

791 F.Supp.3d 1038  
United States District Court, N.D. California.

Andrea BARTZ, Andrea Bartz Inc., Charles Graeber, Kirk Wallace Johnson, and MJ + KJ Inc., Plaintiffs,  
v.  
**ANTHROPIC PBC**, Defendant.

No. C 24-05417 WHA

|

Signed July 17, 2025

### **Synopsis**

**Background:** Authors and loan-out companies brought a putative class action against artificial intelligence (AI) company, alleging that AI company had infringed their federal copyrights by amassing copyrighted books from the internet without paying for copies and by not obtaining licenses to reproduce further copies for large language model (LLM) training. Plaintiffs moved for class certification.

**Holdings:** The District Court, [William H. Alsup](#), J., held that:

authors who had transferred all their rights to loan-out companies were not adequate class representatives;

loan-out companies were adequate representatives;

author who retained rights to his works was adequate representative;

proposed co-lead law firms for class representatives were adequate class counsel;

infringement claims could be shown with classwide evidence as to AI company's alleged conduct in downloading books through peer-to-peer downloads from decentralized pirate libraries for which catalogs of bibliographic metadata existed;

infringement claims could not be shown with classwide evidence as to alleged downloading of books from decentralized pirate library for which no bibliographic metadata existed; and

plaintiffs' claims met superiority requirement for certification.

Motion granted in part and denied in part.

**Procedural Posture(s):** Request or Application for Class Certification.

### **Attorneys and Law Firms**

[Justin A. Nelson](#), Pro Hac Vice, [Alejandra C. Salinas](#), Collin Fredricks, Pro Hac Vice, Susman Godfrey L.L.P., Houston, TX, [Reilly Todd Stoler](#), Amelia Haselkorn, [Daniel M. Hutchinson](#), [Jalle H. Dafa](#), Lieff Cabraser Heimann & Bernstein, LLP, San Francisco, CA, Anna Josefina Freymann, Danna Z. Elmasry, Jacob S. Miller, [Rachel Geman](#), Lieff Cabraser Heimann & Bernstein, LLP, New York, NY, [CeCe Cole](#), Pro Hac Vice, [Scott Jonathan Sholder](#), Pro Hac Vice, Cowan, DeBaets, Abrahams & Sheppard LLP, New York, NY, [James Craig Smyser](#), Pro Hac Vice, [Samir Doshi](#), Pro Hac Vice, Susman Godfrey LLP, New York, NY, [Samuel Issacharoff](#), Pro Hac Vice, New York, NY, Betsy A. Sugar, Pro Hac Vice, Wesley Dozier, Lieff Cabraser

Heimann and Bernstein, LLP, Nashville, TN, [Jordan Connors](#), Pro Hac Vice, Susman Godfrey L.L.P., Seattle, WA, [Michael Brent Adamson](#), Pro Hac Vice, [Michael Gervais](#), Rohit Dwarka Nath, [Molly Alana Karlin](#), Susman Godfrey L.L.P., Los Angeles, CA, for Plaintiffs Andrea Bartz, [Kirk Wallace Johnson](#).

Justin A. Nelson, Pro Hac Vice, [Alejandra C. Salinas](#), Collin Fredricks, Pro Hac Vice, Susman Godfrey L.L.P., Houston, TX, [Reilly Todd Stoler](#), Amelia Haselkorn, [Daniel M. Hutchinson](#), [Jalle H. Dafa](#), Lieff Cabraser Heimann & Bernstein, LLP, San Francisco, CA, [Brandt Silverkorn](#), Edelson PC, San Francisco, CA, Anna Josefina Freymann, Danna Z. Elmasry, Jacob S. Miller, [Rachel Geman](#), Lieff Cabraser Heimann & Bernstein, LLP, New York, NY, [CeCe Cole](#), Pro Hac Vice, [Scott Jonathan Sholder](#), Pro Hac Vice, Cowan, DeBaets, Abrahams & Sheppard LLP, New York, NY, [James Craig Smyser](#), Pro Hac Vice, [Samir Doshi](#), Pro Hac Vice, Susman Godfrey LLP, New York, NY, [Samuel Issacharoff](#), Pro Hac Vice, New York, NY, Betsy A. Sugar, Pro Hac Vice, Wesley Dozier, Lieff Cabraser Heimann and Bernstein, LLP, Nashville, TN, [Corey Miller](#), Pro Hac Vice, [Jeffrey M. Gould](#), Pro Hac Vice, [Matthew J. Oppenheim](#), Pro Hac Vice, Oppenheim & Zebrak LLP, Washington, DC, [J. Eli Wade-Scott](#), Pro Hac Vice, [Jay Edelson](#), Pro Hac Vice, Edelson PC, Chicago, IL, [Jordan Connors](#), Pro Hac Vice, Susman Godfrey L.L.P., Seattle, WA, [Michael Brent Adamson](#), Pro Hac Vice, [Michael Gervais](#), Rohit Dwarka Nath, [Molly Alana Karlin](#), Susman Godfrey L.L.P., Los Angeles, CA, for Plaintiff Charles Graeber.

Justin A. Nelson, Collin Fredricks, Susman Godfrey L.L.P., Houston, TX, [Rohit Dwarka Nath](#), [Michael Brent Adamson](#), Pro Hac Vice, [Michael Gervais](#), [Molly Alana Karlin](#), Susman Godfrey L.L.P., Los Angeles, CA, Amelia Haselkorn, [Daniel M. Hutchinson](#), [Jalle H. Dafa](#), [Reilly Todd Stoler](#), Lieff Cabraser Heimann & Bernstein, LLP, San Francisco, CA, [Brandt Silverkorn](#), Pro Hac Vice, Edelson PC, San Francisco, CA, Betsy A. Sugar, Pro Hac Vice, Lieff Cabraser Heimann and Bernstein, LLP, Nashville, TN, [Corey Miller](#), Pro Hac Vice, [Jeffrey M. Gould](#), Pro Hac Vice, [Matthew J. Oppenheim](#), Pro Hac Vice, Oppenheim & Zebrak LLP, Washington, DC, [J. Eli Wade-Scott](#), Pro Hac Vice, [Jay Edelson](#), Pro Hac Vice, Edelson PC, Chicago, IL, [James Craig Smyser](#), [Samir Doshi](#), Susman Godfrey LLP, New York, NY, [Rachel Geman](#), Lieff Cabraser Heimann & Bernstein, LLP, New York, NY, [Samuel Issacharoff](#), Pro Hac Vice, New York, NY, for Plaintiffs MJ + KJ, Inc., Andrea Bartz, Inc.

Douglas Andrew Winthrop, Jessica Lim Gillotte, Joseph Richard Farris, Pieter S. De Ganon, Arnold & Porter Kaye Scholer LLP, San Francisco, CA, [Kathleen R. Hartnett](#), Cooley LLP, San Francisco, CA, [Andrew Michael Gass](#), [Joseph Richard Wetzel](#), Latham & Watkins LLP, San Francisco, CA, [Daniel M. Hutchinson](#), Lieff Cabraser Heimann & Bernstein, LLP, San Francisco, CA, [Daralyn J. Durie](#), [Ramsey Fisher](#), Jackson Lane, Morrison & Foerster LLP, San Francisco, CA, Aditya Vijay Kamdar, Fitz Beckwith Collings, Pro Hac Vice, [Mary Prendergast](#), Morrison & Foerster LLP, Washington, DC, [Alexander J. Kasner](#), [Ephraim A. McDowell](#), Pro Hac Vice, Cooley LLP, Washington, DC, [Allyson C. Myers](#), [Oscar Daniel Ramallo](#), [Ryan M. Nishimoto](#), [Angel Tang Nakamura](#), Arnold & Porter Kaye Scholer LLP, Los Angeles, CA, Mark Alan Lemley, Lex Lumina LLP, Los Angeles, CA, [Whitney Rose O'Byrne](#), Morrison & Foerster LLP, Los Angeles, CA, [Assad Hussain Rajani](#), Arnold & Porter Kaye Scholer LLP, Palo Alto, CA, Estayvaine Bragg, Arnold & Porter Kaye Scholer LLP, San Francisco, CA, for Defendant.

## ORDER ON CLASS CERTIFICATION

[WILLIAM ALSUP](#), United States District Judge

### \*1045 INTRODUCTION

This civil action exemplifies the classic litigation that should be certified as a representative action, for the entire class stands aggrieved by defendant's downloading of their books from pirate libraries on the internet. It will be straightforward to prove the classwide wrong done. In short order, counsel can prove defendant violated the Copyright Act by doing Napster-style downloading of millions of works. Defendant, however, objects that it will be a \*1046 litigation nightmare to determine the specific works it downloaded and, as to each, who is the rightful victim of the infringement.

