

MORRISON FOERSTER

2100 L STREET, NW
 SUITE 900
 WASHINGTON
 DC 20037
 TELEPHONE: 202.887.1500
 FACSIMILE: 202.887.0763
 WWW.MOFO.COM

MORRISON & FOERSTER LLP
 AUSTIN, BERLIN, BOSTON, BRUSSELS,
 DENVER, HONG KONG, LONDON, LOS
 ANGELES, MIAMI, NEW YORK, PALO
 ALTO, SAN DIEGO, SAN FRANCISCO,
 SHANGHAI, SINGAPORE, TOKYO,
 WASHINGTON, D.C.

Writer’s Direct Contact
 +1 (202) 887-6945
 CMHomer@mofocom

November 13, 2024

Via ECF

Hon. Ona T. Wang
 United States District Court
 Southern District of New York
 500 Pearl Street
 New York, NY 10007-1312

Re: *Authors Guild, et al. v. OpenAI, Inc., et al.* No. 23-cv-8292-SHS
Jonathan Alter, et al. v. OpenAI, Inc., et al. No. 23-cv-10211-SHS

Dear Judge Wang:

OpenAI submits this status report regarding custodial statistics in response to the Court’s order (Dkt. 252) at the October 30, 2024 status conference regarding Plaintiffs’ motion to compel additional custodians (Dkt. 193, 230). OpenAI has already agreed to search and produce documents from 24 custodians; Plaintiffs seek 8 more. The 8 disputed custodians are: Chelsea Voss, Jan Leike, Ilya Sutskever, Andrew Mayne, Shantanu Jain, Qiming Yuan, Jong Wook Kim, and Cullen O’Keefe. Collectively, these 8 additional custodians have over 206 gigabytes of data comprising over 886,000 documents. The below chart reflects the custodial data and documents segregated for each of the 8 disputed custodians.

Custodian	Size (GB)	Files
Voss, Chelsea	20.33	118,511
Leike, Jan	8.96	78,428
Sutskever, Ilya	49.35	209,889
Mayne, Andrew	10.24	31,708
Jain, Shantanu	8.56	266,011
Yuan, Qiming	89.37	34,274
Wook Kim, Jong	8.54	55,636
O’Keefe, Cullen	11.21	92,145
<i>Total</i>	206.56	886,602

MORRISON FOERSTER

Hon. Ona T. Wang
 November 13, 2024
 Page 2

The parties have also been negotiating custodial search terms. At present, OpenAI’s proposed search terms, across the *existing custodians*, would require OpenAI to review more than 460,000 documents. Plaintiffs’ proposed terms, across the *existing custodians*, would require OpenAI to review over 1 million documents – far exceeding the Court’s target. See 10/30/2024 Hearing Tr. at 74:14-75:9 (COURT: “I’m going to impose right now a ceiling of 500,000 hits.”). These competing hit reports are summarized below.

Custodians	OpenAI Proposed Terms		Plaintiffs' Proposed Terms	
	Size (GB)	Files (incl. families)	Size (GB)	Files (incl. families)
24 existing custodians - total hits	359.06	462,492	682.87	1,090,800
8 disputed custodians - total hits	375.26	255,498	539.71	395,334
8 disputed custodians - de-duplicated hits	188.01	73,693	233.87	111,893
8 disputed custodians - duplication %		71.16%		71.70%
Total 32 custodians (de-duplicated)	547.1	536,185	916.7	1,202,693

Included in the above chart is OpenAI’s estimate of the duplication rate between the 8 disputed custodians and the existing 24 custodians. The duplication rate for both OpenAI and Plaintiffs’ proposed terms is 71%. Under Plaintiffs’ proposed search terms, the 8 additional custodians would add over 111,000 hits, requiring OpenAI to review over 1.2 million documents after de-duplication, more than double the Court’s target. Even under OpenAI’s proposed search terms, the additional 8 custodians would add over 70,000 hits, requiring OpenAI to review over 536,000 documents, still more hits than the Court’s target. In addition, for the productions to date, only around 25% of documents hitting on OpenAI’s proposed search terms have been responsive.

Pursuant to the Court’s instructions regarding prioritizing discovery at the October 30 hearing, and taking into account the substantial volume of hits, high custodial duplication rates, and low document responsiveness rates to date, OpenAI will continue to meet and confer with Plaintiffs in an attempt to reach agreement regarding both the 8 disputed custodians and a final set of targeted search terms.

HORRISON FOERSTER

Hon. Ona T. Wang
November 13, 2024
Page 3

Respectfully submitted,



Carolyn M. Homer