Research on Data Intellectual Property Protection and Value Assessment in the AI Era: A Tri-Jurisdictional Comparative Study of China, the United States, and the European Union

Augusto·Ben·Anle

European Institute of International Management and Entrepreneurship，Brandenburger Tor 1, 10178 Berlin, Deutschland，Germany

Abstract

As artificial intelligence reshapes the global economic infrastructure, data has ascended from a by-product of digital activity to the *primary factor of production* in the twenty-first century. Yet the very legal regimes tasked with protecting data-derived value remain profoundly fragmented. This paper conducts a comparative institutional analysis of how China, the United States, and the European Union construct data intellectual property (IP) protection frameworks, and how those divergent architectures interact with market mechanisms for valuing data assets. Drawing on three paradigmatic cases—*(1) hiQ Labs v. LinkedIn (U.S.)*, *(2) the EU's sui generis database right regime under the Database Directive and subsequent Data Act reform*, and *(3) China's emerging data-IP registration pilot system exemplified by the Shanghai and Zhejiang platform-data disputes*—this study identifies systematic patterns: the U.S. relies on a decentralized, litigation-driven mosaic (trade secrets, CFAA, contract, copyright) that maximizes allocative flexibility at the cost of transactional uncertainty; the EU deploys a rights-centric, regulatory-dense architecture (database right + GDPR + Data Act) that secures individual and producer protections but imposes heavy compliance frictions on data monetization; and China pursues a state-guided, registration-enabled property-rights experiment ("three-rights" bifurcation + data IP pilots) designed to engineer a tradable data-asset class from above. The paper then develops an interaction-mechanism framework showing how each jurisdiction's legal DNA shapes which valuation methodologies (cost / income / market) dominate, what discount rates the market applies, and where liquidity emerges or stalls. Business and policy implications are drawn for firms operating across jurisdictions in the AI supply chain.

Keywords: data intellectual property; data asset valuation; AI; comparative law; China; United States; European Union; hiQ v. LinkedIn; data governance

1. Introduction

"Data is the new oil." The aphorism is tired, but the underlying truth is sharper: in the AI era, data is not merely like oil — it is more accurately like land. It must be surveyed, titled, enclosed, transferred, zoned, and taxed. And, as with land in the early modern period, the question of who holds what rights over data determines who captures the surplus value of the AI economy.

Since the OECD's 2019 Ministerial Declaration on "Going Digital" and the WTO's ongoing e-commerce negotiations, a consensus has emerged that data requires *some* form of proprietary scaffolding if it is to be traded efficiently. But there is no global agreement on *what kind* of proprietary scaffolding. The United States, the European Union, and China — the three gravitational centers of the AI ecosystem — have each built radically different institutional answers to the same question: **how do you protect investments in data without throttling the downstream flow that AI itself demands?**

This paper argues that understanding data IP protection cannot stop at doctrinal description. One must trace the **feedback loop** between *legal form* and *market valuation*: the way a jurisdiction defines (or refuses to define) data rights directly determines which valuation methodologies are usable, what discount rates the market prices in, and ultimately which actors capture the rent from AI-training data, licensed datasets, and data-driven mergers.

The motivation is pragmatic as much as scholarly. For business-school audiences, the stakes are corporate: M&A due diligence for data-intensive targets, fair-value accounting of data assets under evolving standards, cross-border compliance budgeting, and strategic IP positioning in AI supply chains. For policymakers, the question is whether fragmented regimes produce innovation-harming data nationalism or manageable pluralism.

1.1 Research Questions

1. **Architectural:** How do the China–U.S.–EU legal frameworks for data protection differ in their underlying logic — and what does each treat as the *protected interest* (privacy? trade secrecy? investment? personality? sovereignty)?
2. **Comparative-doctrinal:** What are the operative legal instruments in each jurisdiction that function *as de facto IP* over commercially valuable data, even

when no formal "data IP" statute exists?

3. **Case-empirical:** What do three landmark case trajectories reveal about how each system actually *operates on the ground* when commercial data conflicts arise?
4. **Mechanism:** How does each legal architecture shape data-asset valuation practice — i.e., the cost-income-market triad — through transaction-cost channels, enforceability premia/discounts, and liquidity constraints?

1.2 Methodology

This is a qualitative, comparative-institutional paper organized around **three paradigmatic case studies**, selected to represent the dominant protective logic of each jurisdiction:

#	Jurisdiction	Case / Institutional Episode	Core Legal Logic
I	**United States**	***hiQ Labs v. LinkedIn*** (N.D. Cal. 2017; 9th Cir. 2019; *Van Buren* remand; final dissolution ca. 2022)	CFAA-as-intrusion-statute; contract/TOS; trade-secret adjacency; competition policy pressures on "information monopolies"
II	**European Union**	**The sui generis database right → *Ryanair v. PR Aviation* → Data Act / Data Governance Act reforms**	Positive legislative entitlement (Database Directive 96/9/EC); GDPR overlay; regulated data-access/intermediation mandates
III	**China**	**Data-IP Registration Pilot System (CNIPA**	Anti-unfair-competition protection of "substantial data

		2022–present) + Platform data misappropriation doctrine (*Taobao v. Guodu* line)	accumulation"; nascent "data product IP" registration; "three-rights" structural framing (*Data Twenty Articles*)

Each case is analyzed not as a self-contained legal curiosity but as a **market institution** that alters the risk-return calculus of data-intensive business models.

2. Conceptual Foundations: What Is "Data IP" in the AI Era?

Before comparing regimes, we must clarify what we mean by *data intellectual property*. Data resists traditional IP categorization, and this resistance is the source of the entire problem.