## STATEMENT

At issue on class certification is whether it will be so impractical to identify the millions of books downloaded by defendant Anthropic PBC from pirate libraries — let alone to match them to the authors, publishers, and the like who own the copyrights in them — that class certification should be denied.

### 1. THE PIRATED BOOKS.

In early 2021, a co-founder of Anthropic downloaded 196,640 unauthorized copies of copyrighted books from the pirate library known as *Books3*. This library conveniently packaged for mass download pairs of each book's extracted text and filename, enabling the books to be readily rebuilt into separate files or reviewed (see Opp. Expert Iyyer ¶¶ 20–21, 23–24, 27 n.10, 39).<sup>1</sup> The pirate behind *Books3* publicly downplayed any “copyright backlash” from these efforts (see Amd. Compl. ¶ 40). Anthropic's co-founder downloaded the pirated books to avoid the trouble of paying for them, hoping they might prove useful for training large language models (LLMs) or for something else.

Later that year, Anthropic wanted still more. But centralized pirate libraries were being shut down by the government. So, in June 2021, Anthropic's co-founder used the infamous BitTorrent protocol to copy books peer-to-peer from decentralized copies of another pirate library — Library Genesis, or *LibGen* (see ¶ 45). He set a computer program to pause torrenting if disk space got “exhausted” (Br. Exh. 19 at -0389775). These five million copies again came from ebooks, this time remaining in file types like .epub and .pdf to be “viewed as eBooks.” They generally had covers including title and authors (see ¶¶ 45, 51 n.25, 60).

As pirate libraries were still being shuttered, third parties online copied *LibGen* and improved upon it to create a new pirate library, simply called “Z-Library.” The FBI shut down Z-Library, too (Br. Exh. 38). But by then, Z-Library had itself been copied or “mirrored” to create the Pirate Library Mirror, or *PiLiMi*. When Anthropic's co-founder saw *PiLiMi* was ready for torrenting, he messaged Anthropic coworkers: “[J]ust in time!” One answered: “zlibrary my beloved” (Br. Exh. 20 at -0391318).

So, in July 2022, Anthropic torrented at least two million copies of pirated books from distributed copies of *PiLiMi*. These copies were similarly of .epub, .pdf, and .txt files of digital books (¶ 46). Because *LibGen* and *PiLiMi* had a common lineage, they had many books in common, too. Appreciating this overlap, about one month before this download, Anthropic engineers had compared Anthropic's five million copies torrented from *LibGen* against *PiLiMi*'s seven million copies available to torrent, and torrented only the two million copies “that [we]re not in Lib[G]en,” as the co-founder explained (Br. Exh. 20 at -0391318).

Plaintiffs thus allege that up to seven million copies of books were stolen by Anthropic in this Napster-like way from all three pirate libraries. Although neither side has systematically tallied the overlap in the books copied from each pirate library, all agree there remains obvious overlap **\*1047** at least in popular titles obtained from each (see Br. 9; Opp. 23 (estimating damages from five million unique works); Reply 6 n.8; ¶¶ 93–102 & n.35–38 (explaining that 65, 70, and 78 of 82 popularly listed titles tested appeared in each set)). Anthropic retained all these copies of pirated books in a central library or general data area “forever” (see, e.g., Br. Exh. 12 at -0144509).

Starting in 2024, instead of pirating more books, Anthropic purchased and scanned books. To do so, Anthropic's chief executive officer and head of training first discussed the “cost to buy all the books in [L]ib[G]en” and the “probability for any given item we can buy it” (Reply Exh. 43 at -0227005). Then its purchasers populated a database with complete metadata for books the company wished to but had not yet purchased (Reply Exh. 45 at - 0365971).

### 2. THE MEANS TO IDENTIFY BOOKS.

Anthropic itself needed to identify the works it copied. It did so using common metadata and methods including hashing. There were several reasons:

*First*, early on, Anthropic realized it had pirated multiple copies of some books into its central library and that using multiple copies of one text to train an LLM caused the model to memorize the text and perform less well. *Second*, Anthropic held in its central library everything from textbooks to trade fiction. It wanted to understand how the breadth and depth of categories selected for recopying to train LLMs affected performance, if at all. *Third*, other parties sometimes asked Anthropic to remove “discrete corpora” of books (see Br. Exh. 1 at 7–8; see also Br. Expert Zhao ¶ 72). *Fourth*, Anthropic was concerned that its completed AI service would regurgitate copyrighted training content, so Anthropic needed to match its LLMs’ output against known works and bar such output (Reply Exh. 46 at -0533421). *Fifth*, once Anthropic stopped using pirate-sourced copies of books to train LLMs, its models suffered a “performance hit.” It wanted to “recover” from this by finding books at least as good as the “old-books quality,” meaning the pirated ones (see MSJ Opp. Exh. 20 at -0391439). This prompted an effort to identify books to purchase (but purchase only once) (see also Br. Exh. 12 at -0144507 (“deduping”); Reply Exh. 45 at -0365971). *And*, sometimes its researchers “want[ed] to search for a book” (Reply Exh. 45 at -0365941).

So, Anthropic obtained catalogs of metadata corresponding to the books it torrented from LibGen and PiLiMi, and some metadata for Books3. In the catalogs — as Anthropic’s co-founder pointed out to colleagues when reviewing one — were ISBNs (Br. Exh. 19 at -0389774). An *ISBN* is an Industry Standard Book Number for published books issued by a private organization since 1970. It can be included on copyright registrations. Each ISBN is a string of numbers — having substrings defined to flag each book’s language, region, publisher, and unique title (at the edition level). Anthropic used ISBNs to identify books quickly and to unpack from them “intelligence about editions and languages” (Reply Exh. 45 at -0366000). Similarly, an *ASIN*, or Amazon Standard Identification Number, is issued by the online retailer to identify any item sold there, including ebooks or self-published books. Anthropic’s metadata sometimes used ASIN. A *hash value* is another critical piece of metadata contained in those catalogs. Anthropic used hash values, too.

In hashing, a user starts with a source input: either a raw file, the text extracted from that file, a paragraph of any text, or so on. This source input is passed through \*1048 a *hash function* or algorithm to produce a unique *hash value*, or short alphanumeric signature, which can be used directly or saved for later in a database alongside information known about that input. This can be repeated for many source inputs. Then, a test input is passed through the same algorithm. If its resulting hash matches one of the source hashes, then the test input matches that source input. The algorithm’s *threshold* for a match can be adjusted to recognize perfect matches or to tolerate fuzzy ones. The hash function is sensitive to differences in the content, but also sensitive to differences in formatting or file-type to the extent those are passed through without normalization — all are just ones and zeroes to it. Finally, such algorithms can be complemented by other programming to do things in sequence and at scale, such as proceeding file by file, extracting and normalizing text from each file to divide into many sub-inputs to hash separately, and comparing each file’s resulting set of hashes alone and in combination against those of other files, and so on (Br. Expert Zhao ¶¶ 53–67; e.g., Br. Exh. 3 at 4).

\* \* \*

In our litigation, both sides used methods like those above to identify works by named plaintiffs that Anthropic had downloaded from pirate libraries. Their ISBNs were furthermore used to find certificates of copyright registration, and thereby to identify relevant plaintiffs as at least beneficial owners of the exclusive right to copy books at issue — a process plaintiffs propose can be followed for all class members.

### **3. THE MEANS TO IDENTIFY COPYRIGHT REGISTRATIONS, PRESUMPTIVE RIGHTSHOLDERS, AND TRANSFERS.**

Plaintiffs must further show, in brief, that they either own the copyrights in their works outright, or else exchanged the exclusive right to reproduce copies of their book for royalties in return. Our named plaintiffs are three writers and two so-called “loan-out” companies used for marketing books — one of those loan-outs being fully owned by one writer, the other co-owned by another and his spouse (“Writers” and “Loan-Outs”). This order finds that all except Author Bartz and Author Johnson are at least

beneficial owners of books stolen from the two websites.<sup>2</sup> Only works in the pirate-sourced books are discussed here. Note well that an author starts with the legal title to the entire copyright (or else his employer), and if in exchange for royalties he transfers the legal title to the copyright or to the exclusive right to make copies he nevertheless remains the beneficial owner of that right.

#### ***A. WRITER CHARLES GRAEBER.***

Writer Charles Graeber penned two non-fiction books at issue: *The Good \*1049 Nurse: A True Story of Medicine, Madness, and Murder*, and *The Breakthrough: Immunotherapy and the Race to Cure Cancer*.

Certificates of registration from the United States Copyright Office show the works were registered effective as of 2013 and 2018, the same years first published. (Anthropic's downloading began in 2021.) Each named Writer Graeber as "Author" and "Claimant" (the latter meaning the holder of legal title in the entire copyright at registration), while various agreements were separately recorded (Opp. Exhs. 15–16 (certs.); Br. Expert Zhao ¶ 66 (downloads); MSJ Opp. Exh. 6 at 4 (download dates)).

