Rillor

RESEARCH REPORT

Copyright and model training: what U.S. courts and the Copyright Office have said

Training data copyright in U.S. law, from the statute and the Copyright Office's training report to court rulings on headnotes, books, news, images and code, each with its date.



Rillor

Contents

01 Summary

02 Background

03 Method

04 Findings

05 Implications

06 Limits of this analysis

07 Sources

Summary

U.S. courts and the Copyright Office have weighed copying for model training under the four fair use factors of the Copyright Act, and on 8 November 2026 their answers differ with the tool, the source of the works and the market at stake.

- On 29 September 2026 the Third Circuit affirmed a ruling that ROSS Intelligence's use of 2,243 Westlaw headnotes to train a non-generative legal search tool infringed Thomson Reuters's copyright and was not fair use [8]. Authors Alliance notes that the ruling binds federal courts in Pennsylvania, New Jersey, Delaware and the Virgin Islands [12].
- In June 2025 two judges in the Northern District of California ruled for model developers on fair use in training language models on books, at summary judgment and each on its own record [13, 14]. One sent to trial the pirated copies of books that its developer kept in a central library [13]. The other said its ruling affected only the thirteen authors before it [14].
- How works were obtained counts too. The Copyright Office says the knowing use of pirated or illegally accessed works "should weigh against fair use without being determinative" [23], and one June 2025 ruling held that using pirated copies of books for a permanent, general-purpose library was not a fair use [13].
- A class settlement in one of the June 2025 cases set a $1.5 billion fund, about $3,000 per work before costs and fees, and received final approval on 20 July 2026 [16, 17]. Its release covers past conduct only and states that it is not a license to train models on any copyrighted works [16, 17].
- The Copyright Office released its training report as a pre-publication version on 9 May 2025, and its web page still listed it that way on 8 November 2026 [23, 24]. The report reads the fourth fair use factor, the effect on the market for the work, to include market dilution, the risk that generated content competes with works of the same kind as those in the training data, and disfavors unlicensed uses where licensing markets are available [23].
- The sources split on licensing markets. The Third Circuit counted a market for licensing headnotes as training data against ROSS [8], and both June 2025 rulings in the Northern District of California on training with books said book authors are not entitled to a market for licensing their works for training [13, 14].
- Rulings on motions to dismiss in image and news cases let copyright claims go forward without deciding fair use [19, 20]. On 16 September 2026 the Ninth Circuit affirmed the dismissal of programmers' Digital Millennium Copyright Act claims about code outputs and held their training theory forfeited [22].

Background

What copyright protects and what fair use weighs

The owner of a copyright holds exclusive rights, among them the rights "to reproduce the copyrighted work in copies or phonorecords", "to prepare derivative works based upon the copyrighted work" and "to distribute copies or phonorecords of the copyrighted work to the public" [1]. Section 107 of the Copyright Act provides that fair use, for purposes "such as criticism, comment, news reporting, teaching (including multiple copies for classroom use), scholarship, or research, is not an infringement of copyright" [1]. Courts weigh four factors, which the statute lists in this order [1]:

Factor	The statute's words
1. Purpose and character	"the purpose and character of the use, including whether such use is of a commercial nature or is for nonprofit educational purposes"
2. Nature of the work	"the nature of the copyrighted work"
3. Amount used	"the amount and substantiality of the portion used in relation to the copyrighted work as a whole"
4. Market effect	"the effect of the use upon the potential market for or value of the copyrighted work"

Why model training reached the courts and the Office

Training copies works. In the Copyright Office's description, developers make copies by downloading works, moving them across storage media, converting them to other formats and creating modified versions or filtered subsets [23]. By May 2025 the Office counted "dozens of lawsuits" in the United States on how fair use applies [23]. It had launched its own initiative in early 2023 and published a notice of inquiry on 30 August 2023 [24, 25].

This report is for teams that acquire or license training data, for data holders that license it out and for whoever keeps the records of where data came from.

Method

What was examined. What U.S. courts, the Copyright Office and the Justice Department have said about copying works to train models, with each ruling's court, date and procedural stage.