2.1 The Categorical Problem

Traditional IP regimes sort creations into bins:

Regime	Protects	Requirement	Duration	Data Fit?
Copyright	Original expression	Creativity / originality	Life + 50–70 yrs	Poor — facts & raw data are uncopyrightable (*Feist v. Rural*, U.S.; EU Database Directive distinguishes structure from content)
Patent	Functional inventions	Novelty, non-obviousness, utility	20 yrs	Typically inapplicable to data *per se*
Trade Secret	Confidential	Reasonable	Perpetual	Partial fit —

	commercial info	secrecy measures	while secret	works for *non-public* data but collapses once data is shared, scraped, or derived
Trademark	Source identifiers	Distinctiveness	Renewable	Irrelevant to data *content*

Raw data — temperature readings, clickstream logs, satellite imagery pixel-values — are **facts**, and facts are unownable under the classical *res nullius* principle. Yet the *investment* required to collect, clean, label, structure, enrich, and maintain a dataset is undeniably capital. AI makes this tension acute: training a frontier model may consume $100M+ in data-acquisition and curation costs, yet the underlying raw facts remain, legally speaking, "nobody's."

2.2 From "Data as Fact" to "Data as Crystallized Investment"

Three conceptual moves have emerged globally to get around the fact/expression barrier:

1. **The *sweat-of-the-brow* / investment-protection theory** (EU sui generis right): you don't own the facts, but you own the *prohibitive investment you made to assemble them*, giving rise to a time-limited extraction right.
2. **The *unfair-competition / misappropriation* theory** (China's AUCL Art. 2 / judicial precedent; some U.S. state common law): even absent a formal right, free-riding on another's data accumulation without equivalent effort is actionable when it causes demonstrable harm.
3. **The *contractual-encirclement* strategy** (Silicon Valley default): since property isn't available, wrap data in ToS / API keys / CFAA threats / trade-secret labeling, and enforce through contract + cybersecurity law.

None is a perfect substitute for a clean property title. But each produces *different market behaviors* — which is exactly why the comparative angle matters.

2.3 Defining "Data IP" for This Paper's Purposes

Operational definition: *Data IP* refers to the ensemble of legal techniques — statutory, judge-made, regulatory, and administrative — that confer *enforceable exclusionary or attributional benefits* upon those who invest in the collection, processing, and structuring of data, sufficient to make data credibly excludable, transferable, and therefore valuatable as an asset on a balance sheet.

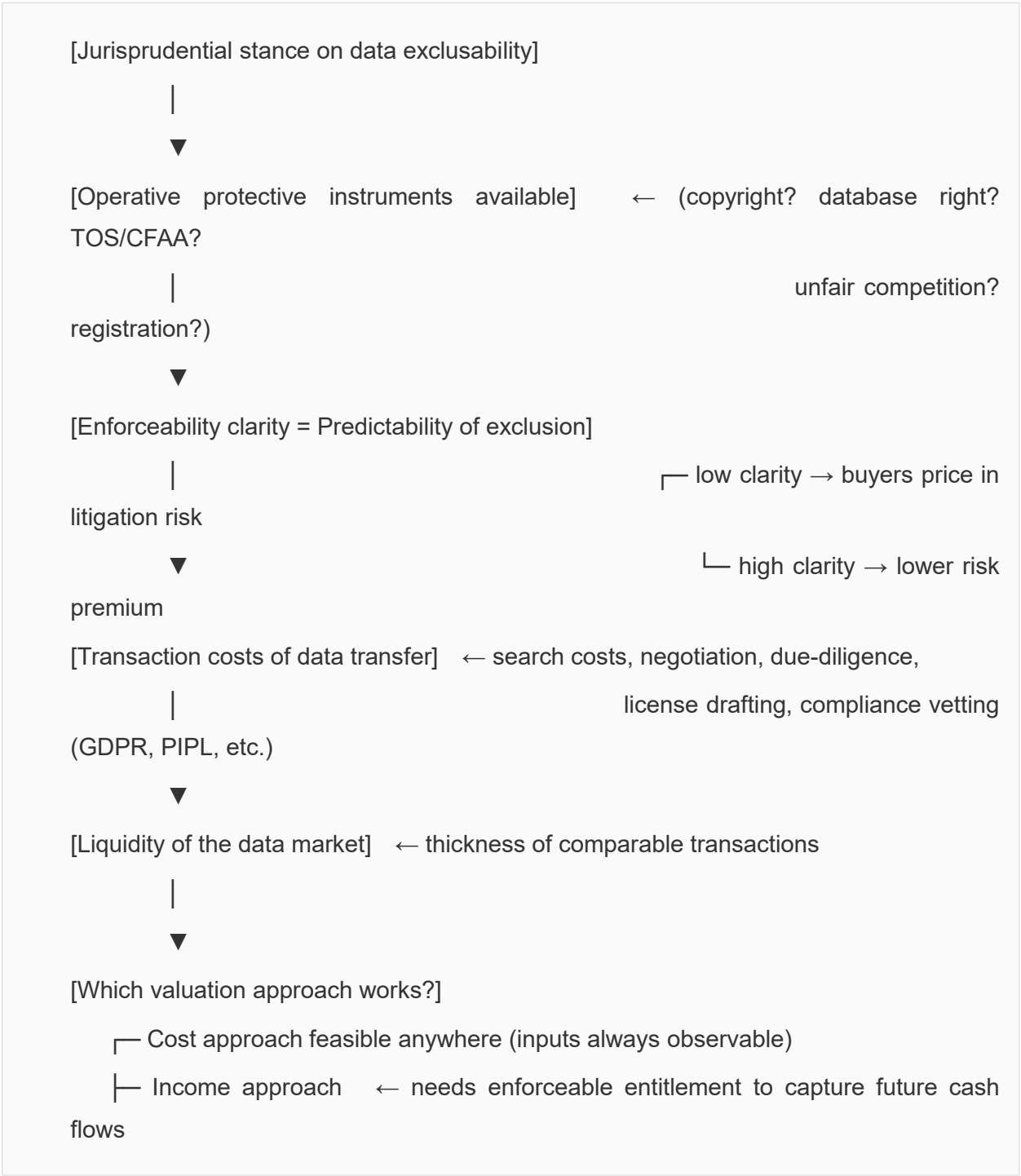
This deliberately broad definition captures the reality that *no jurisdiction* currently has a unified "Data Copyright." What exists instead are **functional equivalents** whose economic effect is to create (or deny) IP-like protection.

