[Counsel on signature page]

UNITED STATES DISTRICT COURT

NORTHERN DISTRICT OF CALIFORNIA

SAN JOSE DIVISION

CONCORD MUSIC GROUP, INC., ET AL.,

Plaintiffs,

v.

ANTHROPIC PBC,

Defendant.

Case Number: 5:24-cv-03811-EKL-SVK

JOINT DISCOVERY DISPUTE STATEMENT REGARDING ANTHROPIC'S CONFIDENTIALITY DESIGNATIONS

Judge Eumi K. Lee
Magistrate Judge Susan van Keulen

Pursuant to Section 8 of the Court's Civil Standing Order, and L.R. 37-1 and 37-2, Plaintiffs ("Publishers") and Defendant Anthropic PBC ("Anthropic") respectfully submit this Joint Discovery Dispute Statement. The Parties seek the Court's intervention in resolving a dispute regarding Anthropic's confidentiality designations of the "raw text [and] extracted text . . . from LibGen and PiLiMi data sets," and Publishers' challenges to those designations as improper. The Parties' lead counsel met and conferred to try to resolve this dispute via videoconference on Sept. 23, 2025 and by email. The Parties have been unable to resolve the issues below. The current fact discovery deadline is 20 days away. ECF No. 416. Publishers cannot offer a compromise as to the pirated files Anthropic downloaded from LibGen and PiLiMi, given the lack of a basis for Anthropic's confidentiality designations of those materials. As to the two challenged designations remaining as part of this dispute, Anthropic cannot offer a compromise, because the raw text and extracted text at-issue reveal proprietary trade secret information about the data processing techniques and extraction decisions that Anthropic has made to develop its state of the art generative AI models.

I. Publishers' Position: Anthropic should withdraw its improper confidentiality designations of the pirated files it downloaded from LibGen and PiLiMi.

Anthropic continues to abuse the Protective Order in this case by over-designating information that clearly does not quality for "Attorneys' Eyes Only" or "Confidential" protection.

The Court has already admonished Anthropic for its improper over-designation of materials under the Protective Order. *See, e.g.*, May 13, 2025 Hearing Tr. 60:6-14 (the Court's expressing "dismay" that "[Anthropic] failed to acknowledge that there clearly is an issue here of over-designation," and emphasizing that "the parties are admonished, and the parties stipulate, to take care to limit any confidential designations to specific material that qualifies under the appropriate standards"); ECF No. 377 at 7 (emphasizing that "Publishers' concerns regarding [confidentiality] over-designation by Anthropic are not unfounded," and down-designating certain information over-designated by Anthropic as "Attorneys' Eyes Only").

Ignoring those instructions, Anthropic has improperly designated as "Attorneys' Eyes Only" all of the "raw text [and] extracted text . . . from LibGen and PiLiMi data sets" that the

Court ordered it to produce in response to Publishers' BitTorrent-related discovery requests. ECF No 431 at 1.

But ***none*** of these files constitute or contain any of Anthropic's confidential information. Rather, these files are illegal pirated copies of copyrighted works owned by others, which Anthropic downloaded via torrenting from third-party shadow libraries LibGen and PiLiMi. The "raw text" files are the actual torrented files that Anthropic illegally downloaded, while the "extracted text" files are simply copies of the text of those torrented files. These illegal files are not confidential, as they can be accessed through LibGen and PiLiMi by others who, like Anthropic, are willing to break the law by downloading them from these pirate libraries. Anthropic in particular has no basis for asserting confidentiality over these files, given that it did not create them and it has no rights to them. And given the nature of peer-to-peer infringement, Anthropic distributed copies of these same torrented files to unknown third parties during the torrenting process, further undermining any confidentiality claims as to those files. *See, e.g.*, *Columbia Pictures Indus., Inc. v. Fung*, 710 F.3d 1020, 1027 (9th Cir. 2013) ("[E]ach [BitTorrent] user is both downloading and uploading several different pieces of a file from and to multiple other users."); *Strike 3 Holdings, LLC v. Doe*, 2019 WL 1277561, at *1 (N.D. Cal. Mar. 20, 2019) ("BitTorrent users . . . simultaneously download and upload pieces of the file from and to other users.").

