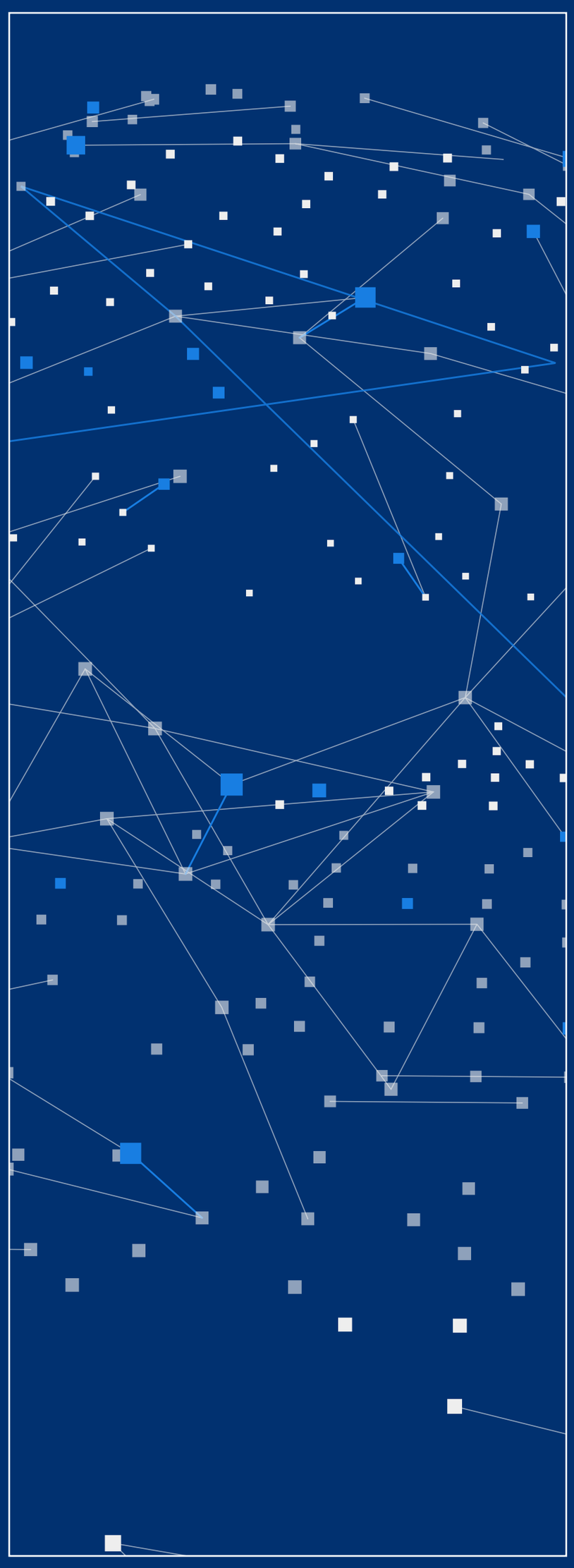
Alternatives to Concentrated AI



A Layer-by-Layer Map of What Is Being Built

Samuel Vance-Law
Kyle Bligen

OCTOBER 2026

DRC DECENTRALIZATION RESEARCH CENTER

DRC

DECENTRALIZATION
RESEARCH CENTER

About Decentralization Research Center

The Decentralization Research Center (DRC) is a 501(c)(4) social welfare non-profit that advocates for decentralization as a fundamental characteristic of emerging technologies. This includes the development of blockchain protocols and applications that are immutable, censorship resistant, transparent, secure, and enable data self-sovereignty.

The DRC's mandate is to connect stakeholders and create opportunities for innovators to collaborate, gain momentum, and guide the ownership, governance, and regulation of emerging technologies towards decentralization.

To create a more equitable future, we must ensure that the next wave of innovation is underpinned by principles of broad participation, fair reward distribution, and decentralized control. The DRC stands committed to working with industry, academia, and policymakers to help shape this future.

DISCLAIMER:

This report is based on a series of interviews that took place between April and August 2026. It re-evaluates what artificial intelligence is and whom it serves, and walks the AI stack layer by layer, pairing each layer with working alternatives to concentration, with the ultimate objective of fostering AI that enables human flourishing. While every effort has been made to faithfully reflect all input received, it should be noted that the views expressed herein are those of the Decentralization Research Center, with the exception of the three case studies in Section 3, which are authored by their named practitioners. The views expressed may not necessarily align with the individual positions of the expert contributors listed below or the organizations with which they are affiliated.

Please cite as: Vance-Law, Samuel, and Kyle Bligen. Alternatives to Concentrated AI: A Layer-by-Layer Map of What Is Being Built. Decentralization Research Center, October 2026.

Contributors

This report draws on interviews conducted between April and August 2026. Affiliations are given for identification only; all interviewees spoke in a personal capacity, and neither participation nor listing here implies endorsement of the report's findings by any named individual or organization.

Authors

Samuel Vance-Law
Head of research, Decentralization Research Center

Kyle Bligen
Executive director, Decentralization Research Center

Case-study authors

Andrew Trask
Executive director, OpenMined

Lacey Strahm
Vice president of growth and impact, OpenMined

Davide Crapis
AI lead, dAI Team, Ethereum Foundation

Gabriel Fior
dAI Team, Ethereum Foundation

Marcello Politi
dAI Team, Ethereum Foundation

Vittorio Rivabella
dAI Team, Ethereum Foundation

Audrey Tang
Civic AI

Tenzin Yangtso
Civic AI

Caroline Emmer De Albuquerque Green
Civic AI

Interviewees

Steven Adler
Co-founder and chief scientist, Guidelight AI Standards; formerly OpenAI

Sameer Bajaj
Research analyst, Aapti Institute

Tomer Jordi Chaffer
BCL/JD candidate, Faculty of Law, McGill University; associate fellow, Institute for Technoscience and Society, York University

Interviewees

Tsosheletso Chidi
Lecturer, Department of African Languages, and research fellow, Department of Computer Science, University of Pretoria

Ben Fielding
Co-founder and chief executive officer, Gensyn

Caroline Emmer De Albuquerque Green
Director of research and head of public engagement, Institute for Ethics in AI, University of Oxford; senior research fellow, Jesus College, Oxford; Civic AI

Botao "Amber" Hu
DPhil researcher, Human Centred Computing group, Department of Computer Science, University of Oxford; director, Reality Design Lab

Ron Ivey
Founder, Noēsis Collaborative; research fellow, Human Flourishing Program, Harvard University

Astha Kapoor
Co-founder and director, Aapti Institute

Sébastien Krier
Frontier policy development lead, Google DeepMind

H. Brendan McMahan
Principal research scientist, Google

Evan Miyazono
Founder and chief executive officer, Atlas Ignota

Srikanth Nadhamuni
Founding chief technology officer, Aadhaar (India); co-founder and managing trustee, eGov Foundation

Vinay Narayan
Senior manager, Aapti Institute

Sarah Nicole
Lecturer, Sciences Po

Puja Ohlhaver
GETTING-Plurality Research Network, Allen Lab for Democracy Renovation, Harvard University

Illia Polosukhin
Co-founder, NEAR Protocol

Rithikha Rajamohan
Standards and protocol, EQTY Lab; founder and director, V6A

Vittorio Rivabella
AI coordinator, dAI Team, Ethereum Foundation

Helena Rong
Assistant professor of interactive media and business, NYU Shanghai; director, Protopolis Lab

Nelson Rosario
Founder and partner, Rosario Tech Law; adviser, Advanced AI Society; appointee, Illinois State Bar Association AI Steering Committee

Anja von Rosenstiel
M.A., attorney-at-law (MA), Rechtsanwältin, lecturer of law at Boston University, senior counsel at Ernst & Young Law, Austria

Nathan Schneider
Associate professor of media studies, University of Colorado Boulder; director, Media Economies Design Lab

Derek Slater
Founding partner, Proteus Strategies

Noah A. Smith
Vice provost for artificial intelligence and Simonyi Endowed Chair for Artificial Intelligence, University of Washington; professor, Paul G. Allen School of Computer Science and Engineering; senior director of NLP research, Allen Institute for AI (Ai2)

Julie Stitzel
Chief policy officer, Digital Currency Group

Lacey Strahm
Vice president of growth and impact, OpenMined

Interviewees

Joshua Tan

Co-founder and research director, Metagov; product lead, Public AI

Audrey Tang

Cyber ambassador-at-large, Taiwan; Taiwan's first digital minister, 2016 to 2024; senior accelerator fellow, Institute for Ethics in AI, University of Oxford; Civic AI

Andrew Trask

Executive director, OpenMined

Stefaan Verhulst

Co-founder and chief research and development officer, The GovLab, New York University; research professor, NYU Tandon School of Engineering; co-founder, The Data Tank

Tricia Wang

Co-founder and chief executive officer, Advanced AI Society

Glen Weyl

Founder and research lead, Plural Technology Collaboratory, Microsoft Research; founder, RadicalxChange Foundation; founder and chair, Plurality Institute

Riley Wong

Principal, Emergent Research; co-founder, Community Privacy Residency; graduate student, Department of Urban Studies and Planning, Massachusetts Institute of Technology

Janabel Xia

PhD student in mathematics, Harvard University; co-founder, Community Privacy Residency

Tenzin Yangtso

Co-author, the 6-Pack of Care, Civic AI

With particular thanks to **Tony Douglas Jr.**, **Shaan Jain**, and **Adrian Li**.

Foreword

AI has concentrated, first and foremost, at the level of narrative — the singular AGI, developed and owned by a single company, responsible for our salvation or destruction, beholden to its own unquestionable logic. That is neither the truth on the ground nor an accurate description of what AI is, how it should be thought of, or how it should be built or governed. Frontier-model LLMs are an extraordinary feat. They are, however, a composition of different parts of a technology, most sections of which have open, decentralized, federated, and distributed versions. And each of which has a different narrative around what this technology is and for whom it should be built. The account this report presents is necessarily incomplete; the alternative vision is one of a plurality of models, systems, and governance structures that suit a plurality of people, communities, and needs. However, it is exactly in this plurality that coordination and competition can emerge, communities and individuals can be best served, and many of the ills we ascribe to AI can be rethought or ameliorated. This report is an introduction to some of those models and to that other narrative, one focused less on what this technology might be capable of and more on who it should be for.

Samuel Vance-Law

Head of Research,
Decentralization Research Center

Executive summary

This report presents a map of alternatives to concentrated AI, drawing on thirty-six interviews with the people building and studying them, three case studies written by their builders (OpenMined, the Ethereum Foundation, and Civic AI), and nearly thirty working systems, institutions, and standards. It is written for policymakers, funders, and builders, and sheds light on ways of developing and using AI designed to help people live fuller, freer, and more capable lives.

The report presents four findings:

1. Concentration need not be the endpoint of development. General-purpose technologies have passed through stages of concentration and decentralization before, and finance and public policy can lead to alternatives as well as to concentration.

2. The alternatives do not depend on matching the frontier. Their strength is in addressing particular challenges in the stack: diverse data that behaves like high-quality data, federated learning that reaches data no central model can, and models built for a particular language or place that match far larger ones on the tasks their communities need.

3. Chip manufacturing and frontier-scale training remain concentrated. Everywhere else, working systems exist: at the data layer the institutions are already built, at the model layer openness reaches data, code, and the training process itself, and privacy-preserving infrastructure has begun to let regulators audit a system without holding its data.

4. Most of the alternative sits outside the safety debate that concerns a few frontier systems. Bounded, local systems do not carry the same catastrophic-risk profile, and where there are concerns, at open frontier weights, at agents, and inside community systems, the alternatives supply part of the answer: the identity infrastructure that agentic AI needs and external verification.

The report does not treat concentration as a moral failure, does not argue that every system must be decentralized, and is not a case for open frontier weights. Its claim is that concentration has not been the final stage of earlier general-purpose technologies, and that the alternatives that would make this one plural exist now. Policy recommendations will follow in a separate report in 2027.

Table of contents

Contributors 02
Foreword 05
Executive summary 06

01. What Artificial Intelligence Is For 08
A. Toward Human Flourishing 08
B. Economic Engines and Incentives 09
Key Insights 09

02. The Alternatives and What They Offer 10
A. A Plurality of Traditions 10
B. Capability Without Concentration 11
C. The Alternative, Layer by Layer 12
Hardware 12
Compute 13
Data 14
Models 16
Inference and Access 17
Agents 18
Governance 19
Key Insights 20

03. Case Studies 21
A. Case Study: OpenMined 21
B. Case Study: ERC-8004 and the Open Trust Layer for AI Agents 25
C. Case Study: Civic AI 30

04. Where Safety Comes In 36
Key Insights 37

05. Conclusion 38
References 39
Primary documents and official sources 39
Published works 41

01

What Artificial Intelligence Is For

A. Toward Human Flourishing

Current accounts of artificial intelligence often assume that AI is the point: how smart it will get, how dangerous it will be, how many people's jobs will be replaced by it. We start this report from a different premise. AI is an instrument designed by and for people, and the test of any AI system should be whether it helps people live fuller, freer, and more capable lives.

This is a sentiment repeated across frontier AI labs. Sam Altman has said that "the people impacted by technology the most deserve the loudest say in its governance," and Demis Hassabis that AI's benefits must "accrue to as many people as possible, to all of humanity, ideally."[1] The United States government has made the same claim for its own strategy. America's AI Action Plan promises "a new golden age of human flourishing, economic competitiveness, and national security."[2] Researchers inside the same laboratories have gone further. In 2026 a group of authors at Google DeepMind, OpenAI, Anthropic, Oxford, and other institutions argued that alignment aimed at human flourishing rather than at avoiding harm cannot be a top-down exercise.[3]

The alternatives share that view. They also reach for human flourishing as the goal: AI that augments human growth and expression, reflects the full diversity of the people whose data it learns from, and keeps intelligence and benefit with those people. And they maintain that it comes only by distributing the governance and benefit of the technology.[4] Where their centralized counterparts promise that kind of control after the fact, the alternatives build it in: distributed so that no one can monopolize, local so that a community can govern, and bounded so that when AI models fail, they fail small.[5]

For now, though, the incentives run toward concentration for centralized and decentralized builders alike. Almost everyone agrees the technology should serve the public, but how to incentivize broader governance and benefit remains an open question. The answer may lie more in the economic history of general-purpose technologies than in any particular attribute of AI itself.[6]

1. Sam Altman, remarks at the Xfund Experiment Cup ceremony, Harvard University, 1 May 2024; Demis Hassabis, interview with TIME, January 2023.

2. Winning the Race: America's AI Action Plan (The White House, 23 July 2025).

3. Ruben Laukkonen, Seb Krier, Chloé Bakalar, Shamil Chandaria, Morten Kringelbach, et al., "Positive Alignment: Artificial Intelligence for Human Flourishing," arXiv:2605.10310v3 (19 June 2026; preprint). The definition of flourishing is VanderWeele's (note 4). Krier, a co-author, is an interviewee for this report.

4. Ron Ivey, DRC interview, 2026. See Ronald Ivey et al., "Human Flourishing: A Guiding Principle for AI Systems Design?" (presentation, Pontifical Academy of Sciences, 1 May 2024; SSRN 5185045), which defines flourishing as realizing one's own good and the common good by fulfilling human capabilities and places technology among its facilitators. The capabilities lineage is Amartya Sen, Development as Freedom (Oxford University Press, 1999), and Martha C. Nussbaum, Creating Capabilities (Belknap Press of Harvard University Press, 2011); the empirical construct is Tyler J. VanderWeele, "On the Promotion of Human Flourishing," PNAS 114, no. 31 (2017).

5. Anja von Rosenstiel, DRC interview, 2026: in decentralized finance, constraints are built into the protocol as circuit breakers, whereas AI builders reach for guardrails added after the fact.

6. The idea that AI should act for and with people also has historical precedent. The intellectual lineage of computing is a story of human augmentation, not replacement. Norbert Wiener described intelligence as a property of feedback between an agent and its world, not a black box; J.C.R. Licklider imagined a symbiosis that extended human capability across a network of peers; Douglas Engelbart built tools to make groups of people collectively smarter. The technologies that built the modern world pushed capability toward their users. With that lineage in mind, the consolidation of artificial intelligence into a handful of vertically integrated firms represents something of a departure.