3. Analytical Framework: The Legal–Valuation Interaction Mechanism

3.1 The Core Proposition

Proposition: The form of a jurisdiction's data-protection regime is not exogenous to data valuation — it is *constitutive of it*. Legal architecture determines (a) whether data is perceived as an *asset class* or a *liability/risk pool*, (b) which valuation methodologies converge on credible numbers vs. diverge into noise, and (c) what liquidity — i.e., trade velocity — the market can support.

3.2 The Mechanism Chain



Corollary: Even technically sound valuation mathematics (DCF, multi-period excess earnings, Shapley-value-style contribution splits) collapse if the *underlying entitlement* is too fuzzy for a buyer to feel confident they're buying something real.

4. Tri-Jurisdictional Comparison of Data-Protection Architectures

4.1 The United States: A Market-Led Patchwork

The U.S. has **no federal data-property statute**. There is no American equivalent of

the EU's Database Directive and no centralized registry of data rights. Instead, data protection is an emergent property of overlapping doctrines:

4.1.1 Copyright's "Facts Exception" (*Feist*)

The Supreme Court's *Feist Publications v. Rural Telephone Service* (1991) definitively established that **raw facts are uncopyrightable**; only the *creative selection, coordination, or arrangement* may qualify. For most machine-generated, AI-relevant datasets (sensor logs, scrape corpora, transaction records), *Feist* strips copyright down to thin protection over the database's expressive "skin," leaving the valuable data-meat unprotected.

4.1.2 Trade Secrets (DTSA / UTSA)

The **Defend Trade Secrets Act (2016)** and state Uniform Trade Secrets Acts protect "information … [with] independent economic value … not generally known … and … subject to reasonable secrecy measures." This is the *workhorse* of U.S. commercial data protection:

- **Strength:** Injunctive relief, damages, attorney-fee shifts, ex parte seizure orders.
- **Weakness:** The moment data leaves the perimeter (via leak, reverse-engineering from public outputs, or independent derivation), trade-secret status evaporates. It protects *secrecy*, not *value-from-sharing*.

4.1.3 The CFAA (Computer Fraud and Abuse Act, 18 U.S.C. §1030)

Originally an anti-hacking statute, the CFAA has been litigated as a data-access control weapon. Its "exceeds authorized access" language was historically stretched to target scraping and ToS violations — until the Supreme Court's *Van Buren v. United States* (2021) narrowed it to *true intrusions* into areas where access was technically barred, not mere policy/ToS breaches. This created the opening for *hiQ v. LinkedIn* to stand.

4.1.4 Contract & Terms of Service

Platforms rely heavily on **contractual encirclement**: ToS provisions prohibiting scraping, API-use limits, and browse-wrap agreements. U.S. courts generally enforce ToS as contract — but contract creates *privity-only* protection (only binds parties who agreed), not *erga omnes* exclusion. A downstream buyer who never clicked "I Agree" may not be bound.

4.1.5 State-Level Privacy (CPRA/CCPA in California, etc.)

The U.S. privacy patchwork adds a *compliance tax* on data use rather than a property right. Notably, CPRA creates rights to opt out of "sharing" and "selling," indirectly

forcing firms to track data provenance — which, paradoxically, improves the *audit trail* needed for valuation.

4.1.6 Summary Judgment on the U.S. Model

Feature	U.S. Position
Formal data property right?	❌ No
De facto protection	✅ Trade secret + contract + CFAA (narrowed) + occasional copyright
Governing logic	Litigation-driven, pro-innovation, info-flow-default
Transaction certainty	Low–Medium (rights are fuzzy, case-specific)
Effect on valuation	Income approach works *inside the firm*; market approach hampered by non-transferability

Business takeaway: In the U.S., data value is primarily realized **within the vertically integrated platform** (use it yourself) rather than through **clean arms-length sale** of raw data titles — because clean titles don't exist.

4.2 The European Union: A Rights-Centric Regulatory Architecture

The EU takes the *opposite constitutional posture*: data protection is not an afterthought but a **foundational regime** combining *fundamental-rights privacy* (GDPR) with *positive economic rights* (Database Directive, Data Act).

4.2.1 The Sui Generis Database Right (Directive 96/9/EC)

Article 7 of the Database Directive creates a **time-limited property-like right** (typically 15 years from publication) granting the maker of a database that shows *qualitatively or quantitatively substantial investment* in obtaining, verifying, or presenting its contents the right to prohibit:

- unauthorized **extraction** (permanent transfer of contents to another medium), and
- unauthorized **re-utilization** (making contents available to the public by any means).

Key features:

- It protects the **investment**, not originality.
- Applies even to *uncopyrightable facts*, solving the *Feist* gap.
- But it is *not perpetual* and requires continuous "substantial new investment" to refresh.
- Has been criticized as under-enforced and conceptually mismatched to cloud/API realities.

4.2.2 GDPR as a Superimposed Data-Governance Layer

The **General Data Protection Act** (GDPR, 2016/679) treats personal data through the lens of *individual rights* (Art. 5 principles: purpose limitation, data minimization, storage limitation, integrity/confidentiality). Crucially:

- **Legal basis for processing** (consent, legitimate interest, contract, etc.) must be established before data has any commercial usability.
- **Art. 20 (Data Portability)** and **Art. 9 (Special Categories)** create carve-outs and minefields.
- For AI training: GDPR makes *personal* data risky; many firms respond by **synthesizing, aggregating, or pseudonymizing** — which alters the very nature of the "asset" being valued.

4.2.3 The Data Act (Regulation EU 2023/2854) & Data Governance Act (Regulation EU 2022/868)

These newer instruments shift the EU away from *pure enclosure* toward **regulated access**:

- The **Data Act** mandates making IoT/connected-product data available to users and, under conditions, to third-party service providers.
- The **Data Governance Act** creates frameworks for **data intermediation services** (neutral brokers) and for re-using certain public-sector data.

