

Commons Collective.

What would it take for people to negotiate with AI
at the scale at which AI negotiates with society?

Collective bargaining institutions for the AI data commons.
A research agenda, institutional design, and testable economic model.

DISCUSSION DRAFT • v0.1



Scan to open the project
[ethical-tech-
colab.github.io/commons-collective/](https://ethical-tech-colab.github.io/commons-collective/)

Abstract

Research question. Under what conditions can collective bargaining institutions (CBIs) convert dispersed contributions to AI into effective agency, fair compensation, and sustained funding for shared knowledge, without privatizing the public domain or turning privacy into a luxury?

Thesis. A viable institution should aggregate negotiating mandates rather than indiscriminately aggregate data. It should distinguish three functions: buying AI services collectively; bargaining over uses that contributors can actually authorize; and financing public goods that no individual contributor owns. We call the coordinated arrangement **Commons Collective**. Its consumer-facing expression is a **digital Costco**: a community-governed buyers' and suppliers' club that uses scale for people, while maintaining distinct purchasing, contributor-licensing, and commons-finance functions. This is an analogy, not a proposed affiliation with Costco.

Method and contribution. This discussion draft combines a targeted literature and primary-source review with institutional comparison, legal issue-spotting, a technical permission-and-receipt architecture, and reproducible financial scenarios. It proposes a rights-aware knowledge object, a separated-governance federation, a commons-benefit covenant, and a staged empirical program. It does not report completed fieldwork, demonstrated bargaining gains, or a validated willingness-to-pay estimate.

Finding at the level of design, not empirical proof. Small individual data payments are unlikely to be a sufficient theory of collective agency. The stronger proposition combines purchasing savings, enforceable service conditions, contributor compensation where authority exists, representation for affected noncontributors, and public-good finance. Every component must survive a test against its own costs and possible harms.

Status: 18 September 2026; version 0.1; AI-assisted synthesis; **not peer-reviewed**. All monetary scenarios and institutional designs below are proposals unless identified as evidence. No funding, partnership, regulatory approval, or participant endorsement is claimed.

1. Research circle and ethical commitments

Session provenance and attribution

The requested research circle is **Tomicah Tillemann, Beth Goldberg, Yorke Rhodes, Glen Weyl, Nick Vincent, Ilan Strauss, Sonam Jindal, and Anna Tumadottir (Creative Commons, as identified in the brief)**. Names reflect the brief, subsequent corrections, and the printed photograph where legible; affiliation details, attendance, authorship, and approval require participant confirmation. Inclusion is not a claim that any individual wrote, reviewed, or agrees with this paper.

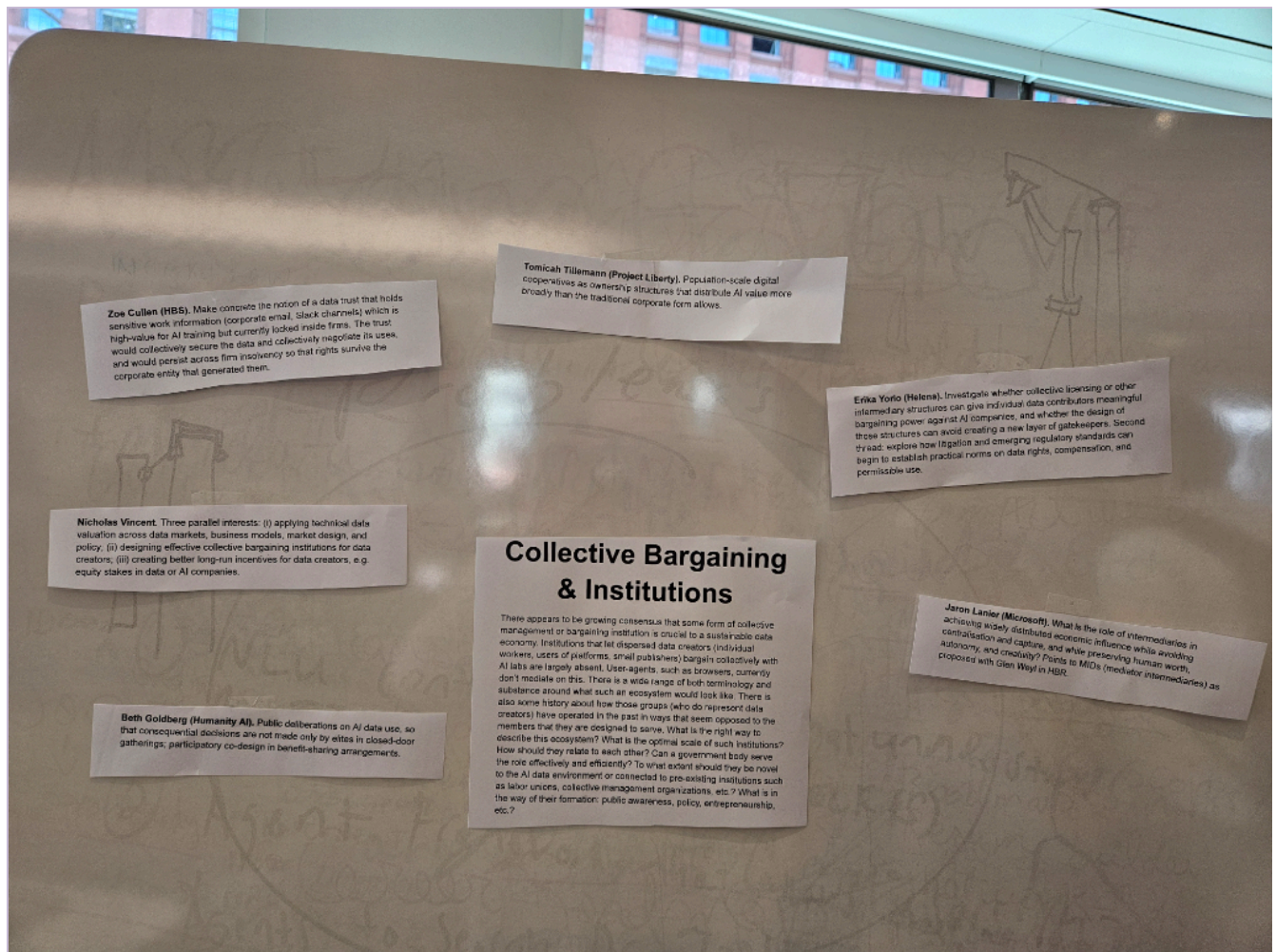
Roundtable reference: [Roundtable on Mechanisms for a Sustainable AI Data Economy | AI Disclosures Project](#). The convening page describes the 18 September 2026 working session at Microsoft Research New York and its focus on mechanisms across pre-training, post-training, and inference. It supplies context for this research, not participant endorsement of this draft. ^[S96]

At the user's request, the [workshop source record](#) now includes the original supplied photograph, a crop of every printed card, and AI-assisted visual transcriptions. The overall group brief appears below; each attributed direction is also placed beside the relevant research. The names and affiliations on the cards are preserved as printed, not treated as verified spoken quotations or endorsements. The photograph's association with 18 September 2026 comes from the supplied context, not image metadata. ^[S97]

The user also supplied an excerpt titled *Blueprint for a Collective Bargaining Institution for Data in the Public Interest*. Its [coverage review](#) distinguishes source wording, editorial interpretation, additions, and premises requiring qualification. The pasted excerpt was reviewed; the authenticated Google Doc and its other tabs were not independently accessed. Its author line is a source record, not automatic authorship of this publication. ^[S99]

Original group brief and directional source map

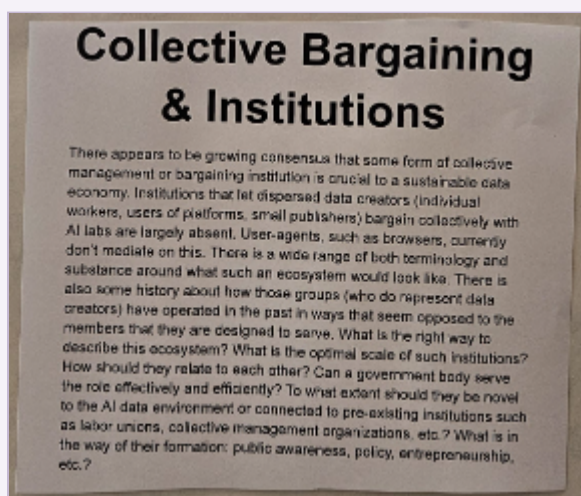
The photograph poses institutional questions; its claims about consensus, absent institutions, or browser mediation are not adopted as independently established findings. Original wording is retained so readers can inspect the framing rather than infer it from our synthesis.



User-provided photograph of the Collective Bargaining & Institutions workshop whiteboard, containing one group brief and six individually attributed printed cards. Published at the user's request. Copyright in the photograph and depicted material remains with its respective rights holders. No photographer attribution or Creative Commons license is established by the supplied image. ^[S97]

OVERALL GROUP BRIEF / SOURCE TRANSCRIPTION

Collective Bargaining & Institutions



Attribution as printed: No individual attribution printed. Open the crop to inspect it.

There appears to be growing consensus that some form of collective management or bargaining institution is crucial to a sustainable data economy. Institutions that let dispersed data creators (individual workers, users of platforms, small publishers) bargain collectively with AI labs are largely absent. User-agents, such as browsers, currently don't mediate on this. There is a wide range of both terminology and substance around what such an ecosystem would look like. There is also some history about how those groups (who do represent data creators) have operated in the past in ways that seem opposed to the members that they are designed to serve. What is the right way to describe this ecosystem? What is the optimal scale of such institutions? How should they relate to each other? Can a government body serve the role effectively and efficiently? To what extent should they be novel to the AI data environment or connected to pre-existing institutions such as labor unions, collective management organizations, etc.? What is in the way of their formation: public awareness, policy, entrepreneurship, etc.?

This records a printed card, not a verified spoken quotation or endorsement. ^[597]

▼ Editorial alignment and coverage

The overall institutional brief: describe the ecosystem, identify legitimate representation, choose scale and relationships, compare existing and new institutions, and explain how they form.

The brief is now reproduced in section 1. Its questions are traced across institutional comparison, regulatory options, the research program, and the federation blueprint; these remain research questions rather than demonstrated outcomes.

- **Define the institutional ecosystem and its relation to existing organizations**

[3. The commons problem and the grazing analogy](#) · [5. From digital Costco to a bargaining federation](#) · [20. Follow-on research: a Commons Collective Framework Agreement](#)

- **Compare scale, coordination, public authority, and institutional formation**

[16. Research program and falsifiable hypotheses](#) · [23. Institution blueprint: a Commons Collective Federation](#) · [29. Replicable sector collectives: blueprint and templates](#)

[View source context and full mapping](#)

Printed attribution	Editorial direction summary	Primary research placement
Zoe Cullen (HBS).	A workplace-data trust for collective protection and negotiation, with a specific continuity question when the originating firm becomes insolvent.	20. Follow-on research: a Commons Collective Framework Agreement Read the aligned card
Tomicah Tillemann (Project Liberty).	Population-scale digital cooperatives as a route to broader ownership and distribution of AI value.	5. From digital Costco to a bargaining federation Read the aligned card
Erika Yorio (Helena).	Collective licensing and intermediaries should improve real bargaining power without becoming new gatekeepers; litigation and emerging regulation can be studied as routes to practical norms.	13. Regulatory framework: present rules and proposals Read the aligned card
Nicholas Vincent.	Connect data valuation and market design to effective contributor institutions and long-term incentives, including the possibility of equity participation.	7. Business model and economic discipline Read the aligned card

Printed attribution	Editorial direction summary	Primary research placement
Beth Goldberg (Humanity AI).	Public deliberation and participatory design should shape consequential AI-data decisions and benefit-sharing arrangements, rather than leaving decisions to closed elite forums.	16. Research program and falsifiable hypotheses Read the aligned card
Jaron Lanier (Microsoft).	Assess intermediaries by distributed economic influence and protection of human worth, autonomy, and creativity, while resisting centralization and capture; retain the card's reference to MIDs and related work.	23. Institution blueprint: a Commons Collective Federation Read the aligned card

Alignment with repository-stated Ethical Tech CoLab principles

A pinned public-repository snapshot of Ethical Tech CoLab's website content identifies **Field-grounded**, **Prototype-first**, and **Open by default** as working principles. This is stored About-page copy, not verified wording on a rendered public principles page: the [About route is disabled in the cited source](#). The [public website](#) provides organizational context. The separate ethical-AI paper emphasizes institutions, human-rights constraints, avoidance of harm, and participation by affected people. We use these as design references, not an invented official Ethical Tech CoLab charter or an endorsement of this research. ^[S01] ^[S02]

Source-stated principle or commitment	Operational requirement for this project	Evidence needed before expansion
Field-grounded	Compensated design sessions with consumers, creators, community stewards, and people affected by data about them	Decisions changed by participants; accessible redress; inclusion of people who refuse participation
Prototype-first	Test membership, mandate comprehension, refusal, and settlement with fictional or low-risk material before acquiring personal data	Measured understanding, actual negotiated terms, reconciled payments, and adverse-event records
Open by default	Publish protocols, evaluation methods, code, budgets, and negative findings	Reproducible calculations and independent scrutiny, without disclosing private data
Human-rights floor and do-no-harm	Do not exchange essential services or fundamental protections for permission to train models	Independent ethics review and a practical way to participate without selling sensitive information
Affected-person participation	Represent nonmembers, correspondents, and communities, not only monetizable rightsholders	Accessible deliberation, funded participation, conflict disclosures, and a grievance process

Interpretive boundary: openness applies to methods and accountability, not to confidential records. A member vote cannot waive another person's rights or legitimate community governance.

2. Method, scope, and limits

This is a **scoping synthesis and design proposal**, not a systematic review. Sources were selected for direct relevance to common-resource governance, data labor and leverage, applicable legal instruments, open licensing, philanthropic intent, provenance, consumer privacy, and energy demand. The evidence register distinguishes what each source supports from what it does not establish. Primary legal texts and official institutional publications take precedence over commentary for descriptions of rules and programs.

For a shorter standalone argument, read the [prioritized first-draft paper](#). The [library-led institutional example](#) applies the design to one hypothetical open-education collective. Both are separate working drafts with their own abstracts, limits, and bibliographies; neither replaces the evidence boundaries of this living research record.

AI usage disclosure: the [model-by-model usage audit](#) documents a bounded local-ledger snapshot for the production of these materials using CoLab's usage-calc calculations. It distinguishes recorded requests and token charges from dollar-conversion assumptions, actual billing, human labor, and unobserved activity. The raw conversation and individual request records are not published; usage volume does not establish the quality or truth of this research.

The principal legal comparison is the European Union and United States. It is not a complete global survey, and rules differ within and beyond these jurisdictions. The original scoping review used sources accessed on 18 September 2026; later additions carry their individual access dates in the source register. Legislation, guidance, litigation, and product settings may change. A pilot needs a jurisdiction-specific legal memorandum immediately before launch, including current amendments and implementation schedules.

We use three epistemic labels in substance:

- **Evidence:** a proposition grounded in a cited source, including the existence or wording of a policy.
- **Interpretation:** a reasoned implication that is not itself directly established by the source.
- **Proposal:** a testable design choice, budget, threshold, or future legal intervention.

The publication provides inspectable settlement arithmetic and a fictional object, not a functioning cooperative, production data vault, compliance certification, payment processor, or deployed training-control system. Source selection remains vulnerable to Anglophone, institutional, and Global North bias. Later work should include multilingual community-led inquiry and a preregistered literature-search protocol.

3. The commons problem and the grazing analogy

What carries over from public-land grazing

The analogy is useful because an individually profitable activity can depend on collectively maintained resources. Public-land grazing offers an institutional vocabulary of access conditions, monitoring, stewardship, and charges. Commons scholarship supplies a deeper lesson: durable governance depends on legitimate boundaries, rules fitted to context, participation, accountability, and nested institutions, rather than an inevitable choice between privatization and unconstrained access. ^[S03] ^[S04]

For AI, the underlying commons includes maintained reference works, public research, language communities, public archives, software documentation, cultural institutions, and the social practices that continually renew knowledge. A developer may capture commercial value while underfunding this renewal.

Where the analogy breaks

Forage is physically subtractable: one animal's consumption reduces what remains. Copying information ordinarily does not use up the original. Economic models of nonrival data identify gains from broader use, but those results do not imply that every kind of sharing is socially desirable. AI-related scarcity appears in maintenance labor, attention, trustworthy provenance, privacy, community authority, energy, and the economic viability of future creation. Knowledge commons therefore require their own institutional analysis rather than a literal enclosure-and-rent model. ^[S03] ^[S25] ^[S31]

There is no single global owner of "the data commons" able to invoice every AI model. Nor is the absence of a universal mechanism the absence of all mechanisms: copyright licensing, subscriptions, public appropriations, philanthropy, procurement conditions, and existing collective management already allocate some benefits. Their coverage, legitimacy, and distribution are incomplete.

Proposal: commons-benefit covenant. A buyer voluntarily contracts to fund specified shared resources alongside any licensed inputs. The payment purchases stewardship, reliable access, curation, evaluation, or a contractual benefit commitment; it does not invent exclusive ownership of public facts. A mandatory levy would require an independently justified statutory basis. A covenant binds its parties, not the entire world.

Analogy	Useful institutional idea	Boundary that must not be crossed
Public-land grazing	Access terms and resource stewardship	Treating copied public knowledge as physically depleted forage
Labor union	Representation and bargaining over conditions	Assuming all contributors are employees with labor-law exemptions
Consumer cooperative	Pooled demand, service standards, patronage accountability	Requiring members to surrender data to obtain essential services

Analogy	Useful institutional idea	Boundary that must not be crossed
Collective rights organization	Mandates, licensing, reporting, distributions	Licensing rights the institution does not control
Public library or commons steward	Access, preservation, and inclusive public benefit	Letting only paying contributors define the public interest

4. A spectrum, not a single data commodity

The requested sequence is **mass-public -> private -> professional -> synthetic**. It is a useful conversational map, but not a legally ordered ladder. A professional work can be public; private records can contain copyrighted works; synthetic records can reveal private information or incorporate protected expression. Describe each asset along independent axes.

The four principal archetypes

Cate gory	Examples	Potential bargaining counterparty	What can legitimately be offered
Mass - publi c	Public-domain texts, public facts, openly licensed reference collections	Stewards, publishers, maintainers, public bodies, or procurement coalitions	Curation, service quality, access infrastructure, attribution compliance, voluntary commons support; no new ownership of public-domain material
Priva te	A personal archive, household messages, nonpublic workplace records	Individuals, organizations with authority, and affected third parties	A narrowly authorized processing service or limited access, subject to privacy, confidentiality, and contractual constraints
Profe ssion al	Commissioned writing, photography, design, specialist annotations	Creators, employers, assignees, publishers, or mandated representatives	Precisely identified rights or services after chain-of-title and employment/assignment review
Synt hetic	Generated text, simulated records, distilled outputs, model-derived datasets	Generator operator plus any relevant upstream rights and affected persons	A verified product with lineage, utility, confidentiality, and license review; not an automatic clean-rights substitute

Required descriptive axes

1. **Access:** public, gated, confidential, or highly restricted.
2. **Legal authority:** public domain, specified license, rightsholder mandate, employment assignment, contract, or unresolved.
3. **Privacy and sensitivity:** nonpersonal, personal, special-category/sensitive, or community-sensitive.
4. **Origin and transformation:** human-authored, observed, inferred, synthetic, or mixed, with derivation records.
5. **Use:** inference, retrieval, evaluation, fine-tuning, foundation-model training, or redistribution.
6. **Structure:** structured rows; semi-structured records; unstructured text/audio/image/video; relational or graph data; embeddings and model-derived representations.
7. **Renewal and quality:** static versus streaming; freshness; validation; contamination risk; maintenance costs.
8. **Collective context:** affected communities, shared authorship, correspondents, and restrictions on cultural use.

Implication: price must attach to a defined bundle of authority, service, quality, and obligations, not to an undifferentiated "data point." Greater structure may lower integration costs, but does not confer rights. An embedding is not evidence that copyright, privacy, or confidentiality obligations have vanished.

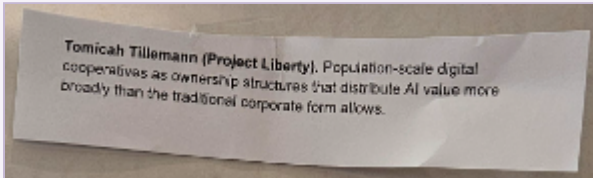
Research on recursive synthetic-data training shows failure modes under studied conditions; it does not establish that all synthetic data are defective. Synthetic generation can be useful when composition, evaluation, and retained real-world reference data are appropriate. This makes lineage and task-specific evaluation more important, not less. ^[S22]

Similarly, the EDPB rejects an automatic inference that a model trained on personal data is anonymous. Whether extraction or disclosure is reasonably likely needs assessment; this does not imply the opposite categorical claim that all models or synthetic outputs contain personal data. ^[S34]

5. From digital Costco to a bargaining federation

PHOTOGRAPHED DIRECTION / SOURCE TRANSCRIPTION

Tomicah Tillemann (Project Liberty).



Attribution as printed: Tomicah Tillemann (Project Liberty).. Open the crop to inspect it.

Population-scale digital cooperatives as ownership structures that distribute AI value more broadly than the traditional corporate form allows.

This records a printed card, not a verified spoken quotation or endorsement. ^[S97]

▼ Editorial alignment and coverage

Population-scale digital cooperatives as a route to broader ownership and distribution of AI value.

Member governance, buyer/supplier organization, and benefit return were already central. The ownership-versus-governance distinction and the trade-off between population scale and accountable representation are now made explicit.

- **Design ownership and governance that distribute benefits rather than merely enlarge membership**
[5. From digital Costco to a bargaining federation](#) · [8. Payment, disbursement, and public-good return](#)
- **Compare nested, sectoral, and population-scale arrangements**
[23. Institution blueprint: a Commons Collective Federation](#) · [29. Replicable sector collectives: blueprint and templates](#)

[View source context and full mapping](#)

The institution to build is a proposed Commons Collective Federation (CCF): a member-governed organization that buys accountable AI services, represents specifically authorized contributors, and funds independent commons stewardship. Its first product is an enforceable service agreement, not a universal data pool. The [institution blueprint](#) specifies the functions, mandates, governance, and bounded starting sequence. It supplies institutional support for a [pro-human stack](#): technology whose human agency claims can be tested and enforced.

Two different forms of scale

The brief's "digital Costco" is interpreted as a **digital Costco**, a membership-based aggregation analogy. Costco's own reporting supports the commercial relevance of membership and scale; it does not validate a consumer-owned AI institution or imply that Costco itself is a cooperative. ^[S16]

Demand-side scale: households collectively procure AI assistance, negotiate prices, portability, accessibility, retention limits, and dispute handling. A credible ability to switch suppliers is central. Purchasing leverage exists even when members provide no training data.

Supply-side scale: authorized contributors collectively negotiate compensation, attribution, permitted uses, audit, and remedies. Data-labor and data-leverage research motivates this possibility, but the bargaining effect depends on substitutability, model dependence, coordination, and legal authority. Neither a large membership count nor a large dataset proves scarcity or buyer willingness to pay. ^{[S05] [S06]}

Contribution is not coordinated withholding. The supplied notes distinguish relatively easy participation in multiple collectives from the harder task of sustaining a collective refusal. This is an organizing hypothesis, not proof that contribution is effortless or a data strike effective. Multiple memberships preserve choice but can complicate overlapping mandates, allocation, and credible future-supply commitments. Any coordinated refusal needs a lawful scope, defined target and duration, a decision process, support for participants, and protected dissent. It cannot retract compliant CC permissions, disregard contracts, or damage shared resources; nonemployee coordination requires competition review. ^{[S06] [S13] [S26] [S99]}

Population scale also differs from ownership and accountability. Compare direct mass membership, nested local/sectoral bodies, and federated representation on meaningful control, decision costs, inclusion, and benefit distribution. A large user count does not establish democratic ownership or prove that a conventional corporate form has been displaced.