Sources. 27 sources, each opened on 8 November 2026: the statute as the Copyright Office publishes it [1, 2], four Supreme Court opinions [3, 4, 5, 6], the Office's Compendium, report, web page and notice of inquiry [7, 23, 24, 25], opinions, orders and dockets of federal courts [8, 9, 10, 13, 14, 15, 19, 20, 22, 26, 27], filings and settlement records [11, 16, 17, 18, 21] and one commentary by a group that filed an amicus brief supporting ROSS [12].

How. Each opinion and order was read in full text, and every quote is taken from the page cited. Each ruling is given with its court, date and stage: motion to dismiss, summary judgment, settlement approval or appeal. Parties that develop or sell models are described by role, and their names appear only in the case titles in Sources. Thomson Reuters and ROSS Intelligence are named because their case is the subject studied. Where a quoted passage names a party or a model, the name is replaced in square brackets.

Findings

1. The statute: rights, damages and works of the government

Section 105(a) reads: "Copyright protection under this title is not available for any work of the United States Government" [1]. The same subsection lets the government receive and hold copyrights "transferred to it by assignment, bequest, or otherwise" [1].

Statutory damages are set per work, "not less than $750 or more than $30,000 as the court considers just", and up to $150,000 where the infringement "was committed willfully" [2]. The class settlement in Finding 7 is also counted per work, at about $3,000 for each work on its list [17]. The section on fair use adds that a work's being unpublished "shall not itself bar a finding of fair use" [1].

2. The Supreme Court frame: originality, transformation and purpose

Originality. Feist Publications v. Rural Telephone Service, decided on 27 March 1991, holds that "The sine qua non of copyright is originality", which requires independent creation and "at least some minimal degree of creativity" [4]. Facts "do not owe their origin to an act of authorship" and so are not copyrightable, and "no one may copyright facts or ideas" [4].

Transformation. Campbell v. Acuff-Rose Music, decided on 7 March 1994, frames the first factor as asking whether the new work merely "supersede[s] the objects" of the original creation "or instead adds something new, with a further purpose or different character, altering the first with new expression, meaning, or message", and so "whether and to what extent the new work is 'transformative'" [3]. The more transformative the new work, the less other factors such as commercialism count [3].

Purpose. Andy Warhol Foundation v. Goldsmith, decided on 18 May 2023, says the first factor "focuses on whether an allegedly infringing use has a further purpose or different character, which is a matter of degree, and the degree of difference must be weighed against other considerations, like commercialism" [5]. The Third Circuit quotes Warhol's rule that where an original and a secondary use "share the same or highly similar purposes, and the secondary use is of a commercial nature, the first factor is likely to weigh against fair use, absent some other justification for copying" [8].

3. Legal text: government edicts and private headnotes

In Georgia v. Public.Resource.Org, decided on 27 April 2020, the Supreme Court held under the government edicts doctrine that officials empowered to speak with the force of law "cannot be the authors of" the works they create in the course of their official duties [6]. It cited Banks v. Manchester (1888) for the rule's application to judges [6].

The Copyright Office's Compendium, in a chapter dated 28 January 2021, applies the doctrine to "binding opinions issued by any federal, state, local, or territorial court", and to "any concurrence, dissent, syllabus, headnote, statement of the case, or other 'non-binding, explanatory legal materials'" a judge creates in a

judicial capacity [7].

The line falls at authorship. A headnote a judge writes falls under the doctrine [7], and headnotes written by a private publisher's editors are what Thomson Reuters v. ROSS decided [8]. Thomson Reuters disclaims copyright in "any part of the original work prepared by a United States Government officer or employee as part of that person's official duties", such as the opinions its headnotes summarize [8].

4. Thomson Reuters v. ROSS in the District of Delaware, 2025

ROSS built a search engine that "would respond to plain-language legal questions with relevant passages of text from judicial opinions", with access to "approximately ten million uncopyrighted judicial opinions" [8]. LegalEase sold ROSS "roughly 25,000 Bulk Memos", which ROSS used to train the tool [9]. Each memo "presented a legal question and four to six judicial opinion passages in response" [8].

On 11 February 2025 Circuit Judge Bibas, sitting by designation, revised his 2023 summary judgment opinion [9]. Of 2,830 headnotes compared, he found actual copying of 2,243 [8, 9]. "It is undisputed that Ross's AI is not generative AI (AI that writes new content itself)", he wrote, and "only non-generative AI is before me today" [9]. He found that the second factor "goes for Ross", and held as a matter of law that the use of the 2,243 headnotes infringed and was not fair use [8, 9]. In a memorandum opinion of 23 May 2025 he explained his earlier order granting ROSS's motion for interlocutory appeal of both questions, originality and fair use, and a stay pending appeal [10].