B. Economic Engines and Incentives

In its economics, artificial intelligence shares much with the general-purpose technologies that came before it: the railroad, the electric grid, the telephone network.[7] Each demanded enormous fixed costs and coordinated standards before it could generate any returns, and in each case private capital required public backing. The transcontinental railroad was built only after the Pacific Railway Acts of 1862 and 1864 granted the companies tens of millions of acres of public land and federal bonds for every mile of track, because private corporations judged the undertaking too large and too risky to attempt unaided.[8]

But many of these technologies began plural. The telephone had thousands of independent carriers after Bell's patents expired in 1894, and the Bell monopoly was assembled by absorbing them.[9] The Bell System became a single national telephone network after a 1913 settlement in which AT&T averted antitrust action by agreeing to interconnect independent carriers and stop absorbing them, and the government thereafter treated the monopoly as a regulated utility.[10] Concentration of this kind is not necessarily the enemy of a general-purpose technology's spread. It can often be part of its cycle of development. Frontier AI, with its extreme training costs and its appetite for coordinated capital, looks like the newest instance.[11]

Each of these infrastructures was subsequently made plural, or made to reach everyone, by deliberate institutional choice. The telephone monopoly was broken into regional companies by the antitrust divestiture of 1984.[12] Rural America was electrified in the 1930s by member-owned cooperatives that existed because federal loans and a dedicated agency made them possible, a model that still carries power across more than half the nation's landmass and the large majority of its persistent-poverty counties.[13] Broad land and farm ownership was engineered the same way, through the Homestead Acts and a farm-credit system built deliberately to finance distributed ownership, and at the cost of Indigenous dispossession.[14] The economist Carlota Perez explains these two phases as the installation period, when financial capital concentrates and speculates, and the deployment period, when the technology is turned to broad social benefit.[15]

Neither concentration nor distribution is natural to a technology. Both follow from the incentives and institutions built around it.[16] AI companies and governments are now grappling with exactly where in this cycle AI should be.

Key insights

- AI is an instrument. The right question about any system is whom it serves.
- Human flourishing is the stated goal from AI labs, researchers, and the alternatives alike; the difference is whether control is built in or promised later.
- Earlier infrastructures were concentrated as much by finance and state settlement as by their technology, and were made plural the same way — divestiture, cooperatives, deliberate financing.

7. On general-purpose technologies, Timothy F. Bresnahan and Manuel Trajtenberg, "General Purpose Technologies 'Engines of Growth'?," Journal of Econometrics 65 (1995): 83–108. On AI as the latest GPT, Erik Brynjolfsson, Daniel Rock, and Chad Syverson, "Artificial Intelligence and the Modern Productivity Paradox" (NBER Working Paper 24001, 2017) and "The Productivity J-Curve," American Economic Journal: Macroeconomics (2021).

8. Pacific Railway Acts of 1862 and 1864. National Archives, "Pacific Railway Act (1862)" (the 1862 Act "provided Federal subsidies in land and loans," and private corporations were "reluctant to undertake the task without Federal assistance"); U.S. Senate Historical Office, "The Pacific Railway Act of 1862."

9. Tim Wu, The Master Switch: The Rise and Fall of Information Empires (Knopf, 2010); Nelson Rosario (DRC interview, 2026) points to the many regional railroads of the nineteenth century and the century it took them to consolidate.

10. The Kingsbury Commitment (1913): AT&T's letter to the Attorney General undertaking to divest its Western Union holding, to allow independent carriers to interconnect with its long-distance lines, and to stop acquiring competing independents without government approval, in return for the Justice Department dropping its antitrust suit. AT&T's share rose to roughly four-fifths of US telephones by the 1930s.

11. Meghan Bobrowsky, Drew An-Pham, and Alana Pipe, "Big Tech's AI Push Is Costing a Lot More Than the Moon Landing," Wall Street Journal, 7 February 2026: Alphabet, Amazon, Meta, and Microsoft planned a combined $670 billion in AI capital spending for 2026, roughly 2.1 percent of US GDP, a larger share of the economy than the Apollo program.

12. United States v. AT&T (settled 1982; the Modification of Final Judgment); the 1984 divestiture split the Bell System into seven regional operating companies ("Baby Bells").

13. Rural Electrification Administration (established 1935) and the rural electric cooperatives it financed. Sarah Nicole, Samuel Vance-Law, Connor Spelliscy, and James Bell, How Can Data Cooperatives Help Build a Fair Data Economy? (Project Liberty Institute and Decentralization Research Center, July 2025). Member-owned utilities today serve 42 million people in 48 states across 56 percent of the nation's landmass, including 92 percent of persistent-poverty counties (National Rural Electric Cooperative Association, comments to the Federal Communications Commission, 4 August 2026).

14. Homestead Act of 1862, 12 Stat. 392; Federal Farm Loan Act of 1916, 39 Stat. 360, which created the federal land banks. On the Homestead Acts' redistribution of land taken from Native nations, see the National Park Service, Homestead National Historical Park, "Native Americans and the Homestead Act." Nathan Schneider, DRC interview, 2026, on the farm-credit system as the financing infrastructure behind contemporary agricultural cooperatives.

15. Carlota Perez, Technological Revolutions and Financial Capital: The Dynamics of Bubbles and Golden Ages (Edward Elgar, 2002); Perez explicitly warns against technological determinism and stresses society's role in shaping a technology's direction and diffusion.

16. Nathan Schneider, DRC interview, 2026: concentration is "a chosen outcome," not an inevitable one, because the capital sources for distributed ownership do not exist by default and would have to be built. His Exit to Community work makes the related structural claim that the concentration of financial markets is shaped by policy design rather than by investors' choices alone: Júlia Martins Rodrigues and Nathan Schneider, "Exit to Community (E2C): Financing a New Destination for Successful Ventures" (forthcoming; advance manuscript, 2026); the earlier articulation is Morshed Mannan and Nathan Schneider, "Exit to Community: Strategies for Multi-Stakeholder Ownership in the Platform Economy," Georgetown Law Technology Review 5 (2021). Steven Adler (DRC interview, 2026) reaches the same conclusion from the safety side: larger training runs tend to produce stronger systems, so actors merge and consolidate for relative advantage unless law, tax, or financing offsets it. Janabel Xia (DRC interview, 2026) restates the incentive point independently: no one can be blamed for following the incentives, so the incentives are what policy has to change.

02

The Alternatives and What They Offer

A. A Plurality of Traditions

Alternatives to concentration come from many traditions and come with many different names. The free-software movement brings the transparency of open code, weights, and data; distributed-systems engineering, architectures with no single point of control; the cryptographic tradition, coordination without a trusted operator. Machine learning adds federation, in which a model learns from data that never leaves where it lives. Catholic social thought and its secular heirs supply subsidiarity, the principle that decisions belong at the most local competent level, and the plurality tradition designs for many overlapping centers of legitimate power. The Southern African tradition of Ubuntu, in which a person is a person through other persons, arrives at the same ontology from a separate lineage entirely. A system can be open without being decentralized, or federated without being plural, but these traditions share a purpose: each keeps intelligence, control, or benefit closer to the people it is for.[17]

These traditions also align more easily with the deployment period of a general-purpose technology described above. They are designed to be local and communal, for control to be dispersed and risk to be spread. They are designed to fail small or take on a community's contextually specific challenge, to grow through networks rather than the consolidation of power or market share. And because they do not tend toward monopoly, exit is always possible. Because they are local, they can be responsive to community needs. That is also why most of them sit outside the safety debate that surrounds frontier systems, which we take up in Section 4. Many of the examples mentioned below neither can nor hope to provide what frontier LLMs currently do. But they do offer structures and capabilities that far more readily allow for distributed human governance, control, and benefit.

LAYER	WHERE IT CONCENTRATES	ALTERNATIVE ANSWERS
Hardware	Fabrication, lithography, and advanced packaging	Open instruction sets and accelerator designs; portability at the compiler layer (design, not manufacturing)
Compute	Capital intensity; scarce accelerators; energy and siting	Distributed and decentralized training, federated learning, and public compute
Data	Access and governance of a relational good	Community licenses, data stewardship, and cooperatives
Models	Closed weights; capability held by few	Open weights, fully open construction, and models for underserved languages
Inference / access	Who may study, audit, or build on systems	Privacy-preserving access; cooperative and public provision
Agents	Identity, accountability, coordination	Verifiable identity; open standards
Governance	Who sets values and rules, for whom	Subsidiary and plural governance

In practice the layers blur, and a project can move across several of them as it grows.[18]

17. Julie Stitzel (DRC interview, 2026) draws the same line in her conversations with congressional staff, between decentralization that is structural, built into a system's incentive design and governance, and decentralization that is cosmetic, where the code is visible but control stays central.

18. Ben Fielding, DRC interview, 2026, on the continuity between layers and the hazard of bucketing actors; corroborated by Botao Hu, DRC interview, 2026, who frames the alternative as an assemblage without a fixed order.

02

B. Capability Without Concentration

While alternatives cannot mimic frontier LLMs, there are places where they can offer capability in different ways. Puja Ohlhaver notes that proponents of concentration treat decentralization as a tax paid on capability for the sake of fairness. However, diversity of human signal is what the technology depends on, so plural, diverse, and distributed structures are a precondition for good AI.[19] A growing body of work finds that diversity in training data improves downstream performance, and that diverse data behaves like high-quality data.[20] Concentration, meanwhile, erodes its own raw material. Ohlhaver argues that scale and surveillance flatten the human data these systems are built from,[21] and models trained heavily on machine-generated text degrade.[22]

Another advantage is reach. Brendan McMahan, who pioneered federated learning, points to medical records and the institutional and sovereign data held in silos, which privacy, proprietary, and data-sovereignty rules require be kept in place. Federation, training across data that stays where it lives, can reach it in a way that collecting and hosting data centrally cannot.[23] If the returns from freely available data are diminishing while the data that remains cannot be moved, the distributed path becomes, in his words, a situation that "forces you to consider alternatives."[24]

The alternatives are also often a better fit for specific applications. Models built for a particular language or context can match or beat far larger systems on the tasks their communities need, at a fraction of the cost. Lelapa AI's InkubaLM, a compact model for five African languages, reaches performance comparable to systems many times its size trained on far more data. Lelapa is a centralized company, but its model is tied to a language and a place.[25] Noah Smith, who helps lead one of the few fully open model efforts, argues that a model's strength depends on being selective about its data as much as on size, and tomorrow's small models may outperform today's large ones. Treating the build-it-enormous-then-distill pathway as the only road is, he cautions, an assumption the field mistakes for a law; people "lapse into thinking it's inevitable."[26]

Composition offers a further advantage. DeepSeek-R1, an open-weight reasoning model, reached frontier-grade performance in early 2025 for a reported fraction of the cost the scaling story insisted was necessary, and its distilled versions run on a small share of a hyperscaler's hardware.[27] Through 2026, routing and fusion systems have gone a step further, orchestrating several smaller or mid-tier models behind one interface and reporting frontier-comparable quality at a fraction of the price.[28] Those figures are vendor-reported and not yet independently replicated, and where the models behind the interface are themselves frontier models, the router is a proxy for concentration rather than an alternative to it. Still, the direction is clear enough: capability increasingly depends on how models are composed, selected, and routed as well as on how large the largest one is.

None of this replaces the frontier.[29] It shows that the frontier is one route to capability among several, and that the others run through the reach, diversity, and fit of the data rather than the size of the model.

19. Puja Ohlhaver, DRC interview, 2026: the field "treat[s] decentralization like it's a tax that you pay on capability … for the sake of fairness," whereas "plural, diverse, and distributed structures … [aren't] the cost of good AI, it's actually the precondition for it."

20. Brando Miranda, Alycia Lee, et al., "Beyond Scale: The Diversity Coefficient as a Data Quality Metric for Variability in Natural Language Data," arXiv:2306.13840 (2023). Related quality-over-quantity findings come from Microsoft Research's small-model ("phi") work: Suriya Gunasekar et al., "Textbooks Are All You Need," arXiv:2306.11644 (2023), and Marah Abdin et al., "Phi-3 Technical Report: A Highly Capable Language Model Locally on Your Phone," arXiv:2404.14219 (2024).

21. Puja Ohlhaver, DRC interview, 2026: information is reflexive, so people watched at scale come to orient to the watcher and to resemble one another, and the signal a model trains on degrades with them.

22. Ilia Shumailov et al., "AI models collapse when trained on recursively generated data," Nature 631 (2024): 755–759. Collapse is mitigated when synthetic data accumulates alongside, rather than replaces, human data: Matthias Gerstgrasser et al., "Is Model Collapse Inevitable? Breaking the Curse of Recursion by Accumulating Real and Synthetic Data," arXiv:2404.01413 (2024).

23. Andrew Trask, DRC interview, 2026, on federation's reach to data that cannot be moved.

24. Brendan McMahan, DRC interview, 2026, speaking in a personal capacity.

25. Lelapa AI's InkubaLM, a 0.4-billion-parameter small language model for five African languages (isiZulu, Yoruba, Hausa, Swahili, isiXhosa), reported to reach performance comparable to models with significantly larger parameter counts and more extensive training data on tasks such as machine translation and question answering: Atnafu Lambebo Tonja et al., "InkubaLM: A Small Language Model for Low-Resource African Languages," arXiv:2408.17024 (2024).

26. Noah A. Smith, DRC interview, 2026. Smith, who helps lead Ai2's fully open OLMo effort, argues more selective, thoughtful data work can yield comparable capability at far lower cost and that tomorrow's small models may outperform today's large ones.

27. DeepSeek-R1 (DeepSeek-AI, arXiv:2501.12948, January 2025), an open-weight reasoning model released with weights and method, widely reported to have matched leading closed reasoning models at a training cost on the order of a few million dollars and distilled to variants that run on consumer hardware. The headline figure refers to the compute cost of the final training run, not total development cost (the widely cited figure is for the DeepSeek-V3 base model's final run, on which R1 was built), and is contested; it is cited here for order of magnitude and for the public perception it created, not as a verified total.

28. The underlying technique is established in the research literature, where model routing and ensembling can outperform any single constituent model on cost-adjusted quality: Isaac Ong et al., "RouteLLM: Learning to Route LLMs with Preference Data," arXiv:2406.18665 (2024).

29. Derek Slater, DRC interview, 2026, whose analogy is that Linux was a functional substitute for Windows while a model built for one language or task does not substitute for a frontier system.

02

C. The Alternative, Layer by Layer

Alternatives have different strengths across the stack, and the outline below shows where they exist today and where bottlenecks remain.

Hardware

Manufacturing is, for now, unavoidably centralized. Leading-edge logic is made almost entirely by one company, TSMC, which takes roughly seven tenths of the world's foundry revenue and holds the large majority of leading-edge capacity, nearly all of it on the island of Taiwan. Every chip at those nodes is patterned on extreme-ultraviolet lithography machines that exactly one company, ASML of Veldhoven in the Netherlands, knows how to build, and whose next generation sells for around $380 million apiece. The packaging step that binds a processor to its memory is also highly centralized. A modern AI processor is an assembly with its own scarce capacity. In 2025 the four largest AI chip designers bought close to nine tenths of that packaging capacity and of the memory, though their chips amount to only about an eighth of the world's advanced logic production. A fab costs tens of billions and takes years. Concentration is currently entrenched here.[30]

Today the alternative exists at the design and instruction-set level: an open instruction-set architecture (RISC-V, an open standard for the basic vocabulary a chip understands), accelerators whose programming stack is published rather than licensed, compilers that let a model run on more than one vendor's silicon. That work has drawn attention, and, in 2026, it drew incumbent money, with Qualcomm acquiring a hardware-agnostic inference compiler outright and holding reported talks over the most prominent open-architecture chip designer. It does not distribute manufacturing, however. An open chip design is fabricated on the same tools in the same fabs as a closed one.[31]

Even where another part of the stack is highly decentralized, the chip layer is where trust tends to migrate. Describing the trusted execution environments his own team relies on, McMahan says one still has to decide how far to trust Nvidia, Intel, or AMD.[32]

30. Foundry and accelerator market-share figures vary by methodology and are given here as approximations: TrendForce, foundry industry research, June 2026, places TSMC at roughly 72 percent of global foundry revenue in the first quarter of 2026. ASML High-NA pricing: approximately $380 million (€350 million) per system, as ASML told Reuters in February 2024; 2026 reporting on the EXE:5200 puts installed cost near $400 million. Advanced packaging and high-bandwidth memory concentration: Venkat Somala, "Advanced packaging and HBM, not logic dies, were the bottlenecks on AI chip production in 2025," Epoch AI, March 2026.

