Rillor

RESEARCH REPORT

Training data licensing deals: what public filings and reports disclose

Most data licensing deals for model training are announced without a price. What SEC filings, results statements and reporting show about parties, scope, structure, term and revenue.



Rillor

Contents

01 Summary

02 Background

03 Method

04 Findings

05 Implications

06 Limits of this analysis

07 Sources

Summary

Data licensing deals for model training reach the public record mainly as announcements without prices, as licensors' revenue lines and as prices given in press reports.

- Most public data licensing deals carry no price. A January 2026 paper found 57 of 73 deals disclose no revenue [30], and on 8 November 2026 three of the 23 entries in Ithaka S+R's tracker of publisher deals carried a deal size [29].
- Reddit's registration statement of 22 February 2024 gives a contract value and a term: January 2024 arrangements worth $203.0 million in aggregate, over two to three years [1]. Neither it nor Reddit's later filings name a licensee [1, 2, 3, 4].
- Reddit's contracted revenue still to be recognized, mostly from long-term content licensing contracts, fell from $252.9 million at 31 December 2024 to $62.4 million at 30 September 2026 [2, 4]. Reddit's quarterly report of 30 October 2026 says its agreements with the two partners behind substantially all of its licensing contract value are approaching the end of their current terms [4].
- Licensors book licensing income in revenue lines that arrive in lumps. Shutterstock's data offering revenue fell 91% in the third quarter of 2026, and it says that revenue varies with metadata delivery timing [9].
- Prices tied to identified buyers come from reporting. Reuters put Reddit's deal with a search and model developer at about $60 million a year in February 2024, citing one unnamed source [7], and Engadget reported in May 2024 a Wall Street Journal estimate of $250 million over five years in cash and credits for News Corp's deal with a model developer [24].
- Deal structures vary: access over a term through an API [1], archive access with recurring fees for new content [18], cash with credits [24], equity taken as payment [8] and per-title fees for books [29].
- Payment to the people who made the content is rare in the record. Six of 73 public deals mention a revenue share with contributors [30], and 7 of 24 major deals compensate the original creator [31].
- Disclosure rules reach a deal's price only in narrow cases. SEC rules let a company leave ordinary-course contracts unfiled and redact confidential terms that are not material [32], and Form 8-K asks for the parties and material terms only of a material agreement outside the ordinary course [33]. The European Commission's template for public summaries of training content asks whether data was licensed, with no counterparty or price [34], and California's AB 2013 asks for the sources or owners of the datasets and whether they were licensed, with no price [35].

Background

What counts as a data licensing deal

In this report, a data licensing deal is an agreement under which a holder of content, such as posts, articles, books, images or their metadata, gives another company access to it for a fee or other consideration, for model training or for uses next to it such as retrieval, display and analysis. The parties use several names, among them "data licensing arrangements" at Reddit [1], "Data Access agreements" at Informa [18] and partnerships or agreements in the parties' announcements [5, 6, 22].

Why the record draws attention

Licensors' filings place these deals beside unlicensed use. Reddit's registration statement says "some companies have constructed very large commercial language models using Reddit data without entering into a license agreement with us" [1], and its September 2026 quarterly report lists the risk that third parties use Reddit content "obtained without our authorization" [4]. The New York Times Company recorded $13.3 million in 2025 and $10.8 million in 2024 of pre-tax litigation costs for lawsuits alleging "unlawful and unauthorized copying and use of the Company's journalism and other content", in a year when its licensing revenue rose [25].

The U.S. Copyright Office's report on generative training, released as a pre-publication version on 9 May 2025, records commenters' view of the market. By the end of 2023 they reported licensing in "music, vocal recordings, and news reports", and pointed to deals between a model developer and the Associated Press and Shutterstock, and Getty Images' collaborations with two technology companies [28]. The Office adds that public reporting since then "reflects AI licensing for images and audio-visual works, academic and nonfiction publishing, and news publishing", along with aggregators offering collective licensing [28]. Its status page still listed Part 3 as a pre-publication version on 8 November 2026, with a final version to follow and no substantive changes expected [36].

This report is for teams that license data in, for holders that license it out and for anyone reading licensors' filings. It is research, never legal advice, and it offers no view on any company's stock or prospects.

Method

What was examined. What public records disclose about data licensing deals: the parties, what is licensed, the structure, the term, the value and how licensors report the revenue, together with the rules that set what must be disclosed.