5. Thomson Reuters v. ROSS in the Third Circuit, 29 September 2026

Judges Restrepo, Montgomery-Reeves and Bove heard argument on 11 June 2026, and the opinion of the court, by Judge Montgomery-Reeves, was filed on 29 September 2026 [8]. The court held "that Thomson Reuters's materials are copyrightable and that ROSS's use was not fair" [8]. A footnote says "AI here refers to machine learning technology in which a software program analyzes data to learn how to perform a task" [8].

Originality. "All 2,243 independently created headnotes at issue in this appeal have 'some creative spark'" [8]. The editors "made creative decisions about which points of law were important enough to include in a headnote and how to word those points of law", while judicial opinions "remain 'free for publication to all'" [8].

The four factors.

Factor	What the court found	Weight
1. Purpose and character	Training was "an intermediate step" that "arguably presents a slight degree of difference in use", and the shared "ultimate purpose" made the use "minimally transformative, at best"	Against [8]
2. Nature of the work	The headnotes "are more factual than fictional"	Favors [8]

Factor	What the court found	Weight
3. Amount used	Each headnote is a work, so for each one taken "ROSS copied an entire work", and copying was "not necessary" because the opinions were freely available	Against [8]
4. Market effect	"the market for licensing headnotes as text to train AI is rapidly developing", and ROSS "usurped Thomson Reuters's opportunity to enter that derivative market"	Against [8]

ROSS had argued that it took "only 0.08% of Thomson Reuters's 28 million headnotes" [8]. On the market, the court noted that Thomson Reuters "is using its headnotes as training data for its own AI search products", and that its not licensing them to others "does not disprove that a market exists to do so" [8].

The precedents it set apart. The search feature in a 2015 book-scanning case in the Second Circuit "served a completely different function", while ROSS's use, as ROSS admits, "aims to replace Westlaw" [8]. In the cases on intermediate copying of computer code, "copying was necessary to access the unprotected functional aspects of computer code", and ROSS "had access to the underlying judicial opinions" [8]. "Unlike necessity, ease is not a justification for copying" [8]. A footnote adds that "to the extent good faith remains relevant, that consideration weighs against fair use" [8].

The limits it stated. A footnote notes the Justice Department's statement of interest in a generative case (Finding 8) and says "The concerns raised in that separate case do not apply here", because ROSS's platform "cannot generate original expression" [8]. The court also wrote that "this does not give ROSS carte blanche to violate copyright law merely because it incorporates AI" [8].

After the ruling. Authors Alliance, which filed an amicus brief supporting ROSS, writes that "No one should treat this decision as the last word on LLM training" [12]. It argues that treating each headnote as a whole work turns the third factor against almost any use of abstracts, annotations or summaries, and that "Noncommercial researchers who do not compete with their sources remain on firmer ground" [12]. On 3 November 2026 ROSS petitioned for panel rehearing or rehearing en banc, saying the decision "warrants the full Court's review" [11].

6. Generative training on summary judgment: two June 2025 rulings

Two judges of the Northern District of California ruled two days apart on training language models on books, and both ruled for the developers on fair use, by different routes [13, 14].

	Authors against a model developer	Thirteen authors against another model developer
Date and judge	23 June 2025, Judge Alsup [13]	25 June 2025, Judge Chhabria [14]

	Authors against a model developer	Thirteen authors against another model developer
Training use	"exceedingly transformative" and a fair use [13]	Partial summary judgment for the developer on its record [14]
How works were obtained	Pirated library copies sent to trial with damages [13]	A separate distribution claim continued [14, 15]
Market	No entitlement to a licensing market for training [13]	Dilution the "potentially winning argument", with no evidence offered [14]

The first ruling. The order states that the developer "pirated over seven million copies of books" and also "purchased millions of print copies" [13]. It held that training on the books "was exceedingly transformative and was a fair use under Section 107 of the Copyright Act" [13]. Elsewhere it calls the use transformative, "spectacularly so", and notes that "Authors do not allege that any LLM output provided to users" infringed [13]. On the pirated copies it held that "Creating a permanent, general-purpose library was not itself a fair use excusing [its] piracy", and it set those copies and the damages, "actual or statutory (including for willfulness)", for trial [13].