This represents a subtle but profound shift: the EU is saying, *"We recognize your data investment, but the socially optimal rule is not absolute lock-up — it is managed sharing via trusted intermediaries."*

4.2.4 Summary Judgment on the EU Model

Feature	EU Position
Formal data property right?	✓ Yes — sui generis database right (limited-term, investment-based) +

	database copyright (originality-based)
Privacy/fundamental rights overlay?	✓✓ GDPR — most stringent in the world
Governing logic	Rights-centered, ex-ante regulatory, dignitarian
Transaction certainty	Medium–High on *entitlement*; High on *compliance burden*
Effect on valuation	Stronger title → market approach more feasible *for compliant datasets*; but GDPR frictions inflate effective cost base (→ cost approach rises; income approach discounted for regulatory risk)

Business takeaway: The EU gives you the *strongest formal claim* over your database — but the *narrowest corridor* through which you can actually operationalize it profitably. Valuation must price in **regulatory-risk beta**.

4.3 China: State-Guided Property-Rights Engineering

China's approach is the most *dynamic* of the three — less settled in doctrine, more active in institutional design.

4.3.1 Constitutional-Strategic Framing: "Data as Factor of Production"

The December 2020 **"Data Twenty Articles"** laid out the governing philosophy:

- Data should be treated as a **factor of production**
- Property rights should be **tripartite**: *data-resource holding right* / *data-processing use right* / *data-product operation right*— later nomenclature-refined to **holding / use / operation** Emphasis on **"use-rights separation"** rather than Western-style monolithic ownership

This is consciously modeled on China's successful rural land-reform insight: don't resolve the metaphysical "who owns it?" question; instead, **unbundle use, transfer, and income rights** to make the asset tradable.

4.3.2 The Anti-Unfair-Competition Path (AUCL)

Before formal data-IP crystallized, Chinese courts protected platform data under **Anti-Unfair Competition Law (AUCL)** — notably:

- *Taobao v. Guodu Network*: the Hangzhou Internet Court recognized that platforms'

substantial investment in gathering, cleaning, and structuring transaction/behavioral data generated a protectable competitive interest against free-riding scrapers

- The logic: not "we own the facts" but "you cannot *misappropriate the commercial fruit* of our labor without equivalent effort."

4.3.3 Data-IP Registration Pilots (CNIPA, 2022–Present)

Since 2022, the **China National Intellectual Property Administration (CNIPA)** has rolled out **data-IP registration pilots** across 17 provincial-level regions (the original 8 + 9 added in the 2024 batch). Key design features:

Element	Design
Object	Lawfully obtained, processed data collections with "intellectual-achievement attributes" and practical value
Subject	The *data processor* — whoever performed the labor of obtaining/storing/processing
Review	*Voluntary*, mostly **formality examination** (+ prior notarization / blockchain timestamp recommended)
Effect	Certificate serves as **prima facie proof** of holding/use/operation rights; supports pledge financing, licensing, asset entry on books
Scale	>15,600 registrations by Oct. 2024 across pilot platforms; leading provinces: Zhejiang, Fujian, Jiangsu (~90%); Shanghai alone issued 800+ certificates, generating >¥195B in economic value via licensing/trading

This is, in effect, **administrative-title-making**: the state creates a registry that *functions like IP office registration* (think patent office, but for curated datasets),

thereby transforming a fuzzy equitable interest into a *marketable instrument*.

4.3.4 Overlay: DSL + PIPL + Data Security Law

China's **Data Security Law (DSL, 2021)** and **Personal Information Protection Law (PIPL, 2021)** add:

- A **data-classification regime** (core / important / general data)
- **Cross-border transfer restrictions** for "important data"
- **PIPL consent / lawful-basis requirements** for personal information

So, like the EU, China layers privacy/security onto data economics — but unlike the EU's rights-centrism, China's implementation is **state-coordinated**, with data-localization pressures and national-security framing prominent.

4.3.5 Summary Judgment on the China Model

Feature	China Position
Formal data property right?	⚙ Under construction — "three-rights" framework + CNIPA registration pilots = *soft-title* system
Privacy/security overlay?	✓ DSL + PIPL (enforcement can be stringent, including criminal referral for data-black-market conduct)
Governing logic	State-guided asset-making: make data *legible, registrable, bankable*
Transaction certainty	Rising (registration helps, but legal status remains *experimental*/soft-law)
Effect on valuation	Cost approach + income approach actively promoted via official valuation guidance; registration enables *market approach* by creating a recorded-trade ecology

4.4 Synthesis: A Comparative Matrix

Dimension	**United States**	**European Union**	**China**

Ontological stance on "data ownership"	No ownership in facts; protection via *contract* + *secrecy* + *anti-intrusion*	Positive legislative entitlement (sui generis DB right) + fundamental-rights privacy ceiling	"Three-rights" unbundling ; data as *factor of production* to be titled
Core operative instruments	Trade Secret (DTSA), CFAA (narrowed), ToS/Contract, thin Copyright, State privacy (CPRA)	Database Directive (Art. 7), GDPR, Data Act, DGA, Copyright (Art. 3)	AUCL (misappropriation doctrine), DSL/PIPL, CNIPA Data-IP Registration Pilots
Exclusion strength	Medium (contract-bound; secrecy-dependent)	High formal right / Med–High practical access-mandates	Medium–High (growing via registration; anti-scraping enforced under AUCL + cybercrime)
Allocative default	Flow-maximizing (pro-innovation, anti-information-monopoly)	Dignity & rights-maximizing (privacy-first, then managed sharing)	State-steered (strategic sectoralism; data sovereignty + market-build)
Valuation environment	Income approach (internal); market approach fragile	Income + incipient market (for compliant datasets); heavy risk-discount	Cost→Income→Market pipeline actively engineered; registration = credit-device
Best-fit business strategy	Monetize *behind* the API; vertical integration; trade *services* not *data*	Compliance-as-barrier-to-entry → premium on well-governed	Register early in pilot zones; leverage certificate for

		datasets; intermediation models	financing/trading; align with sector plans

5. Case Studies

Case I: *hiQ Labs v. LinkedIn Corp.* — The U.S. Paradigm of Contract, CFAA, and the "Public Data" Boundary

5.1 Factual Background

hiQ Labs, a small analytics startup, scraped **publicly visible** LinkedIn member profile data (public-profile fields only — no login walls) to build two B2B products: *Keeper* (predicting which employees were flight risks) and *Skill Mapper* (skill-gap mapping). LinkedIn sent a cease-and-desist citing CFAA, DMCA, misappropriation, and breach of ToS. hiQ sued pre-emptively for declaratory relief and a **preliminary injunction** to keep LinkedIn from blocking access.