Sources. 36 sources in five evidence classes, each opened on 8 November 2026.

Class	What it is	Sources
Filing	A party's own SEC filing, or a results statement filed with a regulator	[1, 2, 3, 4, 8, 9, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 23, 25]
Announcement	A party's own press release or blog post	[5, 6, 22, 26]
Reporting	Reputable reporting that states its sourcing	[7, 10, 21, 24, 27]
Compilation	A tracker or research paper that compiles public deals	[29, 30, 31]
Rule	A law, rule, regulator form or official report	[28, 32, 33, 34, 35, 36]

How. SEC filings were read from EDGAR's own copies, and each figure is given as its source prints it, with its date. The Reuters report was read in its syndicated copy [7], The Wall Street Journal's estimate in Engadget's report of it [24] and the figure from Informa's May 2024 trading update in The Conversation's report [21]. Figures marked "derived" are arithmetic on a source's numbers. Each licensor is named as its source names it, each model developer by type only, and where a filing withholds a counterparty, no name from elsewhere is attached to its figures. Expected recognition of contracted revenue is given as filed, with its filing date.

Out of scope. What data costs per unit, the clauses of a license grant, the legal analysis of copyright and training, and litigation settlements.

Findings

1. Announcements state scope and rarely a price

The five announcements reviewed name their parties and say what is licensed. None states a price [5, 6, 13, 22, 26].

Deal, as announced	Date	What the announcement says	Term stated
Reddit and a search and model developer [5]	22 Feb 2024	The developer "now has access to Reddit's Data API", to "display, train on, and otherwise use" Reddit content	None
Reddit and a model developer [6]	16 May 2024	The developer accesses Reddit's Data API to bring Reddit content into its products, Reddit builds on the developer's models and the developer becomes a Reddit advertising partner	None
News Corp and a model developer [22]	22 May 2024	The developer "has permission to display content from News Corp mastheads in response to user questions and to enhance its products", from current and archived content	"Multi-year"
The Associated Press and a model developer [26]	13 Jul 2023	The developer licenses "part of AP's text archive", and AP draws on the developer's technology and product expertise	None
Dotdash Meredith and a model developer, in IAC's results release [13]	7 May 2024	Attributed content in the developer's chat responses, the content used to improve the developer's models, and work on Dotdash Meredith's D/Cipher advertising product	None

Two announcements draw limits. The search and model developer's post says the partnership does not change its use of "publicly available, crawlable content for indexing, training, or display" in its own products [5]. News Corp's release says "The partnership does not include access to content from any of News Corp's other businesses", and the titles it covers include The Wall Street Journal, the New York Post and The Times [22].

Compilations show the same at larger scale. On 8 November 2026 the 23 entries in Ithaka S+R's tracker were all scholarly or book publishers, and three carried a deal size: Wiley at $23m and Taylor & Francis at $10m and $75m [29]. Jia and co-authors analyzed 73 public data deals: 57 disclose no revenue, 16 disclose an amount, and the 15 with a quoted revenue volume sum to $677.3 million, on the floor of each range [30]. Their top deal types are news (26), images (16), academic content (15) and user-generated content (14), and more than 40% of known transactions had one of four technology companies as the buyer [30]. Oderinwale and Kazlauskas documented 78 deals from 2020 to 2025 and note that "Most data transactions remain private" [31].

2. Filings report licensing as revenue lines

In the filings reviewed, licensing shows up as revenue for a period inside a line of the licensor's results, and rarely with a counterparty.

Licensor and filing	Figure as filed	Period	What the line includes	Counterparty named
John Wiley & Sons, results releases [15, 16]	"AI licensing revenue" of $23 million and $40 million, then "AI revenue" of $49 million	Fiscal years ended April 2024, 2025 and 2026	Agreements with large technology companies, then "landmark partnerships" with two named companies and "a growing roster of corporate customers"	Two partners for fiscal 2026, with no amount
Informa, results statements [18, 19]	Over $75m expected, then "$75m+ of data access revenue which is non-recurring"	2024	Taylor & Francis data access agreements with two technology companies	No
Shutterstock, 10-K [8]	Data, Distribution, and Services of $137.3 million, $175.3 million and $203.3 million	2023, 2024 and 2025	Metadata licenses, Giphy, Shutterstock Studios and other services	No
Getty Images, results release [11]	Other revenue of $5.8 million, down 58.5%	Third quarter of 2025	Compared with a prior-year five-year deal with data licensing rights	No
IAC, 10-K [14]	Licensing and Other revenue up $32.9 million, or 28%	2025	Apple News+, content syndication partners and a full year of revenue from a model developer	A model developer, with no amount
The New York Times Company, 10-K [25]	Licensing revenues up $14.4 million	2025	"commercial agreements with third-party digital platforms"	No

Wiley described its buyers by type in fiscal 2025, reporting an "AI content licensing project this quarter with a third large tech company" [15], and named two partners in fiscal 2026 [16]. Its fiscal 2026 release says "lifetime AI revenue surpassed $110 million" [16].

