FOR LAW FIRMS

Work-level research for AI copyright litigation

Trace your clients' books through dated library catalogs and Copyright Office records, with file identifiers, registration matches and timing screens for counsel to review.

Litigation support, not legal advice.

$50 per work · no minimum.

Each work comes with its registration, catalog lines and § 412 timing, reviewed against the Copyright Office record, as a spreadsheet and PDF exhibits. Rate card · Request a quote for custom work

What we produce for each work

One line from a live report, with five factual fields useful in work-level pleading and exhibit analysis. Every report and defendant list is built from lines like this.

The Lost Night

- **LIBGEN FICTION** **Bartz, Andrea** 2019 · English · Crown/Archetype · ISBN 9780525574712
  in 17 of 25 LibGen catalog snapshots: first Feb. 2020, last June 3, 2025 · added to the catalog June 12, 2019 (LibGen's own record)
- **LibGen fiction ID 2157462** · **EPUB** · MD5 `8b2a0eab82719cfd5ab89c3e4fd6e3de`
- torrent `f_2157000` · file 12 MiB, spans 4 pieces of 4.0 MiB, so no one piece holds the whole file (the largest part in one piece is 33% of the file)
- Registration match TX0008858704 (Nov. 7, 2019) · match: ISBN (confirmed by surname or title)

  TX0008858704 — THE LOST NIGHT: A Novel.
  - **Publication:** published Feb. 26, 2019
  - **Authorship:** Bartz, Andrea — author
  - **Claimants:** Andrea Bartz Inc., (organization)
  - **Recorded documents:** none in the Office's bulk data through June 26, 2025
- **Passes the date screen (§ 412): Anthropic, Books3 (early 2021); Anthropic, LibGen (June 2021); Meta, LibGen (October 2022)** · and 3 more documented events

- **Catalog entry:** the record as LibGen listed it, the date it was added, and which of 25 dated catalogs list it.
- **Exact file:** catalog number and format, with the MD5, the fingerprint of one specific file, where the catalog records it.
- **Torrent:** the named torrent that carried the file, and how the file sits in its pieces. When a file lies inside one piece, every verified copy of that piece holds the whole file; whether one uploader sent all of it has to come from that uploader's records.
- **Registration:** the Copyright Office record, with the claimant named on it (the owner when registered, not necessarily today), dates and any recorded transfers.
- **§ 412 timing** against each defendant's documented or alleged download date. Separately, a U.S. work generally can't be sued on until the Copyright Office registers or refuses it (17 U.S.C. § 411(a)). Works first published outside the United States are exempt from that requirement but not from § 412, so they can still claim actual damages and are not in our "works" counts. Our registration data runs through 2025.

Each work is also screened against the *Bartz* Works List criteria.

The fight in the AI copying cases has moved from whether a company copied a pirate library to which works were in it, who owns them, and when. Motions to dismiss now test ownership, timeliness and distribution work by work. PayAuthors supplies the work-level facts from the libraries' own published catalogs and the U.S. Copyright Office's records.

Screening a client list? Ask for a quote and we will screen the whole list.

What the record already says

Courts and the defendants themselves have supplied the collection-level facts. Each statement below is labeled by its source, and each source links to the document.

STATEMENT	SOURCE	KIND
Anthropic's co-founder "used the infamous BitTorrent protocol" to copy LibGen in June 2021.	*Bartz v. Anthropic*, N.D. Cal., ECF No. 244 at 2	Court order
Anthropic compared PiLiMi with its LibGen copies and torrented only the copies "that [we]re not in Lib[G]en."	*Bartz*, ECF No. 244 at 3	Court order
For settlement purposes only, the parties stipulated that every Works List work was downloaded by August 10, 2022, the agreed Download Date.	*Bartz*, ECF No. 418 at 49	Settlement stipulation
Contributory liability requires intent, shown by inducement or by a service tailored to infringement.	*Cox Communications v. Sony Music*, No. 24-171 (U.S. Mar. 25, 2026)	Supreme Court
Meta did not dispute, at the pleading stage, that uploading and downloading are infringing acts.	*Strike 3 Holdings v. Meta*, N.D. Cal., ECF No. 61 at 7	Court order
Final approval of the $1.5 billion settlement for the works on the Works List; judgment entered (July 20, 2026).	*Bartz*, ECF No. 680	Court order
The court let the authors add a claim over BitTorrent uploading (Mar. 25, 2026).	*Kadrey v. Meta*, N.D. Cal., Dkt. 700	Court order
Meta admits downloading parts of datasets, including Books3, LibGen, Z-Library and Sci-Hub, by direct download and torrenting, and that some data was uploaded while it torrented.	*Sullivan v. Meta*, N.D. Cal., Dkt. 85 ¶¶46, 49–52 (Sept. 29, 2026)	Company filing (answer)
NVIDIA admits downloading data for AI research from Anna’s Archive’s datasets page and from libgen.rs, without dates.	*Nazemian v. NVIDIA*, N.D. Cal., ECF No. 321	Company filing (answer)
OpenAI says it trained GPT-3 and GPT-3.5 on “compilations of books from Library Genesis.”	*In re OpenAI Copyright Litigation*, S.D.N.Y., Dkt. 1720	Company filing (brief)

