

Buyer programmes for company data: eligibility, prices and payment terms of 37 organisations, October 2026

Working paper

Yuma Heymans

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

Abstract

Since 2026 AI data vendors, marketplaces and intermediaries have published programmes through which an ordinary company can sell or license its internal records for training AI models. This paper compares the published terms of 37 organisations with a public way in, as compiled in an open dataset reviewed on 16 October 2026, together with a register of the contract clauses 16 of them publish and a dated log of 110 market events. Twenty-four take company records, 17 of them buying for their own account. Of those 24, 14 publish a floor (median \$100,000) and 9 a ceiling (median \$1 million; highest \$5 million). Eligibility rules are sparse: six publish a headcount minimum (mostly 20 to 40 employees), two a minimum of three years, three a rule about country or language, and eight say they take companies that are closing or closed. Where a time to payment is stated it runs from 7 days after the data passes review to about 60 days, and several programmes pay only on the buyer's acceptance. Exclusivity is stated by few and deletion terms by six; one programme's terms say deletion does not require deleting trained models. Ten publish structured introduction fees, up to \$100,000 per deal. No published ceiling is backed by an independently reported payment of that size: one programme has payments confirmed by sellers, at tens to hundreds of thousands of dollars, and seven report their own. We describe the data, the classification and the limits of comparing posted terms.

Keywords: AI training data; data licensing; company records; buyer programmes; data markets; contract terms; posted prices; referral fees

1. Introduction

Developers of AI models have long bought data from vendors that collect or make it to order. In 2026 a different route opened [2]: programmes that invite ordinary companies to sell or license the records they already hold, such as email, chat, support tickets, CRM activity, documents and source code. Most are run by vendors that buy the records for their own account, turn them into training data or simulated workplaces, and sell the result to AI laboratories. Others are marketplaces that license on the holder's behalf for a share of each sale, intermediaries, and wind-down services for closing companies [1, 2].

These programmes publish more than prices: who may apply, how payment is triggered and timed, what licence the buyer takes, and what they pay whoever introduces a seller. These terms decide what a headline figure is worth, and they differ more than the figures do [3]. They are rare posted terms in a market that otherwise discloses little.

A companion register describes the vendors that sell AI training data [4]; this paper concerns the buyers alone. It asks five questions of the public record as of 18 October 2026. What does each programme buy, and from whom? What does it publish about price? Who is eligible, by size, age, country and status? When and on what condition does it pay, and what licence, exclusivity and deletion terms does it publish? And

how often are its advertised ceilings backed by reported payments? Contract terms are described as published; nothing here is legal advice.

2. Data and method

2.1 Sources

The paper draws on four open datasets compiled by the author's organisation and released under CC BY 4.0, each row tied to the page or report it was read from, with a date and the source's own words [1, 5, 6, 7].

Buyers. The table of organisations with a public way to sell them data (37 rows, reviewed 16 October 2026, pages read between 11 and 16 October) [1]. Each row records what the organisation does with what it buys, the records it wants, what it publishes about price (with the lowest and highest dollar figure it names), how and when it pays, who removes personal data, its eligibility rules (as structured fields and in its own words), the kinds of data it takes, and what it pays introducers (with structured terms where they can be expressed as numbers). Thirty-six rows rest on the organisation's own pages and one on named press. Each row also carries a "track record" tier, the one field that is the compiler's judgement: confirmed (money moving that someone else confirms, or \$5 million or more raised from named investors, and a named team), partial (some independent evidence but no deal anyone else confirms) or own claims only [1].

Buyer terms. A register of the contract clauses 16 organisations publish, each clause quoted verbatim with its page and reading date (11 to 17 October 2026), classified by term (licence, sublicensing, exclusivity, earlier licences, acceptance, refunds, de-identification, payment, deletion, resale), with a list of the terms each organisation does not publish [5]. Fourteen of the 16 are in the buyer table; two (a large technology company's content pilot and a smaller licensing firm) are not.

Market log. A dated log of 110 events in company-data sales (reviewed 17 October 2026): programmes, price ranges, referral terms, intermediaries, transactions and legal or policy events, each with a basis of stated, self-reported or reported [6].

Negotiation levers. A table of 69 findings about how such deals are priced and negotiated, each with quoted sources; it is used here for the outputs of the buyers' public estimators [7].

