and Communications relating to Your knowledge of the alleged use of Your Published Works for training ChatGPT.

REQUEST FOR PRODUCTION NO. 18:

All Documents and Communications relating to any arrangement or agreement by which You permitted a third party to reproduce or display Your Asserted Works, or any part thereof, in

print or on the Internet, regardless of whether the arrangement or agreement was commercial or non-commercial.

REQUEST FOR PRODUCTION NO. 19:

All Documents and Communications relating to Your investigation of the claims alleged in the Complaint.

REQUEST FOR PRODUCTION NO. 20:

All Documents and Communications relating to the creation of Exhibit J of the Complaint, including but not limited to Documents and Communications with any third party or Agent.

REQUEST FOR PRODUCTION NO. 21:

Documents sufficient to show each of the OpenAI accounts created or used by any Person who participated in or was aware of Your use of GPT Services to generate any of the outputs cited in or referred to in the Complaint.

REQUEST FOR PRODUCTION NO. 22:

Documents sufficient to show each of the prompts You have entered into GPT Services, including without limitation Documents sufficient to show any system prompts used, the parameters used in connection with each prompt (including, but not limited to, temperature, model, maximum length, stop sequences, top p, frequency penalty, presence penalty), the date and time on which that prompt was entered, the user account used, and each resulting output.

REQUEST FOR PRODUCTION NO. 23:

Documents sufficient to show the process for obtaining each GPT Services output cited or referred to in the Complaint, including without limitation the full chat log, the prompts used, any system prompts used, the parameters used in connection with each prompt (including, but

not limited to, temperature, model, maximum length, stop sequences, top p, frequency penalty, and presence penalty), each and every output generated by GPT Services as a result of each prompt and parameter combination, the time and date of those queries, and the user account.

REQUEST FOR PRODUCTION NO. 24:

All Documents and Communications between You and other news, media, or writers' organizations and publishers, as well as any other named plaintiff in *Tremblay v. OpenAI, Inc.*, No. 3:23-cv-03223 and *Silverman v. OpenAI, Inc.*, No. 3:23-cv-03416 in the Northern District of California, and *Authors Guild v. OpenAI Inc.*, No. 1:23-cv-08292, *Alter v. OpenAI Inc.*, No. 1:23-cv-10211, *Basbanes v. Microsoft*, No. 1:24-cv-00084, *The Intercept Media, Inc. v. OpenAI, Inc.*, No. 1:24-cv-01515, and *Raw Story Media, Inc. v. OpenAI, Inc.*, No. 1:24-cv-01514 in the Southern District of New York, directly or through any Employee, Agent, or third party, related to copyright and artificial intelligence.

REQUEST FOR PRODUCTION NO. 25:

All Documents and Communications between You and other news, media, or writers' organizations and publishers, directly or through any Employee, Agent, or third party, relating to decisions about whether to license works to OpenAI for purposes of training ChatGPT.

REQUEST FOR PRODUCTION NO. 26:

All Documents and Communications relating to any injury or harm You claim to have suffered as a result of the conduct alleged in the Complaint.

REQUEST FOR PRODUCTION NO. 27:

Documents sufficient to show Your total revenue, broken down by source, including without limitation advertising revenue, revenue from affiliate links, subscription revenue by

subscription type, whether personal or commercial, online or print) for each month beginning January 2015 to the present.

REQUEST FOR PRODUCTION NO. 28:

Documents sufficient to calculate or estimate per-display revenues associated with Your Asserted Works by source, including without limitation revenues from subscriptions, advertising revenues (including cost per impression), and revenues from affiliate links. Each metric shall include both (i) the total number since the work's publication and (ii) a breakdown by month from month of publication to the present.

REQUEST FOR PRODUCTION NO. 29:

Documents sufficient to show the total page views for each of Your Asserted Works for each month, beginning with the month of the work's publication to the present.

REQUEST FOR PRODUCTION NO. 30:

Documents sufficient to show the total page views for each of Your Asserted Works for each hour of each day during the 14-day period following its publication.

