

Which books were in the collections they took? *Answer it work by work.*

Catalog-level evidence for AI copyright litigation, built only from public metadata.

THE SHIFT

The fight has moved from “did they copy?” to “which works, whose, and when?”

Which works

Defendants contest each title: is it in the collection they took, and which file?

Whose

Registrations name publishers, loan-outs and pen names. Standing is argued work by work.

When

Registration timing (§ 412) and limitations both turn on dates you must document per work.

WHAT THE RECORD ALREADY SAYS

Courts and defendants have supplied the collection-level facts

Statement	Source	Kind
Anthropic’s co-founder “used the infamous BitTorrent protocol” to copy LibGen (June 2021)	<i>Bartz v. Anthropic</i> , N.D. Cal., ECF 244 at 2	Court order
Anthropic torrented only the PiLiMi copies “that [we]re not in Lib[G]en”	<i>Bartz</i> , ECF 244 at 3	Court order
Every Works List book was downloaded by Aug. 10, 2022	<i>Bartz</i> , ECF 418 at 49	Stipulation
Contributory liability requires intent, shown by inducement or a service tailored to infringement	<i>Cox Communications v. Sony Music</i> (U.S. 2026)	Supreme Court
Uploading and downloading are infringing acts (not disputed by Meta at the pleading stage)	<i>Strike 3 Holdings v. Meta</i> , N.D. Cal., ECF 61 at 7	Court order

What remains to prove is work-level: that a given owner’s book was in the specific files and torrents taken.

WHAT PAYAUTHORS PRODUCES

For each work: the registration, the catalog entry, the file, the torrent

1

Registration

Copyright Office record, claimant, dates, recorded transfers

2

Catalog entry

LibGen catalog number, date added, and which dated catalogs list it

3

Exact file

MD5 checksum and format, the fingerprint of one specific file

4

Torrent

The named torrent that carried it, and where the file sits in its pieces

Plus a registration-date screen (§ 412): was the work registered before each defendant's documented or alleged download date? And a screen against the Bartz Works List criteria.

THE AUTHOR REPORT

One search, a PDF and CSV counsel can work from

Each report links every match to the Copyright Office record so a paralegal can confirm it. Matching is automated and not perfect: in a software-assisted check of 100 random matches against the Office record, 91% were the same work and 72% the exact edition (about ± 6 points). Expect some missed records and some matches that are not the client’s.

Summary by work and by defendant

Catalog lines with ID, MD5, format and torrent

Torrent piece data for each file

Registration details and the registration-date screen (§ 412)

Sources and SHA-256 checksums

SCALE

7.78M

LibGen files indexed by catalog number and MD5 across 25 dated catalogs, 2015 to 2025

7.05M

of those files located inside their torrents, with piece data

578,874

registrations in our Bartz Works List screen (a work can have several). The official list, under seal, has 482,460 works

Sources: LibGen catalogs from 2015 to 2025, PiLiMi releases 1 and 2, the Books3 file list, Z-Library records via Anna’s Archive, and Copyright Office registration and recordation data.

The screen rebuilds the published Works List criteria; the official list is under seal.

Most files fit inside a single torrent piece

Collection	Files located	File wholly inside one piece	At least half in one piece
LibGen fiction	2,637,140	79%	97%
PiLiMi release 1	2,179,884	74%	93%
LibGen non-fiction	4,311,279	26%	61%

Defendants argue that BitTorrent uploads carried only fragments. Where a file sits wholly inside one piece, any upload of that piece carried the whole file. These figures do not show that a defendant uploaded a particular piece; that has to come from the defendant’s own records or other evidence.

Read from the torrents’ own file lists, using public archived copies; no swarm is joined and no book file is read. An expert can recheck every figure with a short standard-library script.

Pieces are mostly 4 MiB (LibGen 1–8 MiB). Figures cover located files in the current dataset.