2.2 Computation and classification

Counts, medians and the tables below were computed from the JSON files with a short Python script. Figures are reported as the organisations publish them, with no conversion. Terms are reported in the organisations' own words where possible.

To answer the last question, each of the 24 programmes that take company records was placed in one of four classes by the strongest evidence of payment in the four datasets: (A) a payment for records through the programme confirmed by a seller on the record; (B) a bid, offer or statement in a court record, with no payment shown; (C) the organisation's own statement of payments made, including statements relayed by press; (D) nothing found. The classification uses each buyer row's track-record note and every transaction in the market log [1, 6].

3. Findings

3.1 Who buys, and what

Table 1 sets the 37 organisations against the kinds of data they take. Company records are taken by 24, and 17 of those buy for their own account. Source code (13) follows; 10 of the 13 organisations that take it also take company records. Marketplaces account for most of the demand for media: 8 of the 10 that take video and 7 of the 9 that take images.

Table 1. Organisations by kind and the kinds of data they take

Kind of data	Own account	Marketplace	Intermediary	Wind-down	Work to order	All
Company records	17	2	3	2	0	24
Source code	5	3	2	2	1	13
Video	1	8	1	0	0	10
Images	1	7	1	0	0	9
Audio and speech	1	6	1	0	0	8
Datasets and environments	1	3	2	0	1	7
Text archives	0	5	1	0	0	6
Expert work	2	0	0	0	1	3
Physical-world data	2	1	0	0	0	3
Organisations	18	12	3	2	2	37

Source: computed from [1]. An organisation can take several kinds of data. The rest of the paper concerns the 24 that take company records.

3.2 What they publish about price

Table 2 lists the 24 programmes, largest published ceiling first. Fourteen publish a floor and nine a ceiling. The median floor is \$100,000 and the median ceiling \$1 million; the highest ceiling is \$5 million (Mode's BizBuyData) and the lowest floor \$10,000 (four programmes). The other ten have no published bounds in the table.

Table 2. Programmes that take company records: published figures and eligibility

Programme	Kind	Floor	Ceiling	Eligibility as published	Takes closing or closed companies
Mode (BizBuyData)	Own account	\$100,000	\$5,000,000	20+ office employees (accounting firms 10+, law firms 6+), teams primarily in the US	
Handshake AI	Own account	\$100,000	\$4,000,000	20+ full-time employees and 3+ years in operation	
Polyshares	Own account	\$100,000	\$2,000,000	No size or region rule	
SuperAnnotate	Own account	\$100,000	\$1,000,000+	Any size, anywhere	
micro1	Own account	\$100,000	\$1,000,000+	30+ employees, English; US first, then UK, Canada, other Western markets	Separate track: 50+ at peak
Labelbox (Prism)	Own account	\$50,000	\$1,000,000+	Any industry	
Ooak	Own account	\$20,000	\$1,000,000	20+ employees	Yes
Scale AI	Own account	\$10,000	\$1,000,000+	40+ employees as "preferred starting point"	
Turing (Lazarus)	Own account	\$10,000	\$1,000,000	Demand strongest in US, UK, Canada, EU; code 50,000+ lines with tests	Yes
HUMXN Research Labs	Own account	\$250,000		Skilled trades and manufacturing	
Nyne	Own account	\$100,000		20+ full-time US W-2 employees and 3+ years	
Troveo	Marketplace	\$100,000		No minimum	
Corpus Bridge	Intermediary	\$10,000		Established businesses	
SimpleClosure AssetHub	Wind-down	\$10,000		Startups shutting down; operating companies too	Yes
Appen	Own account			Any size	Answered first
Centific	Own account			Enterprises in any industry	Faster process
Collinear AI	Own account			Holders of the rights; human-written code	
FileYield	Intermediary			No minimum size	
Grepped	Intermediary			No rule; prefers established businesses	
Halluminate	Own account			Not published	
Magma	Marketplace			Companies building AI agents	
Mercor	Own account			Eight named industries	
Sunset	Wind-down			Startups winding down	Yes

Programme	Kind	Floor	Ceiling	Eligibility as published	Takes closing or closed companies
The General Data Company	Own account			Industrial companies	Including those no longer trading

Source: computed from [1]; eligibility condensed from each row's own wording. "+" marks a ceiling the organisation states as open-ended. Grepped's own pages cite \$20,000 to \$5 million per company, which the table records in its text but not as the buyer's own bounds [1].

