EXHIBIT Q

Elana Nightingale-Dawson
elana.nightingaledawson@lw.com
Phone: 202.637.2303

555 Eleventh Street, N.W., Suite 1000
Washington, D.C. 20004-1304
Tel: +1.202.637.2200 Fax: +1.202.637.2201
www.lw.com

LATHAM & WATKINS LLP

FIRM / AFFILIATE OFFICES

Austin	Milan
Beijing	Munich
Boston	New York
Brussels	Orange County
Century City	Paris
Chicago	Riyadh
Dubai	San Diego
Düsseldorf	San Francisco
Frankfurt	Seoul
Hamburg	Silicon Valley
Hong Kong	Singapore
Houston	Tel Aviv
London	Tokyo
Los Angeles	Washington, D.C.
Madrid	

March 19, 2025

VIA EMAIL

Jennifer Maisel
jmaisel@rothwellfigg.com
Rothwell, Figg, Ernst & Manbeck, P.C.
901 New York Avenue, N.W.
Suite 900 East
Washington, D.C. 20001

Re: *The New York Times Company v. Microsoft Corp. et al., Case No. 24-cv-11195-SHS; Daily News, LP et al v Microsoft Corp. et al., Case No. 24-cv-3285-SHS*

Counsel:

I write in response to your January 28, 2025 letter in which you seek information regarding logs of user conversation data, in advance of a discussion on this topic with technical experts. We have addressed below your requests for written information based on our investigation to date. For ease of review, questions 1 through 6 are restated below with OpenAI's responses inline. This information is provided on a voluntary basis as part of an informal exchange designed to facilitate efficient discovery procedures. While accurate to the best of our knowledge at this time, it is possible that our further investigation may uncover different or additional information. OpenAI's investigation is ongoing.

With respect to questions 7 through 14 in your letter, which you asked to discuss with technical experts, OpenAI is willing to engage in such a discussion with respect to questions 7, 8, and 10–13. OpenAI is not willing to engage in such a discussion with respect to question 9, which seeks information about OpenAI's business practices and is thus more appropriately addressed in the context of a deposition. None of the information sought in question 9 is necessary for Plaintiffs to understand, sample, and analyze user conversation data. With respect to question 14, as we have explained to you on numerous previous occasions, OpenAI does not have an existing tool or mechanism that can search for "News Plaintiffs' intellectual property," Jan. 28 Ltr. at 4, across the over 60 billion ChatGPT user conversations OpenAI possesses.

***

Plaintiffs' Question (1): Please be prepared to describe and walk through how the log dataset(s) are stored, including the geographic location(s) of the dataset(s), whether the data is split or sharded (and on what basis), the time period for which data is available, and

LATHAM & WATKINS LLP

the volume of available data in bytes and number of records. This should also include any data retention policies and/or backup storage related to the log dataset.

OpenAI's Response:

ChatGPT Free, ChatGPT Plus, and ChatGPT Pro user conversation data is replicated from Azure Cosmos DBs–the production environment for ChatGPT–to a Databricks data warehouse. The user conversation data replicated from the Azure Cosmos DBs is compressed. As of March 5, 2025, there were approximately [redacted] ChatGPT Free/Plus/Pro conversations stored in the Databricks data warehouse since December 9, 2022. In compressed format, this constitutes approximately [redacted] of data. This data is partitioned into roughly [redacted] shards based on a hash of each [redacted] and [redacted]

ChatGPT Enterprise conversations are stored across multiple Azure Cosmos DBs. This conversation data is compressed. As of March 5, 2025, there were over [redacted] ChatGPT Enterprise conversations stored in Azure Cosmos DBs. In compressed format, and not including uploaded files, this is roughly [redacted] of data. This conversation data is partitioned by [redacted] and sharded. OpenAI does not own or control the user inputs or ChatGPT outputs for ChatGPT Enterprise customers. In the ordinary course of business, OpenAI employees do not have access to ChatGPT Enterprise user conversation data.

