Why There Is No Wikipedia for Alignment

ANONYMOUS AUTHOR(S)

Wikipedia shows that many people will build a shared resource without pay when they can see it used and govern it together. Efforts to involve the public in shaping how language models behave have not produced such a commons. We ask why with a comparative case study of fourteen efforts from 2022 to 2026 to shape or evaluate model behavior, among them OpenAssistant, Aya, Chatbot Arena, Collective Constitutional AI and other public consultations, paid preference datasets and open platforms, and five projects that sustained volunteer contribution for years: Wikipedia, OpenStreetMap, Community Notes, Mozilla Common Voice and Galaxy Zoo. We assess every case on six ingredients of successful peer production synthesized from peer production and commons scholarship, using 90 public documents, and measure contribution in trace data from six projects. Human feedback has been organized as labor, exchange, campaign, consultation or open platform, and none of these forms is a commons. Low barriers to entry, motivation tied to seeing one's work used, and a goal that is never finished accompany the efforts that lasted; a lasting public artifact, common ownership and self-governance accompany the projects whose contributors kept what they made. No effort to shape AI behavior had both sets. OpenAssistant's volunteers came in bursts, and the project announced its completion within a year; Chatbot Arena's 82 million unpaid votes stayed with an operator now valued at $3.1 billion. We trace the gap to the unit of contribution. A preference label judges one model's outputs, disappears into the model it trains and goes stale when the next model arrives, so it cannot accumulate into an artifact that contributors see, revise and own. We outline a constitution commons of written principles and adjudicated cases, held under share-alike licenses and governed by its contributors. We release the criteria, the evidence for every assessment and the analysis scripts.

1 Introduction

Efforts to involve the public in shaping model behavior have not produced a commons. OpenAssistant drew more than 13,000 volunteers in 2023 to write and rank conversations for an open assistant, then announced that it was completed [61, 65]. Chatbot Arena has collected more than 82 million unpaid votes comparing models and is now a company valued at $3.1 billion that releases a small fraction of them [3, 8, 9]. Developers and nonprofits have convened samples of the public to write rules for models [37, 51], and developers publish their models' specifications in the public domain while keeping sole control of their content [6, 91]. OpenAI's call for democratic inputs to AI named Wikipedia as its foremost inspiration [126]. Why is there no Wikipedia for alignment?

We answer with a comparative case study in the manner of Hill [46], who explained Wikipedia's success by comparing it with other early online encyclopedias. We assess fourteen efforts from 2022 to 2026 to shape or evaluate how language models behave, and five projects that sustained volunteer contribution for years (Wikipedia, OpenStreetMap, Community Notes, Mozilla Common Voice and Galaxy Zoo), against six ingredients of successful peer production synthesized from peer production and commons scholarship [15, 40]. Where projects released records of contribution, we measure it directly.

Our contributions are threefold. We synthesize six ingredients of successful peer production from peer production and commons scholarship, in two groups: ingredients that bring contributors in and keep them, and ingredients that keep what contributors make in their hands (Section 3). We assess fourteen efforts to involve the public in shaping AI behavior against these ingredients, with five commons that succeeded as comparators and trace data where projects released it (Section 4). And from that comparison we derive a design for a commons for alignment and analyze its hard problems. Don-Yehiya et al. [35] drew lessons from peer production for an open pool of human feedback; we test such lessons against the record.

2 Related Work

2.1 Human Feedback and Public Input

Human feedback is a form of data labor that developers buy and rarely disclose. Contractors give most of the judgments used in preference tuning [114], much as crowdworkers produce the other data on which machine learning depends [43, 78]. Firms seldom report this labor, and its omission is rarely treated as a form of closedness [118]. Research on data leverage treats the public's contributions as a source of power over technology companies [116], and finds that the more openly data labor is shared, the less its producers can keep companies from benefiting [69]. We treat paid labeling as one of five forms that human feedback has taken.

Developers now publish how their models should behave, but only the developer edits these specifications. OpenAI's Model Spec and Anthropic's constitution for Claude are in the public domain under CC0 [91], and Anthropic says that Claude turns the constitution into synthetic training data, including rankings of responses [6]. The Model Spec's repository has issues turned off and restricts pull requests [90], Anthropic describes no public way to propose changes, and outsiders contribute through personal access [5, 92]. A public, openly licensed text that shapes training therefore exists, and anyone may fork it. The specifications invite no public contribution, so we do not assess them as cases.

Efforts to give the public a say in model behavior have mostly consulted samples of people about rules. Developers and researchers have asked the public to write constitutions and policies [16, 37, 51, 59, 60]. Reviews of participatory AI find that most projects stop at consultation [29, 34], warn of participation-washing [108] and co-optation [17], and describe a "participatory ceiling" set by developers' control of foundation models [110]. Moats and Ganguly [81] show that OpenAI's program assumed participation must scale and reach consensus on abstract principles. We ask what would let members of the public produce and govern the norms that models follow, and we treat these critiques as standards for the answer.

2.2 Peer Production, Commons and AI

Peer production communities already govern machine learning for their own purposes. Wikipedians curate training data for models that score edits, audit their predictions, and have blocked a bot through their own processes [44], and Wikipedia rebuilt the labeling of such data as wiki pages with discussion [64]. Commons scholarship offers a case-study instrument, the Governing Knowledge Commons framework [40, 75, 109], which has been applied to Wikipedia, open source and Galaxy Zoo [74] but rarely to AI. We apply it to the norms that language models follow.

AI systems also deplete the commons they draw on. The Wikimedia Foundation reports falling human visits and heavy load from crawlers [80, 85], and Stack Overflow's activity fell after ChatGPT's release [33]. Model weights organize communities poorly: activity on a model's Hugging Face repository lasts about six weeks on average [22]. A commons for alignment therefore needs an artifact other than weights, and protection from the kind of reuse that is draining Wikipedia's readership.

2.3 Closest Proposals

Three lines of work come closest to a commons for alignment: open pools of human feedback, case law for AI, and community feedback that trains models. Don-Yehiya et al. [35] draw lessons from Wikipedia, OpenStreetMap and open source for an open, community-governed pool of human feedback that would be a living artifact; their unit remains chats and preferences, which we argue cannot form one. Case law for alignment has been proposed as repositories built by researchers [20, 38], as a jurisprudential argument for combining rules and cases [18], as a public constitution

with state "AI courts" whose rulings train models [1], and as courts inside developers [31]; we place adjudication in a volunteer commons outside any developer or state. Li et al. [68] propose training note-writing models on Community Notes ratings, which they call reinforcement learning from community feedback; that loop starts from a functioning commons, which alignment lacks. Platform-run living constitutions have also been tried [25, 104], as has a commons of community-written evaluations [27], and all three are among our cases.

3 What Made Peer Production Work

We use the definition of Benkler et al. [15]: peer production is open creation and sharing online in which goals are set and pursued in a decentralized way, contributors act on mostly non-monetary motives, and governance is separated from property and contract. Work assigned by a requester, paid by the piece, or owned by a firm therefore falls outside it.[1]

Peer production succeeds when work divides into small modules, integration is cheap, and no one appropriates the result. Volunteers withdraw when an integrator takes the residual value of the joint product or excludes them from using it [13, 14]. Comparing the eight online collaborative encyclopedias active before 2001, Hill [46] credits Wikipedia's success to a familiar product ("just an encyclopedia"), low barriers to contribution and low individual ownership of text, and rejects better technology and timing as explanations. Free software forms what Kelty [55] calls a *recursive public*, one that maintains the technical, legal and conceptual means of its own existence. From this work and from commons scholarship [94, 109] we synthesize six ingredients in two groups, which Section 4 uses to assess each case. We describe each ingredient in turn and then say how we read each case's outcome:

- **Mobilizing ingredients** bring contributors in and keep them.
  (1) **Entry.** Anyone can make a small, useful contribution quickly, and cheap review, redundancy, bots and reverts keep quality up [62]. When 136 wikis began requiring accounts, contributions fell, mostly high-quality ones [49]. Openness is calibrated, since Wikipedia protects many of its most viewed pages [48] and its anti-vandalism tools drive newcomers away [45].
  (2) **Motivation.** Contributors act on fun, learning, ideology, standing and use value, see their work used by an audience, and receive recognition [66]. After China blocked Chinese Wikipedia in 2005, contributors outside mainland China, who could still edit, cut their contributions by 42.8% [128], which suggests that contributors work harder when a large audience can see and use what they make. Randomly awarded barnstars and awards raised effort and retention [42, 100]. Much of open source development is also paid work [101], so pay alone does not rule out peer production. However, it harms peer production when volunteers come to see their work as raising someone else's profit [13].
  (3) **Goal.** The project's goal names a product that contributors already know, such as an encyclopedia, so each of them can say in a sentence what is being made [46]. Short shared norms then say what a good contribution to that product is [99]. We add whether the goal can be finished, because contributors have no reason to continue once it is met. An encyclopedia is never finished, so Wikipedia had more to do, while a dataset of a target size can be.
- **Commons ingredients** keep what contributors make in their hands.

