Data Flywheels and Public AI

Nicholas Vincent

2025-09-02

Table of contents

Preface

This is a "mini-book" that discusses "public AI flywheels": software meant to enable people to opt-in to contribute data towards "public AI" causes. The goal of this book is to support efforts to build a transparent, people-centric data collection ecosystem that supports the evaluation and training of public-benefit AI models. If successful, public AI flywheels can create valuable data that materially improves public AI evaluation, research and development. If very successful, these flywheels might also play a role in solving thorny problems around the economics of information in a post-AI age.

This is also a way to organize – and socialize! – some design notes, practical documentation that's out of scope for a single example project's repo, and longer abstract writing on the topic.

This document is organized as such:

- In "Part 1: Concepts", we explore definitions, motivations, and the design space of public AI data flywheels.
- In "Part 2: A Case Study", we discuss one particular implementation of a Minimum Viable Product (MVP) opt-in flywheel meant to accompany a "public AI interface" (hosted interface software that hits various endpoints for "public AI models") that uses a "friendly wrapper around a Git backend" approach
  - This MVP focuses on collecting "notable chats" (good, bad, or interesting). This data provides immediate value for model evaluation and, at scale, can be used for fine-tuning. Importantly, collecting a list of good and bad chats is also immediately fun, so contributors can get some value before we reach a threshold of data volume needed to construct a full benchmark or dataset. We expect key ideas discussed in this doc, and concretized in this project, to generalize to other data types.
  - We also provide details on how a data retention policy for a concrete Public AI Data Flywheel might work, and more generally discuss the role of the data strategy for a "full stack" public AI application: from model endpoints to OpenWebUI interface to flywheel platform.
- The book includes some Appendices with additional information.

Part I

Concepts & Rationale

1 Introduction

> Key insight: a public AI data flywheel is a system that enables a data collection feedback loop that embeds the principles of "public AI" – notably, transparency and accountability.

1.1 What is a data flywheel?

What is a data flywheel? Nvidia gives us this definition: "A data flywheel is a feedback loop where data collected from interactions or processes is used to continuously refine AI models." [1]

In general, a "data flywheel" is a system or set of systems that capture and/or incentivize data. A "flywheel" generally differs from a more general data collection system because the flywheel is embedded into some kind of application (as opposed to e.g. "standalone" data labeling tasks). So, if I just post a Google form to the Internet and say, "Hey, feel free to use this form to send me data!", that's just a form, not a "flywheel".

Generally, most data collection systems lean more towards utilizing either

- "sensor-style collection" (passive, instruments like cameras, microphones, or logging software, all of which lack an active "submit data" step) or
- "form-style collection" (active, requiring somebody to click "submit").

Historically, flywheels tend to imply a passive approach to data collection, but this is not necessarily a requirement. (More on this in Chapter 3).

1.2 What is a public AI data flywheel?

First, what is "public AI"? The public AI network gives us this definition in a whitepaper from (Jackson et al. 2024): AI with

[1] For more examples of blogs on data flywheels, see: (Jason Liu 2024), (Shankar 2024), (Roche and Sassoon 2024).

> "Public Access – Certain capabilities are so important for participation in public life that access to them should be universal. Public AI provides affordable access to these tools so that everyone can realize their potential." "Public Accountability – Public AI earns trust by ensuring ultimate control of development rests with the public, giving everyone a chance to participate in shaping the future." "Permanent Public Goods – Public AI is funded and operated in a way to maintain the public goods it produces permanently, enabling innovators to safely build on a firm foundation."

For more on the public AI concept, see also Mozilla's work (including the web page and paper (Marda, Sun, and Surman 2024)). See also several workshop papers (RegML @ NeurIPS 2023 (Vincent et al. 2023); CodeML @ ICML 2025 (Tan et al. 2025); Workshop on Canadian Internet Policy (Vincent, Surman, and Hirsch-Allen 2025)).

Our focus in this mini-book is building "public AI" flywheels. To summarize heavily – if we try to achieve all the principles laid out in the large body work that tries to define "public AI" (and we should try!), we will face some unique challenges in the implementation of data flywheels.

In building public AI data flywheels, we are trying to create a feedback loop to improve AI by creating and collecting high-quality data (more on this in Chapter 2). However, the public AI principles mean that we likely want to start from a position of very high accessibility and very high accountability relative to other technology organizations and products. This means we need to provide an accessible explanation of exactly what happens to any data a user creates and give people real agency over the shape of the data pipeline. Ideally, public AI builders should also endeavor to make as many components of our stack as close as possible to public goods, which creates challenges around sustaining effort and funding.

Of course, it's worth noting that some particular subset of the broad public (for instance, a particular city or state) could deliberate and make a collective decision that they prefer a more "traditional approach" to data flywheels. Very concretely, we could imagine a state conducting a referendum, and asking the public if they'd like a "public AI" product that follows industry standard practices around data and flywheels (sacrificing some degree of accessibility and/or accountability for other benefits). This might mean that the state deploys an AI chatbot with nearly the same data collection practices and privacy policies as organizations like Google or Anthropic.

In this mini-book, we are taking the stance that it's best to start from a position of leaning heavily towards a highly accessible and accountable flywheel. We start by minimizing data collection and retention ("data minimization"); data that is used directly for AI research and development (R&D) should be provided via an opt-in by highly informed users.

1.3 Core Principles

We can translate the core principles of public AI to the data flywheel domain and arrive at roughly four requirements:

- Transparency for informed consent: Users must be fully informed about the models at play, the organizations who are building models, and the ramifications of any contributions to the flywheel. Ideally, users will also be informed about the training data underlying the models they use. A detailed FAQ and some kind of consent module (ideally going above and beyond standard Terms of Service[2]) are required before any data is shared. To some extent, maximally informed consent will require the active expenditure of resources to improve the public's AI literacy (i.e. we need to build AI literacy focused systems and perhaps even pay people for their attention). We need systems that really do inform people. Luckily, that's something it seems like AI can help with!

- Data Rights: A public AI data flywheel should empower users with control over their data, mirroring GDPR principles and similar regulations (this is also practically important for compliance). This includes the right to access (`Art. 15`), rectify (`Art. 16`), erase when possible (`Art. 17`), and port data (`Art. 20`). One exemplar project we might look to for inspiration around the implementation of data rights and legal terms is Mozilla's Common Voice (Ardila et al. 2019).

  - We note that data rights can conflict with a "fully open" ethos; we will attempt to mitigate these tensions to the best extent possible.
  - We also note that public AI faces some unique challenges with cross-jurisdiction compliance; we discuss this at a high-level later on in Chapter 6.

- Balancing reputation and pseudonymity: To the extent possible, we believe it is valuable to offer people the ability to contribute data with some kind of "real account" attached, so people can earn credit and reputation **if they want to**. But this must be balanced with the benefits of also enabling pseudonymity or even anonymous contribution. See e.g. (McDonald et al. 2019) + blog post, (Tran et al. 2020), (Hwang, Nanayakkara, and Shvartzshnaider 2025)).

  - In our MVP (discussed in Section 2), an account with an OpenWebUI instance is required to make contributions, but users can choose to use a pseudonym (not unique; can for instance be "anonymous"). A hashed user id will be stored for internal purposes, but any public data releases will only use the pseudonym.

- Purpose Limitation & Licensing: Users should be able to specify their preferences for how their data is used (e.g., for public display? for evaluation? for future model training?). This can be captured using (new) IETF AI Use Preferences and Creative Commons

[2] See e.g. Terms we serve with (Rakova, Shelby, and Ma 2023).

Preference Signals, or other approaches that emerge. We will discuss below how this might extend to other preference signal proposals and/or technical approaches to gating data.

– This is critical for answering a likely FAQ around public AI data – if you succeed in creating actually useful training data or new benchmarks, won't private labs just immediately use that data as well?

2 Why collect data?

> Key insights: in general, we want more records that contain high-quality signals and/or observations about the world to be available to public AI organizations for training and evaluation.

If we want to build a data flywheel, it is probably useful to first specify why we want more data! This in turn can help us identify what types of data we want.

At its core, "data" is useful for AI (and for other things!) because it provides information about the world.

In general, it is intuitive that having more information will (generally) lead to better decision-making. [1]. Although there are some scenarios we might come across (or invent) where acquiring information is not helpful – because we might not have "room" in our memory for more data, or some records might not help us at a certain task, or data causes our model to get worse in some sense (some examples of nuance in academic work: (Shen, Raji, and Chen 2024), (Sorscher et al. 2022)) – in general, most people benefit from having more records of high-quality observations and signals (Hestness et al. 2017). [2]

So let's put these more complicated cases aside for now, and make the assumption: in expectation, acquiring more high-quality data (that is "accurate", or reflects "insight") is useful. Oftentimes assessing data's quality, or its truthiness, or its insightfulness, is not at all easy! With this assumption in mind (and hearty caution about the thorniness of truth and insight), we can speak generally about the types of data we might acquire through a flywheel and that data will be useful.

[1] A bayesian might say: data is *evidence* that updates a prior into a posterior via Bayes' rule; the "goodness" of a dataset is how much information (likelihood ratio / bits of surprise) it carries about the hypotheses we actually care about. A frequentist might say: data are *samples* from some process; more (and more representative) samples tighten confidence intervals and reduce estimator variance (roughly with $1/\sqrt{n}$), so sampling design and coverage matter as much as sheer volume.

[2] Classical work provides an information-focused perspective on when/why more data is good: (Wolpert and Macready 2002), (Belkin et al. 2019)

2.1 An overly detailed accounting of all the ways we might generate LLM pre-training data

Speaking at a very low level, LLM pre-training data can come from any sensor or form that creates digital records that contain sequences of tokens. However, we generally don't want any old tokens – we want tokens that contain signals about the world and about people, and that have been organized (typically by people) in a way that captures the underlying structures of our world (or the structures that we people have imposed). In pre-training, it seems we can get away with mixing together many different types of structure. For post-training, we may want specific structure (e.g. data produced by people following specific instructions).

We might further try to describe human-generated data in a very general fashion by saying: data is created when a person does something that leaves a digital trace: typing, speaking into a microphone, using other kinds of digital inputs like buttons, controllers, etc. They might also operate a camera or other sensing instrument that captures signals from the world. We also sometimes may want to use truly "sensor-only data" (e.g., seismic readings), though those sensors are built, placed, funded, and so on by humans.

After typing, a person might use a terminal or GUI to send their inputs into some data structure – by committing code, editing a wiki, responding on a forum, and so on. Often, the person creating a record has a goal and/or a task they want to complete. This might be: ask a question, teach or correct something, build software, file a bug, summarize a meeting, translate a passage, or simply react to some information object (like/flag/skip). Critically, in practice, many high value sources of data also have some upstream social structure and corresponding incentives – institutions, communities, etc. that create meaningful incentives for people to produce records that are accurate, insightful, and so on (Deckelmann 2023), (Johnson, Kaffee, and Redi 2024), (Aryabumi et al. 2024).

