

Licensing company records for AI training: the rights questions buyers ask, and what public sources show

Yuma Heymans

AIData.Broker, <https://aidata.broker> | hello@aidata.broker | 18 October 2026

Abstract

Developers of AI models now license the working records of ordinary companies: support tickets, email, chat, CRM activity, documents and code. Whether a company may license such records is not one question but several, and buyers' counsel ask them in a recognisable order. This article compiles what public sources show on those questions for the United States, the United Kingdom and Canada, as of 18 October 2026. It draws on four open datasets (the terms of 18 business software systems, a 27-item rights checklist, 15 licence terms with their public evidence, and an 83-entry register of lawsuits and rulings), each row tied to a dated source and its own words, and on the statutes, regulator guidance and court filings those rows cite. Five findings stand out. Software vendors generally confirm that customers own their content, but 6 of the 18 systems read bar training AI on data pulled through their APIs. Privacy law in all three countries turns on notice, consent and de-identification, with markedly different standards. Under US copyright law employee communications are generally owned by the employer, yet they carry other people's rights. A court-supervised bankruptcy sale of an airline's records shows how de-identification, carve-outs and objections are handled in practice. And published licence terms diverge widely on purpose, model survival, exclusivity, deletion and indemnity. We close with the questions the sources leave open.

Keywords: AI training data; data licensing; privacy law; de-identification; CCPA; UK GDPR; PIPEDA; employee communications; bankruptcy section 363; terms of service

1. Introduction

Public discussion of AI training data has centred on published works: books, news and images that crawlers can reach. A different market has grown beside it. AI developers and data vendors ask companies to license their internal records, which no crawler can reach, so that models can learn how real work is done. In August 2026, for example, Google agreed to buy de-identified internal business data from the bankrupt Spirit Airlines for \$10 million, in a sale agreement filed with the bankruptcy court [71], and said it planned to use the data "for product development and training its AI models" [77].

For a company considering such a licence, the first question sounds simple: may we? Public sources show that the answer depends on several separate bodies of law and contract, each with its own enforcer. A law firm advising sellers put the sequence as a checklist in October 2026: review customer, partner, vendor and data licence agreements, "identify data held as a service provider or processor, and review platform terms for the systems the data will be extracted from" [30].

This article asks what public sources show about those questions, focusing on the United States, the United Kingdom and Canada. It is descriptive. It reports what statutes, regulators, courts, vendors and buyers have written, and where they are silent. It does not tell any company what it may do, and it is not legal advice.

Section 2 describes the data and method. Section 3 reports findings in six areas: software terms, personal data, employee communications, bankruptcy sales, licence terms, and copyright and recent litigation. Section 4 lists open questions, Section 5 the limitations, and Section 6 concludes.

2. Data and method

The article rests on four open datasets compiled by the author's organisation and published under CC BY 4.0, each row of which records a source URL, the date it was read and a verbatim quotation:

- **Software terms** [1]: for 18 business systems, whether the holder can export its own data, what the vendor's API or developer terms say about training AI on data obtained through them, and a reading of the rights risk of the content. Reviewed 11 October 2026.
- **Rights checklist** [2]: 27 questions a buyer's counsel asks before licensing data, each with a good answer, a red flag, the kinds of data it applies to and a severity (blocks a sale, fixable, or to be disclosed). Reviewed 11 October 2026.
- **Licence terms** [3]: 15 terms a training-data licence turns on, with what buyers ask, the public evidence and any gap in it. Reviewed 11 October 2026.
- **Lawsuits and rulings** [4]: 83 lawsuits, rulings, settlements and regulatory actions over data used to train AI. Reviewed 17 October 2026.

These datasets are summarised on two web pages [5, 6] and applied in six published guides on customer data [7], employee communications [8], bankruptcy sales [9], licence agreements [10], fair use [11] and European data protection [12], and in country pages for the United Kingdom [14] and Canada [13]. Throughout this article, each fact is cited to the original source the dataset row names, with the dataset or guide as the compilation.

The method of collection was reading: statutes from official or widely used legal publishers, regulator guidance, court dockets (mostly through CourtListener), vendors' own terms pages, buyers' own published terms, and press reports where no primary record was available. Press reports are marked as such. The severity of each checklist item, and any statement of "where holders stand" on a licence term, is the compiler's reading, not a finding of any court or regulator.

The method has limits that matter for interpretation. It is a compilation of what is public, so it over-represents disputes that reached court, terms that vendors publish, and buyers that publish their terms. The 18 software systems are a selection of common business tools, not a census. Terms change, and each is reported as read on the date given. Section 5 returns to these limits.

3. Findings

3.1 The holder owns the content; the vendor controls the route out

The vendors whose terms were read generally confirm that the customer owns what its workspace holds. Microsoft's data protection addendum states: "As between the parties, Customer retains all right, title and interest in and to Customer Data" [24]. Atlassian's customer agreement: "As between the parties, Customer owns all intellectual property and other rights in Customer Data" [25]. GitHub's terms: "You own Your Content" [27]. Salesforce commits to make customer data available for export or download on request "within 30 days after the effective date of termination or expiration" of the agreement [28].

Ownership as between customer and vendor is not the constraint the compilation finds most often. The constraint is how the data leaves the system. Of the 18 systems, 13 let the holder export its own data without a limit that matters for a sale, 3 limit the export, and for 2 the terms were unclear. Separately, 6 systems bar using data obtained through their APIs to train AI models, and 2 more bar copying or reselling API data [1].

System	Holder's own export	API or developer terms on AI training
Slack	Limited	Bars training AI on API data
Microsoft 365, Teams, Exchange, SharePoint	Permitted	Bars copying or reselling API data
Google Workspace	Permitted	Bars training AI on API data
Atlassian (Jira, Confluence)	Permitted	Not established (developer terms not retrievable)
GitHub	Permitted	No AI clause found
GitLab	Permitted	No AI clause found
Notion	Permitted	Not established
Salesforce	Permitted	No AI clause found
HubSpot	Permitted	Bars training AI on API data
Zendesk	Limited	Bars training AI on API data
ServiceNow	Permitted	No AI clause found
Intercom	Unclear	Bars copying API data
Zoom	Permitted	Bars training AI on API data
QuickBooks Online (Intuit)	Permitted	Not established
Xero	Limited	Bars training AI on API data
Oracle NetSuite	Unclear	Not established
Figma	Permitted	No AI clause found
Linear	Permitted	No AI clause found

Table 1. Export and API terms of 18 business systems, as read in October 2026 [1].

The API restrictions are explicit. Slack's API terms, effective 10 October 2025, say an app may not "(A) use API Data to train a large language model;" or "(B) bulk export Slack message and file data except where expressly allowed by an additional agreement" [15]. Google's Workspace developer policy prohibits "Transferring, selling, or using user data to create, train, or improve a machine learning or artificial intelligence model beyond that specific user's personalized model", and "Transferring or selling user data to third parties like advertising platforms, data brokers, or any information resellers" [17]. Zendesk's developer terms bar using API data for any purpose other than the application's functionality, "including creating, training, or improving any general AI model, such as, but not limited to, a large language model", and "sanitizing API Data for such purposes" [18]. Zoom forbids using customer content obtained through its APIs "to train, develop, evaluate, test, improve, or modify artificial intelligence or machine learning models" without its written permission [19]. Xero: "API Data cannot be used to train, fine tune, adapt, or enhance any AI Models" [20]. HubSpot's May 2026 developer terms restrict "using data accessed through the APIs to train, fine-tune, or improve AI or machine learning models", with a carve-out for "legitimate single-customer use cases" [21]. Microsoft's API terms forbid users to "Redistribute or resell, or sublicense access to, the Microsoft APIs, any data obtained using the Microsoft APIs" [23].