The wording of the figures matters as much as the numbers. Scale calls its range an "illustrative value per data partnership"; SuperAnnotate's is "indicative"; Turing and Oak publish "up to" a sum; Mode says operating history "could be worth" its range; Nyne says partnerships are "commonly worth \$100K or more"; Polyshares gives "typical" licences; and micro1 publishes tiers of \$100,000, \$500,000 and \$1 million and above by kind of data [1]. Only one of these, Handshake, presents its range as what it has paid: its page says "Payments to date have ranged from \$100K to \$4M" [1].

Six programmes publish estimators that turn a company's size and age into a figure. Three of them compute the same formula: for a US company of 100 employees and 10 years, micro1 shows \$379,290 to \$888,088 and Handshake \$383,000 to \$897,000, while Nyne's page code computes the same figure but displays a random 10 to 15 percent of it [7]. For the same company, Grepped's estimator shows \$37,500 to \$54,000, Oak's \$99,000 to \$139,000 and Troveo's \$800,000 to \$1 million, so that the low ends differ about 21 times [7]. An estimator figure therefore says more about the programme than about the records.

3.3 Who is eligible

Eligibility rules are few and simple. Six programmes publish a minimum headcount: 20 employees (Handshake AI, Nyne, Oak, and Mode for office staff, with 10 for accounting firms and 6 for law firms), 30 (micro1) and 40 as a preferred start (Scale AI). Two require three years in operation (Handshake and Nyne). Three publish a rule about country or language: Nyne requires US W-2 employees, Mode wants teams primarily in the United States, and micro1 wants companies that work and document in English and takes the United States first [1]. Turing does not set a rule but names the United States, the United Kingdom, Canada and the European Union as where its demand is strongest [1]. Four name target industries: Mercor eight (legal, healthcare, finance and insurance, professional services, manufacturing, retail and e-commerce, real estate, logistics), The General Data Company industrial companies, HUMXN skilled trades and manufacturing, and Magma companies that build AI agents [1].

Country also enters through price. The three estimators that share a formula weight a company in the United States at 1, in Canada or Europe at 0.75 and elsewhere at 0.3: for 100 employees and 10 years, micro1 shows \$379,290 to \$888,088 in the United States, \$284,468 to \$666,066 in Europe and \$113,787 to \$266,426 elsewhere [7]. A company outside the United States may therefore be eligible on every published rule and still be priced at less than a third of the US figure.

Eight organisations say they take companies that are closing or have closed: four in structured rules (Turing, Oak, SimpleClosure and Sunset) and four in their own words (micro1's separate track, Appen, Centific and The General Data Company) [1]. micro1's track for shut-down companies asks for 50 or more employees at peak and substantial Teams or Slack history, and lists the countries it takes [6]. Oak states that a liquidator or administrator can sign [1].

Across all 37 organisations, 25 publish no structured eligibility criterion [1]. The absence of a rule does not mean any company qualifies: several programmes assess each case, and Scale's job advertisement for its programme says the team is still defining "the first version of how these transactions get priced" [5].

3.4 When they pay

Table 3 lists the programmes that state a time. The fastest stated payments are counted from the buyer's own review: Grepped pays "within 7 days of your data passing review" and says most companies go from form to payment in two to three weeks [5]. Labelbox gives "from first call to first wire in under 21 days" [1]. At the other end, Polyshares pays net 30 to 60 days after the data is approved and anonymised, and Mode's launch release expects payment in about 60 days, with a letter of intent within 14 days and 15 to 30 days of collection and de-identification [1, 6].