In other words, institutions and communities create incentives so that as people type (or otherwise digitize information), they don't just produce random sequences or the same common sequences repeatedly (or we might have an Internet of web pages that all say "I like good food"; don't we all…)

Moving to a more high-level overview, we might begin to categorize LLM training data:

- Human-authored natural language: blogs, books, encyclopedias, news, forums, Q&A, transcripts (talks, meetings, podcasts), documentation, and manuals.
    - And now, some non-human-authored natural language (synthetic versions of any of the above).
- Code: source files, perhaps with licenses and provenance, issue threads, commit messages.
- Semi-structured text: tables, markup, configs (HTML/Markdown/LaTeX/YAML/JSON) that carry schema and relationships.

- Multimodal pairs (for VLM/ASR pretraining): image+text, audio+text, video+text, and associated captions/alignment.
  - Here, the pairing is a critical characteristic that makes this data unique. This implies somebody has looked at the each item in the pair and confirmed a connection (though paired data can be produced in an automated fashion).
- Metadata about data: records that describe characteristics of other records. language, domain/topic tags, timestamps, links, authorship/attribution, license, AI preference signals.
  - Quality signals: dedup scores, perplexity filters, toxicity/PII flags, heuristic or model-based ratings—used to weight or exclude.

Some specific tasks that might create especially useful data include:

- Asking a model a question and marking the response "good" or "fail", optionally with a short note about *why*.
- Corrections/edits: rewriting a wrong answer; adding a missing citation; supplying a step-by-step solution.
- Pairwise preferences: "A is better than B because …" (useful for preference learning/DPO).
- Star ratings / rubrics: numeric or categorical grades on axes like factuality, helpfulness, tone, safety.
- Tagging according to some taxonomy: topic ("tax law"), language ("id-ID"), difficulty ("HS"), license (CC-BY-SA), and AI preference signals.
- Synthetic tasks: user-written prompts + *ideal* references (gold answers, test cases, counterexamples).
- Multimodal: an image with a caption; an audio clip with a transcript; a diagram with labeled parts.
- Programmatic contributions: code snippets with docstrings/tests; minimal reproductions of a bug.
- "Negative" structure: anti-patterns, jailbreak attempts, hallucination catalogs.

Of course, a key data for many AI systems is "implicit feedback": clicks, dwell time, scroll/hover, skips/abandonment. This data is typically collected via a "sensor" (logging software), not something users actively contribute through a form.

3 A Democratic Data Pipeworks

3.1 How does data move from people to AI models — and where can we insert governance levers?

This is a summary of a longer Data Leverage post.

To further motivate the idea of data contribution with public AI principles, it's worth a brief discussion of what the overall "data pipeworks" of the AI industry looks like from a zoomed out view.

Key takeaways

- Modern AI can be understood as a five-stage pipeworks: (1) Knowledge & Values -> (2) Records -> (3) Datasets -> (4) Models -> (5) Deployed Systems.
- Treating AI as a cybernetic system puts feedback and control at the center. Contributors can steer outcomes by shaping data flow (more on the next chapter).
- Human factors dominate AI capabilities because they shape what gets recorded upstream. Interfaces, sensors, and incentives are therefore core AI R&D.
    - some trends may shift this – RL in real life, #todo cite experiental learning
- Properties of data create collective action problems (social dilemmas) that require markets, coalitions, and policy to fix.
- For public AI flywheels, thinking in terms of data pipeworks reveals "insertion points" to add transparency, consent, rights, and preference signals so democratic inputs actually move the system.

3.2 Why a "pipeworks" view?

Most technical AI work zooms in on a clean optimization problem. But questions about who benefits, who participates, and how AI affects society live upstream and downstream of that problem. A "Data Pipeworks" view describes the end-to-end flow by which human activity becomes records, then datasets, then models embedded in systems that act on the world, and thereby change the future data we can collect.

This view pairs naturally with cybernetics/control: identify system state, actuators, sensors, and feedback loops; then decide which loops to strengthen or dampen.

3.3 Five stages of data

1. Knowledge & Values (Reality Signal): Humans (and the physical world) generate the latent "signal" AI tries to model (facts, preferences, norms). We don't presume computability; we note its existence to emphasize sampling implications.
2. Records (Sampling Step): Interfaces and sensors transform activity into structured records (forms, clicks, edits, uploads, buttons, cameras, microphones). Design choices here shape what becomes legible to AI. Key idea: generally, any particular sampling instance either leans more towards "sensor" or "form".
3. Datasets (Filtering & Aggregation): Organizations filter, label, merge, and license records under social, economic, and legal constraints. This determines coverage, bias, and what's even available to learn from.
4. Models (Compression): Learning compresses datasets into input–output mappings. Modeling choices are path-dependent on Stages 1–3; data defines the feasible hypothesis space.
5. Deployed Systems (Actuation): Models are embedded in products, workflows, or infrastructure, producing value and externalities. Deployment feeds back by first, and foremost, **changing the actual world**. Deployment also alters incentives therefore affects future record creation.

Design note: small, well-placed interventions upstream can dominate large downstream tweaks.

3.4 Why this matters for governance and alignment

- Human factors are primary. The distributions the AI field is optimizing over are created, not discovered. Interfaces, defaults, prompts, consent flows, and incentives shape the topology of AI work.
- Social dilemmas are inevitable. Contributing high-quality records to a shared system is a collective action problem (free-riding, failure to reach critical mass). Today's "dictator solution" (opaque scraping) collapses when people gain data agency.
- Data leverage (next chapter) is the steering wheel. Individuals and groups can alter records, licenses, and access. This allows people to steer model behavior by modulating data flow rather than model internals.
- Pluralism becomes measurable. Tracing contributions lets us quantify relative weight of individuals and communities, enabling pluralistic governance and new not

3.5 Where to place the levers (for public AI flywheels)

- Stage 1 to 2 (Knowledge to Records): invest in interfaces and sensors with informed consent; design contribution prompts and micro-tasks; support pseudonymity and reputation choices. Aim to raise signal quality and widen participation. Note that there will be an omnipresent tension between informed consent and "frictionless" contribution. Can be resolved to some degree by building trust between public and public AI operators.
- Stage 2 to 3 (Records to Datasets): attach licenses and AI preference signals per record; validate, de-duplicate, and redact PII; publish partitioned releases. Make rights legible and keep high-trust, high-reuse bundles. Leaderboards, grants, bounties, governance hooks (votes, preferences) to sustain contributions and invite further steering.
- Stage 3 to 4 (Datasets to Models): enable data markets and coalitions, attribution, and sampling weights; build evaluation sets tied to provenance. Align training with community intent and enable bargaining. (more in this in the next section as well).
- Stage 4 to 5 (Models to Systems): publish transparent deployment notes, opt-outs, and model cards tied to data buckets. Surface externalities and set expectations for use.
- Stage 5 to 1 (Feedback loop): try to ensure that AI actually has positive benefits on the world. Improve standards of living, increase health, free-time, well-being etc. so people can become empowered active participants in whatever stage of the pipeline they please.

3.6 Implications for research and practice

Building flywheels is part of a broader agenda to enable a data pipeworks. More in the next chapter on how data contribution through flywheels (including licensed or user-restricted contribution) interplays with data protection, data strikes, markets, etc.

3.7 A compact mental model

- Sensors and interfaces decide what counts.
- Filters and markets decide what persists.
- Compression decides what generalizes.
- Deployment decides what changes next.
- Governance decides who gets to steer.

Public AI flywheels turn that loop into a participatory control system: contributors see consequences, express preferences, and are (hopefully) rewarded for adding high-signal records.

Some useful additional reading that supports these ideas:

- On social dilemmas (Kollock 1998) and collective action theory (Marwell and Oliver 1993)
- On cybernetics ("Cybernetics" 2025)
- on power and progress (Acemoglu and Johnson 2025)
- Technical reference on probabilistic machine learning: (Murphy 2022)
- On influence functions for modern AI systems: https://www.anthropic.com/news/influence-functions
- Reasons to be critical and skeptical: Modeling Complexity (Batty and Torrens 2001), Fallacy of AI functionality (Raji et al. 2022), issues with social simulations (Arnold 2014)
- Viability of technical infrastructures for good data flow: (Fernandez 2023)

4 Flywheels and Bargaining Power

> Key insight: Beyond improving public AI systems, getting public AI data flywheels right can make it easier for people to use data flow as a source of (collective) bargaining power to (1) participate in markets and (2) participate in governance and alignment.

4.1 How can a public flywheel give people real power over AI systems?

Based on Chapter 2, we can arrive at a very obvious argument for a data flywheel: the flywheel will produce data, and that data will make AI better!

But this isn't the only benefit of building flywheels in a "public AI" manner. Doing so can also enhance the amount of agency that people have over data flow, and make "voting with data" possible such that the public has more power to govern and align AI systems.

In short, AI is somewhat unique relative to other technologies, because of its data dependence. Data comes from people. The fact that this powerful technology has a dependency on people from around the world means that AI has a natural "governance lever".

Setting up a public AI data flywheel is thus important not only to improve AI capabilities; success of public AI data flywheels can collectively help to solve some (but not all!) of the thorny governance and alignment challenges that AI poses by fundamentally changing the data pipeworks of AI.

You can read about data leverage via this newsletter or even via this dissertation. For a short summary, of "voting with data to improve alignment", check out this post: Plural AI Data Alignment.

It's worth pulling out two distinct ways that a flywheel can interact with AI and governance:

- A flywheel with no attempt to capture contributor intent or provide data rights may still serve to increase available data, either in fully public repos or in databases accessible by public AI labs. This outcome could still make public models a bit better and help to keep public labs competitive at the margins, but it would not change the bargaining relationship between contributors and model builders.

- A well-governed flywheel that effectively manages the tension between opt-in and friction/ease-of-use can seriously reshape the broader data pipeworks/ecosystem/economy. Ideally this flywheel would also capture provenance, per-item licensing, and per-item AI-use preference (or even enforceable contracts – "you must pay some organization to use this data", or "you must follow this policy around openness, safety, alignment, etc"). Such flywheels would turn contributions into units that can be assembled, priced, withheld, or targeted, opening the door to markets and, if necessary, strikes.

4.2 How an opt-in flywheel enables markets

An opt-in flywheel can create the prerequisites for functioning data markets without turning the project into "just a marketplace."

Critically, on day one of the data flywheel, each contribution is a unit with provenance, license, usage preferences, and minimal schema. There is also the immediate possibility to associate contributions with reputations of contributors or collectives. This is close to something that is legible enough to transact on. While the initial goal would be to promote conscious data contribution towards public AI causes, it is possible that some data contributors could also use the legibility and the organizing effects of the flywheel to also sell some data to private actors. Indeed, a model already exists that enable people to make public contributions that benefit public interest actors while still allowing large private organizations to pay for data contractually: Wikimedia Enterprise. Wikimedia data is open to all, but Wikimedia is able to monetize "enterprise-level access". [1]

As the data flywheel "spins up", a community could form around the open data to build leaderboards, scarcity tags (rare language/domain), and quality scores. This would effectively begin to generate price signals. A bounty board ("need 5k labeled failures in X") would serve to convert demand into targeted supply. An exemplar here would be bounty boards for open source software. While the outputs of such bounty boards are code contributions that become OSS (and thus non-excludable), it's still possible to have market dynamics emerge.