Shutterstock's 2025 annual report describes "increased demand for access to our metadata for machine learning and generative artificial intelligence model training", from customers ranging "from large technology and media companies to smaller start-up organizations" [8]. It names no data customer [8]. Its Data,

Distribution, and Services line was 20.5% of its $989.9 million revenue in 2025 (derived) [8].

Getty Images' third-quarter 2025 release announces "Strategic Partnerships to Integrate Getty Images Content into Emerging AI Platforms" and names no partner [11]. IAC's 2025 annual report says its content licensing royalties include "utilization in large-language models", and attributes part of its growth to a full year of revenue from a model developer, in a partnership that "began in May 2024" [14]. The New York Times Company's licensing growth was "partially offset by the expiration of smaller license agreements", and the filing names no platform [25].

3. Reddit data licensing: one contract value and its run-off

Reddit's registration statement, filed 22 February 2024, is the one filing in this set that states both a contract value and a term [1]: "In January 2024, we entered into certain data licensing arrangements with an aggregate contract value of $203.0 million and terms ranging from two to three years. We expect a minimum of $66.4 million of revenue to be recognized during the year ending December 31, 2024 and the remaining thereafter." Reddit was "in the early stages of allowing third parties to license access to search, analyze, and display historical and real-time data from our platform", under the headings "Data API Access" and "Model Training" [1]. Delivery runs "through continuous access to our data API as well as quarterly transfers of Reddit data over the term of the arrangement" [1]. The fee is "generally a fixed fee or usage-based fee", recognized "generally ratably over the license period" [1].

Later filings, of 13 February 2025, 6 February 2026 and 30 October 2026, place licensing inside other revenue and report the contracted revenue still to come [2, 3, 4].

Date	Other revenue	Contracted revenue still to be recognized	Expected recognition, as filed
31 Dec 2023 [1]	$15.2 million for 2023	Immaterial	None given
31 Dec 2024 [2]	$114.7 million for 2024	$252.9 million	$114.6 million in 2025, $113.2 million in 2026, $25.1 million in 2027
31 Dec 2025 [3]	$140.0 million for 2025	$143.7 million	$118.9 million in 2026, $24.8 million in 2027
30 Sep 2026 [4]	$130.0 million for nine months	$62.4 million	$31.9 million in the rest of 2026, $30.5 million in 2027

The contracted figure covers contracts with an original expected duration of more than one year and "consists primarily of long-term content licensing contracts" [2, 3, 4]. Other revenue also holds, in the registration statement's words, "Reddit Premium subscriptions, and products within our user economy" [1], and in the 2025 report "products sold directly to users" [3]. It was 8.8% of total revenue in 2024 and 6.4% in 2025 (derived) [2, 3]. In the third quarter of 2026 it was $48.0 million, against $35.6 million a year earlier [4].

The 31 December 2024 figure is larger than the January 2024 contract value, and the 2024 annual report attributes the rise in other revenue to "content licensing agreements executed in 2024" [1, 2]. The 2025 report cites agreements "executed in 2024 and 2025" [3].

Since the 2024 annual report, Reddit has said that "substantially all of the contract value associated with our licensing revenue is derived from two of our partners" [2, 3]. The quarterly report filed on 30 October 2026 adds: "Our agreements with those partners are approaching the end of their respective current terms" [4]. Its forward-looking statements include "whether to enter new content licensing deals and whether to renew existing ones" [4]. The 2024 report notes terms "including API performance requirements" [2]. The filings describe the partners Reddit explores licensing with only by type, "ranging from marketing intelligence companies to enterprise-scale technology companies" [2], and in 2025 "specialized providers in the financial services industry" too [3]. None of the four names a licensee [1, 2, 3, 4].

4. Licensing revenue arrives in lumps

The filings show three ways of booking the money: over the term, on delivery and upfront.