5.2 The Legal Arc

Stage	Holding / Significance
N.D. Cal. 2017 (Chen J.)	Granted hiQ's PI. Key reasoning: giving a private platform unilateral power to wall off *public* data risks creating an **"information monopoly"** disserving the public interest. CFAA targets *intrusions*, not browsing public pages.
9th Cir. 2019 (per curiam)	Affirmed. Held: where data is **publicly accessible** without credential gates, accessing it cannot be "without authorization" under CFAA. The court read CFAA as an *anti-intrusion* statute, not a *misappropriation* statute.
Supreme Court (2021)	Vacated and remanded in light of *Van Buren v. United States* (which narrowed "exceeds authorized access"). On remand, 9th Circuit **reaffirmed** the PI.
Endgame (~2022)	LinkedIn and hiQ **settled**. By then hiQ had

	largely ceased operations; the district court dissolved the PI noting hiQ's lack of continuing business. LinkedIn's ToS/contract claims remained *technically alive* as a path to control scraping going forward.

5.3 Business-School Interpretation: What This Case *Actually* Means

Too many summaries frame *hiQ* as "scraping is legal in America." That is wrong. What *hiQ* actually establishes is far more nuanced:

1. **The "Public Data" Carve-Out is Narrow.** The court emphasized that hiQ accessed data *anyone with a browser could see* — no password, no token, no bypass of technical barriers. Scrape *behind* a login wall, or after a ToS-ban backed by technical measures, and the analysis changes sharply.
2. **CFAA Is Not Your Title System.** The case is *not* a property-rights victory for data collectors. It is a *limit-on-a-criminal-statute* ruling. LinkedIn's *civil* claims (ToS breach, misappropriation under California common law, copyright in selection/arrangement) remained viable weapons — just not the CFAA hammer.
3. **The Real Lesson: Contract + Architecture > Doctrine.** After *hiQ*, sophisticated platforms responded not by suing more but by **architecting barriers**: stronger authentication, rate-limiting, CAPTCHA, dynamic DOM structures, API gateways, and — crucially — **requiring OAuth/login** for formerly public endpoints, thereby moving data behind the credential wall where CFAA's "authorized access" logic re-engages.
4. **Valuation Implication:** Because U.S. data rights are *fuzzy*, the market discounts the **transferable value** of a dataset that lives on a platform you don't control. Value accrues to *use* (analytics, ad targeting, recommendation engines *inside* the ecosystem), not to *sale of the dataset corpus*. The *hiQ* saga illustrates why: hiQ built a business on LinkedIn-shaped data, but when the pipe was contested, hiQ had **no alienable asset** to sell or factor — only a contingent access stream that evaporated.

5.4 Connecting to the Interaction Mechanism

Link	Explanation
Low title clarity → high buyer risk	A purchaser of hiQ-style data cannot be sure

	LinkedIn won't cut access or sue tomorrow
→ Market approach fails	No liquid market in "LinkedIn-public-profile derivatives" because each bundle's enforceability is case-specific
→ Income approach survives *only* inside the platform	Value is captured as *service revenue* (B2B analytics subscription), never as *data asset sale*
→ Cost approach is the only "safe" book-value floor	hiQ's costs (scraper infra, QA turkers, algorithms) are real, but they don't map to a tradable capitalized multiple

Case II: The EU Sui Generis Database Right, *Ryanair v. PR Aviation*, and the Shift Toward Managed Access (Data Act / DGA)

5.1 The Sui Generis Database Right in Action

The EU's **Database Directive 96/9/EC** was the world's most ambitious attempt to create a *legislative property right* in data-as-investment. Its Article 7 says:

The maker of a database … shall have the right to prevent extraction and/or re-utilization of the whole or a substantial part … evaluated qualitatively or quantitatively.

This was designed for exactly the kind of situation *hiQ* presents: someone takes the fruits of your substantial data-collection investment without sharing the cost.

***British Horseracing Board v. William Hill* (ECJ 2004)**

The ECJ clarified that the *quantitative/substantial* test looks at the **effort and investment in obtaining/verifying** data, not the *creative* test of copyright. This cemented the "sweat-of-the-brow" theory into positive EU law.

***Ryanair Ltd v. PR Aviation BV* (CJEU 2015)**

Ryanair operated a website publishing its low-fare flight schedules — technically accessible to all — and explicitly prohibited third-party commercial use in its ToS. PR Aviation scraped the site for a meta-search engine. Ryanair claimed *both* database right and copyright.

The CJEU held:

- Ryanair's schedule database **did qualify** as a protected database (substantial investment)
- BUT: because Ryanair chose to **publish the data publicly online**, it **exhausted** its right to prohibit *re-utilization* under Art. 7 — **unless** it imposed effective technical restrictions
- However, under the **copyright** layer (Art. 3 originality test), the *selection/arrangement* might still be enforced, but Ryanair's was too thin

Key takeaway: Even the EU's strongest data-entitlement right has a **publication-exhaustion** flank. If you put it out there without technical access control, your Art. 7 exclusivity drains.

5.2 The Post-GDPR / Data Act Rebalancing

By the late 2010s, Brussels recognized a problem: the *enclosure* logic of Art. 7, layered atop GDPR's risk-aversion, was producing **data hoarding** — the opposite of an AI-ready single market.