Co-ops/unions/intermediaries can represent contributors, negotiate bundle terms, run audits, and set default preferences. The flywheel provides a starting shared ledger and release cadence that markets need. (Again in some cases, the intermediary may need to "move off" the flywheel and transact directly in a market).

The key idea here is that it's possible to enable market activity under two distinct sets of conditions: one in which data is kept open-but-gated-and-restricted ("markets" for bespoke Wikimedia Enterprise style packages) or by using the flywheel as a stepping stone towards a

[1] That said, there is no doubt that for certain types of data, some people will need prevent their data from ending up in any public repositories in order to monetize effectively. The public AI data flywheel is only suitable for certain categories of data (in short: content that could be at home in a peer produced knowledge commons). Other types of data may be managed by complementary markets and sharing approaches.

more "property-like" market (people organize using the flywheel community or use preference signals as exemplars, then form a data intermediary to collectively bargain directly with data users).

4.3 How an opt-in flywheel enables strikes (or credible refusals)

A data strike here means a coordinated, temporary withdrawal or constraint on high-signal contributions or releases, or retroactive deletion of data (which in some cases, with legal support, could trigger legally enforced retraining https://cyberscoop.com/ftc-algorithm-disgorgement-ai-regulation/ – though TBA on how this will play out in 2025 onwards).

In data flywheel where contribution is opt-in, non-participation is the default. This makes strikes easy. If the processing cadence is public, strikers might even use the release pipeline provides a natural "valve" for cadence changes or strikes. Additionally, Everyone sees dependence on fresh contributions (e.g., evaluation drift), and this visibility creates leverage.

There are many variants of data strikes in a flywheel ecosystem:

- Quality freeze. Contributors keep using systems but withhold labeled "good/fail" chats or corrections for a period.
- Selective embargo. A community with scarce data (language/domain) pauses releases or flips new records to "evaluation-only."
- Preference shift. New contributions change AI-use preferences to deny training unless a stated condition is met (funding, governance, attribution).
- Rate limit. Collectives cap monthly volume to force negotiations on price or terms.

What a strike cannot do (and shouldn't promise):

- Undo past licenses. Items released under irrevocable terms (e.g., CC0, CC-BY) remain available.
- Prevent copying entirely. Public releases can be mirrored; anti-scraping reduces risk but does not eliminate it.
- Guarantee compliance outside the ecosystem. Preference signals work when counterparties agree to honor them or when law/policy backs them.

5 Flywheel design space

> Key insight: There is a broad spectrum of technical implementation of the flywheel, ranging from traditional database-on-a-company-server to low-friction-peer-production (our preferred MVP) to radical approaches (e.g. truly federated data access).

5.1 Purpose of this section

This section gives more context about the many ways we might build flywheels, and lays out alternative governance paths and a future work (in particular, a focus on futures that involve healthy data markets, data intermediaries, federated learning, etc.)

We also discuss why we think an approach that includes a minimal retention frontend + opt-in flywheel platform can serve as a pragmatic bridge to more advanced approaches. For instance, we can use the patterns and concepts discussed here to move towards independently governed data co-ops, eventual federated learning, etc.

5.2 General overview of approaches to flywheels

First, let's lay out a toy model of data "creation" and "flow" (this will come up again Part 2, when we walk through the flow for a real flywheel app).

In Chapter 2 we talked about the numerous combinations of sensors, forms, task settings, social structure from institutions, communities, etc. that might exist, and in the Appendix we discuss a number of formats and types of data for LLMs in particular.

To summarize, an AI developers might collect or use some of the following kinds of data:

- Simple Signal: Binary feedback (/), star ratings, or flags
- Annotated Conversation: Full chat with user corrections, ratings, or notes
- Preference Pair: A/B comparisons between responses
- Examples: User-created prompts and ideal responses
- Structured Feedback: Form-based input (error type, severity, correction)
- Multimodal Bundle: Text + images + voice + metadata
- More advanced structured data …

Further, the creation of data could be prompted at several points in time:

- Proactive: User initiates contribution unprompted (e.g., "Share this chat" button)
- Reactive: System prompts based on signals (e.g., after trigger word, patterns in usage behavior, ask "What went wrong?")
- Passive: Automatic collection with prior consent (e.g., telemetry, browser extension)
- Scheduled: Regular prompts (e.g., weekly "best conversations" review)
- Task-Based: Specific requests for data types (e.g., "Help us improve math responses")

This choice will likely impact the level of "friction" users experience, roughly:

- Zero-Friction: Purely passive
- Almost zero-friction: Purely passive with some regular re-consenting process (monthly or yearly "checkup" on sharing settings)
- Low-Friction: One-click actions with no interruption
- Medium-Friction: Multi-click actions or actions that redirect to separate interface
- High-Friction: Multi-step process, account creation, or technical skills required

Data might also be processed at one or more points in time. In practice, there is likely be some degree of "processing" at various steps, but it is important to clarify this to users. This might involve

- Pre-submission: Client-side processing before data leaves user's device
- On-submission: Real-time processing during the contribution flow
- Post-submission: Batch processing after data is received
- Pre-publication: Review and processing before making data public
- On-demand: Processing happens when data is accessed/downloaded

So, person visits an AI interface (e.g. visits a chatbot product on a website). They sit down, enter a prompt, and then react to the Output (take the information and do something with it, follow up, leave positive or negative feedback, etc.). This is our canonical object of interest: a prompt ("Input"), response ("Output"), and optional follow up data (feedback, more queries and responses, etc.).

Typically, this data must live, for some time, on the user's device. It must also be processed by an AI model ("inference"), which involves sending a payload to a hosted service or some local endpoint (if e.g. user is running open weights on their own device). It may or may not be stored on the server/system (we'll use these interchangeably for now to refer to all the devices controlled by the organization running each module) where the interface is hosted. It may or may not be stored by the server/system where the model is hosted. And finally, a flywheel may send that data to a third location.

This final data could live in a centralized database (e.g. traditional relational database), a public repository (e.g. GitHub, HuggingFace), totally local, or even in some kind of distributed network (IPFS, BitTorrent).

Finally, the resulting flywheel-produced data might be accessed in a number of ways: direct download, API access, a static site with download features, some kind of gated access (using HuggingFace, or other impelmentation), some hybrid of the above, some kind of Wikimedia Enterprise package, etc.

So we have five useful questions for classifying flywheel designs:

- Where data lives: …
- When prompted: …
- When processed: …
- How accessed: …
- Friction level: …

5.3 Must-have elements of a flywheel

Summarizing the above once more, a flywheel designer must:

- choose to either share-by-default or not-share-by-default (note we use this language as "opt in" and "opt out" can actually get confusing in these discussions); if we just say "that app uses opt-in", it can be unclear if that means users are opted in by default or users must opt in to contribute). Is the default behavior upon first sign up or first app use for the initial data to be shared, or not?
  - share-by-default means that the actual infrastructure of the app is *not* decentralized (for a truly decentralized app, e.g. an interface + model I run entirely locally, share-by-default is impossible)
  - not-share-by-default allows for several avenues for opting in
    - App is still centralized, but user toggles a setting. Must trust the operator in this case.
    - Some additional software running on the user's machine does the contributions at user's behest (browser extension that hits API endpoint when user asks it to)
      - But the user is responsible for installing this software
    - User does the sharing manually (user exports and then upload their export; contributions are made via manual peer production-style contributions)
- choose a level of transparency. Does the app interface, the settings page, the browser extension, etc. tell the user exactly what is happening to data and how it might be used?
- choose how to motivate users. What kinds of motivations do UX choices appeal to? Are users paid? Is there a reputation system for contributions?
- choose whether to process contributions. For what: PII? Security concerns? Sensitive content or values conflicts? If yes, choose when to process (on user's device, as part of some "approval process", at regular intervals after data has already been published)

- choose how to share or publish contributions

If designed carefully, many of these flywheel approaches can be integrated together.

5.4 Some Categories of Architectural Models

With all these design choices in mind, it will be useful to describe the general approaches we might take to build a data flywheel.

5.4.1 Standard "PrivateCo" Web App

An obvious option is to simply build a hosted "standard" "PrivateCo" / start-up style web app. If Netflix is successful because of its flywheel, why not just build a public AI data flywheel that looks like a private tech company's product from a technical perspective? Indeed, in some contexts it may make sense share-by-default and simply use the data generated by users directly for R&D (training, eval, etc.).

In this case, the "flywheels" just reads data from the existing production database. While one could argue that the Terms of Service for many existing tech products do mean that users have gone through a kind of informed consent process, there are also serious downsides to the status quo. Many would argue that standard practice in tech (long, difficult to read Terms of Service and Privacy policy documents; opacity about exact details of data collection and usage; general challenges in conveying the complexity of modern data pipelines) make it hard for the standard PrivateCo Web App model to offer truly informed consent for data contribution. (For more on general issues with ToS, see e.g. Fiesler, Lampe, and Bruckman 2016 #todo add more of the "classics" of this genre.)

While some users might even prefer this approach, we believe this would **not** be a good starting place for a public AI data flywheel. We also believe it's important to communicate to users how the public AI interface differs standard practices (for instance, how does a public AI model differ in terms of data use from e.g. using ChatGPT, Gemini, or AI overviews via search).

The defining characteristics of this approach is that data is held by a private entity at all times. Under this approach, we can collect all types of signals, mix proactive and reactive data collection, use telemetry freely, process data whenever we want. It's highly likely under this approach, data from a flywheel would live in centralized, privately governed database.

Answering each of the questions posed above:

- Where data lives: private database
- When prompted: flexible
- When processed: flexible
- How accessed: flexible; likely API

- Friction level: flexible; likely low

It's also likely we would want to follow corporate practices in locking down the final data, which makes this a bad choice for maximizing publicly visible output. Put simply: while an interesting idea in theory, we probably can't run an AI product that has a prod database that is openly readable by the public.

Within the broad umbrella of taking a "Standard PrivateCo Web App" approach to data flywheels, some archetypes might include:

- Telemetry heavy approach (imagine an LLM chat app with no feedback buttons, but lots of data is collected re: dwell time, conversation length, user responses, etc.)
- Feedback heavy approach (imagine an LLM chat app where the UX is heavily focused on asking users to use thumbs up / thumbs down buttons, or presenting users with frequent A/B test responses)
- Hybrid approach

5.4.1.1 Turn sharing off by default, but offer a one click opt in

One approach to building a flywheel that retains some of the benefits of the "traditional" model is to simply ask users to opt in with one click to make all data open for R&D, and perhaps even public. This approach might lead to a smaller pool of users, but a big pool of data from the users who are willing to make the large commitment to opt in.