On the market, it compared the authors' complaint to one that "training schoolchildren to write well would result in an explosion of competing works": "This is not the kind of competitive or creative displacement that concerns the Copyright Act" [13]. A market for licensing books for training "could develop", it said, and "such a market for that use is not one the Copyright Act entitles Authors to exploit" [13].

The second ruling. The order opens by saying that in most cases training on protected works without permission is likely illegal, "Although the devil is in the details" [14]. It calls the schoolchildren comparison an "inapt analogy" [14]. The authors' theories of regurgitation and a lost licensing market "are clear losers", and "the plaintiffs are not entitled to the market for licensing their works as AI training data" [14]. The developer's expert "could get no model to generate more than 50 words and punctuation marks (that is, 'tokens') from the plaintiffs' books" [14].

Had the authors "presented any evidence that a jury could use to find in their favor" on market dilution, "factor four would have needed to go to a jury", the court wrote, and they "presented no meaningful evidence on market dilution at all" [14]. The ruling "does not stand for the proposition that [the developer's] use of copyrighted materials to train its language models is lawful" [14]. A claim that the developer "unlawfully distributed their protected works during the torrenting process" continued [14], and on 25 March 2026 the court let the authors add a contributory infringement claim, based on the allegation that the developer uploaded works while downloading them through BitTorrent [15].

7. The class settlement: what the court approved and what it is not

The first case settled. The long-form notice, ordered on 18 November 2025, says the suit claimed infringement "by downloading works in two allegedly pirated online datasets" [17]. It describes "a Settlement Fund of $1.5 billion" and "approximately $3,000 per work, prior to the deduction of any costs, fees, and expenses" [17]. The settlement requires the developer to destroy the books it downloaded from the two datasets, subject to preservation obligations [17].

After a hearing on 14 May 2026, with "submissions from fifty-four class members and putative class members", the court granted final approval on 20 July 2026 [16]. "As of April 16, 2026, 91.3 percent of the Works List had been claimed, for a total of 440,490 works of the 482,460 Works on the Works List" [16]. The claim deadline was 30 March 2026 [18].

The release covers claims "broadly related to past AI inputs, namely pirating and copying their works in the past up to the point of any AI output" [16]. Class members "do not give up claims about past AI outputs, nor claims of any kind about future conduct (on or after August 25, 2025)" [16]. The notice states: "The release does not constitute, in any respect, a license to torrent, scan, or train AI models on any copyrighted works, or to create infringing outputs from AI models" [17].

The administrator's site, viewed on 8 November 2026, says the parties have told the court "that the Effective Date has passed, assuming that no late-filed appeal is permitted to go forward", and that it expects to send the first payments on or before 15 November 2026 [18].

8. Rulings in other cases, at the pleading stage and on appeal

The image and news orders below decide only whether claims may go forward [19, 20].

Date	Court	Parties, by role	What was decided
12 August 2024	N.D. Cal., motion to dismiss	Visual artists against image model developers	Copyright Act claims go forward, Digital Millennium Copyright Act (DMCA) claims dismissed with prejudice [19]
4 April 2025	S.D.N.Y., motion to dismiss	News publishers against a model developer and its commercial partner	Contributory and older direct infringement claims go forward, misappropriation and some DMCA claims dismissed [20]
16 September 2026	Ninth Circuit, appeal from dismissal	Programmers against a code-hosting platform and the developers of its code tools	DMCA dismissal affirmed, training theory held forfeited [22]

Images. The court found that "The plausible inferences at this juncture are that [the model] by operation by end users creates copyright infringement and was created to facilitate that infringement by design", a point to be "tested at a later date" [19]. "The allegations of induced infringement are sufficient" [19].

News. The defendants moved to dismiss the contributory infringement, DMCA and misappropriation claims, direct infringement claims for conduct more than three years before the complaints, trademark dilution claims in one action and the "abridgment" claims of direct infringement in another [20]. Letting a DMCA claim proceed where a work "is not 'close to identical' to the original would risk boundless DMCA liability", the court wrote [20]. The motions did not raise fair use, and the order did not decide it [20].