Over the term. Reddit recognizes its licensing revenue "generally ratably over the license period" [1].

On delivery. Shutterstock expects its line to fluctuate "based on the timing of delivery of large metadata sets in our Data business" [8]. In the third quarter of 2026 the line fell 76% to $15.8 million from $65.7 million, "primarily driven by a decline in our data offering, which decreased by 91%" [9]. Over nine months it fell 46% to $93.0 million, with the data offering down 60% [9].

Upfront or accelerated. Getty Images' other revenue fell 58.5% in the third quarter of 2025 because the prior year held "a 5-year creative content deal with data licensing rights which was largely recognized upfront" [11]. Its October 2026 lender presentation says about $40 million was recognized "on an accelerated basis in Q4 2025", from "large multi-year AI licensing deals signed in 2025" [12].

Non-recurring years. Taylor & Francis grew total revenue 12.8% in 2024 against underlying growth of about 3.5%, and Informa called the $75m+ of data access revenue non-recurring [19]. In 2025 those revenues "were lower than the exceptional levels achieved in 2024", and no 2025 figure is given [20].

Prior-year comparisons. Wiley's Learning segment fell 5% in the fourth quarter of fiscal 2025 against "a $23 million AI licensing agreement in the prior year" [15], and 6% in the fourth quarter of fiscal 2026 on "lower AI licensing revenue in Academic and Professional" [16]. In the quarter ended 31 July 2026, "AI licensing revenue

totaled $14 million", against $29 million a year before, and Research Solutions fell 30% at constant currency on the comparison "($16 million vs. $4 million this quarter)" [17].

5. Reported prices and their sourcing

Deal	Figure	Reported by	Sourcing
Reddit and a search and model developer	About $60 million per year	Reuters, 22 Feb 2024 [7]	One of three unnamed people. Both companies declined to comment
News Corp and a model developer	$250 million over five years in cash and credits	The Wall Street Journal, via Engadget, 22 May 2024 [24]	The Journal's estimate
Shutterstock's licensing to "AI companies"	$104 million in 2023	Bloomberg, 4 Jun 2024 [10]	The chief executive, in an interview
Shutterstock's "anchor customers"	About $10 million in annual revenue each	Bloomberg [10]	An earlier Shutterstock statement, as Bloomberg reports it
Informa and a software company	More than £8 million for initial access, then unspecified recurring payments for three years	The Conversation, 23 Jul 2024 [21]	Not stated beyond Informa's May 2024 trading update on the deal
Informa and a software company	$10m	Ithaka S+R tracker [29]	Tracker entry
HarperCollins and an "undisclosed AI company"	$2,500 to each author who opts in	Publishers Weekly, 18 Nov 2024 [27]	An email from an author's agent, shared on Bluesky by the author
HarperCollins and a buyer named by "an anonymous source to Bloomberg"	$5,000 per book, split evenly, for three years	Ithaka S+R tracker [29]	Tracker entry

None of these figures appears in the filings reviewed. The Reuters sources "were not authorized to speak to media and declined to be identified" [7]. Engadget wrote that "Neither company disclosed the terms of the deal" [24]. News Corp owns The Wall Street Journal [22], and its results release of 4 November 2026 names

two model developers as partners with no licensing figure [23]. Bloomberg also reports a six-year deal between a model developer and Shutterstock in 2023 [10]. The agent's email, as Publishers Weekly quotes it, says "HarperCollins has been required to keep this company's identity confidential" [27]. Jia and co-authors' figures for Reddit and News Corp also come from press reports, CBS News for Reddit and the Associated Press for News Corp, not from the parties' filings [30].

6. How the deals are structured

Structure	As the source describes it	Source
Access over a term	Continuous Data API access plus quarterly transfers, two to three years, fixed or usage-based fee	Reddit [1]
Archive plus new content	Non-exclusive access to archive content, with "some additional recurring annual fees over the next three years for access to newly created knowledge"	Taylor & Francis [18]
Initial and recurring payments	An initial payment, then recurring payments for three years	Informa, as reported [21]
One-time payment	"One time payment, nonexclusive"	Wiley, as the tracker records it [29]
Display in the buyer's products	Content shown in response to user questions, and attributed content in chat responses	News Corp [22], Dotdash Meredith [13]
Cash plus credits	Cash and the buyer's API credits	News Corp, as reported [24, 31]
Equity as payment	Equity instruments received "in addition to cash for revenue contract consideration"	Shutterstock [8]
Per title, opt-in	A fee per book split between author and publisher, for three years	HarperCollins, as the tracker records it [29]
Retrieval and citation	Retrieval for another company's platform or search engine, and content that the tracker says "won't be used to train"	NEJM Group, Springer Nature, De Gruyter Brill, Johns Hopkins University Press and AAAS, as the tracker records them [29]