This would mean the service operates in a privacy maximizing fashion for most users, but for users who opt-in, researchers with access can "read from the production database" as if they were doing research setting (of course, large private organizations do lots of internal security practices and researchers are typically not *literally* reading from prod without multiple stages of approval, anonymization, etc.)

5.4.2 Git/Wiki Platform

Another option to build a "very active flywheel" (that arguably stretches the definition because friction will be very high) is to just use peer production or version control software (a "wiki" like Wikipedia or "git" approach) and just ask people to make their contributions using existing contribution avenues (for instance, "editing" a wiki page or making a "pull request" to a version-controlled git repository).

As a very simple and concrete example, this might mean creating a "flywheel" that starts as a blank GitHub repository or blank Wiki page, with lots of open calls and personal asks for people to make "pull requests" to just "stick some data into the repo". Over time, if people decide to contribute, you might end up with some high quality content in the repo, though this is likely to very dependent on who contributes, how motivated they are, etc.

If we choose this approach, we do likely constrain our answers to the above design questions:

- Where data lives: Public repository
- When prompted: Proactive (user initiates)
- When processed: Pre-submission (user does it) + optional CI/CD validation
- How accessed: Probably via direct download (e.g. download raw data from GitHub), options to add web interface
- Friction level: High (technical knowledge required)

This approach has maximum transparency, built-in versioning, and low cost. But, it is likely to exclude non-technical users and has very high friction even for technical users.

Example Stack: some combo of MediaWiki, GitHub, GitLab, HuggingFace + CI/CD validation

5.4.2.1 Export-based approach

A relate idea is to build a flywheel that leverages existing export features and export mechanisms. Instead of adding feedback buttons or telemetry, flywheel designers could simply create a static site that lets users manually upload exported data from various apps. This would require manual effort (and some friction could be reduced via careful attention to UX, adding features to help standardize data, etc.) but could be powerful in jurisdictions with portability/export rights.

This could lead to something that looks very much like a peer production process, but with a heavily simplified set of actions (primarily, just export, perhaps with some filtering or curation, and upload).

5.4.3 Web app with intuitive actions that "wrap" Git-style contributions

The option described in Part 2 of this mini-book is to use a Git/Wiki approach, but build a web app with features that allow users to take low friction in-app actions (clicking a special button, entering special command, etc.) that writes to a Wiki / Git repo on the contributor's behalf. We could also build a system so that users can effectively commit data to the source control / wiki system automatically (e.g., "Every day, run an anonymization script on my chat history and then write the output as a new file to a shared, version-controlled server"). In other words, these are ways to "contribute data as if it were a GitHub PR or a Wikipedia edit" without having to learn the exact interfaces of GitHub or Wikipedia.

- Where data lives: Public repository
- When prompted: Proactive or reactive
- When processed: On-submission via serverless function
- How accessed: Git access + static site generation

- Friction level: Low (automated complexity)
- Pros: Transparency + usability, serverless scaling
- Cons: Cold starts, API rate limits, complex error handling
- Example Stack: Vercel/Netlify + Hugging Face API

5.4.4 Browser Extension

We could also implement a flywheel that relies on users downloading a browser extension. This only reflects a data ingestion choice: can be used with various backend choices above.

The browser extension could faciliate: - direct contributions to a database - wrapper on top of git-style actions (i.e. instead of a web app that helps users write to a git-repo via the HuggingFace API, have a browser extension that does so!)

5.4.5 Federated Learning Model

One radically different approach might involve building a flywheel that contributes information via a federated learning. In this world, the flywheel is not actually about sharing raw data directly, but instead about sharing model weights.

- Where data lives: User devices (distributed)
- When prompted: Passive with consent
- Information object: Model gradients or aggregated statistics
- When processed: Pre-submission (on-device)
- How accessed: Only aggregated model updates available
- Friction level: Zero after setup
- Pros: Maximum privacy, no data transfer, infinite scale
- Cons: Complex implementation, limited debugging, device requirements

5.4.6 Other experimental approaches

Other approaches to building a flywheel might involve more radical approaches to decentralizing the actual data storage, for instance using peer to peer protocols, various crypto/web3 approaches to data sovereignty, etc.

5.5 Frontier approaches: data cooperatives, federated learning, and more

#todo: this could be made crisper.

In many cases, users may want to have data governed by community organizations (e.g., organized by domain/region/language) that hold rights and decide release cadence, licensing defaults, and benefit policies.

Practically, taking a collective/intermediary focused approach has the potential to massively reduce user friction / attention costs. One vision for a low friction data intermediary approach is: users spend some time, say, once a year choosing which intermediaries to join. Upon joining, they can choose to delegate key decision-making and participate in intermediary governance as suits their desires and needs. If joining process is good + governance is good, can achieve good outcomes!

We note that if an implementation of the flywheel is built on top of open-source software, communities can easily choose to deploy their own instance and their own data flywheel and effectively operate entirely parallel, self-governed instances. If they also choose to share opt-in data via similar licensing and preference signal approaches, such datasets could be easily merged – but with fine-grained adjustments to precise details (e.g., slight modifications on retention, access, release cadence, content moderation, adn so on.) Of course, data co-ops may choose to use quite different technical stacks. This approach is just one among many.

5.5.1 Interactions between these flywheels

#todo

5.5.2 Transitioning for opt-in flywheel to federated learning

It may be possible to also move from an opt-in data flywheel approach to a federated learning-first approach. Here, model training occurs across user or institutional nodes; only gradients/updates (with privacy tech) are centralized. The dataset remains partitioned or local; central custodian minimized. This approach would:

- Reduces central data custody and breach surface
- Aligns with data-residency and institutional constraints
- Enables "learning from data that can't leave"

But has some major downsides / existing barriers:

- Harder reproducibility and data auditability
- Complex privacy stack (secure aggregation, DP, client attestation)
- Benchmarking must be redesigned (federated eval)

This is a bigger leap, but we believe it's important to begin to think about how the implementation of the Public AI Data Flywheels might support communities wishing to transition towards an FL approach.

One rough sketch might look like: * Build the MVP defined in Chapter 2 * Ship license + AI-preference metadata (MVP). * Maintain gated HF releases and public leaderboards/full data access. * Publish provider-payload transparency and link to provider terms (no guarantees). * Process deletions via HF mechanisms when possible; keep our mirrors in sync. * Phase 1 — Co-op pilots * Charter one or two community co-ops; define bylaws, scope, and release cadence. * Spin up many instances of interface + flywheel combos (can fork software directly, or use similar approaches) * Establish a concrete sharing / merging plan * And beyond! * Once several independent data communities, are operated, it might be possible to move from lightweight sharing and merging to more serious federation with technical guarantees. Perhaps this might start with federated evaluation and then move to federated training. Much more to do here, out of scope for this document.

6 Ethics and Compliance

Public AI data flywheels face numerous ethics and compliance challenges.

This mini-book does NOT provide specific legal advice. We do discuss and link to terms of service used by various platforms.

In Part 2, we provide some examples of platform specific data policy terms.

6.1 Ethics

6.1.1 Flywheel-particular challenges

There is a large literature on harms from AI and sociotechnical systems more generally. We provide a longer set of references at the end of this section.

The top ethics priority for a Public AI Data Flywheel (PAIDF) is figuring out informed consent, and balancing consent and friction. One worst case scenario for a public AI organization is that the flywheel is set up in a way that erodes user trust and ultimately hinders the broader public AI mission.

While designing ethical systems normally involves some degree of multiplicity (there is rarely a single "most ethical solution" for a given group of people), our overall stance is that informed consent can be achieved by maximizing user information about data use and taking a fundamentally opt-in approach.

Beyond consent, a number of other interesting ethics challenges arise. We describe them first, and then discuss the intersection between building an ethical flywheel and a compliant flywheel.

In particular, there are three flywheel specific concerns, that primarily stem from the very general nature of modern AI data.

First, it is possible that data that is contributed via the flywheel could create serious security concerns (contributing a chat that includes an injection attack). Second, data that is contributed could create privacy concerns (PII and sensitive strings, from email, names to API keys). And third, data that is contributed may be seen as expressively harmful or leading to representational harms. That is, some users might produce data that is very offensive to other

users. This is likely inevitable in a large enough system, and so public AI flywheel designers must plan with values conflict in mind.

In short, when we open up a form to the world, people may enter things (even in good faith) that creates security risks, violates privacy, or violates social norms.

There are also a set of ethical risks that arise from downstream AI systems that we build/improve with flywheel data. While these are not the focus of this mini-book, it is critical to keep them in mind. A non-exhaustive list includes:

- allocative harms: outputs affect access to opportunities or resources (moderation, ranking, credit scores)
- privacy harms at the model layer (distinct from data layer): re-identification, doxxing, accidental leakage of personal or sensitive data
- security harms (distinct from data layer): prompt injection and data exfiltration via model behavior; poisoning of training or eval sets
- IP and contract harms: misuse of copyrighted or licensed content; violations of platform terms
- AI-driven expressive harms: a system produces content that demeans, stereotypes, or legitimizes abuse against some group
- AI-driven representational harms: skewed data makes groups invisible or mischaracterized (e.g., images that underrepresent darker skin tones; code comments that assume a single gender)

6.1.2 Flywheel-specific high-level goals

To balance these ethical challenges, we might organize our design around high-level goals that often appear in AI regulation and ethical discussions. These might include "purpose limitation" (European Union 2016) (our flywheel should try to collect only data that is necessary for the stated task – evaluating and improving AI systems) and "proportionality" (we should weigh utility of data collection against the likelihood and severity of harm; to some extent, because the flywheel leans opt-in, some decision-making is delegated to contributors). Considering the more general set of AI harms above, we may also want to specifically acquire or filter data in a way that helps achieve fairness goals.

Typically, you will see works attempt to classify high-risk data which should be treated differently. Examples include:

- faces, voices, gait, or other biometrics
- images of minors or contexts involving schools and hospitals (Federal Trade Commission 2013; U.S. Department of Education 1974; U.S. Department of Health and Human Services 2000).
- intimate or medical contexts, support forums, addiction and mental health groups
- government IDs, financial records, geolocation trails, and precise timestamps

- credential artifacts: API keys, cookies, session tokens, SSH keys, access logs
- content from communities with clear norms against scraping or model training

A flywheel designer likely wants to avoid collecting this kind of data, but getting 100% precision will be nearly impossible, because some of the most interesting AI outputs (especially failure cases) may involve high-stakes scenarios. A flywheel that completely bans contributions related to cybersecurity or human health risks collecting "excessively bland" data.

6.1.3 Levers for solving these ethics challenges

The flywheel designer can use several avenues for attempting to pre-empt some of the above challenges. In terms of informed consent, this comes down to the implementation of a usable, informative module for consent and the exact UX for opting in and out. In terms of security and privacy, this mainly comes down to implementing filtering/curation at various stages. In terms of values conflict, the designer may employ filtering, but also take a normative or sociotechnical approach (leaning on peer production-style talk pages, moderation, community-generated rules, etc.).