[1] Benkler's 2002 definition was broader and counted harvested judgments such as NASA's Clickworkers and Google's PageRank [13].

(4) **Artifact.** Contributions accumulate in a lasting object that anyone can see, trace and improve, and that coordinates further work [58]. Persistent, public records of work and debate drove the creation and adoption of Wikipedia's rules [115].
(5) **Ownership.** Rights in the individual contributions and in the whole they are assembled into, such as the full encyclopedia or code base, are held in common, through open or share-alike licenses and a community or nonprofit holder [87]. While forking alone does not let contributors correct a project's governance [113], licenses guarantee the right to fork [86].
(6) **Governance.** Contributors make and change the rules, including the rules by which contributions become outcomes, and settle disputes through processes they consider legitimate [39, 94, 102]. Self-governance tends toward oligarchy, as a small group present from the start consolidates authority [105].

We read each case's outcome as its trajectory in Table 1: whether unpaid contribution lasted, and who kept what contributors made. Following Sanfilippo et al. [102], we take each project's purpose from its organizers' words. A project that stopped after meeting its aim therefore ended as planned, and the Goal ingredient records that such an aim ends contribution. Most peer production fails whatever its design; by 2006, 63% of 107,747 SourceForge projects had been abandoned, and in 2010 the median wiki on Wikia had six contributors [103, 105]. Lacking an ingredient is therefore the norm, and the comparison asks which ingredients the lasting projects shared.

4 Method

We adapt Hill's comparative design to a different question. Hill asked why one of eight similar attempts succeeded [46]; we ask why no attempt in a domain has, so we compare across domains.

4.1 Cases and Assessment

We included efforts from 2022 to 2026 that invited the public to help shape or judge how language models behave, and that are documented well enough to assess on each ingredient. We found the efforts through the related literature, the datasets used in preference-tuning research, and searches for public-input processes and volunteer data projects. The fourteen efforts (Table 1) group similar projects, so that the ten projects funded by Democratic Inputs to AI form one case. They are three volunteer data projects (OpenAssistant, Aya and Data Is Better Together); Chatbot Arena, which collects votes from the people who use it; six public consultations (Collective Constitutional AI, Democratic Inputs to AI, OpenAI's collective alignment studies, Meta's Community Forums, the Collective Intelligence Project's Global Dialogues and STELA); two open platforms (Weval and Community Models); paid research datasets such as PRISM and Community Alignment; and a baseline that shows how model developers do this work themselves, paid annotation. Volunteer projects that built pretraining corpora or models, such as LAION-5B, BigScience's BLOOM and EleutherAI's models, fall outside the criteria, since they set no aim for how models should behave, and so do developers' published specifications, which invite no public contribution (Section 2). We chose five comparators that sustained volunteer contribution for years, each resembling human feedback in one respect: Wikipedia, the reference case; OpenStreetMap, a commons that companies came to depend on and join; Community Notes, contested judgments made by volunteers on a platform, which now train an AI model; Common Voice, volunteer training data for machine learning; and Galaxy Zoo, volunteer labels.

Terminology. We sort the efforts into five forms by who contributed, whether they were paid, and who kept and decided what they made. In *labor*, contractors or recruited participants are paid, and the developer or research team

that paid them keeps the product or releases it frozen. In *exchange*, people contribute as a by-product of using a free service, and the service's operator keeps what they contribute. In a *campaign*, volunteers build an openly licensed dataset with a termination point. In *consultation*, a sample of the public that the organizer recruits advises on rules, and the organizer decides how that input becomes rules. On an *open platform*, anyone may contribute without pay to an artifact with no termination point, and the platform's operator sets the rules for what the artifact accepts.

Documents and assessment. We assessed each case as present, partial or absent on each ingredient by criteria for each ingredient. The commons criteria adapt the Governing Knowledge Commons questions [109], and the mobilizing criteria draw on the studies of Section 3. An ingredient assessed as absent describes the project's design and says nothing about whether the project met the purpose its organizers stated. The assessments rest on 90 public documents collected up to 18 October 2026: 35 research publications, 27 announcements, documentation pages and project pages, 21 repository pages, dataset cards and data services, 6 policies, licenses or terms, and 1 commentary. We assigned each assessment from the 90 documents, citing the ones that support it.

4.2 Trace Data

Where projects released records of contribution, we computed the same three measures: contributors per month, active contributors per month, and return rate. We used OpenAssistant's full message export [93], the daily Community Notes data as of 17 October 2026 [125], Common Voice's release statistics [84], Aya's annotator identifiers [24], a released sample of 106,134 Chatbot Arena votes [70], and Wikipedia's monthly counts of registered, non-bot editors of articles [119]. Following Wikimedia's definition of an active editor, an active contributor makes five or more contributions in a month. A contributor returns if they contribute again 30 or more days after their first contribution, counted among those whose first contribution came at least 30 days before the end of the data. The criteria, the assessments with their evidence, and the scripts that reproduce every number are available at **[TODO:** repository link**]**.

5 Findings

Across the cases, human feedback has been organized in five ways (Table 1), and none of them is a commons. The mobilizing ingredients line up with which efforts lasted, and the commons ingredients line up with who ended up keeping what contributors made. No effort to shape AI behavior had both.

5.1 Labor: Feedback as a Purchase

- Developers buy most of their feedback. InstructGPT used about 40 contractors from Upwork and Scale AI, screened partly for agreeing with the researchers [95], and Llama 2's vendor-supplied annotators had to pass a test of agreement with the team [114]. HH-RLHF didn't screen on agreement [12].
- The labels behind InstructGPT and Llama 2 were never released.
- The money is unevenly spread. Workers in Kenya labeling toxic text for OpenAI through a contractor took home $1.32 to $2 an hour [96], while the vendor Surge AI reported over $1 billion in revenue for 2024 [117].
- Research datasets also pay (£9 an hour in PRISM, $28–90 in Community Alignment) and release their data openly [127], but as PRISM's authors put it, they "cannot grant participants the power to change behaviours of deployed LLMs" [57, App. B.9].
- Open pipelines now use AI judges for much of this work [30, 67, 112].
- Labor has neither set of commons ingredients. It lasts as long as someone pays, and the payer owns the result.

Table 1. Cases assessed on the six ingredients of Section 3: ● present, ◐ partial, ○ absent. The first three are the mobilizing ingredients; the last three are the commons ingredients. Developers' annotation covers InstructGPT, HH-RLHF and Llama 2, and research datasets cover PRISM and Community Alignment.

	Mobilizing			Commons			
Case	Entry	Motiv.	Goal	Artif.	Own.	Gov.	Trajectory
Labor – paid contributors, whose work is kept by the payer; releases frozen							
Developers' annotation	○	○	◐	○	○	○	Sustained by payment; labels mostly unreleased
Research datasets	○	○	◐	◐	◐	○	Collected once or in two waves; released frozen
Campaign – volunteers build an open dataset with a termination point							
OpenAssistant	●	●	◐	◐	●	◐	Bursts in 2023; announced completed in October 2023
Aya	●	●	◐	◐	◐	◐	Collected in 2023; unchanged since March 2024
Data Is Better Together	●	◐	◐	◐	◐	◐	Sprints of weeks in 2024–25; completed
Exchange – users of a free service give input that the operator keeps							
Chatbot Arena	●	◐	●	◐	○	○	82 million votes; now a $3.1 billion company
Consultation – a recruited sample of the public advises, and the organizer decides							
Collective Constitutional AI	○	○	◐	◐	○	◐	One deliberation in 2023; research models only
Democratic Inputs to AI	◐	○	◐	◐	○	◐	Ten pilots reported in 2023
OpenAI collective alignment	○	○	◐	◐	◐	○	Repeated studies; six edits to the Model Spec in 2025
Meta Community Forums	○	○	◐	◐	○	○	Recurring since 2022, new sample each time
CIP Global Dialogues	○	○	◐	◐	◐	○	Nine rounds of about 1,000 people since 2024
STELA	○	○	◐	◐	◐	○	Eight focus groups in 2022–23
Open platform – anyone adds to a lasting artifact, under rules the operator sets							
Weval	◐	◐	◐	●	●	◐	Since 2025; pull requests from eight accounts
Community Models	◐	◐	◐	◐	○	◐	Three communities in 2024–25; models little used
Comparators – projects that sustained volunteer contribution for years							
Wikipedia	●	●	●	●	●	●	Sustained since 2001; editors peaked in 2007
OpenStreetMap	●	●	●	●	●	●	Sustained; heavy corporate editing
Community Notes	◐	●	●	●	◐	◐	10,000+ monthly authors since 2023; trains X's AI
Common Voice	●	◐	●	◐	◐	◐	Releases since 2019; inflow at 3% of 2022 peak
Galaxy Zoo	●	●	●	◐	◐	◐	100 million classifications by 2009