Public-good representation: people who use open knowledge and communities affected by AI should influence commons spending without first proving ownership of profitable data. This function is not reducible to either consumer purchasing or seller representation.

Relational data-governance scholarship provides a reason for this separation: the interests of people affected by population-level inferences are not exhausted by the preferences of the original contributors. ^[S25]

Proposed three-part architecture

Institution	Principal mandate	Money and authority boundary
Member purchasing cooperative	Buy AI services on behalf of consumers; negotiate enforceable service conditions	Dues and service budgets; no automatic authority to license members' data
Contributor bargaining association	Represent creators and authorized data stewards for specified transactions	Revocable, scoped mandates; separate licensing accounts and conflict procedures
Independent commons trust or nonprofit fund	Support shared knowledge and affected communities	Restricted-purpose funds; public reporting; neither a universal rightsholder nor a proprietary gatekeeper

The federation may share an open protocol, counsel, and procurement infrastructure. It should not share unrestricted data custody by default. A "data trust" is a governance description whose legal implementation depends on jurisdiction; it does not, by its name, prove fiduciary duties, representativeness, or authority. Data-trust scholarship is useful precisely because delegation and fiduciary accountability need explicit design. ^[S07]

Governance and collective agency

Proposed charter: one member, one vote within defined constituencies; elected, recallable representatives; compensated deliberative panels; buyer and supplier chambers; and an independent commons and affected-party council. The federation's constitutional forum coordinates these constituencies while each legal entity retains its own duties. No investor acquires additional votes by purchasing more data or funding the operator.

Material changes to permitted use require approval by the affected constituency and an individual refusal path where individual authorization is needed. New use does not inherit permission from an old vote. Budget rules, representation, and extraordinary decisions should be constitutional; routine implementation should remain delegated so the institution can operate.

A rights-and-ethics steward can pause unsafe processing pending independent review. Its mandate, appeal route, conflicts, and response deadlines must be published. Sponsors cannot veto adverse findings. Paid staff may administer the institution but cannot secretly alter member mandates.

The test of **advancing collective agency** is not total dollars paid. It is whether members understand choices, can refuse without disproportionate penalty, can change negotiators, can switch suppliers, can contest uses, and can redirect surplus toward shared priorities.

A buyers' club, a suppliers' club, and a community

The session's extension sharpens the analogy: a successful club assembles both demand and supply. Buyers gain scale and better terms; suppliers gain predictable demand, lower transaction costs, and a voice in the institution. The aim is **fair, sustainable total value**, not the lowest buyer price at any cost. Evaluate price alongside quality, reliability, contributor livelihoods, privacy, access, and commons maintenance. Two-sided-market theory supports examining these interactions, not assuming that scale alone makes every participant better off. ^[S23]

The phrase **community-governed Soho House** describes shared purpose, community norms, and social infrastructure. It is not a claim that Soho House or Costco is a cooperative or that their suppliers currently hold these proposed voting rights. The intended community is not organized around wealth, status, or exclusivity. No affiliation or endorsement is implied. ^[S16]

Free opt-in does not mean unfunded services

Proposal: make basic participation free, including deliberation, applicable voting rights, and core community resources. Optional dues buy specified purchasing support, managed services, or other cost-bearing benefits. Sponsored participation and waivers can widen access, without granting sponsors votes. Free membership does not promise unlimited free AI consumption.

Fund the model through optional dues, disclosed procurement-service charges, transparent transaction fees, and grants for public goods. Budget cross-subsidies explicitly. Public knowledge can benefit nonpayers by design; intensive private services may need a subscription or subsidy. Measure utilization, renewal, governance costs, and subsidy dependence before claiming a scalable free tier. **Payment buys services, not extra citizenship; membership never requires a data sale.**

Supplier voice must have consequences

The **buyer chamber** represents members consuming the club's services. The **supplier chamber** represents contributors, creators, curators, maintainers, and community stewards supplying the club. A commercial AI vendor selected by procurement is not automatically a voting supplier: a sales contract does not confer constitutional rights. Any overlapping role requires separate qualification, disclosure, and recusal.

The **commons and affected-party council** represents interests beyond the immediate transaction. Its powers should be enumerated, supported by written reasons and an appeal route, rather than an unlimited veto over ordinary operations. The reason for this third body is substantive: people affected by data use may be neither the buyer nor the contributing seller. ^[S25]

Decision	Proposed decision-maker	Protected condition
Fees and service tiers	Each chamber approves charges on its constituency; both approve shared charges	Protect free basic participation and voting; disclose cross-subsidies and waivers
Membership and catalog admission	Joint committee applies published eligibility, quality, and conduct criteria	Written reasons, independent appeal, and no incumbent-protecting exclusion
Procurement and vendor selection	Delegated buyer committee within an approved budget	Recusal; no procurement decision can silently expand contributor permissions
Data uses and contributor terms	Relevant rights holders authorize participation; supplier chamber approves material program changes	Required individual/community authorization survives; council review does not consent for absent people
Commons budget	Both chambers approve allocations with council participation in priorities	Restricted commitments cannot be diverted without specified approvals
Charter and protected rights	Supermajority in each chamber after notice and deliberation	Council assent for changes to its protected remit; no influence purchased through dues
Deadlock and exit	Time-limited mediation followed by the last valid rule, only within its continuing lawful scope	Silence is not consent; expired permissions do not continue; exit preserves earned payment claims

Supplier protections can include timely payment, transparent deductions, and bounded workload commitments. Coordinated price floors require jurisdiction-specific competition review; calling the institution a club does not create an exemption. ^[S13]

Members with overlapping roles can register a primary voting chamber for a fixed term while retaining their separate authorization rights. Representation must not scale with sales, spending, or donations. Duplicate-membership controls should be proportionate, privacy-preserving, minimally retained, and appealable; neither public identity disclosure nor proof of wealth is an acceptable default.

Dual chambers can constrain domination but also create delay and factional capture. Delegate routine operations, reserve cross-chamber approval for consequential choices, publish conflicts, and measure governance costs. A committee structure is valuable only if members gain usable agency.

6. Archetypes and the multi-sided marketplace

A person may occupy several roles. Alice may buy an AI assistant, publish open educational material, and license professional writing. Bob may consume that material, contribute corrections, and live near a data center. Membership must not collapse these identities into one blanket permission.

Archetype	Need and source of leverage	Benefit and main risk
Ordinary consumer	Reliable affordable AI; aggregated purchasing and switching	Savings and stronger terms; risk of tying basic services to data permission
Professional writer or artist	Livelihood, credit, control; differentiated work and rightsholder authority	Contract income and enforceable limits; risk of substitution, underpayment, or publisher conflicts
Open contributor or maintainer	Continued access and maintenance funding; trusted collective infrastructure	Commons grants and improved stewardship; risk of enclosure or payment by popularity
Small business or public-service buyer	Affordable compliant tools; procurement coordination and local knowledge	Lower transaction costs; risk of vendor lock-in and uncertain liability
Community or Indigenous steward	Authority over culturally situated knowledge; governance legitimacy	Negotiated protocols and funded stewardship; risk of individual contracts overriding collective relationships
Model developer or application provider	Usable data/services, predictable permissions, quality, and reputation	Lower clearance and integration costs; risk of unverifiable mandates and holdout costs
Affected noncontributor	A say in profiling, representation, environmental burdens, and downstream effects	Redress and voice; risk of exclusion because there is no monetizable asset to sell

Local Contexts' Traditional Knowledge Labels illustrate how community protocols can travel with cultural materials. They are not a blanket substitute for law or community engagement, and a bargaining platform should not rebrand them as conventional commercial licenses. ^[S20]

Market-design problem. More participants can improve matching and bargaining but also make the intermediary a new gatekeeper. Prefer nonexclusive mandates, portable records, interoperable catalogs, transparent fees, and access to competing representatives. A procurement committee must not obtain individual sellers' confidential reserve prices. A contributor association must not exploit access to consumer behavior for discriminatory pricing.

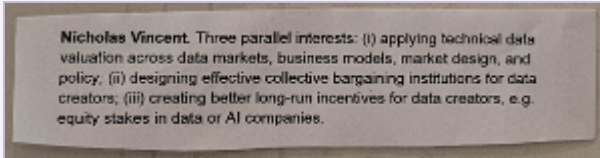
Two-sided-market theory explains why the distribution of charges across participants matters, not only the platform's total revenue. Dues that deter low-income consumers, excessive contributor commissions, or buyer fees that prevent adoption can each undermine the other sides. That theoretical insight does not establish the optimal price schedule for this proposed institution. ^[S23]

The first market should be narrow enough to establish authority and inspect benefits: for example, a curated educational retrieval service with adult contributors, original non-sensitive materials, and a nonprofit/public-interest buyer. Avoid health, children's, workplace-surveillance, and humanitarian-beneficiary data in the initial pilot.

7. Business model and economic discipline

PHOTOGRAPHED DIRECTION / SOURCE TRANSCRIPTION

Nicholas Vincent.



Attribution as printed: Nicholas Vincent.. Open the crop to inspect it.

Three parallel interests: (i) applying technical data valuation across data markets, business models, market design, and policy; (ii) designing effective collective bargaining institutions for data creators; (iii) creating better long-run incentives for data creators, e.g. equity stakes in data or AI companies.

This records a printed card, not a verified spoken quotation or endorsement. ^[S97]

▼ Editorial alignment and coverage

Connect data valuation and market design to effective contributor institutions and long-term incentives, including the possibility of equity participation.

Valuation, business models, and collective bargaining already have dedicated analysis.

Creator equity is now distinguished from cash, royalties, and the separate public sovereign-fund proposal, with unresolved governance and legal trade-offs.

- **Data valuation, business models, market design, and policy**
[6. Archetypes and the multi-sided marketplace](#) · [7. Business model and economic discipline](#) · [27. Private institutions and the commons service model](#)
- **Effective contributor bargaining institutions**
[5. From digital Costco to a bargaining federation](#) · [20. Follow-on research: a Commons Collective Framework Agreement](#) · [23. Institution blueprint: a Commons Collective Federation](#)
- **Long-run contributor incentives and possible equity stakes**
[8. Payment, disbursement, and public-good return](#) · [19. Data dividends, Pigouvian taxes, and commons credits](#)

[View source context and full mapping](#)

Revenue sources should remain distinguishable

1. **Member dues:** fund representation and purchasing services, with waivers or sponsored membership to prevent income-based exclusion.
2. **Transparent service charges:** compensate curation, rights verification, delivery, and audit. Do not disguise them as royalties for public-domain information.
3. **Licensing commissions:** a possible separately negotiated variant, not a deduction implemented in the current Lab's contributor contract; apply only to authorized uses with prior disclosure.

4. **Procurement administration:** a disclosed fee or dues-funded service, not an undisclosed supplier kickback.
5. **Commons contracts and grants:** fund nonexclusive public infrastructure, independent evaluation, and underserved participants.

These are institutional design options. A proposed income source is not booked by the financial summary unless the selected detailed model actually includes it. In particular, current contributor-agency receipts are not an additional operator revenue stream.

The initial operator should not depend on selling profiles, ads targeted from private archives, or speculative token appreciation. Venture capital may fit some interoperable infrastructure services, but a return requirement can conflict with non-extractive membership governance. Any investor participation must be subordinate to the mission and cannot buy contributor permission.

One financial source, detailed model and summary

Follow the Money is a read-only rollup of the selected Simulation Lab source. It must match the same period, currency, account perimeter, receipts, costs, and liabilities exactly. It is not a second calculator that assumes licensing income or applies a different distribution rule. Inputs belong in the Lab; the summary aggregates its output and identifies its source.

The current startup view models investor-funded ramp-up, staff, acquisition and support costs, member prices, and an explicit commons covenant. Its forecast is distinct from the operating-agent model's simulated payment journal. A forecast must not be called a cleared receipt, and neither is evidence of a real institution or transaction.

The current published reference is the same AI Commons startup projection used by the Simulation Lab and Follow the Money: month 0 through 36, source identity `c54691b8aa456b8c`. These are conditional planned cash flows, not executed payments or an empirical forecast.

Operator source / use	Cumulative USD	Accounting treatment
Available investor capital, financing not revenue	\$6,000,000.00	financing
Projected administration/service fees	\$9,781,305.00	earned-revenue
Total uses, including commons transfers	\$10,776,509.75	Derived from the Lab's monthly cost rows
Of those uses: commons transfer	\$1,956,261.00	20% of service fees, not investor capital
Closing projected operator position	\$5,004,795.25	Sources minus uses, not distributable profit
Reconciliation difference	\$0.00	Must match the detailed model exactly

Staffing, acquisition, the user ramp, and every cost line are editable in the [Simulation Lab](#). The [Follow the Money rollup](#) reads that model's output; it has no separate financial assumptions. [Published trial evidence](#) records the distinct, frozen design and holdout study.

These calculations concern paid services, not civic membership or an entitlement to sell other people's information. Free participation, support, accessibility, and governance costs still need funding. Fees must be evaluated against comparable service quality and actual switching costs, not merely against a headline discount.

Equal member payments are a governance choice, not an estimate of marginal contribution. A more granular valuation scheme might reward effort or quality but adds measurement, gaming, and dispute costs. No precise attribution from an individual record to every generated output is assumed. Externalities also mean that one person's sale can affect others who never agreed to sell. [\[S24\]](#)

Separate five meanings of "data value": buyer willingness to pay; contribution to a specified model-performance metric; the contributor's willingness to accept disclosure; production and maintenance cost; and net social benefit. Data Shapley addresses a particular performance-allocation problem, not an intrinsic monetary value. Acquisti, John, and Loewenstein's privacy experiment shows why willingness to accept or pay can depend on context and endowment, not simply reveal a stable price. Neither measurement alone establishes a fair dividend. [\[S27\]](#) [\[S30\]](#)

8. Payment, disbursement, and public-good return

Long-run incentives and possible contributor equity

The photographed Vincent card explicitly raises equity stakes in data or AI companies as one possible long-run incentive. Treat this separately from a fee for current work, a royalty, cooperative patronage, or the public sovereign-fund proposal discussed later. No instrument is universally appropriate. Compare cash flow, valuation, dilution, vesting, liquidity, risk, tax/securities treatment, and who controls decisions. Contributors must not be forced to accept speculative equity instead of promised pay, and investor voting power must not silently replace civic membership rights. This is a research and legal-review question, not an offer of securities or evidence that individual data creates an ownership share in a model. ^[S97]

Contract-specific flows implemented in the Lab

Flow	Implemented accounting	Governing constraint
Investor capital	Financing into the operator, not earned revenue	Does not create permission, a dividend entitlement, or guaranteed demand
Buyer payment for AI/provider service	Provider consideration, not operator revenue	Compare equivalent service and enforce the buyer's mandate
Administration/service fee	Operator revenue; the configured A3 covenant transfers its stated share to the independent commons fund	The published reference is 20% of these fees, not 20% of every incoming dollar
Authorized contribution or maintenance entitlement	Full contracted contributor amount enters the agency ledger; disbursement or outstanding liability is recorded separately	No default operator or commons deduction from money already owed to a contributor
Commons maintenance funding	Restricted funding and its permitted payments are shown in the commons ledger	A transfer is not proof of completed public benefit
Reserves	Cash retention and the model's stated liquidity target, where modeled	A reserve target is not an additional expense or an automatic percentage of licensing revenue

For example, an accepted and funded \$10 contributor entitlement is \$10 owed to that contributor under the current agency contract. It is not silently split into \$4/\$3/\$2/\$1. Administration fees and their commons covenant are separate flows. A different negotiated allocation would require a different named contract implemented consistently in both detailed and summary views.

A real payment path would require

Invoice -> cleared receipt -> contract reconciliation -> reserve and restriction entries -> eligibility check -> payout instruction -> payment confirmation -> member statement -> dispute window -> independent audit.

Use regulated financial-service providers as applicable. Minimize identity collection, but do not promise anonymous disbursement where payment, tax, anti-fraud, or sanctions rules require verification. Keep identity records separate from content and public receipts. Failed payments remain member liabilities, not revenue; unclaimed funds need a jurisdiction-specific policy. Publish aggregate audited totals, not names or bank details.

The executable models use integer cents and explicit contract rounding. Each account must satisfy opening cash plus incoming cash minus outgoing cash equals closing cash. Agency cash must remain matched by unpaid contributor entitlements; internal transfers are eliminated in a combined institutional view. Unpaid invoices are not cleared cash, and failed payments do not become operator income.

Follow the Money exposes those reconciliations for a chosen cumulative window. Operating-run figures are regenerated from the completed run's configuration and checked against its recorded checksum before a ledger is displayed. Startup figures are aggregated from the same monthly projection rows and checked against the Lab's detailed budget. Downstream fund balances or contributor payments that a projection does not model are labeled **not modeled**, never invented as zero or calculated from an unrelated percentage.

Purchasing savings and contributor earnings remain separate stakeholder outcomes. Neither is a universal data dividend. A person who contributes no licensable material may still deserve full voting rights and meaningful purchasing benefits.

Public-good allocation

Proposal: an independently overseen fund supports open reference maintenance, community archives, accessibility, translations, provenance tools, and legal or technical support for underrepresented contributors. Publish award criteria, reviewer conflicts, funding amounts, deliverables, and renewal decisions. Include minimum support for valuable but low-traffic resources rather than allocating entirely by clicks or model usage.

Participatory budgeting can establish priorities. Matching funds may encourage diverse support but can be gamed by fake identities or coordinated donations; any matching mechanism needs a capped pilot and fraud review. Avoid requiring recipients to grant exclusive rights to qualify. Measure maintenance and public access, not merely funds disbursed.

9. Creative Commons as an aggregation model

Creative Commons demonstrates the value of standardized permission infrastructure. It is not automatically an aggregator, collecting society, payment processor, or owner of everything carrying a CC license. Existing CC permissions cannot simply be replaced with a retroactive AI royalty obligation. CC licenses are irrevocable for recipients who comply with their terms, and permissions differ by license and by the rights involved. ^[S08] ^[S26]

Proposed CC-compatible role, not a claim of CC commitment:

- A discoverable registry that preserves the exact license, version, attribution information, provenance, and rights caveats.
- Shared metadata and contributor mandates for **new** services or permissions the contributor actually controls.
- A voluntary "support the commons" covenant for commercial users, independent of whether the underlying open content already permits reuse.
- A separate settlement and grant-disbursement institution with audited governance, rather than pretending that a license badge routes money.
- Optional public-interest access to curated collections and validation services, without restricting lawful access to the underlying open works.

Newer CC discussions about signaling preferences for AI use should be evaluated on their own terms; a preference or social signal is not automatically a modification of a pre-existing license or an enforceable obligation on every downstream user. ^[S09]

For a new proprietary work, Alice might offer a paid, narrowly specified license while separately licensing another version openly. Rights clearance must establish what remains available for negotiation. For an existing CC BY work, a buyer can fund curation, delivery, verification, or commons stewardship, but the institution cannot represent that the buyer needs a new permission merely because an AI system is involved.

Recommendation: begin with interoperable metadata and voluntary commons financing. Do not claim that Creative Commons, its staff, or its community has accepted the proposed commercial or governance role.

10. Gmail, private AI, and meaningful opt-in

The Gmail analogy is about an assistant operating over information entrusted for a defined service. Google's published Gmail privacy explanation describes processing for features such as spam detection and smart composition, user controls, and a distinction between email content and advertising. It does not establish a universal rule for all Google AI products, every account type, or foundation-model training. ^[S17]

Correction to a tempting shortcut: opting into an AI feature is not equivalent to giving unrestricted training rights over an entire inbox. Consumer, education, enterprise, and separately connected AI services may have different terms and controls. Any factual claim about a specific product's defaults requires a current product-and-account-specific check.

For a cooperative assistant, separate at least four choices:

1. **Task execution:** temporarily access only the records needed to answer the current request.
2. **Personalization:** retain explicitly chosen preferences or an index for that member.
3. **Service improvement:** reuse narrowly specified telemetry or feedback for disclosed purposes.
4. **Model training:** a separate decision specifying scope, recipient, retention, and any compensation.

No training permission should be inferred from the first three in the proposed design. An incoming message from Bob creates obligations toward Bob; Alice's custody of the message is not proof that she can license all its content or personal information. GDPR, where applicable, requires a legal basis and purpose analysis; consent is not the only legal basis, and calling something "opt-in" does not by itself make consent valid. ^[S10]

The consumer interface must make refusal as easy as acceptance, disclose whether a service still works when permission is declined, and explain what withdrawal can and cannot undo. A future-access block is technically different from deleting a source, purging a retrieval index, or proving removal of influence from an already trained model.

11. New knowledge objects and technical architecture

A knowledge object is a governed bundle

Proposal: the unit of exchange is not a naked file but a **knowledge object**: content or a pointer, plus assertions about origin, authority, permitted use, evaluation, benefit allocation, and review. The object is not itself proof that those assertions are true.

Object or record	Minimum fields	What it enables and what it does not prove
Knowledge object	Identifier/version, content location, origin type, derivation links, structure, quality limitations	Discovery and traceability; not ownership
Authority and mandate record	Grantor, evidence, authorized representative, scope, term, territory, withdrawal state	Scoped delegation; not rights over absent parties
Purpose-bound access grant	Recipient, allowed operation, expiry, retention, onward-transfer conditions	Runtime policy checks; not prevention of every copy
Use receipt	Object version, contract ID, purpose, timestamp, delivery/metering facts	Contract reconciliation; not causal attribution to every output
Benefit obligation	Currency, split, eligible pool, restrictions, settlement status	Auditable liabilities; not automatic receipt of funds
Evaluation and impact record	Task utility, provenance gaps, privacy tests, distributional and environmental findings	Informed review; not an all-purpose ethical score
Synthetic lineage record	Generator/model version, source mixture where known, evaluation set, contamination findings	Risk analysis; not a guarantee of clean rights or anonymity

Reuse established representations where they fit. W3C PROV represents entities, activities, agents, and derivations. Verifiable Credentials can express issuer claims in interoperable forms, with privacy implications that require careful design. Neither establishes that the issuer's legal claim is correct or that the credential must be public. [\[S18\]](#) [\[S19\]](#)

The accompanying JSON example is deliberately fictional, has no real content or cryptographic proof, and is not a complete schema or executed contract. It separates retrieval permission from prohibited training and does not promise historical model erasure.

Proposed system boundaries

Member-controlled source or steward

- > local selection and third-party/sensitivity review
- > catalog metadata + verified authority evidence
- > member-approved bargaining mandate
- > contract and purpose policy
- > authenticated access gateway / approved computation
- > buyer's bounded retrieval, evaluation, or training environment
- > signed or otherwise tamper-evident use receipts

```
-> segregated settlement ledger
-> member statements + independent commons awards + audit
```

Prefer data minimization and local or trusted compute-to-data patterns where practical. Enclaves, access gateways, and federated learning can reduce particular exposure risks; none eliminates inference, malicious-recipient, or governance risks. Training may genuinely require exporting approved material. If so, disclose that rather than claiming all data remain local.

Use encrypted transport and storage, scoped short-lived credentials, least-privilege roles, deletion schedules, and independent incident response. A public ledger should contain aggregate or carefully designed commitments, not raw content, personal identifiers, or easily guessed hashes of private messages. Provenance itself can leak identity.

