

Data-Licensing Businesses: Royalty Valuation for AI Training and Inference

A Transaction Framework for Volume, Exclusivity, Refresh and Audit Rights

Authored by Chennakeshav Adya

Independent Researcher

September 2026

Abstract

Data-licensing businesses sit between raw information and artificial-intelligence systems that depend on reliable rights, coverage and continuing access. Their value cannot be read from dataset size or a headline royalty rate. It depends on the legal and technical definition of the licensed asset, the uses granted to each customer, the volume actually consumed, the degree of exclusivity, the frequency and quality of refresh, the strength of provenance, and the licensor's ability to measure and audit use. A contract written for a static database can fail when a customer trains a model once, deploys the resulting weights repeatedly and creates derived outputs that persist after the licence ends.

This paper develops a transaction framework for valuing businesses that license data for AI training, evaluation, retrieval and inference. It separates the corpus, rights record, curation layer, delivery system, refresh operation, usage-control system and customer contract. It then links those assets to billable units, renewal behaviour, gross margin, legal durability and collected cash. Cost, income and market approaches are applied through an evidence-weighted reconciliation rather than a single royalty benchmark.

The analysis draws on WIPO guidance on intellectual-property valuation and licensing; IFRS requirements for intangibles, business combinations, fair value and revenue; OECD work on data value and hard-to-value intangibles; the United States Copyright Office report on generative-AI training; European Union data, copyright, privacy and AI legislation; UK regulatory guidance; NIST and ISO governance standards; and competition and consumer-protection materials [1-50]. These sources define relevant questions and controls. They do not determine ownership, lawfulness, fair value, tax treatment or an arm's-length royalty for a particular company.

An illustrative acquisition demonstrates the method. A target reports USD 84 million of revenue and USD 18 million of EBITDA. Management assigns USD 96 million of enterprise value to the data-licensing platform. The framework produces a USD 31 million replacement-cost indication, a USD 72 million income indication and a USD 55 million to USD 90 million market corridor. An evidence-weighted conclusion of USD 67 million reflects rights coverage, measured use, customer concentration, refresh obligations, auditability and the useful life of the underlying corpus. Every amount is a management assumption used solely to demonstrate the framework.

The central conclusion is that royalty value resides in an enforceable and renewable right-to-cash system. Dataset scale, apparent uniqueness and customer logos are inputs. Sustainable value arises when the licensor can prove what it controls, define permitted use, measure consumption, refresh the asset, detect leakage, enforce restrictions and retain customers whose economics justify the price. Gaps should be converted into price adjustments, escrow, indemnities, earn-outs, licence redesign and funded post-close controls.

JEL Classification: G24, G34, K11, K24, L86, M41, O34

Keywords: data licensing, royalty valuation, AI training, AI inference, exclusivity, refresh rights, audit rights, data provenance

Introduction

AI systems consume data in several economically distinct ways. A developer can use records to train or fine-tune a model, evaluate performance, build a retrieval index, populate a knowledge service, generate embeddings or support repeated inference. Each use can create a different duration of benefit. Training may create model weights that continue to generate value after the source files are deleted. Retrieval may require continuing access to current records. Evaluation can be episodic. Inference can turn a licensed source into millions of customer-facing outputs.

These distinctions change valuation. A fixed annual fee may be reasonable for a bounded reference product and inadequate for a corpus that supports a global model. A usage royalty can align consideration with adoption and remain difficult to verify when the contract lacks measurement architecture. Exclusivity can protect a customer's advantage and destroy the licensor's remaining market. Refresh commitments can strengthen recurring revenue and create a costly obligation to acquire, clean and validate new information.

Copyright, database rights, contract, confidentiality, privacy, competition rules and sector regulation can overlap. The United States Copyright Office's work on generative-AI training records substantial disagreement over fair use, licensing feasibility and market development [13-15]. The European Union combines copyright exceptions and reservations, data-access rules, privacy law and AI obligations [17-22]. The UK Information Commissioner's Office treats large-scale web scraping for model development as high-risk processing and expects controllers relying on legitimate interests to evidence necessity and balance [23-25]. A valuation should therefore start with the rights actually held and the uses actually granted.

This paper is designed for boards, founders, corporate-development teams, investors, lenders, valuation specialists and licensing executives. It provides a decision system for transactions and commercial redesign. It does not provide legal, accounting, tax, data-protection, competition, technical-security or valuation advice. Specialists should review the relevant assets, contracts, jurisdictions and facts.

1 Define the licensed asset before selecting a royalty

A data business is rarely one asset. The corpus may contain original observations, licensed third-party records, public material, customer-contributed information and synthetic or inferred fields. A curation layer can add taxonomies, labels, entity resolution, confidence scores and domain rules. Software can provide search, delivery, APIs, embeddings, access control and metering. The company may also possess contracts, provenance evidence, contributor relationships and operating know-how.

The first diligence task is to identify the unit of account. A licence may grant access to a named dataset, a field subset, an API response, a continuous feed, a derived indicator or a managed service. The legal right and the technical delivery object can differ. A customer may never receive source records and can still obtain a commercially valuable output. Conversely, possession of files does not establish the right to grant training, derivative or sublicensing rights.

The buyer should map every component to control, transferability and income. Identifiability and control are relevant to accounting under IAS 38 and IFRS 3 [1-2]. WIPO's valuation materials emphasise legal protection, transferability, future benefits and separable income [5-9]. Transaction value can include assets that are not separately recognised in the seller's balance sheet, yet valuation should still avoid double counting the same cash flow across data, software, brand and customer relationships.

The product catalogue should be reconciled to the asset register. Commercial names often persist while their contents, sources and rights change. The buyer should select material products and compare the

current customer description, contractual schedule, technical schema and released version. A mismatch can cause customers to receive fields or uses that the rights register does not support. It can also conceal valuable assets that have never been priced separately.

Control should be tested in practice. The diligence team should ask who can export the corpus, alter labels, approve a new source, create a derivative feed and grant a customer exception. It should inspect access records and approval history. A company can possess strong contracts while allowing uncontrolled internal copying. It can also operate disciplined technical controls with incomplete contributor terms. Both dimensions affect transferability and remediation cost.

Table 1 Data-licensing asset and rights map

Component	Evidence to inspect	Economic role	Critical valuation question
source corpus	acquisition contracts notices permissions lineage and retention rules	supplies observations and coverage	can the buyer continue every material use after closing
rights record	licences reservations consents jurisdiction and change-of-control terms	defines enforceable commercial perimeter	which rights are owned controlled time-limited or revocable
curation layer	taxonomies annotations validation rules version history and contributor terms	increases usability and differentiation	how much advantage is repeatable without current specialists
delivery system	APIs files indexes embeddings access controls and service levels	converts assets into customer use	is delivery scalable secure measurable and transferable
refresh operation	sourcing cadence cleaning acceptance tests and release records	maintains relevance and renewal	what cash and rights are needed to sustain promised freshness
usage-control system	identity metering logs audit reports and anomaly detection	supports billing and enforcement	can the licensor prove billable use and detect leakage
customer contracts	fields purposes territories duration pricing and termination	converts rights into cash	do terms match actual use and reported revenue quality

Proposed diligence record; legal accounting tax privacy and technical conclusions require specialist review.