Code. The Ninth Circuit held that the tools "do not 'remove or alter' copyright management information (CMI) from a copy of an existing protected work but instead create new works that never contained that information" [22]. On the theory that the defendants removed CMI "at the training stage", the court wrote: "We decline to consider the input theory because plaintiffs forfeited it" [22].

The Justice Department. On 1 September 2026 the United States filed a statement of interest under 28 U.S.C. 517 in consolidated cases against a model developer in the Southern District of New York [21]. It states "a strong interest in this Court rejecting any argument that training LLMs on copyrighted texts violates copyright law", and says "the use of copies to train LLMs is extraordinarily transformative" [21]. Noting that "Each stage may present distinct questions of copyright law", it focuses on the training stage [21]. It calls the second June 2025 ruling's "application of the fourth fair-use factor" "deeply flawed", and "takes no position on whether a licensing regime would be financially or logistically feasible" [21].

9. The Copyright Office: the inquiry and Part 3

The notice of inquiry of 30 August 2023 said the Office "is undertaking a study" of issues including "the use of copyrighted works to train AI models" [25]. The Office received more than 10,000 comments by December 2023 [24]. It published Part 1 of its report, on digital replicas, on 31 July 2024 and Part 2, on the copyrightability of outputs, on 29 January 2025, and released Part 3, on training, as a pre-publication version on 9 May 2025 [24]. On 8 November 2026 its web page listed "Part 3: Generative AI Training (Pre-publication)" and said the Office expects no substantive changes in the analysis or conclusions of a final version [24]. The report gives an analytical framework "Without opining on specific cases" [23].

Where the copying happens. "The steps required to produce a training dataset containing copyrighted works clearly implicate the right of reproduction" [23]. Training "may result in model weights that contain copies of works in the training data", and then later copying of the weights "could also constitute prima facie infringement" [23]. Retrieval-augmented generation (RAG) "also involves the reproduction of copyrighted works" [23].

The first factor. In the Office's view, training a foundation model on a large and diverse dataset is often transformative [23]. "Where a model is trained on specific types of works in order to produce content that shares the purpose of appealing to a particular audience, that use is, at best, modestly transformative" [23].

Access. The knowing use of pirated or illegally accessed works "should weigh against fair use without being determinative" [23]. "Copyright owners have a right to control access to their works, even if someone seeks to obtain them in order to make a fair use" [23].

The fourth factor. "While we acknowledge this is uncharted territory, in the Office's view, the fourth factor should not be read so narrowly" [23]. "The speed and scale at which AI systems generate content pose a serious risk of diluting markets for works of the same kind as in their training data" [23]. Where licensing markets are available to meet training needs, the Office says unlicensed uses are disfavored, and where barriers to licensing prove insurmountable, there is "no functioning market to harm" and "the fourth factor may favor fair use" [23].

Licensing. "Some AI systems have now been trained exclusively on licensed or public domain works", though "The growing licensing market does not itself establish that voluntary licensing is feasible at scale for all AI training needs" [23]. Research projects "may well qualify as fair use and therefore would not require licenses" [23].

The conclusion. How far training uses are fair depends, in the Office's words, "on what works were used, from what source, for what purpose, and with what controls on the outputs" [23]. The Office concludes that "making commercial use of vast troves of copyrighted works to produce expressive content that competes with them in existing markets, especially where this is accomplished through illegal access, goes beyond established fair use boundaries" [23]. Citing "the robust growth of voluntary licensing", the Office called government intervention "premature at this time", and said "alternative approaches such as extended collective licensing should be considered to address any market failure" [23].

10. Where the sources agree and disagree

Question	Third Circuit	First June 2025 ruling	Second June 2025 ruling	Copyright Office	Justice Department
Licensing market as harm	Counted against ROSS [8]	No entitlement for authors [13]	No entitlement for authors [14]	Unlicensed uses disfavored where licensing is available [23]	No position on feasibility [21]
Market dilution	Factor four turned on a derivative licensing market [8]	Competition theory rejected [13]	"potentially winning argument" [14]	Counted under the fourth factor [23]	Application called "deeply flawed" [21]

Question	Third Circuit	First June 2025 ruling	Second June 2025 ruling	Copyright Office	Justice Department
Pirated or illegal access	Good faith, if relevant, against ROSS [8]	Library copies sent to trial [13]	Distribution claim continued [14, 15]	Weighs against, not determinative [23]	Not addressed. The statement focuses on the training stage [21]

Licensing markets. The Third Circuit accepted a market for licensing headnotes as training data even though Thomson Reuters had not licensed them to others [8]. Both June 2025 rulings declined to give book authors such a market [13, 14].