Table 3. Programmes that state when they pay

Programme	Trigger	Stated time
Grepped	Data passes review	Within 7 days of review; two to three weeks from form to payment for most companies
Handshake AI	Delivery milestones agreed per deal	Pricing within 12 to 24 hours of the details it needs; whole process typically 7 to 14 business days
Labelbox (Prism)	Not stated	First call to first wire in under 21 days; quote within 48 hours of scoping
Mercor	Delivery of the anonymised dataset	Typically within 2 to 4 weeks of the signed agreement
Ooak	Not stated; paid to the company, not to individuals	2 to 4 weeks
Troveo	Each sale (revenue share)	Signed deal to first payout in under a month, for in-demand datasets
Collinear AI	Delivery confirmed complete	About a week from first call to signature
Turing (Lazarus)	One payment after sample review and final offer	Intake call within one business day
Polyshares	Data approved and anonymised	Offer within one to two weeks of intake; net 30 to 60 days
Mode (BizBuyData)	Final offer accepted, data collected and de-identified	Letter of intent within 14 days; 15 to 30 days of processing; payment expected in about 60 days
micro1	Written acceptance by micro1	30 business days to reject after delivery; silence counts as acceptance; invoice only after acceptance

Source: [1, 5, 6], in the organisations' words as condensed there.

The trigger is as important as the time. Appen states that "payment is on acceptance rather than submission", Centific that payment "is made when the data is accepted", and Nyne that payment follows once the dataset is "delivered and accepted" [1, 5]. micro1's master terms give it 30 business days after delivery to reject the data, deem silence to be acceptance, and allow the seller to invoice only after written acceptance [5]. Under such terms, about six weeks can pass after delivery before the seller may invoice. Revenue-share marketplaces pay only as the data licenses, and SuperAnnotate pays "on delivery, and on each relicense" [1].

3.5 Licence, exclusivity and deletion

The clause register shows which terms are published at all. Of the 16 organisations it covers, 13 publish something about payment, 11 about the licence, 10 about who removes personal data and 5 about resale; but only 4 publish a clause on exclusivity, 2 on sublicensing, 2 on deletion, 2 on refunds, 1 on acceptance and 1 on earlier licences [5]. Ten list exclusivity, and ten the licence term, among the terms they do not publish anywhere the compiler read [5]. Table 4 brings together what the clause register and the buyer table record on exclusivity and deletion for programmes that take company records, with one other organisation for comparison.

Table 4. Published exclusivity and deletion terms

Organisation	Exclusivity	Deletion
micro1	Non-exclusive, but perpetual, irrevocable and sublicensable "through multiple tiers"; the seller warrants no earlier licence of the same data "on the same or substantially similar terms"	Deletion does "not require Licensee to delete or retrain any model, weights, or derivative works"
Labelbox (Prism)	"Licenses are structured as exclusive by default"	Not published
Polyshares	Its own case study describes a "perpetual exclusive license to the operational record" for \$450,000	Not published
SuperAnnotate	Priced on "rarity, depth, and exclusivity"; sellers approve each buyer and can block competitors	"Retained for an agreed period, then deleted"; deletion "honored on request"
Troveo	"Time-limited and it has to be earned"	Not published
Ooak	Not published	"your raw archive is deleted within 30 days"; only the anonymised dataset is retained for the licence term
Grepped	Not published	"Original datasets are deleted once processing is done"
Nyne	Not published	"retention and deletion terms set per engagement, including deletion and suppression workflows"
The General Data Company	Not published	The link between real and invented names is deleted after a person checks the result
Mercor, Handshake AI, Scale AI, Turing, SimpleClosure	Not published	Not published
Google Content Offer Pilot (not in the buyer table)	"This is a non-exclusive opportunity"	Not published

Source: [1, 5, 6] and the pages they cite, as read on 11 to 17 October 2026.

Three patterns follow. First, exclusivity is mostly left to negotiation: among the programmes for company records, one states exclusivity as its default, one prices it, one describes it in a case study, one says it must be earned, one takes a non-exclusive licence with a warranty of no earlier similar licence, and the rest say nothing. Second, deletion terms concern the copy, not the models: the one programme that addresses trained models says deletion does not reach them [5]. Third, some licences reach further than their

exclusivity terms suggest. micro1's licence is non-exclusive but perpetual, irrevocable and sublicensable through several tiers, so the buyer may grant the rights on; SimpleClosure's terms let whoever acquires the assets license or sell them on "without further consent from or compensation"; by contrast Grepped says the data "is never resold to competitors or published", and Defined.ai's licence to its own buyers is "non-exclusive, non-sublicensable, and non-transferable" [5]. Two refund clauses stand out: micro1's terms provide that if the data manifest is materially inaccurate "no amounts are payable" and the seller refunds what was paid, a right that survives acceptance; and Kled, a consumer app whose licence is "exclusive to us", requires a contributor who withdraws consent to refund what was paid [5].