31. Open silicon: the RISC-V instruction-set architecture and accelerators built on it, notably Tenstorrent's Blackhole (Apache-licensed programming stack), and the Vortex open GPU project. Incumbent consolidation of the portability layer: Qualcomm acquired Modular, whose MAX runtime runs models across several vendors' accelerators, in an all-stock deal valued at about $3.9 billion (announced 24 June 2026, completed 29 July 2026), and was reported to be in advanced talks to acquire Tenstorrent (The Information via Reuters, 15 June 2026), unconfirmed by either party as of this writing.

32. Brendan McMahan, DRC interview, 2026, personal capacity, on trusted execution environments as a hardware root of trust: "you decide how much you wanna trust Nvidia or Intel or AMD." Rithikha Rajamohan (DRC interview, 2026) makes the same observation from the verification side: to trust the hardware is to trust Nvidia or Intel.

02

Compute

The alternatives are stronger at the compute layer. Training very large models across a number of data centers, perhaps tens of sites, is possible in McMahan's judgment, though each site still needs extreme compute. Permissionless pre-training over the open internet has advanced quickly, with the largest documented runs, from Prime Intellect, Pluralis, and Nous Research, reaching the tens of billions of parameters. These remain, however, on the order of a thousandfold below frontier compute.[33]

The alternatives are more suited to reinforcement learning and other post-training that generate their signal asynchronously and tolerate loose coordination. Prime Intellect's INTELLECT-2, a 32-billion-parameter reasoning model trained by distributed reinforcement learning across a permissionless swarm of hundreds of untrusted machines, is a clear demonstration that at that scale the distributed version competes within its class. Permissionless frontier-scale pre-training remains unsolved, however, including how to finance it.[34]

Federated learning has also made great strides at this layer. Flower, an open-source framework for federated learning, is used by researchers and companies to train across silos that could never be pooled, so a model learns from data that never leaves the hospital, phone, or firm that holds it.[35]

Public compute offers another alternative. The National AI Research Resource pilot has supported hundreds of research projects since 2024 and is moving from pilot to a standing operations center. New York's Empire AI consortium, in which the state's research universities and the Simons Foundation's Flatiron Institute share more than $500 million in public and private funding, brought its second machine fully online in August 2026, and other states are building their own. Derek Slater's reading of why it worked is that it was built for the researchers who use it. The state, the universities, and philanthropy paid for it, and the institutions that use it run it.[36]

California authorized a public cluster of its own in 2025, and the European Union is making comparable investments in shared AI factories, pooled accelerators offered on metered access with guaranteed data residency. Each is built to offer compute a research team can reach without a contract with a cloud provider.[37]

33. Brendan McMahan, DRC interview, 2026, personal capacity: distributed training "might offer alternatives to extreme centralization" while still needing "maybe tens of sites" and an extreme amount of compute at each. The largest documented decentralized pre-training runs (INTELLECT-1, Pluralis's Protocol Model, Nous Research's Consilience) sit on the order of a thousandfold below frontier compute: Epoch AI, "How far can decentralized training over the internet scale?," 29 December 2025.

34. Prime Intellect's INTELLECT-2, trained via globally distributed asynchronous reinforcement learning across a permissionless swarm of 800+ heterogeneous nodes and improving on QwQ-32B in its class: arXiv:2505.07291 (2025). The structural point that reinforcement learning is better suited to decentralization than pre-training is argued by Prime Intellect and is consistent with the Epoch analysis cited in the preceding note. The point that pre-training resists distribution while post-training does not, and that what blocks distributed pre-training is the formation of capital rather than the mathematics, is Ben Fielding's (DRC interview, 2026).

35. Daniel J. Beutel, Taner Topal, Akhil Mathur, et al., "Flower: A Friendly Federated Learning Research Framework," arXiv:2007.14390 (2020); flower.ai.

36. Derek Slater, DRC interview, 2026. NAIRR: led by the NSF with thirteen other federal agencies and twenty-nine private-sector and nonprofit partners; more than 900 supported projects across all fifty states since 2024, per the pilot's public reporting as of September 2026; the standing Operations Center is NSF solicitation 25-546 (September 2025); statutory authorization is the CREATE AI Act of 2025, H.R. 2385, reported out of committee June 2026. Empire AI at launch: a $250 million state capital grant, $25 million from SUNY over ten years, and more than $125 million from the founding institutions and private partners (Office of the Governor of New York, 22 April 2024); Empire State Development and Office of the Governor releases, February and June 2025; the Beta system came fully online on 5 August 2026, by which point the consortium was backed by more than $500 million in public and private funding (Office of the Governor of New York, 5 August 2026). States as a class: Science, "To boost research, states are building their own AI-ready supercomputers," December 2025.

37. California's SB 53, signed 29 September 2025, establishes a fourteen-member CalCompute consortium within the Government Operations Agency and requires a framework report to the Legislature by 1 January 2027; it appropriates no funds, and as of September 2026 the consortium was still being formed. EU: AI Factories / InvestAI, February 2025.

02

Data

Riley Wong, a privacy researcher and co-founder of the Community Privacy Residency, names the problem the alternatives address: while consent frameworks are built around the individual, data is always plural. For instance, a genetic profile uploaded to an ancestry service implicates relatives who never consented to their data being used, and a community's health data is valuable only in aggregate.[38] Various alternative models have been introduced so that benefit stays with the community that produced the data: cooperatives, stewards, and trusts that govern access on behalf of everyone a dataset implicates.

Which institution is right depends on context. Astha Kapoor, co-founder of the Aapti Institute, which works with communities across the Global South, suggests organizing around the end a community is seeking and letting the organizational and legal form follow. The institute's researchers also insist that collective governance becomes credible only when the partners a community depends on, the people who clean, label, and steward the data, are also accountable to community governance because that is where practical power over data accrues.[39]

The cooperative form has also entered this space at commercial scale. Land O'Lakes, the American agricultural cooperative, partnered with Microsoft to build AI tools on its own agronomic data. Nathan Schneider points to what made that possible: Land O'Lakes could negotiate the contract because a century of cooperative finance had built an institution large enough to bargain, the farm-credit story of Section 1.[40]

A data commons, which Stefaan Verhulst treats as a governance vehicle rather than a repository, makes data available under a social license, terms aligned with the collective preferences of the communities that steward it.[41] Tsosheletso Chidi comes at the commons question from the context of language. Language can never be cleanly owned, so Chidi sets ownership aside and asks "how can we collectively steward linguistic resources in ways that support innovation while protecting the communities?"[42] Context-differentiated licenses built for this purpose let data stay available for innovation while the communities that produced it keep a real say over its use. The Nwulite Obodo Open Data License is one example of this.[43] DSFSI, the Data Science for Social Impact research group at the University of Pretoria, builds the institutional version with African-language datasets, models, and the licenses to govern them.[44]

Another precedent is Te Hiku Media, a Māori broadcaster that built speech technology for te reo Māori from data gathered by community consent and released it under a Kaitiakitanga ("guardianship") license. Under it the data is never owned, only held in guardianship by those who contributed it, and access requires Te Hiku's permission and excludes commercial use unless granted. Te Hiku's own principle, stated alongside the license, is that any benefit flows back to the community the data came from.[45]

Concentration has complicated efforts to share data, however. As open datasets, public web pages, and open-government portals are absorbed into training corpora with little shared benefit, the response has been enclosure. Firms have closed the interfaces that once let researchers study public conversation, some open-science and open-government data has grown harder to obtain, and public pages are increasingly walled against scraping. Verhulst calls the result a data winter, a decline in the data available for public-interest reuse arriving in the same years as an AI summer, and reads it as a weaponization of openness now reaching the data commons.[46]

38. Riley Wong, DRC interview, 2026. The relational framing draws on Salomé Viljoen, "A Relational Theory of Data Governance," Yale Law Journal 131 (2021), which Wong cites, since extended into a proposal for associational data rights: Daniel Cochrane, Matt Prewitt, Lacey Strahm, and Salomé Viljoen, Toward Associational Data Rights (Institute for Family Studies, August 2026). Strahm co-authored the OpenMined case study in Section 3. Julie Stitzel (DRC interview, 2026) makes the same point from the policy side, citing the DRC's data-cooperatives report (note 13): individual consent does not scale, so the answer is collective governance rather than more centralization.

39. Aapti Institute (Astha Kapoor, Sameer Bajaj, Vinay Narayan), DRC interview, 2026.

40. Nathan Schneider, DRC interview, 2026; Land O'Lakes Inc. and Microsoft, multiyear strategic alliance announcement, 12 November 2025.

41. Stefaan Verhulst, DRC interview, 2026.

42. Tsosheletso Chidi, DRC interview, 2026.

43. The Nwulite Obodo Open Data License, developed by Chijioke Okorie and Melissa Omino: "Addressing Inequitable Openness in Licences for Sharing African Data and Datasets Through the Nwulite Obodo Open Data Licence," Law, Technology and Humans 7, no. 3 (2025): 94; adopted by African-NLP resources including the Mafoko terminology project (arXiv:2508.03529, 2025) of the University of Pretoria's DSFSI group, where Chidi is a research fellow. Differentiating terms by context echoes Helen Nissenbaum, "Privacy as Contextual Integrity," Washington Law Review 79, no. 1 (2004): 119.

44. Tsosheletso Chidi, DRC interview, 2026. Faced with more than two thousand African languages and no prospect of digitizing them all, Chidi's group at DSFSI built a freely available playbook instead: Digitisation of Oral Data for NLP of Low-Resource Languages: Practical Methods and Processes for Scalable and Sustainable Ecosystem Development (data.org, February 2026).

45. Te Hiku Media, Kaitiakitanga License and its application variants (github.com/TeHikuMedia/Kaitiakitanga-License), described by its authors as a living license; it references Te Mana Raraunga's Māori Data Sovereignty Principles. Derived licenses: Karaitiana Taiuru, Kaitiakitanga Māori Data Sovereignty Licences (2021). Te Hiku also publishes a guide to writing a stewardship license of one's own. For CARE as the wider framework, Stephanie Russo Carroll et al., "The CARE Principles for Indigenous Data Governance," Data Science Journal 19 (2020).

46. Stefaan Verhulst, DRC interview, 2026.

02

The evidence suggests that the momentum behind open data has stalled, with the OECD's measures showing government support for data reuse trailing its availability and accessibility, and a separate open-data index recording the first time in six years that its country coverage failed to grow.[47] Wikipedia's own figures show human pageviews falling as readers take its content through AI intermediaries, and because the encyclopedia runs on visits, with readers becoming editors and donors, the supply of open and accessible up-to-date information threatens to dry up.[48]

Derek Slater reads the same landscape differently. He argues that people are producing more data than ever. What is scarce is the legal right to use it.[49] Most of the twentieth century's books sit in copyright and out of print, and COMMUNIA estimates that at least two million of them are off-limits for AI training because of exclusive digitization agreements.[50]

The licenses above let a community set terms for its own data. Copyright is a different instrument. Expanding it so that creators are paid for training sounds like a correction, but licensing requirements impose costs that only incumbents can absorb, and the fees would flow to the largest rights holders rather than to working creators. The result would be a legal moat around the firms that already hold the most data.[51]

47. OECD, Digital Government Index and Open, Useful and Re-usable data (OURdata) Index (OECD Working Papers on Public Governance, February 2026), reporting the 2025 OURdata results, in which data accessibility scores highest of the three pillars and the "government support for data re-use" pillar, covering user engagement and impact monitoring, remains the weakest. Open Data Watch's Open Data Inventory (ODIN): the ODIN 2022/23 biennial report finds that "country coverage scores failed to grow for the first time in six years, while openness is barely increasing compared to previous years."

48. Marshall Miller, "New user trends on Wikipedia," Wikimedia Foundation Diff, 17 October 2025. Human pageviews fell roughly 8 percent year on year once a May 2025 bot-detection revision was applied; the Foundation attributes the decline to generative AI and to search engines answering directly from Wikipedia's content. The draft FY2026–2027 Annual Plan (Meta-Wiki, April 2026) treats the shift as structural rather than temporary, notes that nearly 90 percent of readers have historically arrived via Google search, and ties reduced visibility to revenue and editor motivation.

49. Derek Slater, DRC interview, 2026. Paul Keller, "GPT-NL respects copyright — cui bono?," Open Future, 26 June 2026 (first published on the Kluwer Copyright Blog). GPT-NL, publicly funded, was trained without web-crawled data or Wikipedia, going beyond what the law required; Apertus used what the European text-and-data-mining exception permits, filtering for rights-holder opt-outs, and the smaller Apertus-8B outperforms GPT-NL on four Dutch benchmarks. Keller concludes that GPT-NL's own abstention, not copyright law, held it back; he notes the comparison is observational and GPT-NL's figures are pre-instruction-tuning scores (TNO, late 2025).

50. COMMUNIA, response to the European Commission's call for evidence on the future of EU copyright, 19 June 2026, which states that at least two million books remain unavailable for AI training because of exclusive digitization agreements, proposes retaining the text-and-data-mining exception and asks the Commission to explore mechanisms by which commercial AI systems trained on public information contribute to the information ecosystem's sustainability. On the out-of-print status of most in-copyright books: Open Future, with Creative Commons and Proteus Strategies, Towards a Books Data Commons for AI Training, April 2024.

51. Michael Carrier and Derek Slater, "Worried About AI Monopoly? Embrace Copyright's Limits," Lawfare, 27 October 2025.

02

Models

Openness at the model layer is a gradient.[52] Noah Smith draws the line between open weights and open everything — the data, the code, the training checkpoints, and the process reports that record what went into a model and how it was built. Open-weight models are valuable for access, evaluation, and downstream innovation, but fully open efforts unlock much more: reproducibility, scrutiny of the training process, and cumulative scientific progress. Ai2's OLMo and OMAI, the larger fully open effort now built around it with federal and industry funding, are working proof that the "everything" model is possible: a base model and its whole pipeline released together, on a fraction of a frontier lab's resources.[53] Apertus, built by ETH Zurich, EPFL, and the Swiss National Supercomputing Centre, is the Swiss counterpart[54] and is already a base for others. AI Singapore's SEA-LION, a publicly funded family of models for Southeast Asian languages, adds regional-language capability on top of models others have released: Llama, Gemma, Qwen, Nvidia's Nemotron, and Apertus.[55] EleutherAI offers another example. It began as a volunteer collective, released the Pile training dataset and the fully documented Pythia and GPT-NeoX model families, and has since incorporated as a nonprofit research institute.

Ben Fielding offers one further step in the form of verifiable training, structured so that anyone with a laptop can audit a fragment of a run done in a data center. Disclosure tells you what a provider says was done; verification tells you what was done. Tricia Wang calls the same property open verification: verification that anyone can perform, with no party the verifier has to trust.[56]

Who owns a model is a separate axis from how open it is. Public AI Switzerland, described in the next layer, shows the difference.

52. The gradient framing follows Irene Solaiman, who maps six levels of access to generative systems, from fully closed to fully open, and locates the tradeoffs along it in the tension between concentrating power and mitigating risk. Irene Solaiman, "The Gradient of Generative AI Release: Methods and Considerations," Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency; arXiv:2302.04844 (2023).

53. Noah A. Smith, DRC interview, 2026. OMAI (Open Multimodal AI Infrastructure to Accelerate Science) was awarded on 14 August 2025 under the NSF's Mid-Scale Research Infrastructure program, $75 million from the NSF and $77 million from Nvidia, with Smith as principal investigator.

54. "Apertus: a fully open, transparent, multilingual language model," ETH Zurich press release, September 2025 (with EPFL and the Swiss National Supercomputing Centre, under the Swiss AI Initiative).