5.2 Campaigns: Open Licenses and a Termination Point

- The volunteer campaigns each built an openly licensed dataset, released it, and stopped [24]:
  - OpenAssistant (LAION) set out to "democratize research on large-scale alignment" [61]. Volunteers wrote, labeled and ranked messages for leaderboard points, under Apache 2.0 [65].
  - Aya (Cohere For AI) collected instruction data in 65 languages from 2,997 registered collaborators in 119 countries, with authorship offered to the most active [23, 107].
  - Data Is Better Together ran few-week sprints where several hundred volunteers rated prompts or images [32, 52].
- On their own terms, they achieved their goal: 556 models on Hugging Face list OpenAssistant's first dataset as training data (as of 18 October 2026), and 95.6% of its surveyed contributors were glad they took part.
- OpenAssistant's activity came in bursts (Figure 1). It had 208,584 messages from 25,133 accounts between 16 January and 5 November 2023. The week after its 5 February call for help had 6,195 contributors, and February alone produced 54% of all messages. There was a smaller bump in April after the first models came out, and from June to October it averaged 142 contributors a week, 2.3% of the peak.
- It kept bringing in new people and kept few of them. Half of contributors wrote a single message, 83% were active on only one day, and 4.5% came back 30 or more days after their first message. In the final seven weeks, 92% of active contributors were new.

- Integration, which Benkler names as the bottleneck [13], fell behind as people left. The work was collaborative (96% of replies answered someone else, and the median finished conversation tree had nine contributors), so every departure left something unfinished. Unreviewed messages went from 1% of those written in February to 39% of those written in October, the median wait for a reply went from 14 hours to three or four weeks, and only 2 of the 995 prompts written from June to October became finished trees. Later messages had less time to get reviewed before the export, but the reply waits show the same decline.
- Contributors couldn't go back to their own work. OpenAssistant closed requests to let authors edit their messages, since edits would invalidate the labels already collected on them [41, 53, 111]. Aya allowed peer revision during the campaign, but the release is a frozen snapshot, and the planned volunteer updates never happened [107].
- When OpenAssistant ended, the only thing left was the dataset. Its founder called collecting the data the one task only the project could do and declared it finished; volunteer moderators, programmers and contributors had moved on [56]. Frozen releases can also end up in narrower products: Aya 23, trained on the Aya dataset, is licensed for non-commercial use only [11].
- So the campaigns had low barriers and motivated volunteers, but their goals could be finished, and they had the commons ingredients only partly.

5.3 Exchange: Sustained Contribution, Kept by the Operator

- Chatbot Arena has had the largest volunteer flow of preference labels. Academic researchers launched it in 2023 as "an open platform for evaluating LLMs based on human preferences" [21]: anyone can use frontier models for free, and in return they vote on which of two anonymous models answered better.
- Votes went from about 240,000 in early 2024 to over 82 million by June 2026 [3]. Voting is a by-product of using the service. In a released sample from mid-2024, two thirds of voter IDs appear only once. So contributors don't need to have a sense of ownership for their artifact for contribution to keep going.
- The operator (LMSYS) kept the votes. It committed in 2024 to sharing 20% of vote data, and its September 2026 policy promises "a fraction"; the released vote-level data come to some hundreds of thousands out of more than 82 million [8, 73]. One analysis estimated that the two leading model providers each got about a fifth of all arena data [106], which the operator partly disputes [72].
- Arena became a company in 2025 and in October 2026 raised $200 million at a $3.1 billion valuation [9, 71].
- Exchange got the mobilizing ingredients from offering a free service, and had none of the commons ingredients: contribution lasted, and the product went to the operator.

5.4 Consultation: The Organizer Decides

- In every consultation, the organizer decided who took part, what they were asked, and how their input turned into rules. Anthropic and the Collective Intelligence Project chose Collective Constitutional AI's population, seed statements, moderation criteria, selection rule and final wording [4, 51]. OpenAI's internal review decided which crowd preferences went into its Model Spec [36], and Meta chose its forums' countries and agenda [19].
- Participants were recruited samples, either through survey panels or paid (up to $540 in OpenAI's 2025 study, $75 an hour in STELA) [16, 89]. Most processes ran once [50, 126], and the recurring ones draw a new sample every round [26, 76].

- Some input did reach deployed systems. OpenAI adopted six Model Spec changes from its 2025 study, all still there in August 2026 [92], and Meta says its 2023 forum shaped product defaults [77]. But participants couldn't see or revise what their input became: OpenAI's participants never read the Spec, and OpenAI didn't run its final proposals by them [36].
- The outputs often have no license. Collective Constitutional AI's data do not have one, and the Democratic Inputs code has no license file despite the grant terms saying they would do so [88].
- OpenAI named Wikipedia and Community Notes as inspirations, but it and the other organizers built consultations, which Arnstein [10] placed on the rung of tokenism and which are the most common form in participatory AI [29, 34].
- None of the consultations had either set of commons ingredients.

5.5 Open Platforms: Commons Ingredients, Few Contributors

- Weval, from the Collective Intelligence Project, is a public-domain repository of evaluations that anyone can propose, and it compares its aim to Wikipedia's [27]. Its repository has pull requests from eight accounts [28]. It had the artifact and the ownership, but not the mobilizing ingredients, so owning the commons wasn't enough to draw people in.
- In the Collective Intelligence Project's Community Models, three partner communities voted on principles for chatbots that follow the resulting constitutions. They then barely used the chatbots, and the organizers concluded that people can just prompt a personal assistant to get what they need [25]. Having a say over a model's rules didn't keep the groups involved when the model wasn't useful to them.
- Both platforms had the mobilizing ingredients only partly, and both drew few contributors.

5.6 Recipes for the Comparators' Success

- Wikipedia and OpenStreetMap had all six ingredients. For Wikipedia this came with oligarchy, a gender gap, and quality control that drives newcomers away [45, 47, 105].
  - English Wikipedia's active editors peaked at 53,095 in March 2007 and have stayed between 53% and 67% of that since 2013 [119]. On the same measure, OpenAssistant was under 4% of its peak within four months (Figure 1).
  - OpenStreetMap is a commons that companies joined: paid teams from ten companies, including Apple, Microsoft and Facebook, have made millions of edits, and the foundation responded with rules for organized editing [2].
  - Both have kept contributors for decades, and the product has stayed theirs.
- Community Notes is the closest thing to a commons for AI feedback.
  - Notes show up under posts on X, all notes and ratings are published daily, and an open-source algorithm decides which notes are shown [120, 125].
  - At least 10,000 people have written notes every month since August 2023, and 60% of authors wrote again 30 or more days after their first note, even though you have to earn the right to write by rating first [121].
  - Its judgments now train AI. X post-trains a Grok model on data about which notes people found helpful, and in June 2026 that model wrote 52% of notes rated helpful [124].

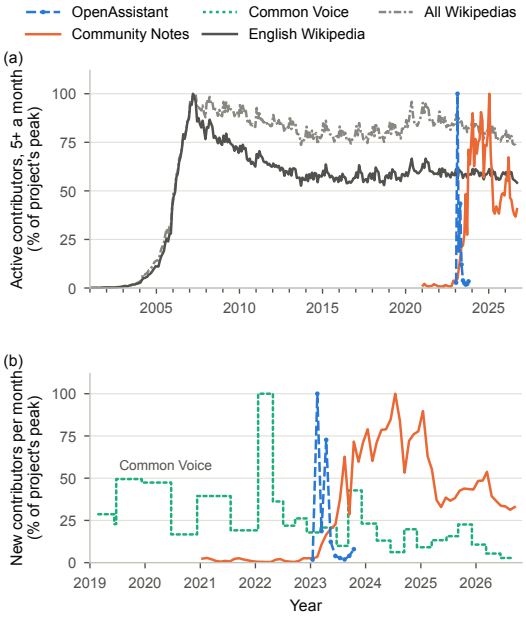


Fig. 1. Contributor activity in volunteer projects, each series as a percentage of its own peak. (a) Contributors with five or more contributions in a month, Wikipedia's definition of an active editor, 2001–2026. (b) New contributors per month, 2019–2026; for Common Voice, the growth in contributors between releases, as a monthly rate. A contribution is a message, a note or an edit, which differ in effort.