These clauses govern apps and connectors that pull data through an API. They do not, on their face, address an export run by the customer's own administrator with the vendor's built-in tool. The compilation's practical reading is therefore that the route out matters: a native export by the holder's own administrator is the customer's data leaving the customer's account, while a buyer's connector is an API client bound by developer terms [1, 2]. Native exports have conditions of their own. Slack makes full exports of private channels and direct messages available only on higher plans and by application: "Workspace Owners and Org Owners must apply to use these

export types" [16]. Zendesk does not enable data exports by default (as read from a search excerpt of its help page) [1]. A law firm writing when Slack changed its terms noted that the new API terms prohibit bulk export, persistent copies and use of the data in large language models [29].

A second pattern concerns vendors' own use of customer content. Atlassian states that metadata contribution is "always on" for organisations whose highest plan is Free, Standard or Premium, and that every customer may opt out of contributing in-app data [26]. HubSpot's customer terms say: "We may also use Customer Data to train our AI models in compliance with our obligations under the Agreement" [22]. GitHub's terms, effective 27 April 2026, grant it a licence to use Copilot inputs and outputs "to develop, train and improve artificial intelligence" unless the user opts out [27]; the compiled row notes that this does not apply where a GitHub Customer Agreement governs [1]. None of this prevents a customer from licensing its own content. The checklist treats it as a disclosure: a buyer paying for exclusivity will want to know whether a platform vendor already holds a model trained on the same material [2].

3.2 Personal data: notice, consent and de-identification

Company records are full of people: customers, prospects, employees, candidates, vendors' staff. All three jurisdictions regulate the use of their information for a new purpose, but they start from different premises.

3.2.1 United States

The United States has no general federal privacy statute. The compiled guide describes the question as five layers: the company's own promises (enforced by the FTC), state privacy laws, sector laws, customer contracts and data broker registration [7].

Prior promises. FTC staff wrote in February 2024 that "A business that collects user data based on one set of privacy commitments cannot then unilaterally renege", and that it may be unfair or deceptive to adopt more permissive data practices "through a surreptitious, retroactive amendment to its terms of service or privacy policy" [31]. The FTC's remedies explain why buyers ask about this first. Its 2021 Everalbum order required deletion of "the models and algorithms it developed by using the photos and videos uploaded by its users" [32]. Its 2022 order against WW International and Kurbo required the companies to "destroy any algorithms derived from the data, and pay a \$1.5 million penalty" [33]. A model trained on records the holder had no right to license is the asset at risk.

State law. California's Consumer Privacy Act defines a sale as making personal information available to a third party "for monetary or other valuable consideration" [34, s. 1798.140(ad)]. Its exemption for employee and applicant data became "inoperative on January 1, 2023" [34, s. 1798.145(m)], and the state's privacy agency confirms that consumers include "California residents that are employees or job applicants, and contacts for business customers, vendors, or independent contractors" [35]. Sensitive personal information includes "The contents of a consumer's mail, email, and text messages unless the business is the intended recipient of the communication" [34, s. 1798.140(ae)].

The compiled guide lists 19 comprehensive state privacy laws in force on 17 October 2026, and four more enacted but not yet in force [7]. Two features matter most for company records. First, most states outside California exclude people acting in a work capacity; Texas, for example, defines a consumer as acting "only in an individual or household context" and excludes "an individual acting in a commercial or employment context" [36]. Second, the de-identification test is close to uniform. California's requires that the holder "(1) Takes reasonable measures to ensure that the information cannot be associated with a consumer or household. (2) Publicly commits to maintain and use the information in deidentified form and not to attempt to reidentify the

information ... (3) Contractually obligates any recipients of the information to comply" [34, s. 1798.140(m)]. Virginia and Texas use the same three conditions [36, 37], and the compiled table finds them in every listed state except Iowa [7]. The second and third conditions are legal, not technical: without a public commitment and a contractual flow-down, records are not de-identified in this statutory sense, however well they were scrubbed.

Customer content. For a software or services company, much of what sits in its systems is held for its customers. California's definition of a service provider requires a written contract that prohibits the provider from "(A) Selling or sharing the personal information" and from using it for other purposes [34, s. 1798.140(ag)]. The checklist treats customer content held under such contracts as a blocker unless excluded, expressly permitted by each customer, or de-identified where the contract does not forbid that [2].

Sector and national-security rules. Sector laws add their own standards; Illinois' biometric statute, for example, provides that no private entity may "sell, lease, trade, or otherwise profit from" a person's biometric identifier [40]. The Justice Department's data security rule restricts transfers of bulk US sensitive personal data to buyers linked to countries of concern, with thresholds such as biometric identifiers on more than 1,000 US persons and personal health or financial data on more than 10,000 [38], and its guidance states that bulk data counts "regardless of whether the data is anonymized, pseudonymized, de-identified, or encrypted" [39]. De-identification therefore does not take records outside that rule.

3.2.2 United Kingdom

The UK GDPR governs any record that identifies a living person. Amendments made by the Data (Use and Access) Act 2025 to purpose limitation, research and transfers came into force on 5 February 2026 [43].

A licence of identifiable records needs a lawful basis, and the compiled pages find legitimate interests the realistic one [12, 14]. The Act's new "recognised legitimate interests" cover matters such as public-task requests, national security and emergencies [12, 41], not commercial licensing. The ICO's guidance accepts commercial interests as legitimate interests but adds: "If you can reasonably achieve the same result in another less intrusive way, legitimate interests doesn't apply" [44]. A new purpose must also pass the compatibility test now in Article 8A, which weighs the link between purposes, the context and relationship, the nature of the data, the possible consequences and safeguards such as pseudonymisation [42]. The ICO's purpose-limitation guidance, updated 23 March 2026, says a new purpose that "is very different from the original purpose; would be unexpected; or would have an unjustified impact" is likely incompatible, and that "you're likely to need to ask for specific consent" for it [45].

On de-identification the ICO takes a relative view: the same information may be "personal data in your hands" but "anonymous in the hands of other parties" who lack the means to identify people [46], and it suggests a "motivated intruder" test as "a good starting point" [46]. The anonymising step is itself processing that needs a lawful basis [47]. For transfers, personal data can go to a US buyer that the Data Privacy Framework list shows as "participating in the UK Extension" [48]; otherwise the ICO points to its transfer agreement or addendum with a transfer risk assessment [49].

One ICO position bears directly on licensing. In its 2024 response on generative AI it said it expects "controllers who develop generative AI to evidence why other available methods for data collection are not suitable" before relying on web scraping [50]. A licence from a holder is such an alternative.

3.2.3 Canada

Canada's federal private-sector law, PIPEDA, is consent-based and has no legitimate-interests ground. Its Schedule 1 states: "When personal information that has been collected is to be used for a purpose not previously identified, the new purpose shall be identified prior to use. Unless the new purpose is required by law, the consent of the individual is required" [52, cl. 4.2.4]. The business-transaction exception does not help a data sale: it does not apply "to a business transaction of which the primary purpose or result is the purchase, sale or other acquisition or disposition, or lease, of personal information" [52, s. 7.2(4)]. The research exception requires, among other things, that it be "impracticable to obtain consent" and that the Commissioner be informed before disclosure [52, s. 7(3)(f)].