Latest from the docket

court orders and company filings, live from our news feed · feed updated Oct. 31, 2026

Oct. 30, 2026	COMPANY FILING	**Suno asks the court to pause the labels' second suit over 60,000 more recordings until summary judgment on fair use is decided in the first case** UMG Recordings, Inc. v. Suno, Inc. (second action)
Oct. 29, 2026	COMPANY FILING	**Google, Microsoft, Meta and Anthropic tell investors the AI copyright claims are real and ongoing; Google drops "may" for "are"** SEC filings of Alphabet, Microsoft, Meta and Anthropic (on the AI copyright suits, including Bartz v. Anthropic and Kadrey v. Meta)
Oct. 28, 2026	COURT ORDER	**Judge Chhabria finds the Hobbs, Older and Sheard author suits against Meta related to Kadrey and takes them over** Kadrey v. Meta Platforms, Inc. / Hobbs v. Meta Platforms, Inc. / Older v. Meta Platforms, Inc. / Sheard v. Meta Platforms, Inc.
Oct. 27, 2026	COMPANY FILING	**Suno asks the court to make Apple sit for a deposition, saying Apple's talks with the labels and its streaming-fraud data go to market harm and Suno's copyright misuse defense** UMG Recordings, Inc. v. Suno, Inc. (subpoena to non-party Apple)
Oct. 23, 2026	COMPANY FILING	**OpenAI tells the court it downloaded books from LibGen and Books3 only to train its models, and asks that fair use be judged by that purpose, not copy by copy** In re OpenAI, Inc. Copyright Infringement Litigation (Authors Guild v. OpenAI; Alter v. OpenAI)

> **We supply the work-level link.** For each incident fixed by a court order or a company's own filing, a report shows whether a given owner's book was in that collection as it stood when it was taken: the catalog snapshot nearest the date, the exact file (catalog ID, and MD5 where the catalog records one) and, where it is archived, the torrent that carried it. Whether a defendant's download included that particular file comes from the defendant's own records.

Scale

PayAuthors for law firms: the data at a glance

22.4M

catalog records searched: LibGen, Z-Library and Books3, plus the Internet Archive's nonprofit lending library (records, not unique books)

31

dated catalogs: LibGen 25 · Z-Library 4 · Books3 1 · Internet Archive 1, Nov. 2015 – Oct. 2026

8.03M

U.S. copyright registrations for literary works, published (TX) and unpublished (TXu), 1978–2025

1.52M

registrations matched to catalog records, each with its method and a link to the Office record

6.95M

LibGen files located inside their torrents, with piece data (fiction 2.64M, non-fiction 4.31M), of 7.8M indexed by catalog number and MD5

578,874

registrations in our *Bartz* Works List screen (a work can have several). The final Works List has 482,460 works (Dkt. 680, July 20, 2026); the list itself is under seal

Data: PayAuthors dataset payauthors-2026-10-01-m64, built Oct. 5, 2026. Records, not unique books. Registration matches are candidates to confirm against the Copyright Office record. Our counts, not a court finding.

File-level records

Every LibGen and Z-Library line in a report names the file the library offered, in the library's own terms, so it can be compared with a defendant's file lists, hashes and torrent logs produced in discovery.