Respecting *The Good Nurse*, in 2007, a publisher agreed to pay royalties to Writer Graeber in exchange for the "exclusive" "right to reproduce" the work in English including in "print, audio, and electronic" formats (Opp. Exh. 17 at -0207). The parties agreed as to how to prosecute infringement (*id.* at -0216). New York law governed (*id.* at -0219). Respecting *The Breakthrough*, substantially similar terms were agreed (Opp. Exh. 18).

Also respecting *The Good Nurse*, film production companies purchased successive options dated in 2013, 2017, and 2019. Each was a period during which its holder could, if milestones were met along the way to making a film, exercise the option to then obtain certain exclusive rights in exchange for additional compensation and royalties to the author (e.g., Opp. Exh. 22 at -04661, -04666, -04667). Our parties treat the last as having been exercised (*cf.* Opp. Exh. 19 at 155–56; Opp. Exh. 22 at -04623 ¶ 2). Writer Graeber reserved any rights he held "to publish and distribute in all languages throughout the world printed versions" of the book "whether hardcover or softcover" or "electronic or digital" (Opp. Exh. 22 at -04653). California law governed. (A term for royalties is purportedly defined in an attachment not in our record, if ever attached; this order need not parse that issue because the agreement did not purport to transfer the relevant rights.)

Again respecting *The Good Nurse*, Author Graeber's solely owned loan-out, Having & Selling LLC, granted other publishers rights to prepare translations and publish the results (Opp. Exh. 20 (Italian); Opp. Exh. 21 (Czech); Opp. Exh. 19 at 164; Reply Graeber ¶ 3).

Writer Graeber is a named plaintiff; his loan-out is not.

#### ***B. WRITER KIRK WALLACE JOHNSON AND LOAN-OUT MJ + KJ INC.***

Writer Kirk Wallace Johnson put two non-fiction titles at issue: *To Be A Friend Is Fatal: The Fight to Save the Iraqis America Left Behind*, and *The Feather Thief: Beauty, Obsession, and the Natural History Heist of the Century*. Loan-Out MJ + KJ, Inc. is co-owned by him and his spouse.

Copyright certificates show the works were registered effective as of 2013 and 2018, the same years as first published. On the first, Writer Johnson was named author and claimant. On the second, Loan-Out MJ + KJ was named author and claimant, and stating Writer Johnson had worked for hire. Other agreements were recorded (*see* Opp. Exhs. 23, 25 (certs.); Br. Expert Zhao ¶ 66 (downloads); MSJ Opp. Exh. 6 at 4 (download dates)).

Respecting *To Be A Friend Is Fatal*, in 2012, a publisher agreed to pay royalties to Writer Johnson in exchange for “the exclusive right to publish the Work” worldwide including in “electronic text” (Opp. Exh. 30 at -04875–76, -04895–96). They agreed as to how to prosecute infringement (*id.* at -04895). New York law governed (*id.* at -04898).

**\*1050** Respecting *The Feather Thief*, in 2015, another publisher agreed to pay royalties to Loan-Out MJ + KJ in exchange for the “exclusive right to print, publish, and sell the Work” including in electronic copies (Opp. Exh. 31 at -05520–21; *see id.* at -05524–25, -05529). The agreement bound the parties to “transfer and permit the recordation of such copyright ownership [if needed to] permit [one or the other] to bring [an infringement] action in his own or its own name” (*id.* at -05526). New York law again governed (*id.* at -05528).

Respecting *The Feather Thief*, Loan-Out MJ + KJ granted a movie studio an option to buy “all right[s] ... other than the reserved rights,” those reserved rights including for “[p]ublishing” including in “e-read” formats (Opp. Exh. 33 at -088, -090).

In an agreement recorded at the Copyright Office this year, Writer Johnson assigned his copyright interests in *To Be A Friend Is Fatal* to Loan-Out MJ + KJ (*see Opp. Exh. 34; Br. Exh. 30 at -04708*).

Writer Johnson and Loan-Out MJ + KJ are named plaintiffs (the spouse is not).

#### **C. WRITER ANDREA BARTZ AND LOAN-OUT ANDREA BARTZ INC.**

Writer Andrea Bartz penned two novels at issue: *The Lost Night: A Novel*, and *The Herd*. Loan-Out Andrea Bartz, Inc. is fully owned by her.

Copyright certificates show the works were registered effective as of 2019 and 2020, the same years published. Each named Writer Bartz as author and Loan-Out Bartz as claimant, with further agreements recorded (Opp. Exhs. 1, 4).

Respecting *The Lost Night*, in 2017, a publisher agreed to pay royalties to Loan-Out Bartz in exchange for “the right to publish ... any and all editions and/or formats of the Work” — “[e]xclusively” (Opp. Exh. 7 at -05290; *see id.* at -05297 (ebooks)). The parties agreed as to how to prosecute infringement (*id.* at -05294). New York law governed (*id.* at -05303). Respecting *The Herd*, in 2018, Loan-Out Bartz entered a substantially similar agreement with the same publisher (Opp. Exh. 5).

For both books, companies purchased options to buy further rights. If exercised, the portion of one agreement in our record states that Loan-Out Bartz was to receive royalties and reserve “print and audio publication rights” (Opp. Exh. 10 at -023). The options for both were returned unexercised (MSJ Br. Exh. 2, Writer Bartz Dep. Tr. 180–81).

Recently, Writer Bartz agreed to transfer to Loan-Out Bartz any residual rights she held in these works (Opp. Exh. 14). Writer Bartz and Loan-Out Bartz are named plaintiffs.

\* \* \*

In August 2024, Writers brought this putative class action complaining that Anthropic had infringed their federal copyrights by amassing copyrighted books from the internet without paying even for the first copy, and further by not obtaining licenses to reproduce further copies for LLM training (Compl. ¶¶ 1–3, 64, 67; *see Amd. Compl. ¶¶ 1–3, 68, 71*). In December 2024, they amended their complaint to include the Loan-Outs.

#### **4. THIS MOTION AND SUMMARY OF CLASS CERTIFICATION.**

Now, plaintiffs move to certify classes of copyright owners in a “Pirated Books Class” and in a “Scanned Books Class.” For the former, they propose two alternatives: The Books3 Pirated Books Class and the LibGen & PiLiMi Pirated Books Class (Br. 1–2).

Named plaintiffs bear the burden of affirmatively demonstrating their compliance **\*1051** with Rule 23. *Wal-Mart Stores, Inc. v. Dukes*, 564 U.S. 338, 350, 131 S.Ct. 2541, 180 L.Ed.2d 374 (2011). The district court may exercise its discretion to modify a class definition under Rule 23(c). See *Gen. Tel. Co. of Sw. v. Falcon*, 457 U.S. 147, 160, 102 S.Ct. 2364, 72 L.Ed.2d 740 (1982). For the same reason, modifying a subclass — even creating one — is “within the district court’s broad power under [the rules] to adopt procedural innovations to facilitate management of the class action.” *Am. Timber & Trading Co. v. First Nat'l Bank of Oregon*, 690 F.2d 781, 786–87 (9th Cir. 1982). To the extent certification turns on expert methods, those methods are subject to a “limited” *Daubert* review. *Lytle v. Nutramax Lab's, Inc.*, 114 F.4th 1011, 1025, 1031 (9th Cir. 2024). And, because it is “in the best position to consider the most fair and efficient procedure for conducting any given litigation,” the district court finally exercises discretion “in weighing the correct mix of factors” to make its decision. *Bateman v. Am. Multi-Cinema, Inc.*, 623 F.3d 708, 712 (9th Cir. 2010) (quotation omitted) (vacating denial of certification). A decision to certify the class commands “noticeably more deference” than a denial. *Abdullah v. U.S. Sec. Assocs., Inc.*, 731 F.3d 952, 956 (9th Cir. 2013) (quotation omitted).

This order certifies only a LibGen & PiLiMi Pirated Books Class. This class is limited to actual or beneficial owners of timely registered copyrights in ISBN/ASIN-bearing books downloaded by Anthropic from these two pirate libraries. And, it is further limited to actual or beneficial owners of the copyright interest under Section 106(1), *i.e.*, the right to make reproductions. The class is not limited to authors or author-like entities, however. A key point is to cover everyone who owns the specific copyright interest in play, the right to make copies, either as the actual or as the beneficial owner.

A rigorous notice process will be required. Notice will be by first-class mail and email to the author, publisher, and copyright owner listed on the copyright certificate for each work recorded. Notice will also be sent by first-class mail and email to all trade and university publishers in the United States. Notice will also be published at least once in a trade journal. And, as a step in the claims process, notice also will be served by the class claimant himself on all others he attests may claim a copyright interest in the work to notify them that he is claiming it — including all those with whom he has contracted concerning the copyright and/or publishing. If disputes arise over ownership, which will be unlikely, the district court or as needed a jury will resolve them. There will only be one recovery, if any, per single work.