Market dilution. The Office treats dilution of markets for works of the same kind as a harm the fourth factor reaches [23]. The second June 2025 ruling called it the potentially winning argument and found no evidence of it in the record [14]. The Justice Department calls that ruling's handling of the factor "deeply flawed" [21].

Generative and non-generative. The Delaware court and the Third Circuit both treated ROSS's tool as non-generative [8, 9]. The June 2025 rulings, the Office's report and the Justice Department's statement concern generative training [13, 14, 21, 23].

Outputs and authorship. Outputs sit apart from training in the record. The first June 2025 ruling noted that the authors did not allege infringing outputs [13], class members kept their output claims under the settlement's release [16, 17], and the Ninth Circuit's ruling concerns outputs [22]. On 18 March 2025 the D.C. Circuit held that "the Copyright Act of 1976 requires all eligible work to be authored in the first instance by a human being", and that "the human authorship requirement does not prohibit copyrighting work that was made by or with the assistance of artificial intelligence" [26]. The Supreme Court denied review on 2 March 2026 [27].

Implications

For teams that acquire or license training data. The rulings and the Office weighed the same facts: which works were used, where they came from, how they were obtained, what the use was for and whether a market to license them existed [8, 13, 14, 23]. Each of those facts can sit in a record kept from the start of an acquisition: the source of each work, the means of access, the terms it came under and the use made of it. Rillor finds the publishers and data holders that hold the data a team needs, agrees use, term and onward sharing in writing before delivery, and delivers each record with the license it came under.

For data holders that license data out. The Third Circuit counted a licensing market for headnotes, citing Thomson Reuters's own use of them as training data [8], while both June 2025 rulings said book authors are not entitled to one for their works [13, 14]. The Office records a growing licensing market [23].

For whoever keeps the records. The class settlement counted works on a list and paid per work [16, 17], and statutory damages are counted per work [2]. The Office notes that model weights may contain copies of works and that retrieval-augmented generation also reproduces works [23].

Limits of this analysis

- This report is research on what statutes, court records and the Copyright Office say. It is never legal advice and gives no view on whether any reader's use is fair.
- The Third Circuit ruling binds federal courts in that circuit [12], and the second June 2025 ruling says it affects only the thirteen authors before it [14].
- Part 3 of the Office's report is a pre-publication version, as its web page showed on 8 November 2026 [24].
- ROSS's petition of 3 November 2026 is covered only as filed [11]. Events in the second June 2025 case after 25 March 2026, and any appeal in the first, are outside the sources read.
- None of the three rulings in Finding 8 decides fair use [19, 20, 22]. Other image, music, news and code cases were not opened.
- Law outside the United States is out of scope.

Sources

1. 17 U.S.C. chapter 1, sections 105, 106 and 107. U.S. Copyright Office, Title 17, Chapter 1, viewed 8 November 2026. https://www.copyright.gov/title17/92chap1.html

2. 17 U.S.C. 504(c), Statutory damages. U.S. Copyright Office, Title 17, Chapter 5, viewed 8 November 2026. https://www.copyright.gov/title17/92chap5.html

3. Campbell v. Acuff-Rose Music, Inc., 510 U.S. 569. Supreme Court of the United States, 7 March 1994, via Legal Information Institute, viewed 8 November 2026. https://www.law.cornell.edu/supremecourt/text/510/569

4. Feist Publications, Inc. v. Rural Telephone Service Co., 499 U.S. 340. Supreme Court of the United States, 27 March 1991, via Legal Information Institute, viewed 8 November 2026. https://www.law.cornell.edu/supremecourt/text/499/340

5. Andy Warhol Foundation for the Visual Arts, Inc. v. Goldsmith, No. 21-869. Supreme Court of the United States, 18 May 2023, viewed 8 November 2026. https://www.supremecourt.gov/opinions/22pdf/21-869_87ad.pdf