An asset register should show dependencies and substitutes. A valuable corpus can depend on one contributor, one public authority or a revocable platform interface. A taxonomy may be openly replicable. A customer list can carry value while contracts remain terminable at short notice. The model should identify the cash lost if each dependency disappears and the cost and time required to replace it.

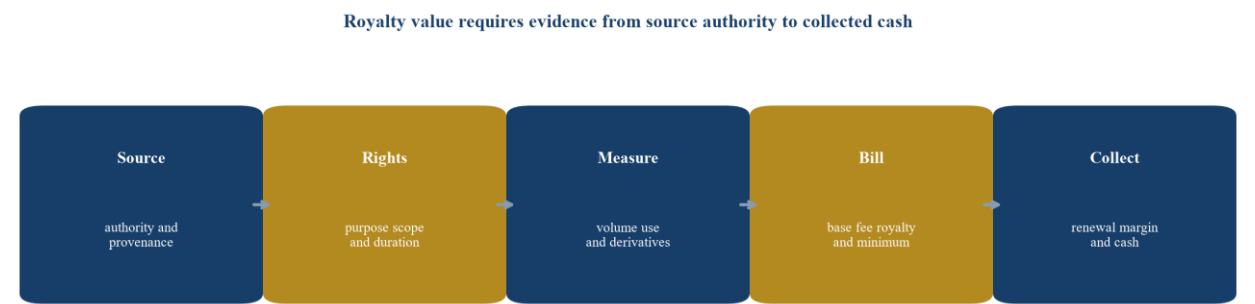
2 Separate training, evaluation, retrieval and inference rights

The label AI use is too broad for a defensible licence. Training changes model parameters. Fine-tuning adapts an existing model. Evaluation tests behaviour. Retrieval creates an index or service that consults information at run time. Inference generates outputs from deployed models. Each activity uses the data differently and can leave a different residual asset with the licensee.

A licence should define whether the customer may create embeddings, labels, features, summaries, models, checkpoints, benchmarks and other derivatives. It should address whether a model trained on the data can remain in service after expiry, whether weights can be transferred within a group, whether outputs can reveal source material and whether the customer can use the data to train a competing data product. The answer directly affects economic duration.

The valuation team should build a rights matrix rather than rely on contract titles. The matrix should link field, territory, entity, purpose, model class, environment, duration, permitted derivatives, retention, audit and termination. It should compare written rights with technical logs and product documentation. Undocumented uses can represent unbilled value, contractual breach or both.

Figure 1 Rights-to-cash evidence chain



Proposed transaction framework; every connection requires asset-specific evidence.

Table 2 AI use-rights matrix

Use	Persistent customer asset	Useful billing evidence	Contract question
pre-training or continued training	model weights checkpoints and learned parameters	training runs tokens records epochs and compute environment	may the trained model operate after licence expiry
fine-tuning	adapted weights and task-specific behaviour	records examples jobs and model identifiers	which models entities and fields are covered
evaluation	scores test results and benchmarks	evaluation jobs dataset versions and reports	can results be published or used to market a model
retrieval	index embeddings cache and citations	indexed records queries responses and active users	must records remain current and attributable
inference	outputs decisions and customer interactions	calls tokens outputs seats or economic events	does the royalty follow calls revenue or deployed models
derived products	labels features analytics and transformed datasets	creation events exports and downstream licences	when does a derivative become a competing substitute

Illustrative contract architecture; counsel should draft terms for the relevant rights and jurisdictions.

The matrix also clarifies termination. Deletion of source files may leave models, indexes and outputs intact. A buyer should identify survival rights, certification requirements, unlearning obligations, transition periods and technical feasibility. Expected post-termination value belongs in the pricing model when the contract grants it. Expected remediation cost belongs in the liability analysis when use must cease.

3 Establish lawful provenance and enforceability

Provenance is a chain of evidence, not a spreadsheet field. It should connect each material record category to its source, acquisition method, permission, restrictions, transformations, current storage and licensed

output. The chain should also identify notices, opt-outs, copyright reservations, database terms, privacy bases and contributor warranties. A buyer should be able to sample from a customer-facing field back to source authority.

The diligence team should segment the corpus. Original observation, commissioned collection, purchased licence, public-sector information, open licence, web-scraped material and customer-supplied data create different rights and risks. Rights can also vary by jurisdiction, date and use. The EU copyright framework includes text-and-data-mining provisions and rights reservations [17]. The Data Act governs aspects of access and use for connected-product and related-service data [18]. GDPR and the EDPB opinion on AI models make personal-data questions fact-specific [21-22].

The US Copyright Office report explains that the availability and practicality of licensing can differ across works and markets [14]. It also records existing collective and direct licensing activity while declining to treat one mechanism as universally appropriate. A transaction model should therefore avoid the assumptions that all training requires a licence or that no licence has value. Counsel should determine the applicable legal position; the valuation should model the cash consequences of each supported rights case.

Sampling should be risk-based. High-revenue products, large source categories, sensitive fields, exclusive grants and rapidly growing uses deserve priority. A sample should include the full chain from source agreement to customer delivery. Exceptions should be expanded until the team can estimate their population effect. Management representation can explain an exception and does not replace the underlying document or system record.

Change of control requires separate attention. Source licences can prohibit assignment, permit termination or require consent. Public-sector and consortium arrangements can include eligibility conditions. Customer licences can contain service continuity, notification and security obligations triggered by ownership change. The model should include the timing, cost and probability of consents as well as the cash effect of a failed consent.

Enforceability has a technical dimension. A contract can prohibit unauthorised use while the licensor lacks logs, identifiers or audit access. Watermarking, canary records, signed releases, customer-specific keys and monitored endpoints can strengthen evidence. These controls do not establish legal liability by themselves. They increase the ability to identify which asset and customer activity require investigation.

Provenance gaps should be quantified. The buyer can classify revenue by verified, remediable, restricted and unsupported rights. Unsupported revenue should not receive the same probability or multiple as income backed by a complete chain. Remediation should include licence acquisition, re-collection, deletion, retraining, customer consent, product redesign and potential interruption.

4 Design the billable unit and volume curve

Volume can be measured in records, fields, files, terabytes, API calls, tokens, training jobs, models, seats, territories, deployments or downstream revenue. Each unit aligns with value differently. Terabytes can reward low-value duplication. Records can differ greatly in scarcity and usefulness. Tokens may be measurable and only loosely connected to customer economics. Revenue royalties align with commercial success and require access to financial evidence.