This order denies certification for a Books3 Pirated Books Class. The books downloaded as part of Books3 came with fewer fields of metadata, less routinely complete files of content, and as a result the identification of titles and authors would be too problematic. This order also denies certification for a Scanned Books Class.

Finally, it does not escape notice that some copyrighted works were copied more than once from across the various sources. Recovery of statutory damages under the Copyright Act is limited to the statutory maximum per work in a given infringement action. As a result, a copyright owner whose work was copied from LibGen and from PiLiMi will be allowed but one recovery.

## ANALYSIS

“Under Rule 23, a class action may be maintained if the four prerequisites of Rule 23(a) are met, and the action meets one of the three kinds of actions listed in **\*1052** Rule 23(b).” *Van v. LLR, Inc.*, 61 F.4th 1053, 1062 (9th Cir. 2023).

The following class definition will be approved:

All beneficial or legal copyright owners of the exclusive right to reproduce copies of any book in the versions of LibGen or PiLiMi downloaded by Anthropic. “Book” refers to any work possessing an ISBN

or ASIN which was registered with the United States Copyright Office within five years of the work's publication and which was registered with the United States Copyright Office before being downloaded by Anthropic, or within three months of publication. Excluded are the directors, officers and employees of Anthropic, personnel of federal agencies, and district court personnel.

## **1. CRITERIA UNDER RULE 23(a).**

Under Rule 23(a), the court must find a proposed class meets (1) numerosity, (2) commonality, (3) typicality, and (4) adequacy as to class representatives and counsel. Here, Anthropic does not contest numerosity or typicality, and treats commonality later within predominance. But Anthropic does not concede adequacy of class representatives (Opp. 23).

### ***A. ADEQUACY OF CLASS REPRESENTATIVES.***

#### ***(i) Conflict with Owners of Other Copyrights?***

Anthropic first contends author- and author-affiliated entities do not adequately represent owners of other copyrights because “academics, researchers, and writers who use [LLMs] disagree with the position taken by plaintiffs and actively use and benefit from LLMs like Claude” (Opp. 2). This is unpersuasive. Whoever would like Anthropic to take their books for free can achieve that result even if the remaining class prevails: They can opt out and license their works to Anthropic or to all the world for nothing. *E.g., Effects Assocs., Inc. v. Cohen*, 908 F.2d 555, 559 (9th Cir. 1990) (individual); *see also Matamoros v. Starbucks Corp.*, 699 F.3d 129, 138–39 (1st Cir. 2012) (class).

And, because the class is limited to those whose books were timely registered, all are entitled to receive statutory damages. There is no conflict between one cohort with that choice and another confined to actual damages — and if one class member thinks her compensable loss is uniquely compelling, she can opt out. *See In re Literary Works in Elec. Databases Copyright Litig.*, 654 F.3d 242, 254–55 & n.6 (2d Cir. 2011).

#### ***(ii) Conflict with Co-Owners or Rightful Claimants of the Same Works?***

Anthropic also suggests conflicts may emerge between beneficial owners and legal owners, or among co-owners, with beneficial owners better represented by named plaintiffs. A beneficial owner of the right to make copies is someone like an author who receives royalties from any publisher's revenues or recoveries from the right to make copies. Yes, the legal owner might be the publisher but the author has a definite stake in the royalties, so the author has standing to sue. And, each stands to benefit from the copyright enforcement at the core of our case however they then divide the benefit. Also, given the terms on which certification is granted herein, there is no risk of double jeopardy of defendant because there will be only one recovery per work. *See, e.g., EMI Christian Music Grp., Inc. v. MP3tunes, LLC*, 844 F.3d 79, 94–95 (2d Cir. 2016). Therefore, the class definition will bind both the legal and beneficial owners of the pirated books.

It is theoretically possible that a dispute could arise, as Anthropic posits, as to the true copyright owner(s) entitled to recover. \*1053 For example, an author might claim the entire award only to have the publisher claim separately that it should have received it (and possibly sue Anthropic all over again). In the judgment and experience of the district judge, this is unlikely to occur except in the rarest of instances. This is because authors and publishers are in business together and will work out the best way to recover.

But if a dispute does occur as a part of our action, the following procedures will prevent it from becoming a problem. *First*, publishers and the book world at large will receive a class notice. That circular will put them on notice of the claims and let them know that all owners of the right to make copies will be bound. *Second*, all class claimants in the claims process will be required to serve notice that they are claiming the award to all others associated with the book. For example, an author would be required to serve notice to the publisher that the author has submitted his claim, so that the publisher will have an opportunity to claim a share from him, or else to submit a competing claim in our action, in which case the district judge or a jury will decide ownership. *Third*, in the very unlikely event such ruinous conflicts become overwhelming, the district court can decertify the class.

### ***(iii) Are Named Plaintiffs Legal or Beneficial Owners?***

Anthropic also argues that Writer Bartz and Writer Johnson are inadequate representatives for any class: Neither is a legal owner of any copyrighted work at issue (they have no exclusive rights). And, neither is a beneficial owner (as during this litigation they transferred any and all their interests to their loan-out corporations and not in exchange for royalties) (Opp. 13–15).

This order agrees neither is a legal owner. It is a closer question whether either is a beneficial one. Our court of appeals has called the author who grants exclusive rights in exchange for royalties the “classic example”—but has expressly left the door ajar to others. *DRK Photo v. McGraw-Hill Glob. Educ. Holdings, LLC*, 870 F.3d 978, 988 (9th Cir. 2017). At bottom, beneficial ownership is a “standing doctrine” to enable suit. *BMI v. Hirsch*, 104 F.3d 1163, 1166 (9th Cir. 1997); *Yount v. Acuff Rose-Opryland*, 103 F.3d 830, 834 (9th Cir. 1996). And, the author who transfers her copyright to her loan-out still stands to benefit from the exclusive right by dividends or enhanced valuation. Indeed, the same would be true even if her corporation were itself receiving only royalties. For reasons like these, other district courts have held such authors to be beneficial owners. See *Drake v. Malouf*, No. 99-CV-0315-G, 1999 WL 1007642, at \*4 (N.D. Tex. 1999) (Judge Joe Fish); cf. *Wildlife Internationale, Inc. v. Clements*, 591 F. Supp. 1542, 1546 (S.D. Ohio 1984) (Judge Carl Rubin). But see *Roberts v. Gordy*, 359 F. Supp. 3d 1231, 1249 (S.D. Fla. 2019) (Judge Kathleen Williams) (declining to follow).

But there is no reason to follow those courts, because this problem is readily resolved regardless. If the author controls the loan-out, she can bring the suit in its name, or effect a transfer from one hand to the other hand by “at anytime simply memorializ[ing] that intent in writing.” *Jules Jordan Video, Inc. v. 144942 Canada Inc.*, 617 F.3d 1146, 1156 (9th Cir. 2010). And, for the same reason, defendant can hardly complain of any purported defects as between an author and that author’s loan-out in the first place: Where one of them plainly has standing, and there is no disagreement as between the two, “it would be unusual and unwarranted to permit a third-party infringer to invoke [the written transfers requirement] to \*1054 avoid suit for copyright infringement.” See *id.* at 1157 (cleaned up).

Because Writer Bartz and Writer Johnson have made the choice to transfer all rights to their loan-outs, they are no longer adequate class representatives. But this situation nevertheless shows why ambiguities between authors and loan-outs will not metastasize.

Loan-Out Bartz and Loan-Out MJ + KJ are on our record adequate representatives. Loan-Out Bartz is the registered claimant of the books penned by Writer Bartz. In exchange for royalties, Loan-Out Bartz granted an exclusive license to reproduce copies. Yes, production companies bought options. But they returned them unexercised. Much the same is true as to Loan-Out MJ + KJ. Specifically as to *The Feather Thief*, Writer Johnson prepared the work for hire for Loan-Out MJ + KJ, which then granted an exclusive license to reproduce copies in exchange for royalties. Separate film options did not purport to transfer those rights a second time.

As for our other named plaintiff, Author Graeber was himself the registered claimant and granted an exclusive license to reproduce copies of his books at issue in exchange for royalties. Yes, he transferred some rights to a filmmaker, but the agreement expressly did not transfer any rights to publish his book — consistent with the above. Nor are rights to prepare Czech or Italian translations and publish the results at issue here.

Loan-Out Bartz, Loan-Out MJ + KJ, and Writer Graeber are adequate class representatives — specifically for the LibGen & PiLiMi Pirated Books Class.