Two responses:

(A) Data Governance Act (DGA, 2022/868):

- Creates **data intermediation services** — neutral, EU-supervised brokers that connect data holders and seekers without the holder losing control
- Enables **re-use of public-sector data** (subject to confidentiality/PII safeguards)

(B) Data Act (2023/2854):

- Mandates **IoT/connected-product data** be made available to the *user* and, under conditions, to third-party service providers
- Introduces **B2B data-sharing obligations** where contractual imbalances exist
- Limits **unfair contractual clauses** that over-block data use

5.3 Business-School Interpretation

Phase	Logic	Effect on Valuation

Art. 7 (1996)	Protect investment → encourages data creation	Supports *income approach*: exclusivity lets you charge license fees
GDPR overlay (2018)	Privacy ceiling → massive compliance overhead	Adds *regulatory-risk discount* to any DCF; inflates cost base
DGA + Data Act (2022–2024)	Managed-access turn → data as *shared infrastructure*	Pushes valuation toward **access-service models** (API tiers, intermediation fees) rather than lump-sum dataset sales

The EU model reveals a fascinating institutional evolution: it began by **creating** a data property right (Art. 7), then spent the next 20 years **building access valves** (GDPR portability, DGA intermediaries, Data Act mandates) because pure enclosure choked the downstream economy.

For AI firms, this means:

- **The cleanest EU-playbook play** = build on **non-personal/public-sector-open** data (data.europa.eu has 1.9M+ datasets) where Art. 7 and GDPR don't bite
- For *personal* data: budget for **pseudonymization/anonymization pipelines** as a *capitalizable cost* (cost approach) — because only anonymized data escapes GDPR's heaviest constraints and becomes a freely tradable input

Case III: China's Data-IP Registration Pilot System and the Platform-Data Misappropriation Line

5.3.1 The Judicial Foundation: AUCL and Platform Data

Before the CNIPA registration system matured, Chinese courts developed a **misappropriation doctrine** under the Anti-Unfair Competition Law to protect platform data.

Representative: The *Taobao / Alibaba v. Guodu / Meijing* line of cases and broader platform-data rulings established the principle that a platform's **substantial investment** in collecting, organizing, and maintaining transactional and behavioral data creates a **protectable interest** against rivals who scrape or mirror that data without equivalent effort.

The AUCL path shares surface DNA with *hiQ* (both ask "did the defendant free-ride on the plaintiff's data-labor?"), but the **default tilt differs**:

- U.S.: default = *flow*, intervention = exceptional (must climb CFAA/copyright/misappropriation hurdles)
- China: default = *platform's interest protected* against "freeloading," especially where the scraper creates **substitute service** that erodes the platform's ecosystem

This is not "absolute ownership" — Chinese doctrine still navigates the *data-as-fact* problem — but it is a **stronger competitive-shield** than the U.S. trade-secret/contract combo alone.

5.3.2 The Registration Pilot System: Turning Equitable Interests into Bankable Titles

The truly distinctive China-development is **administrative registration**:

- **CNIPA** launched data-IP pilots in 8 provinces (2022), expanded to 17 (2023–2024)
- Provincial **Data IP Registration Management Measures** (e.g., Shanxi's 10-department joint issuance) define:
  - **Registrable object:** lawfully obtained, processed data collections showing "intellectual achievement attributes" + practical value
  - **Registrant:** the *data processor* — the entity whose labor/tech transformed raw data
  - **Certificate effect:** *prima facie* evidence of holding/use/operation rights; usable in pledging, licensing, securitization, and as basis for infringement

action

Scale indicators (through 2024–2025):

15,600 registrations across pilot platforms; Zhejiang, Fujian, Jiangsu dominate (~90%)

- Shanghai: 1200+ applications, 800+ certificates issued; cited economic circulation value exceeding **¥195 billion** via licensing, trading, and service arrangements
- Nationally, the "Shu-Zhi-Tong " platform links 8 national trading exchanges; 1126 conversions closed, **¥3.9B** in transaction value reported

5.3.3 Business-School Interpretation

China is essentially performing **institutional derivative trading** on data:

1. **Declare** data a factor of production
2. **Define** who counts as the "processor" entitled to the equity
3. **Register** that entitlement with state-backed evidentiary force
4. **Connect** registration → trading venues → financing channels

This solves, administratively, the problem the U.S. leaves to litigation and the EU solves partially through legislation-but-then-restricts-through-privacy.

Comparison	China's Registration Advantage	China's Registration Risk
Title certainty	✓ Registry > contract-alone	⚠ Still *soft law* (pilot measures, not a national statute)
Liquidity	✓ Direct pipeline to exchange/pledge	⚠ Market is *engineered* — demand may follow policy, not pure price discovery
Privacy/security	DSL+PIPL vetting embedded in review	⚠ Cross-border transfer of "registered" data still constrained for "important data"
Cost basis	Aligns perfectly with **cost approach** (build → register → capitalize)	May over-state value if registered-certificates trade illiquidly

Strategic read: For a multinational AI firm operating in China, the smart move is *not*

to fight the old "does data ownership exist?" battle, but to **play the registry game early** — register your curated training datasets (especially synthetic/non-PII corpora) in pilot zones, because the certificate is becoming the **socially accepted currency** for joint ventures, local partnerships, and regulatory comfort.

6. The Interaction Mechanism: How Legal Form Determines Valuation Form

We now return to the paper's core analytical claim: **legal architecture → valuation methodology → realized market value**.

6.1 Transaction-Cost Channels (Klein–Coase–Williamson Logic Applied to Data)

Channel	U.S.	EU	China
Rights delineation cost (how hard is it to describe what you're selling?)	High (no registry; defined by ToS/agreement)	Medium (Art. 7 defines "database" statutorily, but GDPR adds definitional sprawl)	Declining (registry supplies a defined "data product" ID)
Due-diligence cost (can buyer verify provenance/PII risk?)	Medium–High (depends on seller's internal docs)	High (GDPR audit trail mandatory — good for verification, costly to compile)	Medium (registration requires documented lawful-source + processing history)
Enforcement/remedy predictability	Low–Medium (case-by-case; CFAA uncertain post-*Van Buren*)	Medium–High (statutory rights + CJEU precedent, but damages quantification difficult)	Medium (certificate = prima facie evidence; AUCL backs platform claims; but regime still maturing)
Negotiation cost	High (bespoke contracts; no standard form)	High (DPA/transfer-impact-assessment overhead)	Lower (registered certificates plug into standard exchange/pledge templates)

Net effect: Transaction costs are highest in the U.S. for *third-party dataset sales* (hence why the market stays inside-platform), moderate in the EU (constrained by compliance but supported by clearer entitlements), and *structurally declining* in China's pilot zones (because the registry commoditizes the asset description).