The billable unit should satisfy four tests. It should be observable, attributable to the customer, difficult to manipulate and connected to economic use. The contract should define measurement boundaries, retries, cached calls, batch processing, internal testing, affiliates, contractors, failed requests and derivative use. It should also state who operates the meter and how discrepancies are resolved.

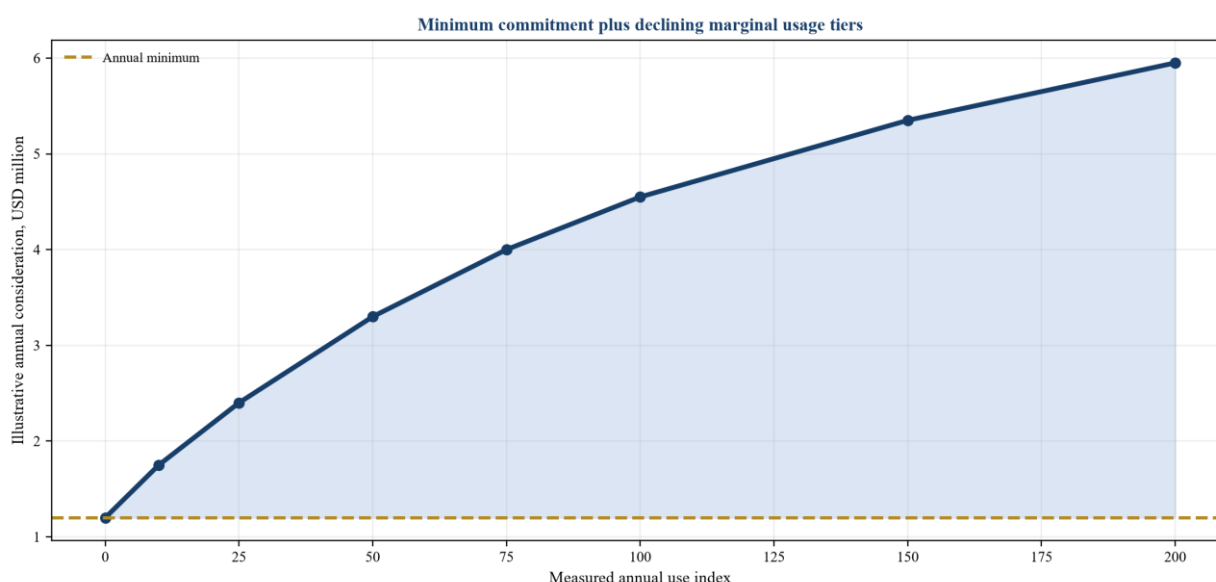
Unit definitions should remain stable across product versions. If a record becomes a document, a document becomes tokens or an API changes pagination, customers can experience an unexpected price change even when their economic use is constant. The licence should include a conversion rule, notice

period and dispute process. Finance should preserve parallel measures during migration so that revenue movement can be separated from metering design.

The seller should demonstrate completeness as well as accuracy. Completeness tests reconcile source-system events to the meter and the meter to invoices. Accuracy tests recompute selected bills from raw logs and contract terms. Access-control tests confirm that events belong to the named customer and licensed product. Exception analysis should cover shared credentials, offline exports, bulk deliveries, test environments and manually adjusted invoices.

A pricing curve can combine a minimum commitment, included volume and marginal tiers. Minimums fund sourcing, availability and assurance. Usage tiers capture adoption and can decline with scale where marginal delivery cost falls. Premiums can apply to scarce fields, low-latency delivery, broader rights or higher assurance. Floors protect the licensor from customers who reserve capacity or exclusivity without using it.

Figure 2 Illustrative volume and royalty curve



Management assumptions solely for method demonstration; amounts are not market benchmarks.

Volume evidence should reconcile contract, meter, invoice, general ledger and cash. Exceptions need explanation. A company can report rapidly growing usage while contract caps prevent additional revenue. It can report annual recurring revenue that includes non-recurring backfiles or implementation. A buyer should cohort customers by pricing model and test realised price per unit, minimum utilisation, overage collection, credits, disputes and renewal.

Inference creates an additional challenge. The source data may influence outputs without being queried directly. Where direct metering is impossible, the parties can use deployed-model counts, active users, revenue share, capacity bands or agreed proxies. The proxy should be reviewed as the product changes. A fixed fee may be the sound choice when measurement cost and dispute risk exceed the value of precision.

5 Price exclusivity and field-of-use restrictions

Exclusivity transfers opportunity from the licensor to the customer. Its value depends on the market actually foreclosed, the duration, field, territory, customer class, model class and availability of substitutes. A global exclusive grant across all AI uses has a different economic effect from a six-month lead in one therapeutic area or one country.

The licensor should model the counterfactual portfolio of customers. Expected lost contribution from excluded customers, strategic option value and concentration risk form the starting point. The customer

may value exclusivity through higher revenue, lower competition, faster market entry or reduced input scarcity. The negotiated premium can fall between licensor loss and customer gain, adjusted for bargaining power and evidence.

Exclusivity should carry performance obligations. Minimum annual consideration, launch milestones, use requirements and reversion can prevent warehousing. Carve-outs can preserve academic use, public-interest use, existing customers, non-competing fields or aggregated products. A most-favoured term may affect pricing across the portfolio even when it is not labelled exclusive.

Table 3 Exclusivity adjustment matrix

Dimension	Narrow grant	Broad grant	Valuation effect to test
duration	short lead period with automatic expiry	multi-year or perpetual	foregone renewal options and technology change
field	named use case or model	all AI and analytical uses	number and value of excluded markets
territory	one country or customer market	global	geographic expansion and enforcement cost
customer scope	one legal entity	affiliates partners and customers	unmetered propagation and concentration
substitutes	several comparable sources	scarce or unique corpus	customer advantage and licensor bargaining power
performance	minimums and reversion	passive reservation	risk of value sterilisation

Proposed decision matrix; percentage adjustments require transaction-specific cash-flow evidence.

The buyer should inspect side letters and commercial practice. A nominally non-exclusive licence can become functionally exclusive when the company promises delayed delivery to competitors, restricts fields for strategic customers or lacks capacity to serve others. A broad grant to an early customer can also impair future fundraising or acquisition value. These burdens belong in the rights register and forecast.