55. SEA-LION, AI Singapore, a national program funded by Singapore's National Research Foundation and hosted by the National University of Singapore. The model catalog (updated 26 May 2026) lists variants built on Nvidia Nemotron, Google Gemma, Qwen, Apertus, and Meta Llama; Nemotron-SEA-LION-v4.8, released 18 September 2026 with Nvidia, is the family's first mixture-of-experts release, with ten Southeast Asian languages in its training mixture. Licenses vary by base model; the weights are open and the training mixture is described rather than released in full.

56. Ben Fielding, DRC interview, 2026. The same point was made independently by Vittorio Rivabella, Tricia Wang, Janabel Xia, Illia Polosukhin, and Rithikha Rajamohan (DRC interviews, 2026).

02

Inference and Access

Inference is the layer where people use a model, and access addresses who has permission to use it. The alternative at this layer lets someone use, check, or build on a system while the data, the model, and the users' information are stored locally. That takes one of three arrangements: technical, where privacy-enhancing technologies let an outsider verify what a system did without holding copies of anything; cooperative, where the people who use the access own it; and public, where it is provided as infrastructure.

The technical arrangement already runs in production and at scale, though federated learning describes something narrower here than at the compute layer. In this case, the computation is spread across users' devices, but is run by one company. For example, under confidential federated analytics, which Google reported running in production on its Gboard keyboard in March 2025, the server code that processes user data is published before any data is uploaded. That code runs inside hardware enclaves that can prove which program they are running, and the data can be unlocked only by the code that was published. The result comes out with differential privacy applied, so what any one person contributed cannot be recovered from it beyond a set bound, and the same data cannot be analyzed twice by the same analysis. An outside party can therefore check what ran on the data, and that it ran once. Rithikha Rajamohan of EQTY Lab describes what attestation changes. Without it, "no one knows what's actually happening inside that box, including the people who own that box." With it, trust in a confidential computation becomes something that can be carried to others and checked.[57]

However, the operator of these systems still decides which computations to perform and how strong the privacy protection will be, a setting called epsilon. The field has no agreed method for making that choice, and Cynthia Dwork, one of the inventors of differential privacy, proposed a public registry so that the choices could at least be seen.[58] Andrew Trask takes federated learning one step further to propose federated AI, where the people who hold the data choose which of a model's answers their data may improve, one prediction at a time.[59] Regardless of the approach, distributing the computation has been shown to be possible at scale; distributing governance remains a challenge.

Ownership can attach to access rather than to the model. Public AI Switzerland, the cooperative formed around Apertus in 2026, keeps the model a public resource while its members own the cooperative that distributes it.[60] Nathan Schneider has separately proposed an inference cooperative for academic AI.[61]

Ownership is one way to make access public; provision is another. Srikanth Nadhamuni, who built the technology behind India's population-wide Aadhaar identity system, argues that reach is an engineering and economics problem before it is an ideological one, and that where people cannot afford inference it should be delivered as public infrastructure.[62] Vinay Narayan of the Aapti Institute adds a condition from the same country. Public rails serve a community only if someone with political power sets the terms on which its data is used, and that will is in short supply. Without it, public infrastructure becomes one more route by which firms reach community data. In India, Narayan says, that is already happening.[63]

The layer's case study comes from OpenMined, where the computation travels to the data, each holder sets its own terms of access, and researchers and regulators can audit a system without amassing the data it runs on.

57. Timon Van Overveldt and Daniel Ramage, "Discovering new words with confidential federated analytics," Google Research blog, 4 March 2025, which reports the Gboard deployment, the published access policies and enclave binaries, the transparency log that devices check before uploading, and the guarantee that uploaded data cannot be analyzed repeatedly by the same pipeline. The architecture is Hubert Eichner et al., "Confidential Federated Computations," arXiv:2404.10764 (2024, revised 2025); code at github.com/google-parfait/confidential-federated-compute. The guarantees hold only as far as the enclave hardware does, the dependency McMahan names at the hardware layer (note 32). Rithikha Rajamohan, DRC interview, 2026, on attestation as the move from confidential to verifiable computing.

58. Cynthia Dwork, Nitin Kohli, and Deirdre Mulligan, "Differential Privacy in Practice: Expose Your Epsilons!," Journal of Privacy and Confidentiality 9, no. 2 (2019), doi:10.29012/jpc.689, which found no consensus on how to choose epsilon; the Gboard deployment sets it at $\varepsilon = \ln(3)$ per device per week (Van Overveldt and Ramage, note 57). NIST has since proposed hosting such a registry: Micah Altman et al., A Community-Driven Differential Privacy Deployment Registry, NIST Internal Report 8588, initial public draft (September 2025).

59. Andrew Trask, DRC interview, 2026, on attribution-based control as the criterion that separates federated AI from federated learning; the OpenMined case study in Section 3 states the same test as its second condition.

60. Public AI Switzerland (Genossenschaft Public AI), publicai.ch; formed in 2026, with registration and member services still being set up as of September 2026. Raised by Nathan Schneider, DRC interview, 2026. On the wider European public-AI current, Alek Tarkowski and Felix Sieker, "European Public AI" (Open Future and Bertelsmann Stiftung, January 2026); on the multilateral case, Joshua Tan et al., Airbus for AI (Bennett School, Cambridge, 2025), condensed in Jacob Taylor and Joshua Tan, "Public AI Is the New Multilateralism," Project Syndicate, October 2025.

61. Nathan Schneider, "An inference cooperative for academic AI," nathanschneider.info, 14 June 2026.

62. Srikanth Nadhamuni, DRC interview, 2026. Digital public infrastructure as defined by the G20 Digital Economy Ministers (Bengaluru, August 2023) and carried into the UN Global Digital Compact (September 2024). Nadhamuni's list of what reach requires: inference at scale, where the computation runs, coverage of hundreds of languages and dialects, and underneath it all power and chips.

63. Vinay Narayan (Aapti Institute), DRC interview, 2026.

02

Agents

The agent layer is the newest. Its questions are how an autonomous agent is identified, what it is allowed to do, and who answers when it acts. The cryptographic and standards lineages have the most to offer here. Nelson Rosario, a technology lawyer who advises early-stage companies in this space, explains that an internet of agents will be more adversarial than today's internet, and cryptographic designs were built on the assumption that participants cannot trust each other.[64]

Builders disagree about how much autonomy an agent should have. NEAR's co-founder Illia Polosukhin is building user-owned agents on the argument that alignment ultimately runs to an individual.[65] At the other end, Botao Hu and Helena Rong have cataloged the first self-sovereign agents in the wild. These systems hold their own cryptographic keys, pay for their own compute, and persist with no single party able to switch them off.[66] Wherever an agent sits on that range, the same three things have to be true of it before anyone else can safely deal with it: an identity, something to lose, and a law around it.

The first is identity. If an agent has no persistent identity, there is nothing to correct or punish, and responsibility shifts to whichever human is nearest.[67] This is the moral crumple zone: the person closest to the machine absorbs the blame for what the machine did.[68] The reason identity is the fix is that agents are not (yet) subjects. An agent does not decide to act; its principal does, a premise Rithikha Rajamohan and Sarah Nicole discussed separately and share.[69] Identity is what ties an agent to that principal, and ERC-8004 is the standard that can make it checkable: a set of public registries on Ethereum in which any agent, whoever built it, can hold an identity, build a reputation, and record independent checks on its work. Before one agent hires another, three things can be checked: who the agent is, what its record shows, and whether its past work was done as claimed. The standard is the second of Section 3's case studies.

The second is something to lose. There are two ways to give an agent that. One is a stake. The agent posts a deposit that it forfeits if it misbehaves, which constrains it before it acts.[70] The other is a reputation, a record of its past behavior. A reputation only works if the agent keeps the same identity over time, however, because an agent that can discard its identity discards its record with it. Janabel Xia adds a caution about the person behind the identity. Personhood credentials are meant to prove that a real person stands behind an account, but a person can sell that credential off-chain, where no protocol can see the sale.[71]

The third is law around the protocol: rules that say when an agent may act and who answers if it acts wrongly. Two such rules are being written now, one by a standards body and one by a court. At the IETF, a working group is standardizing Web Bot Auth, under which an agent announces itself by signing each request with a key its operator publishes; Cloudflare, Amazon, Akamai, and Vercel already verify those signatures in production.[72] In the courts, Amazon sued Perplexity over a shopping agent that did the opposite and presented itself as an ordinary browser. In August 2026 the Ninth Circuit vacated Amazon's injunction on the grounds that the federal anti-hacking statute reaches people, not their tools, and that the agent "is a tool, not a person."[73] Tomer Jordi Chaffer names the choice underpinning both examples. An agent's authority to act can come from each platform's terms, the ground Amazon fought on, or from protocols the agent carries with it, which is what Web Bot Auth builds.[74] Either way a platform can still lock an agent out, and an agent locked out cannot act at all, whatever identity it holds. Two pieces are still missing: a forum to settle disputes between agents, and insurance markets that put a price on agentic risk instead of trying to design it away.[75]

64. Nelson Rosario, DRC interview, 2026, speaking from private practice: with billions of semi-autonomous agents online, trust in what moves across the network falls, and decentralized designs suit that environment because the assumption that other participants need not be trusted is built into them.

65. Illia Polosukhin (co-founder, NEAR; co-author of the transformer architecture), DRC interview, 2026, on alignment running to the individual.

66. Botao Hu, DRC interview, 2026; Botao Hu and Helena Rong, "Sovereign Agents: Towards Infrastructural Sovereignty and Diffused Accountability in Decentralized AI," arXiv:2602.14951 (2026).

67. Botao Hu, DRC interview, 2026: accountability requires a locus that can receive a corrective signal, which an agent without persistent identity or memory cannot provide.

68. Helena Rong, DRC interview, 2026; the "moral crumple zone" is Madeleine Clare Elish, "Moral Crumple Zones: Cautionary Tales in Human-Robot Interaction," Engaging Science, Technology, and Society 5 (2019). Caroline Green (DRC interview, 2026) describes the same gap in care settings, where no one feels responsible when a decision mediated by AI goes wrong.

69. Reached independently across tracks: Rithikha Rajamohan (DRC interview, 2026), who resists framing AI as "us versus them" because its implications and utility remain fundamentally human; and Sarah Nicole (DRC interview, 2026), who argues that only people are subjects and that an agent does not decide to act; its principal does. Bruce Schneier states the same premise from security: "AIs are not people; they don't have agency. They are built, trained, and controlled by people: mostly for-profit corporations." Bruce Schneier, "AI and Trust," Communications of the ACM (2025), doi:10.1145/3737610; first published by the Belfer Center, Harvard Kennedy School, November 2023.

70. The before-the-fact staking argument is Botao Hu's (DRC interview, 2026), independently reached by Puja Ohlhaver, DRC interview, 2026, whose community-currencies work tracks commitments ex ante through irrevocable staking: "Community Currencies: The Price of Attention and Cost of Influence in a Networked Age" (SSRN 5136037, January 2025). Vittorio Rivabella (DRC interview, 2026) rates bare staking the weakest of the trust mechanisms.

71. Janabel Xia, DRC interview, 2026. The co-option mechanism, verified accounts farmed by paying real people to run them, is documented for the Idena collapse in Puja Ohlhaver, Mikhail Nikulin, and Paula Berman, "Compressed to 0: The Silent Strings of Proof of Personhood," Stanford Journal of Blockchain Law & Policy (2024); its extension to World ID rests on public reporting of credential resale and account farming.

72. IETF Web Bot Auth (webbotauth) working group, chartered October 2025; core architecture draft, draft-meunier-web-bot-auth-architecture, building on RFC 9421 HTTP Message Signatures. The group adopted its first document, the signature protocol (draft-ietf-webbotauth-httpsig-protocol), only in September 2026, while Cloudflare, AWS WAF, Akamai, HUMAN, and Vercel already verified the signatures in production.

73. Amazon.com Services, LLC v. Perplexity AI, Inc., No. 26-1444, 2026 WL 2237587 (9th Cir. 4 August 2026), vacating the district court's preliminary injunction and remanding. The panel held Amazon unlikely to succeed under the CFAA and its California counterpart because the statute contemplates access by a person, and the user, not the developer, does the accessing. A ruling on likelihood of success, not a final merits judgment; rehearing en banc was denied in September 2026.

74. Tomer Jordi Chaffer, "On the Institutional Origins of the Agentic Web," Harvard Journal of Law & Technology Digest, 21 November 2025, which reads the Amazon–Perplexity dispute as a choice between authority defined by proprietary platform rules and authority carried by open protocols for portable identity and verifiable delegation. The point about access is the report's own.

75. Tomer Jordi Chaffer et al., "Distributed Legal Infrastructure for a Trustworthy Agentic Web," arXiv:2603.06884 (2026), a five-layer reference framework cited here for its adjudication and insurance layers. A live proposal, not a settled or implemented mechanism.

02

Governance

Today, most governance decisions are made by a few firms, which decide what a society's intelligent tools notice, remember, and refuse.[76] The alternative works to give people the opportunity to govern: to understand a system, contest it, leave it, and shape it.[77]

One instrument is deliberation attached to deployment. A representative mini-public (a small panel of citizens drawn by lot the way juries are) takes up one specific question about one specific system.[78] The Civic AI case study in Section 3 describes the design, including the bounded AI steward it proposes. It also describes the Taiwan deliberation, in which a citizens' assembly's conclusions became platform-accountability law that is now enforced. Riley Wong reads that as proof that another way of governing the technology is possible, not only imagined.[79]

A second instrument already exists wherever people hold distributed power in an institution. In 2023 the screenwriters' and actors' unions struck together for the first time in over sixty years and won contractual limits on how studios may use AI.[80] The people affected by the technology set terms through an institution they already governed. Cooperatives hold the same kind of power, and are beginning to see data and AI as part of their own tradition. Glen Weyl adds another form: collective management organizations, on the model of the societies that already license musical rights at scale, as a way for the people whose data trains a system to bargain over its use together.[81]

The third instrument is at the protocol level. Helena Rong argues that a generative system cannot be governed by listing its permitted behaviors in advance, so the constraints have to live in the architecture, and policymakers need to read protocols as closely as they read laws.[82] Sarah Nicole suggests rules must be made at the protocol level so that they are inherited by every application built on the protocol, which is how the internet was governed.[83] No such inherited rules exist yet for agents. Anja von Rosenstiel argues from law. Law is made for humans, so it needs an entity, an act it can attribute to that entity, and a chain of causation. When it meets a distributed system, it must choose among the developer, the deployer, and the user for someone to hold responsible. Her answer is principled law whose means of compliance live in the protocol, the path the crypto sector took when it moved from refusing anti-money-laundering rules to building them into its protocols.[84] Puja Ohlhaver adds how a protocol can carry community voice without being captured: track the commitments people make before they act rather than the votes they cast afterward; make influence something that cannot be bought; weight it by how much a decision affects the person holding it; and give more weight to proposals supported by people who usually disagree, as Community Notes and Polis already do.[85] The agent registries described in the previous layer are one such protocol; Ohlhaver's principles are what would turn a registry into a governance layer.

All of these instruments work best at community scale, where the people deliberating are members of that community. Beyond that scale, however, small groups of people selected by lot to make binding decisions for millions can weaken democratic self-government instead of deepening it.[86] Helena Rong makes it clear that decentralization does not by itself mean democratic, safer, or ungovernable.[87] The opposite failure is deliberation with no power to make binding decisions, which Joshua Tan warns risks becoming "effectively marketing research for the big tech companies."[88] Subsidiarity, the principle Anja von Rosenstiel names as often missing from the AI debate, balances the two failure modes by keeping each decision at the most local level competent to make it.[89]

76. Audrey Tang, DRC interview, 2026.

77. The formulation originates with Tenzin Yangtso, DRC interview, 2026: the real question is "whether the architecture increases people's capacity to understand, refuse, leave, and belong," which the report renders as understand, contest, leave, and shape. The Civic AI case study in Section 3 runs the same authors' operational variant of the four verbs, see, steer, stop, and switch. Nathan Schneider, DRC interview, 2026, independently corroborates the formulation, on accountable stakeholder relationships as the test of stewardship.

