UNITED STATES DISTRICT COURT
SOUTHERN DISTRICT OF NEW YORK

IN RE: OpenAI, INC., COPYRIGHT INFRINGEMENT LITIGATION This Document Relates To: 1:25-cv-04315-SHS-OTW	25-md-3143 (SHS) (OTW) **DECLARATION OF LACY H. KOONCE, III IN SUPPORT OF ZIFF DAVIS'S MOTION FOR SUMMARY JUDGMENT**

I, Lacy H. Koonce, III, pursuant to 28 U.S.C. §1746, hereby declare as follows:

1. I am a partner at Klaris Law PLLC, located at 161 Water Street, New York, NY, 10038, and counsel to Plaintiffs Ziff Davis, Inc., Ziff Davis LLC, IGN Entertainment, Inc., Everyday Health Media, LLC, Mashable, Inc., and CNET Media, Inc. (together "Ziff Davis") in the above-captioned action. I make this declaration in support of Ziff Davis's Rule 56 Motion for Summary Judgment.

2. Ziff Davis is submitting certain pieces of voluminous materials in support of its Motion for Summary Judgement. However, I understand that, due to the size and nature of these documents, they cannot be filed electronically on PACER. Instead, Ziff Davis has mounted this evidence on a hard drive ("the Drive") and is providing it to the Court. Details about opening the Drive have been sent by email to the Court.

3. The Drive is broken up into three buckets: (1) "Asserted Works" (2) "Training Data" and (3) "CMI Removal". Taken together, these materials are being submitted to provide comprehensive support for the facts that are relevant for Ziff Davis's claims on Summary Judgement: that (1) Ziff Davis owns its Asserted Works, that (2) those works appear in OpenAI's

training data, and (3) that copyright-management information was stripped from the articles that OpenAI trained on.

4. Attached hereto as Exhibit A is a slipsheet for the Drive, which provides further information about how the files are organized and named, as detailed below.

The "1 - Asserted Works"

5. As explained in the Declaration of Robert Taylor ("Taylor Decl."), Ziff Davis undertook a process to identify only copyrighted articles that (1) are owned by Ziff Davis and (2) match to an OpenAI training record as an "Exact Match" or "Revised Version."[1] I understand that there are 1,636,586 such articles ("ZD Asserted Works").

6. Ziff Davis has created Rule 1006 Summary Spreadsheets ("ZD Rule 1006 Summaries") for each of its 20 brands. Taylor Decl. ¶¶ 47–55. The ZD Rule 1006 Summaries demonstrate, on a per-work basis, how Ziff Davis owns and holds valid copyright registrations for each of the articles it is asserting. Id. The ZD Rule 1006 Summaries are stored on the Drive in a subfolder called "ZD Rule 1006 Summaries". Ex. A.

7. Attached hereto as Exhibit AA is a true and correct copy of the CNET Rule 1006 Summary. Attached hereto as Exhibit BB is a true and correct copy of the Mashable Rule 1006 Summary. Attached hereto as Exhibit CC is a true and correct copy of the ZDNET Rule 1006 Summary. Attached hereto as Exhibit DD is a true and correct copy of the Lifehacker Rule 1006

[1] As explained by Mr. Delorey in his declaration, Ziff Davis's "Exact Match" comparison indicates that two normalized texts, (excluding the first and last 50 characters and excluding any CMI removed) contain identical text or have directional word containment equal to or greater than 95%. "Exact Match" can be divided into the following four subcategories for additional specificity: (i) Byte-for-byte: the normalized texts are byte for byte identical; (ii) CMI stripped: the article text (excluding title, author name, publication name, copyright notice, and date) is identical to the training text, or the two match with a word-level Jaccard of at least 0.95. This encompasses instances in which the only difference between the texts is that OpenAI removed Copyright Management Information; (iii) Complete with chrome: identical article text (excluding the first and last 50 characters) is contained within the training text (the training text may contain additional content); (iv) High containment: all other matches in which at least 95% of the article's words appear in the training text and the training text is at least 85% the length of the article. Ziff Davis's "Revised Version" comparison indicates directional word containment is equal to or greater than 80% and 1-gram Jaccard similarity greater than 50%.