3.6 Introduction fees

Fifteen of the 37 organisations publish something about paying whoever introduces a seller, and ten in a form that can be expressed as numbers (Table 5). All ten take company records; nine buy for their own account and one is an intermediary.

Table 5. Structured introduction fees

Organisation	Terms as published	Condition
Mode (BizBuyData)	Greater of \$50,000 or 6% of the licence fees paid to the company (first five qualified US referrals); 6% outside the US	After the transaction completes
Mercor	5% of the deal, up to \$100,000 (shown in its referral app; its public page names no amount)	After the company's first delivery
HUMXN Research Labs	10% of the deal, up to \$100,000 per deal	Deals signed within 12 months; paid within 30 days of closing
Grepped	10% of what it pays the owner, up to \$100,000; 2% of the fees of referrers a referrer brings	Referrer may not help the owner sell any other way for 12 months after the terms end
micro1	Up to \$50,000 per company; 7.5% of the price where no amount is posted; \$75,000 field on its posting for shut-down companies	Once the referred company's partnership is paid or its data accepted, by type of partnership
Scale AI	Up to \$50,000	Only if Scale signs a partnership agreement
The General Data Company	Up to \$50,000	When a data partnership is signed
Polyshares	6% of the price, paid on top of it, to brokers registered in its portal (as its co-founder states publicly)	First record of the company keeps the credit
Handshake AI	\$10,000 per successful referral	After the partner completes its data transfer
Turing (Lazarus)	Up to \$10,000	Per referral

Source: [1, 6]. Five more publish fees in words only: Nyne (an earlier reported amount is unconfirmed), Sieve ("varies by partnership"), Wirestock (10% of a referred creator's earnings, up to \$25), DataVendor (5% to 10% out of its own share) and SimpleClosure (up to \$1,000 for a referred shutdown, nothing published for data sales) [1].

The terms create incentives a seller should see. Fixed sums are large against the floors the same programmes publish: by the arithmetic of their own figures, Mode's \$50,000 minimum equals half of its \$100,000 floor, and Scale's maximum of \$50,000 is five times its \$10,000 floor [1]. Some terms bind the

introducer to the buyer: Grepped's referral terms forbid a referrer from arranging "the sale or licence of a referred Owner's data other than through Grepped" [6]. Others reward recruitment of further introducers, at 2% (Grepped) or 20% of the recruited partner's payout (micro1) [1, 6]. The terms also move quickly: micro1's referral programme offered up to \$25,000 per company at its launch in June 2026 and \$50,000 by October [6]. Two organisations outside the buyer table publish similar terms: a licensing intermediary pays 10% of the licensing payment up to \$100,000 per business [6], and a smaller licensing firm says it pays introducers 20% of what the business receives [5].

3.7 Advertised ceilings and reported payments

Table 6 sets each of the nine published ceilings beside the strongest evidence of payment in the datasets.

Table 6. Published ceilings and the payments on record

Programme	Published ceiling	Payments on record	Class
Mode (BizBuyData)	\$5,000,000	None; the programme opened on 13 October 2026, and the parent's half-year filing to the SEC does not mention data licensing [9]	D
Handshake AI	\$4,000,000	Its page: "Payments to date have ranged from \$100K to \$4M"; no seller has confirmed a sale	C
Polyshares	\$2,000,000	Its track-record page: four anonymous licences of \$110,000 to \$450,000; no seller confirms them	C
micro1	\$1,000,000+	Its own court filing: more than 50 data transactions in 45 days, with no sellers, prices or customers named; an unsigned \$12.5 million proposal for a bankrupt airline's archive	B
Ooak	\$1,000,000	Its page: deals to date from \$20,000 to several hundred thousand dollars	C
Turing (Lazarus)	\$1,000,000	Its chief executive, as reported: five to ten failed-startup codebases bought at an average in the tens of thousands of dollars	C
Labelbox (Prism)	\$1,000,000+	None in the datasets	D
Scale AI	\$1,000,000+	None; "no seller has confirmed a deal through it"	D
SuperAnnotate	\$1,000,000+	None; "no deal anyone confirms"	D

Source: [1, 6] and the reports and filings they cite [8, 9]. Classes as defined in Section 2.2.