For access requests, fail closed when authority, purpose, or mandate status is unresolved. For payment failures, report an actionable error and preserve the liability. Do not create success-shaped receipts for failed delivery or failed settlement.

An agent trace can mix public material, private communications, professional work, third-party content, and synthetic outputs in the same record. Classify and assess authority at the appropriate component and use level; possession of the log does not clear every span or tool result for donation or redistribution. The supplied notes' shareback idea is therefore a purpose- and permission-specific process, not a default upload of everything. [\[S10\]](#) [\[S99\]](#)

Divergence at every infrastructure node

A request can use an open protocol yet still concentrate identity, attention, payment, and negotiating power elsewhere in the stack. The relevant question at **every node** is: who sets the rules, who can refuse or switch, who receives the money, and who can audit the result?

[Open the full divergence chart and accessible comparison](#). The branches show **design alternatives and concentration risks**, not a factual finding of misconduct by any named provider. Cloudflare is an illustrative CDN provider; example.com is an illustrative application domain.



The middle lane traces the technical request. The side branches trace alternative distributions of **control, information, rents, and benefits**, not two physically exclusive networks. A centralized service can fulfill a commons-serving contract; a decentralized or open-source service can still extract value. Ownership labels alone do not decide the branch.

For Alice's request to example.com, this means choosing the browser or agent's delegation policy; examining the application's business model; negotiating edge/CDN logging and portability; limiting identity and database access; specifying compute and model use; verifying receipts; and directing payments through an accountable settlement institution. Bob's interests as a correspondent, recipient, or affected community member must remain visible even when he is not the paying customer.

At each contract handoff, record an **enforceable condition**, an **exit or refusal mechanism**, a **beneficiary**, and an **independent verification method**. If a claimed commons branch lacks these, label it an aspiration rather than a demonstrated return to the commons.

The companion page's [node-by-node tipping points](#) identifies the countervailing power, organizing actor, enforceable instrument, evidence, and remaining capture risk at each handoff. These are proposed causal mechanisms to test, not measured transition thresholds: open infrastructure alone does not force an incumbent to share control or surplus.

The notes attribute to Tomicah a distinction among market mechanisms, policy interventions, non-market action, and philanthropic cold-start capacity. The [ten-node mechanism matrix](#) applies all four lenses to every node. They can combine: a contract may need lawful public rules, volunteer stewardship, and seed funding. A grant does not create authority, prove buyer demand, or entitle a donor to control the institution. ^[S99]

Treat the notes' fear/greed framing as a limited motivation hypothesis, not an exhaustive account or a warrant to manufacture fear. Reciprocal service, professional identity, community purpose, and expected gains can also matter. Likewise, the notes' IETF concern is a proposed case study of participation resources, agenda-setting, conflicts, and outcomes, not an established finding of institutional capture.

12. Alice and Bob: end-to-end flows

A. Alice buys an assistant; Bob is her correspondent

1. **Join:** Alice joins the purchasing cooperative using minimal membership information. She chooses a subsidized or paid tier; training permission is not a condition of membership.
2. **Compare:** she sees a matched-price comparison, retention terms, accessibility, supplier exit options, and what the cooperative earns.
3. **Authorize a task:** Alice asks the assistant to find an appointment in selected messages. She sees the folders, time window, recipient service, and retention period before access.
4. **Minimize:** the connector excludes unrelated messages and flags sensitive material. Bob's information is limited to what the authorized service requires; it is not silently enrolled in a licensing pool.
5. **Execute:** the supplier receives a purpose-bound token. The assistant performs retrieval or inference under the agreed terms, not general model training.
6. **Receive evidence:** Alice receives the answer and a comprehensible activity receipt: what was accessed, for what purpose, by whom, and how to contest it.
7. **Exit:** she revokes future access, requests applicable deletion, exports settings, or switches supplier. The institution reports actual deletion scope and limitations rather than promising that a previously trained model has "forgotten" her.

Outcome to test: comprehension and control with no compulsory data sale. There is no assumed royalty for this private-assistant flow.

B. Alice licenses a field guide; Bob uses an educational tool

1. Alice is the fictional author of original, non-sensitive material. The association verifies her authority and any assignment, collaborator, or source restrictions.
2. She grants a one-year, nonexclusive mandate for a **retrieval service**, not training. The contract specifies permitted excerpts, attribution, retention, recipient, fees, and remedies.
3. The fictional educational provider evaluates Alice's specifically authorized retrieval-service offer. If it accepts and funds a \$10 entitlement in the chosen scenario, the payment has a named contract and recipient. No receipt is invented when the offer is rejected. Bob's assistant request identifies the relevant collection and permitted operation.
4. A receipt records delivery against the contract. It does **not** claim to calculate the contribution of every sentence to the model's internal weights.
5. The full contracted \$10 is recorded in the contributor-agency ledger and either paid to Alice or retained as cash matched by a payable. Operator fees, the commons covenant, and verification costs are separately identified. The money rollup shows those same entries rather than imposing a second allocation.
6. Bob can inspect provenance and report incorrect attribution. Alice can dispute a use or revoke future mandate renewals, subject to the signed contract and applicable rights.
7. If the buyer wants training access, it must negotiate a separate permission and review the changed risk. No technical flag alone converts the old contract into the new one.

Outcome to test: actual willingness to pay for provenance, quality, and permitted access, together with enforceable contributor limits.

C. Alice publishes openly; Bob supports the commons

Alice has already released an article under CC BY. Bob's developer can exercise the permissions that license grants while meeting applicable terms. The association does not revoke that permission or bill retroactively. Instead, Bob's organization can pay for a curated, well-maintained collection and voluntarily contribute to a commons fund. The fund might support translation, maintenance, and accessibility, including resources with little commercial demand.

Outcome to test: additional public-good funding without reduced lawful access. A successful test must compare funding gains against the cost and friction of the institution, not merely celebrate a new payment channel.

D. Alice and Bob coordinate book and music services

The supplied notes attribute this example to Nick. It is an **illustrative research proposal**, not a case report about existing collecting organizations or a claim that either has particular statutory powers. Here Alice represents a hypothetical book collective-management organization (CMO), and Bob represents a music CMO. ^[S99]

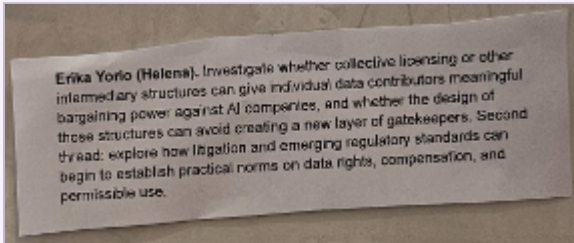
1. **Verify mandates:** identify represented rights, jurisdictions, exclusions, and overlapping representation.
2. **Choose one service:** define a bounded educational retrieval offering; do not assume training is included.
3. **Map interfaces:** exchange permission-safe catalog schemas and authority fields without pooling works or confidential member records.
4. **Compare appraisal methods:** discuss curation cost, coverage, reliability, and task utility, not automatically confidential prices, margins, bids, or customer terms.
5. **Review information exchange:** obtain jurisdiction-specific competition advice. Independent intermediaries, aggregation, or a clean room can reduce exposure but do not create immunity. ^[S13]
6. **Allocate responsibilities:** specify delivery, provenance, attribution, corrections, costs, and benefits while preserving separate rights and accounts.
7. **Choose a lawful negotiation route:** negotiate jointly only within verified authority and an appropriate legal basis; otherwise share infrastructure while retaining independent commercial decisions.
8. **Pilot and reconcile:** test service quality, member understanding, payment allocation, complaints, and exit; publish only permission-safe methods and results.

Success is reduced coordination cost and better service without unauthorized rights expansion or harmful information exchange. It does not require a universal valuation of books versus music.

13. Regulatory framework: present rules and proposals

PHOTOGRAPHED DIRECTION / SOURCE TRANSCRIPTION

Erika Yorio (Helena).



Attribution as printed: Erika Yorio (Helena).. Open the crop to inspect it.

Investigate whether collective licensing or other intermediary structures can give individual data contributors meaningful bargaining power against AI companies, and whether the design of these structures can avoid creating a new layer of gatekeepers. Second thread: explore how litigation and emerging regulatory standards can begin to establish practical norms on data rights, compensation, and permissible use.

This records a printed card, not a verified spoken quotation or endorsement. ^[S97]

▼ Editorial alignment and coverage

Collective licensing and intermediaries should improve real bargaining power without becoming new gatekeepers; litigation and emerging regulation can be studied as routes to practical norms.

Licensing, competition, rights, and gatekeeper risks were covered. A dated litigation-and-standards research track is added without asserting unverified case outcomes or universal new compensation rights.

- **Test bargaining gains against the risk of a replacement gatekeeper**
[6. Archetypes and the multi-sided marketplace](#) · [17. Failure modes, counterarguments, and safeguards](#) · [21. Reputation, validation, and quality ranking](#)
- **Track legal and regulatory norm formation with jurisdiction and status made explicit**
[13. Regulatory framework: present rules and proposals](#) · [20. Follow-on research: a Commons Collective Framework Agreement](#)

[View source context and full mapping](#)

This section is issue-spotting, not legal advice. "Collective bargaining" is an institutional objective, not an automatic legal status. Laws protecting employee bargaining do not necessarily protect coordinated pricing by independent creators or businesses. The parties, market, activity, jurisdiction, and scope of any exemption matter.

For example, the U.S. NLRA excludes independent contractors from its employee definition. A valuable contribution to AI does not itself establish employment, but exclusion from that statute also does not prove that all collective negotiation is prohibited by other law. ^[S33]

Area	Existing framework or constraint	Proposed intervention and unresolved question
Privacy and data protection	GDPR distinguishes lawful bases, controller/processor responsibilities, special-category rules, rights, and conditions of consent	Recognize accountable representatives and standardized mandates without making rights transferable commodities; establish responsibility for joint/overlapping processing ^[S10]
EU data governance	The Data Governance Act provides frameworks for specified intermediaries and recognized data-altruism organizations; it is not a universal ownership or royalty law	Determine which legal entity and service fall in scope; separate paid licensing from any altruism representation and neutral intermediation requirements ^[S11]
EU AI transparency and copyright	The AI Act creates obligations for general-purpose AI providers, including copyright-policy and training-content-summary provisions	Use relevant disclosures for accountability, without equating a summary with an item-level payment ledger or a new royalty right ^[S12]
Copyright and open licenses	Rights, exceptions, reservations, contracts, and license conditions differ; U.S. training/fair-use questions are context-sensitive	Improve mandate verification and audit; do not claim all training is necessarily licensed or necessarily exempt ^{[S08] [S14] [S26]}
Competition and antitrust	EU guidance for certain solo self-employed collective agreements is limited; it is not a blanket exemption for data cartels or every licensing arrangement	Explore a narrow, supervised safe harbor for qualifying bargaining, with portability, transparency, access, and anti-collusion safeguards ^[S13]
Payments and consumer protection	Real distribution can trigger payment-services, tax, anti-fraud, advertising, and unclaimed-funds duties	Use appropriate licensed providers, plain-language statements, restricted accounts, and no guaranteed-dividend marketing
Public goods and procurement	Public bodies can support research and stewardship within their lawful powers; a global commons charge is not presumed to exist	Pilot voluntary benefit covenants and lawful procurement conditions; evaluate any future levy through legislation, incidence analysis, and public accountability

Date-sensitive AI Act reading

Under the enacted AI Act's phased schedule, general-purpose-model obligations began applying on **2 August 2025**, with specific transitional treatment for models placed on the market before that date, including an **August 2027** compliance deadline under the relevant transition. The general application date and specific duties must not be conflated. Scope, exceptions, amendments, and current implementation guidance require a fresh check for the actual model and provider before procurement. ^[S12]

September 2026 update: the enacted **Digital Omnibus on AI, Regulation (EU) 2026/1744**, entered into force on **27 July 2026**. It changes relevant high-risk Annex III application dates to **2 December 2027** and Annex I dates to **2 August 2028**, without replacing the principal GPAI copyright/summary timetable discussed above. Describing all AI obligations as delayed would be incorrect. ^[S32]

The EU copyright directive separately distinguishes qualified research text-and-data mining from a broader lawful-access exception subject to appropriately expressed rights reservations. National implementation and the facts matter; neither a "research" label nor public accessibility settles the issue. ^[S35]

The report does not assert that a transparency duty creates a right for every person to be paid. Nor does it assume that publicly accessible information is free of privacy obligations or that a CC license covers every personality, trademark, or confidentiality issue.

A proportionate reform package

Near term: model mandates; standard use receipts; independent audits; interoperability; accessible complaints; consumer switching; public-interest procurement experiments; and funded legal support for smaller contributors.

Medium term, subject to legislative authority: clarify qualifying representative institutions and narrowly tailored competition-law treatment; require meaningful records and remedies for specified uses; evaluate statutory collective-management or commons-funding instruments with safeguards for open knowledge and small developers.

Do not start with: a universal individual property right over all observable facts, a presumption that every inference incurs a royalty, compulsory licensing of intimate communications, or a token that purports to sell rights the issuer cannot grant.

Litigation and standards as norm formation: the photographed Yorio card makes this a separate research track. Maintain a dated, jurisdiction-specific record of the claim, decision or settlement, appeal status, and narrow implications for permissions, compensation, and remedies. Do not turn one decision, an advisory report, or a proposed standard into a universal training rule. Evaluate whether the resulting intermediary improves contributor terms or simply becomes a new gatekeeper. No new case outcome is asserted by adding this research question. ^[S97]

14. Energy, data centers, and community bargaining

Digital abundance is not physical abundance. The IEA's 2025 analysis estimates global data-center electricity use at roughly **415 TWh in 2024** and projects roughly **945 TWh in 2030 in its Base Case**. These are data-center totals, not AI-only measured consumption; the 2030 number is a scenario projection, not an observation or a fresh 2026 forecast. ^[S21]

Proposal: a parallel infrastructure bargain. Consumer and public-sector purchasing coalitions can request verifiable energy and water reporting, workload efficiency, demand flexibility where technically feasible, and community-benefit commitments from suppliers. Host communities need their own representation: they should not depend on owning training data to have a voice in local burdens.

Negotiable commitment	Verification	Tradeoff or limit
Shift suitable batch workloads to lower-stress periods	Metered workload schedules, grid conditions, and service-level reports	Training/checkpointing and interactive latency constraints differ
Fund additional clean supply and enabling infrastructure	Contract terms, project completion, location, and time-matching evidence	Annual renewable certificates alone do not prove hourly carbon-free operation or local additionality
Avoid shifting infrastructure costs onto households	Regulator-approved tariffs, interconnection contributions, public cost-allocation analysis	Contract design and utility rules are jurisdiction-specific
Water stewardship and heat reuse where viable	Site-specific water data, basin conditions, engineering and demand studies	Benefits depend on climate, cooling design, nearby heat users, and seasonal demand
Local benefit agreements and independent complaints	Public commitments, delivery milestones, representative community review	A payment does not substitute for environmental compliance or consent where required

Report energy per useful service alongside absolute consumption. Efficiency improvements can lower cost and increase total demand; a per-query metric alone may conceal rebound. Do not invent precise per-output carbon accounting when the supplier exposes only aggregate data.

Philanthropic capital should support independent measurement, community technical assistance, and procurement expertise before subsidizing speculative data-center assets. Capital-intensive energy projects need qualified engineering, interconnection, environmental, and financial diligence; this report does not establish project feasibility or recommend a particular investment.

Data-center and data-commons coalitions may share procurement, technical assistance, and accountability methods, as suggested in the supplied notes. They must not redirect a host community's priorities without authority or treat a distant knowledge grant as an offset for local environmental duties. Local mitigation and a commons contribution require distinct evidence and approvals. ^[S99]

15. A philanthropic strategy relevant to Humanity AI

Evidence and identification. Humanity AI is the ten-philanthropy initiative announced on 14 October 2025 with a **\$500 million commitment over five years**, including aligned grantmaking and a pooled fund. This is an announced collective commitment, not proof that all money has been disbursed, is pooled into one account, or has been awarded to this project. It is not treated as a commercial investment fund or a presumed partner. ^[S15]

A current, separate opportunity: on 10 September 2026 Humanity AI announced a **\$10 million call-wide competitive grant program**. Published terms describe awards of **\$75,000-\$1 million** over up to three years, focused on U.S.-based community-led work, with the required nonprofit lead or appropriate fiscal sponsorship. The stated deadline is **21 October 2026 at 7:59 p.m. ET**. Research-only projects are out of scope: this paper would need to become a genuinely community-led action pilot, not merely an application to write more research. Verify current eligibility and timing before applying. ^{[S28] [S29]}

Proposal: a hypothetical **\$10 million, three-year catalytic portfolio**. This is a strategy for discussion with Humanity AI or similarly aligned philanthropies, not a funding request approved by any named organization.

The proposed portfolio's \$10 million is **not** the same as the actual call's \$10 million envelope, and it should not be presented as a plausible single grant under the published \$1 million maximum. A bounded eligible pilot could be one component of a separately financed portfolio.

Workstream	Allocation	Instrument and purpose
Field research and compensated deliberation	\$1.5m	Grants for community-led design, accessibility, translation, and baseline evidence
Cooperative pilots and member support	\$2.0m	Milestone grants for two or three bounded pilots, including fee-waiver support
Open protocol and reference implementation	\$2.0m	Grants for portable mandates, receipts, authority review, and settlement interoperability
Independent legal and policy infrastructure	\$1.0m	Grants for jurisdictional analysis, templates, redress design, and public-interest representation
Commons-finance matching experiments	\$1.5m	Capped, audited grants for nonexclusive public goods; no assumption of permanent subsidy
Energy and host-community bargaining	\$1.0m	Grants for independent metering, utility expertise, and community representation
Evaluation, replication, and contingency	\$1.0m	Independent evaluation and controlled contingency, with conflict-free release decisions
Total	\$10.0m	A proposed allocation, not a committed budget

Stage gates

- **Months 0-6: up to \$2m discovery.** Map rights and market structure; conduct compensated research; establish representation and ethics review; negotiate at least two independently credible buyer prospects. Do not collect sensitive records to make the pilot seem concrete.
- **Months 7-18: up to \$4m pilot.** Release funds against signed bounded contracts, understood mandates, an approved legal analysis, credible complaints handling, and a tested settlement ledger. Compare against individual or status-quo procurement.
- **Months 19-36: up to \$4m conditional replication.** Require independently verified benefit after costs, no unresolved material rights violations, actual switching/exit, and transparent accounting. Preserve a funded wind-down path.

The stage totals and workstream totals describe the same \$10m envelope, not two separate budgets. Allocate each grant to both a workstream and a stage in the actual ledger.

Grants first; investment only when justified

Non-excludable standards, independent oversight, and community participation are appropriate grant candidates. A recoverable grant or program-related investment might fit a revenue-producing, mission-locked service once counsel establishes eligibility and realistic repayment economics. Do not promise a PRI classification, assume a foundation can take any security, or model token appreciation as a repayment source.

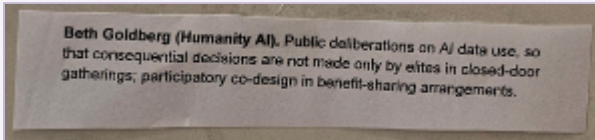
Stop or redesign if a pilot depends on coerced permission, cannot establish authority, produces benefits below costs without a justified public-good rationale, cannot attract genuine buyer demand, or excludes affected nonmembers from meaningful recourse. Protect evaluation independence and prohibit sponsors from suppressing unfavorable results.

Philanthropy should buy learning and institution-building capacity, not merely a temporary appearance of a market. Publish the subsidy per participant and a credible account of what survives when grants end.

16. Research program and falsifiable hypotheses

PHOTOGRAPHED DIRECTION / SOURCE TRANSCRIPTION

Beth Goldberg (Humanity AI).



Attribution as printed: Beth Goldberg (Humanity AI).. Open the crop to inspect it.

Public deliberations on AI data use, so that consequential decisions are not made only by elites in closed-door gatherings; participatory co-design in benefit-sharing arrangements.

This records a printed card, not a verified spoken quotation or endorsement. ^[S97]

▼ Editorial alignment and coverage

Public deliberation and participatory design should shape consequential AI-data decisions and benefit-sharing arrangements, rather than leaving decisions to closed elite forums.

Affected-person participation, paid design research, member governance, and commons budgeting were covered. Their connection to public co-design beyond invited participants is now explicit.

- **Include affected publics in consequential data-use decisions**

[1. Research circle and ethical commitments](#) · [5. From digital Costco to a bargaining federation](#) · [16. Research program and falsifiable hypotheses](#)

- **Co-design benefits and objectives with people affected**

[8. Payment, disbursement, and public-good return](#) · [24. Commons-authored RL rewards and regulatory design](#) · [25. A pro-human stack: working definition and test](#)

[View source context and full mapping](#)

The central empirical question is whether an institution changes bargaining outcomes and control, not whether participants express abstract enthusiasm for fairness.

The Goldberg card makes public deliberation and benefit-sharing co-design explicit. Compensated research should include affected people beyond the invited workshop or paying membership, preserve dissent, and record which decisions changed. A consultation count alone does not establish influence, representation, or approval of a proposed allocation. ^[S97]

The [open research questions and work register](#) records the current follow-on questions, next steps, and explicit statuses. It is the canonical work source for the [live overview presentation](#); presentation text is drawn from this report rather than maintained as a separate slide draft.

Hypothesis	Study and comparison	Primary outcome and failure signal
H1: pooled procurement benefits consumers	Compare matched supplier offers for individual and collective procurement; randomize invitation where ethical and feasible	Net savings at comparable quality and privacy; failure if lower price reflects worse service or hidden permission
H2: scoped mandates improve agency	Usability study comparing a purpose-specific mandate with typical bundled terms	Comprehension and effective refusal; failure if users misunderstand training scope or withdrawal
H3: collective supply improves contributor terms	Compare matched contributor cohorts or phased representative access	Net compensation and enforceable nonprice terms after transaction costs; failure if gains depend only on temporary subsidy
H4: commons covenants add funding without enclosure	Compare supported and matched open resources; inspect licensing and access changes	Additional maintenance/access improvements; failure if lawful reuse is chilled or grants replace existing support without net benefit
H5: transparent governance limits capture	Publish participation, candidate access, conflicts, concentration, and complaints	Members can change decisions and representatives; failure if sponsors or large contributors control outcomes
H6: infrastructure bargaining reduces externalized costs	Evaluate metered pilots and comparable procurement/site cases	Credible changes in cost allocation and environmental performance; failure if annual marketing claims substitute for measurable improvement

Proposed initial design

Recruit approximately **200 adult consumers and 40 professional or community contributors** for compensated design and usability research across at least two different socioeconomic contexts. These counts are feasibility targets, **not a statistical power calculation**. Use staged low-risk experiments before any commercial distribution. Determine a confirmatory sample size from a prespecified minimum meaningful effect and observed variance; do not infer causality from an underpowered convenience sample.

For procurement comparisons, match feature sets, support, privacy terms, and contract duration. Measure outcomes before and after participation and include refusal/dropout data. Where randomization is unavailable, preregister confounding risks and treat estimates as observational. Separate the effect of discounts from selection into membership and from donor subsidy.

Proposed progression thresholds: at least 80% of tested participants correctly identify whether training is permitted and how to withdraw; every payment is reconciled; all scoped-access denial cases pass; and no unresolved high-severity privacy or authority failure remains. These are design gates to be debated with affected people, not externally established standards. Report confidence intervals and subgroup differences rather than declaring success from a single average.