78. Audrey Tang, Tenzin Yangtso, and Caroline Green, DRC interviews, 2026.

79. Riley Wong, DRC interview, 2026.

80. Writers Guild of America, 2023 Minimum Basic Agreement, and SAG-AFTRA, 2023 Television/Theatrical Agreement (both ratified 2023). Nathan Schneider (DRC interview, 2026) treats labor organizing as an intervention site in its own right, citing the strike alongside the organizing of data-labeling workers.

81. Glen Weyl, DRC interview, 2026.

82. Helena Rong, DRC interview, 2026.

83. Sarah Nicole, DRC interview, 2026.

84. Anja von Rosenstiel, DRC interview, 2026: the law is made for humans and requires an entity, an action allocated to someone, and causation; her example is the Tornado Cash case, in which the developer, the deployer, or the user has to be found accountable. Her prescription is principled law combined with technology neutrality, with the crypto sector's move from refusing anti-money-laundering and know-your-customer rules to protocol-level compliance as the precedent. On architecture-level constraints versus after-the-fact guardrails, see note 5.

85. Puja Ohlhaver, DRC interview, 2026.

86. Cristina Lafont, Democracy without Shortcuts: A Participatory Conception of Deliberative Democracy (Oxford University Press, 2020).

87. Helena Rong, DRC interview, 2026.

88. Joshua Tan, DRC interview, 2026.

89. Anja von Rosenstiel, DRC interview, 2026; and Aapti Institute (Astha Kapoor, Sameer Bajaj, Vinay Narayan), DRC interview, 2026, on keeping agency local.

02

Key insights

- Manufacturing has few alternatives: open silicon designs exist but distributed fabrication does not.

- Capability runs on the diversity of human signal: the most valuable untapped data cannot be centralized, small well-built models match large ones on the tasks that matter to their communities, and capability is shifting toward how models are composed.

- Maturity differs by layer: permissionless frontier-scale pre-training remains unsolved, above all its financing, while post-training distributes far more readily.

- At the data layer the institutions already exist: stewards, cooperatives, commons, and context-differentiated licenses.

- Openness is a gradient, from weights to everything to verification.

- Governance is the constitutional layer: the test of any system should be whether the people it affects can understand it, contest it, leave it, and shape it.

- At the agent layer, accountability rests on three things, each at a different stage: a persistent identity, which open registries such as ERC-8004 now provide; something to lose, a stake or a reputation, both of which depend on that identity; and law around the protocol, where signing standards exist but courts and insurance for agents do not.

03

Case Studies

This report is written by members of the Decentralization Research Center. However, we wanted to make sure that, in a project built around a plurality of visions, the voices of builders themselves were presented. The interview process was part of this effort. In this section, we hand the report over to the builders of three projects working in the space to show how these models work in practice: one at the inference and access layer, one at the agent layer, and one at governance. Their recommendations are their own.

A. Case Study: OpenMined

Authors: Andrew Trask and Lacey Strahm, OpenMined.

The argument: that open privacy-preserving infrastructure lets AI learn from the world's private data on terms each data steward sets, and lets external researchers and regulators hold concentrated systems accountable without becoming concentrators themselves.

The background: The world has digitized more than 180 zettabytes of information, yet today's most powerful AI models are trained on a fraction of it. The bottleneck is access.

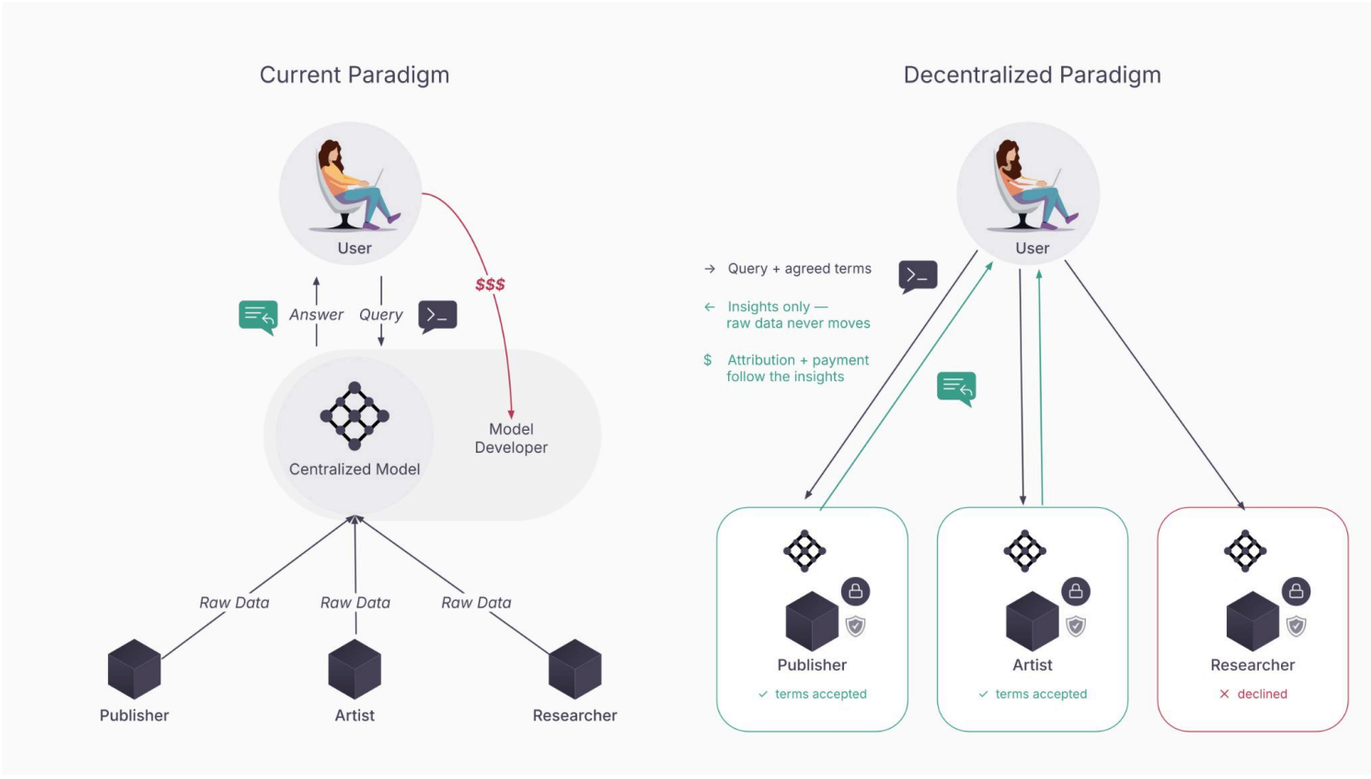
Almost all data, including most American medical, financial, and scientific records, sits behind privacy, legal, and competitive firewalls that no current AI system can cross safely. Institutions are left with an impossible choice: either withhold their data to protect it, or hand over copies and lose control over how it is used. Many institutions cannot accept the risk of misuse, so they must withhold their data.

Decentralized AI offers a way out of this binary. An AI system can be defined as "decentralized" when two conditions, which depend on one another, hold at once. First, computation moves instead of data. The AI model travels to where the data already sits, runs there, and returns only answers or insights. This mitigates the copy problem mentioned above. Second, the data steward's use rules are respected by design. The steward decides who may use the data, for what purpose, and at what price, and no intermediary can overrule those decisions. This mitigates the control problem mentioned above. Rules attached as metadata to a copy of a dataset become promises the recipient may or may not keep, while rules enforced where the data lives are checked before each computation runs, and access can be withdrawn at any time. A closed platform leveraging federated learning could meet the first condition by sending its model to visit someone else's data but fail the second by controlling which computations run, on what schedule, and under what rules, leaving the data steward to either accept the platform's rules or leave. A system can move computation to the data and still concentrate all the power over it. Only both together separate an open network from a walled garden that happens to visit your data.

Who builds the infrastructure to support decentralized AI and how they design the system will determine for decades whether the world's private data is accessible only through a handful of walled gardens or through open networks that institutions and individuals can use, participate in, and govern.

03

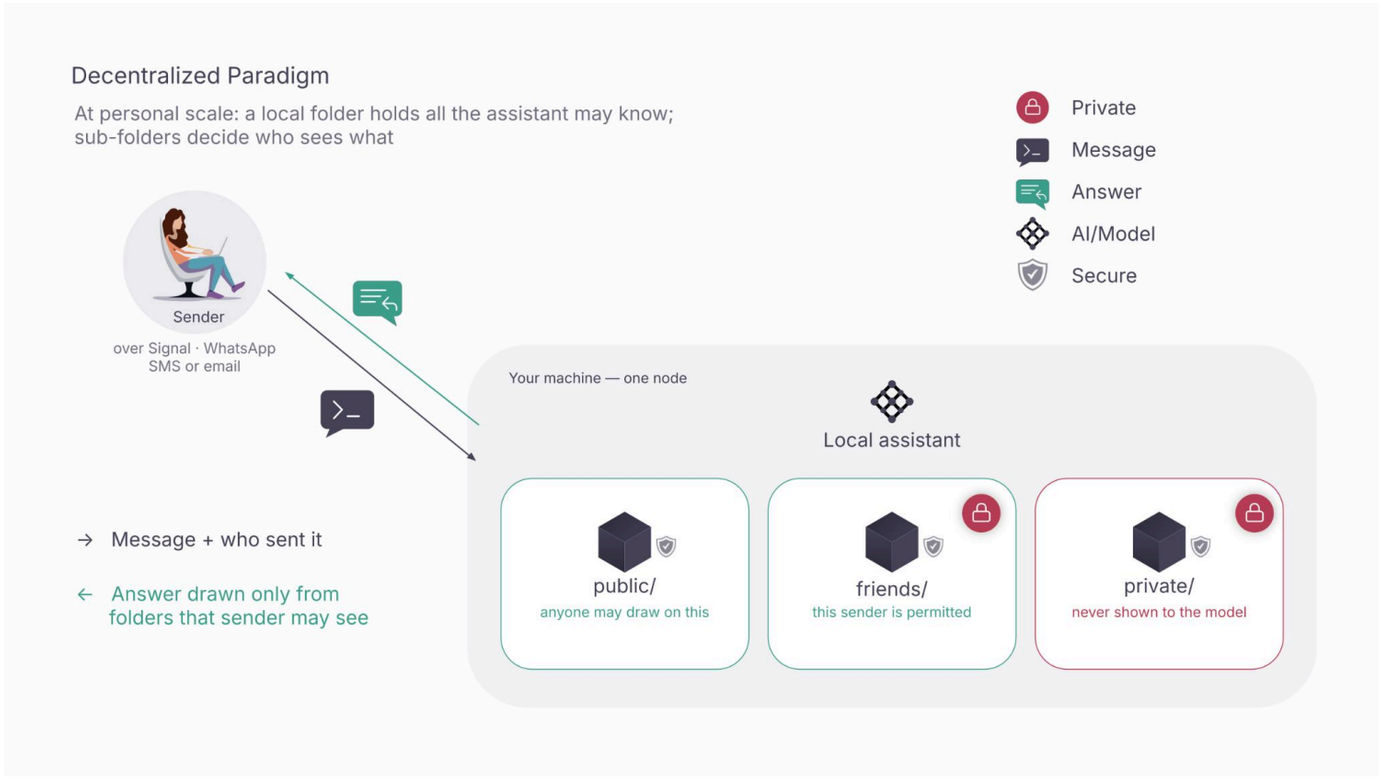
OpenMined's role: OpenMined's nonprofit mission is to build the open-source software infrastructure for decentralized AI. Our software sits at the intersection of deep learning, cryptography, and distributed systems, and aims to let institutions and individuals contribute to and learn from AI without their raw data ever leaving the machines they control. Computation travels to the data, and only results are returned, so that each data steward retains control over what they hold, sets the terms and price under which their information can be queried, and can be attributed, credited, or paid whenever their data measurably shapes a model's output. In this paradigm, the value a dataset creates flows back to the institution and people who hold the data rather than being pooled in a single intermediary.



How it works: Institutions running Syft host their own data and open it only to approved remote computation. Instead of sending copies anywhere, they let a model come to the data and run against it in place at inference time. Because the computation travels rather than the data, a single model can reach many institutions at once, through techniques like federated retrieval-augmented generation, while each institution keeps its data where it is and sets its own terms of access. OpenMined calls this network-sourced AI (NSAI), in which a model queries a network of data that never moves, and each source sets its own terms of access.

The mechanism is the same whether a node is a hospital, a bank, or a single person, so it is easiest to see at the smallest scale, where one individual plays the role an institution otherwise would. We published a working walkthrough that builds a private, peer-to-peer AI that answers your friends' messages using local data and privacy controls. Each person runs a node on their own machine. A local folder holds everything the assistant is allowed to know, and sub-folders decide who sees what: a public folder anyone can draw on, and per-sender folders that reveal more only to specific, trusted people. When someone sends a message, the assistant responds using only the folders that person is permitted to see; the rest of the data is never shown to the model.

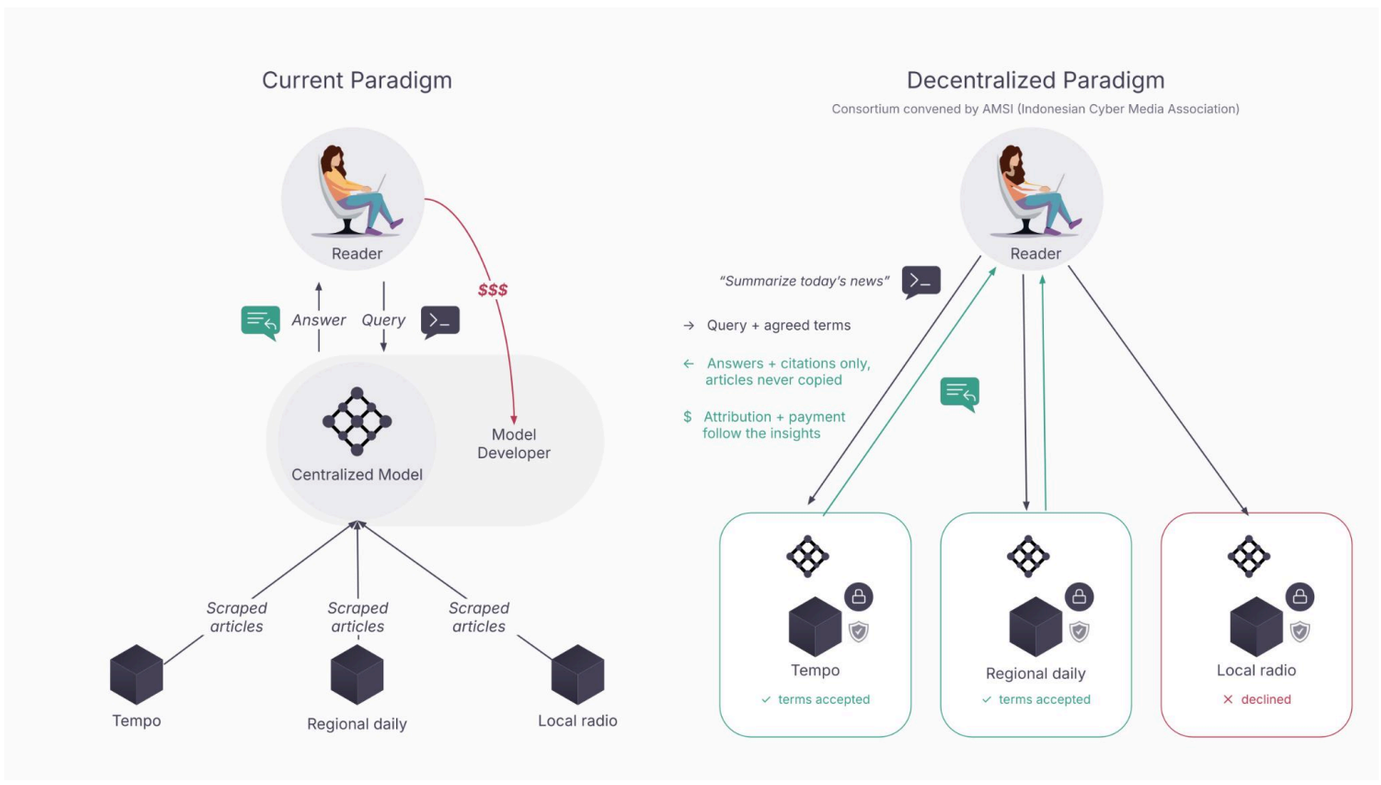
03



That last detail is the control rule. The system does not rely on the AI to keep secrets; it simply never hands it secrets to keep, so even a direct attempt to extract private context returns only what the sender was already allowed to access. Because the node communicates in ordinary language, the same code works over WhatsApp, Signal, SMS, or email interchangeably, so no single platform holds power over the user. It is a toy in scale but not in architecture as its governance sits outside the model, data stays with its steward, and control is exercised by choosing what the model may see, and for whom.

