

**Lieff
Cabraser
Heimann &
Bernstein**
Attorneys at Law

SUSMAN GODFREY

June 6, 2025

VIA ECF

Judge William H. Alsup
Phillip Burton Federal Building & U.S. Courthouse
450 Golden Gate Avenue, 16th Floor Clerk's Office
San Francisco, CA 94102

RE: *Bartz et al. v. Anthropic PBC*, Case Number 3:24-cv-05417 (N.D. Cal.)
Motion to Compel Production of Anthropic Founder Records and the
Transcript of Benjamin Mann's March 3, 2025 Deposition

Dear Judge Alsup:

Pursuant to paragraph 35 of this Court's Supplemental Order to Order Setting Initial Case Management Conference, Plaintiffs request that the Court compel non-party OpenAI to produce highly relevant documents from Dario Amodei and Benjamin Mann, two former OpenAI employees who left to found Anthropic. Mann and Amodei led OpenAI's pirating of books from Library Genesis ("LibGen") from 2019 to 2020. That is the very same shadow library that *Anthropic* used to source pirated datasets shortly after its founding in 2021. Plaintiffs therefore seek (1) the transcript and exhibits of the March 3, 2025 deposition of Mann taken by the plaintiffs in the pending copyright class action against OpenAI¹; and (2) Mann and Amodei's documents and communications regarding pirated datasets from January 1, 2019 to December 31, 2020 while they were working at OpenAI.²

Background

On January 28, 2025, plaintiffs in copyright litigation against OpenAI and Microsoft³ moved to compel the deposition of Mann, explaining that a key allegation in their case concerned OpenAI's use of two pirated books datasets and that Mann and Amodei had primary responsibility for acquiring and using these datasets while at OpenAI. *See In re: In re Motion to Compel*

¹ *See In re: In re Motion to Compel Compliance with Non-Party Subpoenas*, No. 3:25-mc-80017 (N.D. Cal. Feb. 13, 2025) Dkt. 11 (granting motion to compel deposition of Benjamin Mann in *In re OpenAI ChatGPT Litigation*, No. 3:23-03223 (N.D. Cal.) and *Authors Guild et al v. OpenAI Inc. et al*, No. 1:23-cv-08292 (S.D.N.Y.)).

² These dates correspond with Mann's tenure as a member of OpenAI Technical Staff and Amodei's as Vice President of Research (and Research Director before that).

³ *Authors Guild et al v. OpenAI Inc. et al*, No. 1:23-cv-08292 (S.D.N.Y.).

Judge William H. Alsup
June 6, 2025
Page 2

Compliance with Non-Party Subpoenas, No. 3:25-mc-80017 (N.D. Cal. Feb. 13, 2025) Dkt. 1; Dkt. 17 at 3. Those books datasets were sourced from LibGen. *See* Ex. 4 (“[T]he use of books1 and books2 for model training was discontinued in late 2021 . . .”); “The Unbelievable Scale of AI’s Pirated-Books Problem,” *The Atlantic*, Mar. 20, 2025, <https://www.theatlantic.com/technology/archive/2025/03/libgen-meta-openai/682093/> (“OpenAI has used LibGen in the past . . . a spokesperson for OpenAI said, ‘The models powering ChatGPT and our API today were not developed using these datasets. These datasets, created by former employees who are no longer with OpenAI, were last used in 2021.’”); Ex. 5 (discussing “Libgen1” and “Libgen2” datasets). On February 13, 2025, Judge Illman compelled Mann to testify at a deposition in the OpenAI action, including about his acquisition of pirated datasets for training LLMs. *See* Dkt. 11, 3:25-mc-80017.⁴

On March 15, Plaintiffs served a Rule 45 subpoena upon OpenAI⁵ seeking documents about these topics, OpenAI refused to produce anything, and the parties have held two Zoom meet and confers. Ex. 1.⁶ Plaintiffs offered to narrow the scope of the Requests to (1) the transcript and exhibits of Mann’s March 3, 2025 deposition; and (2) Mann’s documents and communications regarding pirated datasets while employed at OpenAI. *See* Ex. 3 at 2–3. On June 5, OpenAI rejected Plaintiffs’ narrowed proposal simply because Plaintiffs reserved the right to seek additional transcripts and documents if doing so became necessary. *Id.* at 2. This motion follows.

Plaintiffs’ Motion to Compel Should Be Granted

The material is relevant. Because Anthropic’s acquisition and use of pirated books datasets from shadow libraries like LibGen is at the core of this case, their key executive’s prior use of those datasets—and knowledge of their illegality—is highly relevant. *See, e.g.*, Pls. MSJ Op. Br., Dkt. 158 at 1–4, 11; Pls. Mot. Class Cert Br., Dkt. 125 at 5–7, 21; FAC, Dkt. 70 ¶¶ 18, 44–45.