What to measure

- **Agency:** comprehension, effective refusal, time to switch, successful mandate withdrawal, electoral participation, and complaints resolved.

- **Distribution:** median and lower-quartile net benefits, fee-waiver coverage, contributor concentration, and outcomes for people receiving no royalty.
- **Business viability:** independently verified buyer demand, acquisition and support costs, retention, annual subsidy, and realistic path to sustainability.
- **Commons health:** maintenance hours, accessibility, translation, public availability, steward autonomy, and underfunded resources supported.
- **Integrity:** authority errors, third-party exposure, payment discrepancies, conflicts of interest, and procurement/seller-information separation.
- **Environment:** absolute and service-normalized energy/water metrics, grid context, ratepayer impacts, and community-reported burden.

Preregister methods and publish negative results. Seek institutional ethics review where applicable. Participation compensation must not depend on consenting to commercial reuse.

17. Failure modes, counterarguments, and safeguards

Failure mode or counterargument	Why it matters	Proposed safeguard and remaining uncertainty
Data are too substitutable to create leverage	A buyer can switch datasets or generate substitutes	Test actual buyer demand early; lead with procurement and quality services; do not promise universal data income
Contributors receive pennies while operators grow	Revenue may be small relative to coordination costs	Publish net unit economics and compensation concentration; cap opaque fees; stop unviable schemes
Privacy becomes something poor people sell	High need can undermine meaningful choice	Separate benefits from data permission, fund waivers, exclude high-risk categories initially
One person's disclosure harms others	Data often describe relationships and groups	Third-party review, community representation, purpose limitation, and remedies; no assumption that consent eliminates externalities ^[S24]
The intermediary becomes a new monopoly	Scale can entrench the representative	Nonexclusive mandates, portability, term limits, external audit, and alternative providers
An open commons becomes a tollbooth	Rights claims may chill lawful reuse	Exact license preservation, public-domain safeguards, voluntary service/covenant distinctions
Technical receipts are mistaken for proof	Signed claims can be false; training influence is difficult to attribute	Independent authority verification, sample audits, explicit claim limits, and enforceable contracts
Democratic procedures are captured or too slow	Participation consumes time and resources	Compensated deliberation, delegated routine operations, recall, and transparent constitutional boundaries
Buyers refuse public-good funding	Voluntary contributions may be unstable	Test covenant demand, procurement leverage, and public support; evaluate legislative proposals separately
Energy benefits are overstated	Efficient AI can still raise total demand	Absolute metrics, location/time context, independent verification, and public cost-allocation review

The strongest counterargument is that a new institution adds overhead without creating additional bargaining power. The research program must be able to conclude that a cooperative is inappropriate for a given setting. In that case, direct public funding, conventional consumer protection, stronger labor institutions, or investment in existing commons stewards may be better than inventing a data market.

18. A ninety-day starting sequence

Days 1-30: mandate and listening. Confirm the research circle and editorial ownership; select a pilot jurisdiction and domain; commission legal issue-spotting; compensate community sessions; inventory candidate rights and existing institutions; publish conflicts and a preregistered research protocol. Do not ingest an inbox or scrape a corpus merely because it is technically accessible.

Days 31-60: institutional and technical rehearsal. Test plain-language membership and scoped mandates with fictional objects. Obtain comparable procurement offers and bounded contributor-service offers. Rehearse failed authorization, third-party objection, revocation, failed payment, appeal, and supplier exit. Verify separate accounts and a proposed public-good allocation process.

Days 61-90: independent go/no-go review. Evaluate comprehension, buyer interest, legal authority, distribution, and costs. Publish evidence and failure cases. Proceed only to a bounded pilot with explicit limits; otherwise revise the institution or redirect funds to existing stewards.

Immediate deliverables provided here: the public research text; source register with claim limits; a fictional governed object; executable simulations with reconciled financial summaries; a public trial-evidence archive; and a publication workflow that rebuilds and deploys on every push to the repository's main branch. They are infrastructure for the research, not evidence that its proposed institutions already work.

19. Data dividends, Pigouvian taxes, and commons credits

The perspective: separate five objectives

A **Pigouvian tax** is a corrective tax on an activity that imposes costs on others, designed to make decision-makers account for those external costs. ^[S36] A useful proposal must distinguish a **data dividend** (a distribution to a defined population), **compensation** (payment for particular work or authorized use), **commons finance** (support for shared resources), **harm correction** (making decision-makers face external costs), and **rent sharing** (redistributing returns above a defined benchmark).

These objectives can coexist, but they do not justify identical instruments. A dividend may be financed from tax receipts or public investments without calculating every recipient's contribution to AI. A Pigouvian charge should correspond to marginal external harm, using credible measurement or a defensible proxy. A tax on economic rent instead targets exceptional returns. Commons maintenance has positive spillovers; supporting it does not automatically cancel an unrelated privacy, environmental, or labor harm. ^[S36]

Our assessment: a capped **commons-reinvestment credit** is worth testing as public-finance policy. It should not become an unmeasured "good behavior" score, a way to buy permission to violate rights, or a reason to relabel every profit tax as Pigouvian.

Fifty percent of what?

The user's **50% is an illustrative parameter, not an economically estimated optimum**. The tax base must be specified first. Consider a fictional company with \$100m annual revenue, \$90m allowable costs, and \$10m profit:

Base	Hypothetical 50% charge	Main issue
Revenue: \$100m	\$50m	Can exceed profit and tax loss-making firms; cascading charges and entry effects require analysis
Profit: \$10m	\$5m	Requires accounting, loss, allocation, and profit-shifting rules
Economic rent: \$10m profit less an assumed \$6m normal return	\$2m	More directly targets excess returns, but measuring the benchmark is difficult
Marginal external harm	Not determined by these figures	Requires a separate harm measure; profitability is not a reliable universal proxy

A company can be profitable with relatively low external harm, or loss-making with substantial harm. Consequently, a flat 50% of revenue or profit is not automatically Pigouvian. The scenarios are not recommendations that these rates or bases are efficient.

Which "Bernie's tax"?

The strongest matching verified proposal is **Bernie Sanders's American AI Sovereign Wealth Fund Act**, announced **18 June 2026**. It proposes a **one-time levy paid in newly issued equity**, designed to produce **50% public ownership after issuance** in covered AI companies. This is not a recurring 50% tax on revenue or profit. The announcement and linked bill describe public ownership and governance; they are evidence of a proposal, not its enactment. ^[S38]

Other possible meanings of the shorthand include Sanders and Khanna's **March 2026 proposal for a 5% annual billionaire wealth tax**, and Sanders's **October 2025 robot-tax proposal** directed at corporations replacing workers with machines, for which the cited announcement specifies no rate. These are different policies. None of the cited proposals establishes the Wikimedia-contribution credit considered here. ^{[S39] [S40]}

Interpretation: public equity could combine financial participation with voting power, making it closer to institutional collective agency than a cash dividend alone. But a public fund also creates risks of political capture, conflicts between financial returns and regulatory duties, valuation uncertainty, and dependence on distributable company cash. Equity value is not spendable tax revenue, and a stated distribution target is not guaranteed investment income. Constitutional, corporate, competition, and cross-border implementation would need separate legal analysis.

The announced fund valuation and dividend projections are sponsor estimates, not forecasts adopted by this report. The tax-and-credit illustration below is an independent proposal; it must not be presented as a provision of Sanders's bill.

A capped credit: trace every dollar

A **credit** reduces tax due; a **deduction** reduces the taxable base. Under a hypothetical 50% tax, a \$10m deduction would ordinarily reduce liability by \$5m, whereas a dollar-for-dollar \$10m credit reduces it by \$10m. Actual availability and interactions depend on legislation. ^[S37]

Proposed toy rules: normalize the specified base to a nonnegative amount, B; assess gross liability $L = 0.50 \times B$; verify eligible incremental commons spending E; allow $C = \min(E, 0.40 \times L)$; and collect net tax $T = L - C$. The 40% credit cap is our illustrative safeguard, not supplied evidence of an optimal rate. Assume no refund, no carryforward, and no additional deduction for the same expenditure.

Defined base B	Gross tax L	Eligible spending E	Credit C	Treasury payment T	Tax plus spending
\$100m	\$50m	\$10m	\$10m	\$40m	\$50m
\$100m	\$50m	\$20m	\$20m	\$30m	\$50m
\$100m	\$50m	\$30m	\$20m	\$30m	\$60m
\$10m	\$5m	\$10m	\$2m	\$3m	\$13m
\$0	\$0	\$10m	\$0	\$0	\$10m

The cap preserves **60% of pre-credit liability for the treasury**, not a fixed \$30m minimum for every firm. A zero or negative underlying base produces no liability or refundable credit under this simplified design; real loss treatment needs its own rules. The repository's integer-cent model tests the baseline, cap, zero base, and invalid inputs.

The first row is the central tradeoff: \$10m moves to eligible commons work instead of the treasury, while the company's combined outlay remains \$50m. The mechanism redirects finance; it does not automatically generate additional spending relative to collecting the full tax. Public accounts should recognize the forgone revenue as a tax expenditure. If lawmakers earmark part of the remaining tax for a data dividend, they must specify recipients, allocation, administration, and safeguards against overstating uncertain receipts.

Wikimedia contributions: additional public benefit, not reputation credits

Wikimedia is an example of a commons steward, **not a participating or tax-credit-approved recipient**. Its gift policy preserves review and refusal and generally favors unrestricted support. A donor must not buy editorial control, privileged access, or a say over who receives public knowledge. ^[S41]

Any pilot should require:

1. **Additionality:** exclude prior commitments and ordinary commercial spending relabeled as philanthropy; publish a baseline and anti-manipulation rules.
2. **Accepted, verified value:** recognize accepted cash support or independently assessed deliverables, not self-priced datasets, bulk edits, nominal token values, or unverifiable volunteer hours.
3. **No double counting:** coordinate deductions, credits, matching grants, and multiple firms claiming the same contribution.
4. **Anti-laundering controls:** identify related parties, subtract reciprocal commercial benefits, audit pass-throughs, and recover improperly claimed credits.
5. **Recipient independence:** preserve refusal and independent priorities; prohibit donor conditions that compromise editorial or community governance.
6. **Inclusive allocation:** maintain direct public funding and representation for communities unable to attract corporate donors or afford accreditation.
7. **Revenue protection:** cap offsets, report recipients and tax expenditures, review effectiveness, and preserve legislative control over public budgets.
8. **Protection of existing freedoms:** never make ordinary open-content reuse conditional on a payment merely to manufacture a creditable service.

For a genuine harm tax, credits should normally correspond to demonstrable reduction of **the same harm**, not an unrelated donation. A separate commons credit within a redistribution or rent-sharing instrument is conceptually cleaner. This avoids giving a large emitter, for example, permission to ignore local costs because it supports a popular global knowledge project.

What the shared privacy paper contributes

In **What Is Privacy Worth?**, Acquisti, John, and Loewenstein studied choices between a \$10 anonymous gift card and a \$12 identified card, varying endowment and presentation. The field experiment finds that privacy choices depend strongly on those conditions. It therefore cautions against reading a single willingness-to-accept response as an intrinsic price. ^[S30]

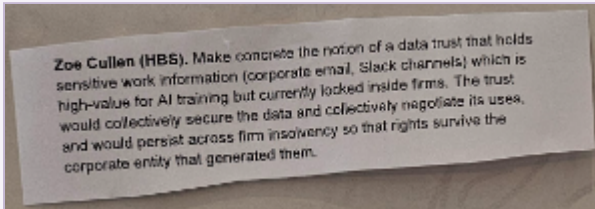
It does **not** establish a universal \$2 privacy price, a fair data dividend, an AI training value, or an optimal tax rate. For Commons Collective, the practical implication is to test comprehension and choice under multiple neutral presentations, protect refusal, and never treat payment as evidence that privacy concerns have been fully compensated.

Recommendation: test a narrow, independently audited commons-reinvestment credit alongside transparent public-good grants, while analyzing the equity-fund, rent-tax, and harm-tax alternatives separately. Judge success by incremental public benefit, distribution, enforceable agency, administrative cost, and forgone treasury revenue, not corporate participation counts or a "good behavior" badge.

20. Follow-on research: a Commons Collective Framework Agreement

PHOTOGRAPHED DIRECTION / SOURCE TRANSCRIPTION

Zoe Cullen (HBS).



Attribution as printed: Zoe Cullen (HBS).. Open the crop to inspect it.

Make concrete the notion of a data trust that holds sensitive work information (corporate email, Slack channels) which is high-value for AI training but currently locked inside firms. The trust would collectively secure the data and collectively negotiate its uses, and would persist across firm insolvency so that rights survive the corporate entity that generated them.

This records a printed card, not a verified spoken quotation or endorsement. ^[S97]

▼ Editorial alignment and coverage

A workplace-data trust for collective protection and negotiation, with a specific continuity question when the originating firm becomes insolvent.

Private and relational data, scoped mandates, and continuity were already discussed. The workplace-data and insolvency aspiration is now an explicit follow-on question; survival of rights is not assumed to be legally guaranteed.

- **Separate access to workplace information from authority to reuse it**
[4. A spectrum, not a single data commodity](#) · [10. Gmail, private AI, and meaningful opt-in](#)
- **Test collective negotiation, secure stewardship, and insolvency continuity**
[20. Follow-on research: a Commons Collective Framework Agreement](#) · [23. Institution blueprint: a Commons Collective Federation](#)

[View source context and full mapping](#)

Status: an illustrative, non-binding research framework, not legal advice or a production-ready agreement. The gaps below are **not resolved in this study**; this is not a claim that no sector, jurisdiction, or existing organization has addressed them.

A statutory labor collective bargaining agreement operates within applicable employment law. Other constituencies may need a **multi-party service, licensing, and governance agreement** instead. The label does not create employment status, rights ownership, a duty to bargain, or competition immunity. Local counsel must establish the legal route for actual parties and activities. ^[S33] ^[S13]

Which agreements still need investigation?

Constituency	Candidate agreement and negotiable terms	Authority boundary and research gap
Ordinary AI consumers	Collective service agreement covering price, accessibility, reliability, profiling limits, portability, and complaints	Establish valid representation and provider commitments; membership alone is not a lawful basis for processing ^[S10]
Nonemployee creators	Licensing/services agreement covering represented works, permitted uses, remuneration, attribution, and reports	Only administered rights can be licensed; preserve exceptions and existing permissions; assess coordinated terms under competition law ^{[S13] [S35]}
Open/public-domain stewards	Stewardship and infrastructure agreement covering curation, preservation, delivery, and public benefits	No invented ownership or retroactive CC restriction; distinguish service fees from claimed royalties ^[S26]
Private/relational data participants and affected nonmembers	Purpose-limited access and governance agreement with independent impact review	A contributor cannot waive others' privacy; an affected-person council cannot substitute for lawful processing ^{[S10] [S25]}
Synthetic/derived-data participants	Provenance and downstream-use agreement covering generation methods, validation, reuse, and disclosure controls	Derivation proves neither rights clearance nor anonymity; map actual rights and contractual obligations rather than invent exclusivity ^{[S10] [S34]}
Host communities	Community-impact agreement with an accountable operator covering monitoring, mitigation, complaints, and benefits	Validate representation and public-law requirements; proximity alone does not confer rights over training data

Existing sectoral agreements, licensing collectives, cooperative and trust instruments, and community-benefit arrangements should be comparators. The next research step is to identify their coverage and limits, not assume institutional absence.

Workplace-data trusts and institutional continuity

The Cullen card specifically proposes secure collective stewardship of sensitive work information, collective negotiation of use, and continuity through firm insolvency. This is not solved merely by creating a separate ledger or naming a trust. Assess employer and worker authority, correspondents' rights, confidentiality, trade secrets, custody, permitted purposes, creditor interests, successor responsibilities, and the treatment of contracts in the selected jurisdiction. ^{[S07] [S10] [S97]}

Distinguish an institution's continued existence from continued lawful access to information or enforceability of a mandate. The research should test what can legitimately survive a sale, closure, insolvency, or operator replacement; it does not promise bankruptcy-proof rights, irrevocable worker consent, or deletion from already trained models.

An annotated agreement architecture

Use a short master agreement with independently amendable schedules. Each promise must identify an accountable party, evidence of performance, a remedy, and funding for enforcement.

1. **Schedule A: membership and representation.** Define constituencies, elections, mandates, dissent, and affected-nonmember participation. Trustee duties need an appropriate legal structure, not simply a title. [\[S07\]](#) [\[S25\]](#)
2. **Schedule B: authority and scope.** Identify parties, rights, objects, exclusions, and whether authority derives from ownership, agency, service provision, or statute.
3. **Schedule C: permitted uses.** Specify retrieval, embeddings/indexing, pretraining, fine-tuning, evaluation, inference, and onward disclosure separately. A new use needs its own authority analysis.
4. **Schedule D: payment and public-good allocation.** Separate licensing, stewardship, and services; define eligible receipts, fees, reserves, contribution splits, payment dates, and audited liabilities. Rewarded licensing must not be mislabeled as DGA altruism. [\[S11\]](#)
5. **Schedule E: evidence and audit.** Define authorization references, object versions, recipients, purposes, retention, access controls, and independent inspection. Receipts are evidence to evaluate, not proof that all underlying rights were valid.
6. **Third-party and mandatory rights.** Require provenance checks, lawful processing, incident escalation, and confidential-data protections. Payment does not cure unlawful sourcing. [\[S10\]](#)
7. **Schedule F: withdrawal and remediation.** Distinguish future supply, contractual termination, deletion, and model remediation. State surviving permissions, legally required cessation, and technically verifiable duties.
8. **Schedule G: grievance and funded enforcement.** Set accessible complaint channels, response deadlines, independent review, proportionate remedies, and a protected enforcement budget; preserve statutory routes.
9. **Conflicts and competition compliance.** Disclose interests, separate incompatible roles, and limit coordination to counsel-approved scope. DGA status is not competition immunity. [\[S11\]](#) [\[S13\]](#)
10. **Portability and exit.** Specify usable formats, representative switching, continuing payment liabilities, and safeguards for other people's information.
11. **Term and renegotiation.** Trigger review for new purposes, material model changes, acquisitions, and relevant legal changes. Do not treat nonresponse as renewal.
12. **Law and jurisdiction.** Identify forums, governing law, cross-border responsibilities, and mandatory protections that contractual choices cannot remove.

Four non-binding clause sketches

Authority limitation. A representative may act only within the verified mandate and rights identified in Schedules A and B. Joining the framework does not establish ownership of contributed material, authority over nonmembers, employee status, or a competition exemption.

Open-knowledge safeguard. No payment or restriction is asserted solely because information is publicly available or in the public domain. Existing compliant CC permissions and independently applicable exceptions remain unaffected. Any purchased service is identified separately.

Affected-person protection. A collective decision does not authorize processing inconsistent with applicable law. The affected-party council may challenge a use, but its review does not consent on behalf of absent people or extinguish individual remedies.

Withdrawal without false promises. Withdrawal ends future supply according to Schedule F. Retained-record deletion, termination, and model-remediation duties are evaluated separately against applicable law, surviving licenses, and validated capabilities. This distinction does not permit continued processing when cessation is legally required. [\[S10\]](#) [\[S26\]](#)

These are design sketches, not clauses to sign without local adaptation. Every operative duty would need precise definitions, deadlines, evidence standards, allocation of liability, and an enforceable remedy.

A focused follow-on work package

Compare existing instruments in one selected jurisdiction and two bounded domains, then convene compensated consumer, supplier, steward, and affected-person sessions. Produce an authority map, a redlined framework, plain-language member explanations, and an independently evaluated negotiation simulation.

Go/no-go tests: verified mandates and rights chains; no majority waiver of protected rights; a defensible competition-law route; comprehensible retrieval-versus-training permissions; practical exit and payment reconciliation; and demonstrated capacity to fund complaints and enforce remedies. If an essential promise cannot be evidenced or enforced, do not proceed to general deployment. Move only to a bounded voluntary pilot that can be stopped.

21. Reputation, validation, and quality ranking

What to borrow from Ethereum agent examples

Proposal: add contextual reputation to provenance through separately signed, challengeable evidence about particular objects, versions, providers, and tasks. Do not turn "trust" into a single transferable score.

ERC-8004: Trustless Agents is a researched comparator, not an identified implementation of the Ethereum example mentioned in the supplied notes. The implementation intended by those notes remains unverified. At the research cutoff, the inspected official specification remains **Draft**. It proposes identity, reputation, and validation registries: ERC-721-based agent identity, contextual feedback, and requests to designated validators. Its normative interface language operates within a proposal; it is not proof of final standardization. Payments are outside its scope. ^[S42] ^[S99]

The curated contracts repository contains validation implementation code while warning that validation design remains under discussion and revision. Thus validation is neither simply absent nor established as a finalized, uniformly deployed mechanism. This report has not verified a chain deployment or contract address and does not present the example below as ERC-conformant. ^[S43]

Five layers, five different questions

Layer	Question	What it cannot establish alone
Identity	Which actor or organizational role controls this identifier and signs?	Honesty, legal authority, or one-human-one-identity
Provenance	What origin, transformation, contribution, and custody history is recorded?	Truth of that recorded history
Claim verification	What evidence supports this particular rights, source, or delivery assertion?	Universal ownership or freedom from third-party claims
Task validation	Does this version meet stated requirements for this use?	Fitness for every other task or permission for a different use
Reputation	What can be inferred from accumulated, scoped observations?	A permanent or context-free judgment of a person or organization

A hash demonstrates correspondence with committed bytes, not truth, authorship, ownership, or permission. A token identifying an agent does not transfer copyright in its outputs or training sources. Signed false statements remain false.

Maintain distinct identifiers for objects and immutable versions; authors, providers, and stewards; evaluators and methods; and observations with status histories. A proposed observation record contains:

```
subjectId, subjectKind, versionCommitment, claimOrMetric,
taskScope, outcome, unit, methodVersion, evidenceReference,
```

```
issuerId, observedAt, expiresAt, conflictDisclosure,
useProofReference, supersedes, statusReference, signature
```

An evidence registry holds the records; a reputation view interprets them under a published policy. Expose recency rules, raw and effective sample counts, uncertainty, reviewer selection, conflicts, and unresolved disputes. Do not average licensing evidence, translation accuracy, and delivery reliability as if they were the same quantity. "Insufficient evidence" is different from "poor performance." New versions and new owners do not automatically inherit old ratings.

ERC-8004's contextual tags, reviewer-filtered summaries, and off-chain aggregation are useful precedents, not an obligation to use a particular score. Ethereum Attestation Service supplies another comparator for schemas, issuer identity, linked statements, expiry, and revocation. Applications must check those conditions explicitly rather than treat the existence of an attestation as proof of validity for their use. [\[S42\]](#) [\[S44\]](#)

Alice's evidence and Bob's experience

1. Alice registers a version commitment and a separate rights claim. Controlled storage retains sensitive evidence.
2. A qualified, conflict-controlled reviewer checks the specific authority claim and records the evidence, jurisdictional assumptions, limitations, and expiry. It does not certify ownership for all purposes.
3. Bob chooses a permitted task and acceptance criteria. An evaluator tests that exact data and agent version, identifying the method and limits.
4. After an authorized interaction, Bob can submit feedback tied to a replay-resistant use receipt. Legitimate free use can qualify; purchasing is not the only credible interaction. The receipt proves an interaction, not independence or honesty.
5. Alice can respond, and an independent appeal can record a counter-observation. Revocations, corrections, and contested status affect future decisions without pretending historical events never occurred.
6. Payment follows the agreed contract, including any specified acceptance condition. A changing reputation score cannot silently rewrite an earned payment claim.