DEFENDANT LISTS

Ready lists of works, by defendant and theory

List	Works	Basis
Anthropic: Books3 works not on the Works List, by our screen	4,515	<i>Bartz</i> record
Anthropic: LibGen works not on the Works List, by our screen	5,461	<i>Bartz</i> record
Anthropic: PiLiMi works not on the Works List, by our screen	1,629	<i>Bartz</i> record
Anthropic: works opted out of the Bartz settlement (court’s list, no timing check)	1,802	Court order (<i>Bartz</i> final approval)
Meta: LibGen	493,479	Court opinion (<i>Kadrey</i>)
Meta: Z-Library records added Aug. 2022 to Aug. 2023, via Anna’s Archive	221,052	Court opinion (<i>Kadrey</i>); Meta’s answer in <i>Elsevier</i>
DeepSeek: same Z-Library records, via Anna’s Archive	209,902	Company paper (DeepSeek-VL)
NVIDIA: same Z-Library records, via Anna’s Archive	217,965	Allegation; answer admits Anna’s Archive downloads, undated
OpenAI: LibGen, 2018	315,229	Court order; scope alleged
OpenAI: LibGen torrenting, 2019–2020	406,913	Allegation

Counted once per company (each company’s lists combined, one award per work): Anthropic 9,584 · Meta 616,230 · NVIDIA 266,677 · OpenAI 406,939.

Works = books with a registration dated before the copying could have begun (17 U.S.C. § 412), counting copies, formats, editions and translations of one book once. Every match is a candidate to confirm against the Office record. One award per work against each company, however many copies or collections (17 U.S.C. § 504(c)(1)): a company’s lists overlap and are not additive. Each list carries catalog IDs, MD5s, torrents and piece data. Basis shows what the defendant-side link rests on.

Ways to work with us

Paid pilot: \$1,000

Up to 50 works, chosen from our lists or by registration number. Every line reviewed by PayAuthors against the Office record; one correction pass; five business days.

Case support

Complaint exhibits for your works: catalog entries, MD5s, torrents, piece data and ownership screens, with a method note.

Expert support

Scripts and checksums so your expert can re-run the method. Declarations on how the data was built, as separate expert work.

Defendant lists

Lists of works by defendant and theory, as spreadsheets, for intake and case planning.

Client code packs

Codes your clients redeem for reports marked “Prepared for” your firm.

Five free reports

Request a code for five free reports.

Independent research: the same data, method and standards for counsel on either side and for rights holders. Custom work is confidential to the firm that commissions it.

ONGOING MONITORING

We check the AI copyright cases every hour, so your evidence stays current

New filings

Orders, complaints and motions in the AI copyright cases, checked hourly and logged with their source.

Timeline updates

Documented download dates go into the report timeline and the defendant lists, graded as court finding, company statement or allegation.

New catalogs

Newly archived catalog snapshots and torrent metadata are added in dataset releases.

METHOD AND LIMITS

Metadata only, from sources you can cite

We never download, open or share a pirated book, and never join a BitTorrent swarm. Torrent files are read for their file lists only, from public archived copies.

Measured accuracy. Against the Bartz opt-out list (Ex. J, 1,749 TX and TXu registrations), the matcher links 94.0% of the works found in the catalogs to the exact registration and 96.0% to some registration of the work. Against 7,001 textbook registrations in publishers' works lists: 93.1% and 96.9% of those found.

How often a match is right. In 100 random matches checked against the Office record, 91% were the same work and 72% the exact edition (about ± 6 points).

Checksums: every source file is recorded by SHA-256.

Verifiable: each match links to the Copyright Office record.

Independent: not affiliated with or endorsed by the U.S. Copyright Office.

Research support, not legal advice.

PRICING

Free to check. Reports from \$65.

Name check Free Counts of matching works, and the titles to confirm	Standard report \$65 Matches, catalog lines, timeline	Registration Details \$100 Adds the Office status record and recorded transfers	Upgrade \$35 Turns a Standard report into Registration Details for the same name
--	--	--	---

Firm pilot: \$1,000 for up to 50 works, reviewed line by line. Case support, defendant lists, code packs and expert declarations are priced per engagement.

If a report includes another person’s books, we correct it or refund it.

Request a code for five free reports.

payauthors.com • support@payauthors.com