The designer has the least leverage to directly control downstream model harms, but can have some influence via further training data filtering, helping to document data produced by the flywheel, etc.

6.2 Compliance

In general, data protection regimes impose responsibilities on anyone operating a platform.

Most likely, any public AI data flywheel will also be connected to some frontend (e.g., hosted OpenWebUI instance) and some backend (model provider). These distinct systems are likely to have their own data-related responsibilities, depending on exactly how they hold or process data.

6.2.1 Risks

In terms of compliance risks, some issues may emerge because of contributor mistakes: users may post personal data that evades whatever filtering/curation the designer has implemented. In some way, PII, secrets, or identifiers may make it into the flywheel's data repo. Further, even when users make contributions via pseudonym, unique phrasing, contextual clues, etc. can deanonymize. If we use salted contributor hashes, these are still stable identifiers across contributions which creates some small risk as well.

In general, a major risk with an approach that creates publicly accessible data is the potential for permanence via forks and mirrors. Removed data can persist in external forks, local

clones, or third-party mirrors outside this project's control. Further, while repo history can be rewritten and monthly files reissued, but downstream models may already have trained; unlearning is best-effort and not guaranteed.

Risks may also stem from the use of various vendors. Hosting providers (e.g. Vercel and similar services, any caching databases uses, any APIs used) may retain request logs; this could be outside the flywheel designer's control.

In some cases, contributions that create "security-related ethical risks" (e.g. a chat in which an LLM provides instruction for conducting some kind of attack) could also create compliance risks. This creates some continuous burden on maintainers. The same is true of offensive content or privacy violations. Even with consent and public repos, some jurisdictions treat certain content types as sensitive or restricted.

6.3 Further reading:

First: works that taxonomize harms (Shelby et al. 2023; Weidinger, Mellor, et al. 2021; Blodgett et al. 2020)

Allocative harms: outputs affect access to opportunities or resources (moderation, ranking, credit-like inferences) (Barocas and Selbst 2016; Obermeyer et al. 2019).

Works that discuss expressive harms and representative harms (Shelby et al. 2023; Weidinger, Mellor, et al. 2021; Buolamwini and Gebru 2018; Grother, Ngan, and Hanaoka 2019; Crawford and Paglen 2019; Blodgett et al. 2020).

On data that has actual security concerns (contributing a chat that includes an injection attack) (OWASP 2023; Hubinger et al. 2024; Carlini et al. 2024).

On PII and sensitive strings (from email, names to API keys) (Carlini et al. 2019, 2021).

Reg and legal examples: (McCallister, Grance, and Scarfone 2010; European Union 2016; Illinois General Assembly 2008; "Rosenbach v. Six Flags Entertainment Corp." 2019).

- privacy harms: re-identification, doxxing, accidental leakage of personal or sensitive data (Sweeney 2000; Narayanan and Shmatikov 2008)

Further reading on:

- proportionality: weigh utility against the likelihood and severity of harm (*ISO/IEC 23894:2023 Information Technology—Artificial Intelligence—Risk Management* 2023; NIST 2023).
- respect for context: treat data according to the social norms of its origin community (Nissenbaum 2004; Jo and Gebru 2020).

- transparency: explain collection, uses, and the limits of control in clear language (Mitchell et al. 2019; Gebru et al. 2018; Holland et al. 2018).
- accountability: assign owners, metrics, and escalation paths (NIST 2023; European Union 2024).
- fairness and non-discrimination: measure and mitigate disparate impacts (Barocas and Selbst 2016; Selbst et al. 2019; Obermeyer et al. 2019; Bender et al. 2021).
- security harms: prompt injection and data exfiltration via model behavior; poisoning of training or eval sets (OWASP 2023; Carlini et al. 2024).
- IP and contract harms: misuse of copyrighted or licensed content; violations of platform terms (U.S. Copyright Office 2024; Creative Commons 2023).

7 Upstream data and data contribution

Data flywheels / contribution pathways are one part of the broader "data strategy" for an AI product or organization. Another key factor is making the full public AI pipeline is transparent is telling users about upstream data. Typically, the terms of service for an application or flywheel try to tell users where the data will go; but it can also be useful to tell users about where the data/AI come from.

7.0.1 AI builder attribution

At a high-level: in each interaction between users and a public AI system, we want to attribute the organization who did the hard work of prepping a model. Ideally, we also want to attribute the original data creators, though in some cases practical constraints make this hard.

- The custom text, branding, etc. within an AI interface can provide organization- specific, with the goal of making sure all model builders are happy. Can even highlight other interfaces/endpoints, something private AI systems are less likely to do.
- Important to get this right so that model developers don't "back out" of the inference MVP and just switch to their own sovereign interfaces

7.0.2 Data attribution

Another way that public AI platforms can differentiate themselves from private AI is by heavily emphasizing data attribution. This might involve showing users data cards, incorporating features like OlmoTrace (Jiacheng Liu et al. 2025), etc.

7.1 Why does upstream matter?

Telling users about upstream data is a key part of system-wide transparency. Transparency on both fronts (model builders, data) has the potential to provide further incentive to users to provide data in the first place (because, e.g., they specifically want to support one of the organizations providing models or data).

There are a number of other exciting conncetions between data valuation/attribution, collective action in data (algorithmic collective action, data leverage), and flywheels.

Part II

Case Study: Low friction peer production

8 The OpenWebUI Action MVP

8.1 Overview

Our first working flywheel is deliberately simple: an OpenWebUI Action lets people opt in to share selected chats directly from the interface, and those contributions become pull requests to a HuggingFace dataset. By building on OpenWebUI's accounts and controls, we avoid asking contributors to learn a new product or workflow, while still anchoring the data pipeline in a transparent source-control backend.

The design goal is to keep friction low without abandoning provenance. Git-backed contributions provide an auditable history, clear authorship models, and a natural venue for discussion and review. At the same time, the action abstracts the mechanics of branching and committing so the contribution experience feels like a normal chat, not a dev tool.

We experimented with two architectures on the way to this MVP. The earlier attempt staged submissions in a private "waiting room" repository and processed them asynchronously. It offered control but made it hard for contributors to see their impact. The current approach, which raises a pull request per contribution, lets contributors and reviewers observe exactly what was added and why, and it makes the curation process legible to the public.

8.2 Components

Three pieces make the flywheel work in practice. The chat frontend at https://chat.publicai.co/ provides the familiar place where conversations happen. An OpenWebUI action, implemented in Python, packages a contribution when a user chooses to share, pulling in the conversation, the user's persistent preferences, and some basic metadata. Finally, the action talks to a HuggingFace dataset repository with a write token and opens a pull request that can be triaged manually or by scripts. In effect, the web app functions as a thin wrapper over "make a pull request with a typed JSON payload," but that thin wrapper is precisely the difference between a usable experience and a developer-only workflow. Power users who prefer to operate in the open can always submit direct pull requests from their own HuggingFace accounts; the action does not preclude that path.

8.3 Contribution Flow

First-time setup is intentionally lightweight. A user signs in to OpenWebUI, opens the Controls (Valves, Functions), and enables sharing. In the same place, they select a default license (for example CC0-1.0, CC-BY-4.0, or CC-BY-SA-4.0), choose an AI preference signal such as `train-genai=n;exceptions=cc-cr`, and decide how they want to be named in public artifacts: their username, a custom pseudonym, or a generic "anonymous." Users who prefer to skip extra prompts can also opt into automatic feedback collection.

With preferences set, contributing looks like any other chat until the moment of consent. After finishing a conversation with any model, the user clicks the Public AI Data Flywheel action. The action shows a concise confirmation screen that restates the current settings, previews the exact content to be shared, and explains what will happen next. The user can edit optional feedback or add context, then confirm. The action assembles a single, well-typed JSON record: the messages, a summary of the model and usage metadata, the selected license and preference signal, the chosen attribution, and a random contribution id. That record becomes the body of a pull request to the dataset repository.

The pull request is the central coordination object. Reviewers can comment, merge, or request changes; automated checks can validate format and run policy scanners; and anyone can see the history that led to the decision. If the contribution is accepted, the data lands in the public dataset with the same license and preference signals the user selected. The PR thread itself preserves the reasoning and any follow-up, which becomes part of the public provenance trail.

8.4 Privacy, Attribution, and Preference Signals

We want people to get credit when they want it, and privacy when they do not. Each contribution carries an attribution string drawn from the user's chosen setting. Public releases display this attribution string but not the underlying OpenWebUI account id. Today, if you select "pseudonym," that pseudonym is a deterministic function of your OpenWebUI account id (unsalted), which means contributions under a pseudonym can be linked to each other over time. We chose determinism to make it easy to accrue credit, support light-weight moderation, and keep analytics simple without exposing account identifiers. If you want the least linkability, choose "anonymous."

Legal terms and downstream use are encoded explicitly. The action stores the contributor's default license and the AI-use preference signals they select. Preference signals, such as `train-genai=n`, are not a silver bullet, but they provide a machine-readable expression of intent that downstream consumers can honor in tooling and policy. This makes it straightforward to build filters, gates, and dashboards that keep training-only or evaluation-only subsets separate, and it answers a recurring question about public datasets: what prevents private

labs from silently absorbing everything? Preference signals and licensing do not prevent misuse on their own, but they make honoring the public's choices the easiest path for responsible actors—and they create structure for accountability conversations when that trust is broken.

Preset options (Content-Usage expressions using CC signals as exceptions): - Training family: - `train-genai=n` (deny training) - `train-genai=n;exceptions=cc-cr` (deny training unless Credit) - `train-genai=n;exceptions=cc-cr-dc` (deny training unless Credit + Direct Contribution) - `train-genai=n;exceptions=cc-cr-ec` (deny training unless Credit + Ecosystem Contribution) - `train-genai=n;exceptions=cc-cr-op` (deny training unless Credit + Open) - General AI use family: - `ai-use=n` (deny AI use) - `ai-use=n;exceptions=cc-cr` (deny AI use unless Credit) - `ai-use=n;exceptions=cc-cr-dc` (… unless Credit + Direct Contribution) - `ai-use=n;exceptions=cc-cr-ec` (… unless Credit + Ecosystem Contribution) - `ai-use=n;exceptions=cc-cr-op` (… unless Credit + Open)

Default preset: `train-genai=n;exceptions=cc-cr` (deny training unless Credit).

8.5 What We Tried First (and Why We Changed It)

The earliest prototype wrote contributions to a private area named `_waiting_room/` and processed them in batches. A periodic job validated files, ran PII scanning, moved clean items forward, and quarantined anything suspect to a private area. The model mirrored a traditional ingestion pipeline and felt safe and controlled, but it carried an important cost: contributors could not immediately see that they had made a contribution, nor could they link to it, discuss it, or watch it progress. In a public AI context, those shortcomings matter. Visibility is not just a nice-to-have; it is how people learn what the system values and how it behaves.