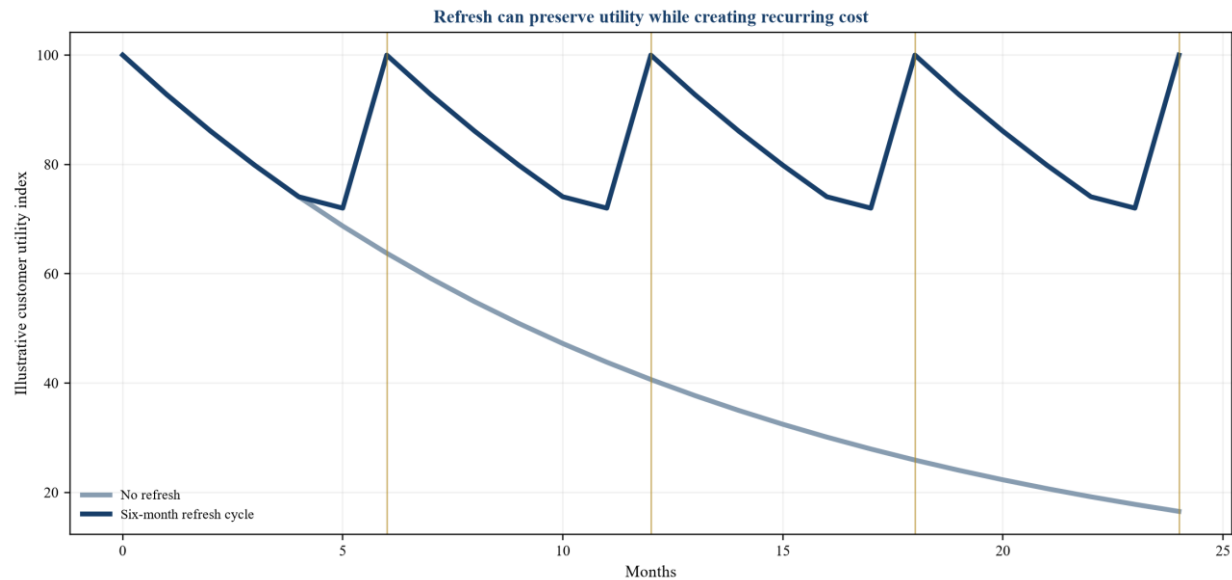
6 Value refresh, maintenance and economic decay

Data can decay because the represented population, products, language, prices, behaviour, regulation or physical system changes. The rate varies by use. Historical geological observations can remain useful for decades. fraud patterns, product catalogues and market prices can change quickly. A valuation should estimate economic life by field and customer decision rather than use one amortisation period for the entire corpus.

Refresh is an operating system. It includes sourcing, permission, ingestion, cleaning, reconciliation, labelling, quality testing, documentation, release and customer migration. The cost should be separated from one-time corpus creation. A company may report high gross margin while capitalising or excluding essential refresh labour. The model should include the cash required to keep the licensed promise current.

Refresh can support recurring revenue when customers need continuity, comparable versions and rapid updates. The buyer should test release punctuality, defect rates, lineage, customer adoption and price realisation. A promised monthly feed that arrives late or changes definitions without a bridge can reduce customer trust and switching cost.

Figure 3 Illustrative decay and refresh economics



Management assumptions solely for method demonstration; actual decay and refresh value require customer-specific evidence.

Table 4 Refresh economics and evidence

Refresh driver	Evidence	Cash-flow effect	Diligence response
source availability	contributor renewals API continuity and collection rights	acquisition cost and interruption risk	probability-weight future releases
data drift	field-level stability and customer task performance	renewal and product utility	model economic life by field
processing effort	labour compute error correction and validation	gross margin and working capital	normalise recurring maintenance cash
release quality	timeliness defects rollbacks and documentation	service credits and retention	cohort acceptance and renewal
schema continuity	version bridges identifiers and deprecation policy	customer migration cost	value switching cost only when evidenced
regulatory change	notices permissions deletion and access restrictions	remediation and product redesign	create jurisdiction-specific scenarios

Proposed operating review; the useful life and required cadence should be tested for each licensed field.

The useful-life assumption should connect to refresh economics. A corpus with high decay can still be valuable when the company has durable sourcing relationships and an efficient refresh process. A static corpus with low decay can be impaired by a rights expiry. The value belongs to the combined system, with each dependency tested separately.

7 Build audit, reporting and enforcement into the product

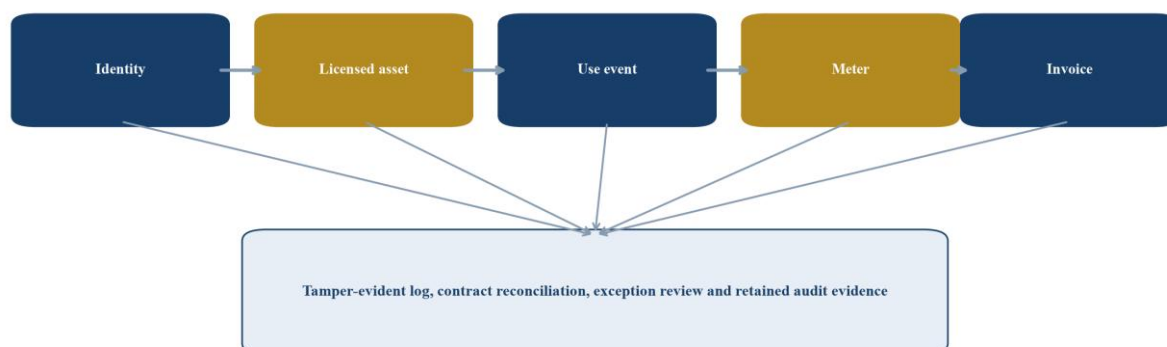
Royalty contracts require evidence. Audit rights written in a contract provide limited protection when the licensor cannot identify use. Metering should record customer, credential, dataset version, permitted purpose, model or application where feasible, volume unit, time, location and export. Records should be tamper-evident, retained for the contractual period and reconciled to invoices.

Customer reporting should match the economic trigger. A revenue royalty requires definitions of gross and net revenue, bundled products, discounts, refunds, related parties, currency and allocation. A model-

based fee requires a registry of deployed models and descendants. A training fee requires agreed treatment of reruns, experiments and checkpoints. The parties should define error thresholds and cure procedures.

Audit design should be proportionate. Continuous technical reporting can protect both parties and reduce intrusive financial audits. Independent assurance can test the metering system, access controls and reconciliation. The licence should address notice, frequency, auditor confidentiality, record access, cost allocation, underpayment thresholds and survival after termination.

Figure 4 Usage measurement and audit architecture



Proposed control architecture; implementation should reflect product design privacy security and contract obligations.

Enforcement should be modelled as expected cash and cost. Detection probability, contractual remedies, jurisdiction, customer solvency, evidence quality and relationship impact matter. A punitive remedy with poor detection and collection can be worth less than accurate real-time billing. The objective is to align use and consideration while preserving a workable customer relationship.

8 Reconstruct revenue quality and royalty economics

Reported revenue should be rebuilt from customer contracts and cash. The analysis should separate backfile licences, subscriptions, usage royalties, exclusivity premiums, refresh fees, implementation, custom collection, analytics and support. Each stream has a different renewal basis, delivery cost and recognised asset.