Nor is Anthropic's use of these illegal files to train its AI models in any way confidential. To the contrary, the facts of Anthropic's illegal downloading of millions of copyrighted works from shadow libraries LibGen and PiLiMi and its use of those pirated files for AI training has been discussed in detail in Judge Alsup's publicly available opinions in the *Bartz v. Anthropic* case. *See, e.g.*, ECF No. 411 at 2-3. These facts have also been reported widely in the public news media. *See, e.g.*, *Anthropic Agrees to Pay $1.5 Billion to Settle Lawsuit With Book Authors*, N.Y. TIMES (Sept. 5, 2025), https://www.nytimes.com/2025/09/05/technology/anthropic-settlement-copyright-ai.html ("[The *Bartz* court] found that Anthropic had illegally acquired millions of books through online libraries like Library Genesis and Pirate Library Mirror that many tech companies have used to supplement the huge amounts of digital text needed to train A.I. technologies. When

Anthropic downloaded these libraries, the judge ruled, its executives knew they contained pirated books."). These illegally torrented files in no way qualify as a confidential AI training dataset under the Protective Order, as Anthropic apparently claims.

There is absolutely no basis for Anthropic to seek to maintain these documents as "Attorneys' Eyes Only" or "Confidential." *See, e.g.*, ECF No. 377 at 8 (the Court's recognizing that "confidentiality is not appropriate" for otherwise public material).

Anthropic cannot invoke the Protective Order to conceal its infringing activity. Judge Chhabria emphatically rejected similar over-designations by another AI company in a copyright infringement case, given that they were "not designed to protect against the disclosure of sensitive business information that competitors could use," but rather "to avoid negative publicity." *Kadrey v. Meta Platforms, Inc.*, No. 23-cv-03417-VC (N.D. Cal. 2025), ECF No. 373.

Anthropic fails to make anything close to the required particularized good cause showing to justify the above confidentiality designations of the pirated files it downloaded from LibGen and PiLiMi. Just as there is "a strong presumption in favor of access to court records," *Foltz v. State Farm Mut. Auto. Ins.*, 331 F.3d 1122, 1135 (9th Cir. 2003), Anthropic's "[p]retrial discovery is presumptively public," *Doe 1 v. Wolf*, 2020 WL 2745266 (N.D. Cal. 2020). Anthropic can overcome this presumption only upon a "particularized showing of good cause." *San Jose Mercury News, Inc. v. U.S. Dist. Ct.*, 187 F.3d 1096, 1103 (9th Cir. 1999). Even if it shows "particularized harm" from disclosure, the Court weighs that against the public interest and whether redactions could allow disclosure. *In re Roman Cath. Archbishop of Portland*, 661 F.3d 417, 425 (9th Cir. 2011). The Anthropic discovery above does not meet the very high standard for "Attorneys' Eyes Only" treatment under the Protective Order, which limits such protections to only "[i] extremely sensitive 'Confidential Information or Items,' [ii] disclosure of which . . . would create a substantial risk of serious harm . . . [iii] that could not be avoided by less restrictive means," like redactions. ECF No. 293, ¶ II.12 (emphasis added).

Anthropic cannot cloak this key discovery in confidentiality designations and obstruct Publishers' ability to use it in this case. Without the required showing to justify these designations,

Anthropic cannot justify the onerous restrictions it seeks to impose on Publishers' ability to access and utilize this critical discovery.

Accordingly, Anthropic should remove all confidentiality designation of the pirated files it downloaded from LibGen and PiLiMi.

II. Anthropic's Position

Publishers have ginned up an avalanche of last-minute, overreaching discovery disputes without engaging meaningfully in any discussion of compromise on their own requests and refusing to provide even the most basic discovery to Anthropic. Instead, they have insisted on declaring impasse on every issue, presumably to support Publishers' bid to extend the discovery schedule and turn this case into *Bartz II*. See Dkt. 435. This is one such dispute.