#### ***B. ADEQUACY OF CLASS COUNSEL.***

The proposed co-lead law firms of Lieff Cabraser Heimann & Bernstein, LLP and Susman Godfrey LLP are experienced in class actions, including copyright (Br. Attorneys Geman & Nelson ¶¶ 30, 34, 46, 47). They represent related parties in co-pending litigation against other AI companies (*id.* ¶¶ 24–26). They do “not hesitate to go to trial” (*id.* ¶ 53; *see id.* ¶ 30). They will fulfill their duty to their class clients. *See Ellis v. Costco Wholesale Corp.*, 657 F.3d 970, 985 (9th Cir. 2011).

Adequacy is met. And, if the commonality prong is also met — as incorporated into the predominance inquiry below — then all of Rule 23(a) is met.

#### **2. CRITERIA UNDER RULE 23(b)(3).**

Under Rule 23(b)(3), the court must “find[ ] that the questions of law or fact common to class members predominate over any questions affecting only individual members, and that a class action is superior to other available methods for fairly and efficiently adjudicating the controversy.” The rule does not impose a threshold requirement that a class be ascertainable, although similar questions may arise to the extent implicated by other requirements. *Briseno v. ConAgra Foods, Inc.*, 844 F.3d 1121, 1124 nn.3–4 (9th Cir. 2017).

#### ***A. PREDOMINANCE.***

“First, we identify which questions are central to the plaintiffs’ claim. Second, we determine which of these questions are common to the class and which present individualized issues. Third, we analyze whether the common questions predominate ....” *DZ Rsv. v. Meta Platforms, Inc.*, 96 F.4th 1223, 1233 (9th Cir. 2024).

This is the classic case for certification. Anthropic downloaded millions of book files. This can and should be proven on a classwide basis. A list of works stolen in this downloading can be constructed using **\*1055** the methods set forth in the statement above. Instead of millions of separate lawsuits with millions of juries, we will have a single proceeding before a single jury, *Napster* style.

##### ***(i) Copying Original Elements?***

###### ***(a) LibGen?***

There is no question Anthropic torrented five million files from LibGen. Instead, Anthropic contends that infringement of plaintiffs’ books cannot be shown with classwide evidence because without individualized investigation no one can be sure what it even copied, whether they were the books they said they were, or someone else’s books, or something else like empty files. Plaintiffs propose relying on the catalog of bibliographic metadata listing the five million books — a catalog that Anthropic separately torrented and adopted to account for what it copied — with titles to be confirmed as needed by common hashing methods and occasional manual review.

Plaintiffs point specifically to a LibGen-provided book catalog. As Anthropic’s co-founder waited for the LibGen-sourced books to torrent, he separately downloaded and started exploring the LibGen-sourced book catalog purporting to describe those books

(*see* Br. Exh. 19). Each row represented a book and each column a different type of metadata respecting it (¶ 83). The co-founder pointed out to his team two types of metadata in particular: “Identifier,” and “MD5.” As he restated them, “Identifier is the ISBN,” and “MD5 corresponds to [each file’s] hash” (Br. Exh. 19 at -0389774). Anthropic’s expert later further explained what the MD5 hash means: “An md5 hash is ... used in the LibGen database to uniquely identify books[, and] is generated by passing the raw binary contents of a file through MD5, a cryptographic hash function” (¶ 51 n.23 (emphasis omitted)). If the same book file were passed through the same MD5 hash function, the very same hash value would result as that listed in the LibGen book catalog (*see* ¶¶ 51, 83). Other metadata on the same row as each book’s ISBN and MD5 hash were the book’s “Title,” “Author,” “Edition,” “Publisher,” and “Year,” and so on (Br. Exh. 19 at -0389774). This evidence could be offered to a jury to support the contention that Anthropic adopted and relied upon this list of the books it copied, and thereby admitted the list of books it copied.

Even Anthropic’s own expert stated in his declaration that a list of errors found in LibGen over the span of several years has lengthened by about one post for every 400 books — suggesting an error rate of less than one percent (*compare* ¶ 49, *with* ¶ 45). Such a per work list could come in via an expert at trial. Defendant would be free to prove up the exceptions and errors.

It remains, however, plaintiffs’ burden to produce such a per work list in the form of a CD with a table having columns for title, author, publisher, copyright registration number, copyright-registered author (frequently the author), copyright-registered claimant (frequently the author or author’s loan-out), and ISBN and/or ASIN.<sup>3</sup> Class notice will be mailed to those listed (once is enough per recipient). The class will be limited to the books on the list for which there is complete information — meaning actual and beneficial owners of an exclusive copyright to reproduce any book on the list. The list will also be offered as evidence at trial. Plaintiffs will be allowed reasonable discovery and time to prepare **\*1056** the list. They must omit from the list any book for which the pirate site provided only an empty file. They must correct the title on any book copied from the pirate site for which the pirate catalog provided the wrong title. Although this order certifies a class, it does so on the condition that plaintiffs prepare and file a comprehensive, per work list and do so by **NOON ON SEPTEMBER 1, 2025**.

Anthropic makes various objections. None persuade. What counts is not whether plaintiffs will prevail at trial, but whether they will be able to mount common evidence and methods in support of their claims. Anthropic raises challenges to these forms of proof that are themselves best addressed through common evidence and methods.

*First*, Anthropic submits the sworn declaration of one employee who swears that “Anthropic did not create this metadata” and that “Anthropic has never relied on this metadata to identify specific books in its LibGen Dataset” (Opp. Employee Tucker ¶ 16). The jury will hear that testimony and its cross-examination as common evidence as to every class member. The challenge is itself what could be raised in any one class member’s case, *e.g.*, *Range Rd. Music, Inc. v. E. Coast Foods, Inc.*, 668 F.3d 1148, 1154 (9th Cir. 2012), and here it can be done common to all at once.

*Second*, Anthropic argues that sometimes its infringement was *de minimis* because, whatever its expectations, it in fact obtained partial books due to errors in torrenting. One might suppose this criticism would be readily substantiated and then at our trial prompt manual work-to-work comparison. Not so — or not on this record. After using computer-aided methods to rapidly identify thousands of purportedly partial books downloaded as part of Books3 (¶ 87–88), and after trawling through 13,733 complaints collected over several years by internet users of the LibGen library (¶ 49), Anthropic’s expert still failed to provide a single example of a purportedly partial book in the LibGen-sourced books acquired by Anthropic. Moreover, stealing a page of a copyrighted work is still a violation.

Relatedly, Anthropic submits that sometimes the pirate librarians’ extractions or its own torrenting got botched so it got no book at all (¶¶ 59–60). He points to ten examples in five million files that omitted any book (¶ 89). Then he elsewhere explains that whether a book available for torrenting was in fact torrented exactly as expected can be easily checked by comparing the canonical MD5 hash values available in the LibGen book catalog against the computed MD5 hash values of the files as in fact torrented by Anthropic (*see* ¶ 51 & n.23; *cf.* ¶ 72). This is a solvable problem that will not overwhelm the trial.

Finally, Anthropic argues that some metadata is mismatched, such that instead of pirating one book wholesale, it pirated another. To be clear, for each changeling so far discovered, Anthropic's expert readily found its spitting image in an "Actual Content Match" (¶¶ 51–52). So the problem is not whether our jury will be subjected to detailed infringement analysis. These books, too, were *in toto* infringed, but apparently switched. Instead, the problem is whether rare changelings can be identified using common methods so that, among other things, the right copyright owner(s) can be compensated. Neither expert sized up how common this problem might be. When asked to estimate if all errors in the LibGen metadata amounted to more than one percent, Anthropic's expert demurred. This failure to substantiate the rate of otherwise speculative errors — particularly for the party \*1057 possessing all the information — cuts against defendant and not the putative class at this point. *Lytle*, 114 F.4th at 1030. Regardless, this issue, too, cries out for common methods to solve, and sooner than trial. There are several.

For one thing, there are the corroborative sources of common metadata. Recall that Anthropic engineers chose *not* to torrent from the PiLiMi library those books already torrented from the LibGen library, and yet that the metadata for the books *not* torrented remains in the PiLiMi book catalog. And, recall that Anthropic's chief executive officer and its head of training talked about purchasing books to *replace* the LibGen-sourced books, and populated a database listing books to purchase. The first created a list of cross-checked metadata, the second a fresh list of metadata that may be coextensive. The three sets together may catch the changelings.