IFRS 15 establishes a framework for identifying contracts and performance obligations, determining transaction price, allocating it and recognising revenue when obligations are satisfied [4]. Sales-based or usage-based royalties associated with licences of intellectual property receive specific treatment [44-45]. Accounting conclusions require the actual arrangement. Transaction diligence should still reconstruct the commercial pattern even when the financial statements comply.

The buyer should examine contract assets, deferred revenue, unbilled use, minimum commitments, credits and true-ups. Annual recurring revenue should exclude one-time backfills and services unless a supported recurring component exists. Usage growth should be reconciled to realised price and collections. A customer that increases calls while moving into lower tiers may contribute less incremental gross profit than headline volume suggests.

Revenue recognition and cash economics should be presented side by side. A multi-year prepaid licence can produce strong near-term cash and a continuing delivery obligation. A usage contract can show rising accrued revenue while invoices remain disputed. A minimum commitment can protect revenue and create

future credits that reduce later billing. The diligence model should show booking, recognised revenue, invoice, collection, deferred balance, usage and remaining obligation for each material contract.

Renewal analysis should distinguish contractual renewal from commercial continuation. An automatically renewed agreement can have declining use and weak collection. A customer that signs a new contract after a competitive review provides stronger evidence of willingness to pay. The buyer should inspect churn reasons, downsell, field reductions, expanded rights, price changes, concessions and the time required to obtain customer approval.

Table 5 Illustrative revenue-quality bridge

Revenue component	Reported revenue	Recurring or renewable evidence	Normalised revenue used in valuation
annual corpus access and minimums	31	contract renewal and collection history	29
metered training and inference use	24	reconciled logs invoices and cash	21
refresh and maintenance	12	release acceptance and renewal	11
exclusivity premiums	7	term-limited with performance conditions	4
backfiles and historical archives	5	episodic	1
implementation and custom collection	5	labour-dependent	3
total	84	evidence-weighted reconstruction	69

Management assumptions solely for method demonstration; USD million unless stated.

Customer concentration deserves a rights-based view. A large customer can provide validation and create broad restrictions, preferential pricing or dependency. The analysis should measure revenue, contribution, receivables, unbilled use, exclusivity burden and source dependencies for each major customer. Renewal probability should reflect customer use and switching evidence rather than logo prominence.

Gross margin should include source royalties, contributor payments, cloud delivery, metering, refresh labour, quality assurance, privacy and security controls, customer support and dispute resolution. Excluding these costs can turn a data operation into an apparent software business. The forecast should show the investment required to maintain rights and utility.

9 Develop a replacement-cost indication

Replacement cost asks what a capable market participant would spend today to create an asset with equivalent utility, rights and readiness. Historic development cost is a starting record and can include failure, obsolete work and founder effort that a replacement team would organise differently. Cost does not establish value when income is weak or rights are defective.

The estimate should include lawful source acquisition, collection, contributor contracting, cleaning, annotation, taxonomy, quality testing, provenance records, privacy review, security, delivery software, metering, customer validation and time to commercial readiness. Opportunity cost during the build period can matter. Economic obsolescence should reflect substitutes, lower-cost acquisition and changes in customer demand.

The replacement object must be defined carefully. Recreating files without contributor relationships, refresh capability or accepted customer workflows may not replace the business. Recreating every historical record may also be unnecessary when a smaller current corpus produces equivalent customer outcomes. The cost model should target equivalent service potential.

Table 6 Illustrative replacement-cost indication

Component	Current replacement expenditure	Obsolescence or reuse adjustment	Indicated value
source acquisition and contributor rights	14	20%	11.2
cleaning curation taxonomy and labels	11	25%	8.3
provenance privacy and legal evidence	5	10%	4.5
delivery metering and access controls	8	35%	5.2
customer validation and commercial readiness	6	70%	1.8
total replacement-cost indication	44	blended	31.0

Management assumptions solely for method demonstration; USD million and rounded.

The indicated USD 31 million is a management assumption. It does not include the full value of profitable customer relationships or a strategic premium. It also does not validate the underlying rights. If a key source cannot be replaced lawfully, the cost approach may understate scarcity and overstate reproducibility at the same time. The income and market evidence should therefore be considered independently.

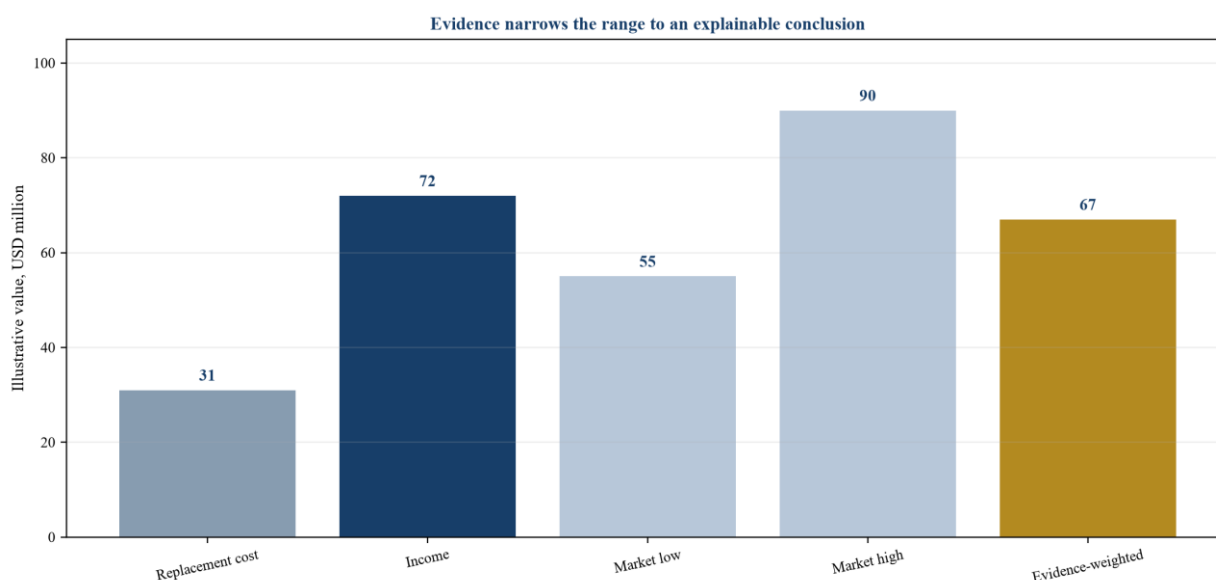
10 Build the income and relief-from-royalty indications

The income approach forecasts cash attributable to the licensed asset and discounts it for time and risk. The model should start with normalised revenue by right and use, deduct source and delivery costs, fund refresh and assurance, and reflect customer attrition, price, volume, exclusivity and legal scenarios. Cash flows associated with software, brand and customer relationships should not be counted twice.