The pull-request-based MVP retains the benefits that mattered—validation and quarantine are still possible as part of PR checks—while restoring legibility. In practice, the contribution becomes a public artifact immediately, but one that can be refined before it becomes part of the canonical dataset. This change turned out to be the simplest way to align usability with accountability.

8.6 Safety and Review

Contributions pass through a few layers of basic safeguards. The action can run in a mock mode for testing without sending anything upstream. Automated checks scan for common types of sensitive information—emails, phone numbers, government identifiers, payment instruments—and flag or block contributions that appear risky. When there is doubt, reviewers can ask for edits in the PR or move the contribution to quarantine for a closer look. OpenWebUI's existing rate limiting helps keep abuse manageable without inventing a new throttle.

The goal is not to promise perfect redaction, but to reduce the chance that obviously sensitive material lands in the public dataset. The confirmation screen and the surrounding documentation set expectations clearly: do not share private or confidential content, and prefer synthetic or anonymized examples when in doubt. This guidance, coupled with visible review in PRs, encourages a culture of care without putting the entire burden on automation.

8.7 Why This MVP Matters

Flywheels live or die by their first loop. An MVP that lets people see and understand their impact builds early momentum and attracts contributors who care about quality. Pull requests give us a natural unit of credit, discussion, and iteration; they also create space for lightweight governance. Over time, we can add small, high-leverage improvements—rubrics for labeling, better previews, richer preference signals—without changing the mental model. People chat, opt in, and their contribution shows up where the public can review it.

This approach also sets up the rest of the pipeline. Because every artifact is a typed JSON with explicit license and use preferences, it is straightforward to derive evaluation sets, training splits, or dashboards that track coverage. And because the data moves through PRs, the same infrastructure can support community-led benchmarks, model audits, and discussions about edge cases. The infrastructure is simple on purpose, but it is pointed at the right problems: provenance, participation, and practical control over data use.

9 The Life of a Chat

The goal of this Section is to provide a worked example of what exactly happens to one of your chats in a system like the public ai utility interface.

First, what happens when you visit the site?

- You visit publicai.co
- You sign up
- Either you create a username and password (these are stored in our database) OR you sign in with Google
- Either way, you now have an account. This "account data" is stored in database, but it's never available to outside partners. It's just stored there so you can log in, keep track of your chats, etc. On other platforms your account might also be linked to third party data, used to create an advertising profile, etc. (Although we're not trying to be scary here: AI-focused companies like OpenAI generally treat your account data the same as we do for now. It's companies that also sell ads or other products where things get a bit more fuzzy).

Ok, now you have an account. Here's what it looks like in our database:

- #todo: screenshot what an example chat in a database look like. let users see "admin view" (using fake data, of course!)

Now you start a new chat!

- You write a "Prompt" (aka an "Input")
  - This might be: "Tell me something similar about Switzerland and Singapore". Let's use this as our running example
  - Note that in LLM systems, you have a lot of freedom when you write your prompt, so you might accidentally or intentionally include personal information or otherwise sensitive information
  - While we provide some tools (and are working on shipping more tools) to help with this, there's no right or wrong answers as to whether a chat is 100% "safe" or "intended". Perhaps you *want* to tell the AI that you're a man located in western Canada; you probably don't want to give it your credit card number (though there are exceptions if you want an agent to buy something on your behalf).

  - So now you have a prompt. It's just what computer scientists call a "string" – it's a bunch of text

- If you're using the Inference Utility, we send your Prompt to a "compute endpoint" so you can get a corresponding output

  - Basically, your prompt goes "into a machine" and an "Output" comes out the other end
  - The inference utility takes advantage of donated compute from a number of organizations
  - In some cases, we send it directly to the AI operator's compute cluster for "inference"
  - In other cases, we have a copy of the AI operator's "model weights" on our own compute and we send the prompt there
  - Exactly what we do depends on the model you're using, how many others are using the Utility at the same time, etc.
  - In general, we don't want you to have to worry about this, but if you're interested it's all open source

- So now the AI model takes your "Prompt"

  - To learn more about exactly what happens – it's more or less just a LOT of number crunching – check out 3Blue1Brown videos
  - We send the Output back to you!

- Your chat history keeps track of all the Input/Output pairings, organized by "Chats"

  - By default, nobody outside of the Utility Interface can see your chats. We only access them internally as needed to operate the service and investigate security or legal issues.
  - We may share high-level, aggregate statistics about usage (for example, total volume or broad topic distributions) without exposing individual chat content.
  - You can opt in to two separate programs: (1) Researcher Access, which allows vetted research partners to analyze your chats for public-interest evaluation and R&D without making them public; and/or (2) the public Data Flywheel, which lets you contribute specific chats to a public repository under a license you choose. These are independent choices—you can enable either, both, or neither.

- You can delete Chats at any time

- You can export all your chats to store them somewhere else, pass them to other AI products, etc.

10 Terms and Privacy

For reference, this chapter include the full text of the the Terms and Privacy document for publicai.co (Sep 12, 2025.)

10.1 1. Terms of Service

10.1.1 1.1 About these Terms

When we say "Company", "we", "our", or "us" in this document, we are referring to the nonprofit 501(c)(3) **Metagov Inc**.

When we say "Services", we mean our web application, available via the website https://chat.publicai.co/ and our API Gateway Developer Portal.

When we say "Partners", we mean the external organizations that provide the AI models that our Services use.

When we say "You" or "your", we are referring to the people that own an account with our Services.

We may update these Terms of Service ("Terms") in the future. We will share past versions of our Terms. Whenever we make a significant change to our policies, we will refresh the date at the top of this page and notify users of the Services directly via the website.

When you use our Services, now or in the future, you are agreeing to the latest Terms. There may be times when we do not exercise or enforce a right or provision of the Terms; however, that does not mean we are waiving that right or provision. These Terms do contain a limitation of our liability.

If you violate any of the Terms, we may terminate your account. That's a broad statement and it means you need to place a lot of trust in us. We do our best to deserve that trust by being open about who we are, how we work, and keeping an open door to your feedback.

10.1.2 1.2 Account Terms

You are responsible for maintaining the security of your account and password. The Company cannot and will not be liable for any loss or damage from your failure to comply with this security obligation.

You are responsible for all content posted to and activity that occurs under your account, including content posted by and activity of any users in your account.

You must be a human. Accounts registered by "bots" or other automated methods are not permitted.

You must be at least 18 years old to use this Service.

10.1.3 1.3 Description of the Service

The Services include hosted applications, demos, and other AI-powered features that let you submit prompts and other content ("Inputs") and receive generated responses ("Outputs"). The Services are designed primarily for research and educational use and are not a replacement for human decision-making. Your Inputs and related metadata may be transmitted to external AI providers you enable to generate Outputs.

10.1.4 1.4 Responsible Use Guidelines

Use the Services responsibly. They are intended for research and education and not as a substitute for professional judgment. In addition to applicable law, you agree not to use the Services for the following disallowed use cases (examples are illustrative, not exhaustive):

- Violence and threats: Incitement or encouragement of violence; promoting self-harm; sexual exploitation (including sexualization of minors); hate speech.
- Antisocial and antidemocratic uses: Harassment or doxing; insensitive content targeting victims; intentionally sowing division; perpetuating harmful stereotypes; attempts to characterize identity (e.g., inferring race or gender); graphic sexual or torture depictions; political manipulation.
- Deceit: Fraud, phishing, or evading the law; spam; misrepresentation (e.g., passing automated content as human without disclosure); misinformation that causes harm.
- Security or privacy attacks: Spearphishing; creating or disseminating malware; attempts to extract personal information or defeat model safeguards.
- Unsafe automation: Posting to social media without human oversight; systems that hide that content is generated by AI.
- Decision-making without human review: AI-based social scoring by public authorities; systems making consequential decisions about people (credit, employment, education, housing, insurance, legal, or medical) without adequate human-in-the-loop review.

- Other harms: Manipulative redirection of attention; tools for plagiarism or academic dishonesty; political campaigning or lobbying.

10.1.5 1.5 Third-Party Providers

Our Services may route Prompts and related data to external providers. When using our Services, you will need to select specific AI models and AI model providers to use. In some cases, your Prompt will be processed via a subprocessor like a cloud services organization that provides compute resources. In other cases, your Prompt will be sent directly to an AI developer to produce an output. You are responsible for reviewing and complying with applicable third-party terms. We do not control providers' retention, training, or other uses once data is sent to them.

Our AI model providers and cloud service providers are listed below under "Subprocessors".

10.1.6 1.6 Availability and Changes

Our Services may change, suspend, or discontinue at any time without liability. We may update features, models, or integrations periodically. Your use of the Services is at your sole risk. We provide these Services on an "as is" and "as available" basis. We do not offer service-level agreements (SLAs) for our Services.

10.1.7 1.7 IP; Inputs and Outputs

As between you and us, and to the extent allowed by law, you own your Inputs. Subject to your compliance with these Terms, you own the Outputs you generate. Outputs may not be unique; the same or similar Outputs may be generated by others. You grant us a limited, non-exclusive license to use, reproduce, and process your Inputs and Outputs as needed to provide, secure, and improve the Services (and, if you explicitly opt in, to contribute to the Data Flywheel).

10.1.8 1.8 Feedback License

If you provide feedback, you grant us a perpetual, worldwide, royalty-free license to use it to improve the Services.

10.1.9 1.9 Enforcement and Appeals

We have the right to suspend or terminate your account and refuse any and all current or future use of our Services for any reason at any time. Suspension means you will not be able to access the account or any content in the account. Termination will furthermore result in the deletion of your account or your access to your account, and the forfeiture and relinquishment of all content in your account. We also reserve the right to refuse the use of the Services to anyone for any reason at any time.

10.1.10 1.10 Export Controls and Sanctions

You must comply with applicable export, re-export, and sanctions laws. You represent you are not prohibited from receiving the Services.

10.1.11 1.11 Disclaimers

The Services, including any AI features, are provided "as is" and "as available" for research and educational use. Outputs may contain errors or appear authoritative while being wrong. To the maximum extent permitted by law, we disclaim all warranties (including merchantability, fitness for a particular purpose, and non-infringement). Use independent judgment and verify important results.

Outputs are for informational purposes only and are not a substitute for professional advice. The Services do not provide medical, legal, financial, or other professional services. Do not rely on Outputs without independent verification by a qualified professional.

You will not use the Services in high-risk environments where failure could lead to death, personal injury, or environmental or property damage (including medical diagnosis/treatment, autonomous vehicles, critical infrastructure, or weapons).

Do not use any Output as the sole source for decisions that have legal or material effects on a person (including credit, employment, education, housing, insurance, legal, or medical decisions).