That leaves anonymisation as the main route for archives collected without consent to licensing, and the federal standard is demanding. In March 2026 the Privacy Commissioner found that information is anonymised only if "there is no serious possibility that the information retained may be reidentified, either alone or in combination with other available information", and that "stripping accounts of names, phone numbers and email addresses alone is insufficient" [53]. Quebec adds its own layer: use for another purpose needs consent unless an exception applies [55, s. 12]; communication to a third person needs consent, expressly for sensitive information [55, s. 13]; information is anonymised only if it "irreversibly no longer allows the person to be identified directly or indirectly" [55, s. 23], under a regulation that requires supervision by a qualified person [56]; and before personal information leaves Quebec, the enterprise "must conduct a privacy impact assessment" and sign a written agreement [55, s. 17].

Canada does not bar transfers to the United States. The Commissioner's guidance says PIPEDA "does not prohibit organizations in Canada from transferring personal information to an organization in another jurisdiction", but that "The transferring organization is accountable for the information in the hands of the organization to which it has been transferred" [54]. A bill that would change the federal framework, Bill C-36, was at second reading when read and is not law [57].

Question	United States (California as the binding state for most company records)	United Kingdom	Canada (federal, with Quebec)
Basis for a new use of identifiable records	Notice and opt-out for a sale; opt-in for sensitive data in most states [34, 7]	Legitimate interests with a written test; specific consent likely for an unexpected purpose [44, 45]	Consent for a purpose not previously identified; no legitimate-interests ground [52]
Work contacts and staff	Covered in California since 2023; mostly excluded elsewhere [34, 35, 36]	Covered as personal data [12]	Varies by employer and province; not compiled here
De-identification standard	Reasonable measures, public commitment, contractual flow-down [34]	Relative identifiability; motivated intruder test [46]	No serious possibility of re-identification [53]; irreversible in Quebec [55]
Transfer to a US buyer	Federal bulk-data rule screens foreign buyers [38, 39]	UK Extension listing, or transfer agreement with risk assessment [48, 49]	Permitted; sender stays accountable; Quebec impact assessment [54, 55]

Table 2. Personal-data questions for a licence of company records, by jurisdiction, as read in October 2026.

3.3 Employee communications

Internal email, chat and documents raise questions that privacy law alone does not answer.

Ownership. Under US copyright law a work "prepared by an employee within the scope of his or her employment" is a work made for hire [59], and "the employer ... is considered the author" and owns the

copyright unless a signed instrument says otherwise [60]. Contractors' work is not covered unless assigned, which is why the checklist asks whether every contributor signed an assignment [2].

Access. The federal Stored Communications Act's access offence does not apply to conduct authorised "by the person or entity providing a wire or electronic communications service" [61], and its bar on divulging contents applies to providers of service "to the public" [62]. The compiled guide therefore reads a company's own export from its own systems as authorised access, while a hosting vendor may divulge contents only within the Act's exceptions [8]. The compiled guide reads the Wiretap Act's provider exception and one-party consent rule as narrower than they first appear, since an employee's consent speaks only for that employee [8, 63].

Notice. At least four states require notice before employers monitor electronic communications: New York [64], Connecticut (as amended with effect from 1 October 2026) [8, 65], Delaware [66] and Maine [67]. These laws govern notice, not licensing, and a monitoring notice written for security does not tell staff their messages may be licensed [8].

Recordings, organising and privilege. California makes it an offence to record a confidential communication "without the consent of all parties" [68]. Staff discussion of pay and conditions is protected activity under section 7 of the National Labor Relations Act [69]; the checklist recommends excluding union and HR channels but records that no source was found applying section 7 directly to the sale of employee communications [2]. Privileged legal advice is the content that removing names cannot fix: the federal rule protecting inadvertent disclosure applies only when the disclosure is "made in a federal proceeding or to a federal office or agency" [70]. A law firm advising sellers recommends excluding "privileged material, HR and personnel files, payroll and tax records, health and benefits information, union-related communications, customer and partner confidential information" entirely rather than relying on de-identification [30].

Other companies' information. The Spirit Airlines sale shows the problem at scale. The flight attendants' union stated that the asset schedule included approximately 100 million emails across 80,000 email accounts, together with OneDrive, SharePoint and Teams items (as reported in a summary of the docket) [74]. After the auction, an aircraft-engine supplier filed a limited objection with a cross-motion to enforce confidentiality agreements [73]. The court-appointed consumer privacy ombudsman's report states that the ombudsman "did not investigate or reach any conclusion about whether the data offered for sale or the deidentification process to be followed presents any risks to affected employees or third parties" [72]. A letter signed by 119 members of Congress asked the chief executives of Spirit and Google to halt the transfer of information related to former Spirit employees until protections are in place, according to a press release from Representative Steven Horsford's office as reported by PYMNTS [75].

In the United Kingdom and Canada, staff messages are personal data like any other, and the questions in Section 3.2 apply to them. The checklist applies extra caution to employee consent in the UK, citing European guidance that consent is unlikely to be freely given in employment [2]; for Canada, the compilation does not cover provincial employee-privacy statutes (Section 4).

3.4 Court-supervised sales of records in bankruptcy

US bankruptcy law has a specific rule for personal information. If the debtor disclosed "a policy prohibiting the transfer of personally identifiable information about individuals to persons that are not affiliated with the debtor", the trustee may sell or lease such information only consistently with that policy or after a consumer privacy ombudsman is appointed and the court approves [78]. The ombudsman is appointed "not later than 7 days before the commencement of the hearing" [79]. The protected category is narrow: information provided "in connection with obtaining a product or a service from the debtor primarily for personal, family, or household

purposes" [80], which, on the compiled guide's reading, leaves employee and business-contact data outside it [9].

Precedents. The FTC's 2000 settlement with the failed retailer Toysmart allowed a sale of its customer list only "as a package which includes the entire Web site, and only to a 'Qualified Buyer'" in a related market [81]. In RadioShack's 2015 case, the FTC told the ombudsman its concerns "would be greatly diminished" if the information was not sold as a standalone asset, the buyer was in substantially the same lines of business and agreed to be bound by the privacy policies [82]; a coalition of 38 states negotiated an agreement under which "the overwhelming bulk" of the consumer data would be destroyed [83]. In 23andMe's 2025 case, the ombudsman recommended separate, affirmative consent as "the optimal solution from a consumer privacy perspective" [85]; the court declined to impose it, holding that "the effect of the proposed transaction on consumer privacy is not so significant as to justify court intervention", and approved a sale to TTAM Research Institute after a post-auction proposal of \$305 million topped Regeneron's \$256 million [84]. The court also held that section 363(b)(1) does not let a debtor ignore state data-transfer restrictions: "The case law resolves this gap in favor of compliance" [84].

Spirit Airlines (2026). The Spirit case is the first in the compilation in which the asset was a company's working records sold for AI training rather than a customer list.