None of the nine ceilings is backed by an independently reported payment of that size. One programme reports its own payments up to its ceiling (Handshake), three report their own payments well below it (Polyshares, Ooak, Turing), one has a court filing that states a count without amounts (micro1), and four have nothing on record.

Across all 24 programmes the picture is similar (Table 7). One programme, the wind-down service SimpleClosure, has payments confirmed by sellers on the record: a founder who netted roughly \$25,000 for a closed startup's code base [10], and a closed transcription company whose former chief executive told the press that its 13-year archive brought "hundreds of thousands of dollars" [11]. Two have statements or bids in a court record without a payment: micro1's filing and Mercor, which is the \$7.5 million alternate

bidder for the airline's archive [12]. Seven report their own payments or revenue, and fourteen have nothing on record.

Table 7. Programmes for company records by the strongest evidence of payment

Class	Programmes	Count
A. Payment confirmed by a seller on the record	SimpleClosure AssetHub	1
B. Statement or bid in a court record, no payment shown	micro1, Mercor	2
C. The organisation's own statement of payments or revenue	Handshake AI, Polyshares, Oak, Turing, Troveo, Grepped, Sunset	7
D. Nothing found	Mode, Scale AI, SuperAnnotate, Labelbox, Nyne, HUMXN, Corpus Bridge, FileYield, Appen, Centific, Collinear AI, Halluminate, Magma, The General Data Company	14
All		24

Source: classified from [1, 6] as described in Section 2.2. Troveo's own total of payments to licensors covers all its lines, and named creators confirm its video payouts, but its line for company records is described as newer and unconfirmed [1].

The largest dollar amounts in court records are bids, not payments: a \$10 million winning bid by a technology company that is not in the buyer table, still awaiting approval, and the alternate bid of \$7.5 million [12]. Headline commitments are larger still, and are not payments either: micro1 has announced a commitment "to spend \$1 billion over the next 12 months" acquiring and licensing operational data, financed with capital whose size and form are not disclosed [6, 13]; Mode's launch release is headlined as a "\$5 Billion" initiative whose body does not say what the figure counts [6]; and HUMXN describes a programme "backed by \$50 million" without naming a backer [6].

The track-record tier should not be read as evidence of payments for records. Twelve of the 24 programmes are rated "confirmed", but the tier can rest on funding from named investors; for nine of those twelve the row's own note says that no seller has confirmed a sale of records, or that the records line is unconfirmed [1].

4. Limitations

Posted terms, not contracts. The datasets record what organisations publish. Contracts can differ, and the terms that matter most to a seller (licence scope, exclusivity, indemnities) are often negotiated per deal and not published.

A moving record. The pages were read between 11 and 17 October 2026 and change often. In that week alone, one programme raised its stated ceiling from \$300,000 to \$1 million and another dropped a statement about who carries the legal risk of a transfer [1, 6]. A later reading may differ.

Coverage. Buyers that acquire records only through private approaches, AI laboratories that buy directly, and any programme the compiler did not find are absent.

Disclosure and promotion. Ceilings and estimator outputs are written to attract sellers. Payments are disclosed when someone chooses to disclose them, and most deals in this market are not disclosed. The absence of a confirmed payment is not evidence that a programme has paid nothing.

Classification. The evidence classes and the track-record tier are the compiler's judgements, applied by stated rules. Eligibility in Table 2 is condensed from each row's wording, and a short form loses detail.

Small numbers. Medians over 14 floors and 9 ceilings are descriptive. They are not estimates of what a company would be paid.

Position of the compiler. The author's organisation acts as an agent for companies that sell or license data, and several organisations in these tables are its potential counterparties or offer similar services. The quote behind every row and the open datasets are intended to let readers check each classification.

5. Conclusion

Thirty-seven organisations now publish a way for others to sell them data, and 24 take ordinary companies' records. Their published figures cluster around a \$100,000 floor and a \$1 million ceiling, with a small number of higher ceilings, and their eligibility rules are few: a headcount of 20 to 40 for most companies where one is stated, three years for two programmes, and country rules that mostly favour the United States, reinforced by estimators that price the rest of the world lower.