10.1.12 1.12 Limitation of Liability and Indemnity

You expressly understand and agree that the Company shall not be liable, in law or in equity, to you or to any third party for any direct, indirect, incidental, lost profits, special, consequential, punitive or exemplary damages, including, but not limited to, damages for loss of profits, goodwill, use, data or other intangible losses (even if the Company has been advised of the possibility of such damages), resulting from: (i) the use or the inability to use the Services; (ii) the cost of procurement of substitute goods and services resulting from any goods, data,

information or services purchased or obtained or messages received or transactions entered into through or from the Services; (iii) unauthorized access to or alteration of your transmissions or data; (iv) statements or conduct of any third party on the service; (v) or any other matter relating to these Terms or the Services, whether as a breach of contract, tort (including negligence whether active or passive), or any other theory of liability.

To the extent permitted by law, our aggregate liability will not exceed the greater of USD $100 or the fees paid by you to us for the Services in the 12 months before the event. The foregoing does not limit liability for willful misconduct or where not permitted by law. Exclusive venue and governing law: Commonwealth of Massachusetts.

You agree to indemnify us against claims arising from your use of the Services, your content, or violation of these Terms or laws.

10.1.13 1.13 Governing Law and Dispute Resolution

These Terms are governed by **Massachusetts** law, excluding its conflict-of-laws rules. Disputes will be resolved in **Massachusetts**.

10.1.14 1.14 Notices of Infringement (DMCA)

If you believe content available through the Services infringes your rights, please submit a notice under our DMCA/notice-and-takedown process to our designated contact. Valid notices should include sufficient detail to identify the material and your rights claim. We may notify the affected user and, where appropriate, allow a counter-notice.

10.2 2. Privacy Policy

This Privacy Policy explains how we collect, use, share, and protect information when operating the Services.

If you do not agree with the terms of this Privacy Policy, you must not access or use any Services.

10.2.1 2.1 Scope and Roles

"Metagov" is the organization that acts as the "controller" of personal data; external providers act as separate controllers under their terms.

10.2.2 2.2 Data We Process

We collect: - **Account data**: Email, username, profile settings. Optionally, you may store your name and birth date. - **Chat data**: Prompts, outputs, and any other content you upload such as text, files, images, and audio. You may also create metadata about a chat, such as "feedback" or "tags". - **Communications**: If you communicate with us, such as via email, we may collect personal data such as your name, contact information, and the contents of the messages you send. - **Log Data**: We may collect information that your browser or device automatically sends when you use our Service. This may include the details about your device, your IP address, your browser type, browser settings, and the date and time of your request. - **Usage Data**: We may collect information about your use of the Services, such as the features you use and the actions you take, your time zone, country, the dates and times of access, and your user agent and version. - **Cookies**: We use cookies to improve your experience with our Services. If you use our Services without creating an account, we may store some of the information described in this policy with cookies, for example to help maintain your preferences across browsing sessions. - **Account Information**: We offer the ability to log in using your Google account, if you elect to do so. If you log in with your Google account, the personal data we collect may depend on your Google privacy settings.

If, during your use of our Services, you provide us with any personal data relating to a third party (e.g. your spouse, children, parents, and/or friends), by submitting such personal data to us, you represent to us that you have obtained the consent of such third party to you providing us with their personal data, and for the collection, use and disclosure of their personal data for all purposes set out herein and by or for the benefit of the persons referenced herein.

Our backend uses the source-available "OpenWebUI software", and exact database implementation details are available via the OpenWebUI source code.

10.2.3 2.3 Special Program: Data Flywheel (Opt-In)

With consent, you can submit contributions (implemented as "pull requests" to a Hugging Face dataset repository) to a public dataset under open licensing. Withdrawal stops future collection, but past public releases generally cannot be fully retracted. The legal basis for flywheel participation is consent, with legitimate interests for safety where applicable. An in-product consent module explains operator, storage, license, AI-use preferences, irreversibility, and rights.

Licenses: when you contribute, you must choose a license for each item: `CC0-1.0`, `CC-BY-4.0`, or `CC-BY-SA-4.0`. Your chosen license governs downstream use of that public contribution.

Attribution: public releases display an attribution string you select in-product (your username, a pseudonym, or "anonymous"). If you choose a pseudonym, today it is a deterministic, unsalted value derived from your OpenWebUI account id; this enables consistent credit and moderation but makes contributions under the same pseudonym linkable to each other. If you

prefer the least linkability, choose "anonymous." We do not publish your underlying account id.

AI preference signals: you may attach AI-use preferences (e.g., "do not train") to each public contribution. These are machine-readable expressions of intent that responsible downstream users should honor; they are not a technical enforcement mechanism.

See more in the Program-Specific Terms for the Flywheel.

10.2.4 2.3A Researcher Access (Additional Opt-In)

Separately from public contributions, you may enable "Researcher Access" so vetted research partners can analyze your chats for public-interest evaluation and model development without making your chats public.

- Scope: when enabled, Researcher Access covers chats in your account that are not marked as temporary and have not been deleted at the time of access. It does not publish your chats. Account identifiers are not shared with partners.
- Purpose limits: evaluation, quality assurance, topic modeling, safety analysis, and related research tasks. No public redistribution of raw chat text.
- AI-use preferences: where applicable, we transmit machine-readable AI-use preferences associated with your account or contribution and instruct partners to honor them (e.g., no training when you select "DoNotTrain"). These are policy signals; they are not a technical enforcement mechanism.
- Recipients: limited to our research team, which may include researchers from partner institutions. Research team is bound by purpose limits and required safeguards.
- Retention: partners must store data only as long as necessary for the approved research purpose and delete upon completion or within 180 days, whichever is sooner, unless longer retention is required by law and disclosed to us.
- Withdrawal: turning off Researcher Access stops future sharing. We will instruct partners to delete unneeded copies and derived artifacts where feasible, but prior analyses and irreversibly aggregated statistics may persist.

This program is optional and independent from the public Data Flywheel. You may enable either, both, or neither.

10.2.5 2.3B Responsible Asset Access (Planned)

We are exploring an additional, consent-based program that would provide controlled access to certain underlying assets using privacy-preserving tooling (for example, OpenMined's DataSite). Participation would be strictly opt-in and gated by explicit, in-product consent with purpose limits, auditing, and required safeguards. Today, there are four tiers for handling

your data and contributions (private/temporary, aggregate-only, Researcher Access, and public contributions). If launched, this would become a fifth tier. We will update these Terms and present a dedicated consent flow before enabling any such program.

10.2.6 2.4 How We Use Your Data

We use information to: - Provide, secure, and maintain the Services - Prevent abuse and enforce our terms - Comply with legal obligations - With your consent, contribute to the Data Flywheel program

In order to operate the Services, we send your Prompts to subprocessors so that you can receive an AI Output. These subprocessors are listed in this document (under "Subprocessors") and include both model providers (the organization who built the AI models our Services use) and cloud service providers.

Your data protection: - We do not sell or rent your personal data. We share only what is necessary with subprocessors to operate, secure, and improve the Services as described in this Policy. - We may send your Prompts to outside providers as required to operate the Services. - Individual chat sharing to researchers only occurs if you opt in to the Researcher Access program; public release only occurs if you opt in to the Data Flywheel program. - Security reviews are conducted under legitimate interests with strict access controls.

We may share **aggregated, non-identifying statistics** with Partners, which may include: - Academic research institutions - Open source AI development communities
- Nonprofit organizations aligned with our mission

What aggregate data includes: - Total usage volume, growth metrics, user retention, engagement patterns - General topic distributions (e.g., "15% educational queries") - Model performance statistics - Safety and abuse prevention metrics

What it NEVER includes: - Individual usernames, emails, or identifiers - Specific chat content, prompts, or files - Any data from fewer than 100 users - Information that could reasonably identify you

This aggregate sharing helps improve AI systems for everyone while protecting your privacy.

10.2.7 2.5 Retention, Deletion, and Export

We temporarily retain technical logs (such as system activity and usage records) to help us ensure security, monitor performance, and improve our services. These logs are stored for no longer than 90 days. After that period, we permanently delete the raw logs and retain only aggregated, anonymized data that cannot reasonably be used to identify you.

Deleted chats are removed from active systems within 30 days, subject to legal holds (duration of legal proceedings plus 1 year). Safety/abuse data retained as necessary.

Depending on your region, you may have rights to access, rectification, erasure, portability, restriction/objection, and consent withdrawal. We make these available via two avenues: self-service within our Service for export and deletion, and by contacting us via email using [**hello@publicai.co**].

Currently, our Service supports exporting all chats as JSON directly from the Settings menu. Users can also delete all chats directly from the Settings menu.

For Data Subject Access Requests (DSAR), you can submit requests to **hello@publicai.co**. Response within 30 days (extendable by 60 for complexity). Identity verification required.

Our Services offer users the option to contribute to external projects, such as a "Public AI Data Flywheel". These offerings involve making public contributions which are handled under a separate retention and publication policy. Participation will be gated via an explicit opt-in and consent procedure.

10.2.8 2.6 Incident Response & Breach Notification

We investigate incidents, mitigate risk, and notify affected users and authorities without undue delay and, where GDPR applies, within 72 hours when required.

We assess legal requests for validity, seek to narrow scope, and provide notice unless prohibited. Data may be placed under legal hold until resolved.

10.2.9 2.7 Changes to this Policy

We may update this Policy. Material changes will be communicated directly on our website.

10.2.10 2.8 Cookie Policy

We use persistent first-party cookies and some third-party cookies to store certain preferences, make it easier for you to use our applications, as well as support some analytics.

A cookie is a piece of text stored by your browser. It may help remember login information and site preferences. It might also collect information such as your browser type, operating system, web pages visited, duration of visit, content viewed, and other click-stream data. You can adjust cookie retention settings and accept or block individual cookies in your browser settings, although our apps won't work and other aspects of our service may not function properly if you turn cookies off.

10.2.11 2.9 International Users, Transfers, and GDPR

We are a US-based nonprofit. We welcome users from around the world but want to be transparent about our limitations:

What we DO commit to: - Honor all data deletion requests within 30 days - Practice data minimization (we only collect what's needed) - Provide data export tools - Never sell or rent your personal data - Respond to privacy requests at hello@publicai.co

What we DON'T currently offer: - Formal Data Processing Agreements (DPAs) - EU Standard Contractual Clauses - EU/UK representatives - Localized data storage outside the US

If you're in the EU/UK and these formal mechanisms are required for your use case, this service may not be suitable for you. We're a small nonprofit focused on advancing open AI research with consideration for privacy rather than enterprise compliance frameworks. Organizations or individuals requiring formal GDPR compliance documentation should not use our Service.

For the purposes of the EU General Data Protection Regulation (GDPR), we act as a controller for personal data processed by the Service, including your chats (account data and chat data), participation in Researcher Access, and contributions to the Public AI Data Flywheel. This means we determine how such data is collected, processed, redacted, licensed (for public contributions), and released.

When you interact with external model providers through our Service, those providers act as independent controllers of your data. They determine their own purposes and means of processing, including whether to retain prompts, responses, and related metadata for research, safety, or service improvement. We do not direct or control how they handle your data, and no Data Processing Agreements are in place for those providers.