Element	What the record shows	Source
Price and form	\$10,000,000, "with delivery of the Assets to be made in deidentified form"	Sale agreement, Dkt. 1463 [71]
Auction	Google opened at \$5 million; Mercor offered \$5.2 million with Google's form or \$7 million if it could de-identify with its own tools; the debtors treated \$5.2 million as the higher bid	Docket summary [74]
Alternate bid	Mercor at \$7,500,000	Dkt. 1463 [71]
Late proposal	micro1 at \$12.5 million	Fortune [76]
Exclusions	Passenger systems "will not be subject to deidentification and will not be provided to the successful bidder in any form, deidentified or not"	Ombudsman report, Dkt. 1684 [72]
Standard	The parties "voluntarily agreed to follow the deidentification standard adopted in the California Consumer Privacy Act"	Dkt. 1684 [72]
Testing	The de-identification agent "shall perform and document quality-control testing" for direct and indirect identifiers, free text and record linkage	Dkt. 1684 [72]
Privilege	The assets "shall not include any communications, documents, or other information protected by any attorney-client privilege"	Dkt. 1463 [71]
Ombudsman's view	"The potential privacy losses or costs to consumers have been eliminated or mitigated"	Dkt. 1684 [72]
Independence	On micro1's proposal: it "proposed to conduct the deidentification themselves and not to use an independent party"; the ombudsman asked to research it further	Dkt. 1684 [72]
Status	Hearing adjourned to 6 November 2026; not approved when read	Dkt. 1703 [73]

Table 3. The Spirit Airlines records sale in the court record, as read on 18 October 2026.

Two features of the record are worth noting for licences outside bankruptcy. Structure was priced: the estate preferred a lower bid with third-party de-identification over a higher one that let the bidder scrub the records itself [74]. And the de-identification standard came from California's statute, including its legal conditions, adopted voluntarily [72].

AI-era cases and out-of-court routes. Character.AI filed for chapter 11 on 12 October 2026 to facilitate a court-supervised sale, telling users their information "continues to be protected under our Privacy Policy" [86]. State attorneys general said they would seek, among other things, "limits on using private conversations for AI training or other commercial purposes without valid consent" and a chance for consumers to delete their information before any sale [87, 88]. Outside court, in an assignment for the benefit of creditors or a dissolution, there is no ombudsman and no sale order; the company's privacy promises, state law and contracts apply unchanged, and a sale of all or substantially all assets of a Delaware corporation needs a shareholder resolution [9, 89].

3.5 The licence terms buyers ask for

The licence decides what a price buys. The compiled register records 15 terms; for 2 of them the public record is thin [3, 6]. Published terms diverge widely. At one end, micro1's master terms take "a worldwide, royalty-free, fully paid-up, nonexclusive, irrevocable, perpetual license, with the right to sublicense through multiple tiers" [90]. At the other, Handshake says it "is contractually restricted to using the data for AI model training and evaluation only" and that the company "retains ownership of its data" [91].

Term	What buyers ask, per published terms and templates	Public evidence	Gap in the record
Purpose and field of use	Training, fine-tuning, evaluation, sometimes retrieval	NVIDIA's data agreement limits use to "internal training" [93]	None recorded
Survival of trained models	Perpetual right to keep trained models	micro1: deletion does "not require Licensee to delete or retrain any model, weights, or derivative works" [90]; a template: "silence grants none" [94]	None recorded
Term	Two to three years in disclosed deals	Reddit's 2024 contracts: \$203.0 million aggregate value, "terms ranging from two to three years" [95]	None recorded
Exclusivity	Sometimes exclusive; often a warranty of no prior licence	Founders told Epoch AI exclusive deals are "roughly 4-5x more expensive" [97]; micro1 asks the holder to warrant no prior transaction on "the same or substantially similar terms" [90]	No public source on typical exclusivity window lengths
Price structure	Fixed fee, minimum plus variable, per unit, revenue share	HarperCollins: "\$5,000 fee per title, which will be split 50-50" (Authors Guild) [96]	None recorded
Warranties	Title, lawful collection, non-infringement, privacy	A template uses a knowledge-qualified warranty tied to a rights schedule [94]; SimpleClosure asks the holder to warrant all required notices and consents [92]	None recorded
Indemnity and caps	Licensor indemnity for third-party claims	NVIDIA's form runs the indemnity from recipient to provider [93]	No public source on typical cap levels
Deletion of the copy	Delete the dataset, not the models	NVIDIA: delete "within fourteen (14) days" of termination [93]	None recorded
Audit and certification	Records and certification; buyers resist pipeline audits	Spirit: certification filed before each transfer [72]	None recorded
No re-identification, flow-down	Accepted when the holder insists; required by the CCPA test	California's third condition [34]	None recorded
Output limits	Caps on verbatim reproduction	HarperCollins: no more than "200 consecutive words and/or 5 percent of a book's text" [96]	None recorded
Who de-identifies	Some buyers offer to do it after receipt	Counsel: "Select (or jointly select) the de-identification provider, approve the protocol, and require certification" [30]	None recorded

Table 4. Selected licence terms, compiled from published agreements, templates and disclosed deals [3, 10].

Three observations follow from the table. First, remedies can be sharp: micro1's terms provide that if the data manifest is materially inaccurate, "no amounts are payable to Licensor, Licensor will refund any amounts already paid", a right that "survives Licensee's acceptance" [90]. Second, licences outlive the licensor's independence: in one public acquisition agreement, the company being sold had to represent whether it had let any company or customer data "(including deidentified data)" be used as training data by anyone else [98]. Third, buyers' disclosure duties may name the holder in general terms: California's AB 2013 requires developers of public generative AI systems to post "The sources or owners of the datasets" and "Whether the datasets were purchased or licensed by the developer" [99], although early disclosures stayed at the level of categories such as "nonpublic data obtained from third-party partners" according to one law firm's review [100].

3.6 Copyright and recent litigation

For private records that a buyer cannot reach without the holder, copyright is rarely the central question; contract and privacy are [11]. It matters in two ways. Archives often contain third-party content (downloaded reports, vendors' documents), which the checklist treats as a blocker until removed or licensed [2]. And the case law sets the value of the alternative to a licence.

The US rulings split. In *Bartz v. Anthropic*, the court granted "summary judgment for Anthropic that the training use was a fair use" but sent pirated library copies to trial [101]; the case then settled. In *Kadrey v. Meta*, the court found fair use "on this record", stressing that the ruling "only affects the rights of these thirteen authors" [102]. On 29 September 2026 the Third Circuit held in *Thomson Reuters v. Ross* that "Thomson Reuters's materials are copyrightable and that ROSS's use was not fair", in a case about a non-generative legal search tool [103]. The Copyright Office's pre-publication report says copying from pirate sources "to generate unrestricted content that competes in the marketplace, when licensing is reasonably available, is unlikely to qualify as fair use" [104].

In the United Kingdom, the High Court held in *Getty Images v. Stability AI* that "the model weights are not themselves an infringing copy and they do not store an infringing copy" [105]; an appeal on secondary infringement is expected in November 2026 [4]. The UK's text and data analysis exception covers only computational analysis "for the sole purpose of research for a non-commercial purpose" [108]. In Canada, the government's consultation report on copyright and generative AI said it "will consider options" on text and data mining, with no exception enacted [58].

Recent litigation also shows contract claims standing apart from copyright. In *Reddit v. Anthropic*, a federal court remanded Reddit's contract and related claims to state court, holding them not preempted by the Copyright Act because they rest on extra elements such as Reddit's user agreement [106]; a state judge later allowed three of Reddit's five claims to proceed (as reported) [107]. The register's privacy entries include two proposed Illinois class actions filed on 9 October 2026 alleging that delivery-robot companies retained face and hand scans for AI training (allegations only, as reported) [4].

4. Open questions

The sources compiled leave the following questions open as of 18 October 2026. Each is stated as the sources state it.