  - But X sets the rules, the data have no stated license, and API users can't train foundation models on X content [122, 123]. Monthly note authors dropped from 45,398 in January 2025 to 26,586 in September 2026 as AI-written notes grew [79].
  - It has the mobilizing ingredients and a public artifact, but the platform owns it.
- Common Voice depends on a funded steward. Volunteers have built a public-domain speech corpus of 42,593 hours in 295 languages as of September 2026 [7, 84], and contributors talk about giving their languages a future in technology [54]. Mozilla pays staff and took corporate investment [82], field reports describe tedious work kept going by small material rewards [63], new contributors now arrive at about 3% of the 2022 rate, and since 2025 Mozilla has asked users not to mirror the data [83].
- Galaxy Zoo shows that labeling itself can keep volunteers. By 2009, more than 200,000 people had made over 100 million galaxy classifications [98], mostly to contribute to science [97]. Scientists ran the project, but the data were public, volunteers moderated the forum, and some volunteers became paper authors [74]. Contributors could see what their labels led to.

5.7 Two Patterns of Success, and the Unit of Contribution as Point of Failure

- Every case that kept contributors for years, Chatbot Arena included, had a clear goal that is never finished, plus the other two mobilizing ingredients at least partly (motivation tied to seeing the work used, or for

Arena, a service people wanted). The campaigns ended with their datasets, labor and consultations lasted while organizers paid, and the open platforms drew few people.

- Only Wikipedia and OpenStreetMap had all three commons ingredients. Where an operator held the artifact, as with Arena, contributors' work became the operator's asset. Where the artifact was a frozen release, as with the campaigns, others could take it into narrower products.
- No effort to shape AI behavior had both sets. Community Notes and Weval came closest, each with most of one set and part of the other.
- These are associations, with three limits: the comparators are older and funded differently than the AI efforts; nineteen cases can't separate the ingredients from funding or from what organizers intended; and the trace data show changes in contribution, not their causes.
- The unit of contribution explains why no AI effort had the commons ingredients. Take one OpenAssistant ranking:
  - Rankings train a reward model, which trains an assistant, so nothing in the released assistant shows which rankings shaped it.
  - The trained model belongs to whoever trained it, which is the kind of appropriation Benkler predicts drives volunteers away [13].
  - When the next model writes different replies, the ranking no longer describes what a developer needs judged. This is why Meta collected Llama 2's comparisons in 14 weekly batches [114].
  - If the author of a reply edited it, its rankings would no longer apply, which is why OpenAssistant closed requests for editing (Section 5.2).
- Galaxy Zoo's labels don't have these problems, because they describe stable objects, add up to a public catalogue and lead to visible science. Community Notes ratings don't either, because they judge notes that are part of a public record.
- Labor and exchange use up preference labels, and campaigns collect them until they have enough. A commons needs a unit of contribution that accumulates. Wikipedia's own community came to the same conclusion about labeling and rebuilt it as wiki pages with discussion [64].

6 A Constitution Commons

- Articles and cases, versioned, with discussion pages (meeting the Artifact ingredient)
- A stream of new behavior to adjudicate (meeting the Goal ingredient)
- Contributions of every size; bringing one's own cases; domain projects organized around shared expertise (meeting the Entry and Motivation ingredient)
- Compile reports showing output of the Artifact; public evaluation of every model (meeting the Motivation ingredient)
- Bridging, juries, sampled panels, forks for voice backed by exit (meeting the Governance ingredient)
- Share-alike, dispersed copyright, nonprofit steward (meeting the Ownership ingredient)

References

[1] Gilad Abiri. 2025. Public Constitutional AI. *Georgia Law Review* 59, 2 (2025), 601–670. arXiv:2406.16696

[2] Jennings Anderson, Dipto Sarkar, and Leysia Palen. 2019. Corporate Editors in the Evolving Landscape of OpenStreetMap. *ISPRS International Journal of Geo-Information* 8, 5 (2019), 232. doi:10.3390/ijgi8050232

[3] Anastasios Angelopoulos. 2026. Arena Reaches $100M in 8 Months. Arena blog, https://arena.ai/blog/arena-100m-revenue. Dated 2026-06-29; last updated 2026-07-08. Accessed 2026-10-18.

[4] Anthropic. 2023. Collective Constitutional AI: Aligning a Language Model with Public Input. https://www.anthropic.com/news/collective-constitutional-ai-aligning-a-language-model-with-public-input. Published 17 October 2023. Accessed 2026-10-18.

[5] Anthropic. 2026. Claude's Constitution. https://www.anthropic.com/constitution.

[6] Anthropic. 2026. Claude's New Constitution. https://www.anthropic.com/news/claude-new-constitution. Published 22 January 2026. Accessed 2026-10-18.

[7] Rosana Ardila, Megan Branson, Kelly Davis, Michael Kohler, Josh Meyer, Michael Henretty, Reuben Morais, Lindsay Saunders, Francis Tyers, and Gregor Weber. 2020. Common Voice: A Massively-Multilingual Speech Corpus. In *Proceedings of the Twelfth Language Resources and Evaluation Conference*. European Language Resources Association, Marseille, France, 4218–4222. https://aclanthology.org/2020.lrec-1.520/

[8] Arena Team. 2026. Arena Leaderboard Policy. Arena blog, https://arena.ai/blog/policy. First published 2024-03-02; last updated 2026-09-01. Accessed 2026-10-18.

[9] Arena Team. 2026. Arena Raises $200M Series B at $3.1B Valuation. Arena blog, https://arena.ai/blog/series-b. Dated 2026-10-08; last updated 2026-10-13. Accessed 2026-10-18.

[10] Sherry R. Arnstein. 1969. A Ladder of Citizen Participation. *Journal of the American Institute of Planners* 35, 4 (1969), 216–224. doi:10.1080/01944366908977225

[11] Viraat Aryabumi, John Dang, Dwarak Talupuru, Saurabh Dash, David Cairuz, Hangyu Lin, Bharat Venkitesh, Madeline Smith, Jon Ander Campos, Yi Chern Tan, Kelly Marchisio, Max Bartolo, Sebastian Ruder, Acyr Locatelli, Julia Kreutzer, Nick Frosst, Aidan Gomez, Phil Blunsom, Marzieh Fadaee, Ahmet Üstün, and Sara Hooker. 2024. Aya 23: Open Weight Releases to Further Multilingual Progress. arXiv:2405.15032

[12] Yuntao Bai, Andy Jones, Kamal Ndousse, Amanda Askell, Anna Chen, Nova DasSarma, Dawn Drain, Stanislav Fort, Deep Ganguli, et al. 2022. Training a Helpful and Harmless Assistant with Reinforcement Learning from Human Feedback. arXiv:2204.05862

[13] Yochai Benkler. 2002. Coase's Penguin, or, Linux and "The Nature of the Firm". *The Yale Law Journal* 112, 3 (2002), 369–446. doi:10.2307/1562247

[14] Yochai Benkler. 2006. *The Wealth of Networks: How Social Production Transforms Markets and Freedom.* Yale University Press, New Haven, CT. doi:10.12987/9780300127232

[15] Yochai Benkler, Aaron Shaw, and Benjamin Mako Hill. 2015. Peer Production: A Form of Collective Intelligence. In *Handbook of Collective Intelligence*, Thomas W. Malone and Michael S. Bernstein (Eds.). MIT Press, Cambridge, MA, 175–204. Author preprint: https://mako.cc/academic/benkler_shaw_hill-peer_production_ci.pdf.