The terms that most affect a seller are the least often published. Payment is commonly tied to the buyer's acceptance, exclusivity is rarely stated, the one deletion term that mentions trained models leaves them untouched, and a perpetual, sublicensable licence can matter more than a non-exclusive label. The evidence of payments is thin: no advertised ceiling is matched by an independently reported payment, and the payments sellers have confirmed are in the tens to hundreds of thousands of dollars. Three disclosures would make these markets easier to read: realised payments by programme with the size of each seller, published standard contracts, and a statement of how many applicants each programme has paid.

6. Data availability

The datasets are available under CC BY 4.0 at <https://aidata.broker/datasets/buyers.json>, <https://aidata.broker/datasets/market-log.json> and <https://aidata.broker/datasets/negotiation-levers.json>, with CSV alongside; the published clauses are at <https://aidata.broker/records/clause> and in the offer check at <https://aidata.broker/tools/offer-check> [1, 5, 6, 7]. A snapshot of 18 October 2026 that includes the buyer table, the market log and the levers table is deposited at <https://doi.org/10.5281/zenodo.23452930>; the clause register is not in it. The script that computes the tables is available from the author on request.

7. Conflict of interest

The author is chief executive of AIData.Broker, which compiles and publishes the datasets analysed here and acts as an agent for organisations that sell or license data to AI developers. Several organisations in the tables are potential counterparties of the author's organisation, and some intermediaries in them offer similar services.

8. Funding

The datasets and this analysis were produced by AIData.Broker. No other funding was received.

9. Use of AI tools

Generative AI tools were used to assist with the scripted analysis, with checking the sources and with drafting this manuscript. The author reviewed the text, the tables and every cited source and takes responsibility for the content.

References

All URLs were accessed on 18 October 2026.

[1] AIData.Broker, "Buyers of AI training data" (dataset, JSON, reviewed 16 October 2026), CC BY 4.0. <https://aidata.broker/datasets/buyers.json> ; browsable at "Who buys AI training data, and what each one publishes," <https://aidata.broker/buyers> . The rows cite the organisations' own pages, among them: Mode (<https://bizbuydata.com/50k>), Handshake AI (<https://joinhandshake.com/ai/data-partnerships/>), Polyshares (<https://polyshares.com/> and <https://polyshares.com/track-record>), SuperAnnotate (<https://www.superannotate.com/data-partnerships>), micro1 (<https://www.micro1.ai/data-partnerships>), Labelbox (<https://prism.labelbox.com/>), Oak (<https://www.oakdata.com/partner>), Scale AI (<https://scale.com/data-partnership>), Turing (<https://lazarus.turing.com/>), HUMXN (<https://humxn.io/sell-company-data>), Nyne (<https://nyne.ai/sell-company-data/>), Troveo (<https://www.troveo.ai/business-data>), Corpus Bridge (<https://corpusbridge.com/>), SimpleClosure (<https://simpleclosure.com/blog/insights/what-your-assets-are-worth/>), Appen (<https://www.appen.com/enterprise-data>), Centific (<https://www.centific.com/enterprise-data>), Collinear AI (<https://collinear.ai/data-acquisition>), FileYield (<https://fileyield.com/sell-data>), Grepped (<https://www.grepped.ai/business/>), Halluminate (<https://www.halluminate.ai/data-partnerships>), Magma (<https://magmahq.ai/>), Mercor (<https://mercor.com/data>), Sunset (<https://sunsethq.com/product>) and The General Data Company (<https://thegeneraldata.com/data-partnerships/>).