1. **Native exports and API terms.** The API restrictions in Section 3.1 address data obtained through APIs. None of the terms read says in terms whether a licence of data exported through the vendor's own admin tool is affected, and Slack frames full exports as a legal and regulatory compliance tool [1, 16].
2. **The employer as "intended recipient".** Whether an employer counts as the intended recipient of messages its staff exchange, for the purpose of California's sensitive-information category, is not settled [8, 34].
3. **Exports as monitoring.** Whether pulling mailboxes into an export and reviewing them is "monitoring" under the state notice laws is not addressed in the sources read [8].
4. **Labour law.** No source was found applying section 7 of the NLRA directly to the sale of employee communications; the checklist's caution is an inference [2].
5. **Raw audio and video under BIPA.** Whether raw recordings, as opposed to extracted templates, are biometric identifiers is contested in pending cases [2].
6. **Bankruptcy and employees.** Whether the Spirit sale will be approved, and on what terms for employee and vendor data, is pending at a hearing set for 6 November 2026 [73]. The ombudsman's mandate did

not reach employees [72].

7. **AB 2013 specificity.** The statute does not say how specific "sources or owners" must be, and a challenge to it was pending when read [10, 99].
8. **UK research.** Whether training a commercial model qualifies as scientific research under the 2025 Act's broadened definition is untested [12, 51].
9. **Canada.** Whether Bill C-36 will add a legitimate-interests ground is open [57]; how the Commissioner's 2026 anonymisation standard applies to free-text records has not been tested in a published finding [53]; and provincial employee-privacy statutes were not compiled.
10. **UK and Canadian insolvency.** The compilation covers US bankruptcy in depth and does not cover how UK administrators or liquidators, or Canadian trustees and receivers, sell records containing personal data.
11. **Market terms.** No public source states typical exclusivity window lengths for corporate data deals, or typical indemnity caps in AI data licences [3].
12. **De-identified data and national security.** The Justice Department counts bulk data regardless of de-identification [39]; how this applies to heavily scrubbed free text is not addressed in the sources read.

5. Limitations

Compilation, not survey. The datasets record what is public. Disputes that settle privately, licences that are not disclosed and buyers that publish no terms are under-represented. The 18 software systems are common business tools chosen by the compiler, not a sample.

Survivorship and disclosure bias. Disclosed deal terms come disproportionately from public companies (Reddit's prospectus), large publishers and court-supervised sales. Small private licences of company records rarely leave a public trace.

Small numbers. Several findings rest on one case. The Spirit sale is the only court-supervised sale of working records for AI training in the compilation, and it was not approved when read.

Dates. Every term, statute and docket entry is reported as read on its stated date. Vendor terms in particular change without notice, and several cited items (the Spirit hearing, the Getty appeal, Bill C-36) were pending.

Reading, not advice. Severity ratings and statements of where holders stand are the compiler's readings. Some items rest on press reports or search excerpts and are marked as such. The analysis has not been reviewed by counsel in any of the three jurisdictions.

Interest. The author's organisation compiled the datasets and acts for data holders in licensing transactions. The datasets are open, and every row links to its source so that readers can check the readings.

6. Conclusion

Public sources do not support a single answer to whether a company may license its records for AI training. They support a sequence of questions, and on several of them a consistent pattern.

Ownership between customer and software vendor is rarely in doubt; the vendors say the content is the customer's. The route out of the system is where restrictions concentrate, with a third of the systems read barring AI training on data obtained through their APIs. Personal data is the central constraint in all three jurisdictions, and the three regimes differ in kind: notice and opt-out with a three-part de-identification test in California, a documented legitimate-interests and compatibility analysis in the United Kingdom, and consent or a strict anonymisation standard in Canada. Employee communications are generally the employer's property

under US copyright law, but they carry the rights of the people in them. The one court-supervised sale of working records for AI training on record shows the pieces assembled: consumer systems carved out, privileged material excluded, an independent agent applying a statutory de-identification standard with certification before each transfer, and objections from employees and vendors that the bankruptcy rule does not address. Published licence terms vary enough that two offers at the same price can grant very different rights.

The open questions in Section 4 are where the next statutes, rulings and disclosed agreements will add most.

Competing interests

The author is the chief executive of AIData.Broker, which compiled the datasets cited and acts as agent for data holders in licensing transactions.

Note

This article describes public sources as read on the dates given. It is not legal advice. A licence of records should be reviewed by counsel who have read the records and the contracts that govern them.

References

Compilations

- [1] AIData.Broker, "What software terms allow for AI training data" (dataset, reviewed 11 October 2026, CC BY 4.0). <https://aidata.broker/datasets/software-terms.json> (accessed 18 October 2026).
- [2] AIData.Broker, "Rights checklist for selling data for AI training" (dataset, reviewed 11 October 2026, CC BY 4.0). <https://aidata.broker/datasets/rights-checks.json> (accessed 18 October 2026).
- [3] AIData.Broker, "Terms of an AI training data licence" (dataset, reviewed 11 October 2026, CC BY 4.0). <https://aidata.broker/datasets/licence-terms.json> (accessed 18 October 2026).
- [4] AIData.Broker, "Lawsuits and rulings over AI training data" (dataset, reviewed 17 October 2026, CC BY 4.0). <https://aidata.broker/datasets/cases.json> (accessed 18 October 2026).
- [5] AIData.Broker, "Can it be sold? The rights behind AI training data." <https://aidata.broker/rights> (accessed 18 October 2026).
- [6] AIData.Broker, "What a data licence turns on." <https://aidata.broker/licence-terms> (accessed 18 October 2026).
- [7] AIData.Broker, "Is It Legal to Sell Customer Data? A US Guide for AI Training," 17 October 2026. <https://aidata.broker/guides/is-it-legal-to-sell-customer-data> (accessed 18 October 2026).
- [8] AIData.Broker, "Can a Company License Employee Emails and Slack for AI Training?," 17 October 2026. <https://aidata.broker/guides/employee-emails-and-chats-ai-training> (accessed 18 October 2026).
- [9] AIData.Broker, "Selling Customer Data in Bankruptcy: Section 363 Explained," 17 October 2026. <https://aidata.broker/guides/selling-customer-data-in-bankruptcy> (accessed 18 October 2026).
- [10] AIData.Broker, "What Is a Data Licensing Agreement? The Terms for AI Training," 17 October 2026. <https://aidata.broker/guides/what-is-a-data-licensing-agreement> (accessed 18 October 2026).