The [fictional observation sketch](#) makes these fields inspectable. It contains no real evaluation, signature, validated rights claim, or usable reputation rating.

Do not publish personal data, private purchase histories, or guessable hashes of sensitive content. Keep evidence access controlled; if public anchoring is justified, assess appropriate hiding commitments. A blockchain is optional: the first pilot could use signed records in an ordinary database with independently auditable exports.

Is Costco the verification and reputation marketplace?

Only as a limited analogy. Costco is a curated membership retailer/merchant, not a pure neutral two-sided marketplace. Its role in assembling an assortment and a buying relationship is useful to this design, but the research does not attribute supplier democracy, blockchain registries, or independent validation to actual Costco. [\[S16\]](#)

The proposed club could bundle discovery, rights clearance, testing, delivery histories, and dispute resolution. Commercially bundling these functions does not make them independent. A platform paid for a sale should not be the sole unchallengeable judge of that seller's quality. Separate evidence custody, evaluator selection, ranking policy, appeals, and independent audit; allow competing qualified evaluators.

Does Costco rank the best products to the top?

The official Costco Business Delivery search configuration inspected here includes **Best Match**, a descending-score setting, and other sorting vocabulary. It does not disclose the scoring inputs or prove that the highest-quality products rank first. We could not establish the exact proprietary algorithm. This finding is limited to the inspected Business Delivery configuration, not every Costco channel. ^[S45]

Distinguish **assortment curation**, **search/review sorting**, **commercial merchandising**, and **empirically validated task ranking**. Popular, highly rated, heavily viewed, best-selling, and suitable for a particular purpose are not equivalent.

How a data commons should rank data and serving agents

First apply hard eligibility gates. Check applicable authority, license compatibility, required consent or other authorization, privacy restrictions, and security constraints for the requested use and versions. Missing or disputed mandatory evidence means exclusion or review, not a small penalty that popularity, donations, or payment can overcome.

Then evaluate three distinct candidates:

Candidate	Task-specific evidence
Dataset or knowledge-object version	Target-task utility, relevant population/domain coverage, appropriate freshness, provenance-evidence completeness and verification status
Serving-agent version	Groundedness against authorized sources, reliability, latency distribution, and actual cost under a specified workload
Data-agent combination	End-to-end outcomes, including retrieval failures, unsupported answers, version compatibility, and interaction effects

Freshness is task-dependent: an old historical source is not defective merely because it is old. Excellent data cannot make an unreliable agent trustworthy, and a capable agent cannot make unsuitable data adequate.

Proposed ranking trace: task profile -> eligibility-policy version -> eligible candidates -> evaluation records -> task-local ordering -> explanation. Publish metrics, normalization, weights where used, tie-breaks, methods, and conflicts. Show evaluation dates, exact versions, sample counts, uncertainty, and whether evidence is independent or provider-supplied. A task-local suitability score is acceptable if its components and assumptions remain inspectable; a universal reputation score is not.

In a fictional query, Bob requests a grounded answer within a latency and cost budget. Alice's collection passes permission checks. A more popular collection missing the required authority is excluded. Bob compares remaining collections and compatible agents on the displayed metrics,

with an explanation of the ordering and the option to choose different legitimate tradeoffs.

Validate the ranker, not only the candidates. Test whether ordering predicts successful held-out task outcomes rather than clicks or purchases. Audit coverage and unequal exposure. Keep advertisements visibly separate; membership spending and Wikimedia donations are not quality evidence.

The supplied notes attribute to Ilan a quality/preference-signal proposal, with payment and public-good aims in tension. A proposed resolution is an open core of provenance, methods, versions, limitations, and challengeable evaluation summaries, with separately priced assurance, integration, freshness, or delivery guarantees. Do not sell favorable findings or preferential ranking. Shared funding should make baseline evaluation accessible so lack of budget is not mistaken for low quality. Payment for a maintained signal is separate from permission to crawl or reuse underlying material; existing rights and lawful access still govern. ^{[S26] [S42] [S61] [S99]}

Threat-model self-reviews, Sybil identities, fake transactions, reciprocal review rings, collusion, bribery, validator capture, and feedback loops that entrench incumbents. Use conflict disclosures, duplicate-use checks, reviewer-correlation analysis, independent samples, appeals, and bounded discovery opportunities for newcomers **only among eligible candidates**. Wallet counts are not independent-person counts, and tokenizing reputation does not resolve that problem. ^[S42]

22. Participation incentives and the future AI data club

Waze: useful service, contribution, and recognition

Waze illustrates **service reciprocity supplemented by gamification and community participation**, not, in the reviewed documentation, a market paying contributors per observation.

Separate three contribution modes. **Passive driving/location information** supports speed calculations and routing while the user receives navigation. **Active reports and confirmations** describe traffic, hazards, closures, and whether an alert remains current. **Map editing and community work** involve more deliberate corrections, localization, and beta testing. Their effort, risks, and incentives differ. ^[S46] ^[S48]

Current official documentation awards participation points, including five for a completed drive, six for an ordinary report, and three for a map edit. These are documented points, not promised cash. The current schedule must not be projected backward across the product's entire history, and the sources do not isolate how much each reward caused participation. ^[S47]

Interpretation: useful navigation, low-friction contribution, visible recognition, and community roles can reinforce participation. A governed data club should first identify the service contributors actually want, rather than assume micropayments are the only motivating force. It should not copy incentives that encourage unnecessary reports, invasive location collection, or unsafe interaction while driving.

Weather Underground: a related exchange with real equipment costs

Weather Underground's personal weather station network offers a comparable **contribute-and-benefit** pattern. The official overview describes local conditions, dashboards, historical archives, and shared observations. The participation path includes buying compatible equipment, installing it, and registering it. Sensor quality and connectivity are explicit considerations. ^[S49] ^[S51]

This is not evidence of a company purchasing every reading. Station owners can incur equipment, installation, connectivity, and maintenance costs while receiving useful information and participation benefits. The reviewed official descriptions do not promise per-reading payment or demonstrate contributor control of the marketplace.

Nor is volume equivalent to quality. Weather Underground's installation guidance identifies siting as fundamental: sunlight can distort temperature, while structures can obstruct rain or alter wind measurements. More badly situated stations need not outperform fewer well-maintained stations filling important coverage gaps. Calibration, timestamps, spatial distribution, maintenance history, and comparisons with reference measurements should accompany counts. That last proposition is a research inference, not a quantified assessment of the entire network. ^[S50]

Mechanism	Waze analogy	Weather Underground analogy	Proposed lesson for a governed club
Immediate reciprocal value	Navigation improves with relevant observations	Owners can inspect local and historical conditions	Supply a useful contributor-facing service before assuming willingness to sell
Contribution burden	Passive information differs from deliberate reports and editing	Hardware and maintenance can be substantial commitments	Match support and compensation to effort, scarcity, and responsibility
Recognition and belonging	Points and community roles	Participation in a shared observation network	Recognition can complement, not replace, fair support for sustained work
Quality bottleneck	Relevance, timeliness, confirmation, and accurate maps	Siting, sensor quality, connectivity, and coverage	Reward verified useful coverage, not raw volume
Governance	Community participation is documented; ownership is not established	Contributor network is documented; collective control is not established	Representation, portability, and enforceable rights must be designed explicitly

What is and is not verified about access benefits

The PWS overview supports visualization, local-information, and historical-archive benefits. It does not by itself establish unlimited free API access or a free premium subscription. The Weather Company's current commercial API page advertises a limited 30-day trial; that is a different offer, not proof of a PWS-contributor entitlement. ^[S49] ^[S52]

The contributor API-key page required login, and a reliable dated official record of historical API/subscription terms was not established in this research. Accordingly, we do **not** assert that uploading observations universally earned unlimited API access, either historically or today. This is a verification gap, not evidence that such benefits never existed.

Likewise, neither network's community language proves an open-data license, contributor ownership, or binding voting rights. Those would be additional commitments in Commons Collective.

Projecting the value proposition: dependable observation as a service

For some future AI applications, the scarce input may be **fresh observations of a changing world**, not a larger static corpus. A generated example does not independently discover today's road closure, temperature, equipment failure, or local institutional change. Synthetic data can still be useful for training and testing; the need for new observation depends on the task.

Proposed product: maintained coverage with known measurement limits, corrections and version histories, documented permissions, accountable contributor participation, reliable updates, and integration support. Buyers might value fewer downstream errors and lower verification or engineering costs. Those are hypotheses requiring paid trials, not demonstrated premiums.

The analogs also warn against romanticizing unpaid contribution. Useful reciprocal services may attract participation, but scarce expertise, equipment, sustained maintenance, and difficult coverage can require cash or direct support. Community enthusiasm is not an inexhaustible operating subsidy.

Does a future AI company need the Costco-style model?

Not universally, and not yet established empirically. A company may obtain adequate information from existing open resources, a few direct licenses, in-house collection, conventional suppliers, or task-appropriate simulation. It may need high-quality maintained information without needing this organizational form.

Falsifiable hypothesis: for selected applications requiring current, distributed observations, a governed contributor network delivers better downstream performance at lower **quality-adjusted total cost** than the buyer's best available alternative, while preserving meaningful member agency.

The proposed advantage should be greatest where information becomes stale quickly, missing coverage changes outcomes, contributors hold hard-to-replace access or expertise, and reliable correction matters. The counterhypothesis is that equivalent data are readily available or that coordination, verification, and governance costs consume the gains. Buyers might value reliable delivery while refusing to pay a distinct governance premium.

Three future scenarios, not a single forecast

1. **Scarce, changing ground truth.** The club fills consequential gaps and maintains observations competitors cannot cheaply reproduce. Buyers renew subscriptions for the service. Contributors receive useful services, compensation, or both, and governance helps sustain participation.
2. **Commodity data, competitive service.** Equivalent information is abundant. Revenue depends on cleaning, integration, responsiveness, or specific assurance, not ownership of an ever-larger corpus. The institution survives only if those services justify its overhead.
3. **Public value without sufficient commercial demand.** Better coverage benefits communities but paying buyers do not fund it, or one buyer captures bargaining power. Public procurement, grants, or member support may be warranted. That is a public-infrastructure model, not a demonstrated self-financing marketplace.

A pilot that can reject the hypothesis

Choose one bounded application with measurable freshness and coverage requirements. Compare the club with realistic open-data, direct-licensing, and in-house alternatives on the same held-out downstream tasks. Count acquisition, curation, verification, integration, support, governance, rights review, and payment administration in total cost.

Measure paid willingness to subscribe, renewal after the initial trial, quality and failure rates, missing-coverage outcomes, update latency, contributor retention, full contribution costs, net compensation, switching, complaints, and effective voice. Separate grants from earned revenue and compare subsidized and unsubsidized economics.

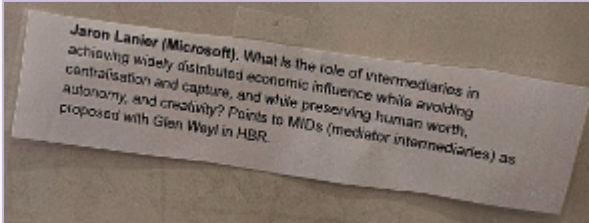
Reject or redesign if apparent value disappears after full costs, buyers do not renew, quality gains fail out-of-sample, contributors bear uncompensated burdens, rights cannot be verified, or meaningful refusal and exit are not possible. If public benefits remain compelling but commercial demand does not, evaluate an explicit public-interest funding model rather than forcing the evidence into a marketplace story.

The strategic question is not "How much data can the club collect?" It is **which dependable service the community can provide better than its alternatives, at a sustainable cost and on acceptable terms.**

23. Institution blueprint: a Commons Collective Federation

PHOTOGRAPHED DIRECTION / SOURCE TRANSCRIPTION

Jaron Lanier (Microsoft).



Attribution as printed: Jaron Lanier (Microsoft).. Open the crop to inspect it.

What is the role of intermediaries in achieving widely distributed economic influence while avoiding centralisation and capture, and while preserving human worth, Autonomy, and creativity? Points to MIDs (mediator intermediaries) as proposed with Glen Weyl in HBR.

This records a printed card, not a verified spoken quotation or endorsement. ^[597]

▼ Editorial alignment and coverage

Assess intermediaries by distributed economic influence and protection of human worth, autonomy, and creativity, while resisting centralization and capture; retain the card's reference to MIDs and related work.

Anti-capture governance, intermediaries, shared benefits, and the pro-human stack were covered. The photographed MID reference is preserved and connected to these tests without treating an acronym as proof of effectiveness.

- **Intermediation that broadens benefits without concentrating control**

[5. From digital Costco to a bargaining federation](#) · [23. Institution blueprint: a Commons Collective Federation](#) · [29. Replicable sector collectives: blueprint and templates](#)

- **Human worth, autonomy, creativity, and contestable mediation**

[25. A pro-human stack: working definition and test](#) · [26. Open-agent incentives and a Commons Collective Agency Starter Kit](#)

[View source context and full mapping](#)

Recommendation: build a Commons Collective Federation (CCF). It is a member-governed institution for purchasing accountable AI services, representing specifically authorized contributors, and supporting independent public-interest stewardship. Its first product should be a competitively negotiated, enforceable AI-service agreement, not a universal data pool or a token.

This is a **proposed institutional blueprint**, not an incorporated organization, validated business, statutory labor union, investment fund, or endorsement by any named researcher. "Federation" describes the governance arrangement; local counsel must establish the legal forms and actual authority.

Functions first, legal entities when justified

Preserve the three-part purchasing, contributor-representation, and commons-funding design, but do not incorporate three entities automatically on day one.

Required function	Concrete responsibility	Boundary that makes it credible
Purchasing cooperative function	Aggregate authorized commitments, solicit bids, and negotiate service, price, portability, and accountability	No commitments beyond members' approved mandates or spending limits
Contributor bargaining association	Maintain the rights-and-mandates register and negotiate specified licensing or contribution schedules	Contributors retain ungranted rights; membership is not blanket consent or data transfer
Independent commons fund	Finance preservation, open knowledge, public-interest evaluation, and participation support	Restricted accounts, independent allocation, and public reporting; not an opaque operator subsidy
Professional operator	Run procurement, administer contracts and records, and support complaints	No authority to expand purposes or waive protected rights
Independent assurance	Inspect authority records, contractual compliance, privacy, and reported benefits	Findings cannot be controlled by the commercial team whose work is evaluated
Regulated payment partner	Process charges and distributions through appropriately authorized services	The federation does not assume that holding or moving other people's money is unregulated

Use separate legal entities where law, fiduciary duties, or credible independence require them. Applicable DGA intermediation and neutrality requirements must be assessed before combining services or reusing intermediated data. Functional separation in a diagram is not a substitute for a required legal separation. ^[S11]

The overall photographed brief asks whether a public body, existing union, collective-management organization, consumer organization, or new institution should perform the role. Compare legal mandate, constituency, scale, accountability, administrative cost, and usable exit before choosing a federation. Public awareness and member organizing, policy or procurement support, entrepreneurial service provision, and adaptation of existing institutions are alternative formation routes to test, not a presumption that a new organization must be created. ^{[S03] [S97]}

The Lanier card's reference to "MIDs (mediator intermediaries)" is preserved as printed, not standardized into an independently verified acronym definition. [A Blueprint for a Better Digital Society](#) is linked as related HBR reading; the accessible publisher preview was not treated as a full review of the article. Intermediaries here must be judged by distributed economic influence, human agency, evidence, and remedies, not their name alone. ^[S98]

Membership, representation, and mandates

The initial buyers could include individuals, libraries, nonprofits, and small organizations. Contributors could include creators, professionals, and organizations with demonstrable authority over specified inputs. Open-knowledge stewards participate as stewards, not presumed owners of public-domain material.

Provide free civic participation, consultation, and complaint access alongside optional paid services. Formal legal membership is a separate matter: if the chosen cooperative form requires a capital contribution, provide lawful inclusion subsidies or another suitable arrangement rather than pretending the requirement disappears. Payment must not buy extra constitutional influence.

Members need information, voting, recall, and exit rights. Affected nonmembers and host communities need representation and complaints access because contributor votes alone do not represent everyone exposed to data use. ^[S25]

Proposed founding governance to test: buyer and contributor/supplier chambers, with a seven-seat steering board composed of two delegates from each chamber and three public-interest representatives, including affected-community representation. Keep the independent commons/affected-party review and reserved approvals described earlier; the steering board is not a substitute for them. Vendor counterparties do not gain governance control through a sales contract.

Test one-year representative terms and six-month, transaction-specific negotiating mandates. These durations and seat allocations are initial design choices for community ratification, not evidence of an optimal constitution. Mandates should identify counterparties, scope, price limits, prohibited concessions, ratification, and expiry. Individual authorization and mandatory protections cannot be overridden by a board vote.

Actual authority comes from lawful membership arrangements, agency, purchasing commitments, and contracts. A trustee label, member count, or "collective bargaining" name does not manufacture authority or employee status. ^{[S03] [S07] [S33]}

What is negotiated, and what creates leverage?

Start with AI-service providers or resellers as counterparties. Negotiate predictable pricing, quality, accessibility, permitted processing, privacy settings, portability, incident response, audit evidence, and remedies. Clear contributor schedules separately, specifying retrieval, embeddings, training, fine-tuning, evaluation, and onward disclosure.

Leverage requires credible commitments and alternatives: a buyer coalition that can purchase or switch; authorized inputs or continuing services that have real value; and reliable provenance, administration, and correction. Reduced transaction costs can benefit the counterparty too. Member count alone is not a credible purchasing commitment, and the public domain must not be enclosed to create artificial scarcity. ^{[S16] [S23] [S26]}

The institution needs professionally competent procurement and contract management, local counsel, funded enforcement, and independent assurance. Fractional appointments may reduce initial overhead. Optional subscriptions, disclosed administration fees, and stewardship services can provide earned income; grants can finance research, inclusion, safeguards, and public goods. The earlier financial model remains hypothetical.

Maintain distinct ledgers for operations, contributor liabilities, restricted grants, reserves, and commons allocations. Publish related-party transactions and allocation rules. Do not silently use contributor money to cover operating losses or describe rewarded licensing as DGA altruism. ^[S11]

Constitutional boundaries

- Preserve public-domain freedoms and existing compliant CC permissions. ^[S26]

- Do not infer permission to process personal information, including third-party information, merely from membership. ^[S10]
- Do not presume employee bargaining protections or competition immunity for nonemployee licensing. ^{[S13] [S33]}
- Distinguish future withdrawal, retained-record deletion, surviving permissions, and technically verifiable model remediation.
- Do not launch binding commitments without a funded, accessible route to complain, investigate, enforce, and obtain a remedy.

A minimum viable institution

Days 1-30: recruit a founding cohort; establish interim representation; commission legal scoping; adopt draft conflicts, mandate, and constitutional rules. A suitable sponsor may host the research pilot under written arrangements before incorporation, subject to counsel. This does not establish tax-exempt eligibility or a sponsor's commitment.

Days 31-60: fund the bounded operator, counsel, assurance, and enforcement work. Seek commitments from two or three institutional buyers, assess a small contributor inventory, and request at least two credible vendor proposals. These are pilot targets, not proof of sufficient bargaining scale.

Days 61-90: rehearse switching, authority review, use receipts, complaints, and payment failure. Ratify one capped purchasing mandate. Approve any contributor schedule separately; refuse unclear authority rather than treating participation as clearance.

Days 90-180: run one purchasing contract and, only if cleared, one small contributor schedule. Publish the negotiated terms where lawful, independently reviewed mandate register, full-cost benefit analysis, portability test, grievance exercise, and lessons.

Stop or redesign if authority remains unresolved, competition compliance lacks a defensible route, third-party rights cannot be protected, vendors reject essential accountability, exit fails, or the benefits do not justify full costs. If public benefits justify explicit subsidy, assess that transparently rather than presenting it as earned commercial viability.

Build the bargaining capacity before the marketplace. Begin with a sponsored, professionally operated federation pilot; incorporate only the structure demonstrated functions and applicable law require.

24. Commons-authored RL rewards and regulatory design

Assessment: commons participation in defining AI objectives is a plausible route to collective agency. A mandatory fixed percentage of a reinforcement-learning reward, however, is a weak standalone policy instrument. A stronger proposal combines public contributions to training signals, documented treatment of those contributions, independent evaluation, and enforceable outcomes.

First distinguish meanings of **reward**. An RL reward is a numerical optimization signal, not contributor pay or a pool of money. Paying reviewers, compensating contributors, licensing materials, or financing commons stewardship requires separate arrangements. An optimization coefficient cannot substitute for fair payment.

What the technical evidence establishes

Christiano and colleagues learned reward functions from comparisons of agent behavior. InstructGPT used human-ranked responses to fit a reward model and optimize a language model. These establish methods for incorporating feedback, not democratic authorization of the preferences being optimized. ^[S53] ^[S54]

Constitutional AI adds a route through written principles: AI-generated critiques, revisions, and preference judgments support supervised training and reinforcement learning from AI feedback, or RLAI. Public participation could therefore shape rubrics and principles as well as individual labels. But an AI judge still interprets those principles, and that interpretation remains contestable. ^[S55]

Collective Constitutional AI offers a relevant bounded experiment: input from 1,002 U.S. adults, with demographic sampling and an AI-familiarity screen, was translated into training principles. The authors report reduced measured bias on their evaluated dimensions alongside broadly comparable results on other reported evaluations. This is not global representation, a complete measure of adherence to public preferences, or proof of democratic legitimacy. ^[S56]

A proposed Commons Reward Trust

Use the name for a **governance function**, not an assertion that a legal trust has been established. Its outputs could include domain-specific deliberated rubrics, consented preference comparisons that retain disagreement, versioned reward models and calibration records, adversarial cases, and separately controlled held-out evaluations.

Each artifact should state the intended population, scope, provenance, exclusions, contributors' terms, revision process, and appeal route. The body producing training signals must not be the sole certifier of their effects. Publish methods and governance while protecting personal data and sensitive test items. Participation should be compensated for effort, not for supplying the operator's preferred answers.

This could be a protected public-interest function associated with the federation, but its independence from commercial negotiation and model evaluation requires explicit design. A public label alone does not prevent state, corporate, or organized-interest capture.

Why a reward share is not a share of agency

Consider the illustrative objective:

$$R = (1 - \alpha) * R_{\text{task}} + \alpha * R_{\text{commons}}$$

Setting `alpha = 0.2` does not mean that 20% of behavior reflects commons preferences. Multiplying `R_commons` by 100 changes its numerical contribution by 100 while leaving the coefficient unchanged. Reward scales, distributions, correlations, sampling, normalization, optimizer behavior, and regularization all affect the result. InstructGPT's objective itself used normalization and a KL penalty, with an additional pretraining term in one variant. ^[S54]

A study should report component distributions and calibration, training settings, ablations, and independently measured effects. It should test several weights rather than advertise a coefficient as a democratic entitlement. More commons weight is not automatically better.

Mandatory rights and safety requirements cannot be traded away in a compensatory average. High helpfulness cannot authorize unlawful disclosure. Use applicable requirements as admissibility conditions, deployment controls, and independently tested constraints. Neither a large penalty nor a "hard constraint" label guarantees actual compliance. Required legal bases and third-party protections remain necessary. ^[S10]

What regulation could require

These are **future policy options**, not a description of an existing commons-reward quota. The research is not a global legal survey; authority and scope require jurisdiction- and domain-specific analysis.