For another thing, there is the common hashing method that plaintiffs' Expert Ben Zhao proposed to spot and solve problems with the metadata — and to stand as its own, common method for identifying and indeed proving books were duplicated in their entirety or at least in generous verbatim chunks. *E.g.*, *Harper & Row, Publishers, Inc. v. Nation Enters.*, 471 U.S. 539, 548–49, 105 S.Ct. 2218, 85 L.Ed.2d 588 (1985) ("300 [to] 400" words of a fact-intensive book); *Thomson Reuters Enter. Centre GmbH v. Ross Intell. Inc.*, 765 F. Supp. 3d 382, 391, 395–96 (D. Del. 2025) (Judge Stephanos Bibas) (paragraphs of headnotes), *appeal docketed*, No. 25-8018 (3d Cir. Apr. 14, 2025). The specific implementation would compute multiple hashes across a class claimant's work and compare the hashes both individually and in bunches against those from Anthropic's LibGen book files. Expert Zhao identified named plaintiffs' own books in LibGen this way.<sup>4</sup>

Anthropic's Expert Mohit Iyyer identified no fundamental flaw in plaintiffs' Expert Zhao's specific implementation of the decades'-old class of hashing methods. And, Expert Iyyer identified no instance where the hashing method as implemented in fact identified a book that Anthropic lacked (false positive). Instead, the expert identified several adjustments all agree could be made during implementation to further improve results (for instance, improving the way text is normalized before being passed through the hash function). The expert's other main point was that it was not clear how high the threshold should be set to identify a match, whether exact or fuzzy. That is a matter of debate the experts can have before the jury. And, where the jury cannot conclude whether there was virtual duplication via the method, they can do so via direct comparison of the copies, with source copies for this comparison to be obtained by class counsel where required for this purpose. It is this order's "predictive judgment" that this method will "bear fruit." *Lytle*, 114 F.4th at 1031.

#### *(b) PiLiMi?*

There also is no question Anthropic downloaded via torrenting two million \*1058 files from PiLiMi. Again, Anthropic contends that infringement of plaintiffs' books cannot be shown with classwide evidence or common methods. Plaintiffs propose relying on the catalog of bibliographic metadata unique to PiLiMi, which again Anthropic torrented, as well as on the common hashing method and occasional manual review.

The above points apply equally well here, especially given the intertwined origins of the LibGen and PiLiMi pirate libraries. When cross-checking the works Anthropic did not want to torrent a second time, it did so against PiLiMi's bibliographic catalog, not yet possessing PiLiMi's book files. It then used metadata from that catalog to make further dissections of what it did download. Anthropic's Expert Iyyer has explained that the PiLiMi book catalog is substantially similar to the LibGen book catalog — except that it still contains the five million metadata entries for the books it decided not to copy. When asked to

estimate whether the errors in this metadata amounted to more than one percent, Expert Iyyer declined. Regardless, all the challenges raised above, if raised again here, can be solved in similar fashion.

\* \* \*

Because the quality of common evidence and common methods for identifying these two intertwined sources is similar — indeed, sometimes the same metadata and always the same methods will be common to both — there is no reason why those whose works were copyrighted in one, the other, or both cannot be efficiently considered in one trial. Furthermore, an analysis of the other issues below would bear the same result, including for damages, where the maximum statutory amount will be per work rather than per copy in this one infringement action.

### *(c) Books3?*

Again, there is no question Anthropic downloaded 196,640 records in the Books3 collection. Again, Anthropic makes the same arguments. This time they are better received.

When Anthropic's co-founder downloaded Books3, he did not download and explore a separate book catalog. The only metadata available for Books3 was the metadata intrinsic to the books' filenames. At their best, the filenames strung together two or three chunks of metadata, like "Title – Author" (see ¶¶ 22–23). There is no evidence Anthropic relied on this metadata in the same way it relied on the other sets of metadata — no doubt because this metadata was and is less comprehensive and noticeably less accurate.

True, some short filenames permitted identification. For example, to verify the accuracy of the first book in the set Anthropic downloaded, with filename "Lindbergh – A. Scott.Berg.epub.txt" (¶ 23), Anthropic's expert "manually compared the opening paragraphs of text [for that book] with the opening paragraphs of text in the Google Books preview of *Lindbergh* by A. Scott Berg and found that the texts appeared to be from the same book" (¶ 25; see ¶ 82). But, as Anthropic uncovered in its business, a title-author combination (absent ISBN) can point to multiple editions of the same book, each a different copyrighted work — such as multiple copyrighted versions of abridged or annotated classics. And, if there is only a title present, or only an author, the problem compounds.

Not all Books3's filenames contained even that much. Many were private mnemonics ("math"), not ready-to-parse metadata ("Mathematics, Substance and Surmise: Views on the Meaning and Ontology of Mathematics" by Ernest Davis and Philip J. Davis) (¶ 28). And, there were mismatched **\*1059** filenames. Anthropic's expert did investigate ten such books and "[i]nferred correct title and author" (¶ 28). But plaintiffs' expert proposed no convincing way to accelerate this process, partly because of the further problems described next.

The book content was also spotty. For 1,284 of 196,640 files, there was no content present — resulting in no book in need of identification and thus solving itself (¶¶ 37–38). But similarly obvious errors (for 442 files) or other issues (for at least 3,312 files) resulted in partial content for another 3,750 files (see ¶¶ 40, 87), these files having the sparse metadata particular to this set.

In all these examples, the point is not merely the portion or number of issues, but the combination of apparently less reliable metadata and less reliable content. Plaintiffs have not made clear how they can solve the chicken-or-egg problem in this instance. Yes, there may yet be ways to speed up identification. Our record suggests Anthropic may have found one (see Reply Exh. 47 at -0389524). But plaintiffs have had plenty of time to do so, too — whether through expert methods or discovery of common evidence. They still have not. This delay prejudices the owners of the other copyrighted works in LibGen and PiLiMi, who already have comprehensive metadata, more complete book content, and a clear path to prosecute.

Because this problem is so fundamental it moots the remaining inquiry, the certification of the Books3 Pirated Books Class is **DENIED**.

***(d) Scanned Books?***

The summary judgment order on fair use revealed that the path to recovery for scanning purchased print books and storing their digital replacements peters out. Same goes for further copying used for training LLMs. Although there is evidence of other copying for other uses, this evidence has not been presented with the requisite specificity and uniformity to warrant class certification of a scanned books class specifically — and meanwhile the same points about prejudicing all the other claimants while we wait apply here. Its certification is **DENIED**.

This order will focus on the LibGen & PiLiMi Pirated Books Class from here forward.

***(ii) Of a Valid, Copyrighted Work?***

Next, plaintiffs must prove that what was duplicated was in fact a valid copyrighted work. This matching process can again be shown in the main through common evidence and methods.

The class is limited to books for which an ISBN or ASIN exists. From the metadata described above, as well as through commercial metadata associated with ISBNs or ASINs, copyright registrations associated with these works can be identified. For every certificate of registration presented by named plaintiffs, for example, there is an associated ISBN (e.g., Br. Exh. 30). And, because the classes are limited to works registered within five years of first publication, these certificates and the facts therein are presumed valid. [17 U.S.C. § 410\(c\)](#).

Anthropic speculates that some registrations might be hard to find or invalid (Opp. 11–12). It fails to substantiate either concern.

Yes, some works share even title and author — as do multiple editions. But Anthropic's own records indicate ISBNs and ASINs are typically unique to each book at the edition level, and so reliably can be linked to a copyright. And, not only do the likely choices of pirate librarians limit the scope of hard-to-find or out-of-print works copied in the first place, but the limitation to works by ISBN and ASIN — the one **\*1060** codified in 1970, the other in the 1990s — further limits the likelihood we will encounter copyright-registered works for which their corresponding registration, if any, cannot quickly be found. (Some self-published authors might use an identifier like ASIN but choose not to register at the Copyright Office at all.)

And, yes, although the copyright registrations and facts therein are presumptively valid, the Copyright Office does not verify every one. Anthropic is entitled to rebut the validity (Opp. 11 n.5). But any “inadvertent mistakes on registration certificates [would] not invalidate a copyright and thus [would] not bar infringement actions, unless the alleged infringer ha[d] relied to its detriment on the mistake.” *Jules Jordan Video*, 617 F.3d at 1156 (cleaned up). Nothing in our record hints at Anthropic having relied on anything in a copyright certificate. So, if an author's loan-out were named instead of the author, or vice-versa, such a mistake would not stall our litigation.

***(iii) Causing Damages?***

Plaintiffs next must prove damages. They propose theories for proving statutory damages and actual damages. This, too, can be shown through common evidence and methods.

The LibGen & PiLiMi Pirated Books Class is limited to works that were registered prior to infringement or within three months of first publication. (The date of effective registration and publication are stated on the certificates of registration and, for reasons above, presumed true.) Because class members' works will meet this scope (by class definition), they will be entitled to elect statutory damages rather than actual damages. [17 U.S.C. §§ 412, 504\(c\)\(1\)](#). To the extent a class member believes her actual damages will surpass the statutory ones, she can intervene to prove it or opt out.

In order to determine a statutory award, a jury must first make findings as to the infringer's mental state (e.g., willfulness), then consider factors including actual damages and deterrence to set an award within a range of values allowed for that mental state — bottoming out at \$200 per work for the reasonable user who believed his use fair. *See id. § 504(c)(2); Desire, LLC v. Manna Textiles, Inc.*, 986 F.3d 1253, 1271–72 (9th Cir. 2021).