1

The library's own ID

The number LibGen or Z-Library gave that one file (LibGen numbers non-fiction and fiction separately).

```
Z-Library ID 5276476
```

2

The MD5 fingerprint

A checksum of the file's exact bytes, from the library's own catalog. A produced copy can be matched to it byte for byte, without anyone opening the book.

```
MD5 721630f8116455be2e24f39246759772
```

3

The torrent and its pieces

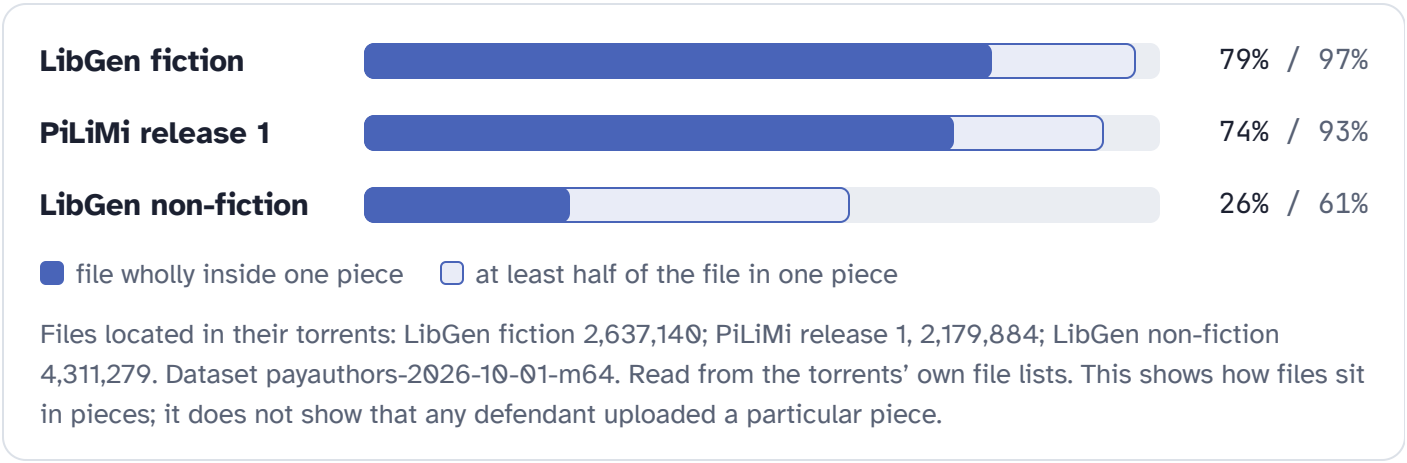
The BitTorrent bundle that carried the file, and the file's size, the piece length, how many pieces it spans, and the largest share of the file inside one piece, from the torrent file itself.

```
pilimi-zlib-5240000-5329999 · file 320 KiB, inside one 4.0 MiB piece
```

Example: *The Herd* by Andrea Bartz, from the sample report. Method and coverage: Methodology.

Distribution claims: a whole book often fits in one torrent piece

Defendants argue that BitTorrent uploads carried only fragments. BitTorrent checks data piece by piece but sends it in small blocks, which a downloader can collect from several peers. When a file lies inside one piece, every verified copy of that piece holds the whole file; whether one uploader sent all of it has to come from that uploader's records. Most LibGen fiction (79%) and PiLiMi release 1 (74%) files lie inside one piece; 26% of LibGen non-fiction files do. The torrents' own file lists do not show that a defendant uploaded a particular piece; that has to come from the defendant's own records or other evidence. We read them from public archived copies; no swarm is joined and no book file is read.



Files located in their torrents: LibGen fiction 2,637,140; PiLiMi release 1, 2,179,884; LibGen non-fiction 4,311,279. Dataset payauthors-2026-10-01-m64. Read from the torrents' own file lists. This shows how files sit in pieces; it does not show that any defendant uploaded a particular piece.

Pieces are mostly 4 MiB (LibGen 1 to 8 MiB). Figures cover located files in the current dataset. An expert can recheck every figure with a short standard-library script we supply.

Defendant lists

Ready lists of registrations by defendant and theory, each with catalog IDs, § 412 timing, and MD5s, torrents and piece data where they are recorded. The basis column says what the defendant-side link rests on. Lists identify works, not people, for analyzing claims of clients you already represent.

Works, not dollars. The lists count works that pass the registration-date screen; they are not damages estimates. If a court awards statutory damages, it sets them per work, from $750 to $30,000, or up to $150,000 for willful infringement (17 U.S.C. § 504(c)), with one award per work for all of one infringer's infringements in the case, or one shared award where infringers are jointly and severally liable; all parts of a compilation or derivative work count as one work (§ 504(c)(1)). So a book on two of a company's lists is still one work, a company's lists must not be added together, and different companies' lists are not a combined total. Whether any work qualifies is for a court. Actual damages and profits (§ 504(b)) are an alternative that must be proven.