Step	Proposed instrument	Evidence and boundary
Procurement pilot	Contract for documented consideration of public rubrics, fairly compensated participation, independent testing, and remedies	Establish feasibility and actual benefit before imposing a general rule
Participation and evidence duties	Require disclosure of whose preferences shaped objectives, exclusions, conflicts, and reasons for accepting or rejecting principles	Disclosure must permit scrutiny without exposing participants or pretending consultation equals consent
Outcome accountability	Specify contextual performance, harm controls, monitoring, incident reporting, and redress	Permit different technical routes to demonstrated compliance; test metrics themselves for manipulation
Narrower process obligations, if justified	Consider stronger requirements after evidence of benefit, cost, enforceability, and safeguards	Any fixed reward mix would need calibration rules, auditability, equivalent-method provisions, review, and a sunset or revision mechanism

An obligation should not vanish merely because a provider uses a different architecture, supervised adaptation, or distillation rather than explicit RL. Conversely, requiring one technical recipe risks locking regulation to a method and encouraging formal compliance without improved outcomes.

Participation, pluralism, and reward hacking

Define the affected population, include people disproportionately exposed to harm, protect minority rights, disclose conflicts, publish dissent, and permit plural domain-specific rubrics. A majority vote cannot waive another person's rights. Minimize and protect participation data; representativeness does not justify indiscriminate collection of sensitive attributes.

Separate the production of feedback from independent evaluation. Gao and colleagues show that stronger optimization of a proxy reward can eventually worsen performance measured by a separate gold reward model in their studied setup. That gold model is itself synthetic, not a measure of human welfare. The result supports caution about proxy optimization, not a claim that a particular public rubric will necessarily fail. ^[557]

Shared benchmarks can leak or be overfit. Restrict access to held-out items where justified, refresh evaluations, examine subgroup outcomes, and make dispute resolution independent of the people paid to deliver high scores.

A bounded first experiment

Start with a low-risk public-library assistant over licensed, non-sensitive material. Develop rubrics for faithful answers, attribution, uncertainty, and accessibility. Compare a baseline, a rubric-only condition, and several commons-reward variants under comparable budgets and multiple runs. Illustrative coefficients might be 0, 0.1, 0.2, and 0.4; they are experiment settings, not proposed universal legal minima.

Preregister held-out tests, subgroup analysis, uncertainty reporting, full costs, and adverse-effect thresholds. Pause for material privacy failures, serious rights violations, unexplained subgroup harm, test leakage, or evaluator manipulation. Do not expand because the training reward alone rises.

The institutional objective is accountable public influence, not a particular coefficient.

Mandating a process or share is defensible only if independently assessed outcomes, rights protections, legitimacy, and costs support it.

25. A pro-human stack: working definition and test

A pro-human stack is a sociotechnical system that expands people's practical capabilities while preserving their ability to understand, authorize, refuse, contest, and leave consequential arrangements. It combines software and infrastructure with representative institutions, enforceable obligations, independent scrutiny, and funded human remedies.

The defining property is **accountable human agency**, not an open-source label, decentralization, or branding alone. Openness can be valuable, but it is neither necessary for every individual component nor sufficient for the whole system.

This is a **proposed working definition for this research**, not official CoLab approval, a statement that the stack has been implemented, or an accreditation scheme. It draws on the repository-stated orientation and the cited ethical-AI paper's treatment of affected people, institutional accountability, human rights, and avoidance of harm while remaining a distinct proposal. ^[S01] ^[S02]

The accountability tuple

For every component, document:

```
Accountable actor -> decision right -> enforceable condition
-> observable evidence -> appeal/remedy -> funded exit/maintenance
```

Name who can change the component's behavior, who bears responsibility, what constrains them, how compliance is checked, how people obtain redress, and who pays when it fails or becomes obsolete. An unnamed "community," a reputation token, or an autonomous agent is not an adequate accountable actor.

Stack layers and evidence requirements

Layer	Required property under this proposed definition	Evidence to inspect
Institution and constitution	Scoped authority, plural representation, protected rights, and independent challenge	Charter, mandates, elections, conflicts, and remedy budgets ^[S03] ^[S25]
Identity and permission	Minimal disclosure, specific delegation, applicable consent, expiry, and revocation	Permission records and tested controls; no universal personal trust score ^[S10] ^[S19]
Harness and agent	Understandable task boundaries, meaningful interruption, and controls on consequential actions	Task traces, approval gates, refusal tests, accessible controls, and recovery exercises
Protocols and tools	Documented interfaces, bounded tool permissions, and replaceable integrations	Permission manifests, interoperability and migration tests, and failure handling

Layer	Required property under this proposed definition	Evidence to inspect
Knowledge and data	Purpose-aware access, provenance, and applicable permissions and exceptions	Source records, license notices, access decisions, and retention schedules ^[S10] ^[S18] ^[S26]
Models and compute	Disclosed limitations, constrained use, and contestable selection	Version records, task evaluations, contractual controls, substitution tests, and incidents
Evaluation and reward governance	Plural input, protected-rights constraints, and independently reviewed objective changes	Protocols, dissent, reward specifications, calibration, and independent findings
Provenance, assurance, and reputation	Traceable claims and contextual, challengeable assessments	Linked evidence, assessor independence, uncertainty, and correction histories ^[S18] ^[S19]
Settlement and commons return	Transparent charges, authorized distributions, and sustained stewardship	Reconciled accounts, allocation rules, receipts, inclusion, and maintenance budgets
Physical infrastructure	Accountable resource use and meaningful host-community recourse	Metered or explicitly estimated energy and water data, methods, impact records, and remedies ^[S21]

Provenance and credentials can support scrutiny. They do not independently establish truth, lawful authority, task suitability, or fairness.

Seven boundaries that cannot be replaced by branding

- 1. Protected rights and meaningful refusal.** People must not be forced to accept unrelated data uses to participate. Explain refusal and its consequences in usable terms. Majority decisions cannot waive third-party privacy or mandatory protections. Rights retain their legal scope and exceptions; this is not a universal data-ownership claim. ^[S10]
- 2. Plural and bounded collective authority.** Include relevant users, contributors, and affected people, not only investors or large suppliers. Make mandates limited, reviewable, and challengeable. Contributor representation does not exhaust the interests of people affected by a model. ^[S03] ^[S25]
- 3. Usable exit and maintained alternatives.** Test export, switching, termination, assistance, and continuity while protecting others' information. Fund migration and maintenance. A nominal right to leave an unusable system is not enough.
- 4. Permission-aware knowledge without enclosure.** Assess public, private, professional, and derived data contextually. Preserve public-domain freedoms, compliant open-license permissions, and lawful exceptions. Tokens, credentials, and platform registration do not manufacture underlying rights. ^[S10] ^[S26]
- 5. Independent assurance and human remedy.** Evaluate specified claims and outcomes in stated contexts. Do not turn service reputation into an automated social-credit score or universal rating of a person. Provide accessible correction, human review, and effective remedies for harms obscured by aggregate success.
- 6. Independent commons-reward governance.** Public contributions to training objectives are a research proposal, not an existing mandate or a guaranteed optimization breakthrough. Keep objective-setting independent of commercial management, protect

affected perspectives and rights floors, and commission independent review. Numerical RL reward and monetary compensation are different; neither measures a person's worth.

7. **Fair accounting, inclusion, and community responsibility.** Disclose fees, costs, distribution, and stewardship allocations. Protect segregated liabilities and enforcement budgets. Optional paid services can fund operations, but protected rights and civic voice must not depend on selling personal data. Address host-community burdens, rather than merely reporting them. ^[S21]

Operational acceptance, not a moral seal

Before making a conformity claim, state the users, tasks, jurisdictions, components, versions, and exclusions being assessed. Test an ordinary useful task and adverse cases: refuse an optional data use; block an unauthorized tool action; challenge an erroneous output; obtain human review; reconcile a charge; and switch suppliers while exporting only authorized information. Include accessibility and an affected-nonmember complaint.

An independent reviewer should inspect the accountability tuples, permission evidence, known failures, actual remedies, resource-accounting methods, and funded maintenance. Missing essential authority, failed refusal or exit, breached mandatory protections, or unfunded promised remedies are no-go conditions. Material changes require renewed assessment.

Use graded, bounded descriptions:

- **Proposed:** a documented blueprint.
- **Design-conformant within stated scope:** specified requirements and operational tests are evidenced.
- **Independently evaluated outcome:** a named evaluator reports particular results, methods, dates, and limitations.

These labels do not create accreditation or establish that the Commons Collective website itself implements the full stack.

A counterexample and a possible hybrid

A fully open-source system that traps user histories, hides charges, or penalizes refusal fails this definition. Conversely, an open harness using a proprietary model and centralized payment service **could** support a bounded conformity claim if enforceable conditions, independent evidence, real substitution, protected rights, and funded remedies satisfy the requirements. Neither architecture qualifies automatically.

The stack is therefore **an allocation of accountable power across the whole system**, supported by technology, not a shopping list of software labeled open.

26. Open-agent incentives and a Commons Collective Agency Starter Kit

Recommendation: choose components for demonstrated agency, task fitness, and sustainable full cost, not the word "open." An open harness can make orchestration inspectable, customization feasible, and supplier switching credible. Those possibilities become collective agency only through tested controls, enforceable terms, and institutions that can exercise them.

Separate the harness from the model and service

A harness coordinates model requests, tools, context, and execution. Pi and Oh My Pi support configurable model/provider connections, so an open harness can call a proprietary API. Conversely, open or downloadable weights can be exposed through a closed managed service. Record openness separately for orchestration, policy configuration, tools, weights, licenses, and training-data documentation. [\[S58\]](#) [\[S59\]](#)

Local tool execution does not imply local model processing. Nor does a downloadable model establish an unrestricted license or complete training-data permission. Model and dataset cards are useful disclosures to inspect, not automatic clearance. Open policy files remain instructions; enforcement depends on operating controls and implementation. [\[S60\]](#) [\[S63\]](#) [\[S64\]](#)

Incentives that should be tested

Potential reason to choose an open or substitutable component	Cost or countervailing consideration	Evidence needed
Privacy and control over context, tools, and routing	Hosted inference may still transmit data; local isolation needs engineering	Observed data flows, contractual terms, access tests, and incident handling
Inspectability and customization	Maintaining a fork, reviewing dependencies, and retaining skills cost money	A reproducible configuration, review responsibilities, and a maintenance budget
Portability and stronger procurement leverage	Nominal alternatives may perform poorly or require expensive migration	Same-task comparisons and a successful, full-cost switch
Collective purchasing and shared improvements	Coordination can add overhead; contributors need support	Net member benefit, accountable decisions, and sustained shared maintenance
Independent evaluation and contestability	Benchmarks can be incomplete, manipulated, or costly	Relevant held-out tasks, uncertainty, conflicts, and replication

Compare quality, latency, setup, repair time, hardware utilization, accessibility, compliance, support, and convenience. **Full cost per accepted task** is more informative than token price alone. A managed proprietary component can be the better choice for a bounded workload if its protections and demonstrated exit meet the federation's requirements.

Where to look: an authoritative shortlist, not a trust seal

Resource	Useful layer or function	What not to infer
Pi Agent Harness , currently resolved by GitHub from <code>badlogic/pi-mono</code> to <code>earendil-works/pi</code>	Inspectable MIT-root-licensed orchestration and provider configuration	Pi explicitly leaves execution isolation to the deployment; do not assume files, processes, network, or credentials are restricted by the harness ^[S58]
Oh My Pi , <code>can1357/oh-my-pi</code>	Pi-derived harness and configurable providers; MIT root code license	No blanket assurance for dependencies, models, performance, or security ^[S59]
Official VS Code agent documentation	Understand harness, model, and execution choices	The phrase "Pi VS Code Agents" is ambiguous; no official product relationship is inferred here ^[S60]
Official MCP Registry , documented as preview	Discover published server metadata and publisher namespaces	Listing or namespace authentication is not a security or quality certificate ^[S61]
MCP architecture specification	Evaluate host/client/server interfaces and responsibilities	Protocol documentation does not prove correct implementation or compatibility with every harness ^[S62]
Hugging Face Model Cards	Inspect declared model uses, licenses, training information, and evaluations	Declarations do not independently establish capability or legal clearance ^[S63]
Hugging Face Dataset Cards	Inspect declared provenance, licensing, content, and bias considerations	A card is not consent or permission for all downstream uses ^[S64]
Inspect AI	Build comparative model evaluations using documented tasks, solvers, and scorers	A framework does not supply a universal ranking or certification ^[S65]

License statements above concern the identified code repositories, not every dependency, model, dataset, or hosted service. Verify the exact version before adoption. No packages or plugins were installed or executed to make this shortlist.

What the proposed starter kit contains

The **Commons Collective Agency Starter Kit** should be a source-controlled collection of configurations, policies, evaluation tasks, component passports, contract templates, and redacted run manifests. It is not one magic package, and the report does not claim that a complete production kit has already been implemented.

A component passport records maintainer and exact release/commit; license and dependency review; permissions and egress; hosting and upstream providers; task-evaluation method, sample, results, and date; full cost and support; export and migration evidence; governance and benefit terms; and incidents, appeals, and an accountable reviewer.

For nontechnical users, a convenient public catalog should offer comprehensible capabilities, permission manifests, reviewed releases, safe defaults, revocation notices, and independent challenge. A single discovery interface need not imply one unaccountable owner: permit

competing mirrors, portable records, and independent evaluators. Prompt packages and tool downloads are not safe merely because a trusted institution lists them. [S61] [S99]

The unfilled component-passport sketch is a starting artifact, not an approved component or runtime enforcement system. Keep credentials and personal information out of such public records. Mark missing evidence as unknown, not as a successful review.

Use eligibility gates before ranking. Establish required authority, rights, security, accessibility, and operational conditions; then compare eligible candidates by task outcomes and cost. Supplier payments cannot buy eligibility or higher quality scores.

Selected trace contribution and permission-safe shareback

Do not invite bulk uploads of raw agent logs. Start locally: select a useful example, remove credentials and unnecessary material, review third-party and employer interests, and approve recipient, purpose, and retention. Retrieved documents, tool results, and hidden context can contain information the user cannot authorize for redistribution; redaction alone does not establish anonymity or clearance. [S10] [S99]

Assess retrieval, evaluation, and training separately. Exclude material with unresolved authority. Share back permission-safe corrections, schemas, methods, and approved derived examples under compatible terms while keeping sensitive evidence access-controlled. A donation creates no missing rights or automatic reversal of completed training. Routing through an agent rather than a browser is not inherently safer; test permissions, egress, retention, isolation, and reviewability. [S26]

A workflow mapped to the agency chart

1. **Agent/browser:** specify the member's mandate, interruption, revocation, and approval boundaries; pin the harness and extensions. Use appropriate isolation and least privilege, and verify prohibited actions are blocked. A permissive code license is not a sandbox.
2. **Protocols/tools:** inventory endpoints, publisher identities, dependencies, permissions, and supported versions. Prefer narrow capabilities to unrestricted shell or database access. Test replacement rather than assuming a common protocol eliminates exit costs.
3. **Model/compute:** compare the same tasks across at least two eligible providers, retaining the current service as a baseline. Include a self-hosted option where feasible. Record routing, retention, failures, latency, and staffing; do not conceal fallback to a different provider.
4. **Database/knowledge:** start with synthetic or approved low-risk material. Apply purpose-limited, preferably read-only access; minimize logs; test export and applicable deletion; preserve others' rights.
5. **Provenance/validation:** link outcomes to source permissions, versions, and evaluation runs. Retain uncertainty and failed tests. Keep independent review distinct from supplier claims.
6. **Settlement/governance:** define purchasing authority, representation, payment approval, public benefits, complaints, and exit in the proposed Commons Collective Framework Agreement. Fund shared maintenance and independent assurance.

At each stage, identify an enforceable condition, beneficiary, refusal or exit path, and verification method. Without them, a claimed collective benefit remains an aspiration.

An adoption experiment, not a predetermined winner

Run a proposed four-week feasibility pilot over two low-risk workflows, with approximately 30 representative cases per workflow and repeated runs where affordable. This is a feasibility target, not a statistical power calculation. Plan the sample required for any subsequent confirmatory claims.

Compare eligible model providers using a fixed harness, then compare harnesses separately to avoid conflating effects. Track accepted-task quality, human repair, latency, full cost, prohibited-action attempts, data egress, and time to switch and recover working state. Publish limitations and negative findings without exposing secrets.

The counterhypothesis is that a proprietary managed service delivers better outcomes at lower full cost under sufficiently portable and accountable terms. The CCF can choose that result if supported. **Collective agency concerns who sets conditions, contests decisions, exits, and shares benefits.** Openness by itself supplies neither supplier democracy nor commons funding.

27. Private institutions and the commons service model

Correct the premise, then solve the narrower problem

The commons does not lack all business models. arXiv combines free reading and submission with institutional membership; Crossref combines membership and service fees with a free public metadata API; Wikimedia Enterprise offers commercial delivery services while maintaining commitments to free use and editorial independence. These examples do not prove that every commons can be commercially self-supporting or that individual contributors are paid. ^{[S80] [S81] [S82] [S83]}

The gap to address: a replicable way to convert demand for maintained, trustworthy information into durable revenue, shared benefits, and contributor agency without enclosing the underlying commons.

"Private institution" is not synonymous with "for-profit." A university, library, foundation, nonprofit, or company may sponsor a sector collective under different legal duties. The CCF should supply a tested operating method and shared infrastructure, not declare those differences irrelevant.

Stewardship and service operation

Keep two functions distinguishable: **commons stewardship** governs participation, permissions, public access, scientific/editorial integrity, and public-benefit allocation; **service operation** handles contracts, delivery, support, employment, billing, and commercial risk.

Initially these may be separate budgets and decision processes in one suitable organization. Use an independent steward and mission-bound operator where law, liability, financing, or credible independence requires separation. Wikimedia's organizational example is not a universal prescription. Do not incorporate multiple entities without a functional reason. ^[S80]

The federation can provide buyer discovery, assurance, procurement, and administrative support for disclosed fees. Sector collectives retain protected decisions. Neither investors nor paying customers should acquire permissions the collective cannot grant, editorial control, or additional constitutional votes through a purchase.

Who pays, and for what?

Sell maintained services: validated updates, quality checks, provenance diligence, integration, evaluations, availability commitments, purchasing support, and appropriately authorized commissioned work. Do not claim retroactive royalties for public-domain material or uses already permitted by open licenses. A commercial service can coexist with a free commons; Wikimedia Enterprise's principles provide one explicit example of that boundary. ^[S83]

Sector	Potential customer	Offering and pricing unit	Main cost	Protected boundary
Open access	Research platforms and libraries	Maintained metadata, correction feeds, and integrations; annual institutional/feed contract	Normalization and support	No charge merely for exercising existing reuse rights
Open data	Application providers and public agencies	Validated current observations and change records; maintained endpoint/site/year	Collection, quality control, hosting	Verified authority and permitted uses; sensitive records protected
Open science	Laboratories and AI developers	Reproducibility services and maintained evaluations; benchmark subscription or commissioned run	Expert review and computation	No purchased scientific outcomes or unauthorized participant data
Open education	Universities and education providers	Accessible resources, curriculum alignment, and integration; institutional support contract	Educator time and accessibility	Preserve resource freedoms and learner privacy
Open cultural heritage	Museums and research platforms	Digitization, contextual metadata, and maintenance; collection contract plus upkeep	Conservation, digitization, community review	Custody of an object is not authority over every use
Open journalism	Newsrooms and information services	Verified feeds, archive delivery, and specialist verification; annual feed plus commissioned work	Reporting, verification, corrections	No purchased editorial decisions or confidential-source disclosure

These are proposed offerings, not verified willingness-to-pay estimates. Each sector must test a specific customer's need and distinguish services from underlying rights.

Three funding obligations

Recurring anchor contracts should cover predictable delivery and ongoing operations. Variable projects should fund their own work and contribute a disclosed margin; do not assume they recur.

Grants and public finance should fund identified public benefits, such as underserved coverage, preservation, public interfaces, participation, or reusable infrastructure. Report these as funded public goods, not evidence of commercial demand.

Contributor funds require protected accounting. An agent cannot count every dollar passing through its accounts as revenue. Unpaid contributor balances, restricted grants, and commons money cannot silently become an operating reserve.

A separate hypothetical annual service-operator case

All amounts below are invented planning assumptions in annual USD, not observed prices, promised revenues, or committed grants. This is a separate maintained-services scenario, not an addition to the consumer-dues example in sections 7-8. Assume service fees are earned and collected within the year.

Six anchor contracts at \$90,000 generate \$540,000. Delivery costs \$42,000 per anchor, comprising \$30,000 contracted curation and \$12,000 infrastructure/support. Six integration projects at \$20,000 generate another \$120,000, with \$8,000 delivery cost each.

Operator cash-planning line	Annual amount
Earned service receipts	\$660,000
Direct service costs	\$300,000
Gross service contribution	\$360,000
Continuing operations, including an assumed \$24,000 CCF shared-services charge	\$240,000
Surplus before commons allocation	\$120,000
Contractually designated transfer to a restricted commons fund	\$60,000
Retained unrestricted cash before omitted obligations	\$60,000

The separate contributor-agency ledger receives \$100,000 of customer pass-through, pays \$80,000, and retains \$20,000 entirely against a contributor payable. **None of that \$100,000 is operator revenue.** It is distinct from the contracted curation already included in direct service costs. Appropriate regulated providers and legal segregation are required; this worksheet does not authorize commingling or holding client funds.

The commons fund receives a hypothetical \$120,000 restricted grant and the \$60,000 operator transfer. It pays \$130,000 for public maintenance and \$40,000 for access and participation, leaving \$10,000 restricted cash.

Across the operator-side and commons ledgers, external receipts are **\$880,000**, external outflows **\$790,000**, and closing cash **\$90,000**: \$60,000 unrestricted, \$20,000 owed to contributors, and \$10,000 restricted. Eliminate the internal \$60,000 transfer on consolidation. Reserve retention is not another expense. The CCF service charge is already within operations, not an additional deduction from the displayed result.

The [machine-readable worksheet](#) records these inputs, ledgers, and reconciliations; automated tests check the arithmetic.

Viable under assumptions does not mean adequately capitalized

The scenario has no annual operating gap under its assumptions, but a three-month unrestricted reserve target based on \$600,000 annual operator cash costs is \$150,000. Only \$60,000 is available, leaving a **\$90,000 reserve-funding gap**, before other launch and liquidity needs.

Sensitivity	Result before omitted obligations
Five anchors and six projects; corresponding delivery costs adjust	\$12,000 retained cash
Six anchors, no integration projects	\$12,000 deficit
No projects: seven \$90,000 anchors, or six anchors at \$92,000	Covers operations and the stated commons allocation before reserve building
Direct service costs rise 20%	\$60,000 baseline surplus falls to zero
Service revenue falls 10% without cost reductions	\$6,000 deficit

Sensitivity	Result before omitted obligations
One anchor is lost and its \$42,000 delivery cost disappears	\$12,000 retained cash
One anchor is lost but delivery costs remain sticky	\$30,000 deficit

The commons program is separately grant-dependent: without its \$120,000 grant, \$170,000 of planned spending exceeds the \$60,000 operator contribution by \$110,000. That public-funding requirement is not evidence that the operator generates \$170,000 of market-financed commons support.

This is a cash-planning model, not a statutory P&L. Tax, VAT/sales tax, depreciation, debt service, launch costs, deferred income, grant conditions, principal-versus-agent treatment, and invoice timing require professional review and a monthly model. The reserve gap does not cover all of those omissions.

Founding and capital discipline

By day 90, seek two independent paid pilots specifying deliverables, acceptance, permissions, cancellation, payment timing, and contributor obligations. Expressions of interest are not sales. By day 180, seek renewal or conversion from actual purchasing budgets and compare quality-adjusted full cost with direct licensing, open alternatives, and in-house provision.