Structure	As the source describes it	Source
Wider commercial ties	The buyer as a Reddit advertising partner, and the buyer working on Dotdash Meredith's advertising product	[6, 13]

Shutterstock held $24.0 million of such equity instruments at 31 December 2024 and $30.5 million at 31 December 2025 and 30 September 2026, valued "based on issuers' recent market transactions at the time the instruments were received from the customers" [8, 9].

Beyond the terms in the table, the sources give a five-year deal in Getty Images' prior year [11], five years in the Journal's estimate for News Corp [24], six years in Bloomberg's report on Shutterstock [10] and two years in Jia and co-authors' description of AP's deal [30].

The compilations describe the market's patterns. Oderinwale and Kazlauskas find "five distinct pricing mechanisms (from per-unit licensing to commissioning)", and that in large enterprise deals "buyers pay fixed fees for time-limited access to curated corpora or feeds. Most licenses are non-exclusive" [31]. Jia and co-authors write that "Where money does change hands it is often a lump-sum buy-out" [30]. The Copyright Office notes that "Direct licensing is most common and most promising with respect to corporate entities with catalogs of high-quality and easily identifiable content" [28]. Ithaka S+R writes that "a standard set of terms or overall model from which to build these deals has yet to emerge" [29].

7. Who gets paid

The record shows little payment to the people who made the content. Jia and co-authors find "documented creator royalties rounding to zero", and that "Of the 73 transactions in Section A only six mention any revenue-share with contributors" [30]. Oderinwale and Kazlauskas find that "of 24 major deals in Table 4, only 7 compensate the original creator of the data" [31].

Where publishers pay authors, the terms differ. Informa described its partnerships as a source of "additional royalties for authors" [18], while The Conversation reported that the authors of the content Informa licensed "do not appear to have been asked or even informed about the deal" [21]. HarperCollins made its program opt-in, with $2,500 to each participating author as Publishers Weekly reports it [27] and an even split of $5,000 per book as the tracker records it [29]. For Johns Hopkins University Press, the tracker summarizes reporting that authors receive an "estimated less than $100 per title per license" [29].

8. What disclosure rules ask

SEC exhibits. Item 601(b)(10) of Regulation S-K requires filing "Every contract not made in the ordinary course of business that is material to the registrant" [32]. A contract that "ordinarily accompanies the kind of business conducted by the registrant" need not be filed unless it falls in listed categories, including "Any

contract upon which the registrant's business is substantially dependent" [32]. A filer "may redact specific provisions or terms" that it "customarily and actually treats" as confidential, when the omitted information is not material [32].

Form 8-K. Item 1.01 asks, for a material definitive agreement not made in the ordinary course, for "the date on which the agreement was entered into or amended, the identity of the parties to the agreement" and "a brief description of the terms and conditions" that are material [33]. Its first instruction ties the ordinary course test to Item 601(b)(10)(ii) [33]. In the filings reviewed, the deals appear in revenue discussions, risk factors, results releases and a lender presentation, and none as an Item 1.01 agreement [1, 2, 3, 4, 8, 9, 11, 12, 13, 14, 15, 16, 17, 23, 25].

EU training content summary. The Commission's template for general-purpose models explains: "limited disclosure of information is required for licensed data given that the rightsholders concerned are parties to the licensing agreements" [34]. Its Section 2.2.1 asks, yes or no, "Have you concluded transactional commercial licensing agreement(s) with rightsholder(s) or with their representatives?", and if yes, the modality of content: text, image, video, audio or other [34]. It asks for no counterparty, price or term [34].

California AB 2013. The law, chaptered on 28 September 2024, requires developers to post documentation on or before 1 January 2026, including a high-level summary with "The sources or owners of the datasets" and "Whether the datasets were purchased or licensed by the developer" [35]. It asks who the sources or owners are and whether data was licensed, with no price or terms [35].

9. Where the sources disagree

Deal counts. The compilations count 73 deals [30], 78 deals from 2020 to 2025 [31] and 23 publisher entries [29], with different scopes and dates. Jia and co-authors' Table 1 lists 16 deals with an amount disclosed, and their Figure 3 sums 15 with a quoted revenue volume [30].