Pursuant to the Court's September 8, 2025 Order, Anthropic produced on September 11 "the raw text, extracted text and metadata from LibGen and PiLiMi data sets." Dkt. 431. Anthropic has informed Publishers that it is not maintaining confidentiality designations over the metadata from the LibGen and PiLiMi datasets–the metadata files were not created by Anthropic and are overinclusive of the datasets that Anthropic actually obtained. But Anthropic has appropriately designated the raw text and extracted text datasets as Highly Confidential – Attorneys' Eyes Only.

Rather than engage on Anthropic's explanation for why this data is appropriately designated confidential, Publishers argue that the raw and extracted text datasets simply cannot be confidential because they were obtained from third parties. Not so. Publishers conflate the *fact* that Anthropic sourced data from certain sources, with the precise *content* of those datasets. The extracted text datasets are processed from various file types contained in the raw files into extracted text according to technical decisions and effected through source code. Removing confidentiality protections over the contents of the extracted text dataset risks allowing a competitor to derive the technical choices made and the method used to extract the datasets. That the process by which Anthropic extracts and processes data is not known to Anthropic's competitors is central to Anthropic's position as an industry leader in generative AI model development. These processes amount to trade secrets, which Anthropic deliberately keeps confidential to maintain its

competitive market standing. Similarly, removing protection over the raw datasets would reveal precisely the contents of the dataset that Anthropic possessed and used as part of its LLM development.

Publishers' argument that the files are not confidential because they can be "accessed through LibGen and PiLiMi by others" is also wrong. The data in the raw and extracted text datasets may be a subset of what can be accessed by others today, but the precise data within the files Anthropic maintained and the technical processing steps used to extract those datasets is not publicly accessible. The contents of the extracted datasets reveal the choices as to certain file and content attributes that were filtered for extraction, and the technical processing steps used to extract readable text from a randomized assortment of files and file types; choices that are unquestionably proprietary and competitively sensitive. As such, and given their competitively sensitive nature, the extracted and raw text datasets are properly maintained as HC-AEO, consistent with the Court's Order. Dkt. 377 at 8[1].

Publishers' position is also inconsistent with the Court's prior order permitting Anthropic to maintain as highly confidential different training datasets used by Anthropic. *See* Dkt. 377 at 8 ("[T]he Court agrees with Anthropic that, as a general matter, the training datasets used to train Claude are competitively sensitive and may properly be maintained as 'HC-AEO' material.").

Without any basis, Publishers declare that Anthropic "distributed copies" of these datasets "to unknown third parties." Publishers cite no evidence that this was in fact the case. They simply cite to caselaw about torrenting generally. But of course, general descriptions from other cases about how torrenting *can* be done say nothing about whether Anthropic has in fact disclosed the extracted text from these data sets to anyone. Publishers' motion would require the Court to make factual findings about the torrenting claim that is not yet part of the case–namely, whether the datasets were "distributed." And there is no argument that anyone torrented or shared the datasets processed through proprietary means to extract the text of some of the raw files.

[1] Publishers' position is also at odds with the position they have taken in the *Bartz* litigation, where Publishers' counsel has noticed in as "Coordination Counsel." In *Bartz*, the plaintiffs sought to seal the final works list, derived from the same datasets, arguing that making the list public would allow "non-class members to submit fraudulent claims." *Bartz v. Anthropic*, Dkt. 421 at 2.

And critically, Publishers have not articulated any compelling reason—or any reason at all—that they need to disclose these datasets more broadly than the current designation allows. Their expert can review the datasets under the Protective Order as he has other Highly Confidential information produced by Anthropic in the litigation.

Tellingly, during the parties' meet and confer on this issue, Publishers refused to engage with Anthropic's explanation for its confidentiality designation, instead suggesting that data from third parties can never be confidential. What is clear is that this dispute—like several others Publishers have insisted on bringing to the Court—is a back-door request for the Court to wade into factual questions related to claims that are not in the case, including arguments related to torrenting and the so-called "pirate libraries" that are the subject of their Motion for Leave to Amend. The Motion for Leave to Amend has not yet been resolved by Judge Lee and Plaintiffs should not be permitted to use a barrage of discovery disputes to ask the Court to adjudicate the facts of their not-yet permitted claims.