Relief from royalty estimates the hypothetical payment avoided by owning an asset rather than licensing it. It requires an appropriate royalty base, rate, tax effect, useful life and discount rate. A headline comparable rate is insufficient. The analyst should compare rights, exclusivity, territory, use, maturity, contribution of other assets and bargaining conditions. WIPO describes income, market and cost approaches and the need for judgement and evidence [5-8].

For a data-licensing company, a multi-period excess earnings method can isolate customer relationships while a relief-from-royalty method addresses a corpus, taxonomy or brand. The business forecast can also be valued directly through discounted cash flow. Reconciliation should explain which cash flows each method captures and how contributory assets are charged.

Figure 5 Evidence-weighted valuation reconciliation



Management assumptions solely for method demonstration; values are not transaction advice or market benchmarks.

Table 7 Illustrative income and transaction case

Item	Base assumption	Downside test	Diligence evidence required
reported revenue	84	84	contracts invoices ledger and cash
normalised renewable revenue	69	58	use logs renewals and exclusion of episodic items
sustainable EBITDA	18	11	source refresh delivery assurance and service costs
five-year revenue growth	12%	4%	customer cohorts contracted minimums and pipeline conversion
terminal growth	3%	2%	useful life refresh capability and market substitutes
discount rate	18%	26%	rights concentration customer risk and execution
income indication	72	41	probability-weighted free cash flow
evidence-weighted conclusion	67	46	reconciliation of cost income market and risk

Management assumptions solely for method demonstration; USD million except percentages and multiples.

The USD 72 million income indication assumes that verified renewable revenue converts into cash after required refresh and control expenditure. The downside case reflects weaker rights coverage, lower renewal, customer concentration and higher discounting. The USD 67 million conclusion gives less weight to the USD 96 million management allocation because evidence has not supported the full value claimed.

11 Use market evidence without importing false precision

Comparable data licences are difficult to interpret. Publicly reported arrangements may omit scope, exclusivity, minimums, source costs, derivative rights, audit terms and strategic relationships. Company transaction multiples can combine software, customer relationships, services and data. A market approach should record differences explicitly rather than convert sparse observations into a precise rate.

Market observations should be stored in a structured comparison record. The record should identify the date, parties, asset, rights, use, geography, duration, exclusivity, consideration, minimums, variable terms, obligations and source quality. Missing fields should remain visible. A press announcement can confirm that an arrangement exists and may provide little evidence about the effective royalty. A complete agreement can still require interpretation because consideration can compensate for several assets and services.

Bid and negotiation evidence can be useful. A rejected proposal may show that a buyer did not accept a price under the offered terms. It does not prove the asset's value without understanding alternatives, budget, timing and negotiation strategy. Executed renewals, collected true-ups and independently audited use generally provide stronger evidence of realised economics.

The analyst should prioritise arm's-length agreements with comparable use, rights and economics. Adjustments can address exclusivity, territory, duration, maturity, renewal, update obligations, customer bargaining power and the contribution of software. WIPO's market-approach guidance emphasises comparability and adjustments [8]. OECD guidance on intangibles and hard-to-value assets highlights uncertainty and the need to connect returns to value-creating functions and risk control [11-12].

Internal licences can provide evidence when negotiated on commercial terms, yet transfer-pricing policy and group incentives require examination. A customer renewal can be more informative than a distant industry rate because it reveals willingness to pay for the actual product. Bids lost on price, discounts, true-ups and exclusivity negotiations also show the boundary of value.

The illustrative market corridor of USD 55 million to USD 90 million should therefore be treated as a range. The lower end reflects narrower rights, concentration and refresh cost. The upper end assumes strong provenance, measurable use, enforceable restrictions and durable renewal. The valuation conclusion should move within the corridor as evidence improves.

12 Convert uncertainty into transaction and licence terms

A valuation gap does not always require a single negotiated price. It can be allocated through structure. Verified rights and contracted minimums can be paid at closing. Consideration tied to future use, renewal or licensing milestones can be deferred. Specific provenance defects can be covered by escrow, indemnity, remediation covenants or exclusion from the acquired perimeter.

Earn-outs should use metrics the parties can measure and control. Revenue can be affected by bundling and discounting. EBITDA can be affected by integration choices. Usage can grow without cash. A combination of collected licence revenue, defined gross margin, named renewals and rights milestones can better align value. Definitions, accounting policy, customer allocation, audit and dispute resolution should be written before closing.

The acquisition agreement should address ownership and authority, contributor and source contracts, privacy compliance, infringement claims, customer use, open-source or public-source obligations, model derivatives, security incidents, audit records and regulatory correspondence. Representations do not replace diligence. Their financial value depends on survival, caps, exclusions, credit support and the ability to prove loss.

Licence redesign can improve the target before or after closing. Products can be split into training, retrieval and inference rights. Broad exclusivity can be replaced with bounded fields and performance minimums. Usage reporting can move into the API. Refresh obligations can receive separate pricing. Customer rights can be matched to the technical architecture.

13 Govern value after signing and closing

Data value can change quickly after a transaction. Source permissions can expire, customer use can expand, products can drift and new regulation can alter processing. The buyer needs a rights and economics control system rather than a static closing binder.

The core register should connect source assets, rights, product versions, customer grants, models or applications, usage, invoices, refresh releases, incidents and renewals. Owners should be assigned across legal, data, product, security, finance and commercial teams. Material changes should trigger review before new use or delivery.

The register should support lineage in both directions. A source owner should be able to identify every product and customer that depends on a licence approaching expiry. A customer manager should be able to identify the sources and restrictions underlying a proposed use. This capability makes renewal, deletion and incident response faster and provides evidence for pricing and board reporting.

Control thresholds should lead to action. A fall in verified-rights coverage can suspend new sales of an affected field. Unbilled use above a defined level can trigger a contract and metering review. Repeated late refreshes can require product repricing or additional investment. A concentration threshold can limit further exclusivity. Each trigger should identify the accountable executive, decision deadline and required evidence.

Integration should preserve customer trust. The buyer should communicate changes in ownership, systems, support and processing where required and should avoid broadening use before authority is confirmed. Customer-facing documentation should remain consistent with contracts and technical reality. A disciplined first renewal cycle can validate the acquisition thesis and expose value gaps early.

Post-close indicators should include verified-rights coverage, metered-use coverage, unbilled use, refresh timeliness, source concentration, customer concentration, renewable gross margin, audit exceptions, rights disputes, deletion or unlearning requests, renewal, net revenue retention and cash collection. Measures should be segmented by product and jurisdiction.

The integration plan should protect evidence. Contract repositories, source records, logs and release archives can be lost during system migration. A buyer should preserve immutable snapshots, define retention and control access. It should also test whether customer consent or notice is required for assignment, new processing or broader group use.

Governance should support commercial speed. A pre-approved rights taxonomy, standard clauses, metering patterns and exception process can shorten negotiations and reduce bespoke obligations. Product teams should know which rights are available before promising a new use. Finance should know which meter supports each invoice. Legal should know when the technical design changes the licence perimeter.