Informa's initial payment. The Conversation reports more than £8 million for initial access [21], and the tracker $10m [29]. The currencies differ, and this report converts neither.

Shutterstock. The chief executive's $104 million for 2023 covers licensing to "AI companies" [10]. The filed Data, Distribution, and Services line for 2023 is $137.3 million and includes Giphy, Studios and other services [8].

AP's deal. AP's release states no term or fee [26]. Jia and co-authors describe it as a "two-year, flat-fee deal (amount undisclosed)" [30].

Vocabulary. Reddit's registration statement says "data licensing" [1], and its later filings "content licensing" [2, 3, 4]. Wiley says "AI licensing revenue" for fiscal 2024 and 2025 [15] and "AI revenue" for fiscal 2026 [16]. Its three annual figures sum to $112 million (derived), in line with "lifetime AI revenue surpassed $110 million" [15, 16].

Implications

For teams licensing data in. The public record rarely shows the permitted uses, the term or onward sharing, so the agreement is where those are settled. Reported prices rest on an unnamed source, an interview or a newspaper's estimate [7, 10, 24], and filed figures are revenue for a period, which says little about any one deal [2, 8, 15]. Holders license for training, retrieval or display [5, 22, 29], so the use a team needs belongs in the grant. Rillor finds the publishers and data holders that hold the data a team needs, agrees the use, the term and onward sharing in writing before delivery, and delivers every record with the license it came under.

For data holders licensing out. Peers book the money over the term, on delivery or upfront, and the choice shapes reported results [1, 9, 11, 12]. Reddit's filings show what concentration in two partners looks like near the end of their terms [4]. Author consent and creator payment draw attention in the reporting and the research [21, 30, 31].

For readers of licensors' filings. Read the line's definition first: Reddit's other revenue includes products sold to users [3], and Shutterstock's line includes Giphy and Studios [8]. Contracted revenue still to be recognized shows how much licensing revenue is already under contract, and when it is expected [2, 3, 4]. A company's own figure and a reported figure can measure different things [8, 10].

Limits of this analysis

- Only public records were reviewed. Private terms are unknown, and Jia and co-authors note their deal set "likely undercounts and is skewed toward larger, English-language, and U.S.-centric agreements" [30].
- No party confirms the reported figures in its filings reviewed here.
- The filings reviewed are a set chosen for each licensor, and the statement that no deal appears under Item 1.01 covers those filings only. The 8-K that carries Getty Images' lender presentation reports an Item 1.01 for a restructuring support agreement, not a licensing deal [12].
- Shutterstock's line holds more than data licensing, and Reddit's other revenue more than licensing [3, 8].
- Informa gives no 2025 figure for Taylor & Francis data access revenue [20].
- One paper's table gives an annual figure for the Dotdash Meredith agreement that no IAC filing reviewed contains, so it is left out [14, 31].
- Company revenue forecasts, including those in Bloomberg's Shutterstock report and Getty Images' lender presentation, are left out [10, 12].
- Deals reported for other news publishers, music labels and platforms are left out where no primary source or reputable report was opened.
- The Copyright Office's Part 3 is a pre-publication version [28, 36].

Sources

1. Reddit, Inc. Registration Statement on Form S-1. U.S. Securities and Exchange Commission, EDGAR, filed 22 February 2024, viewed 8 November 2026. https://www.sec.gov/Archives/edgar/data/1713445/000162828024006294/reddits-1q423.htm

2. Reddit, Inc. Annual Report on Form 10-K for 2024. U.S. Securities and Exchange Commission, EDGAR, filed 13 February 2025, viewed 8 November 2026. https://www.sec.gov/Archives/edgar/data/1713445/000171344525000018/rddt-20241231.htm

3. Reddit, Inc. Annual Report on Form 10-K for 2025. U.S. Securities and Exchange Commission, EDGAR, filed 6 February 2026, viewed 8 November 2026. https://www.sec.gov/Archives/edgar/data/1713445/000171344526000022/rddt-20251231.htm

4. Reddit, Inc. Quarterly Report on Form 10-Q for the quarter ended 30 September 2026. U.S. Securities and Exchange Commission, EDGAR, filed 30 October 2026, viewed 8 November 2026. https://www.sec.gov/Archives/edgar/data/1713445/000171344526000116/rddt-20260930.htm