[16] Stevie Bergman, Nahema Marchal, John Mellor, Shakir Mohamed, Iason Gabriel, and William Isaac. 2024. STELA: A Community-Centred Approach to Norm Elicitation for AI Alignment. *Scientific Reports* 14 (2024), 6616. doi:10.1038/s41598-024-56648-4

[17] Abeba Birhane, William Isaac, Vinodkumar Prabhakaran, Mark Díaz, Madeleine Clare Elish, Iason Gabriel, and Shakir Mohamed. 2022. Power to the People? Opportunities and Challenges for Participatory AI. In *Proceedings of the 2nd ACM Conference on Equity and Access in Algorithms, Mechanisms, and Optimization (EAAMO '22)*. ACM, 1–8. arXiv:2209.07572 doi:10.1145/3551624.3555290

[18] Nicholas A. Caputo. 2025. Alignment as Jurisprudence. *Yale Journal of Law & Technology* 27, 2 (2025), 390–473. arXiv:2605.08416

[19] Samuel Chang, Estelle Ciesla, Michael Finch, James Fishkin, Lodewijk Gelauff, Ashish Goel, Ricky Hernandez Marquez, Shoaib Mohammed, and Alice Siu. 2024. *Meta Community Forum: Results Analysis, April 2024.* Technical Report. Deliberative Democracy Lab, Stanford University. Deposited in the Stanford Digital Repository in 2025. doi:10.25740/jg174tg0725

[20] Quan Ze Chen and Amy X. Zhang. 2025. Case Law Grounding: Using Precedents to Align Decision-Making for Humans and AI. In *Proceedings of the ACM Collective Intelligence Conference (CI '25)*. ACM, 226–238. arXiv:2310.07019 doi:10.1145/3715928.3737487

[21] Wei-Lin Chiang, Lianmin Zheng, Ying Sheng, Anastasios Nikolas Angelopoulos, Tianle Li, Dacheng Li, Banghua Zhu, Hao Zhang, Michael Jordan, Joseph E. Gonzalez, and Ion Stoica. 2024. Chatbot Arena: An Open Platform for Evaluating LLMs by Human Preference. In *Proceedings of the 41st International Conference on Machine Learning (PMLR, Vol. 235)*. PMLR, 8359–8388. arXiv:2403.04132

[22] Madiha Zahrah Choksi, Ilan Mandel, and Sebastian Benthall. 2025. The Brief and Wondrous Life of Open Models. In *Proceedings of the 2025 ACM Conference on Fairness, Accountability, and Transparency*. ACM, 3224–3240. doi:10.1145/3715275.3732206

[23] Cohere For AI. n.d.. Aya – Prizes! Aya project site, https://sites.google.com/cohere.com/aya-en/prizes. Undated; the site banner says the dataset and model will be released in early 2024. Accessed 2026-10-18.

[24] Cohere Labs. 2024. Aya_dataset. Hugging Face dataset card, https://huggingface.co/datasets/CohereLabs/aya_dataset. Repository created 2024-01-31; data uploaded 2024-02-11; last data fix 2024-03-15; license Apache-2.0. Accessed 2026-10-18.

[25] Collective Intelligence Project. 2025. Community Models. https://blog.cip.org/p/community-models. Accessed 2026-10-18.

[26] Collective Intelligence Project. 2026. Global Dialogues. https://globaldialogues.ai. Project site listing rounds from November 2024 to May 2026. Accessed 2026-10-18.

[27] Collective Intelligence Project. 2026. Weval Application Repository. https://github.com/weval-org/app. GitHub repository and README; code under the MIT license. Accessed 2026-10-18.

[28] Collective Intelligence Project. 2026. Weval Evaluation Blueprints. https://github.com/weval-org/configs. GitHub repository; blueprints dedicated to the public domain under CC0 1.0. Accessed 2026-10-18.

[29] Eric Corbett, Remi Denton, and Sheena Erete. 2023. Power and Public Participation in AI. In *Proceedings of the 3rd ACM Conference on Equity and Access in Algorithms, Mechanisms, and Optimization (EAAMO '23)*. ACM, 1–13. doi:10.1145/3617694.3623228

[30] Ganqu Cui, Lifan Yuan, Ning Ding, Guanming Yao, Bingxiang He, Wei Zhu, Yuan Ni, Guotong Xie, Ruobing Xie, Yankai Lin, Zhiyuan Liu, and Maosong Sun. 2024. ULTRAFEEDBACK: Boosting Language Models with Scaled AI Feedback. In *Proceedings of the 41st International Conference on Machine Learning (PMLR, Vol. 235)*. 9722–9744. arXiv:2310.01377

[31] Nathan Darmon and Tom Reed. 2026. Courts for AI Constitutions. Lawfare, https://www.lawfaremedia.org/article/courts-for-ai-constitutions. Accessed 2026-10-18.

[32] Data Is Better Together. 2024. 10k_prompts_ranked. Hugging Face dataset card, https://huggingface.co/datasets/data-is-better-together/10k_prompts_ranked. Repository created 2024-02-22; license field "other", card says license information needed. Accessed 2026-10-18.

[33] R. Maria del Rio-Chanona, Nadzeya Laurentsyeva, and Johannes Wachs. 2024. Large Language Models Reduce Public Knowledge Sharing on Online Q&A Platforms. *PNAS Nexus* 3, 9 (2024), pgae400. doi:10.1093/pnasnexus/pgae400

[34] Fernando Delgado, Stephen Yang, Michael Madaio, and Qian Yang. 2023. The Participatory Turn in AI Design: Theoretical Foundations and the Current State of Practice. In *Proceedings of the 3rd ACM Conference on Equity and Access in Algorithms, Mechanisms, and Optimization (EAAMO '23)*. ACM, 1–23. arXiv:2310.00907 doi:10.1145/3617694.3623261

[35] Shachar Don-Yehiya, Ben Burtenshaw, Ramon Fernandez Astudillo, Cailean Osborne, Mimansa Jaiswal, Tzu-Sheng Kuo, Wenting Zhao, Idan Shenfeld, Andi Peng, Mikhail Yurochkin, Atoosa Kasirzadeh, Yangsibo Huang, Tatsunori Hashimoto, Yacine Jernite, Daniel Vila-Suero, Omri Abend, Jennifer Ding, Sara Hooker, Hannah Rose Kirk, and Leshem Choshen. 2025. The Future of Open Human Feedback. *Nature Machine Intelligence* 7, 6 (2025), 825–835. arXiv:2408.16961 doi:10.1038/s42256-025-01038-2

[36] Tyna Eloundou, Mitchell Gordon, Eddie Zhang, and Sandhini Agarwal. 2025. Collective Alignment: Public Input on Our Model Spec. https://openai.com/index/collective-alignment-aug-2025-updates/. OpenAI, 27 August 2025. Accessed 2026-10-18.

[37] Tyna Eloundou and Teddy Lee. 2024. Democratic Inputs to AI Grant Program: Lessons Learned and Implementation Plans. https://openai.com/index/democratic-inputs-to-ai-grant-program-update/. OpenAI, 16 January 2024. Accessed 2026-10-18.

[38] K. J. Kevin Feng, Quan Ze Chen, Inyoung Cheong, King Xia, and Amy X. Zhang. 2023. Case Repositories: Towards Case-Based Reasoning for AI Alignment. MP2 Workshop at NeurIPS 2023. arXiv:2311.10934

[39] Andrea Forte, Vanesa Larco, and Amy Bruckman. 2009. Decentralization in Wikipedia Governance. *Journal of Management Information Systems* 26, 1 (2009), 49–72. doi:10.2753/MIS0742-1222260103

[40] Brett M. Frischmann, Michael J. Madison, and Katherine J. Strandburg (Eds.). 2014. *Governing Knowledge Commons*. Oxford University Press. doi:10.1093/acprof:oso/9780199972036.001.0001

[41] gabrielmfern. 2023. Proposal of Improvements Both in the Guidelines of the Dataset Collection and in the Tasks for Dataset Collection. GitHub issue 3185, LAION-AI/Open-Assistant, https://github.com/LAION-AI/Open-Assistant/issues/3185. Opened 2023-05-17; replies from maintainers on 2023-05-18 and 2023-05-24. Accessed 2026-10-18.

[42] Jana Gallus. 2017. Fostering Public Good Contributions with Symbolic Awards: A Large-Scale Natural Field Experiment at Wikipedia. *Management Science* 63, 12 (2017), 3999–4015. doi:10.1287/mnsc.2016.2540

[43] Mary L. Gray and Siddharth Suri. 2019. *Ghost Work: How to Stop Silicon Valley from Building a New Global Underclass*. Houghton Mifflin Harcourt, Boston.

[44] Aaron Halfaker and R. Stuart Geiger. 2020. ORES: Lowering Barriers with Participatory Machine Learning in Wikipedia. *Proceedings of the ACM on Human-Computer Interaction* 4, CSCW2, Article 148 (2020), 37 pages. arXiv:1909.05189 doi:10.1145/3415219

[45] Aaron Halfaker, R. Stuart Geiger, Jonathan T. Morgan, and John Riedl. 2013. The Rise and Decline of an Open Collaboration System: How Wikipedia's Reaction to Popularity Is Causing Its Decline. *American Behavioral Scientist* 57, 5 (2013), 664–688. doi:10.1177/0002764212469365

[46] Benjamin Mako Hill. 2013. Almost Wikipedia: Eight Early Encyclopedia Projects and the Mechanisms of Collective Action. Chapter in *Essays on Volunteer Mobilization in Peer Production*, Ph.D. dissertation, Massachusetts Institute of Technology. Draft revision dated 30 October 2014: https://mako.cc/academic/hill-almost_wikipedia-DRAFT.pdf. Accessed 2026-10-18. Listed by the author as a working paper under the variant title "Almost Wikipedia: What Eight Early Online Collaborative Encyclopedia Projects Reveal about the Mechanisms of Collective Action".