- [11] AIData.Broker, "Fair Use and AI Training: What US Courts Have Ruled," 17 October 2026. <https://aidata.broker/guides/fair-use-ai-training> (accessed 18 October 2026).
- [12] AIData.Broker, "Licensing Company Data for AI Training Under the GDPR (2026)," 18 October 2026. <https://aidata.broker/guides/gdpr-licensing-data-for-ai-training> (accessed 18 October 2026).
- [13] AIData.Broker, "Selling data to AI companies from Canada." <https://aidata.broker/countries/canada> (accessed 18 October 2026).
- [14] AIData.Broker, "Selling data to AI companies from the United Kingdom." <https://aidata.broker/countries/united-kingdom> (accessed 18 October 2026).
- Software vendors' terms
- [15] Slack, "Slack API Terms of Service," effective 10 October 2025. <https://slack.com/terms-of-service/api> (accessed 18 October 2026).
- [16] Slack, "Export your workspace data," help article. <https://slack.com/help/articles/201658943-Export-your-workspace-data> (accessed 18 October 2026).
- [17] Google, "Google Workspace API User Data and Developer Policy," last updated 3 September 2026. <https://developers.google.com/workspace/workspace-api-user-data-developer-policy> (accessed 18 October 2026).
- [18] Zendesk, "Zendesk Developer Terms," effective 12 August 2025. <https://www.zendesk.nl/company/agreements-and-terms/zendesk-developer-terms/> (accessed 18 October 2026).
- [19] Zoom, "Zoom API License and Terms of Use," last updated 16 July 2025. <https://explore.zoom.us/en/legal/zoom-api-license-and-tou/> (accessed 18 October 2026).
- [20] Xero, "Xero Developer Platform Terms and Conditions," last updated 4 December 2025. <https://developer.xero.com/xero-developer-platform-terms-conditions> (accessed 18 October 2026).
- [21] HubSpot, "Our ecosystem vision for the agent era, updated Developer Terms," developer changelog, 4 May 2026. <https://developers.hubspot.com/changelog/our-ecosystem-vision-for-the-agent-era-updated-developer-terms> (accessed 18 October 2026).
- [22] HubSpot, "Customer Terms of Service," last modified 16 September 2026, s. 5.3.1. <https://legal.hubspot.com/terms-of-service> (accessed 18 October 2026).
- [23] Microsoft, "Terms of Use: Microsoft APIs," last updated October 2025. <https://learn.microsoft.com/en-us/legal/microsoft-apis/terms-of-use> (accessed 18 October 2026).
- [24] Microsoft, "Microsoft Products and Services Data Protection Addendum," last updated 22 May 2026. [https://www.microsoft.com/licensing/docs/documents/download/MicrosoftProductandServicesDPA\(WW\)\(English\)\(May2026\)\(CR\).docx](https://www.microsoft.com/licensing/docs/documents/download/MicrosoftProductandServicesDPA(WW)(English)(May2026)(CR).docx) (accessed 18 October 2026).
- [25] Atlassian, "Atlassian Customer Agreement," effective 1 October 2026. <https://www.atlassian.com/legal/atlassian-customer-agreement> (accessed 18 October 2026).
- [26] Atlassian, "Data contribution," Atlassian Trust (change effective 17 August 2026). <https://www.atlassian.com/trust/ai/data-contribution> (accessed 18 October 2026).

[27] GitHub, "GitHub Terms of Service," effective 27 April 2026 (including section J.3). <https://docs.github.com/en/site-policy/github-terms/github-terms-of-service> (accessed 18 October 2026).

[28] Salesforce, "Main Services Agreement," 15 October 2026. https://www.salesforce.com/en-us/wp-content/uploads/sites/4/documents/legal/salesforce_MSA.pdf (accessed 18 October 2026).

[29] Hunton Andrews Kurth, "Salesforce locks down Slack data: time to review your Slack API terms," 28 July 2025. <https://www.hunton.com/insights/legal/salesforce-locks-down-slack-data-time-to-review-your-slack-api-terms> (accessed 18 October 2026).

Practitioner commentary

[30] D. M. Goldberg (Frankfurt Kurnit Klein & Selz), "Selling Company Data For AI Training: What To Know Before You Say Yes," Legal 500, 6 October 2026. <https://www.legal500.com/intelligence/united-states/privacy/selling-company-data-for-ai-training-what-to-know-before-you-say-yes> (accessed 18 October 2026).

United States: privacy

[31] Federal Trade Commission, Office of Technology, "AI (and other) Companies: Quietly Changing Your Terms of Service Could Be Unfair or Deceptive," 13 February 2024. <https://www.ftc.gov/policy/advocacy-research/tech-at-ftc/2024/02/ai-other-companies-quietly-changing-your-terms-service-could-be-unfair-or-deceptive> (accessed 18 October 2026).

[32] Federal Trade Commission, "FTC Finalizes Settlement with Photo App Developer Related to Misuse of Facial Recognition Technology" (Everalbum), May 2021. <https://www.ftc.gov/news-events/news/press-releases/2021/05/ftc-finalizes-settlement-photo-app-developer-related-misuse-facial-recognition-technology> (accessed 18 October 2026).

[33] Federal Trade Commission, press release on WW International and Kurbo, March 2022. <https://www.ftc.gov/news-events/news/press-releases/2022/03/ftc-takes-action-against-company-formerly-known-weight-watchers-illegally-collecting-kids-sensitive> (accessed 18 October 2026).

[34] California Consumer Privacy Act, Cal. Civ. Code 1798.100 et seq., ss. 1798.140(ad), (ae), (ag), (m) and 1798.145(m), statute text as published by the California Privacy Protection Agency. https://coppa.ca.gov/pdf/20260101_ccpa_statute.pdf (accessed 18 October 2026).

[35] California Privacy Protection Agency, "Frequently Asked Questions." <https://coppa.ca.gov/faq.html> (accessed 18 October 2026).

[36] Texas Data Privacy and Security Act, Tex. Bus. & Com. Code ch. 541, ss. 541.001(7) and 541.106. <https://statutes.capitol.texas.gov/Docs/BC/htm/BC.541.htm> (accessed 18 October 2026).

[37] Virginia Consumer Data Protection Act, Va. Code 59.1-581. <https://law.lis.virginia.gov/vacode/title59.1/chapter53/section59.1-581/> (accessed 18 October 2026).

[38] 28 CFR 202.205 (bulk thresholds). <https://www.law.cornell.edu/cfr/text/28/202.205> (accessed 18 October 2026).

[39] US Department of Justice, National Security Division, "Data Security Program: Frequently Asked Questions," question 22. <https://justice.gov/nsd/media/1415006/dl> (accessed 18 October 2026).

[40] Illinois Biometric Information Privacy Act, 740 ILCS 14/15(c).
<https://www.ilga.gov/Documents/legislation/ilcs/documents/074000140K15.htm> (accessed 18 October 2026).

United Kingdom

[41] Data (Use and Access) Act 2025, Schedule 4 (new UK GDPR Annex 1).
<https://www.legislation.gov.uk/ukpga/2025/18/schedule/4/enacted> (accessed 18 October 2026).

[42] Data (Use and Access) Act 2025, section 71 (new UK GDPR Article 8A).
<https://www.legislation.gov.uk/ukpga/2025/18/section/71/enacted> (accessed 18 October 2026).

[43] The Data (Use and Access) Act 2025 (Commencement) Regulations 2026, SI 2026/82.
<https://www.legislation.gov.uk/uksi/2026/82/made> (accessed 18 October 2026).

[44] Information Commissioner's Office, "Legitimate interests," A guide to lawful basis. <https://ico.org.uk/for-organisations/uk-gdpr-guidance-and-resources/lawful-basis/a-guide-to-lawful-basis/legitimate-interests/> (accessed 18 October 2026).

[45] Information Commissioner's Office, "Principle (b): Purpose limitation," updated 23 March 2026.
<https://ico.org.uk/for-organisations/uk-gdpr-guidance-and-resources/data-protection-principles/a-guide-to-the-data-protection-principles/purpose-limitation/> (accessed 18 October 2026).

[46] Information Commissioner's Office, "How do we ensure anonymisation is effective?," Anonymisation guidance, 28 March 2025. <https://ico.org.uk/for-organisations/uk-gdpr-guidance-and-resources/data-sharing/anonymisation/how-do-we-ensure-anonymisation-is-effective/> (accessed 18 October 2026).

[47] Information Commissioner's Office, "Introduction to anonymisation," 28 March 2025.
<https://ico.org.uk/for-organisations/uk-gdpr-guidance-and-resources/data-sharing/anonymisation/introduction-to-anonymisation/> (accessed 18 October 2026).

[48] The Data Protection (Adequacy) (United States of America) Regulations 2023, SI 2023/1028, reg. 3(2).
<https://www.legislation.gov.uk/uksi/2023/1028/made> (accessed 18 October 2026).