5. Google expands partnership with Reddit. Google, The Keyword, 22 February 2024, viewed 8 November 2026. https://blog.google/inside-google/company-announcements/expanded-reddit-partnership/

6. Reddit and OpenAI Build Partnership. Reddit, Inc., 16 May 2024, viewed 8 November 2026. https://www.redditinc.com/blog/reddit-and-oai-partner

7. Exclusive-Reddit in AI content licensing deal with Google, sources say. Anna Tong, Echo Wang and Martin Coulter, Reuters, via Yahoo, 22 February 2024, viewed 8 November 2026. https://tech.yahoo.com/ai/articles/exclusive-reddit-ai-content-licensing-041012503.html

8. Shutterstock, Inc. Annual Report on Form 10-K for 2025. U.S. Securities and Exchange Commission, EDGAR, filed 17 February 2026, viewed 8 November 2026. https://www.sec.gov/Archives/edgar/data/1549346/000154934626000008/sstk-20251231.htm

9. Shutterstock, Inc. Quarterly Report on Form 10-Q for the quarter ended 30 September 2026. U.S. Securities and Exchange Commission, EDGAR, filed 5 November 2026, viewed 8 November 2026. https://www.sec.gov/Archives/edgar/data/1549346/000154934626000034/sstk-20260930.htm

10. Shutterstock's AI-Licensing Business Generated $104 Million Last Year. Brody Ford, Bloomberg, via Yahoo Finance, 4 June 2024, viewed 8 November 2026. https://finance.yahoo.com/news/shutterstock-ai-licensing-business-generated-120000890.html

11. Getty Images Reports Third Quarter 2025 Results (Exhibit 99.1 to Form 8-K). Getty Images Holdings, Inc., SEC EDGAR, 10 November 2025, viewed 8 November 2026. https://www.sec.gov/Archives/edgar/data/1898496/000162828025051079/gety-20250930xex991.htm

12. Getty Images Lender Presentation, October 2026 (Exhibit 99.2 to Form 8-K). Getty Images Holdings, Inc., SEC EDGAR, filed 13 October 2026, viewed 8 November 2026.

https://www.sec.gov/Archives/edgar/data/1898496/000121390026108700/ea030834301ex99-2.htm

13. IAC Reports Q1 2024 (Exhibit 99.1 to Form 8-K). IAC Inc., SEC EDGAR, 7 May 2024, viewed 8 November 2026. https://www.sec.gov/Archives/edgar/data/1800227/000180022724000022/ex_991q12024iac-pressrelea.htm

14. IAC Inc. Annual Report on Form 10-K for 2025. U.S. Securities and Exchange Commission, EDGAR, filed 20 February 2026, viewed 8 November 2026. https://www.sec.gov/Archives/edgar/data/1800227/000162828026009997/iaci-20251231.htm

15. Research Growth, AI Licensing, and Cost Reduction Drive Wiley's Fiscal 2025 Results (Exhibit 99.1 to Form 8-K). John Wiley & Sons, Inc., SEC EDGAR, 17 June 2025, viewed 8 November 2026. https://www.sec.gov/Archives/edgar/data/107140/000010714025000071/wly-2025430xex991.htm

16. Research and AI Momentum, Record Margins, and Cash Flow Growth Highlight Wiley's Fourth Quarter and Fiscal 2026 Results (Exhibit 99.1 to Form 8-K). John Wiley & Sons, Inc., SEC EDGAR, 16 June 2026, viewed 8 November 2026. https://www.sec.gov/Archives/edgar/data/107140/000162828026043447/ex991fy26q4.htm

17. Wiley Reports First Quarter 2027 Results (Exhibit 99.1 to Form 8-K). John Wiley & Sons, Inc., SEC EDGAR, 3 September 2026, viewed 8 November 2026. https://www.sec.gov/Archives/edgar/data/107140/000162828026060252/ex991earningsrelease.htm

18. 2024 Half-Year Results. Informa PLC, 24 July 2024, viewed 8 November 2026. https://www.informa.com/globalassets/documents/investor-relations/2024/informa-2024-half-year-results.pdf

19. 2024 Preliminary Full-Year Results. Informa PLC, 6 March 2025, viewed 8 November 2026. https://www.informa.com/globalassets/documents/investor-relations/2025/informa-2024-full-year-results.pdf