[47] Benjamin Mako Hill and Aaron Shaw. 2013. The Wikipedia Gender Gap Revisited: Characterizing Survey Response Bias with Propensity Score Estimation. *PLoS ONE* 8, 6 (2013), e65782. doi:10.1371/journal.pone.0065782

[48] Benjamin Mako Hill and Aaron Shaw. 2015. Page Protection: Another Missing Dimension of Wikipedia Research. In *Proceedings of the 11th International Symposium on Open Collaboration (OpenSym '15)*. ACM, New York, NY, USA, Article 15, 4 pages. doi:10.1145/2788993.2789846

[49] Benjamin Mako Hill and Aaron Shaw. 2021. The Hidden Costs of Requiring Accounts: Quasi-Experimental Evidence from Peer Production. *Communication Research* 48, 6 (2021), 771–795. doi:10.1177/0093650220910345

[50] Saffron Huang. 2023. ccai: Data Processing for the Collective Constitutional AI Project. https://github.com/saffronh/ccai. GitHub repository of Polis exports and processing notebook; no license file; last push 2023-10-17. Accessed 2026-10-18.

[51] Saffron Huang, Divya Siddarth, Liane Lovitt, Thomas I. Liao, Esin Durmus, Alex Tamkin, and Deep Ganguli. 2024. Collective Constitutional AI: Aligning a Language Model with Public Input. In *Proceedings of the 2024 ACM Conference on Fairness, Accountability, and Transparency*. ACM, 1395–1417. doi:10.1145/3630106.3658979

[52] Hugging Face. 2024. Data Is Better Together. GitHub repository, https://github.com/huggingface/data-is-better-together. Repository created 2024-03-11; README lists community efforts as completed. Accessed 2026-10-18.

[53] IllusionDX. 2023. Option to Edit an Assistant Reply after It's Been Sent. GitHub issue 3031, LAION-AI/Open-Assistant, https://github.com/LAION-AI/Open-Assistant/issues/3031. Opened 2023-05-04; closed as a duplicate of issue 1925 (not planned) on 2023-05-28. Accessed 2026-10-18.

[54] Hillary Juma. 2021. Introducing Common Voice Language Reps 2021/2022. Mozilla Foundation blog, https://www.mozillafoundation.org/en/blog/introducing-cvlr-20212022/. Dated 2021-12-17. Accessed 2026-10-18.

[55] Christopher M. Kelty. 2008. *Two Bits: The Cultural Significance of Free Software.* Duke University Press, Durham, NC. doi:10.1215/9780822389002

[56] Yannic Kilcher. 2023. OpenAssistant Is Completed. YouTube video, https://www.youtube.com/watch?v=gqtmUHhaplo. Published 2023-10-24; 11 min 49 s. Accessed 2026-10-18.

[57] Hannah Rose Kirk, Alexander Whitefield, Paul Röttger, Andrew Bean, Katerina Margatina, Juan Ciro, Rafael Mosquera, Max Bartolo, Adina Williams, He He, Bertie Vidgen, and Scott A. Hale. 2024. The PRISM Alignment Dataset: What Participatory, Representative and Individualised Human Feedback Reveals About the Subjective and Multicultural Alignment of Large Language Models. In *Advances in Neural Information Processing Systems*, Vol. 37. 105236–105344.

[58] Aniket Kittur and Robert E. Kraut. 2008. Harnessing the Wisdom of Crowds in Wikipedia: Quality through Coordination. In *Proceedings of the 2008 ACM Conference on Computer Supported Cooperative Work (CSCW '08).* ACM, New York, NY, USA, 37–46. doi:10.1145/1460563.1460572

[59] Oliver Klingefjord, Ryan Lowe, and Joe Edelman. 2024. What Are Human Values, and How Do We Align AI to Them? arXiv:2404.10636

[60] Andrew Konya, Lisa Schirch, Colin Irwin, and Aviv Ovadya. 2023. Democratic Policy Development Using Collective Dialogues and AI. Report produced as part of OpenAI's Democratic Inputs to AI grant program. arXiv:2311.02242

[61] Andreas Köpf, Yannic Kilcher, Dimitri von Rütte, Sotiris Anagnostidis, Zhi-Rui Tam, Keith Stevens, Abdullah Barhoum, Nguyen Minh Duc, Oliver Stanley, Richárd Nagyfi, Shahul ES, Sameer Suri, David Glushkov, Arnav Dantuluri, Andrew Maguire, Christoph Schuhmann, Huu Nguyen, and Alexander Mattick. 2023. OpenAssistant Conversations – Democratizing Large Language Model Alignment. In *Advances in Neural Information Processing Systems 36 (NeurIPS 2023), Datasets and Benchmarks Track.* Neural Information Processing Systems Foundation, Inc., 47669–47681. Author names follow the arXiv record; Crossref gives "Duc Nguyen" and "Zhi Rui Tam". arXiv:2304.07327 doi:10.52202/075280-2064

[62] Robert E. Kraut and Paul Resnick. 2012. *Building Successful Online Communities: Evidence-Based Social Design.* MIT Press, Cambridge, MA. With Sara Kiesler, Moira Burke, Yan Chen, Niki Kittur, Joseph Konstan, Yuqing Ren, and John Riedl. doi:10.7551/mitpress/8472.001.0001

[63] Jan Krewer. 2023. *Creating Community-Driven Datasets: Insights from Mozilla Common Voice Activities in East Africa.* Technical Report. Deutsche Gesellschaft für Internationale Zusammenarbeit (GIZ) GmbH, FAIR Forward, Bonn, Germany. Content review included the Mozilla Foundation. Accessed 2026-10-18. https://www.bmz-digital.global/wp-content/uploads/2023/03/Creating-Community-Driven-Datasets-Report-032023-GIZ-Mozilla.pdf

[64] Tzu-Sheng Kuo, Aaron Lee Halfaker, Zirui Cheng, Jiwoo Kim, Meng-Hsin Wu, Tongshuang Wu, Kenneth Holstein, and Haiyi Zhu. 2024. Wikibench: Community-Driven Data Curation for AI Evaluation on Wikipedia. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems.* ACM, 1–24. arXiv:2402.14147 doi:10.1145/3613904.3642278

[65] LAION-AI. 2023. Open-Assistant. GitHub repository, https://github.com/LAION-AI/Open-Assistant. README completion note added 2023-11-07 (commit 7558fa8e, PR 3724) and amended 2024-01-06 (PR 3743). Code under Apache-2.0. Accessed 2026-10-18.

[66] Karim R. Lakhani and Robert G. Wolf. 2005. Why Hackers Do What They Do: Understanding Motivation and Effort in Free/Open Source Software Projects. In *Perspectives on Free and Open Source Software*, Joseph Feller, Brian Fitzgerald, Scott A. Hissam, and Karim R. Lakhani (Eds.). MIT Press, Cambridge, MA, 3–22. doi:10.7551/mitpress/5326.003.0005

[67] Nathan Lambert, Jacob Morrison, Valentina Pyatkin, Shengyi Huang, Hamish Ivison, Faeze Brahman, et al. 2024. Tülu 3: Pushing Frontiers in Open Language Model Post-Training. arXiv:2411.15124

[68] Haiwen Li, Soham De, Manon Revel, Andreas Haupt, Brad Miller, Keith Coleman, Jay Baxter, Martin Saveski, and Michiel Bakker. 2025. Scaling Human Judgment in Community Notes with LLMs. *Journal of Online Trust and Safety* 3, 1 (2025). arXiv:2506.24118 doi:10.54501/jots.v3i1.255

[69] Hanlin Li, Nicholas Vincent, Stevie Chancellor, and Brent Hecht. 2023. The Dimensions of Data Labor: A Road Map for Researchers, Activists, and Policymakers to Empower Data Producers. In *Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency (FAccT '23).* ACM, 1151–1161. doi:10.1145/3593013.3594070

[70] LMArena. 2025. Arena-Human-Preference-140k. Hugging Face dataset card, https://huggingface.co/datasets/lmarena-ai/arena-human-preference-140k. Repository created 2025-08-01; card license CC-BY-4.0; prompts CC-BY-4.0 and model outputs under providers' terms. Accessed 2026-10-18.

[71] LMArena Team. 2025. LMArena Is Growing to Support Our Community Platform. LMArena blog, https://blog.lmarena.ai/blog/2025/new-beta/. Dated 2025-04-17. Read via Internet Archive snapshot of 2025-08-09. Accessed 2026-10-18.