First, Mann and Amodei’s pirating of datasets while employed at OpenAI—and immediately prior to their founding of Anthropic—bears directly on willfulness. The copyright statute provides enhanced statutory damages in cases of willful infringement, *see* 17 U.S.C. § 504(c)(2), which occurs when “the defendant recklessly disregarded the possibility that its conduct represented infringement” or acted “with knowledge that [its] conduct constitutes copyright infringement.” *Wall Mountain Co. v. Edwards*, 2010 WL 4940778, at *2 (N.D. Cal.

⁴ Judge Illman subsequently granted a motion to compel the deposition of Amodei, *see In re: In re Motion to Compel Compliance with Non-Party Subpoenas*, No. 3:25-mc-80017 (N.D. Cal. Feb. 13, 2025) Dkt. 28, however that deposition has yet to take place.

⁵ Plaintiffs previously sought these materials from Anthropic. *See* Ex. 2. Anthropic stated that it was not in possession, custody, or control of Mann’s transcript. Geman Decl. at ¶ 3.

⁶ At OpenAI’s request, Plaintiffs served an amended Rule 45 subpoena on March 24, 2025 to change the place of compliance to N.D. Cal. *See* Ex. 6.

Judge William H. Alsup
June 6, 2025
Page 3

Nov. 30, 2010) (cleaned up). Evaluating willfulness, therefore, “requires an assessment of a party’s state of mind.” *Id.*; see also *Sega Enters. Ltd. v. MAPHIA*, 948 F. Supp. 923, 936 (N.D. Cal. 1996) (same).

Evidence of Mann and Amodei’s states of mind immediately preceding their founding of Anthropic with respect to the conduct here are likely to evince whether, for example, they “knew there was a risk that” the books were pirated “but nonetheless went forward” with downloading them anyway while at Anthropic. *Friedman v. Live Nation Merch., Inc.*, 833 F.3d 1180, 1187 (9th Cir. 2016) (reversing district court’s grant of summary judgment on issue of willfulness in favor of defendant). So too might Mann and Amodei’s conduct at OpenAI show whether they in fact acted “with knowledge” of the pirated nature of the works. *Wall Mountain*, 2010 WL 4940778, at *2. Further, Mann also will be a critical deponent in this case and Plaintiffs’ obtaining his prior sworn testimony is thus necessary for evaluation of the credibility of his testimony here.

Second, the documents are also relevant to the first fair use factor—the “purpose and character” of Anthropic’s use. These prior experiences may inform whether Anthropic pirated books to obtain commercial or competitive advantages, as well as Anthropic’s understanding of why books were critical training data for their models. See Pls. MSJ Op. Br., Dkt. 158 at 4, 7–10, 14–15; *VHT, Inc. v. Zillow Grp., Inc.*, 918 F.3d 723, 742 (9th Cir. 2019) (“The first factor of the fair use analysis assesses the character of the use, including whether it is commercial in nature”).

OpenAI has not shown any undue burden. OpenAI has the deposition transcript, and it appears to have *already* collected and reviewed the relevant documents from Mann and Amodei as part of the *OpenAI* litigation. See, e.g., *In re Uber Techs., Inc., Passenger Sexual Assault Litig.*, No. 23-md-03084, 2024 WL 4950060, at *2 (N.D. Cal. Dec. 2, 2024) (ordering production of “documents that have already been disclosed to other litigation adversaries” because they “impose[] minimal burden”); *San Francisco Bay Area Rapid Transit Dist. v. Spencer*, No. 04-cv-04632, 2006 WL 2734284, at *2 (N.D. Cal. Sept. 25, 2006) (ordering production of deposition transcripts from other actions over burden objections). Even if there were any meaningful burden, it would not come close to outweighing the relevance of the material sought. Courts frequently consider “relevance as a factor related to undue burden when determining motions to quash a subpoena.” See, *J.T. v. City of San Francisco*, No. 23-cv-06524, 2024 WL 4361579, at *1 (N.D. Cal. Oct. 1, 2024).

For these reasons, Plaintiffs respectfully request an order directing OpenAI to produce (1) the transcript and exhibits of the March 3, 2025 deposition of Mann; and (2) Mann and Amodei’s documents and communications regarding pirated datasets Books3, Bibliotik, The Pile, Library Genesis, Anna’s Archive, and Z-library from the period of January 1, 2019 to December 31, 2020.

Judge William H. Alsup
June 6, 2025
Page 4

Respectfully Submitted,

LIEFF CABRASER HEIMANN
& BERNSTEIN, LLP

SUSMAN GODFREY LLP

/s/ Rachel Geman
Rachel Geman

/s/ Rohit Nath
Rohit Nath

Judge William H. Alsup
June 6, 2025
Page 5

ATTESTATION

Pursuant to Civil Local Rule 5-1(i)(3), I hereby attest that all signatories listed, and on whose behalf the filing is submitted, concur in the filing's content and have authorized the filing.

Dated: June 6, 2025

/s/ Rachel J. Geman