REQUEST FOR PRODUCTION NO. 31:

Documents sufficient to show the number of times each of Your Asserted Works was shared on social media platforms for each hour of each day during the 14-day period following its publication.

REQUEST FOR PRODUCTION NO. 32:

Documents sufficient to show Your monthly expenses relating to the creation of Your Published Works, broken down by category of expense (including, without limitation, travel, research sources, and salaries) and category of content (as described in Paragraphs 33–37 of the Complaint).

REQUEST FOR PRODUCTION NO. 33:

All Documents and Communications relating to licenses or attempts to license Your Asserted Works for reproduction, preparation of derivative works, distribution, display, or performance by businesses (including, but not limited to, artificial intelligence companies, academic institutions, non-profit organizations, or other legal entities), including without limitation enterprise licenses, archive licenses, Generative AI licenses, and licenses for text and data mining.

REQUEST FOR PRODUCTION NO. 34:

Documents sufficient to show the number of Your active subscribers for each month beginning January 2015 to the present.

REQUEST FOR PRODUCTION NO. 35:

All Documents and Communications relating to formal or informal analyses relating to trends in readership or online subscriptions and the causes thereof.

REQUEST FOR PRODUCTION NO. 36:

Documents sufficient to show how You analyze the financial value of content on Your website and the results of such analyses, including without limitation analyses of the financial value of page views and other indicators of financial performance.

REQUEST FOR PRODUCTION NO. 37:

All Documents and Communications relating to formal or informal analyses of the performance of content published on Your website, including without limitation analysis of methods to evaluate that performance.

REQUEST FOR PRODUCTION NO. 38:

All Documents and Communications relating to Your use or potential use of Exclusion Protocols on Your website, including without limitation Documents sufficient to show the time periods during which You implemented such Exclusion Protocols on Your website.

REQUEST FOR PRODUCTION NO. 39:

Documents sufficient to show Your policies and procedures regarding the use of third-party sources (including third-party texts) in creating Your published content.

REQUEST FOR PRODUCTION NO. 40:

All Documents and Communications relating to Your alleged “attempt[] to reach a negotiated agreement with Defendants” as referenced in Paragraph 7 of the Complaint.

REQUEST FOR PRODUCTION NO. 41:

All Documents and Communications relating to any commercial arrangements through which You obtain access to third-party content for use by Your Employees, including without limitation enterprise subscriptions to other news sources, research databases, and online libraries.

REQUEST FOR PRODUCTION NO. 42:

For each of Your Asserted Works, Documents sufficient to show the relevant category of that work as referenced in Paragraphs 33–37 of the Complaint. Categories can include, without limitation, “Investigative Reporting,” “Breaking News Reporting,” “Beat Reporting,” “Reviews and Analysis,” or “Commentary and Opinion.”

REQUEST FOR PRODUCTION NO. 43:

All Documents and Communications relating to any commercial arrangements by which third parties reproduce, distribute, display, or perform Your Asserted Works, including without limitation agreements relating to syndication.

REQUEST FOR PRODUCTION NO. 44:

Documents sufficient to show Your policies and procedures regarding the use of Generative AI by Your Employees.

REQUEST FOR PRODUCTION NO. 45:

All Documents and Communications relating to Your potential adoption of Generative AI for use by Employees of The New York Times Company, including Communications with any Generative AI company regarding that potential adoption.

REQUEST FOR PRODUCTION NO. 46:

Documents sufficient to show any training programs You plan to implement or have implemented regarding Generative AI at The New York Times Company.

REQUEST FOR PRODUCTION NO. 47:

All Documents and Communications relating to Your “A.I. Initiatives” program referenced in Your December 12, 2023 article titled “Zach Seward Is the Newsroom’s Editorial Director of A.I. Initiatives.”

REQUEST FOR PRODUCTION NO. 48:

All Documents and Communications relating to Zach Seward’s role as the editorial director of Your “A.I. Initiatives” program referenced in Your December 12, 2023 article titled “Zach Seward Is the Newsroom’s Editorial Director of A.I. Initiatives.”