[72] LMArena Team. 2025. Our Response to 'The Leaderboard Illusion' Writeup. Arena blog, https://arena.ai/blog/our-response/. Dated 2025-05-09 (originally at blog.lmarena.ai/blog/2025/our-response/); last updated 2025-06-13. Accessed 2026-10-18.

[73] LMSYS Arena Team. 2024. LMSYS Chatbot Arena: Live and Community-Driven LLM Evaluation. LMSYS Org blog, https://lmsys.org/blog/2024-03-01-policy/. Dated 2024-03-01; policy section last updated 2024-05-31. Accessed 2026-10-18.

[74] Michael J. Madison. 2014. Commons at the Intersection of Peer Production, Citizen Science, and Big Data: Galaxy Zoo. In *Governing Knowledge Commons*, Brett M. Frischmann, Michael J. Madison, and Katherine J. Strandburg (Eds.). Oxford University Press, 209–254. doi:10.1093/acprof:oso/9780199972036.003.0007

[75] Michael J. Madison, Brett M. Frischmann, and Katherine J. Strandburg. 2010. Constructing Commons in the Cultural Environment. *Cornell Law Review* 95, 4 (2010), 657–709. https://scholarship.law.cornell.edu/clr/vol95/iss4/10

[76] Meta. 2026. Community Forums at Meta. https://transparency.meta.com/governance/community-forums/. Meta Transparency Center; page updated 5 June 2026. Accessed 2026-10-18.

[77] Meta. 2026. Our 2023 Community Forum on Generative AI. https://transparency.meta.com/governance/community-forums/2023-community-forum-on-generative-AI/. Meta Transparency Center; page updated 16 March 2026. Accessed 2026-10-18.

[78] Milagros Miceli and Julian Posada. 2022. The Data-Production Dispositif. *Proceedings of the ACM on Human-Computer Interaction* 6, CSCW2 (2022), 1–37. doi:10.1145/3555561

[79] Brad Miller, Jay Baxter, Jiansong Chao, Keith Coleman, Sophie Hilgard, and Daniel Ortiz. 2026. Community-Driven API and AI Writer Design for Openly Scaling Community Notes. Submitted 2026-09-30. arXiv:2609.40067

[80] Marshall Miller. 2025. New User Trends on Wikipedia. Wikimedia Foundation, https://wikimediafoundation.org/news/2025/10/17/new-user-trends-on-wikipedia/. Published 17 October 2025. Accessed 2026-10-18.

[81] David Moats and Chandrima Ganguly. 2025. Bringing AI Participation down to Scale. *Patterns* 6, 5 (2025), 101241. arXiv:2407.11613 doi:10.1016/j.patter.2025.101241

[82] Mozilla. 2021. Mozilla Partners with NVIDIA to Democratize and Diversify Voice Technology. The Mozilla Blog, https://blog.mozilla.org/en/mozilla/mozilla-partners-with-nvidia-to-democratize-and-diversify-voice-technology/. Dated 2021-04-12. Accessed 2026-10-18.

[83] Mozilla. 2025. Common Voice Legal Terms. https://commonvoice.mozilla.org/terms/en. Effective 2025-10-31. Accessed 2026-10-18.

[84] Mozilla Common Voice. 2026. Common Voice Datasets. GitHub repository README, https://github.com/common-voice/cv-dataset. Release history and statistics through Scripted Speech v27.0 (cut-off 2026-09-11). Accessed 2026-10-18.

[85] Birgit Mueller, Chris Danis, and Giuseppe Lavagetto. 2025. How Crawlers Impact the Operations of the Wikimedia Projects. Diff (Wikimedia), https://diff.wikimedia.org/2025/04/01/how-crawlers-impact-the-operations-of-the-wikimedia-projects/. Published 1 April 2025. Accessed 2026-10-18.

[86] Linus Nyman and Juho Lindman. 2013. Code Forking, Governance, and Sustainability in Open Source Software. *Technology Innovation Management Review* 3, 1 (2013), 7–12. doi:10.22215/timreview/644

[87] Siobhán O'Mahony. 2003. Guarding the Commons: How Community Managed Software Projects Protect Their Work. *Research Policy* 32, 7 (2003), 1179–1198. doi:10.1016/S0048-7333(03)00048-9

[88] OpenAI. 2023. OpenAI Democratic Input Grant Repository. https://github.com/openai/democratic-inputs. GitHub repository of the ten teams' code and report links; no license file. Accessed 2026-10-18.

[89] OpenAI. 2025. CoVal: Public Input on Model Defaults (Version 2.0). https://huggingface.co/datasets/openai/coval. Hugging Face dataset card; license CC BY 4.0; version 1.0 was titled Collective Alignment 1. Accessed 2026-10-18.

[90] OpenAI. 2025. OpenAI Model Spec. https://model-spec.openai.com/.

[91] OpenAI. 2025. Sharing the Latest Model Spec. https://openai.com/index/sharing-the-latest-model-spec/. Published 12 February 2025. Accessed 2026-10-18.

[92] OpenAI. 2026. Model Spec Changelog. https://github.com/openai/model_spec/blob/main/CHANGELOG.md. In the openai/model_spec repository (CC0 1.0); latest release 2026-08-18. Accessed 2026-10-18.

[93] OpenAssistant. 2023. Open Assistant Conversations Dataset Release 2 (OASST2). Hugging Face dataset card, https://huggingface.co/datasets/OpenAssistant/oasst2. Repository created 2023-12-24; data collected until 2023-11-05; license Apache-2.0. Accessed 2026-10-18.

[94] Elinor Ostrom. 1990. *Governing the Commons: The Evolution of Institutions for Collective Action.* Cambridge University Press, Cambridge. doi:10.1017/CBO9780511807763

[95] Long Ouyang, Jeffrey Wu, Xu Jiang, Diogo Almeida, Carroll L. Wainwright, Pamela Mishkin, Chong Zhang, Sandhini Agarwal, Katarina Slama, et al. 2022. Training Language Models to Follow Instructions with Human Feedback. In *Advances in Neural Information Processing Systems*, Vol. 35. arXiv:2203.02155

[96] Billy Perrigo. 2023. Exclusive: OpenAI Used Kenyan Workers on Less Than $2 Per Hour to Make ChatGPT Less Toxic. TIME, https://time.com/6247678/openai-chatgpt-kenya-workers/. Published 18 January 2023. Accessed 2026-10-18.

[97] M. Jordan Raddick, Georgia Bracey, Pamela L. Gay, Chris J. Lintott, Carie Cardamone, Phil Murray, Kevin Schawinski, Alexander S. Szalay, and Jan Vandenberg. 2013. Galaxy Zoo: Motivations of Citizen Scientists. *Astronomy Education Review* 12, 1 (2013). doi:10.3847/AER2011021

[98] M. Jordan Raddick, Georgia Bracey, Pamela L. Gay, Chris J. Lintott, Phil Murray, Kevin Schawinski, Alexander S. Szalay, and Jan Vandenberg. 2010. Galaxy Zoo: Exploring the Motivations of Citizen Science Volunteers. *Astronomy Education Review* 9, 1 (2010). doi:10.3847/AER2009036

[99] Joseph Michael Reagle, Jr. 2010. *Good Faith Collaboration: The Culture of Wikipedia.* MIT Press, Cambridge, MA. doi:10.7551/mitpress/8051.001.0001

[100] Michael Restivo and Arnout van de Rijt. 2012. Experimental Study of Informal Rewards in Peer Production. *PLoS ONE* 7, 3 (2012), e34358. doi:10.1371/journal.pone.0034358

[101] Dirk Riehle, Philipp Riemer, Carsten Kolassa, and Michael Schmidt. 2014. Paid vs. Volunteer Work in Open Source. In *Proceedings of the 47th Hawaii International Conference on System Sciences (HICSS).* IEEE, 3286–3295. doi:10.1109/HICSS.2014.407

[102] Madelyn Rose Sanfilippo, Brett M. Frischmann, and Katherine J. Strandburg. 2021. Privacy and Knowledge Commons. In *Governing Privacy in Knowledge Commons*, Madelyn Rose Sanfilippo, Brett M. Frischmann, and Katherine J. Strandburg (Eds.). Cambridge University Press, Cambridge, 5–50. doi:10.1017/9781108749978.002

[103] Charles M. Schweik. 2013. Sustainability in Open Source Software Commons: Lessons Learned from an Empirical Study of SourceForge Projects. *Technology Innovation Management Review* 3, 1 (2013), 13–19. doi:10.22215/timreview/645

[104] Ethan Shaotran, Ido Pesok, Sam Jones, and Emi Liu. 2023. Aligned: A Platform-Based Process for Alignment. arXiv:2311.08706