LIST	WORKS	BASIS
Anthropic: Books3 works not on the Works List	5,419	*Bartz* record
Anthropic: LibGen works not on the Works List	28,750	*Bartz* record
Anthropic: PiLiMi works not on the Works List	6,787	*Bartz* record
Anthropic: LibGen works in torrents made June 20, 2021 to Aug. 10, 2022, not on the Works List (caution, see below)	18,904	*Bartz* record (agreed date, settlement only)
Anthropic: works first registered 1931–1977 (renewed, or registered 1964–1977), not on the Works List; candidates	21,806	*Bartz* record; Copyright Office registration and renewal records
Anthropic: works opted out of the *Bartz* settlement (their owners kept the right to sue)	1,802 works	Court order (*Bartz* final approval)
Meta: LibGen	493,479	Court opinion (*Kadrey*)
Meta: Z-Library records added Aug. 2022 to Aug. 2023, via Anna's Archive (early 2024)	221,052	Court opinion (*Kadrey*); Meta's answer in *Elsevier*
DeepSeek: same Z-Library records, via Anna's Archive (by March 2024)	209,902	Company paper (DeepSeek-VL)
OpenAI: LibGen, 2018 (October–November as alleged)	315,984	Court order (the year); months and scope alleged
OpenAI: LibGen torrenting, 2019 to 2020	406,913	Allegation
OpenAI: Books3 (via The Pile), February 2021	78,553	Allegation
NVIDIA: same Z-Library records, via Anna's Archive (August 2023)	217,965	Allegation; NVIDIA's answer admits downloading from Anna's Archive's datasets page, undated

Works = books with a registration dated before the download (17 U.S.C. § 412), counting copies, formats, editions and translations of one book once. Lists overlap and are not additive, and each match is a candidate to confirm. Opt-out works are counted by the court's list, without our timing check. Anthropic's rows count every work with a copy in its downloads that is not on the Works List as collected from the settlement's lookup site, under any registration or edition we can link to it, and that was registered before Anthropic's copying could have begun. Anthropic's count is 34,330 works with a TX or TXu registration and a copy in Anthropic's stated download windows. Another 18,093 such works have a copy only in a LibGen torrent made after Anthropic's stated LibGen window (June 8–19, 2021) and on or before Aug. 10, 2022; they are on its list but are reported only, not counted. **Caution:** Aug. 10, 2022 is the date by which the *Bartz* parties agreed every Works List work was downloaded, for the settlement only and only for Works List works. Anthropic's stated window is an estimate from code dates, and no record shows Anthropic downloaded those later torrents. Works first registered from 1931 to 1977 are shown in their own column for every company. They are matched by title and author to the original registrations for 1964–1977 and to the renewal records for earlier works, both as transcribed by the New York Public Library from the Catalog of Copyright Entries; each is counted once per company and left out if a copy carries a later TX or TXu registration, and for Anthropic also if the work, its renewal or another edition is on the Works List. They are candidates to confirm: in a check of 50 Anthropic matches, 44 were the same work. Registrations that cover only new matter, such as a revision, a translation or illustrations, are included and protect only that new matter. Another 136,767 (Meta) and 119,381 (NVIDIA) pre-1978 works are in Internet Archive lending books those companies are alleged to have downloaded; they are not counted. Other counts cover TX and TXu registrations from 1978 on.

One award per work: each company's lists combined

A book on two of a company's lists is one work and one award. Here each company's lists are combined and each work is counted once. Because statutory damages depend on registration before the infringement commenced (17 U.S.C. § 412), a work is counted only if a registration predates that company's first download of it, across all its lists. A work is on a company's list if it was in that library at the time, whatever its registration date, so a work first copied before it was registered stays out of the count even when a later copy came after registration. The first three columns cover TX and TXu registrations (1978 on); works first registered from 1931 to 1977 are counted once in the last column, as candidates. Another 136,767 (Meta) and 119,381 (NVIDIA) pre-1978 works are in Internet Archive lending books those companies are alleged to have downloaded; they are not counted.