Summary. Attached hereto as Exhibit EE is a true and correct copy of the PCMag Rule 1006 Summary. Attached hereto as Exhibit FF is a true and correct copy of the IGN Rule 1006 Summary. Attached hereto as Exhibit GG is a true and correct copy of the Rock Paper Shotgun Rule 1006 Summary. Attached hereto as Exhibit HH is a true and correct copy of the VG247 Rule 1006 Summary. Attached hereto as Exhibit II is a true and correct copy of the Eurogamer Rule 1006 Summary. Attached hereto as Exhibit JJ is a true and correct copy of the Dicebreaker Rule 1006 Summary. Attached hereto as Exhibit KK is a true and correct copy of the GamesIndustry Rule 1006 Summary. Attached hereto as Exhibit LL is a true and correct copy of the Everyday Health Rule 1006 Summary. Attached hereto as Exhibit MM is a true and correct copy of the Medpage Today Rule 1006 Summary. Attached hereto as Exhibit NN is a true and correct copy of the Diabetes Daily Rule 1006 Summary. Attached hereto as Exhibit OO is a true and correct copy of the bpHope Rule 1006 Summary. Attached hereto as Exhibit PP is a true and correct copy of the Lose It Rule 1006 Summary. Attached hereto as Exhibit QQ is a true and correct copy of the Migraine Again Rule 1006 Summary. Attached hereto as Exhibit RR is a true and correct copy of the Daily OM Rule 1006 Summary. Attached hereto as Exhibit SS is a true and correct copy of the BabyCenter Rule 1006 Summary. Attached hereto as Exhibit TT is a true and correct copy of the What to Expect Rule 1006 Summary.

8. The subfolder called “Source” contains the underlying evidence supporting the accuracy of the ZD Rule 1006 Summaries:

Source	**Produced Ownership Documents** *Contains Produced Staff Lists, Vendor Agreements, IC Agreements, and other Employment Documents*	1,378
	Produced Text from CMS *Produced at ZD0618475*	4
	Full Asserted Work Text *Contains Full Contents of Asserted Work Text*	25

9. I understand that Column (H) of the ZD Rule 1006 Summaries references various freelance contracts, staff lists, vendor lists, and other BATES-stamped documents supporting Ziff Davis's claims of ownership over each of its works, such as ZD0647854 (Staff List) and ZD0516622 (Freelance Agreement). These documents are contained within a subfolder called "Produced Ownership Documents", and each are true and correct copies of how they were produced in this litigation and maintained in the ordinary course. Ex. A.

10. I understand that the Ziff Davis produced the full text for millions of its articles (ZD0618475), which was derived from CMS (content management system) databases maintained in the ordinary course. A true and correct copy of ZD0618475 is contained within a subfolder called "Produced Text from CMS". Because of the reconciliation process that occurred, there are more articles contained within ZD0618475 than in the ZD Asserted Works. Ex. A.

11. I understand that Column (G) or "article text" of the ZD Rule 1006 Summaries contains the full text for each of its Asserted Works. However, because of a limitation with the Excel application, only the first 32,000 characters of a work's full text appear in this column. For the Court's convenience, Ziff Davis is separately providing the full text for each of its Asserted Works in .parquet format, which does not contain the same Excel-limitation. These files are

contained within a subfolder called "Full Asserted Work Text", and each file corresponds to a distinct Ziff Davis brand.

The "2 – Training Data"

12. Ziff Davis is also providing Excel-spreadsheets showing each of OpenAI's training data records found to match to one of the ZD Asserted Works as an Exact Match or Revised Version (the "ZD Training Spreadsheets"). The ZD Training Match Spreadsheets include fields indicating, for each Asserted Work, where matching training record was found, the metrics used for confirming it is a match, and the text of the training record. I understand that the ZD Training Match Spreadsheets are 162 files located in the folder called "2-Training Data". Ex. A.

13. Attached hereto as Ex. B is a true and correct copy of the ZD Training Match Spreadsheets.

14. For the ZD Training Match Spreadsheets, each row represents an asserted work that matched on a training data record. Below is a description for each column provided in the training data spreadsheets:

a. "aw_id" refers to the unique identifier for the individual Asserted Work;

b. "training_data_id" is a unique identifier for OpenAI's training data record;

c. "url" refers to the webpage associated with the asserted work;

d. "title" refers to the title associated with the asserted work;

e. "author" refers to the author associated with the asserted work;

f. "aw_text" refers to the text of the asserted work;

g. "training_data_text" refers to the text in OpenAI's training data record;

h. "opco_bates" refers to the BATES number under which OpenAI produced the training data record;

i. "match_method" refers to the search methodology used to find and match the training data record. These search methods are described in the Supplemental Expert Report of Daniel P. Delorey ¶¶11-12 (available as Exhibit D to the Declaration of Daniel P. Delorey);

j. "category" this column refers to whether the matching training data record is an Exact Match or a Revised Version (as explained above and in SF 1216-1217);

k. "exact_reason" for Exact Matches, this column provides the subcategory within Exact Match that applies to the training data record. *See* SF 1216 for a description of these subcategories;