6.2 Risk-Pricing and Discount Rates

Any income-approach DCF for a data asset must apply a risk-adjusted discount rate:

$$V_0 = \sum_{t=1}^{n} \frac{E(CF_t)}{(1 + r + \pi_t)^t}$$

Where π_t is a **legal-uncertainty premium**:

Jurisdiction	Typical π_t intuition (qualitative)	Driver
U.S.	**+200–500 bps** on transferable-data DCF	Title unclarity; downstream scraping risk; CFAA/contract litigation volatility; *hiQ*-type unpredictability
U.S. (internal-use only)	Baseline (no π or small)	Inside the firm, you don't need to prove title to yourself
EU	**+150–400 bps** on PII-containing datasets	GDPR enforcement risk (up to 4% global turnover), Schrems-II transfer invalidation, DSR fulfillment costs
EU (anonymized/open)	Near-baseline	Once PII stripped to legal standard, Art. 7 may apply but practical risk drops
China	**−100 bps** (policy tailwind) to **+200 bps** (cross-border freeze risk)	Registration *lowers* enforceability risk domestically; but

		DSL/PIPL "important data" rules can *suddenly* block an offshore transfer, wiping value

From a CFO's perspective: the **same dataset** (say, a 10-million-profile consumer-behavior corpus) carries radically different *discounted* NPV depending on which legal container it sits in.

6.3 Which Valuation Approach Flourishes Where — and Why

Recall the **cost–income–market triad** now embedded in China's *Data Asset Valuation Guidance*:

(A) Cost Approach —

$$V \approx \text{Replacement Cost} \times \text{Utility Coefficient}$$

$$V = TC(1+R) \cdot U - \text{depreciation}$$

- **Universal.** Works everywhere because *input costs are always observable*.
- In all three jurisdictions, this is the **floor** — the minimum below which no rational seller transacts.
- China actively promotes this for **data-asset capitalization on balance sheets** (intangible-asset treatment; cost accumulation: collection + cleaning + labeling + governance + storage).
- In the U.S., firms quietly use it for **internal cap-ex / cloud-cost allocation** but shy from publishing "data asset" line items without clearer SEC guidance.
- In the EU, GDPR compliance costs *increase TC* → raising the floor, but also reducing the margin between floor and potential market value.

(B) Income Approach — Multi-Period Excess Earnings & Royalty Reliefs

The income approach seeks to answer: *What is the present value of the future economic benefits attributable specifically to the data asset?* It is the most theoretically aligned with the concept of "property" — but also the most fragile when property rights are fuzzy.

Jurisdiction	Dominant Income-Method Variant	Legal Architecture Driver	Business Implication

United States	**Relief-from-Royalty (RFR)**	Since you cannot easily *sell* the data corpus, you *license* access. RFR calculates value based on what a hypothetical licensee would pay to avoid infringement.	Favors large platforms acting as licensors (e.g., Twitter API pricing). Difficult for startups to justify high multiples for data-heavy acquisitions.
European Union	**Multi-Period Excess Earnings Method (MPEEM)**	Requires a stable, predictable entitlement (Art. 7) to isolate data-contributed cash flows. GDPR forces rigorous data-governance logs, ironically improving cash-flow attribution.	Necessary for M&A due diligence of EU data firms. Buyers demand "data room" proof of GDPR compliance as a condition for accepting the forecast.
China	**Direct Income Capitalization**	CNIPA registration provides a *legal presumption* that the cash flows belong to the registrant, reducing the need for complex apportionment.	Enables data assets to be used as collateral for loans (e.g., Shanghai Pudong Development Bank accepting data-IP certificates as security).

The Legal Constraint on Cash Flows: Under the income approach, the valuator must forecast the **Economic Life (n)** of the data asset. Here, legal divergence is stark:

1. In the **U.S.**, n is limited by the speed of technological obsolescence and the risk of a *hiQ*-style legal challenge. A dataset's economic life may be contractually capped by the ToS.
2. In the **EU**, n is formally capped by the **15-year term of the sui generis right** (renewable upon substantial update). This provides a clear horizon for DCF modeling.
3. In **China**, n is increasingly tied to the **validity of the registration** and the "useful life" defined by the data's application scenario, often supported by government-led scenario libraries.

Discount Rate Sensitivity: As established in Section 6.2, the discount rate (r) acts as the transmission belt for legal risk. A U.S. court's refusal to recognize a data property right (as in *Feist*) does not destroy the data's utility, but it drastically increases the **specific risk premium** within r, crushing the Net Present Value (NPV) for any external investor. Conversely, China's registration system aims to compress r by providing state-backed legal certainty.

(C) Market Approach — Comparable Transactions & the Liquidity Gap

The market approach (Value $\approx$ Price of comparable assets in the market) is the gold standard for valuation because it reflects actual supply and demand. However, it is the most dependent on **standardized property rights**.

The "Data Commodity" Problem: Real estate is a mature asset class because a "house" is a well-defined legal object. Data is not. Without a standardized legal wrapper, every dataset is a unique, non-fungible asset, making comparables nearly impossible to find.

Jurisdiction	Status of Market Approach	Mechanism
United States	**Severely Constrained**	Lack of standardized titles means no liquid exchange. Transactions are bespoke M&A deals (e.g., acquisition of a data-rich startup), not spot-market sales of datasets.