Anthropic has introduced no evidence from which it might be supposed that its mental state differed as to its torrenting of any of these books. To the contrary, it has actively submitted that piracy is for an AI company the fair price to pay. Thus, common evidence can be taken as to the mental state involved in the copying that is the crux of plaintiffs' contentions for these classes. Questions of deterrence would also be general in nature. And, in the alternative, Anthropic has suggested that *it already made recompense* by paying the full price for each work when it later purchased print copies of the same titles. Assuming for the moment that this argument is admissible, it is hard to see how this would not also be evidence in common across the broad category of those for whom it is true.

As for any embedded assessment of actual damages, common evidence and methods again suffice. Anthropic proffered in its motion for summary judgment on fair use that there was no world in which it would have been able to negotiate individual deals. Indeed, our entire record reflects as much — not only for Anthropic, but for every other AI company to have purchased books. Books were in every case purchased according to either a blanket price, a tiered price based on subject matter or the like, or else according to a mass-market price. Debating which of these options is **\*1061** appropriate, and the resulting price to be applied to each book here, is susceptible to common evidence and methods. Indeed, for many if not all the works, one piece of common evidence already exists as to what Anthropic would have paid: the price it later paid to purchase any of the same books (assuming admissible) (*see* Br. Exhs. 12, 22).

The summary judgment order makes clear that Anthropic is liable for all pirated copies regardless of whether they were used for training and regardless of whether another copy of the book was bought and scanned. To the extent a jury thinks these differences should affect the amount of statutory damages the points can be made on a categorical basis rather than an individual one.

#### ***(iv) Plaintiffs Entitled to Recover?***

To recover, a claimant will have to submit a claim under oath stating that it is the owner of the copyright interest infringed and stating the title, author, publisher and ISBN and/or ASIN for the claimed work and further stating that the claimant serve notice to the publisher or any other person with an interest to permit the person to contest the claim. In the district judge's experience and judgment, very few disputes over ownership will arise but if they do they can be resolved by the district judge on a summary judgment or by a jury at trial (or by excluding the work altogether from the class).

Anthropic asserts that until ownership is proven all putative class members lack standing, so its due process rights will be violated to the extent it is made to answer claims that a copyright was infringed. But named plaintiffs plainly have standing and the class will be defined in a way to bind all actual and beneficial owners of every work it stole, so this objection falls away.

Anthropic also argues that managing ownership inquiries at scale will necessarily entail "Trial by Formula" (Supp. Definitions Opp. 2 (citing [Wal-Mart Stores](#), 564 U.S. at 367, 131 S.Ct. 2541)). A "Trial by Formula" classically involves sampling what happened to some class members, assuming the same happened to all, and not allowing rebuttal. That is not the plan. Plaintiffs shall present, among other things, a list of books with their authors, publishers, and ISBN/ASIN identifiers that Anthropic

downloaded. This list will be prepared (or should be) with sufficient care to satisfy *Daubert*, no doubt relying in part on defendant's own records. The class definition is limited to actual and beneficial owners of the copyrights involved. In the district judge's judgment from experience (of fifty years), to repeat, there will be very few competing claimants for the same work. This is because authors and their publishers have ongoing business relationships and they will work out whatever differences (if any) they have over how to divide the recovery.

The core claim of the class here is infringement of the right to reproduce copies under Section 106(1) by way of episodes of mass downloading of timely registered copyrighted books. All the claimants can be made out through sworn declarations subject to challenge, in some cases.

#### **(v) Fair Use?**

To the extent fair use were to recur as an issue at trial, there is no reason to believe it would not be susceptible to treatment by common evidence and common methods.

\* \* \*

At bottom, the major elements of this case concern a common course of conduct that can be established with common evidence and methods. The ownership inquiry is where Anthropic puts its focus — but its objections have been in the \*1062 main addressed by plaintiffs' concession that they seek to enforce rights under Section 106(1) and that the focus of the LibGen & PiLiMi Pirated Books Class is on classic reproductive copying by piracy. "We do not permit a defendant to support its invocation of individualized issues with mere speculation." *Van*, 61 F.4th at 1068. And, Anthropic's option to rebut otherwise completely sufficient evidence by itself "does not preclude the predominance of common questions." Cf. *Blackie v. Barrack*, 524 F.2d 891, 906 n.22 (1975) (early fraud-on-the-market case where defendant enjoyed right to rebut presumption but was unlikely to exercise it or to succeed in doing so). "What a district court may not do is to assume, *arguendo*, that problems will arise, and decline to certify the class on the basis of a mere potentiality that may or may not be realized." *United Steel Workers v. Conoco-Phillips Co.*, 593 F.3d 802, 810 (9th Cir. 2010).

This is how Judge Marilyn Patel, one of the great jurists of this district court, assessed predominance in the *Napster* case, which is very close to our case:

[Although] ownership, registration, and actual damages ultimately require[ ] a work-by-work inquiry, viewing these determinations as purely "individual issues" ignores the fact that the claims of every member of the class are uniformly premised upon the uploading or downloading of a copyrighted work by Napster users. This shared factual predicate in turn gives rise to a host of common legal issues concerning Bertelsmann's involvement in the operation of the Napster network. There can be no serious dispute that these issues are sufficiently "significant" to warrant adjudication of the parties' dispute on a representative rather than individual basis, nor is there any question that considerations of judicial economy heavily favor litigating these common issues once, as part of a single class action, rather than rehashing the same questions of law and fact in each of what could likely amount to thousands of individual lawsuits.

*In re Napster, Inc. Copyright Litig.*, No. C MDL-00-1369 MHP, 2005 WL 1287611, at \*7 (N.D. Cal. June 1, 2005) (citations omitted); see also *Tyson Foods, Inc. v. Bouaphakeo*, 577 U.S. 442, 453, 136 S.Ct. 1036, 194 L.Ed.2d 124 (2016).

Predominance does not require that every issue be provable by a single method of proof. It does not rule out that one group may involve extra issues. It does require that the broad and differing nuances among the class be treatable by categorical proof as

to each nuance. If the nuances are so varied and so many as to be unmanageable by a jury, then predominance is lacking. But if the differences are manageable, then predominance is satisfied. So here.

Common issues and evidence heavily *predominate*.

### ***B. SUPERIORITY.***

This inquiry “at [its] very core” assesses whether the claims could not be resolved but for a class action. *Amchem Prods, Inc. v. Windsor*, 521 U.S. 591, 617, 117 S.Ct. 2231, 138 L.Ed.2d 689 (1997). Superiority is met where there is insufficient “incentive for any individual to bring a solo action prosecuting his or her rights” — and therefore insufficient opportunity for any defendant to conclusively vindicate its rights in turn — while by contrast a class action would “aggregat[e] the relatively paltry potential recoveries into something worth someone’s (usually an attorney’s) labor.” *Ibid.* (quoting *Mace v. Van Ru Credit Corp.*, 109 F.3d 338, 344 (7th Cir. 1997)). We answer this question by considering the interests:

(1) of the judicial system, (2) of the potential class members, (3) of the \*1063 [named] plaintiff[s], (4) of the attorneys for the litigants, (5) of the public at large and (6) of the defendant. The listing is not necessarily in order of importance of the respective interests. Superiority must also be looked at from the point of view [7] of the issues [and legal regime itself].

*Bateman*, 623 F.3d at 713 (quotation omitted); *see also* Fed. R. Civ. P. 23(b)(3).

Here, if not brought as a class action, there will likely be no action at all. And, in any case, only a class action that concentrates plaintiffs’ interests and defendant’s resources into one confrontation will present the claims and defenses with a quality commensurate with the issues’ import. Named plaintiffs already have undertaken extensive discovery and litigation, and have expressed their preference for a class action. For good reason: Defendant is formidable. Few if any potential class members alone

could muster comparable resources [to those of defendant], nor is there any guarantee [any alone] could find counsel willing to work pro bono or on a contingency basis. Pooling these burdens among a group of [plaintiffs] helps more [ ] bring claims. Proceeding as a class action would also maximize efficiency, particularly in light of the considerable discovery the parties have already undertaken.

*See Doe v. Mindgeek USA Inc.*, 702 F. Supp. 3d 937, 952 (C.D. Cal. 2023) (Judge Cormac Carney). Meanwhile, defendant now sits under a cloud of copyright liability — needlessly, assuming it is correct on the facts and on the law — and cannot likely dispel that cloud except by class action. Other stakeholders and the public at large stand to benefit from the clarity only a robustly litigated class action could provide — from authors who wish to know how copies of their works are being used in fact, to other AI companies that would prefer to pirate rather than pay for those works for whatever uses they might have, to people who might not engage in one activity or the other without more clarity.