l. "jaccard_word" refers to the 1-gram jaccard score based on the comparison of the asserted work to the OpenAI training data record;

m. "jaccard_trigram" refers to the 3-gram jaccard score based on the comparison of the asserted work to the OpenAI training data record;

n. "containment_word" refers to the fraction of the asserted work's distinct words appearing in the matching OpenAI training data record;

o. "len_ratio" refers to the ratio of the asserted work's length to the length of the matching OpenAI training data record;

p. "aw_len / oai_len" refers to the ratio of the asserted work's length to the length of the matching OpenAI training data record after normalization (*e.g.*, removing extraneous white space and treating all letters as the same case);

q. "has_16gram" refers to whether the asserted work and OpenAI's matching training data record share a run of 16 consecutive words or more;

r. "goldstein_trigram_jaccard" refers to the 3-gram jaccard score based on the comparison of the asserted work to the OpenAI training data record, using an alternative normalization process that retains the case, and retains certain whitespace.

15. The subfolder called "source" contains the underlying evidence supporting the accuracy of the ZD Training Match Spreadsheets:

source	**Produced OpenAI Training Records** *OPCO_MDL_004898390–416 – HC AEO* *OPCO_MDL_004832874– 884 – HC AEO* *OPCO_MDL_004832887 – HC AEO*	39
	[BRAND]__aw_text_full.parquet (20) [BRAND]__training_data_text_full.parquet (20)	40

16. The files named in the style of "[BRAND]_aw_text_full.parquet" contain, for each brand, the full text of a given ZD Asserted Work and associated AW ID. These are being provided in .parquet format to avoid Excel's character-length limitation, organized by Brand.

17. The files named in the style of "[BRAND]_training_data_text_full.parquet" contain, for each brand, the text of the training records that matched and associated metadata, organized by Brand.

18. Ziff Davis is also providing the underlying training data produced by OpenAI that was used to create the training data spreadsheets. A true and correct copy of these records as they were produced in this litigation are located in the "Produced OpenAI Training Records" subfolder. Ex. A. OPCO_MDL_004898390–416; OPCO_MDL_004832874–884; OPCO_MDL_004832887.

The "3 - CMI Removal"

19. Ziff Davis is also providing spreadsheets showing, for each of the above-described training data records, whether OpenAI's records are missing copyright management information elements ("the ZD CMI Spreadsheets"). The ZD CMI Spreadsheets provide columns showing, for each training data record, whether the copyright notice, author, publisher, title, or terms of service were removed from the copy in OpenAI's training data record. I understand that the ZD CMI Spreadsheets are 46 files located in the folder called "3 – CMI Removal". Ex. A.

20. Attached hereto as Exhibit C is a true and correct copy of the ZD CMI Spreadsheets.

21. Below is a description for each column provided:

a. "cc_page_id" refers to the common crawl page identifier that used to locate the CMI;

b. "cc_fetch_ts" refers to when the common crawl page used was captured;

c. "copyright_cc_text" provides the copyright notice found on the common crawl page;

d. "author_cc_meta," and "author_cc_byline_extracted" provides whether the author was found on the common crawl page;

e. "site_name_cc_text" provides the publisher identity found on the common crawl page;

f. "og_title_cc_text" provides the title found on the common crawl page in the Open Graph field;

g. "html_title_cc_text" provides the title found on the common crawl page in HTML

h. "tos_cc_text" provides the terms of service were found on the common crawl page;

i. "copyright_removed" indicates whether the copyright notice was removed;

j. "author_removed" indicates whether the author was removed;

k. "site_name_removed" indicates whether the publisher identity was removed;

l. "og_title_removed" indicates whether the title found in the Open Graph field was removed;

m. "html_title_removed" indicates whether the title found in the html was removed;

n. "tos_removed" indicates whether the terms of service were removed.

22. Ziff Davis is also providing identifiers for the corresponding Common Crawl page because the CMI removal analysis compares training data records matching on asserted works to common crawl pages. Supplemental Expert Report of Daniel P. Delorey ¶¶ 40–43 (available as Exhibit D to the Declaration of Daniel P. Delorey). I understand that these files are contained in the subfolder called "common-crawl-captures", and are named in the style of "<Brand>_partAofZ". I understand that they are being provided in both Excel format for the Court's convenience and .parquet format to capture the full length of various fields. Ex. A.

23. I declare under penalty of perjury that the foregoing is true and correct.

Executed on September 4, 2026

/s/ Lacy H. Koonce, III

Lacy H. ("Lance") Koonce, III
Klaris Law PLLC
161 Water Street
New York, NY 10038
Telephone: (646) 779-4882
lance.koonce@klarislaw.com