Dated: October 1, 2025

By: /s/ Nicholas C. Hailey

OPPENHEIM + ZEBRAK, LLP
Matthew J. Oppenheim
Nicholas C. Hailey
Audrey L. Adu-Appiah
(admitted *pro hac vice*)
4530 Wisconsin Ave., NW, 5th Floor
Washington, DC 20016
Telephone: (202) 480-2999
matt@oandzlaw.com
nick@oandzlaw.com
aadu-appiah@oandzlaw.com

Jennifer Pariser
Andrew Guerra
Bret Matera
Timothy Chung
Michelle Gomez-Reichman
(admitted *pro hac vice*)
461 5th Avenue, 19th Floor
New York, NY 10017
Telephone: (212) 951-1156
jpariser@oandzlaw.com
andrew@oandzlaw.com
tchung@oandzlaw.com
bmatera@oandzlaw.com
mgomez-reichman@oandzlaw.com

COBLENTZ PATCH DUFFY & BASS LLP
Jeffrey G. Knowles (SBN 129754)
One Montgomery Street, Suite 3000
San Francisco, CA 94104
Telephone: (415) 391-4800
ef-jgk@cpdb.com

COWAN, LIEBOWITZ & LATMAN, P.C.
Richard S. Mandel
Jonathan Z. King
Richard Dannay
(admitted *pro hac vice*)
114 West 47th Street
New York, NY 10036-1525
Telephone: (212) 790-9200
rsm@cll.com
jzk@cll.com
rxd@cll.com

Attorneys for Plaintiffs

Respectfully submitted,

By: /s/ Joseph R. Wetzel

LATHAM & WATKINS LLP
Joseph R. Wetzel (SBN 238008)
joe.wetzel@lw.com
Andrew M. Gass (SBN 259694)
andrew.gass@lw.com
Brittany N. Lovejoy (SBN 286813)
britt.lovejoy@lw.com
505 Montgomery Street, Suite 2000
San Francisco, California 94111
Telephone: +1.415.391.0600

Sarang V. Damle
(admitted *pro hac vice*)
sy.damle@lw.com
Sara Sampoli (SBN 344505)
sara.sampoli@lw.com
555 Eleventh Street NW, Suite 1000
Washington, DC 20004
Telephone: +1.202.637.2200

Allison L. Stillman
(admitted *pro hac vice*)
alli.stillman@lw.com
1271 Avenue of the Americas
New York, New York 10020
Telephone: +1.212.906.1747

Attorneys for Defendant

Respectfully submitted,

By: /s/ Sonal N. Mehta

WILMER CUTLER PICKERING HALE AND DORR LLP
Sonal N. Mehta (SBN 222086)
sonal.mehta@wilmerhale.com
Allison Bingxue Que (SBN 324044)
allison.que@wilmerhale.com
2600 El Camino Real, Suite 400
Palo Alto, CA 94306
(650) 858-6000
Fax: (650) 858-6100

Ari Holtzblatt (SBN 354631)
ari.holtzblatt@wilmerhale.com
Robin C. Burrell (admitted *pro hac vice*)
robin.burrell@wilmerhale.com
2100 Pennsylvania Avenue NW
Washington, DC 20037
(202) 663-6964

Louis W. Tompros (admitted *pro hac vice*)
louis.tompros@wilmerhale.com
60 State Street
Boston, MA 02109
(617) 526-6886

Joseph Taylor Gooch (SBN 294282)
taylor.gooch@wilmerhale.com
50 California Street
Suite 3600
San Francisco, CA 94111
(628) 235-1002

Attorneys for Defendant

SIGNATURE ATTESTATION

Pursuant to Civil L.R. 5-1(i)(3), I hereby attest that concurrence in the filing of this document was obtained from all other signatories of this document. I declare under penalty of perjury that the foregoing is true and correct.

Dated: October 1, 2025

/s/ Bret Matera
Bret Matera