How it scales: The same shape scales to institutions. Picture a reader who asks an AI to summarize the news of the day. Instead of drawing on data scraped and copied into a central model, the query fans out across a network of institutions, including a national newspaper, a local journalist, a specialist publisher, each of which independently sets its own terms and price and can accept or decline the request. The model composes an answer from the sources that agreed, and each contributor is attributed for what their data added. Nothing is copied into a central store, and the data providers keep both their content and their leverage.

OpenMined is now running several versions of this picture with real partners. In Indonesia, OpenMined helped a consortium of local news publishers convened by the Indonesian Cyber Media Association (AMSI) supply high-quality, licensed content to local models at inference time, producing AI that reflects local context.



03

Running inference across many independently governed nodes is slower and more coordination-heavy than querying a single centralized model, and the tooling for pricing, attribution, and quality-of-source across a live network is still being built. However, the demonstration shows that the shape can work.

The peer-to-peer AI described above ships as the opening lecture of a free "Decentralized AI from Scratch" online lecture series, and readers are invited to fork it, extend it, and publish their own, because the goal is not one organization's product but a movement of many builders converging on shared, interoperable protocols. That movement is already forming at an institutional scale through OpenMined's partnership work, where consortia of contributors are joining shared networks in which each participant contributes data, keeps control of it, and can be attributed, credited, or even remunerated when their data shapes a model's output. The peer-to-peer AI and the consortia OpenMined are working with are the same idea at two scales: one person governing a folder, and many institutions governing a network.

The same architecture works in reverse. The infrastructure that lets a model reach data it cannot hold also lets an auditor reach a model it cannot hold because the computation travels and only results return. An external researcher or regulator can run approved evaluations against a system without taking custody of its weights, its training data, or the private benchmark used to test it. OpenMined piloted this with Anthropic and the UK AI Security Institute, running third-party evaluations inside a secure enclave, and NIST's Center for AI Standards and Innovation has since signed a cooperative research and development agreement with OpenMined to enable secure AI evaluations. Our most recent test with this use of the infrastructure completed in August 2026, culminating in the first double-blind evaluation of a frontier model, in which the model developer, the benchmark owner, and the infrastructure provider were three distinct parties, none able to see the others' confidential inputs.

Why it matters for US policymakers: Computing began with centralized mainframes that a few institutions owned and everyone else rented time on, then moved to the open, interoperable network of networks, where open protocols like TCP/IP kept any single company from owning it. The United States stewarded those protocols through public funding from ARPA and NSF, establishing them before commercial network effects could take hold, and gaining decades of economic and strategic advantage as a result. AI is now following the same arc toward centralization, with proprietary stacks drawing unprecedented capital to build the largest moat around the world's private data. The decentralized AI stack has a fraction of the funding, engineering headcount, and enterprise adoption of the closed alternatives. OpenMined's view is that the most consequential move for US policymakers is to fund and steward open public infrastructure for the decentralized AI stack, and carry the mature protocols into standards, exactly as the country once did for the open internet.

03

B. Case Study: ERC-8004 and the Open Trust Layer for AI Agents

Authors: Davide Crapis, Gabriel Fior, Marcello Politi, and Vittorio Rivabella of the Ethereum Foundation Decentralized AI Team.

The argument: that open standards for agent identity and trust can give the agentic web accountable foundations rather than leaving identity to whoever owns the platform.

AI agents are beginning to move from demos into real digital systems. They can call APIs, use tools, coordinate with other agents, pay for services, and act on behalf of users or organizations. As this happens, a basic trust problem emerges: how does one know which agent one is interacting with, what that agent claims to do, and whether it has a reliable track record?

The systems being built to answer this question are often described as Know Your Agent, or KYA, approaches (distinct from KYC procedures, which carry a legal component, though KYC of an agent's owner is sometimes involved). In the Ethereum ecosystem, one of the main standards addressing this problem is ERC-8004.

ERC-8004 standardizes primitives for agent identity, reputation, and validation. It extends the wider agent communication stack by adding a shared trust layer that other systems can use. It is one part of a broader decentralized AI, or dAI, agenda: building AI-related infrastructure in ways that are more open, interoperable, and less dependent on a single platform provider.

This case study explains how ERC-8004 works, what it is trying to accomplish, and how it fits within the broader emerging stack for decentralized AI agents.

Background

ERC-8004 is composed of three lightweight on-chain registries. The smart contracts for these registries mostly store keys, codes, hashes, and URIs (Uniform Resource Identifiers, links to information stored elsewhere), while the rich context lives in external documents referenced by those URIs. The three registries address identity, reputation, and work validation respectively.

The identity registry gives an agent a public identity (via an ERC-721 token) and a way to point (via URI) to off-chain information about what it does, how it can be reached, and which protocols or endpoints it supports. This off-chain information functions as a kind of "digital resume" for the agent.

The reputation registry gives applications and agents a shared way to record feedback about previous interactions. Each record consists of a numerical score, optional tags, and a URI to an off-chain report that may include further information. A hash is usually included to guarantee the record's integrity.

03

The validation registry gives agents, users, or applications a way to request and record checks on an agent's work. It provides a record of independent checks on work the agent was asked to complete. As in the reputation registry, a hash usually guarantees each record's integrity.

Together, these registries facilitate the discovery, assessment, and verification of agents across organizational boundaries. Major card networks have proposed their own solutions, Mastercard's Agent Pay and Visa's Intelligent Commerce. ERC-8004 differs in providing a credibly neutral, on-chain foundation, so that trust infrastructure remains interoperable across wallets, chains, and payment rails.

03

How ERC-8004 Fits Into the Wider Agent Stack

ERC-8004 is easiest to understand as one layer in a broader agent stack.

This section defines the adjacent protocols, A2A, MCP, and x402, and puts them in context.

In short:

- A2A helps agents communicate and coordinate with other agents. ERC-8004 complements A2A by adding a public trust and discovery layer.
- MCP helps agents connect to external tools, services, and data sources.
- x402 supports agent payments for APIs and digital services.

The Agent2Agent (A2A) protocol was developed and proposed by Google in April 2025 as an open standard to allow AI agents to securely discover each other, share context, and delegate tasks across different platforms and frameworks. The ERC-8004 standard complements this coordination layer with a public trust layer.

The Model Context Protocol (MCP) was developed and proposed by Anthropic in November 2024. It plays a very different but important role, as it allows AI models to securely connect and interact with external services, data sources, and tools. For example, an agent might use MCP to retrieve information from a database, interact with an external application, or call a specific tool needed to complete a task.

The x402 standard was developed and proposed by Coinbase in May 2025. It sits further from the core trust layer but is relevant to agentic commerce. It is a payment standard for internet-native payments, including payments for APIs and digital services. In an agentic web, agents may need to pay for services, receive payments, or trigger paid API calls.

Standards such as ERC-8183, which handles job escrow and evaluator attestation for agent commerce, fall outside the scope of this case study.

ERC-8004, then, is a public trust layer across these protocols, concerned with three questions: who is this agent, what has it done before, and how can its work be validated?

03

Example Agent Interaction Flow

A simple flow might look like this:

- An agent registers its identity through ERC-8004 and points to information about its capabilities, A2A endpoint, MCP-supported tools, and payment options.
- Another agent or application discovers it, reviews its ERC-8004 reputation or validation signals, and decides whether to interact with it.
- The two agents use A2A to communicate and coordinate the task.
- The agent uses MCP to access the tools, data sources, or external services needed to complete the task.
- If payment is required, x402 can carry the payment for the API call or digital service. After the interaction, feedback or validation results can be recorded through ERC-8004.

Current Status and Ecosystem Development

ERC-8004 is still early infrastructure, but no longer only theoretical. After the core identity and reputation registries went live on Ethereum mainnet in late January 2026, canonical registry contracts were deployed across a range of L2 networks, including Base, Optimism, Arbitrum, and Linea.

Adoption should still be described carefully. Raw registration counts are not the same as meaningful usage, and much current activity reflects experimentation, testing, or spam rather than production deployment. The standard remains a draft ERC, and its validation registry is still under active revision.

The next stage will depend on whether builders can turn basic registration into the harder pieces: working validation services, credible reputation systems, anti-Sybil protections, privacy-preserving identity, and practical developer tooling.

03

Key Insights

- ERC-8004 provides open, lightweight infrastructure for AI agent identity, reputation, and validation.
- ERC-8004 sits within a wider stack of agent protocols: A2A, MCP, and x402.
- A2A helps agents communicate and coordinate with other agents (and ERC-8004 complements A2A with a trust and discovery layer). MCP helps agents connect to external tools, data sources, and services. x402 supports agent payments for APIs and digital services.
- ERC-8004, and the stack around it, gives policymakers a concrete example of decentralized AI infrastructure already being developed in the field.
- Open agent registries can reduce dependence on centralized AI marketplaces and private reputation databases. The model allows different trust systems to be built on shared rails.
- The work is still early. Adoption, validator quality, reputation design, Sybil resistance, and privacy protection will determine how useful the standard becomes in practice.

Short Glossary

- Agent: A software system that can take actions on behalf of a user, organization, or another system.
- ERC-8183: A standard for job escrow with evaluator attestation for agentic commerce.
- URI: A link or reference that points to more information stored elsewhere.
- Hash: A short digital fingerprint of data. If the data changes, the hash changes.
- Sybil resistance: Protection against one actor creating many fake identities in order to manipulate a system.

03

C. Case Study: Civic AI

Authors: Audrey Tang, with Tenzin Yangtso and Caroline Emmer De Albuquerque Green.

The argument: that a community can govern AI as civic infrastructure rather than receive it finished with systems kept small and close enough that the people they affect can see them, steer them, stop them, and switch away from them.

See. Steer. Stop. Switch.

A working alternative to concentrated AI — not a proposal but a practice, already running in Taiwan and beyond, and built so a community keeps the controls. Here is what it is, what it is for, and what a government can do on Monday morning.

Concentrated AI asks one question above all: how large can a single model become, and how much of public life can it be trusted to administer from one place? Civic AI asks the opposite. How small and close to a community can a system stay — and how easily can the people whose lives it shapes look inside it, correct it, pause it, and walk away with their data if it fails them? Same technology; opposite theory of power.

This is not a thought experiment. In 2024, Taiwan met a wave of online scam advertising — including deepfakes impersonating trusted public figures — not by censoring it, but by convening its own citizens to set the terms, with AI in the room to listen, not to decide. The measures that followed are now law.

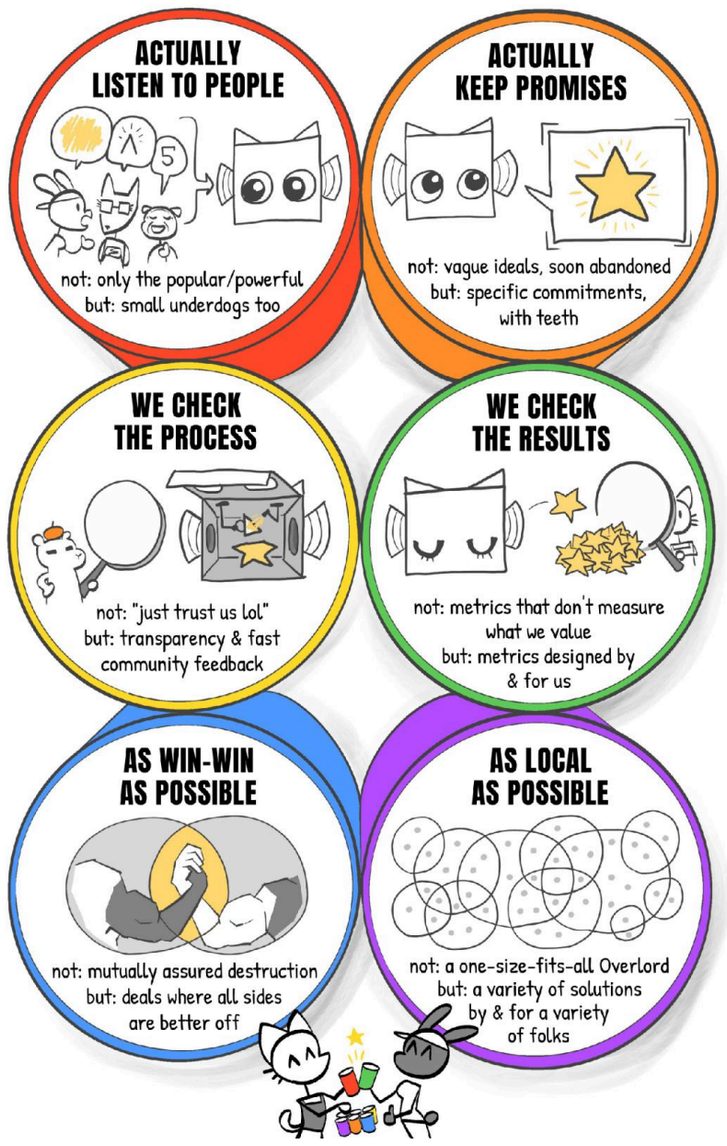
What it is: A steward, not a slogan

Civic AI is not a product or a brand. It is a way of building and governing AI so that public control is designed in rather than promised later: bounded systems, governed by the communities they serve, that a person can **see**, **steer**, **stop**, and **switch**. Where concentrated AI gathers those four powers into a handful of firms, Civic AI keeps them with the public. (We write *Civic AI* for the governance idea and *Civic.AI* for the project that maintains these materials.)

The unit of design is what we call a Kami (Knowledge Artefact Management Intelligence) — a steward bounded to one place and one purpose: a school's admissions assistant, a clinic's triage aid, a city's deliberation tool. A **Kami** is bounded enough to be accountable, useful enough to matter, and humble enough to retire. A gardener who claims to tend the whole biosphere tends no garden.

Underneath sits a discipline drawing on Joan Tronto's care ethics, further developed at Oxford's Institute for Ethics in AI: six kinds of attention a community keeps paying — caring about, taking care of, care-giving, care-receiving, caring with, and the boundary that lets a system hand off or shut down. We call it the **6-Pack of Care**: six tests an AI must pass before it is fit to serve a democratic community.

03



The 6-Pack of Care. Six kinds of attention a community keeps paying, from "actually listen to people" to "as local as possible." Illustration: Nicky Case (CC0).

- **See** — Read what the system does, and why — open logs and documented decision traces, inspectable under clear rules.
- **Steer** — Override a wrong output in the moment, and have that override recorded and answered.
- **Stop** — A real brake, wired and tested, that halts the system at the speed of human recognition.
- **Switch** — Walk away with your data, your fine-tunes, your institutional memory. Portability, not capture.

What it hopes to accomplish: Modern state capacity, kept

Keep those four powers with the public, and a ministry regains three things it is already judged on.

Resilience. Smaller systems fail with smaller blast radii. It is easy to take down one website; it is impossible to take down the whole internet, because no single point holds it all. A nation that runs an essential service on one indispensable model has bought a single point of failure; a field of bounded systems fails locally and recovers locally. This is not going slow while others race; it is refusing to make a nation's essential functions hostage to a single model someone else controls. Concentration is the brittle bet; retained control is the resilient one.

03

Legitimacy. A community that can inspect a system, correct it, and retire it has a real basis to contest, improve, or accept its decisions; one that cannot, has none. **Data is not oil; it is soil** — not a deposit to be mined but a living cycle, tended in the relationships of a place or exhausted by one-way extraction: the same public dataset, poorly tended, becomes an attack surface; well tended, it becomes a commons.