14 Board decision framework and conclusion

The board should require a concise evidence pack before approving an acquisition, investment, financing or material licence. The pack should define the asset, supported rights, permitted uses, customers, billable units, refresh obligations, control gaps, revenue quality, valuation methods, downside cases and transaction protections. Every material assumption should have an owner and evidence status.

Five questions organise the decision. First, can the company prove authority for the data and every commercially important use? Second, can it measure what customers consume and reconcile that use to invoices and cash? Third, can it refresh the product at a cost that preserves margin and customer utility? Fourth, can it detect and address unauthorised use, rights expiry and product change? Fifth, does the price remain supported after unsupported revenue, remediation and concentration are included?

A positive answer requires more than a valuable-looking corpus. It requires a controlled system connecting source authority, rights, delivery, use, billing, collection and renewal. Each link should be tested. A weakness in one link can reduce the value of all others.

The illustrative case begins with USD 84 million of reported revenue, USD 18 million of EBITDA and a USD 96 million management allocation. Normalisation reduces renewable revenue to USD 69 million. Replacement cost indicates USD 31 million, income indicates USD 72 million and market evidence produces a USD 55 million to USD 90 million corridor. The evidence-weighted conclusion is USD 67 million. These management assumptions demonstrate the method and provide no opinion on an actual company.

Royalty valuation for AI data should remain use-specific and evidence-led. Volume, exclusivity, refresh and audit rights are economic terms as well as drafting terms. When they are defined and measured, a buyer can distinguish durable licensing cash from temporary access, unbilled adoption and unsupported claims. The resulting valuation can then inform price, structure, licence design and post-close investment.

Frequently asked questions

What should be valued in a data-licensing company

The valuation should distinguish the corpus, rights record, curation and taxonomy, delivery software, refresh operation, usage-control system, customer contracts and workforce. The analysis should prevent the same cash flow from being counted in data, software, brand and customer relationships.

How should training rights differ from inference rights

Training can create model weights that persist after source access ends. Inference can generate repeated commercial outputs and may use the data directly or indirectly. Contracts and valuation models should define each use, permitted derivatives, survival, measurement and termination.

Which volume unit is most defensible

The useful unit is observable, attributable, resistant to manipulation and connected to economic value. Depending on the product, it can be records, calls, tokens, models, users, deployments, revenue or a combination. Contract, meter, invoice and cash should reconcile.

How should exclusivity be priced

The licensor should estimate contribution and option value lost from excluded markets. The customer should estimate the advantage created by the restriction. Scope, duration, territory, field, substitutes, minimum commitments and reversion determine the supported premium.

Does dataset size determine value

Size can support coverage and can also contain duplication, obsolete records or unusable rights. Value depends on fitness for a customer decision, lawful and transferable rights, quality, scarcity, refresh, delivery, measured use and sustainable cash.

How should refresh obligations enter valuation

The forecast should include recurring source acquisition, cleaning, validation, release and customer-migration cash. Renewal benefit should be supported by accepted releases and customer use. Economic life should be assessed by field and intended use.

Can a public royalty benchmark determine fair value

A benchmark can provide market evidence when rights, use, exclusivity, territory, maturity and economics are comparable. Public terms are often incomplete. The analyst should document adjustments and reconcile market evidence with cost and income indications.

Which post-close controls protect value

Important controls include a linked source-and-rights register, customer-use matrix, tamper-evident metering, contract-to-cash reconciliation, refresh acceptance, incident review, renewal cohorts and board thresholds for unsupported rights, unbilled use and concentration.

References

- [1] IFRS Foundation. IAS 38 Intangible Assets. <https://www.ifrs.org/issued-standards/list-of-standards/ias-38-intangible-assets/>
- [2] IFRS Foundation. IFRS 3 Business Combinations. <https://www.ifrs.org/issued-standards/list-of-standards/ifrs-3-business-combinations/>
- [3] IFRS Foundation. IFRS 13 Fair Value Measurement. <https://www.ifrs.org/issued-standards/list-of-standards/ifrs-13-fair-value-measurement/>
- [4] IFRS Foundation. IFRS 15 Revenue from Contracts with Customers. <https://www.ifrs.org/issued-standards/list-of-standards/ifrs-15-revenue-from-contracts-with-customers/>
- [5] World Intellectual Property Organization. Valuing intellectual property assets. <https://www.wipo.int/en/web/business/ip-valuation>
- [6] World Intellectual Property Organization. Intellectual Property Valuation Basics for Technology Transfer Professionals. <https://www.wipo.int/web-publications/intellectual-property-valuation-basics-for-technology-transfer-professionals/en/index.html>
- [7] World Intellectual Property Organization. The income approach. <https://www.wipo.int/web-publications/intellectual-property-valuation-basics-for-technology-transfer-professionals/en/6-the-income-approach.html>
- [8] World Intellectual Property Organization. The market approach. <https://www.wipo.int/web-publications/intellectual-property-valuation-basics-for-technology-transfer-professionals/en/5-the-market-approach.html>
- [9] World Intellectual Property Organization. Expert determination for intellectual property valuation. <https://www.wipo.int/amc/en/expert-determination/ip-valuation/>
- [10] OECD. The value of data in digital-based business models. https://www.oecd.org/en/publications/the-value-of-data-in-digital-based-business-models-measurement-and-economic-policy-implications_d960a10c-en.html
- [11] OECD. Guidance on Transfer Pricing Aspects of Intangibles. https://www.oecd.org/content/dam/oecd/en/publications/reports/2014/09/guidance-on-transfer-pricing-aspects-of-intangibles_g1g46cf9/9789264219212-en.pdf
- [12] OECD. Guidance on the Approach to Hard-to-Value Intangibles. https://www.oecd.org/en/publications/guidance-for-tax-administrations-on-the-application-of-the-approach-to-hard-to-value-intangibles-beps-action-8_5e402215-en.html
- [13] United States Copyright Office. Copyright and Artificial Intelligence initiative. <https://www.copyright.gov/ai/>
- [14] United States Copyright Office. Copyright and Artificial Intelligence Part 3 Generative AI Training. <https://www.copyright.gov/ai/Copyright-and-Artificial-Intelligence-Part-3-Generative-AI-Training-Report-Pre-Publication-Version.pdf>
- [15] United States Copyright Office. Notice of Inquiry on Artificial Intelligence and Copyright. <https://www.copyright.gov/ai/docs/Federal-Register-Documents-Artificial-Intelligence-and-Copyright-NOI.pdf>
- [16] United States Copyright Office. Economic research datasets. <https://www.copyright.gov/economic-research/usco-datasets/>
- [17] European Union. Directive EU 2019/790 on copyright and related rights in the Digital Single Market. <https://eur-lex.europa.eu/eli/dir/2019/790/oj>
- [18] European Union. Regulation EU 2023/2854 on harmonised rules on fair access to and use of data. <https://eur-lex.europa.eu/eli/reg/2023/2854/oj>
- [19] European Union. Regulation EU 2024/1689 laying down harmonised rules on artificial intelligence. <https://eur-lex.europa.eu/eli/reg/2024/1689/oj>
- [20] European Union. Regulation EU 2022/868 on European data governance. <https://eur-lex.europa.eu/eli/reg/2022/868/oj>
- [21] European Union. Regulation EU 2016/679 General Data Protection Regulation. <https://eur-lex.europa.eu/eli/reg/2016/679/oj>