6. Georgia v. Public.Resource.Org, Inc., No. 18-1150. Supreme Court of the United States, 27 April 2020, viewed 8 November 2026. https://www.supremecourt.gov/opinions/19pdf/18-1150_7m58.pdf

7. Compendium of U.S. Copyright Office Practices, Third Edition, Chapter 300: Copyrightable Authorship, section 313.6(C)(2). U.S. Copyright Office, 28 January 2021, viewed 8 November 2026. https://www.copyright.gov/comp3/chap300/ch300-copyrightable-authorship.pdf

8. Thomson Reuters Enterprise Centre GmbH v. ROSS Intelligence Inc., No. 25-2153, opinion of the court. U.S. Court of Appeals for the Third Circuit, 29 September 2026, viewed 8 November 2026. https://www2.ca3.uscourts.gov/opinarch/252153p.pdf

9. Thomson Reuters Enterprise Centre GmbH v. Ross Intelligence Inc., No. 1:20-cv-613-SB, Memorandum Opinion. U.S. District Court for the District of Delaware, via GovInfo, 11 February 2025, viewed 8 November 2026. https://www.govinfo.gov/content/pkg/USCOURTS-ded-1_20-cv-00613/pdf/USCOURTS-ded-1_20-cv-00613-5.pdf

10. Thomson Reuters Enterprise Centre GmbH v. Ross Intelligence Inc., No. 1:20-cv-613-SB, Memorandum Opinion on interlocutory appeal. U.S. District Court for the District of Delaware, via GovInfo, 23 May 2025, viewed 8 November 2026. https://www.govinfo.gov/content/pkg/USCOURTS-ded-1_20-cv-00613/pdf/USCOURTS-ded-1_20-cv-00613-6.pdf

11. Appellant ROSS Intelligence Inc.'s Petition for Panel Rehearing or Rehearing En Banc, No. 25-2153. U.S. Court of Appeals for the Third Circuit, via CourtListener, 3 November 2026, viewed 8 November 2026. https://storage.courtlistener.com/recap/gov.uscourts.ca3.125346/gov.uscourts.ca3.125346.225.0.pdf

12. Third Circuit affirms in Thomson Reuters v. ROSS: The Court Gets Merger Wrong and Does a Disservice to Fair Use. Dave Hansen, Authors Alliance, 2 October 2026, viewed 8 November 2026. https://www.authorsalliance.org/2026/10/02/third-circuit-affirms-in-thomson-reuters-v-ross-the-court-gets-merger-wrong-and-does-a-disservice-to-fair-use/

13. Bartz v. Anthropic PBC, No. C 24-05417 WHA, Order on Fair Use, Document 231. U.S. District Court for the Northern District of California, via GovInfo, 23 June 2025, viewed 8 November 2026.

https://www.govinfo.gov/content/pkg/USCOURTS-cand-3_24-cv-05417/pdf/USCOURTS-cand-3_24-cv-05417-0.pdf

14. Kadrey v. Meta Platforms, Inc., No. 23-cv-03417-VC, Order on Cross-Motions for Partial Summary Judgment, Document 598. U.S. District Court for the Northern District of California, via CourtListener, 25 June 2025, viewed 8 November 2026. https://storage.courtlistener.com/recap/gov.uscourts.cand.415175/gov.uscourts.cand.415175.598.0.pdf

15. Kadrey v. Meta Platforms, Inc., No. 23-cv-03417-VC, Order Granting Motion for Leave to File Fourth Amended Complaint, Document 700. U.S. District Court for the Northern District of California, copy hosted by Ars Technica, 25 March 2026, viewed 8 November 2026. https://cdn.arstechnica.net/wp-content/uploads/2026/03/Kadrey-v-Meta-Order-Granting-Motion-for-Leave-3-25-26.pdf

16. Bartz v. Anthropic PBC, No. 3:24-cv-05417-AMO, Order Granting Final Approval of Class Action Settlement, Document 680. U.S. District Court for the Northern District of California, 20 July 2026, viewed 8 November 2026. https://assets-us-01.kc-usercontent.com:443/1eeb16db-4934-006e-40a6-38fa91285ebb/36cce252-a32c-4cbb-a624-31dd8ac9574c/2026-07-20%20Order%20Granting%20Final%20_dckt%20680_0_.pdf