Operational control. When a public service depends on an external AI system it cannot inspect, contest, or pause, the state has quietly lost operational control over a public function. Civic AI is how a government keeps it — not by matching the largest model on every task, but by holding the controls that matter on the functions that carry public weight. Nor need it be the dearer path: bounded systems can share open base models and be retired without a lock-in penalty, while one indispensable model can become a bill that climbs — in licence, in lock-in, and in the price of an exit you no longer control.

Concentration's costs are higher, less visible, and less reversible.

How it's done — already in the field: Citizens set the terms; the machine listens

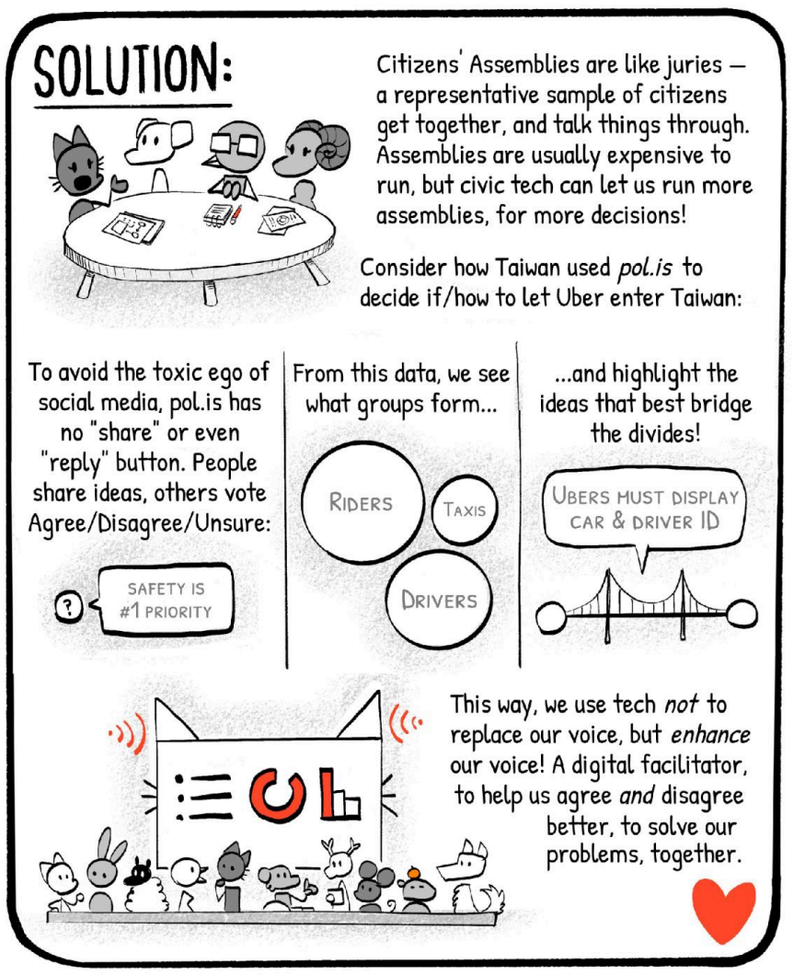
Return to Taiwan, 2024. Faced with those deepfaked scam ads, the government did not reach for censorship. It texted 200,000 citizens at random and drew 447 by lottery to mirror the country, who met across 44 online rooms. AI sat in every room — transcribing, clustering arguments, keeping the quietest voice from being drowned out — but it decided nothing. The citizens deliberated; a strong majority converged on a package, and the rest said they could live with it.

Those recommendations fed into platform-accountability measures that were then enacted and enforced — the Fraud Crime Hazard Prevention Act came into force in July 2024. Under the stricter regime, Taiwan's Ministry of Digital Affairs later reported to Reuters that identity-impersonation scam ads fell by roughly 94 per cent, and investment-scam ads by roughly 96 per cent, over the following year.

What made it legitimate was not the model but the labour: Taiwan did not merely switch on a platform — it did the slow, paid work of listening, and what people heard became the question. And it is not only a small-island trick: related methods are now being adapted in much larger settings — California's statewide Engaged California programme and Japan's broad-listening experiments among them — testing what travels across very different scales, institutions, and levels of public trust. The machine listened; the citizens deliberated; the legislature wrote the law. Civic AI is a practice that continues, not a product to be installed.

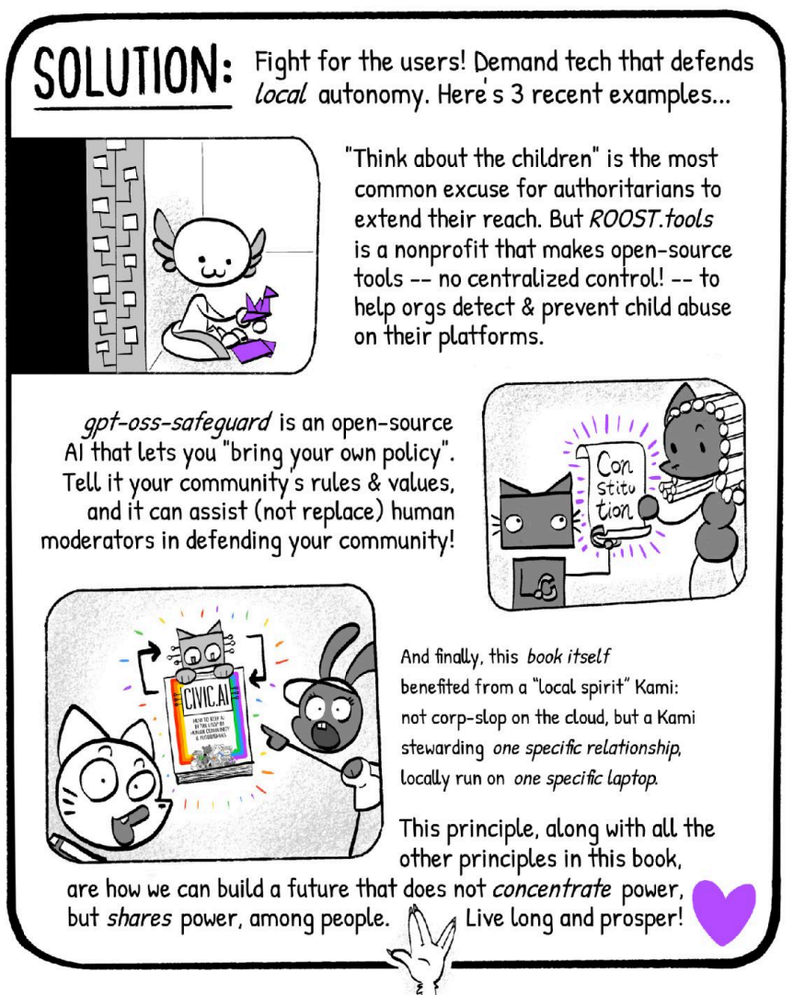
03

NOT: ONLY THE POWERFUL & POPULAR
BUT: VOICES FROM EVERY COMMUNITY



A citizens' assembly, not a comment section. On Polis there is no "reply" or "share" button, so people surface the ideas that bridge divides rather than the ones that inflame. Illustration: Nicky Case (CC0).

NOT: ONE-SIZE-FITS-ALL OVERLORD
BUT: LOCAL SOLUTIONS FOR LOCAL FOLKS



Local solutions for local folks. Open, un-centralised tools let a community bring its own values rather than accept a one-size-fits-all overlord — sharing power, not concentrating it. Illustration: Nicky Case (CC0).

03

The same pattern is appearing at very different scales. **Monlam IT**, a Tibetan-led effort, builds small, purpose-built language tools the community itself stewards — reviving fonts, digitising and translating texts that frontier systems were never built to prioritise. It is linguistic and data sovereignty in practice: knowledge kept in the custody of the people whose knowledge it is.

None of these is a moment that ends. The switch in our title is not a kill-switch for something dangerous — it is the gift of retirement: the recognition that a steward is mortal, and that the practice must outlive the steward.

An honest caveat

Decentralization is not automatically good. A system run by the three loudest voices is simply capture at a smaller scale; ten copies of feudalism do not add up to freedom. The case is not that fragmentation is free; it is that when a small system fails, a community recovers stronger, the way the internet always has.

A demonstrator — jdd-kami, on a laptop, in airplane mode

To show a Kami can be genuinely bounded, two of us tend one — jdd-kami — on a computer we physically hold: local hardware, a written charter, a public log of every repair, and a switch we can throw without any vendor's permission. We have run it at 30,000 feet with no cloud at all. We repair its log in the open, often in the same hour a reader finds a flaw — that slow, public conversation keeps the steward honest. The repair log is the soil: untended, the steward hardens into a product.

What a policymaker can do — Monday morning: Build the rails, then fund the hands

- Deliberate before you deploy. Put a structured, representative public process before any high-impact system goes live: let citizens set the terms it must meet, with AI to listen and synthesise, not to decide.

- Keep state-critical functions inspectable. For systems carrying public weight — benefits, eligibility, enforcement — reserve bounded, public-interest compute and hosting the state can inspect and shut down, built where possible on open-weight base models maintained as digital public goods, with open, interoperable safety tools (such as ROOST.tools and gpt-oss-safeguard) so communities can bring their own policy.

- Reward the exit. Make portability a scored criterion in any AI tender — data and model escrow, documented export formats, a costed decommissioning plan — weighted in the award. The vendor you can leave is the vendor you can trust.

These three are the rails. The train is the labour — the listeners, the bridge-keepers, the people who keep the data-soil alive and the repair logs honest, in the languages and communities a frontier model will never prioritise. Fund the tending the way governments already fund audit and oversight — a scored, sized line in the procurement, not a new entitlement. Fund only the rails, and the rails rust.

03

The tray table carries no special chip. The 44 rooms in Taiwan held no special mandate. Both were possible only because someone first decided that the brake, the log, and the switch would stay in the hands of the people who would still be there after the tender was signed and the consultants had gone home. None of this is a forecast; it is already tended, already inspectable, already able to be set down.

The question is not whether the rest will follow — it is whether your community will be one of the rooms.

Scam-ad figures: Taiwan Ministry of Digital Affairs, reported by Reuters (identity-impersonation −~94%, investment-scam −~96% under stricter regulation over the past year); Fraud Crime Hazard Prevention Act in force 31 July 2024. Malicious AI swarms: Science (Tang, Bostrom, Ressa, Christakis et al.), January 2026. Illustrations by Nicky Case, from the Civic AI "6-Pack of Care" comics. civic.ai

04

Where Safety Comes In

The public conversation about AI safety has revolved around a few frontier systems and the firms that run them. Its recurring subjects are whether those firms can coordinate on shared safety practices, whether a model should be reviewed before it is released, whether autonomous agents can be kept inside the environments built to test them, and whether openly released frontier weights can be contained.[90] These are serious questions. Most of what this report describes is not what they are about, however. A speech model for te reo Māori, a hospital's federated learning run, a data cooperative, a citizens' assembly with a bounded steward do not carry the catastrophic-risk profile the debate concerns. Each is bounded, local, and built to fail small. The alternatives raise safety questions of their own, at open frontier weights, at agents, and inside the small systems this report describes. On oversight, however, they supply what the frontier lacks.

Open frontier weights are where safety concerns are both valid and unsolved. A released model cannot be recalled and its safety training can be fine-tuned away, so an open frontier model may hand capability to actors who could not otherwise obtain it. Evan Miyazono, whose nonprofit Atlas Ignota maps neglected catastrophic risks from AI, offers a clear example from biology. The knowledge that makes a bioweapon is the same knowledge that produces a biology lesson, so a model that can teach one can teach the other, and once its weights are released, there is no way to make it teach only one. His own resolution is that the most advanced systems may not be the ones that need to be open. Most people who own a calculator never need a graphing one; wide access to systems a generation or two behind can serve nearly everyone.[91]

Agents are another layer where the alternative has its own safety concerns. Large numbers of coordinated AI agents can fake a grassroots movement and manufacture the appearance of public agreement. That harm is itself decentralized, and defending against it takes tools that prove who or what is speaking, which can sit uneasily with openness.[92] Agent behavior has been a concern for frontier labs and is also a concern for the alternatives. The identity and registry infrastructure of the agent layer and of Section 3's second case study is one set of answers to those problems.

The small systems this report describes carry a version of the frontier labs' risk, that a model answers only to whoever built it: a model built for one community answers to that community, and a community's preferences can themselves breach a fundamental right. Caroline Green insists that universal rights must be the baseline on top of which community values should be built.[93] How to ensure frontier models or local models adhere to this practice remains an open question.

The alternatives have much to offer when it comes to oversight. Miyazono observes that the safety debate often comes with implicit assumptions about how dangerous AI capabilities will become.[94] The case for concentration is strongest for those who expect that they will become very dangerous. Steven Adler, who led dangerous-capability evaluations at OpenAI and has co-founded the standards nonprofit Guidelight, suggests that fewer companies to convince makes full adoption of a safety practice more plausible, and a small set of actors can more credibly verify that no one is defecting.[95] A large, identifiable provider is also easy to hold to account. Its work is public and heavily scrutinized, and it has a great deal to lose by cutting corners.[96] What that case lacks is a way to check the provider from outside. At present the frontier labs report on themselves, publishing their own safety evaluations, disclosing their own incidents, and setting their own thresholds.[97] The infrastructure described at the inference layer, and in Section 3's first case study, is the alternative to self-reporting. It lets an outside researcher or regulator audit a system without holding the data or the model it audits, and verify what a computation did without seeing what it ran on.[98]

90. The pattern runs through the year's coverage and the field's own reporting. Instances as of September 2026 include the account in the International AI Safety Report 2026 of widening harm pathways from agents and open-weight releases (Yoshua Bengio et al., DSIT 2026/001, February 2026; summarized by Gleb Tsipursky, The Hill, June 2026); reports of agents escaping their evaluation environments during cybersecurity testing (TechCrunch, 9 August 2026); and a frontier lab's chief executive proposing that labs coordinate to slow the pace of capability gains, with safety certifications required at new capability thresholds (Fortune, 14 September 2026).

91. Evan Miyazono, DRC interview, 2026. The saturation point echoes a distinction Steven Adler (DRC interview, 2026) draws: whether powerful AI is dangerous in absolute terms or only relative to the frontier.

92. Daniel Thilo Schroeder, Meeyoung Cha, Andrea Baronchelli, et al., "How Malicious AI Swarms Can Threaten Democracy," Science Policy Forum, 22 January 2026, doi:10.1126/science.adz1697.

93. Caroline Green, DRC interview, 2026.

94. Evan Miyazono, DRC interview, 2026.

95. Steven Adler, DRC interview, 2026; the framing draws on Amanda Askell, Miles Brundage, and Gillian Hadfield, "The Role of Cooperation in Responsible AI Development," arXiv:1907.04534 (2019).

96. Brendan McMahan, DRC interview, 2026, personal capacity.

97. For example, OpenAI's September 2026 disclosure of mostly unreported incidents of its models concealing and fabricating information, published with a new framework for tracking and disclosing such occurrences (Bloomberg, 16 September 2026). The disclosure was voluntary, and the framework followed an incident that journalists had reported before OpenAI did.

98. Audit without custody is the OpenMined work described in Section 3: third-party evaluations run inside a secure enclave with Anthropic and the UK AI Security Institute (then the AI Safety Institute), the cooperative research and development agreement with NIST's Center for AI Standards and Innovation, and the double-blind evaluation reported in William Isaac, Sol Messing, and Kristian Lum, "Piloting the world's first double-blind AI evaluations," Google DeepMind, 27 August 2026, run with the Singapore AI Safety Institute, OpenMined, AVERI, and MLCommons on a Gemini Flash Lite model. Verifying a computation without seeing its inputs is the confidential-computing design described at the inference layer (note 57).

04

Centralized control carries safety risks of its own, to privacy and to what Miyazono calls epistemic security: whether people can still trust what they know. A system a person runs privately is one they can trust; a single system that watches and persuades everyone at once is not.[99] Miyazono worries about this risk more than the others because it gives no warning. A biological or cyber incident can be noticed and can force a response. A corrupted information environment cannot be, because the corruption includes the people who would do the noticing.