Track payment timeliness, participation, decisions changed by members, complaints, public access, and migration alongside revenue. Publish bounded exit obligations and protect earned payments. Stop or redesign when customers do not renew, permissions remain unresolved, buyers demand editorial control, contributor costs are hidden, or operations depend on undisclosed permanent subsidy.

Use grants for institution-building and demonstrated public goods. Consider recoverable working capital only when contracts and credible economics support repayment, subject to legal review; program-related-investment eligibility is not automatic. Openness and member control can limit exclusivity, investor control, and margins. Some operators may support patient bounded-return capital; others belong in membership- or grant-funded public infrastructure.

Founding checklist: identify the customer problem; separate paid service from underlying rights; assign steward/operator duties; calculate full costs and recurring requirements; protect restricted and contributor money; set paid-pilot, renewal, concentration, and shutdown gates; and disclose capital needs and repayment assumptions.

The approach does not solve viability by declaration. It makes the business and public-benefit bargain **testable, financeable where justified, and accountable when subsidy remains necessary**.

28. The quasi-commons tension: contribution without control

The important tension is distributed contribution without corresponding shared control.

It is not established that every Claude Code interaction changes model weights, or that AI contributors outnumber everyone who has ever contributed to Wikimedia. Those stronger claims require evidence and comparable definitions.

"Conscription" can express concern about unequal bargaining power and meaningful choice, but should not become a blanket allegation of involuntary training use. Accounts, settings, feedback mechanisms, authority, and actual downstream processing matter.

Context, memory, feedback, and weights are different

Anthropic describes Claude Code as an agentic harness around models, tools, and an execution environment. Prompts, retrieved files, tool results, corrections, and test outcomes can inform subsequent responses. Project instructions in `CLAUDE.md`, auto-memory, and saved sessions can support future interactions without proving a change to underlying model weights. ^[S91]

Distinguish **session context, persistent memory, tool traces, operational telemetry, explicit feedback, and material selected for future training**. Collection or retention in one category does not establish movement into another. Training eligibility is not proof of dataset inclusion, an update, or a traceable contribution to a model capability. ^{[S90] [S92] [S93]}

Provider policy varies by account and mechanism

The official documentation reviewed on 18 September 2026 distinguishes:

- **Consumer Free, Pro, and Max:** model-improvement settings can permit account data, including associated Claude Code sessions, to train future models. The privacy article also describes explicit opt-ins, feedback, and particular safety-review purposes. Safety-related use must not be generalized into unrestricted training permission. ^{[S90] [S92]}
- **Commercial Team, Enterprise, and API use:** the documented default is no model training on commercial inputs or outputs. Explicit feedback, separately authorized sharing, or programs such as the Development Partner Program can create exceptions. Identify the actual account, authentication route, and agreement rather than relying on the product name. ^{[S90] [S93]}
- **Different feedback paths:** thumbs-up/down feedback may make the associated conversation available for training under the relevant privacy terms. The documented Claude Code session-quality survey, including its separate optional transcript-sharing step, is described differently and as not usable for model training. Do not flatten these into one rule. ^{[S90] [S92] [S93]}

These are statements of provider policy, **not an independent audit of implementation**. They establish neither universal training nor universal exclusion. Recheck the applicable current terms before making a deployment decision.

Two Alice cases

Commercial task: Alice uses a commercial API for employer work. Her prompts and tool outputs assist the task; absent an applicable opt-in or feedback exception, the documented default is no model training on those inputs and outputs. That does not mean processing is local, eliminate every retention question, or establish her authority to send the employer's material. It does mean task usefulness alone is not evidence of a future weight contribution. ^{[S90] [S91] [S93]}

Optional improvement contribution: Alice uses a consumer account with model improvement enabled and submits relevant feedback. The conversation can become eligible for future training under the documented arrangements. Selection into a specific run is still not established. Her choice cannot automatically authorize use of a colleague's personal information, third-party code, or confidential client material. ^{[S90] [S92]}

Audit the claims before counting contributors

Proposed claim	What the evidence supports	Limit and next measurement
Every interaction augments weights	Context and tool results inform subsequent responses ^[S91]	Contextual adaptation is not a demonstrated parameter update; distinguish inference, memory, collection, and training records
All users contribute training data	Consumer opt-in and feedback routes exist ^{[S90] [S92]}	Defaults and mechanisms differ; measure eligibility and actual selected contribution separately
AI contributors exceed all-time Wikimedia contributors	Wikimedia publishes defined participation metrics ^{[S94] [S95]}	No comparable all-time person totals were established; agree period, scope, contribution threshold, and deduplication
Dispersed inputs make a governed commons	Many actors can provide useful inputs	Contribution does not establish shared ownership or decision rights; inspect governance, benefits, and exit

Wikimedia's active-editor metric counts registered, non-bot editors making at least five edits in a month. Its newly registered-user metric counts specified account creation, not editing or unique human contributors. Neither establishes an all-time count of everyone contributing across Wikimedia. ^{[S94] [S95]}

Likewise, AI users, paid accounts, eligible training participants, selected contributors, and paid evaluators are different populations. Comparing cumulative AI accounts with monthly active Wikimedia editors would be invalid. The reviewed evidence does **not** establish the "more than ever contributed" proposition.

A defensible study would specify Wikipedia versus all Wikimedia projects, a common period, qualifying activity, anonymous contributions and bots, and cross-account/project duplication. Provider-side counts should separate authorized eligibility, actual inclusion, and paid evaluation work. Use privacy-preserving aggregate measurement and disclose uncertainty; do not construct unnecessary cross-platform identity tracking to produce a dramatic headline.

A critical working concept, not automatic ownership

Use **quasi-commons** here to describe dispersed contributions that may support a widely used infrastructure while control remains concentrated. This is a critical working term, not a legal classification. Public inputs, private weights, and public benefits can coexist without a governed commons in which participants have decision rights.

Counterarguments belong in the analysis. Users obtain useful services; providers fund research, compute, engineering, and safety; terms and choices vary; and supplying an input does not automatically create an ownership share in the whole model. The institutional question is what representation, compensation, assurance, and bargaining arrangements are justified, not whether every interaction already creates a property claim. Commons and relational-governance scholarship informs that inquiry without deciding it automatically. [\[S03\]](#) [\[S25\]](#)

The Commons Collective response

1. Separate necessary processing from optional model improvement and make the consequences of account settings understandable.
2. Negotiate evidence-use receipts that identify authority, purpose, processing category, and known inclusion decisions, rather than pretending to attribute every model capability to a particular record.
3. Use collective procurement to obtain meaningful defaults, feedback controls, assurance, remedies, and tested exit.
4. Pay deliberate elicitation and evaluation work, and negotiate commons-benefit covenants without asserting rights over third-party or public-domain material.
5. Give affected groups a contestable role in objectives and evaluation while preserving minority protections and applicable individual rights.

Explain what withdrawal can stop and what it cannot reliably reverse after processing or training.

Accountable contribution under legitimate authority, rather than rhetorical conversion of every user into an involuntary owner or unpaid worker, is the testable institutional objective.

29. Replicable sector collectives: blueprint and templates

The [companion research paper](#) defines a replicable institutional grammar: scoped procurement, contributor agency, independent commons stewardship, professional operations, independent assurance, and appropriate payment services. It proposes autonomous sector collectives connected by voluntary protocols, not a franchisor or a universal licensing cartel.

Map, adapt, and federate before founding. Forking a document does not create an organization or authority. Create a new body only when existing institutions cannot legitimately supply the missing functions and a constituency, viable operator, safeguards, and exit budget are evidenced.

Sector	Distinct scope	Foundational boundary
Open access	Scholarly publication, preservation, and access infrastructure	Free reader access does not eliminate production costs or create authority over all research materials ^[S70]
Open data	Documented datasets, reuse, and maintenance	Open permissions and machine-readable access do not override privacy or third-party rights ^[S71]
Open science	Research methods, code, instruments, evaluation, and outputs	Broader than publication access; justified restrictions and participant/community obligations remain ^[S72]
Open education	Open resources, adaptation, accessibility, and teaching support	OER openness is not permission to extract learner records or waive safeguarding ^[S73]
Open cultural heritage	Community-directed description, preservation, and approved access	Custody, legal rights, public-domain status, and community protocols must be assessed separately ^[S20]
Open journalism	Public-interest reporting and enabling services	Not all reporting, notes, or sources are open; editorial and source independence are protected ^[S74]

The editable [replication pack](#) includes:

- [Collective charter](#): purpose, decision gates, representation, authority, and federation limits.
- [Bargaining mandate](#): specified principals, rights, purposes, counterparties, term, withdrawal, and continuing obligations.
- [Governance and benefits](#): protected decisions, segregated liabilities, independent appeal, succession, and wind-down.
- [Pilot and evaluation](#): cost-inclusive comparisons, evidence gates, stopping rules, and publication of negative results.
- [Business model](#): customer and payer evidence, service pricing, separate ledgers, capital, reserves, and renewal tests.
- [Unfilled sector profile](#): six distinct adapters with local fields unresolved and all essential approvals not-reviewed.

Original templates and research are CC BY 4.0; original software/tests are MIT. Cited works retain their terms. The profile is expressly invalid for operation; no file creates an entity, opens an account, processes money, certifies compliance, or authorizes representation. Local legal,

financial, community, and professional validation is part of the work, not a box solved by cloning.

The [source-linked overview](#) also loads the companion paper's current abstract and the work register on opening. Changes to the source paper, main research, or work data flow through the normal publication workflow without maintaining a separate slide narrative.

30. Conclusion

Commons Collective is best understood as an effort to align **voice, authority, service, and benefit**, not as a universal auction of personal data. Consumer purchasing power can operate without data sales. Creator bargaining can operate without enclosing the public domain. Commons funding can operate without pretending to identify every person's contribution to a model. Communities can demand a say in physical infrastructure without first becoming data vendors.

The institutional wager is that coordinated, accountable representation can negotiate better terms than isolated consent screens. The ethical constraint is that new bargaining power must not be purchased by overriding others' rights or rebuilding extraction inside a cooperative wrapper. The scientific obligation is to test that wager, disclose its costs, and remain willing to reject it.

Recommended first move: a small, member-governed AI purchasing pilot linked to an independently evaluated, rights-cleared educational retrieval service and a transparent commons covenant. Keep private-assistant processing, creator licensing, commons spending, and energy-community bargaining distinct enough to audit, but coordinated enough to improve collective agency.

A WORKING, INSPECTABLE PROTOTYPE

Follow the money.

A read-only rollup of the selected Simulation Lab source. The summary and detailed model use the same cents, financial policies, and cumulative window.

Change assumptions in the Lab, then use **Inspect money flows** to bring that exact scenario here. No independent allocation percentages are applied.

[Open the startup plan in the Lab](#) · [Inspect the 14,000-trial study and its data](#)

Startup projection, month 0 through 36

Conditional projection, not executed payments

Source identity: c54691b8aa456b8c

Sources, including financing

\$15,781,305.00

Uses, including commons transfers

\$10,776,509.75

Closing operator cash / projected position

\$5,004,795.25

Financing included in sources, not revenue

\$6,000,000.00

Received / projected service-fee revenue

\$9,781,305.00

Unpaid operator obligations

Not modeled / no applicable value

\$15,781,305.00 sources - \$10,776,509.75 uses = \$5,004,795.25. Reconciliation difference: \$0.00.

Commons allocation is 20% of administration fees, not of investor capital or provider subscriptions.

► Inspect the reconciled accounts, transfers and liabilities

Transferred to the commons in this window

\$1,956,261.00

Capital needed for the full planning horizon, including reserve and contingency

\$4,151,298.47

Last paid cohort: price after fee per member/month

\$19.11

Last paid cohort: saving versus equivalent retail per member/month

\$0.89

- This is the operator perimeter. AI-provider subscription payments are not operator revenue.
- The forecast models transfers to an independent commons fund, not its downstream cash balance or contributor distributions.
- Unpaid liabilities are not modeled in this cohort cash plan. A negative projected position is a funding gap, not negative executed cash.

- Reserve and contingency are funding requirements, not extra uses of funds. No automatic 40/30/20/10 distribution is applied.

All calculations and file reads stay in your browser. An explicit “Inspect money flows” or “Open this source” action uses one-time storage in this tab, removed when the destination reads it. No scenario is uploaded. If local sharing is unavailable, download and import its JSON.

[Explore the infrastructure divergence diagram](#). All model results are conditional on synthetic assumptions, not validated forecasts.

EVIDENCE REGISTER

Sources & their limits

Primary materials and academic research. Access dates are recorded with each source. A source supports the stated claim, not every proposal built upon it.

S01 / 2026

Ethical Tech CoLab. [Repository snapshot: mission, vision, and working principles \(stored About-page copy\)](#) · [Ethical Tech CoLab public website \(organizational context\)](#)

The public repository copy lists Field-grounded, Prototype-first, and Open by default. Its mission concerns realizing human potential through technology; its vision specifically concerns technology shaped by affected people.

Limit: This is a pinned TypeScript content file, not JSON or a rendered principles page. The About route is disabled in the cited snapshot and in the current implementation inspected on the access date. The record supports stored repository wording, not verified public-page display, formal charter status, or endorsement of this project. The context link is the public homepage, not evidence that this wording appears there. Accessed 2026-09-20.

S02 / 2026

Ethical Tech CoLab. [What Is Ethical AI?](#)

The CoLab's July 2026 paper frames ethical AI institutionally, with human rights, avoidance of harm, ethical deliberation, and affected-person participation.

Limit: Cited for the CoLab's normative framework, not as independent validation of the present proposals or of every historical claim in that paper. Accessed 2026-09-18.

S03 / 2009

Elinor Ostrom. [Beyond Markets and States: Polycentric Governance of Complex Economic Systems](#)

Nobel lecture distinguishes resource subtractability and exclusion and describes collective-choice arrangements, monitoring, sanctions, and conflict resolution. Printed pages 412-413 and 422-423.

Limit: Knowledge and pasture have different resource economics. Institutional findings are not a universal template or an endorsement of an AI-data proposal. Accessed 2026-09-18.

S04 / Undated; accessed 2026

U.S. Bureau of Land Management. [Livestock Grazing on Public Lands](#)

Describes permits or leases, eligibility, authorized grazing privileges, and fees, including animal-unit months as a forage measure.

Limit: A publicly managed, subtractable resource regime; it establishes neither public ownership of internet content nor authority to license others' data. Accessed 2026-09-18.

S05 / 2018

Imanol Arrieta-Ibarra, Leonard Goff, Diego Jimenez-Hernandez, Jaron Lanier, and E. Glen Weyl. [Should We Treat Data as Labor? Moving beyond 'Free'](#)

Argues for treating data partly as labor and for countervailing power against dominant data buyers. AEA Papers and Proceedings 108: 38-42.

Limit: An institutional argument, not a demonstrated viable cooperative, universal data price, or legal finding of employee status. Publisher abstract and bibliographic record verified. Accessed 2026-09-18.

S06 / 2021

Nicholas Vincent, Hanlin Li, Nicole Tilly, Stevie Chancellor, and Brent Hecht. [Data Leverage: A Framework for Empowering the Public in its Relationship with Technology Companies](#)

Examines public power arising from dependence on human-generated data, including withholding and conscious contributions to alternatives. Warns of collateral harm to public commons. FAcCT 2021.

Limit: A framework and synthesis, not a contemporary frontier-model bargaining trial. Originally posted as a 2020 preprint, revised in 2021. Accessed 2026-09-18.

S07 / 2019

Sylvie Delacroix and Neil D. Lawrence. [Bottom-up data Trusts: disturbing the 'one size fits all' approach to data governance](#)

Proposes plural independently administered trusts with fiduciary obligations and negotiation tied to beneficiaries' purposes. International Data Privacy Law 9(4): 236-252; doi:10.1093/idpl/ipz014.

Limit: A proposed model, not universally recognized ownership of personal data or an international trust regime. Accessible university full text notes limits on erasure and bargaining. Accessed 2026-09-18.

S08 / Undated; accessed 2026

Creative Commons. [Frequently Asked Questions](#)

Explains CC's standardized legal tools and sharing mission, the relationship to copyright, and differences and limits across licenses.

Limit: General institutional guidance, not legal advice, a universal determination of AI-training legality, or evidence that CC operates the proposed payment intermediary. Accessed 2026-09-18.

S09 / Undated; accessed 2026

Creative Commons. [CC Signals](#)

Describes a developing commons-friendly framework for communicating expectations about AI use and advancing reciprocity alongside existing licensing tools.

Limit: A developing institutional framework; a preference signal is not automatically a new legal obligation or a retroactive amendment to existing CC licenses. Accessed 2026-09-18.

S10 / 2016

European Union. [Regulation \(EU\) 2016/679 \(General Data Protection Regulation\)](#)

Articles 5-9, 17, 21, and 80 address processing principles, lawful bases, consent, specified rights, and representative enforcement; recital 26 distinguishes anonymity from pseudonymity.

Limit: Not a general property or collective-licensing right. Withdrawal does not invalidate prior lawful processing; erasure and other rights have conditions and exceptions. Accessed 2026-09-18.

S11 / 2022

European Union. [Regulation \(EU\) 2022/868 \(Data Governance Act\)](#)

Recognizes data-cooperative services and regulates specified intermediation and data-altruism recognition. Articles 2(15)-(16), 10-12, and 18-19; applicable from 24 September 2023.

Limit: No new personal-data processing basis or competition-law immunity. Altruism excludes rewards beyond specified cost compensation; neutrality and organizational requirements matter. Accessed 2026-09-18.

S12 / 2024

European Union. [Regulation \(EU\) 2024/1689 \(Artificial Intelligence Act\)](#)

Article 53(1)(c)-(d) establishes GPAI copyright-policy and public training-content-summary duties; Articles 111(3) and 113 set relevant phased dates. Read with the 2026 amendment.

Limit: Does not establish universal royalties or an item-level public training ledger. The limited open-source exception does not remove those two duties. Accessed 2026-09-18.

S13 / 2022

European Commission. [Guidelines on the application of Union competition law to collective agreements regarding the working conditions of solo self-employed persons](#)

Communication 2022/C 374/02 distinguishes specified working-condition agreements outside Article 101 from categories subject to Commission nonintervention.

Limit: Guidance, not new social rights or a blanket data-licensing exemption. Paragraphs 17-18 limit scope, including exclusion of mere asset exploitation. Accessed 2026-09-18.

S14 / 2025

U.S. Copyright Office. [Copyright and Artificial Intelligence, Part 3: Generative AI Training](#)

Discusses training and exclusive rights, contextual fair-use analysis, licensing markets, and targeted policy intervention. Printed pages 99 and 105-108.

Limit: The 9 May 2025 pre-publication advisory report, not a statute or court judgment. Its litigation discussion is not a verified September 2026 docket update. Accessed 2026-09-18.

S15 / 2025-10-14

Omidyar Network. [Humanity AI launches \\$500 million commitment to mobilize civil society in shaping the future of AI](#)

Official founder announcement of ten philanthropies committing \$500 million over five years, with aligned grantmaking and a pooled fund managed by Rockefeller Philanthropy Advisors.

Limit: An announced collective commitment is not evidence of completed disbursement, a single available funding pot, or an award to Commons Collective. Accessed 2026-09-18.

S16 / Undated; accessed 2026

Costco Wholesale Corporation. [Company Profile](#)

Describes membership warehouses, membership fees, and lower-price positioning; identifies Costco as a publicly traded corporation.

Limit: Costco is not a member-owned cooperative. Its company description does not validate pooled AI procurement, data sales, or democratic governance. Accessed 2026-09-18.

S17 / Undated; accessed 2026

Google. [Learn how Gmail keeps your information secure & protects your privacy](#)

Gmail's published explanation describes processing for smart features and spam detection, user controls, and a distinction between email content and advertising.

Limit: A provider's product-specific policy statement, not an independent audit or a blanket policy covering every AI service, account type, setting, or model-training use. Accessed 2026-09-18.

S18 / 2013

W3C Provenance Working Group. [PROV-DM: The PROV Data Model](#)

The Recommendation models provenance using entities, activities, agents, derivations, and related structures for interoperable exchange.

Limit: A representation standard; it does not establish ownership, verify the truth of a recorded assertion, or supply a compensation formula. Accessed 2026-09-18.

S19 / 2025

W3C Verifiable Credentials Working Group. [Verifiable Credentials Data Model v2.0](#)

Provides an interoperable model for issuer claims and discusses correlation, data minimization, and credential privacy risks.

Limit: A credential's validity does not establish the substantive truth or legal sufficiency of its claims; privacy depends on implementation. Accessed 2026-09-18.

S20 / Undated; accessed 2026

Local Contexts. [Traditional Knowledge \(TK\) Labels](#)

Community-developed labels communicate Indigenous protocols and responsibilities for culturally situated knowledge and materials.

Limit: Not a generic commercial license, a replacement for community engagement, or proof of a platform's authority to commercialize cultural materials. Accessed 2026-09-18.

S21 / 2025

International Energy Agency. [Energy and AI: Energy demand from AI](#)

Estimates global data-center electricity consumption around 415 TWh in 2024 and projects around 945 TWh by 2030 in its Base Case, with substantial uncertainty.

Limit: Figures concern total data centers, not AI alone; 2030 is a scenario projection from the 2025 report, not observed consumption or a new 2026 forecast. Accessed 2026-09-18.

S22 / 2023; v3 revised 2024

Ilya Shumailov, Zakhar Shumaylov, Yiren Zhao, Yarin Gal, Nicolas Papernot, and Ross Anderson. [The Curse of Recursion: Training on Generated Data Makes Models Forget](#)

Analyzes model-collapse mechanisms in recursive learning from generated data under the paper's theoretical and experimental conditions.

Limit: A version-specific preprint reference; it does not show that every synthetic-data method fails or establish copyright or privacy clearance. Accessed 2026-09-18.

S23 / 2003

Jean-Charles Rochet and Jean Tirole. [Platform Competition in Two-Sided Markets](#)

Models participation and price structure across platform sides, including profit-maximizing and nonprofit arrangements. Journal of the European Economic Association 1(4): 990-1029.

Limit: Stylized platform theory, not evidence that an AI-data cooperative will succeed. The university record also links a 2002 working-paper version. Accessed 2026-09-18.

S24 / 2022

Daron Acemoglu, Ali Makhdoumi, Azarakhsh Malekian, and Asu Ozdaglar. [Too Much Data: Prices and Inefficiencies in Data Markets](#)

Models how one person's disclosure can reveal information about others, depress prices, and generate excessive sharing. AEJ: Microeconomics 14(4): 218-56.

Limit: A theoretical model, not a measured estimate of the externality in the proposed pilot. Low prices do not imply low social stakes. Accessed 2026-09-18.

S25 / 2021

Salome Viljoen. [A Relational Theory of Data Governance](#)

Analyzes population-level data relationships and interests beyond individual consent or payment; warns that contributor-only trusts can omit affected noncontributors. Yale Law Journal 131: 573-654, especially pages 598 and 645-648.

Limit: Normative legal scholarship, not enacted institutional authority or empirical validation of the proposed federation. Accessed 2026-09-18.

S26 / 2013; version 4.0

Creative Commons. [Attribution 4.0 International: Legal Code](#)

Provides a nonexclusive, royalty-free, irrevocable license over licensed rights, with terms, independent exceptions, limits on additional restrictions, and breach/reinstatement provisions.

Limit: CC BY 4.0 specifically. Privacy/publicity and other rights remain separately addressed; no universal answer to every AI use or permission over rights the licensor does not hold. Accessed 2026-09-18.