17. Notice of $1.5 Billion Proposed Class Action Settlement, Bartz v. Anthropic PBC, No. 3:24-cv-05417 (N.D. Cal.), long-form notice. Settlement Administrator, 18 November 2025, viewed 8 November 2026. https://assets-us-01.kc-usercontent.com:443/1eeb16db-4934-006e-40a6-38fa91285ebb/ce109a57-390c-4d76-ad65-5807d4830fe5/ANT%20-%20Long-Form%20Notice_Updated%20Deadline%201.8.2026.pdf

18. Anthropic Copyright Settlement. JND Legal Administration, Settlement Administrator, viewed 8 November 2026. https://www.anthropiccopyrightsettlement.com/

19. Andersen v. Stability AI Ltd., No. 23-cv-00201-WHO, Order Granting in Part and Denying in Part Motions to Dismiss First Amended Complaint, Document 223. U.S. District Court for the Northern District of California, via GovInfo, 12 August 2024, viewed 8 November 2026. https://www.govinfo.gov/content/pkg/USCOURTS-cand-3_23-cv-00201/pdf/USCOURTS-cand-3_23-cv-00201-2.pdf

20. The New York Times Company v. Microsoft Corporation et al., No. 23-cv-11195 (SHS), Opinion, Document 514. U.S. District Court for the Southern District of New York, via CourtListener, 4 April 2025, viewed 8 November 2026. https://storage.courtlistener.com/recap/gov.uscourts.nysd.612697/gov.uscourts.nysd.612697.514.0.pdf

21. Statement of Interest of the United States, In re: OpenAI, Inc. Copyright Infringement Litigation, No. 25-md-3143. U.S. Department of Justice, filed in the U.S. District Court for the Southern District of New York, via CourtListener, 1 September 2026, viewed 8 November 2026. https://storage.courtlistener.com/recap/gov.uscourts.nysd.640396/gov.uscourts.nysd.640396.1682.0.pdf

22. J. Doe 1 v. GitHub, Inc., No. 24-7700, opinion. U.S. Court of Appeals for the Ninth Circuit, 16 September 2026, viewed 8 November 2026. https://cdn.ca9.uscourts.gov/datastore/opinions/2026/09/16/24-7700.pdf

23. Copyright and Artificial Intelligence, Part 3: Generative AI Training (pre-publication version). U.S. Copyright Office, May 2025, released 9 May 2025, viewed 8 November 2026. https://www.copyright.gov/ai/Copyright-and-Artificial-Intelligence-Part-3-Generative-AI-Training-Report-Pre-Publication-Version.pdf

24. Copyright and Artificial Intelligence. U.S. Copyright Office, web page, viewed 8 November 2026. https://www.copyright.gov/ai/

25. Artificial Intelligence and Copyright, notice of inquiry, 88 FR 59942. U.S. Copyright Office, Library of Congress, Federal Register, 30 August 2023, viewed 8 November 2026. https://www.federalregister.gov/documents/2023/08/30/2023-18624/artificial-intelligence-and-copyright

26. Thaler v. Perlmutter, No. 23-5233. U.S. Court of Appeals for the District of Columbia Circuit, 18 March 2025, viewed 8 November 2026. https://media.cadc.uscourts.gov/opinions/docs/2025/03/23-5233.pdf

27. Thaler v. Perlmutter, No. 25-449, docket. Supreme Court of the United States, petition denied 2 March 2026, viewed 8 November 2026. https://www.supremecourt.gov/search.aspx?filename=/docket/docketfiles/html/public/25-449.html

About this report

Published by Rillor on 11 November 2026. Superintelligence (SI) tools assist research and drafting, and every figure cites its source. Rillor builds and secures SI systems, supplies the data they learn from and delivers the infrastructure they run on. We apply SI to financial and compute markets.

Cite as: Rillor. Copyright and model training: what U.S. courts and the Copyright Office have said. 11 November 2026. https://rillor.com/insights/copyright-and-model-training

Work with Rillor

Send the data you need, the publishers or holders you know of, and the use the license must cover. Contact us at rillor.com/contact?type=acquisition or contact@rillor.com. More on acquisition and licensing: rillor.com/data#acquisition