[105] Aaron Shaw and Benjamin Mako Hill. 2014. Laboratories of Oligarchy? How the Iron Law Extends to Peer Production. *Journal of Communication* 64, 2 (2014), 215–238. doi:10.1111/jcom.12082

[106] Shivalika Singh, Yiyang Nan, Alex Wang, Daniel D'souza, Sayash Kapoor, Ahmet Üstün, Sanmi Koyejo, Yuntian Deng, Shayne Longpre, Noah A. Smith, Beyza Ermis, Marzieh Fadaee, and Sara Hooker. 2025. The Leaderboard Illusion. In *Advances in Neural Information Processing Systems 38 (NeurIPS 2025), Datasets and Benchmarks Track*. Neural Information Processing Systems Foundation, Inc., 86910–86964. arXiv:2504.20879 doi:10.52202/085713-2620

[107] Shivalika Singh, Freddie Vargus, Daniel D'souza, Börje F. Karlsson, Abinaya Mahendiran, Wei-Yin Ko, Herumb Shandilya, Jay Patel, Deividas Mataciunas, Laura O'Mahony, Mike Zhang, Ramith Hettiarachchi, Joseph Wilson, Marina Machado, Luisa Souza Moura, Dominik Krzemiński, Hakimeh Fadaei, Irem Ergun, Ifeoma Okoh, Aisha Alaagib, Oshan Ivantha Mudannayake, Zaid Alyafeai, Vu Minh Chien, Sebastian Ruder, Surya Guthikonda, Emad A. Alghamdi, Sebastian Gehrmann, Niklas Muennighoff, Max Bartolo, Julia Kreutzer, Ahmet Üstün, Marzieh Fadaee, and Sara Hooker. 2024. Aya Dataset: An Open-Access Collection for Multilingual Instruction Tuning. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. Association for Computational Linguistics, Bangkok, Thailand, 11521–11567. doi:10.18653/v1/2024.acl-long.620

[108] Mona Sloane, Emanuel Moss, Olaitan Awomolo, and Laura Forlano. 2022. Participation Is Not a Design Fix for Machine Learning. In *Proceedings of the 2nd ACM Conference on Equity and Access in Algorithms, Mechanisms, and Optimization (EAAMO '22)*. ACM, 1–6. arXiv:2007.02423 doi:10.1145/3551624.3555285

[109] Katherine J. Strandburg, Brett M. Frischmann, and Michael J. Madison. 2017. The Knowledge Commons Framework. In *Governing Medical Knowledge Commons*, Katherine J. Strandburg, Brett M. Frischmann, and Michael J. Madison (Eds.). Cambridge University Press, Cambridge, 9–18. doi:10.1017/9781316544587.002

[110] Harini Suresh, Emily Tseng, Meg Young, Mary L. Gray, Emma Pierson, and Karen Levy. 2024. Participation in the Age of Foundation Models. In *Proceedings of the 2024 ACM Conference on Fairness, Accountability, and Transparency*. ACM, 1609–1621. arXiv:2405.19479 doi:10.1145/3630106.3658992

[111] Syphikor. 2023. There Should Be a Way to Tweak the Messages. GitHub issue 1925, LAION-AI/Open-Assistant, https://github.com/LAION-AI/Open-Assistant/issues/1925. Opened 2023-03-02; closed as not planned 2023-06-12. Accessed 2026-10-18.

[112] Team Olmo, Allyson Ettinger, Amanda Bertsch, Bailey Kuehl, David Graham, David Heineman, et al. 2025. Olmo 3. arXiv:2512.13961

[113] Nathaniel Tkacz. 2011. The Politics of Forking Paths. In *Critical Point of View: A Wikipedia Reader*, Geert Lovink and Nathaniel Tkacz (Eds.). Number 7 in INC Reader. Institute of Network Cultures, Amsterdam, 94–109. https://pure.uva.nl/ws/files/1341841/105459_cpov_reader_7.pdf

[114] Hugo Touvron, Louis Martin, Kevin Stone, Peter Albert, Amjad Almahairi, Yasmine Babaei, Nikolay Bashlykov, Soumya Batra, Prajjwal Bhargava, Shruti Bhosale, et al. 2023. Llama 2: Open Foundation and Fine-Tuned Chat Models. arXiv:2307.09288

[115] Fernanda B. Viégas, Martin Wattenberg, and Matthew M. McKeon. 2007. The Hidden Order of Wikipedia. In *Online Communities and Social Computing (OCSC 2007)*, Douglas Schuler (Ed.). Lecture Notes in Computer Science, Vol. 4564. Springer, Berlin, Heidelberg, 445–454. doi:10.1007/978-3-540-73257-0_49

[116] Nicholas Vincent, Hanlin Li, Nicole Tilly, Stevie Chancellor, and Brent Hecht. 2021. Data Leverage: A Framework for Empowering the Public in Its Relationship with Technology Companies. In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency (FAccT '21)*. ACM, 215–227. doi:10.1145/3442188.3445885

[117] Milana Vinn and Krystal Hu. 2025. Exclusive: Scale AI's Bigger Rival Surge AI Seeks up to $1 Billion Capital Raise, Sources Say. Reuters, https://www.reuters.com/business/scale-ais-bigger-rival-surge-ai-seeks-up-1-billion-capital-raise-sources-say-2025-07-01/. Published 1 July 2025. Accessed 2026-10-18.

[118] David Gray Widder, Meredith Whittaker, and Sarah Myers West. 2024. Why 'Open' AI Systems Are Actually Closed, and Why This Matters. *Nature* 635, 8040 (2024), 827–833. doi:10.1038/s41586-024-08141-1

[119] Wikimedia Foundation. 2026. Wikimedia Analytics REST API: Editors by Activity Level (editors/aggregate). https://wikimedia.org/api/rest_v1/metrics/editors/aggregate/. Monthly counts of registered, non-bot editors of content pages, per Wikipedia edition, 2001-01 to 2026-09. Data retrieved 2026-10-18.

[120] Stefan Wojcik, Sophie Hilgard, Nick Judd, Delia Mocanu, Stephen Ragain, M. B. Fallin Hunzaker, Keith Coleman, and Jay Baxter. 2022. Birdwatch: Crowd Wisdom and Bridging Algorithms Can Inform Understanding and Reduce the Spread of Misinformation. Preprint. arXiv:2210.15723

[121] X. 2024. Rating and Writing Impact. Community Notes Guide, https://communitynotes.x.com/guide/en/contributing/writing-and-rating-impact. Page source first committed 2023-05-04, last updated 2024-11-13. Accessed 2026-10-18.

[122] X. 2025. AI Note Writer API. Community Notes Guide, https://communitynotes.x.com/guide/en/api/overview. Page source first committed 2025-07-01, last updated 2026-09-25 (github.com/twitter/communitynotes, documentation/api/overview.md). Accessed 2026-10-18.

[123] X. 2026. Community Notes. GitHub repository, https://github.com/twitter/communitynotes. Repository created 2021-01-25; Apache-2.0; README describes which pull requests are accepted. Accessed 2026-10-18.

[124] X. 2026. The Community Writer. Community Notes Guide, https://communitynotes.x.com/guide/en/api/community-writer. Page source first committed 2026-09-25. Accessed 2026-10-18.

[125] X. 2026. Downloading Data. Community Notes Guide, https://communitynotes.x.com/guide/en/under-the-hood/download-data. Change log entries through 2026-09-18. Accessed 2026-10-18.

[126] Wojciech Zaremba, Arka Dhar, Lama Ahmad, Tyna Eloundou, Shibani Santurkar, Sandhini Agarwal, and Jade Leung. 2023. Democratic Inputs to AI. https://openai.com/index/democratic-inputs-to-ai/. OpenAI, 25 May 2023. Accessed 2026-10-18.

[127] Lily Hong Zhang, Smitha Milli, Karen Jusko, Jonathan Smith, Brandon Amos, Wassim Bouaziz, Manon Revel, Jack Kussman, Yasha Sheynin, Lisa Titus, Bhaktipriya Radharapu, Jane Yu, Vidya Sarma, Kris Rose, and Maximilian Nickel. 2026. Cultivating Pluralism In Algorithmic Monoculture: The Community Alignment Dataset. In *International Conference on Learning Representations (ICLR)*. 147076–147215. arXiv:2507.09650

[128] Xiaoquan (Michael) Zhang and Feng Zhu. 2011. Group Size and Incentives to Contribute: A Natural Experiment at Chinese Wikipedia. *American Economic Review* 101, 4 (2011), 1601–1615. doi:10.1257/aer.101.4.1601