S27 / 2019

Amirata Ghorbani and James Zou. [Data Shapley: Equitable Valuation of Data for Machine Learning](#)

Allocates contributions to predictive performance for a specified learning algorithm and evaluation measure, with computational approximations. ICML, PMLR 97: 2242-2251.

Limit: Not an intrinsic cash price, moral entitlement, or share of corporate revenue. Exact computation is exponentially costly, and fairness assumptions may not fit every setting. Accessed 2026-09-18.

S28 / 2026-09-10

Humanity AI. [Humanity AI Announces \\$10 Million Open Call for Grants Supporting Community-Led Projects](#)

Announces a \$10 million competitive call for U.S.-based community-led work in education, labor and economy, and humanities, arts, and culture.

Limit: The \$10 million is the call-wide envelope, not an individual award or funding for this report's separate hypothetical strategy. Application conditions may change. Accessed 2026-09-18.

S29 / 2026

Humanity AI. [Open Call for Proposals: Communities Leading on AI](#)

Lists \$75,000-\$1 million grants, terms up to three years, U.S. community focus, nonprofit/fiscal-sponsorship requirements, and a 21 October 2026 deadline at 7:59 p.m. ET.

Limit: Research-only and primarily AI-literacy/adoption projects are out of scope. The call describes grants, not a PRI program. Check live terms before applying. Accessed 2026-09-18.

S30 / 2013

Alessandro Acquisti, Leslie K. John, and George Loewenstein. [What Is Privacy Worth?](#)

A field experiment finds that endowments and presentation order affect consumer privacy choices. *Journal of Legal Studies* 42(2): 249-274.

Limit: Context-dependent privacy preferences, not a universal price for personal data or an estimate of AI training value. The repository upload date differs from the 2013 publication date. Accessed 2026-09-18.

S31 / 2020

Charles I. Jones and Christopher Tonetti. [Nonrivalry and the Economics of Data](#)

Models simultaneous use of data by multiple firms, gains from nonrival use, and inefficiencies from hoarding. *American Economic Review* 110(9): 2819-58.

Limit: Model-dependent welfare results, not a claim that unrestricted sharing is always beneficial or will finance its own stewardship. Accessed 2026-09-18.

S32 / 2026

European Union. [Regulation \(EU\) 2026/1744 \(Digital Omnibus on AI\)](#)

Enacted 8 July 2026 and in force 27 July 2026; changes relevant Annex III high-risk dates to 2 December 2027 and Annex I dates to 2 August 2028 while retaining the principal GPAI timetable discussed here.

Limit: Not a delay of all AI obligations. Interpret the specific application/transition provisions and obtain current model- and use-specific legal review. Accessed 2026-09-18.

S33 / 1935, as amended

United States. [National Labor Relations Act: 29 U.S.C. 152](#)

Section 152(3)'s definition of employee excludes independent contractors and other specified categories, limiting assumptions about data-union labor-law status.

Limit: Exclusion from this statute alone does not establish that every collective negotiation is unlawful under other law; actual employment classification requires factual analysis. Accessed 2026-09-18.

S34 / 2024

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](#)

Models trained on personal data are not automatically anonymous; analysis considers reasonably likely extraction/disclosure and contextual lawful-basis assessment. Executive summary, pages 2-4.

Limit: A regulatory opinion, not a statute or a finding that every model or synthetic dataset contains personal data. Accessed 2026-09-18.

S35 / 2019

European Union. [Directive \(EU\) 2019/790 on copyright and related rights in the Digital Single Market](#)

Articles 3-4 address text-and-data mining and lawful access, with rights reservations under Article 4; Article 7 protects the specified research exception against conflicting contracts.

Limit: Requires national implementation and fact-specific analysis. Research labels do not automatically qualify; public accessibility is not unlimited permission. Accessed 2026-09-18.

S36 / 2010

Thomas Helbling, International Monetary Fund. [Back to Basics: What Are Externalities?](#)

Explains external costs, Pigouvian taxation, positive externalities, and public goods, supporting a distinction between harm correction and commons subsidy.

Limit: An economic explanation, not an estimate of an optimal AI tax rate or a justification for a uniform 50% tax. Accessed 2026-09-18.

S37 / 2026

U.S. Internal Revenue Service. [Credits and deductions](#)

Explains that credits reduce tax due while deductions reduce taxable income.

Limit: Does not establish an existing AI-commons or Wikimedia contribution credit, corporate eligibility, or the proposed calculation's legal validity. Accessed 2026-09-18.

S38 / 2026-06-18

Office of Senator Bernie Sanders. [Sanders Introduces Legislation to Create \\$7 Trillion AI Sovereign Wealth Fund](#)

Announces the American AI Sovereign Wealth Fund Act, a one-time equity-paid levy seeking 50% public ownership of covered AI companies. Linked proposed section 5000E(b) specifies newly issued equity and post-issuance ownership.

Limit: An introduced proposal, not evidence of enactment. Fund valuation and dividend amounts are sponsor projections, not independently verified forecasts; no Wikimedia-credit mechanism is established. Accessed 2026-09-18.

S39 / 2026-03-02

Office of Senator Bernie Sanders. [Sanders and Khanna Introduce Legislation to Tax Billionaire Wealth and Invest in Working Families](#)

Describes the Make Billionaires Pay Their Fair Share Act and a proposed 5% annual billionaire wealth tax.

Limit: Different base and mechanism from an AI-company revenue, profit, or equity levy. Sponsor-supported projected receipts are not realized revenue. Accessed 2026-09-18.

S40 / 2025-10-06

Office of Senator Bernie Sanders. [Sanders Releases Report on Big Tech Oligarchs' War Against Workers, Warns AI Could Eliminate Nearly 100 Million U.S. Jobs](#)

Includes a proposal for a robot tax on corporations replacing workers with machines; the announcement does not specify a tax rate.

Limit: A policy proposal. The headline's job-loss figure is a sponsor-reported scenario, not adopted here as a validated forecast; it does not establish a 50% rate. Accessed 2026-09-18.

S41 / 2025 revision

Wikimedia Foundation. [Gift policy](#)

Describes gift review, a preference for unrestricted support, additional scrutiny of large or restricted gifts, and the Foundation's right to refuse donations.

Limit: Not an endorsement of the proposed tax credit or designation as an eligible recipient under any new scheme. Recipient independence and actual law remain controlling. Accessed 2026-09-18.

S42 / 2025; inspected revision 2026

ERC-8004 authors and Ethereum ERC contributors. [ERC-8004: Trustless Agents](#)

Draft specification proposes identity, reputation, and validation registries; includes contextual feedback and off-chain aggregation, and explicitly separates payments from scope.

Limit: Draft at the research cutoff, not a final standard or evidence of any specific deployed contract. Identity and integrity commitments do not prove capability, authority, or freedom from Sybil manipulation. Accessed 2026-09-18.

S43 / 2026 snapshot

ERC-8004 contracts contributors. [ERC-8004: Trustless Agents - curated contracts implementation](#)

Implementation includes validation requests/responses while its README flags ongoing validation-registry discussion and anticipated specification updates.

Limit: Source existence is not proof of finalized design, independent audit quality, economic security, or uniformly deployed behavior; no deployment address was verified. Accessed 2026-09-18.

S44 / 2026 snapshot

Ethereum Attestation Service contributors. [Ethereum Attestation Service contracts](#)

Schemas and attestations represent issuers, linked records, timestamps, expiry, revocation, and custom data. Applications must evaluate status and policy explicitly.

Limit: Attestation infrastructure, not a universal reputation algorithm or truth verifier. The inspected isAttestationValid implementation checks existence, not every condition an application needs. Accessed 2026-09-18.

S45 / Dynamic page; accessed 2026

Costco Business Delivery. [Search | Costco](#)

Delivered official search configuration contains Best Match with descending-score ordering and other sorting vocabulary, including price, ratings, and best sellers.

Limit: Configuration does not disclose scoring inputs or establish quality optimization, visible controls in every session, or identical behavior across Costco consumer channels. The proprietary ranking algorithm was not verified. Accessed 2026-09-18.

S46 / Undated live documentation; accessed 2026

Waze / Google. [About Waze](#)

Describes how driving with Waze open contributes real-time information used for speeds, routing, and road information, alongside community reports.

Limit: A product description, not an independent causal evaluation of participation incentives or proof of contributor ownership or open licensing. Accessed 2026-09-18.

S47 / Undated live documentation; accessed 2026

Waze / Google. [Earn points on Waze](#)

Documents current participation points for driving, reports, confirmations, and map contributions.

Limit: Points are not a promise of cash compensation. The current schedule does not establish historical schedules or the causal effect of points on participation. Accessed 2026-09-18.

S48 / Undated live documentation; accessed 2026

Waze / Google. [Join our community](#)

Describes map-editor, localization, and beta-testing communities and contributions to map accuracy and currency.

Limit: Community participation does not establish corporate ownership, binding contributor votes, or an open-data license. Accessed 2026-09-18.

S49 / Undated live documentation; accessed 2026

Weather Underground. [PWS Network Overview](#)

Describes local conditions, visualization, historical archives, and sharing benefits for personal weather-station owners, with a participation path involving equipment purchase and installation.

Limit: Not proof of payments per reading or independent quality improvement. Inconsistent live network-size figures were not used; API and subscription entitlements are not established by this page. Accessed 2026-09-18.

S50 / Undated live documentation; accessed 2026

Weather Underground. [Installing your Personal Weather Station](#)

Explains how station placement affects temperature, wind, rainfall, and humidity measurements, including sunlight and obstruction problems.

Limit: Installation guidance, not evidence that every contributing station complies or that all readings receive equivalent quality control. Accessed 2026-09-18.

S51 / Undated live documentation; accessed 2026

Weather Underground. [Personal Weather Station Buying Guide](#)

Discusses compatible equipment, sensor quality, connectivity, and hardware choices for contribution to the station network.

Limit: Some product examples may be old. The guide neither quantifies typical contributor expenditure nor establishes compensation for observations. Accessed 2026-09-18.

S52 / Undated live documentation; accessed 2026

The Weather Company. [Weather Data APIs](#)

Describes commercial weather APIs and a limited 30-day trial.

Limit: A commercial product offer, not evidence of a personal weather-station contributor's API or subscription entitlement. Historical contributor-access terms were not verified. Accessed 2026-09-18.

S53 / 2017

Paul Christiano, Jan Leike, Tom B. Brown, Miljan Martic, Shane Legg, and Dario Amodei. [Deep reinforcement learning from human preferences](#)

Section 2 describes learning a reward predictor from human comparisons of behavior and optimizing an agent against the predicted reward.

Limit: Experiments concern simulated locomotion and Atari tasks, not representative public governance, democratic legitimacy, or regulatory authority. Accessed 2026-09-18.

S54 / 2022

Long Ouyang and colleagues. [Training language models to follow instructions with human feedback](#)

Documents instruction fine-tuning, preference rankings, reward modeling, and PPO; section 3.5 discusses normalization, KL regularization, and the combined objective.

Limit: Reported effects depend on the studied models, prompts, and labeler populations. A reward coefficient does not imply an equivalent share of behavioral influence. Accessed 2026-09-18.

S55 / 2022

Yuntao Bai and colleagues. [Constitutional AI: Harmlessness from AI Feedback](#)

Demonstrates principle-guided critique and revision and AI-generated preference judgments used in training and reinforcement learning from AI feedback.

Limit: Developer-selected principles and AI judges do not remove human governance choices. The paper treats principle selection as exploratory, not democratically settled. Accessed 2026-09-18.

S56 / 2024

Saffron Huang, Divya Siddarth, Liane Lovitt, Thomas I. Liao, Esin Durmus, Alex Tamkin, and Deep Ganguli. [Collective Constitutional AI: Aligning a Language Model with Public Input](#)

FACCT study connects a public-input process involving 1,002 U.S. adults to constitutional training and comparative model evaluation. DOI:10.1145/3630106.3658979.

Limit: U.S.-focused, screened participation and interpretation choices do not establish global representativeness, democratic legitimacy, or comprehensive adherence to public principles. Accessed 2026-09-18.

S57 / 2022 preprint

Leo Gao, John Schulman, and Jacob Hilton. [Scaling Laws for Reward Model Overoptimization](#)

Studies divergence between an optimized proxy reward and a separate gold-model score under reinforcement learning and best-of-n sampling.

Limit: The synthetic gold-model signal is not ground-truth human welfare; the study does not eliminate the gap between preference labels and human intent. Accessed 2026-09-18.

S58 / Repository inspected 2026

Pi maintainers. [Pi Agent Harness](#)

The canonical repository reached from badlogic/pi-mono documents configurable model/provider connections and an MIT root code license, and explicitly assigns execution isolation to the deployment.

Limit: An inspectable harness, not a guarantee of model openness, security, dependency licensing, or restricted access to files, processes, networks, and credentials. Review the exact version and deployment. Accessed 2026-09-18.

S59 / Repository inspected 2026

Oh My Pi maintainers. [can1357/oh-my-pi](#)

Identifies the project as a Pi fork, provides configurable model/provider endpoints and overrides, and includes an MIT root code license.

Limit: Does not independently establish performance, dependency safety, legal clearance, or an official relationship with VS Code. The root license does not license every connected model or service. Accessed 2026-09-18.

S60 / Documentation approved 2026-09-16**Microsoft.** [Build with agents in VS Code](#)

Official documentation distinguishes agent harnesses, model choices, and execution environments, including the separation of local tool execution from model hosting.

Limit: Product documentation, not an independent benchmark or a statement that all model processing stays local. It does not establish an official Pi-VS Code product relationship. Accessed 2026-09-18.

S61 / Preview documentation inspected 2026**Model Context Protocol Registry contributors.** [The MCP Registry](#)

Describes a preview metadata registry for open- and closed-source MCP servers, namespace authentication, and the separation of discovery from downstream code-security scanning.

Limit: Publisher authentication and registry inclusion are not security, quality, provenance, or legal-clearance certificates for a server. Accessed 2026-09-18.

S62 / 2026-07-28 specification**Model Context Protocol contributors.** [Architecture - Model Context Protocol](#)

Defines host, client, and server architecture and responsibilities for controlled tool and resource integration.

Limit: A protocol reference, not proof that a specific adapter or deployment implements all responsibilities; exact supported versions and behavior require testing. Accessed 2026-09-18.

S63 / Live documentation inspected 2026**Hugging Face.** [Model Cards](#)

Describes model documentation and metadata for intended uses, licenses, training information, limitations, and evaluation.

Limit: Declared information is not independent verification or a blanket license. Downloadable weights do not establish training-data clearance or unrestricted reuse. Accessed 2026-09-18.

S64 / Live documentation inspected 2026**Hugging Face.** [Dataset Cards](#)

Describes dataset documentation, declared licensing, contents, use considerations, and potential biases.

Limit: A card does not prove complete provenance, consent, or permission for every intended use; dataset-specific terms and evidence remain necessary. Accessed 2026-09-18.

S65 / Documentation inspected 2026**UK AI Security Institute and Inspect AI contributors.** [Inspect AI: Framework for Large Language Model Evaluations](#)

An MIT-licensed evaluation framework with datasets, solvers, scorers, and multiple model-provider integrations.

Limit: A testing framework, not a universal quality or safety certification. Conclusions depend on task selection, grading methods, and integration fidelity. Accessed 2026-09-18.

S70 / 2002

Budapest Open Access Initiative. [Budapest Open Access Initiative: original declaration](#)

Defines open access to scholarly literature, proposes self-archiving and open-access journals, and distinguishes free reader access from the costs of producing and disseminating scholarship.

Limit: A declaration, not legislation, evidence of this proposal's effectiveness, or a blanket license for all research materials. Accessed 2026-09-18.

S71 / Version 2.1; accessed 2026

Open Knowledge Foundation. [Open Definition 2.1](#)

Defines open knowledge through public-domain status or open permissions, access, machine readability and open formats; open-license requirements include non-discrimination and no royalty condition.

Limit: A definition and interoperability standard, not authority to release personal data, waive community obligations, or override third-party rights. Accessed 2026-09-18.

S72 / 2021

UNESCO. [Recommendation on Open Science](#)

Adopted 23 November 2021; frames open science across knowledge creation, evaluation and communication, including infrastructure and societal engagement, while recognizing justified restrictions for privacy, confidentiality and sacred or secret Indigenous knowledge.

Limit: A recommendation to Member States, not a universal license, directly applicable authorization, or proof that every research output should be public. Accessed 2026-09-18.

S73 / 2019

UNESCO. [Recommendation on Open Educational Resources \(OER\)](#)

Adopted 25 November 2019; defines OER as public-domain or openly licensed learning, teaching and research materials permitting no-cost access and reuse, and promotes inclusive and equitable access and sustainable support.

Limit: A recommendation, not authority over student records, an exemption from children's rights, or evidence of learning gains from this proposed pilot. Accessed 2026-09-18.

S74 / 2019; living guidance accessed 2026

Committee to Protect Journalists. [Digital safety kit](#)

Provides primary journalist-safety guidance on risk assessment, source and communication protection, accounts, devices, and reviewing third-party services' handling of sensitive information.

Limit: Practical safety guidance, not a guarantee of source confidentiality, legal privilege, or approval of the proposed journalism collective. Accessed 2026-09-18.

S80 / Undated; accessed 2026

Wikimedia Enterprise / Wikimedia Foundation. [About Wikimedia Enterprise](#)

Describes enterprise-grade APIs, data-quality signals, developer support, continued free services, and operation through Wikimedia LLC, a Foundation-owned entity.

Limit: An organizational example, not proof that separate entities are always necessary, that the proposed CCF model will succeed, or that individual contributors receive these payments. Accessed 2026-09-18.

S81 / Undated; accessed 2026

arXiv. [Membership](#)

Institutional membership supports a platform free for researchers to read and submit to, with member benefits including recognition, statistics, and governance eligibility.

Limit: A funding arrangement, not evidence of individual-contributor payments or sufficient AI-specific commercial demand for another institution. Accessed 2026-09-18.

S82 / Undated; accessed 2026

Crossref. [Fees](#)

Documents membership and registration fees, a free public metadata API, and optional Metadata Plus subscriptions for more predictable service.

Limit: Fees and programs may change. Service differentiation does not create royalties over underlying public knowledge or demonstrate payments to individual authors. Accessed 2026-09-18.

S83 / Undated; accessed 2026

Wikimedia Enterprise. [Wikimedia Enterprise/Principles](#)

Operating principles distinguish paid service from restrictions on free use, exclusive access, or customer editorial influence, and describe transparent mission-directed use of receipts.

Limit: Project-specific commitments, not universal legal rules, independently audited financial results, or endorsement of Commons Collective. Accessed 2026-09-18.

S90 / Undated living documentation; accessed 2026

Anthropic. [Data usage - Claude Code](#)

Documents consumer/commercial distinctions, improvement opt-ins, the Development Partner Program, and distinct feedback and session-quality-survey pathways.

Limit: Provider policy, not an implementation audit or proof that a particular record entered a training run. Retention does not by itself establish training use. Accessed 2026-09-18.

S91 / Undated living documentation; accessed 2026

Anthropic. [How Claude Code works](#)

Describes an agentic harness around models and tools, with context management, project instructions, auto-memory, and saved sessions.

Limit: Contextual adaptation and persistent instructions do not establish universal or immediate updates to underlying model weights. Accessed 2026-09-18.

S92 / Undated living documentation; accessed 2026

Anthropic. [Is my data used for model training? - Consumer accounts](#)

Describes consumer model-improvement choices, explicit opt-ins, feedback handling, and specified safety-review uses.

Limit: Applies to the specified products/accounts, not every commercial plan or interface. Safety-related uses are not a blanket permission for unrestricted generative-model training. Accessed 2026-09-18.

S93 / Undated living documentation; accessed 2026

Anthropic. [Is my data used for model training? - Commercial accounts](#)

States that commercial inputs and outputs are not used for model training by default, with explicit feedback or separately authorized sharing capable of creating exceptions.

Limit: Does not make every feedback mechanism identical or replace review of the actual agreement, account, authentication route, and program. Accessed 2026-09-18.

S94 / 2026 definition

Wikimedia contributors. [Research:Wikistats metrics/Active editors](#)

Defines the metric as registered, non-bot editors with at least five edits in a given month, including redirect-page edits.

Limit: Not an all-time count of unique human contributors, registered accounts, or every kind of Wikimedia participation. Accessed 2026-09-18.

S95 / 2026 definition

Wikimedia contributors. [Research:Newly registered user](#)

Specifies self-created account registration as a user-acquisition metric, with exclusions and limitations concerning attached/proxy-created accounts, alternate accounts, and later bot classification.

Limit: Registration does not prove editing, unique personhood, or a contribution to a knowledge commons. Accessed 2026-09-18.

S96 / 2026

AI Disclosures Project. [Roundtable on Mechanisms for a Sustainable AI Data Economy](#)

Describes the 18 September 2026 working session at Microsoft Research New York on mechanisms and incentives across AI pre-training, post-training, and inference, including working-group design of proposals.

Limit: Convening context, not a verified transcript, authorship record, or endorsement of this Commons Collective discussion draft by the organizers or participants. Accessed 2026-09-19.

S97 / 2026 context supplied by user

User-provided workshop photograph; card attributions preserved as printed. [Collective Bargaining & Institutions: group brief and six attributed direction cards](#)

Records the printed group brief and cards attributed to Zoe Cullen, Tomicah Tillemann, Erika Yorio, Nicholas Vincent, Beth Goldberg, and Jaron Lanier, with original-frame and crop evidence.

Limit: AI-assisted visual transcription, not verified spoken quotations, independent proof of the cards' premises, or endorsement. The event-date association is contextual, not inferred from image metadata; source rights are not relicensed. Accessed 2026-09-19.

S98 / 2018-09-26

Harvard Business Review (publisher). [A Blueprint for a Better Digital Society](#)

Publisher metadata identifies the article and publication date; its accessible introduction discusses dissatisfaction with surveillance-funded digital services and the need for an economic shift.

Limit: Bibliographic metadata and public preview only, not a full-text review. Included as related reading, not verification of every MID detail or endorsement of the proposed institution. Accessed 2026-09-19.

S99 / 2026-09-18

Session-note authors as credited in the user-supplied excerpt. [Blueprint for a Collective Bargaining Institution for Data in the Public Interest: supplied note excerpt](#)

The supplied draft raises contribution versus withholding, mixed traces, node-level incentives, market/policy/non-market/philanthropic mechanisms, sector institutions, CMO collaboration, and public quality signals. Its author line and specific attributions are recorded in the source review.

Limit: Review is of the user's pasted excerpt, not authenticated access to the complete Google Doc or other tabs. Draft premises and attributed suggestions are not adopted as verified empirical findings or approval of this publication. Accessed 2026-09-19.

[Download the machine-readable source register](#)

ETHICAL TECH COLAB

COMMONS COLLECTIVE

Evidence, interpretation, and proposals are distinguished throughout. No participant endorsement is implied.

AI-assisted research synthesis commissioned from a session brief. Human review, authorship confirmation, legal review, and empirical validation remain necessary. Names identify the requested research circle, not signatories.

Views are not official positions of NYU, Microsoft, Creative Commons, Humanity AI, or any participant or partner. Experimental research; not legal, investment, or financial advice.

Original research: [CC BY 4.0](#). Software: MIT. Workshop photographs, crops, and quoted source wording retain their respective rights and are not relicensed by this publication. No analytics, cookies, external fonts, or personal-data collection. Explicit financial-view sharing uses one-time storage in this tab; scenarios are not uploaded.

[Propose a correction or contribute evidence](#)

[AI usage by recorded model](#)