- [22] European Data Protection Board. Opinion 28/2024 on certain data protection aspects related to the processing of personal data in the context of AI models. https://www.edpb.europa.eu/our-work-tools/our-documents/opinion-board-art-64/opinion-282024-certain-data-protection-aspects_en
- [23] United Kingdom Information Commissioner's Office. The lawful basis for web scraping to train generative AI models. <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/>
- [24] United Kingdom Information Commissioner's Office. Guidance on AI and data protection. <https://ico.org.uk/for-organisations/uk-gdpr-guidance-and-resources/artificial-intelligence/guidance-on-ai-and-data-protection/>
- [25] United Kingdom Information Commissioner's Office. Anonymisation pseudonymisation and privacy enhancing technologies guidance. <https://ico.org.uk/for-organisations/uk-gdpr-guidance-and-resources/data-sharing/anonymisation/>
- [26] National Institute of Standards and Technology. Artificial Intelligence Risk Management Framework 1.0. <https://doi.org/10.6028/NIST.AI.100-1>
- [27] National Institute of Standards and Technology. Artificial Intelligence Risk Management Framework Generative Artificial Intelligence Profile. <https://doi.org/10.6028/NIST.AI.600-1>
- [28] National Institute of Standards and Technology. Privacy Framework. <https://www.nist.gov/privacy-framework>
- [29] National Institute of Standards and Technology. Cybersecurity Framework 2.0. <https://doi.org/10.6028/NIST.CSWP.29>
- [30] International Organization for Standardization. ISO IEC 5259-1 Artificial intelligence data quality for analytics and machine learning. <https://www.iso.org/standard/81088.html>
- [31] International Organization for Standardization. ISO IEC 42001 artificial intelligence management systems. <https://www.iso.org/standard/81230.html>
- [32] Federal Trade Commission. Artificial intelligence accomplishments. https://www.ftc.gov/system/files/ftc_gov/pdf/ai-accomplishments-1.17.25.pdf
- [33] Federal Trade Commission. Artificial intelligence and algorithmic tools. <https://www.ftc.gov/business-guidance/technology/artificial-intelligence>
- [34] United Kingdom Competition and Markets Authority. AI Foundation Models update report. <https://www.gov.uk/government/publications/ai-foundation-models-update-report>
- [35] European Commission. Competition in generative AI and virtual worlds. https://competition-policy.ec.europa.eu/about/reaching-out/calls-contributions/call-contributions-competition-virtual-worlds-and-generative-ai_en
- [36] European Commission. Common European data spaces. <https://digital-strategy.ec.europa.eu/en/policies/data-spaces>
- [37] OECD. Recommendation of the Council on Artificial Intelligence. <https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449>
- [38] OECD. Enhancing Access to and Sharing of Data. https://www.oecd.org/en/publications/enhancing-access-to-and-sharing-of-data_276aaca8-en.html
- [39] United Kingdom Intellectual Property Office. Copyright and artificial intelligence. <https://www.gov.uk/government/consultations/copyright-and-artificial-intelligence>
- [40] United Kingdom Government. Data ethics framework. <https://www.gov.uk/government/publications/data-ethics-framework>
- [41] World Intellectual Property Organization. Exchanging Value Negotiating Technology Licensing Agreements. https://www.wipo.int/edocs/pubdocs/en/licensing/906/wipo_pub_906.pdf
- [42] World Intellectual Property Organization. Trade secrets. <https://www.wipo.int/tradesecrets/en/>
- [43] World Intellectual Property Organization. Technology transfer and access to markets. <https://www.wipo.int/en/web/technology-transfer/access-market>
- [44] IFRS Foundation. Post-implementation Review of IFRS 15 Licensing. <https://www.ifrs.org/content/dam/ifrs/meetings/2024/february/iasb/ap6c-ifrs-15-pir-licensing.pdf>
- [45] IFRS Foundation. IFRS 15 educational session. <https://www.ifrs.org/content/dam/ifrs/meetings/2016/september/wss/education-session/edu-ifrs15.pdf>
- [46] United States Internal Revenue Service. Transfer Pricing Examination Process. <https://www.irs.gov/businesses/international-businesses/transfer-pricing-examination-process>
- [47] European Commission. Data Act explained. <https://digital-strategy.ec.europa.eu/en/policies/data-act>
- [48] European Commission. Model contractual terms and standard contractual clauses under the Data Act. <https://digital-strategy.ec.europa.eu/en/library/model-contractual-terms-and-standard-contractual-clauses-data-act>
- [49] European Union Agency for Cybersecurity. Data protection engineering. <https://www.enisa.europa.eu/publications/data-protection-engineering>
- [50] International Organization for Standardization. ISO IEC 27001 information security management systems. <https://www.iso.org/standard/27001>

About the Author

Chennakeshav (CK) is a corporate finance and investment banking executive with 25+ years of global experience in deal origination, structuring and execution across M&A, growth capital and corporate strategy. He has led value-creation mandates for founders, corporates and funds — bridging the boardroom view to hands-on execution and close.

His career spans Morgan Stanley, HSBC, Lloyds Banking Group, EWEC, ADQ portfolio companies and Emirates Growth Fund, across TMT, real estate, fintech, deeptech, cleantech, infrastructure and energy. He has partnered with C-suite leaders, private equity and venture funds, sovereign wealth funds and family offices to finance complex fund raises and scale-up ventures, and has led M&A due diligence, post-merger integration and business-transformation initiatives to create value.

At Matchpoint Partners he is Managing Partner, leading the firm's corporate finance, M&A and capital-raising practice. He holds an MBA from London Business School, an engineering degree from VTU and a Master of Laws (LLM, in progress) from UCL London.

An active start-up mentor, CK mentors at Techstars, DIFC FinTech Hive, Startup Grind, Founder Institute and IN5, serves as Entrepreneur Mentor in Residence (EMiR) at London Business School, and judges the Entrepreneurship World Cup.

<https://www.linkedin.com/in/ckadya/>

<https://www.matchpoint-partners.com/team/ck-adya.html>

This paper is part of a continuing series on the structure of private and alternative markets. The views expressed are the author's own. The paper is for information only, describes market structure in general terms, and does not constitute investment, legal, tax or regulatory advice or a recommendation in respect of any security, vehicle or counterparty.