20. 2025 Full-Year Results. Informa PLC, 12 March 2026, viewed 8 November 2026. https://www.informa.com/globalassets/documents/investor-relations/2026/informa-2025-full-year-results-statement.pdf

21. An academic publisher has struck an AI data deal with Microsoft, without their authors' knowledge. Wellett Potter, The Conversation, 23 July 2024, viewed 8 November 2026. https://theconversation.com/an-academic-publisher-has-struck-an-ai-data-deal-with-microsoft-without-their-authors-knowledge-235203

22. News Corp and OpenAI Sign Landmark Multi-Year Global Partnership. News Corp, 22 May 2024, viewed 8 November 2026. https://newscorp.com/2024/05/22/news-corp-and-openai-sign-landmark-multi-year-global-partnership/

23. News Corporation Reports First Quarter Results for Fiscal 2027 (Exhibit 99.1 to Form 8-K). News Corp, SEC EDGAR, 4 November 2026, viewed 8 November 2026. https://www.sec.gov/Archives/edgar/data/1564708/000156470826000248/release-q1fy2027.htm

24. OpenAI will reportedly pay $250 million to put News Corp's journalism in ChatGPT. Pranav Dixit, Engadget, 22 May 2024, viewed 8 November 2026. https://www.engadget.com/openai-will-reportedly-pay-250-million-to-put-news-corps-journalism-in-chatgpt-214615249.html

25. The New York Times Company Annual Report on Form 10-K for 2025. U.S. Securities and Exchange Commission, EDGAR, filed 27 February 2026, viewed 8 November 2026. https://www.sec.gov/Archives/edgar/data/71691/000007169126000011/nyt-20251231.htm

26. AP, Open AI agree to share select news content and technology in new collaboration. The Associated Press, 13 July 2023, viewed 8 November 2026. https://www.ap.org/media-center/press-releases/2023/ap-open-ai-agree-to-share-select-news-content-and-technology-in-new-collaboration/

27. HarperCollins Inks AI Licensing Deal for Nonfiction Books. Publishers Weekly, 18 November 2024, viewed 8 November 2026. https://www.publishersweekly.com/pw/newsbrief/index.html?record=5076

28. Copyright and Artificial Intelligence, Part 3: Generative AI Training (Pre-Publication Version). U.S. Copyright Office, 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

29. Generative AI Licensing Agreement Tracker. Ithaka S+R, published 12 October 2024, sheet last updated 7 August 2026, viewed 8 November 2026. https://sr.ithaka.org/our-work/generative-ai-licensing-agreement-tracker/

30. A Sustainable AI Economy Needs Data Deals That Work for Generators. Ruoxi Jia, Luis Oala, Wenjie Xiong, Suqin Ge, Jiachen T. Wang, Feiyang Kang and Dawn Song, arXiv 2601.09966, 15 January 2026, viewed 8 November 2026. https://arxiv.org/abs/2601.09966

31. The Economics of AI Training Data: A Research Agenda. Hamidah Oderinwale and Anna Kazlauskas, arXiv 2510.24990, version 3, 2 August 2026, viewed 8 November 2026. https://arxiv.org/abs/2510.24990

32. Regulation S-K, Item 601(b)(10), Material contracts (17 CFR 229.601). Electronic Code of Federal Regulations, up to date as of 5 November 2026 (Title 17 last amended 19 October 2026), viewed 8 November 2026. https://www.ecfr.gov/current/title-17/chapter-II/part-229/subpart-229.600/section-229.601

33. Form 8-K, Item 1.01 Entry into a Material Definitive Agreement. U.S. Securities and Exchange Commission, Form SEC 873, revision 02-25 (February 2025), viewed 8 November 2026. https://www.sec.gov/files/form8-k.pdf

34. Explanatory Notice and Template for the Public Summary of Training Content for general-purpose AI models, C(2025) 8311 final. European Commission, 5 December 2025, viewed 8 November 2026. https://ec.europa.eu/newsroom/dae/redirection/document/118480

35. AB 2013, Generative artificial intelligence: training data transparency, Chapter 817, Statutes of 2024. California Legislature, chaptered 28 September 2024, text as published by CalMatters Digital Democracy, viewed 8 November 2026. https://calmatters.digitaldemocracy.org/bills/ca_202320240ab2013

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

About this report

Published by Rillor on 9 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. Training data licensing deals: what public filings and reports disclose. 9 November 2026. https://rillor.com/insights/training-data-licensing-deals

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