For OpenAI's API Platform, [redacted] The API completions stored in Databricks are [redacted] On March 5, 2025, there were over [redacted] individual API completions stored. In uncompressed format, this is roughly [redacted] of completion data. This data is [redacted]

OpenAI has addressed extensively its retention policies with respect to user conversation data, including in its filings at ECF 423 and 425. We will not rehash those details here.

We do not understand the basis for your request for the geographic location in which user conversation data is stored. Such information is not necessary to understand, sample, or analyze user conversation data. Nor is it relevant to this litigation.

Plaintiffs Question (2): The output log that was produced was limited to Retrieval Augmented Generation (RAG). Please provide sample output logs for non-RAG user sessions using ChatGPT, as well as sample output for the API and custom GPTs as we have previously discussed.

OpenAI's Response:

OpenAI will produce an API sample as requested. OpenAI does not have a way to identify "non-RAG user sessions" in ChatGPT user conversation data, but we have created a non-RAG sample and will produce it.

LATHAM & WATKINS LLP

Plaintiffs Question (3): The produced data contains extensive nested fields. Please provide field, subfield, and value descriptions at each level. What other fields/subfields/values may be contained in the log dataset but are not represented in this produced sample data?

OpenAI's Response:

A consolidated description of the fields, subfield, and values is not kept in the ordinary course of business. User conversation data logs are semi-structured and contain hundreds of fields. OpenAI does not have documentation detailing the meaning of all the fields in conversation logs. We are willing to consider questions regarding specific fields. That said, there is no central mechanism for creating and including fields.

Plaintiffs' Question (4): Prior to the expert discussion, to better focus the parties' time and resources, please produce additional documentation about the log files and the data/fields contained in those log files not produced already, including at least: schema related documentation and any files related to analyzing or processing of the logs, definitions of each taxonomy and classifier category, example mappings between user input/system output and taxonomy classifications, conversation flows/decision trees, and rules for applying classifiers in conversations, lists of all possible event types (user/system/error/etc.), and any related event groupings, and event taxonomy documentation/schemas. If such additional documentation has been produced, please identify the production numbers of such materials.

OpenAI's Response:

OpenAI has identified, based on a reasonable investigation, one document it has already produced (OPCO_NEWS_0827050) that addresses this Question. OpenAI is happy to discuss what, exactly, Plaintiffs are hoping to accomplish with the documentation sought to see if there is a way to provide additional information or documentation.

Plaintiffs Question (5): Please provide a full list of the referenced domains associated with Plaintiffs, and classifications of those domains (e.g., copyright, business, legal case number, etc.), and the basis for such classifications. Please also identify when the domains were added to the [redacted] and [redacted] lists, the difference between these lists, and the effect of including a domain on these lists.

OpenAI's Response:

The [redacted] field contains URLs, whereas the [redacted] field contains domains. The lists have existed since at least August 21, 2024. When a domain is included in the [redacted] field, ChatGPT Search will not access content on the domain in relation to a user's query. When a URL is included in the [redacted] field, the URL will not be displayed in a Search result and ChatGPT will not access content from the URL in relation to a user's query. The remainder of the information sought is either protected by attorney-client and/or work product privilege or,

LATHAM & WATKINS LLP

if not privileged, more properly addressed in the context of a deposition, as none of this information is necessary to understand, sample, or analyze user conversation data. Additionally, OpenAI has already produced numerous documents containing these lists including: OPCO_NEWS_0192796, OPCO_NEWS_0855146, OPCO_NEWS_0429066, OPCO_NEWS_0620902, and OPCO_NEWS_0429067.

Plaintiffs Question (6): Please provide technical documentation about classifiers (e.g., as indicated by but not limited to field [redacted], and associated values, such as, but not limited to, [redacted] and capability classes (e.g., [redacted]

OpenAI's Response:

OpenAI has not identified, based on a reasonable investigation, documentation of the sort Plaintiffs are requesting. OpenAI will be prepared to address classifiers generally, and the specific classifiers identified, during the parties' upcoming discussion.

Best regards,

/s/ Elana Nightingale-Dawson
Elana Nightingale-Dawson
of LATHAM & WATKINS LLP