European Union	**Nascent**	Emerging data intermediation services (DGA) aim to create a marketplace. However, GDPR compliance costs and liability concerns limit the volume of "clean" comparables.
China	**Actively Engineered**	CNIPA registration + regional exchanges (Guiyang, Beijing, Shanghai) create **standardized data products**. The "Shu-Zhi-Tong" platform provides actual transaction prices, enabling the market approach to become viable for the first time globally.

Conclusion of the Mechanism Analysis: The legal form dictates the valuation form. The U.S. system, optimized for innovation diffusion, yields high **use value** but low **exchange value**. The EU system, optimized for rights protection, yields moderate use value but high compliance costs that depress multiples. China's emerging system is explicitly designed to maximize **exchange value** by manufacturing the legal prerequisites for a market.

7. Discussion: Strategic Implications for AI-Facing Firms

7.1 For Multinational AI Developers (e.g., OpenAI, Anthropic, Google DeepMind)

- **Training Data Sourcing:** The U.S. legal environment allows for broad scraping of public data, but creates **litigation risk** regarding outputs and downstream use. The EU requires strict **provenance tracking** (GDPR/Data

Act). China requires **localization and registration** of key datasets.

- **IP Strategy:** In the U.S., focus on **trade secrets** for model weights and **patents** for novel architectures, while treating training data as a consumable input rather than a capital asset. In China, register your curated, synthetic, or non-personal training datasets immediately to secure a defensive moat and enable future financing.

7.2 For Data Intermediaries and Brokers

- **Productization:** To succeed in the EU, productize data as a **managed service** (API access) rather than a bulk file transfer to mitigate GDPR liability.
- **Standardization:** In China, align your data offerings with the CNIPA registration criteria (lawful source, substantial processing, clear utility) to gain access to the state-backed trading exchanges.

7.3 For Financial Institutions and Investors

- **Due Diligence:** When valuing a data-driven target in the U.S., heavily discount the "data asset" line item unless it is underpinned by strong trade-secret protocols. In the EU, scrutinize the GDPR compliance budget as a capital expenditure. In China, treat the CNIPA registration certificate as equivalent to a patent in terms of asset quality.
- **Risk Modeling:** Incorporate **jurisdictional risk premiums** into DCF models. A dataset located in California (CCPA) or Virginia (CDPA) carries different regulatory risk profiles than one in Shanghai or Frankfurt.

8. Conclusion

This paper has argued that the debate over "data ownership" is a distraction. The critical question for the AI economy is not *who owns it*, but *what legal architecture best facilitates the valuation and exchange of data as a factor of production*.

Through a comparative analysis of China, the United States, and the European Union, we find three distinct institutional equilibria:

5. **The United States** operates a **litigation-driven, contract-based system**. It excels at maximizing the *flow* of data for domestic AI development but fails to provide the property-rights certainty required for robust secondary markets. Value is captured through **vertical integration** and **service delivery**, not asset trading.
6. **The European Union** operates a **rights-centric, regulatory-intensive system**. It provides the clearest *formal* property rights (sui generis database right) but overlays them with such significant privacy and access obligations (GDPR, Data

Act) that the net value extracted from data is heavily taxed by compliance costs.

7. **China** is pioneering a **state-engineered, registration-based system**. By decoupling data use rights from nebulous ownership concepts and creating administrative titles (CNIPA registration), it is attempting to solve the "market failure" of data valuation. While still experimental, it offers the most promising pathway to treating data as a **bankable, tradable capital asset**.

The interaction mechanism is clear: **legal clarity reduces transaction costs, lowers discount rates, and enables the market approach to valuation.** As AI continues to scale, the jurisdictions that successfully engineer this legal-clarity pipeline will capture the lion's share of the data-finance economy. For businesses, the imperative is to map their data strategies onto these divergent legal landscapes, recognizing that the same data asset commands fundamentally different values depending on the flag it flies.

9. References

(Note: This section would contain full academic citations. Below is a representative sample formatted for a business/economics journal.)

4. **Acemoglu, D., & Restrepo, P.** (2019). Automation and new tasks: How technology displaces and reinstates labor. *Journal of Economic Perspectives*, 33(2), 3-30.
5. **Bessen, J.** (2018). *AI and Jobs: The Role of Demand*. Boston University School of Law Working Paper.
6. **China National Intellectual Property Administration (CNIPA).** (2023). *Notice on Determining the Second Batch of Data Intellectual Property Pilot Locations*.
7. **European Commission.** (2023). *Regulation (EU) 2023/2854 of the European Parliament and of the Council on harmonised rules on fair access to and use of data (Data Act)*.
8. **Goldman, E.** (2019). The hiQ v. LinkedIn Litigation and the Future of Web Scraping. *Santa Clara High Technology Law Journal*, 35(4), 467-502.
9. **Handke, C.** (2006). *Copyright vs. Database Right*. Review of Economic Research on Copyright Issues, 3(2), 3-20.
10. **Huang, Y.** (2021). *The Rise of Data as a Factor of Production: China's Evolving Data Governance Regime*. Carnegie Endowment for International Peace.
11. **Kerber, W.** (2016). *The Economic Aspects of Data Ownership*. Max Planck Institute for Innovation & Competition Research Paper No. 16-13.
12. **Mayer-Schönberger, V., & Ramge, T.** (2018). *Reinventing Capitalism in the Age*

of Big Data. Basic Books.

13. **Posner, E., & Weyl, E. G.** (2018). *Radical Markets: Uprooting Capitalism and Democracy for a Just Society*. Princeton University Press.
14. **Samuelson, P.** (2022). *Information Property Law 2.0*. Berkeley Technology Law Journal, 37(1), 1-68.
15. **State Council of the People's Republic of China.** (2020). *Opinions on Building a Data Foundation System to Better Play the Role of Data Elements* (Data Twenty Articles).
16. **Weber, R. H.** (2015). *Realizing a Data Economy*. Swiss Review of International Economic Relations, 70(3), 257-275.
17. **Zech, H.** (2016). *The Law of Data and Algorithms*. Oxford Journal of Legal Studies, 36(4), 811-839.