COMPANY	WORKS ON ITS LISTS	ON TWO OR MORE LISTS	COUNTED (REGISTERED IN TIME)	PRE-1978 WORKS (CANDIDATES)
Anthropic Books3, LibGen and PiLiMi works not on the Works List; works whose only copy is a LibGen torrent made through Aug. 10, 2022 are on its list but not counted (caution)	62,537	8,176	**34,330**	21,792
Meta LibGen, Anna's Archive and Books3	625,735	148,757	**617,167**	41,130
NVIDIA Books3 and Anna's Archive (allegations)	274,043	29,992	**267,678**	57,884
OpenAI LibGen 2018 (court order; months and scope alleged), LibGen torrenting 2019–2020 (alleged) and Books3 2021 (alleged)	432,371	337,675	**422,577**	24,321

About the opt-out figures

What we work from

Beyond the library catalogs, an investigation can use any lawful source:

- torrent and magnet files, with their file lists, piece hashes, sizes and MD5s;
- court exhibits, and discovery that counsel shares with us under the case's protective order;
- company papers, model cards and dataset documentation;
- archived web pages;
- authors' and publishers' own records;
- tests of what AI models produce, run against the author's own legitimate copy of the book.

Ongoing monitoring

We check for new filings in the AI copyright cases every three hours, and post items after review, so an item can appear later than its filing. New court rulings, company statements and allegations are logged with their sources and graded; documented download dates go into the report timeline and the defendant lists after review, and newly archived catalogs are added in dataset releases. The latest items are on our News page. For firm clients we can flag developments that touch their clients' works.

Ways to work with us

Work-level research: $50 per work

Any number of works, chosen from our defendant lists or by registration numbers you give us. For each work: the registration and its Copyright Office record; each catalog line with the library's ID and MD5 for LibGen and Z-Library files, and the torrent and piece layout where the torrent is archived; the registration-date screen against each company; and a note on any edition or same-name risk. Every line is reviewed by PayAuthors against the Office record before delivery, and you get one correction pass. Delivered as a spreadsheet and PDF exhibits with a method note; we give a delivery date with the quote.

Case support

Complaint exhibits for your works: catalog entries, and MD5s, torrents and piece data where the library and its archived torrents record them, with ownership screens and a method note an expert can check.

Expert support

Your expert gets our scripts and checksums to re-run the method from the same sources.

Content checks

Some questions need a look inside a pirate copy: which edition it is, or whether the copyright page and other copyright management information were removed (17 U.S.C. § 1202). A person acting for the rightsholder does these checks, only for the rightsholder's own works, with their permission and at counsel's direction. Only that one file is downloaded, directly, never seeded, and every step goes in a chain-of-custody log. We supply the written protocol, the target list (each file's library ID, MD5 and torrent, and the edition the Copyright Office record describes) and a checklist, and we match the file's MD5 against the library's catalog. With the rightsholder's written permission, we can also receive the reviewer's copy, check it ourselves and quote what the case needs, such as the copyright notice or the edition statement. Without that permission, we review the completed checklist and never receive the book.

Defendant lists

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

Who we work with. Our published reports are available to anyone, with the same standards for everyone. Custom and case work is done for one side only: we do not take work for a party adverse to an existing client in the same matter. Custom work for a firm is confidential to that firm. Firm work is done under a short written engagement.

Rate card

Law firm work is priced per work.

SERVICE	PRICE
Work-level research: registration, catalog lines, § 412 timing, reviewed, as spreadsheet and PDF exhibits	$50 per work, no minimum
Case support, defendant lists, content-check protocols and review	By engagement

Firm work runs under a short written engagement covering confidentiality, the deliverable, corrections and limits of liability.

Method and limits

We never download or share a pirated book, and never join a BitTorrent swarm. We open one only in a content check: the rightsholder's own reviewer downloads it, and we receive it only with the rightsholder's written permission. Each file is identified by the library's own MD5 and its torrent's piece hashes, so a copy produced in discovery can be checked against them byte for byte (how). Every source file is recorded by SHA-256 checksum, and each registration match links to the Copyright Office record so it can be confirmed. Matching is automated and not perfect: expect some missed records and some matches that are not your clients'. A catalog listing shows that a file was in a library on a date; tying it to a particular download takes the defendant's own evidence. PayAuthors is not a law firm, does not refer clients to lawyers, and is not affiliated with or endorsed by the U.S. Copyright Office. See the Methodology for sources and checksums.

> **Supporting research materials are preserved for review.** Where legally permissible, underlying source records and research materials can be made available for inspection by counsel, experts, or other qualified reviewers, under a short written engagement. What is preserved

This page is general information, not legal advice.

© 2026 PayAuthors.com. All rights reserved. Court records quoted here are public records; each quotation is cited to its source.

Charts, figures and tables on this site are © 2026 PayAuthors.com. You may share them unaltered with credit and a link. Reuse terms.