REQUEST FOR PRODUCTION NO. 49:

Documents sufficient to show any past and present efforts or future plans to utilize Generative AI in Your reporting and presentation of content, including but not limited to print articles, online articles, blog posts, videos, and graphics.

REQUEST FOR PRODUCTION NO. 50:

All Documents and Communications relating to the training of Your artificial intelligence

satellite images tool referenced in the February 19, 2024 Reuters Institute article, “New York Times publisher A. G. Sulzberger: ‘Our industry needs to think bigger’” including without limitation any base models and training data You used.

REQUEST FOR PRODUCTION NO. 51:

All Documents and Communications relating to the training of Your artificial intelligence tool referenced in the February 20, 2024 Axios article, “Exclusive: NYT plans to debut new generative AI ad tool later this year” including without limitation any base models and training materials You used.

REQUEST FOR PRODUCTION NO. 52:

All Documents and Communications relating to The New York Times’ API that “allows researchers and academics to search Times content for non-commercial purposes” referenced in Paragraph 46 of the Complaint.

REQUEST FOR PRODUCTION NO. 53:

Documents sufficient to show Your total revenue generated from licenses through the Copyright Clearance Center each year from 2015 to the present.

REQUEST FOR PRODUCTION NO. 54:

All Documents and Communications relating to The New York Times English Gigaword (LDC2003T05) of the Linguistic Data Consortium of the University of Pennsylvania, including but not limited to all articles included therein, any licensing arrangements, and any negotiations between You and the Linguistic Data Consortium or University of Pennsylvania.

REQUEST FOR PRODUCTION NO. 55:

All Documents and Communications relating to The New York Times Annotated Corpus (LDC2008T19) of the Linguistic Data Consortium of the University of Pennsylvania, including

but not limited to all articles included therein, any licensing arrangements, and any negotiations between You and the Linguistic Data Consortium or University of Pennsylvania.

REQUEST FOR PRODUCTION NO. 56:

All Documents and Communications relating to any efforts You have undertaken to separate Your reporting on OpenAI and Your activities underlying the preparation and prosecution of this lawsuit.

REQUEST FOR PRODUCTION NO. 57:

Documents sufficient to show people most knowledgeable about Your use (including use by Your Agents, Employees, or third parties at Your direction) of GPT Services to derive any of the outputs cited or referred to in the Complaint and any other outputs on which You plan to rely upon in support of Your claims in this Litigation.

REQUEST FOR PRODUCTION NO. 58:

All Documents and Communications between You and any United States government or regulatory agencies relating to Generative AI, including this Litigation.

REQUEST FOR PRODUCTION NO. 59:

All Documents referenced or cited in preparing Your responses to any of OpenAI's Interrogatories and Requests for Admissions.

REQUEST FOR PRODUCTION NO. 60:

All Documents You intend to rely on in this Litigation.

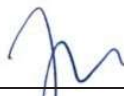
REQUEST FOR PRODUCTION NO. 61:

All Documents You intend to provide to Your experts in this Litigation.

Dated: March 8, 2024

LATHAM & WATKINS LLP

By: _____


LATHAM & WATKINS LLP
Joseph R. Wetzel
joseph.wetzel@lw.com
Andrew M. Gass (*pro hac vice*)
andrew.gass@lw.com
505 Montgomery Street, Suite 2000
San Francisco, CA 94111
Telephone: 415.391.0600

Sarang V. Damle
sy.damle@lw.com
555 Eleventh Street, NW, Suite 1000
Washington, D.C. 20004
Telephone: 202.637.2200

Allison L. Stillman
alli.stillman@lw.com
1271 Avenue of the Americas
New York, NY 10020
Telephone: 212.751.4864

*Attorneys for Defendants OpenAI, Inc.,
OpenAI LP, OpenAI GP, LLC, OpenAI,
LLC, OpenAI Opco L.L.C., OpenAI Global
LLC, OAI Corporation, LLC, and OpenAI
Holdings, LLC.*