[2] AIData.Broker, "micro1, Handshake, Turing, Mercor: Data Programs Compared," guide, 17 October 2026. <https://aidata.broker/guides/company-data-buyer-programs-compared>

[3] AIData.Broker, buyer board page, "Who buys AI training data, and what each buyer pays," updated 16 October 2026. <https://aidata.broker/buyers>

[4] AIData.Broker, "Who sells AI training data, and where each gets it," vendor register, updated 15 October 2026. <https://aidata.broker/vendors>

[5] AIData.Broker, buyer terms: the clauses 16 organisations publish, each with its quote, page and reading date (11 to 17 October 2026). <https://aidata.broker/records/clause> ; also in the offer check, <https://aidata.broker/tools/offer-check> . Clauses quoted here come from, among others: micro1, "Data License Master Terms," <https://www.micro1.ai/legal/data-license-master-terms> ; Grepped, <https://www.grepped.ai/business/> ; Nyne, <https://nyne.ai/data-partnerships> and <https://nyne.ai/sell-company-data/> ; Oak, <https://www.oakdata.com/partner> ; SimpleClosure, "Terms of Service," <https://simpleclosure.com/tos/> ; Defined.ai, "Data License Agreement," <https://defined.ai/data-license-agreement> ; Kled, "Terms of Service (mobile)," <https://www.kled.ai/terms-of-service-mobile> ; Protege, broadcaster case study, <https://withprotege.ai/articles/blog/media-partner-case-study-broadcaster-earns-1m->

in-6-months/ ; Troveo, <https://www.troveo.ai/business-data> ; Google, Content Offer Pilot, <https://contentpilot.google.com/> ; Telegraph Lab, <https://telegraphlab.com/partners> ; Scale AI job posting, <https://job-boards.greenhouse.io/scaleai/jobs/4725509005> .

[6] AIData.Broker, "Market log of company-data sales" (dataset, JSON, reviewed 17 October 2026), CC BY 4.0. <https://aidata.broker/datasets/market-log.json> . Rows used include micro1's referral terms (<https://www.micro1.ai/legal/referral-terms> ; <https://www.micro1.ai/forum/company-data-partnerships-referral-program>), its posting for shut-down companies (<https://req.micro1.ai/post/681c84e1-2628-4382-aeca-a50433f2f72d>), Grepped's referral terms (<https://grepped.ai/assets/grepped-referral-terms.pdf>), Pewter's partner terms (<https://pewterdata.com/partners/terms>), Mode's launch release (Business Wire, read on Yahoo Finance, <https://finance.yahoo.com/technology/ai/articles/mode-inc-launches-5-billion-151600765.html>) and HUMXN's release (<https://humxn.io/humxn-research-labs-wants-to-buy-50-million-worth-of-skilled-trades-and-manufact>).

[7] AIData.Broker, "Negotiation levers" (dataset, JSON), records three-buyer-estimators-share-one-formula, public-estimators-differ-twentyfold and location-and-language-discount, with the estimators they quote: micro1 (<https://www.micro1.ai/data-payout>), Handshake AI (<https://joinhandshake.com/ai/data-partnerships/>), Nyne (page code), Grepped, Oak and Troveo (pages in [1]).
<https://aidata.broker/datasets/negotiation-levers.json>

[8] 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/>

[9] Mode Mobile, Form 1-SA for the half year ended 30 June 2026, U.S. Securities and Exchange Commission. <https://www.sec.gov/Archives/edgar/data/1748441/000149315226046559/form1-sa.htm>

[10] MIT Technology Review, "AI labs are buying data from dead startups to train their models," 13 October 2026. <https://www.technologyreview.com/2026/10/13/1145755/ai-labs-buying-data-from-dead-startups-to-train-models/>

[11] Forbes, "AI's new training data: your old work Slacks and emails," 16 April 2026. <https://www.forbes.com/sites/annatong/2026/04/16/ais-new-training-data-your-old-work-slacks-and-emails/> ; and SimpleClosure, customer story: cielo24, <https://simpleclosure.com/customer-stories/cielo24/>

[12] In re Spirit Aviation Holdings, Inc., No. 25-11897 (Bankr. S.D.N.Y.), Dkt. 1463, Notice of Auction Results, Exhibit A (Sale Agreement) 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>

[13] micro1, "micro1 commits \$1B to enterprise data acquisition," company blog, 9 October 2026. <https://www.micro1.ai/blog/micro1-commits-1b-to-enterprise-data-acquisition>

[14] AIData.Broker, "The AI training data market: prices, deals, buyers and cases (snapshot of 18 October 2026)" [Data set], Zenodo. <https://doi.org/10.5281/zenodo.23452930>

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). Buyer programmes for company data: eligibility, prices and payment terms of 37 organisations, October 2026. AIData.Broker, 18 October 2026. <https://aidata.broker/research/paper-buyers>