[49] Information Commissioner's Office, "A brief guide to international transfers." <https://ico.org.uk/for-organisations/uk-gdpr-guidance-and-resources/international-transfers/a-brief-guide-to-international-transfers/> (accessed 18 October 2026).

[50] Information Commissioner's Office, "The lawful basis for web scraping to train generative AI models," response to the consultation series on generative AI, 2024. <https://ico.org.uk/about-the-ico/what-we-do/our-work-on-artificial-intelligence/response-to-the-consultation-series-on-generative-ai/the-lawful-basis-for-web-scraping-to-train-generative-ai-models/> (accessed 18 October 2026).

[51] Data (Use and Access) Act 2025, section 67 (meaning of scientific research).
<https://www.legislation.gov.uk/ukpga/2025/18/section/67/enacted> (accessed 18 October 2026).

Canada

[52] Personal Information Protection and Electronic Documents Act, S.C. 2000, c. 5, ss. 7(3)(f), 7.2(4) and Schedule 1, cl. 4.2.4. <https://laws-lois.justice.gc.ca/eng/acts/P-8.6/FullText.html> (accessed 18 October 2026).

[53] Office of the Privacy Commissioner of Canada, PIPEDA Findings #2026-001 (Loblaw, PC Optimum), 5 March 2026. <https://www.priv.gc.ca/en/opc-actions-and-decisions/investigations/investigations-into->

businesses/2026/pipeda-2026-001/ (accessed 18 October 2026).

[54] Office of the Privacy Commissioner of Canada, "Guidelines for processing personal data across borders." https://www.priv.gc.ca/en/privacy-topics/airports-and-borders/gl_dab_090127/ (accessed 18 October 2026).

[55] Act respecting the protection of personal information in the private sector, CQLR c. P-39.1, ss. 12, 13, 17 and 23, as amended by S.Q. 2021, c. 25. https://www.publicationsduquebec.gouv.qc.ca/fileadmin/gazette/pdf_encrypte/gaz_entiere/2202-A.pdf (accessed 18 October 2026).

[56] Regulation respecting the anonymization of personal information, O.C. 783-2024, ss. 4 and 7 (in force 30 May 2024). https://www.publicationsduquebec.gouv.qc.ca/fileadmin/gazette/pdf_encrypte/lois_reglements/2024A/106829.pdf (accessed 18 October 2026).

[57] Parliament of Canada, LEGISinfo, Bill C-36 (45th Parliament, 1st session). <https://www.parl.ca/legisinfo/en/bill/45-1/c-36> (accessed 18 October 2026).

[58] Innovation, Science and Economic Development Canada, "Consultation on copyright in the age of generative artificial intelligence: What we heard report," 2025. <https://ised-isde.canada.ca/site/strategic-policy-sector/en/marketplace-framework-policy/consultation-copyright-age-generative-artificial-intelligence-what-we-heard-report> (accessed 18 October 2026).

Employee communications

[59] 17 U.S.C. 101 (definition of work made for hire). <https://www.law.cornell.edu/uscode/text/17/101> (accessed 18 October 2026).

[60] 17 U.S.C. 201(b). <https://www.law.cornell.edu/uscode/text/17/201> (accessed 18 October 2026).

[61] 18 U.S.C. 2701(c). <https://www.law.cornell.edu/uscode/text/18/2701> (accessed 18 October 2026).

[62] 18 U.S.C. 2702(a). <https://www.law.cornell.edu/uscode/text/18/2702> (accessed 18 October 2026).

[63] 18 U.S.C. 2511(2)(a)(i) and (2)(d). <https://www.law.cornell.edu/uscode/text/18/2511> (accessed 18 October 2026).

[64] New York Civil Rights Law 52-c. https://www.nysenate.gov/legislation/laws/CVR/52-C*2 (accessed 18 October 2026).

[65] Connecticut Public Act 26-73 (amending Gen. Stat. 31-48d). <https://www.cga.ct.gov/2026/ACT/PA/PDF/2026PA-00073-R00SB-00472-PA.PDF> (accessed 18 October 2026).

[66] 19 Del. C. 705. <https://delcode.delaware.gov/title19/c007/sc01/index.html> (accessed 18 October 2026).

[67] 26 M.R.S. 620-A. <https://legislature.maine.gov/statutes/26/title26sec620-A.html> (accessed 18 October 2026).

[68] California Penal Code 632(a). https://leginfo.legislature.ca.gov/faces/codes_displaySection.xhtml?lawCode=PEN§ionNum=632 (accessed 18 October 2026).

[69] National Labor Relations Act section 7, 29 U.S.C. 157. <https://www.law.cornell.edu/uscode/text/29/157> (accessed 18 October 2026).

[70] Federal Rule of Evidence 502(b). https://www.law.cornell.edu/rules/fre/rule_502 (accessed 18 October 2026).

Bankruptcy

[71] In re Spirit Aviation Holdings, Inc., No. 25-11897 (Bankr. S.D.N.Y.), Dkt. 1463, Notice of Auction Results, Exhibit A (Sale Agreement, Google LLC) and Exhibit B (Alternate Sale Agreement, Mercor.io), 14 August 2026. <https://storage.courtlistener.com/recap/gov.uscourts.nysb.331039/gov.uscourts.nysb.331039.1463.0.pdf> (accessed 18 October 2026).

[72] In re Spirit Aviation Holdings, Inc., Dkt. 1684, Consumer Privacy Ombudsman supplemental report, 5 October 2026.

<https://storage.courtlistener.com/recap/gov.uscourts.nysb.331039/gov.uscourts.nysb.331039.1684.0.pdf> (accessed 18 October 2026).

[73] In re Spirit Aviation Holdings, Inc., Dkt. 1703, amended agenda for the 14 October 2026 hearing. <https://storage.courtlistener.com/recap/gov.uscourts.nysb.331039/gov.uscourts.nysb.331039.1703.0.pdf> (accessed 18 October 2026).

[74] Chapter11Cases (Stretto), "The Spirit Airlines deidentified data sale: auction results, objections and the September 30 hearing." <https://chapter11cases.com/blogs/news/the-spirit-airlines-deidentified-data-sale-auction-results-objections-and-the-september-30-hearing> (accessed 18 October 2026).

[75] PYMNTS, report on the letter from members of Congress to Spirit Airlines and Google, October 2026. <https://www.pymnts.com/?p=4292562> (accessed 18 October 2026).

[76] Fortune, "micro1 tries to beat Google's bid to buy bankrupt Spirit Airlines data for \$12.5 million," 14 September 2026. <https://fortune.com/2026/09/14/micro1-tries-to-beat-googles-bid-to-buy-bankrupt-spirit-airlines-data-for-12-5-million-highlighting-value-of-data-to-ai-companies/> (accessed 18 October 2026).

[77] The Star (Reuters report), "Google to buy Spirit Airlines business data for \$10 million," 18 August 2026. <https://www.thestar.com.my/tech/tech-news/2026/08/18/google-to-buy-spirit-airlines-business-data-for-10-million> (accessed 18 October 2026).

[78] 11 U.S.C. 363(b)(1). <https://www.govinfo.gov/content/pkg/USCODE-2024-title11/html/USCODE-2024-title11-chap3-subchapIV-sec363.htm> (accessed 18 October 2026).

[79] 11 U.S.C. 332(a). <https://www.govinfo.gov/content/pkg/USCODE-2024-title11/html/USCODE-2024-title11-chap3-subchapII-sec332.htm> (accessed 18 October 2026).