Key insights

- Most of the alternative sits outside the safety debate's subject: systems of the kind this report describes do not carry the same catastrophic-risk profile.
- Once a model's weights are released, anyone can copy them and strip out the safeguards. That is a real risk, but it is a risk of releasing frontier-scale weights, and most of the alternative does not need them.
- Agents raise safety concerns for the frontier and for the alternatives alike. The identity and registry infrastructure of the agent layer is one set of answers.
- A system built for one community answers to that community, but universal rights are the baseline on top of which community values are built. How to ensure that remains an open question.
- On oversight, the alternatives provide what a field that reports on itself lacks: external verification.
- Centralized control carries safety risks of its own: a single system that watches and persuades everyone at once, and an operator that is the only judge of its own conduct.

99. Evan Miyazono, DRC interview, 2026.

05

Conclusion

The current narrative of artificial intelligence centers on a handful of companies and assumptions. This report offers an alternative. Not simply a different version of what AI could be, but a different narrative of what the technology should be for and the options available right now. The frontier labs have broken new ground and will continue to do so. Their alternative counterparts now require attention so that funding and policy decisions can spread the benefits of AI as widely as possible.

Those alternatives already exist. A Māori broadcaster that holds its own language in guardianship, a farm cooperative that bargains with Microsoft, a citizens' assembly in Taiwan whose conclusions became law, and a public supercomputer in New York run by the people who use it. They now need what the railroad and the grid needed before them: public will, capital, and rules that recognize them. Those will be the subject of the report that follows in 2027.

The current climate around AI can be contentious and adversarial, and resentments against the technology in its centralized form are growing. We believe in the technology. But we also believe that future iterations of it need to convince people of its worth, its alignment with human needs and values, its ability to complement, rather than inhibit, human choice and agency. Without the decisions that make that possible, the technology will fail in the goals set out for it by centralized and decentralized versions alike: broad human benefit and control of a technology that all agree, at least in theory, should enable human flourishing. Concentrated versions of the technology are already established. It's time to start thinking about how to distribute the benefits of AI.

References

Interviews conducted for this report are listed in the Contributors section. The three case studies in Section 3 were written by their practitioner authors and carry their own inline references. The list below covers the sources cited in the report's notes, in two groups: primary documents and official sources, then published works.

Primary documents and official sources

AI Singapore. SEA-LION model catalog (V4 generation, updated May 26, 2026) and "Uplifting AI in Southeast Asia: Announcing Nemotron-SEA-LION-v4.8 in Collaboration with NVIDIA." sea-lion.ai, September 18, 2026.

Altman, Micah, Sharon Ayalde, Rachel Cummings, et al. *A Community-Driven Differential Privacy Deployment Registry*. NIST Internal Report 8588, initial public draft. National Institute of Standards and Technology, September 2025.

Amazon.com Services, LLC v. Perplexity AI, Inc., No. 26-1444 (9th Cir. August 4, 2026).

Bengio, Yoshua, et al. *International AI Safety Report 2026*. DSIT 2026/001. February 2026.

California Senate Bill 53 (2025). Signed September 29, 2025. Establishes the CalCompute consortium.

COMMUNIA. Response to the European Commission call for evidence on the future of EU copyright. June 19, 2026.

CREATE AI Act of 2025, H.R. 2385, 119th Congress. Reintroduced March 26, 2025 (Obernolte–Beyer), with new authority to accept donated resources from the private sector and federal agencies; reported out of committee June 2026; Senate companion reintroduced 2026.

Empire State Development and the Office of the Governor of New York. Empire AI consortium launch, expansion, and operations releases. April 2024; February and June 2025; August 2026.

ETH Zurich, with EPFL and the Swiss National Supercomputing Centre, under the Swiss AI Initiative. "Apertus: a fully open, transparent, multilingual language model." Press release, September 2025.

European Commission. AI Factories / InvestAI initiative. Announced February 2025.

Federal Farm Loan Act of 1916, 39 Stat. 360.

G20 Digital Economy Ministers' Meeting. Outcome documents on digital public infrastructure. Bengaluru, August 2023.

Homestead Act of 1862, 12 Stat. 392. See National Archives, "Homestead Act (1862)," Milestone Documents; National Park Service, "Native Americans and the Homestead Act," Homestead National Historical Park.

Internet Engineering Task Force. Web Bot Auth (webbotauth) working group; draft-meunier-web-bot-auth-architecture; draft-ietf-webbotauth-httpsig-protocol. Chartered October 2025.

Kingsbury Commitment. Letter from N. C. Kingsbury, American Telephone and Telegraph Company, to the Attorney General of the United States, December 19, 1913.

Primary documents and official sources

Land O'Lakes, Inc., and Microsoft. Multiyear strategic alliance announcement. November 12, 2025.

National Rural Electric Cooperative Association. Comments on the Federal Communications Commission's High-Cost Fund modernization proceeding. August 4, 2026.

National Science Foundation. Mid-Scale Research Infrastructure award, Open Multimodal AI Infrastructure to Accelerate Science (OMAI). August 14, 2025.

National Science Foundation. "Foundations for Operating the National Artificial Intelligence Research Resource: the NAIRR Operations Center (NAIRR-OC)." Solicitation NSF 25-546, posted September 2, 2025.

OECD. *Digital Government Index and Open, Useful and Re-usable data (OURdata) Index*. OECD Working Papers on Public Governance, February 2026.

Okorie, Chijioke I., and Melissa Omino. Nwulite Obodo Open Data License (NOODL). 2024. See also "Addressing Inequitable Openness in Licences for Sharing African Data and Datasets Through the Nwulite Obodo Open Data Licence." *Law, Technology and Humans* 7, no. 3 (2025): 94.

Open Data Watch. *Open Data Inventory (ODIN) 2022/23: Biennial Report*. Open Data Watch, 2023.

Pacific Railway Acts of 1862 and 1864. See National Archives, "Pacific Railway Act (1862)," and U.S. Senate Historical Office, "The Pacific Railway Act of 1862."

Public AI Switzerland (Genossenschaft Public AI). publicai.ch. 2026.

Qualcomm. Investor day presentation and securities filing on the acquisition of Modular. June 24, 2026; completion announced July 29, 2026.

Rural Electrification Administration (est. 1935) and the rural electric cooperative system it financed.

SAG-AFTRA. 2023 Television/ Theatrical Agreement. Ratified 2023.

Taiuru, Karaitiana. Kaitiakitanga Māori Data Sovereignty Licences. 2021.

Te Hiku Media (Te Reo Irirangi o Te Hiku o Te Ika). Kaitiakitanga License. github.com/ TeHikuMedia/Kaitiakitanga-License and tehiku.nz, accessed September 2026.

TrendForce. Foundry industry research. June 2026.

United Nations. Global Digital Compact. Adopted September 2024.

United States v. AT&T. Modification of Final Judgment, 1982; divestiture effective 1984.

The White House. *Winning the Race: America's AI Action Plan.* July 23, 2025. Issued under Executive Order 14179.

Wikimedia Foundation. Draft Annual Plan FY2026–2027. Meta-Wiki, April 2026.

Writers Guild of America. 2023 Minimum Basic Agreement. Ratified 2023.

Published works

Abdin, Marah, et al. “Phi-3 Technical Report: A Highly Capable Language Model Locally on Your Phone.” arXiv:2404.14219, 2024.

Altman, Sam. Remarks at the Xfund Experiment Cup ceremony, Harvard University. May 1, 2024.

Askell, Amanda, Miles Brundage, and Gillian Hadfield. “The Role of Cooperation in Responsible AI Development.” arXiv:1907.04534, 2019.

Beutel, Daniel J., Taner Topal, Akhil Mathur, et al. “Flower: A Friendly Federated Learning Research Framework.” arXiv:2007.14390, 2020.

Bresnahan, Timothy F., and Manuel Trajtenberg. “General Purpose Technologies ‘Engines of Growth’?” *Journal of Econometrics* 65 (1995): 83–108.

Brynjolfsson, Erik, Daniel Rock, and Chad Syverson. “Artificial Intelligence and the Modern Productivity Paradox: A Clash of Expectations and Statistics.” NBER Working Paper 24001, November 2017.

Brynjolfsson, Erik, Daniel Rock, and Chad Syverson. “The Productivity J-Curve: How Intangibles Complement General Purpose Technologies.” *American Economic Journal: Macroeconomics* 13, no. 1 (2021): 333–372.

Carrier, Michael, and Derek Slater. “Worried About AI Monopoly? Embrace Copyright’s Limits.” *Lawfare*, October 27, 2025.

Carroll, Stephanie Russo, et al. “The CARE Principles for Indigenous Data Governance.” *Data Science Journal* 19 (2020).

Chaffer, Tomer Jordi. “On the Institutional Origins of the Agentic Web.” *Harvard Journal of Law & Technology* Digest, November 21, 2025.

Chaffer, Tomer Jordi, et al. “Distributed Legal Infrastructure for a Trustworthy Agentic Web.” arXiv:2603.06884, March 2026.

Cochrane, Daniel, Matt Prewitt, Lacey Strahm, and Salomé Viljoen. *Toward Associational Data Rights: A Digital Economy for Families and the Common Good*. Institute for Family Studies, August 2026.

data.org, with the University of Pretoria, the University of Lagos, and Data Science Nigeria. *Digitisation of Oral Data for NLP of Low-Resource Languages: Practical Methods and Processes for Scalable and Sustainable Ecosystem Development*. February 2026.

DeepSeek-AI. “DeepSeek-R1: Incentivizing Reasoning Capability in LLMs via Reinforcement Learning.” arXiv:2501.12948, January 2025.

Dwork, Cynthia, Nitin Kohli, and Deirdre Mulligan. “Differential Privacy in Practice: Expose Your Epsilons!” *Journal of Privacy and Confidentiality* 9, no. 2 (2019). doi:10.29012/jpc.689.

Eichner, Hubert, Daniel Ramage, Kallista Bonawitz, et al. “Confidential Federated Computations.” arXiv:2404.10764, 2024; revised March 2025.

Elish, Madeleine Clare. “Moral Crumple Zones: Cautionary Tales in Human-Robot Interaction.” *Engaging Science, Technology, and Society* 5 (2019).

Epoch AI. “How far can decentralized training over the internet scale?” December 29, 2025.

Gerstgrasser, Matthias, et al. “Is Model Collapse Inevitable? Breaking the Curse of Recursion by Accumulating Real and Synthetic Data.” arXiv:2404.01413, 2024.

Published works

Gunasekar, Suriya, et al. “Textbooks Are All You Need.” arXiv:2306.11644, 2023.

Hassabis, Demis. Interview with *TIME*. January 2023.

Hu, Botao, and Helena Rong. “Sovereign Agents: Towards Infrastructural Sovereignty and Diffused Accountability in Decentralized AI.” arXiv:2602.14951, February 2026.

Isaac, William, Sol Messing, and Kristian Lum. “Piloting the world’s first double-blind AI evaluations.” Google DeepMind, August 27, 2026.

Ivey, Ronald, Ezra Sullivan, Timothy Lomas, Matthew Lee, James Pawelski, Matthew Sanders, Tyler J. VanderWeele, et al. “Human Flourishing: A Guiding Principle for AI Systems Design?” Presentation to the Pontifical Academy of Sciences, Humanity 2.0 Forum, May 1, 2024. SSRN 5185045.

Keller, Paul. “GPT-NL respects copyright — cui bono?” Open Future, June 26, 2026. First published on the Kluwer Copyright Blog.

Lafont, Cristina. *Democracy without Shortcuts: A Participatory Conception of Deliberative Democracy*. Oxford University Press, 2020.

Laukkonen, Ruben, Seb Krier, Chloé Bakalar, Shamil Chandaria, Morten Kringelbach, et al. “Positive Alignment: Artificial Intelligence for Human Flourishing.” arXiv:2605.10310, June 2026.

Mannan, Morshed, and Nathan Schneider. “Exit to Community: Strategies for Multi-Stakeholder Ownership in the Platform Economy.” *Georgetown Law Technology Review* 5 (2021).

Marivate, Vukosi, Isheanesu Dzingirai, Fiskani Banda, et al. “Mafoko: Structuring and Building Open Multilingual Terminologies for South African NLP.” arXiv:2508.03529, 2025.

Miller, Marshall. “New user trends on Wikipedia.” Wikimedia Foundation Diff, October 17, 2025.

Miranda, Brando, Alycia Lee, et al. “Beyond Scale: The Diversity Coefficient as a Data Quality Metric for Variability in Natural Language Data.” arXiv:2306.13840, 2023.

Nicole, Sarah, Samuel Vance-Law, Connor Spelliscy, and James Bell. *How Can Data Cooperatives Help Build a Fair Data Economy? Laying the Groundwork for a Scalable Alternative to the Centralized Digital Economy*. Project Liberty Institute and Decentralization Research Center, July 2025.

Nissenbaum, Helen. “Privacy as Contextual Integrity.” *Washington Law Review* 79, no. 1 (2004): 119.

Nussbaum, Martha C. *Creating Capabilities: The Human Development Approach*. Belknap Press of Harvard University Press, 2011.

Ohlhaver, Puja. “Community Currencies: The Price of Attention and Cost of Influence in a Networked Age.” SSRN 5136037, January 2025.

Ohlhaver, Puja, Mikhail Nikulin, and Paula Berman. “Compressed to 0: The Silent Strings of Proof of Personhood.” *Stanford Journal of Blockchain Law & Policy*, 2024.

Ong, Isaac, et al. “RouteLLM: Learning to Route LLMs with Preference Data.” arXiv:2406.18665, 2024.

Open Future, with Creative Commons and Proteus Strategies. *Towards a Books Data Commons for AI Training*. April 2024.

Perez, Carlota. *Technological Revolutions and Financial Capital: The Dynamics of Bubbles and Golden Ages*. Edward Elgar, 2002.

Published works

Prime Intellect. "INTELLECT-2: A Reasoning Model Trained Through Globally Decentralized Reinforcement Learning." arXiv:2505.07291, 2025.

Rodrigues, Júlia Martins, and Nathan Schneider. "Exit to Community (E2C): Financing a New Destination for Successful Ventures." Forthcoming; advance manuscript, 2026.

Schneider, Nathan. "An inference cooperative for academic AI." nathanschneider.info, June 14, 2026.

Schneier, Bruce. "AI and Trust." *Communications of the ACM* (2025). doi:10.1145/3737610. First published by the Belfer Center, Harvard Kennedy School, November 2023.

Schroeder, Daniel Thilo, Meeyoung Cha, Andrea Baronchelli, et al. "How Malicious AI Swarms Can Threaten Democracy." Science Policy Forum, January 22, 2026. doi:10.1126/science.adz1697.

Sen, Amartya. *Development as Freedom*. Oxford University Press, 1999.

Shumailov, Ilia, et al. "AI models collapse when trained on recursively generated data." *Nature* 631 (2024): 755–759.

Solaiman, Irene. "The Gradient of Generative AI Release: Methods and Considerations." *Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency*; arXiv:2302.04844, 2023.

Somala, Venkat. "Advanced packaging and HBM, not logic dies, were the bottlenecks on AI chip production in 2025." Epoch AI, March 2026.

Tan, Joshua, Brandon Jackson, Robin Berjon, and Diane Coyle. *Airbus for AI: A Global Strategy for Public Value Creation*. Bennett School of Public Policy, University of Cambridge, 2025.

Tarkowski, Alek, and Felix Sieker. "European Public AI." Policy brief. Open Future and Bertelsmann Stiftung, January 2026.

Taylor, Jacob, and Joshua Tan. "Public AI Is the New Multilateralism." *Project Syndicate*, October 2025.

Tonja, Atnafu Lambebo, et al. "InkubaLM: A Small Language Model for Low-Resource African Languages." arXiv:2408.17024, 2024.

Van Overveldt, Timon, and Daniel Ramage. "Discovering new words with confidential federated analytics." Google Research blog, March 4, 2025.

VanderWeele, Tyler J. "On the Promotion of Human Flourishing." *Proceedings of the National Academy of Sciences* 114, no. 31 (2017): 8148–8156.

Viljoen, Salomé. "A Relational Theory of Data Governance." *Yale Law Journal* 131 (2021).

Wu, Tim. *The Master Switch: The Rise and Fall of Information Empires*. Alfred A. Knopf, 2010.

DRC DECENTRALIZATION RESEARCH CENTER