Dated: March 8, 2024

MORRISON & FOERSTER LLP

By: _____


Joseph C. Gratz (*pro hac vice*)

JGratz@mofocom

Vera Ranieri (*pro hac vice*)

VRanieri@mofocom

425 Market Street

San Francisco, CA 94105-2482

Telephone: 415.268.7000

Allyson R. Bennett (*pro hac vice*)

ABennett@mofocom

Rose S. Lee (*pro hac vice*)

RoseLee@mofocom

707 Wilshire Boulevard

Los Angeles, California 90017-3543

Telephone: (213) 892-5200

*Attorney for Defendants OpenAI, Inc.,
OpenAI LP, OpenAI GP, LLC, OpenAI,
LLC, OpenAI Opco L.L.C., OpenAI Global
LLC, OAI Corporation, LLC, and OpenAI
Holdings, LLC.*

CERTIFICATE OF SERVICE

I hereby certify that on March 8, 2024, a copy of the foregoing **DEFENDANT OPENAI OPCO, LLC'S FIRST SET OF REQUESTS FOR PRODUCTION OF DOCUMENTS AND THINGS (NOS. 1-61)** was served by E-mail upon the following:

Ian Crosby (<i>pro hac vice</i>) Genevieve Vose Wallace (<i>pro hac vice</i>) Katherine M. Peaslee (<i>pro hac vice</i>) SUSMAN GODFREY L.L.P. 401 Union Street, Suite 3000 Seattle, WA 98101 Telephone: (206) 516-3880 Facsimile: (206) 516-3883	Elisha Barron (5036850) Zachary B. Savage (ZS2668) Tamar Lusztiq (5125174) Alexander Frawley (5564539) Eudokia Spanos (5021381) SUSMAN GODFREY L.L.P. 1301 Avenue of the Americas, 32nd Floor New York, NY 10019 Telephone: (212) 336-8330 Facsimile: (212) 336-8340
Davida Brook (<i>pro hac vice</i>) Emily K. Cronin (<i>pro hac vice</i>) Ellie Dupler (<i>pro hac vice</i>) SUSMAN GODFREY L.L.P. 1900 Ave of the Stars, Suite 1400 Los Angeles, CA 90067 Telephone: (310) 789-3100 Facsimile: (310) 789-3150	Steven Lieberman (SL8687) Jennifer B. Maisel (5096995) Kristen J. Logan (<i>pro hac vice</i>) ROTHWELL, FIGG, ERNST & MANBECK, P.C. 901 New York Avenue, N.W., Suite 900 East Washington, DC 20001 Telephone: (202) 783-6040 Facsimile: (202) 783-6031
Scarlett Collings (<i>Admission pending</i>) SUSMAN GODFREY L.L.P. 1000 Louisiana, Suite 5100 Houston, TX 77002 Telephone: (713) 651-9366 Facsimile (713) 654-6666	
<i>NYT-AI-SG-Service@simplelists.susmangodfrey.com</i>	
<i>Attorneys for Plaintiff</i>	

Annette L. Hurst
**ORRICK, HERRINGTON &
SUTCLIFF LLP**
The Orrick Building
405 Howard Street
San Francisco, CA 94105
Phone: (415) 773-5700

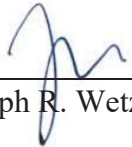
Jeffrey S. Jacobson
**FAEGRE DRINKER BIDDLE &
REATH**
1177 Avenue of the Americas
New York, NY 10036
Telephone: (212) 248-3191
jeffrey.jacobson@faegredrinker.com

NewYorkTimes_Microsoft_OHS@orrick.com

Christopher J. Cariello
**ORRICK, HERRINGTON & SUTCLIFF
LLP**
51 West 2nd Street
New York, NY 10019
Telephone: (212) 506-5000

Jared B. Briant
FAEGRE DRINKER BIDDLE & REATH
1144 15th Street, Suite 3400
Denver, CO 80202
Telephone: (303) 607-3588
jared.briant@faegredrinker.com

Attorneys for Defendant Microsoft Corporation



Joseph R. Wetzel