In opposition, Anthropic offers decisions finding a class action mechanism inappropriate where, for instance, harm was caused “in different individuals in different states by different doctors.” *Zinser v. Accufix Rsch. Inst., Inc.*, 253 F.3d 1180, 1192 (9th Cir.) (quoting favorably *Haley v. Medtronic, Inc.*, 169 F.R.D. 643, 654 (C.D. Cal. 1996) (Judge William Rea)), *amd'd*, 273 F.3d 1266 (9th Cir. 2001). Here, different copyright owners had their copyrighted works downloaded from one of three pirate libraries by one company and kept forever.

Anthropic next argues that “[t]he prospect of ruinous statutory damages — \$150,000 times [five] million books — only compounds the significant due process implications of certification here” (Opp. 23). But our court of appeals has squarely held that it is improper to consider whether or not “defendant’s potential liability [to the class was] ‘enormous and completely out of proportion to any harm suffered by [the named] Plaintiff.’ ” *Bateman*, 623 F.3d at 712–13 (quoting district court). “If the size of a defendant’s potential liability alone was a sufficient reason to deny class certification, [in other words], the very purpose of Rule 23(b)(3) — ‘to allow integration of numerous small individual claims into a single powerful unit’ — would be substantially undermined.” *Id.* at 722 (quoting *Blackie*, 524 F.2d at 899, and further citing a long line of cases in accord). Here, there is a “legislative decision to authorize awards as high as [\$150,000 per work], combined with multiple violations of the statute.” *Ibid.* (cleaned up). That decision is enforceable, and enforceable most efficiently here by class action.

\* \* \*

**\*1064** There is no serious prospect that these claims can be addressed through individual actions. A denial of a motion to certify the class would amount to a concession that copyright owners’ credible allegations of infringement will go unchecked by courts so long as a copyist allegedly violates the Copyright Act not a little but a whole lot. “[T]o do so would eliminate the class action deterrent for those who engage in complicated and imaginative rather than straightforward schemes” — or, in this instance, straightforward piracy but at massive scale. *Cf. Blackie*, 524 F.2d at 903 n.19 (re Rule 10b-5). It cannot be true that as the scope of copying grows, so does the impunity.

Proceeding as a class is superior. The LibGen & PiLiMi Pirated Books Class thus meets [Rule 23\(b\)\(3\)](#). This order need not reach the alternative basis for certifying the class, under [Rule 23\(b\)\(2\)](#) (nor consider certifying issues under [Rule 23\(c\)\(4\)](#)). The class also meets [Rule 23\(a\)](#), so may be certified.

### 3. NOTICE TO POTENTIAL CLASS MEMBERS.

Under [Rule 23\(c\)\(2\)\(B\)](#), a court must direct to class members the best notice that is practicable under the circumstances. That proposed class notice shall provide all the usual language *and the following*:

The notice shall state that any class member shall be bound by the outcome and judgment and shall not (unless it timely opts out) be allowed to relitigate any claim certified for class treatment against Anthropic. Plaintiffs’ counsel shall provide notice to all authors, publishers, copyright-registered authors (if different), and copyright-registered claimants (if different) whose works were downloaded, and to all major, regional, and university presses. Notice shall be by first-class mail, by email, and by publication in at least one national trade journal, all costs of notice to be borne by plaintiffs (but recoverable as costs upon judgment). To assist, by **NOON ON AUGUST 1, 2025**, defendant shall provide the titles, authors, publishers, ISBNs and/or ASINs data to plaintiffs’ counsel to the extent such information about the two pirated sets is in its possession or that of its attorneys. This is a court-ordered interrogatory and must be honored.

The class notice shall further state that upon any judgment favoring the class, only those who submit satisfactory claims shall recover. A satisfactory claim form shall identify the class claimant and the works claimed. It shall state under oath that the claimant owns the copyright to make copies of the book or receives royalties in exchange for having granted to another the exclusive right to make copies of the book. And, it shall further state under oath that the claimant has served a copy of its claim form to anyone with whom it has entered an agreement respecting the copyright and to any other publisher, author, or person he suspects has any copyright interest of any type in the work (so that such persons can submit competing claims if they really think they own the right to make copies). No payment shall be made until **70 DAYS** have passed after mailing of such notice to others. If a competing claim is in fact made, then the district judge or jury if needed will decide ownership.

This will be a claims-made process and the benefit to the class will be determined, in part, by the number of works infringed for which a rightful owner is found, a factor which may influence the amount of any attorney’s fees award.

By **NOON ON AUGUST 15, 2025**, both sides shall submit a single agreed-on proposed form of notice for the judge's review as well as a plan for distribution of notice. Remember, notice will not be sent until the list of works involved (complete with title, \***1065** author, publisher, copyright-registered author (if different), copyright-registered claimant (if different), copyright-registration number, and ISBN and/or ASIN) is submitted by plaintiffs' counsel.

## CONCLUSION

The motion for class certification is **GRANTED** as to one of the alternative pirated books classes, and otherwise denied, as follows:

**1.** The following LibGen & PiLiMi Pirated Books Class is **CERTIFIED**:

All beneficial or legal copyright owners of the exclusive right to reproduce copies of any book in the versions of LibGen or PiLiMi downloaded by Anthropic. "Book" refers to any work possessing an ISBN or ASIN which was registered with the United States Copyright Office within five years of the work's publication and which was registered with the United States Copyright Office before being downloaded by Anthropic, or within three months of publication. Excluded are the directors, officers and employees of Anthropic, personnel of federal agencies, and district court personnel.

Because this order grants certification to this class under [Rule 23\(b\)\(3\)](#), it declines to consider alternative bases for certification proposed by plaintiffs under [Rule 23\(b\)\(2\)](#) or [Rule 23\(c\)\(4\)](#). Defendant is compelled to produce related information described above by noon on August 1, 2025. This certification is contingent upon plaintiffs' satisfactory filing of the list described above by noon on September 1, 2025.

**2.** Certification of a Book3 Pirated Books Class is **DENIED**, as is certification of a Scanned Books Class.

**3.** As Representative Plaintiffs for the LibGen & PiLiMI Pirated Books Class, the following person and entities are **DESIGNATED**: Andrea Bartz, Inc., Charles Graeber, and MJ + KJ Inc.

**4.** As LibGen & PiLiMi Pirated Books Class Counsel, the following law firms are **APPOINTED**: Lieff Cabraser Heimann & Bernstein, LLP and Susman Godfrey LLP.

**5.** Counsel from both sides shall submit a joint proposal for the form of class notice and for the dissemination of class notice to conform with the standards set forth above, by noon on August 15, 2025.

**IT IS SO ORDERED.**

**All Citations**

791 F.Supp.3d 1038

---

## Footnotes

<sup>1</sup> All citations to paragraphs in this order are from the declaration of Anthropic's Expert Mohit Iyyer, unless noted.

<sup>2</sup> From Books3, Anthropic downloaded works by our writers (with at least beneficial owners):

- Andrea Bartz's *The Lost Night* (Andrea Bartz, Inc.);
- Charles Graeber's *The Good Nurse* and *The Breakthrough* (Charles Graeber); and,
- Kirk Wallace Johnson's *To Be A Friend Is Fatal* and *The Feather Thief* (MJ + KJ, Inc.).

From LibGen, Anthropic torrented:

- Andrea Bartz's *The Lost Night* and *The Herd* (Andrea Bartz, Inc.);
- Charles Graeber's *The Good Nurse* (two) and *The Breakthrough* (Charles Graeber); and,
- Kirk Wallace Johnson's *The Feather Thief* (MJ + KJ, Inc.).

From PiLiMi, Anthropic torrented:

- Andrea Bartz's *The Lost Night* and *The Herd* (Andrea Bartz, Inc.);
- Charles Graeber's *The Breakthrough* (Charles Graeber); and,
- Kirk Wallace Johnson's *The Feather Thief* (two copies) (MJ + KJ, Inc.).

Other books were identified in scanned but not pirated books (Br. Expert Zhao ¶¶ 66, 68).

- 3 The table may also include anything else plaintiffs deem important, including the hash values or set of hash values computed for the work according to the plaintiffs' methods, which must withstand *Daubert* and defendant's exacting scrutiny.
- 4 Expert Zhao outlined an algorithm starting with a known book's full text, splitting its sentences, dropping any sentences under ten words in length to avoid matching on common phrases, and hashing those longer sentences — then performing the same on an unknown book. The resulting long-sentence hashes from the two books are compared. If not a perfect match, additional programming compares in essence whether the known-book hashes that do appear among the unknown-book hashes do so in a bunch or else equally spread out across the length of the unknown book — helping to parse whether the unknown book might simply be a different book discussing and quoting the first. If any doubt remains, a human can review and confirm (see Br. Expert Zhao ¶¶ 53–67, Exh. D).