[80] 11 U.S.C. 101(41A). <https://www.govinfo.gov/content/pkg/USCODE-2024-title11/html/USCODE-2024-title11-chap1-sec101.htm> (accessed 18 October 2026).

[81] Federal Trade Commission, "FTC Announces Settlement With Bankrupt Website, Toysmart.com, Regarding Alleged Privacy Policy Violations," 21 July 2000. <https://www.ftc.gov/news-events/news/press-releases/2000/07/ftc-announces-settlement-bankrupt-website-toysmartcom-regarding-alleged-privacy-policy-violations> (accessed 18 October 2026).

[82] Federal Trade Commission, Bureau of Consumer Protection, letter to the RadioShack consumer privacy ombudsman, May 2015.

https://www.ftc.gov/system/files/documents/public_statements/643291/150518radioshackletter.pdf (accessed 18 October 2026).

[83] Office of the Texas Attorney General, "Attorney General Paxton Announces Agreement to Protect Consumer Privacy in RadioShack Case," 20 May 2015.
<https://www.texasattorneygeneral.gov/news/releases/attorney-general-paxton-announces-agreement-protect-consumer-privacy-radioshack-case> (accessed 18 October 2026).

[84] In re 23andMe Holding Co. (Bankr. E.D. Mo.), memorandum opinion approving the sale, 27 June 2025, Dkt. 908.
<https://storage.courtlistener.com/recap/gov.uscourts.moeb.462637/gov.uscourts.moeb.462637.908.0.pdf> (accessed 18 October 2026).

[85] In re 23andMe Holding Co., Report of the Consumer Privacy Ombudsman, 11 June 2025, Dkt. 718.
<https://storage.courtlistener.com/recap/gov.uscourts.moeb.462637/gov.uscourts.moeb.462637.718.0.pdf> (accessed 18 October 2026).

[86] Character.ai, "An Update on Character.ai," 12 October 2026. <https://blog.character.ai/an-update-on-character-ai/> (accessed 18 October 2026).

[87] New York State Attorney General, "Attorney General James Urges Court to Protect Children and Consumers in Character.AI Bankruptcy," 14 October 2026. <https://ag.ny.gov/press-release/2026/attorney-general-james-urges-court-protect-children-and-consumers-characterai> (accessed 18 October 2026).

[88] In re Character Technologies (Bankr. D. Del.), States' Notice of Interest and Reservation of Rights, 14 October 2026, Dkt. 33.
<https://storage.courtlistener.com/recap/gov.uscourts.deb.203004/gov.uscourts.deb.203004.33.0.pdf> (accessed 18 October 2026).

[89] Delaware General Corporation Law, 8 Del. C. 271.
<https://delcode.delaware.gov/title8/c001/sc10/index.html> (accessed 18 October 2026).

Licence terms and deals

[90] micro1, "Data License Master Terms." <https://www.micro1.ai/legal/data-license-master-terms> (accessed 18 October 2026).

[91] Handshake AI, "Data partnerships." <https://joinhandshake.com/ai/data-partnerships/> (accessed 18 October 2026).

[92] SimpleClosure, "Terms of Service." <https://simpleclosure.com/tos/> (accessed 18 October 2026).

[93] NVIDIA, "Data Agreement for Model Training" (2025), as recorded in ScanCode LicenseDB.
<https://scancode-licensedb.aboutcode.org/nvidia-model-training-2025.yml> (accessed 18 October 2026).

[94] ezel.ai, "AI Training Data License Agreement" template. <https://ezel.ai/templates/ai-training-data-license-agreement> (accessed 18 October 2026).

[95] Reddit, Inc., Form S-1, 22 February 2024.
<https://www.sec.gov/Archives/edgar/data/1713445/000162828024006294/reddits-1q423.htm> (accessed 18 October 2026).

[96] Authors Guild, "HarperCollins AI licensing deal," 19 November 2024.
<https://authorsguild.org/news/harpercollins-ai-licensing-deal/> (accessed 18 October 2026).

[97] Epoch AI, "The state of RL environments," Gradient Updates, 12 January 2026. <https://epoch.ai/gradient-updates/state-of-rl-envs> (accessed 18 October 2026).

[98] Ooma, Inc., Form 10-Q, Exhibit 10.1 (Stock Purchase Agreement with FluentStream Holdings, LP), 2025. https://www.sec.gov/Archives/edgar/data/1327688/000119312525312961/ooma-ex10_1.htm (accessed 18 October 2026).

[99] California Civil Code 3111 (AB 2013). https://california.public.law/codes/civil_code_section_3111 (accessed 18 October 2026).

[100] Goodwin, "California's AB 2013 Takes Effect," January 2026. <https://goodwinlaw.com/en/insights/publications/2026/01/alerts-otherindustries-californias-ab-2013-takes-effect> (accessed 18 October 2026).

Copyright and litigation

[101] Bartz v. Anthropic PBC, No. 3:24-cv-05417-WHA (N.D. Cal.), order on fair use, 23 June 2025. <https://storage.courtlistener.com/recap/gov.uscourts.cand.434709/gov.uscourts.cand.434709.231.0.pdf> (accessed 18 October 2026).

[102] Kadrey v. Meta Platforms, Inc., No. 3:23-cv-03417-VC (N.D. Cal.), order on summary judgment, 25 June 2025. <https://storage.courtlistener.com/recap/gov.uscourts.cand.415175/gov.uscourts.cand.415175.598.0.pdf> (accessed 18 October 2026).

[103] Thomson Reuters Enterprise Centre GmbH v. Ross Intelligence Inc., No. 25-2153 (3d Cir.), opinion, 29 September 2026. <https://www2.ca3.uscourts.gov/opinarch/252153p.pdf> (accessed 18 October 2026).

[104] US Copyright Office, "Copyright and Artificial Intelligence, Part 3: Generative AI Training," pre-publication version, 9 May 2025. <https://www.copyright.gov/ai/Copyright-and-Artificial-Intelligence-Part-3-Generative-AI-Training-Report-Pre-Publication-Version.pdf> (accessed 18 October 2026).

[105] Getty Images (US) Inc. v. Stability AI Ltd [2025] EWHC 2863 (Ch), 4 November 2025. <https://caselaw.nationalarchives.gov.uk/ewhc/ch/2025/2863> (accessed 18 October 2026).

[106] Reddit, Inc. v. Anthropic PBC, No. 3:25-cv-05643-TLT (N.D. Cal.), order granting motion to remand, 30 March 2026, Dkt. 60. <https://storage.courtlistener.com/recap/gov.uscourts.cand.452341/gov.uscourts.cand.452341.60.0.pdf> (accessed 18 October 2026).

[107] Mashable, report on the San Francisco Superior Court ruling in Reddit v. Anthropic, 19 September 2026. <https://mashable.com/tech/judge-allows-most-of-reddit-lawsuit-against-anthropic-proceed> (accessed 18 October 2026).

[108] Copyright, Designs and Patents Act 1988, section 29A. <https://www.legislation.gov.uk/ukpga/1988/48/section/29A> (accessed 18 October 2026).

Statements

Use of AI. This paper was prepared with generative AI tools, which compiled and analysed the open datasets and drafted the text. The author takes responsibility for its content.

Licence: CC BY 4.0. Cite as: Heymans, Y. (2026). Licensing company records for AI training: the rights questions buyers ask, and what public sources show. AIData.Broker, 18 October 2026. <https://aidata.broker/research/paper-rights>