For Researcher Access, vetted research partners act as our processors: they process data under our instructions and for limited purposes (evaluation and related research tasks), and they are contractually obligated to implement appropriate safeguards, honor our deletion instructions, and not further disclose your data.

Because model providers are independent controllers, your data rights must be exercised separately with each provider. We encourage you to review their privacy notices and terms to understand how they handle your information.

10.2.12 2.10 Children's Privacy

The Services are not directed to children. We do not knowingly collect personal data from children under 18. If you believe a child provided personal data, contact us for deletion.

10.2.13 2.11 Security Measures

We implement reasonable security measures for a small nonprofit: - Access controls and least privilege principles - MFA on administrator accounts
- Regular security updates - Encrypted data transmission (HTTPS)

Current limitations: - No encryption at rest (planned when OpenWebUI supports it) - No formal security audits or certifications - Backups are retained for 90 days for disaster recovery

This is a research service, not enterprise infrastructure. Please don't use it for sensitive data.

10.2.14 2.12 Regional Rights (US State Supplement)

For residents of California, Colorado, Connecticut, Utah, and Virginia: We do not "sell" or "share" personal information as defined by these laws and do not use personal data for targeted advertising. You have rights to access, delete, correct, portability, and appeal adverse decisions (state-specific). We will not discriminate for exercising rights. Submit requests at hello@publicai.co.

10.2.15 2.13 Subprocessors

Note: Some entities listed below act as independent controllers rather than subprocessors; see Section 2.9 for roles and responsibilities.

Model providers (independent controllers; see 2.9) are: - Swiss AI Initiative (https://www.swiss-ai.org/) - AI Singapore

Cloud service providers are: - AWS - Exoscale - Swiss National Supercomputing Center - Australian National Computational Infrastructure - Cudo Compute - Vercel

Analytics: - Mixpanel

Public data hosting: - HuggingFace

11 Public AI Data Flywheel — User FAQ

This FAQ explains what the data flywheel is, why you might contribute, and how your choices affect privacy, attribution, and downstream use. It is written for people using the OpenWebUI chat at https://chat.publicai.co/.

11.1 TL;DR: Four Data Tiers (Today)

- Tier 1 — Private/Temporary: Use temporary chats or delete chats yourself. Nothing is shared outside the app; stored minimally; you control deletion.
- Tier 2 — Aggregate-Only (default): Chats stay private, but we compute de-identified aggregate stats (e.g., usage counts, broad trends). Nothing is shared outside the app.
- Tier 3 — Researcher Access (opt-in): Vetted partners can privately analyze your non-temporary, non-deleted chats for evaluation/R&D. No public release; account IDs not shared; partners must honor AI-use preferences.
- Tier 4 — Public Contributions (opt-in per chat): You select specific chats to publish to a public dataset (via PR), choose a license, and attach AI-use preferences; attribution can be username, pseudonym, or "anonymous."

You can mix and match: keep default behavior, enable Researcher Access, and still publish only some chats publicly. Note: we are exploring a future fifth tier for responsible, consented asset access using privacy-preserving tools (see note below).

11.2 What is the "data flywheel" and why does it exist?

The flywheel is a simple feedback loop: people can choose to share selected chats, those contributions are reviewed and organized, and the resulting dataset helps evaluate and improve public AI. Using a Git-backed repository makes the process transparent: every contribution has history, discussion, and a clear license. The rationale is to create high-quality, accountable data for public benefit without asking you to learn a new workflow.

11.3 I want the most privacy possible. How do I get that?

By default, we never share your chats, but we do perform aggregate analysis. For instance, we count the total volume of chat and how many users we have. We may also perform aggregate analyses of chat usage: how long are chats, what are common topics, etc.

If you want a high degree of privacy, you can use "temporary chats" or you can delete chats yourself at any time.

11.4 I'm OK with sharing a little if it helps out with public AI. What are my options?

There are three ways to share data, based on your interests and comfort level.

If you use the inference utility without temporary mode or deleting chats, you contribute aggregate data. For instance, the fact you even used the interface at all is useful so we can tell public bodies that there is real interest in using public AI models.

If you are OK with researchers using your data directly for research and development, you can enable Researcher Access. This means your data may be shared with vetted research partners for direct evaluation, topic modeling, and related R&D tasks. Your account info will NOT be shared and the chats will not be made public (unless you choose to, see next section).

Finally, you can choose to contribute specific chats to a public repository. Most of this document covers this option. These chats will be public, but you can select a license and also set AI use preferences (you can ask that only organizations who attribute training data use your data, for instance).

So to summarize, there are four options: - Make use of temporary chats and deletion to maximize privacy - Use the app in "default mode". Your data will never be shared outside the app or made public, but will contribute to aggregate analyses. - Turn on research sharing so public AI researchers can use data directly for R&D. - Pick notable chats to contribute to public repository, with restrictions that you select.

Note on future options: we are exploring a fifth option that would provide responsible, consented access to certain assets using privacy-preserving tools (for example, OpenMined's DataSite). If introduced, this would allow controlled access by vetted parties under explicit consent and safeguards. We will update this FAQ and in-product consent flows before enabling any such option.

11.5 What happens when I share a chat?

When you click the share action, the app shows a confirmation screen with exactly what will be shared and which settings apply. If you confirm, the system creates a single JSON record containing the conversation, light metadata about the model and usage, and your chosen license and AI preference signals. That record is submitted as a pull request to a HuggingFace dataset repository where it can be reviewed. If accepted, it becomes part of a public dataset with the same license and preference signals you selected.

11.6 What is the difference between publishing a chat publicly and opting in to Researcher Access?

Publishing a chat publicly means the content and its license are visible to everyone. Others can read it, discuss it, and use it under the terms you chose. Opting in to Researcher Access grants vetted research partners permission to analyze your chats for evaluation and model development without making them public. When Researcher Access is enabled, researchers may analyze chats in your account that are not marked as temporary and have not been deleted at the time of access. The difference is who can see the content and when. If you only want public publication and do not want non-public research access, keep Researcher Access disabled.

11.7 What privacy protections are in place?

We aim for a balance between usefulness and care. The action warns you not to share confidential or private information and provides a preview before you confirm. Automated checks scan for common identifiers such as emails and financial numbers and can flag items for quarantine. Reviewers can request changes or reject contributions that seem risky. You can choose a pseudonym or "anonymous" for public attribution, and public releases display the attribution string but not your account id. Today the pseudonym is deterministic from your OpenWebUI account id (unsalted), so contributions under a pseudonym can be linked to each other over time. If you want the least linkability, choose "anonymous."

No automated system can guarantee perfect redaction. If you are unsure, please edit the chat before sharing or decline to share. When in doubt, prefer synthetic or anonymized examples. If you want to retract data later on, we will help you do that, but we must warn you that because the data is public it may not be possible to delete any copies that exist.

11.8 What data about me is stored?

For public contributions, the dataset includes the chat content, timestamps, model information, your selected license and AI preference signals, and the attribution string you chose. Internally, we store a random contribution id and, if you choose "pseudonym," we derive a stable pseudonym from your OpenWebUI account id. We do not publish your account information. If you opt in to Researcher Access, vetted partners may analyze your chats (excluding temporary or deleted chats) for evaluation and QA without making them public.

11.9 Who can see my data and where is it stored?

Public contributions live as pull requests and merged records in a HuggingFace dataset repository. Anyone can view merged public content and its history. Items under review are visible to reviewers and, where practical, to the public in PR form. Content that triggers a privacy concern (or any other concerns) may be quarantined for private review. If you enable Researcher Access, vetted partners may analyze your chats privately for evaluation and model development.

11.10 Will my chat be used to train models?

That depends on the AI preference signals and license you choose. Preference signals such as `train-genai=n` express that you do not want a contribution used for training. Responsible users of the dataset should honor these signals and the license terms. Preference signals are not a technical barrier; they function as a clear, machine-readable policy that downstream tools and institutions can enforce. We use them to keep evaluation-only subsets separate from training sets. If you enable Researcher Access, partners must honor these preferences for any analyses they perform.

Preset options (Content-Usage expressions with CC signals as exceptions): - Training family: - `train-genai=n` (deny training) - `train-genai=n;exceptions=cc-cr` (deny training unless Credit) - `train-genai=n;exceptions=cc-cr-dc` (deny training unless Credit + Direct Contribution) - `train-genai=n;exceptions=cc-cr-ec` (deny training unless Credit + Ecosystem Contribution) - `train-genai=n;exceptions=cc-cr-op` (deny training unless Credit + Open) - General AI use family: - `ai-use=n` (deny AI use) - `ai-use=n;exceptions=cc-cr` (deny AI use unless Credit) - `ai-use=n;exceptions=cc-cr-dc` (… unless Credit + Direct Contribution) - `ai-use=n;exceptions=cc-cr-ec` (… unless Credit + Ecosystem Contribution) - `ai-use=n;exceptions=cc-cr-op` (… unless Credit + Open)

Default preset: `train-genai=n;exceptions=cc-cr` (deny training unless Credit).

11.11 How do licenses work here?

You choose a default license in the app (for example, CC0-1.0, CC-BY-4.0, or CC-BY-SA-4.0). That license is recorded with each contribution and governs how others can use it. If you want attribution when others use your content, choose a license that requires it. If you want the broadest possible reuse, choose a permissive option. You can change your default for future contributions at any time.

11.12 Do you perform aggregate analysis of chats?

We may compute aggregate, de-identified statistics across chats to understand system performance, reliability, and safety. Examples include counts, rates, and broad trends. Aggregate analysis is designed to avoid re-identifying individuals and to inform product quality and operations.

If you prefer not to leave any record tied to your account, use temporary chats and avoid saving conversations. You can also delete individual chats and, at any time, delete your account.

11.13 How can I fully opt out?

If you never want your content used in public datasets or internal research, do not enable sharing and do not enable Researcher Access. Use temporary chats so conversations are not saved to your account, or delete chats when you finish. You can delete your account at any time from account settings. Deleting your account removes account-linked data we control. It does not unpublish contributions that were previously made public under an open license.

11.14 Can I change my mind after sharing?

You can withdraw future consent by turning off sharing and disabling Researcher Access. For items already submitted, comment on the pull request or contact support to request changes. Public contributions that have been merged and redistributed under an open license may continue to circulate outside our control; we will reflect removals in our canonical dataset and communicate changes downstream where possible.

11.15 What are my choices for identity and credit?

You can publish under your username, a custom pseudonym, or "anonymous." Public pages display the attribution string you chose. If you want credit linked to a third-party account, you can also submit directly from your own HuggingFace account by opening pull requests yourself.

11.16 Why is the pseudonym deterministic?

A deterministic pseudonym creates a stable identity that lets you accrue credit over time without exposing your account id. It helps reviewers and researchers recognize consistent contributors, makes moderation and feedback easier, and enables simple deduplication and analytics. The tradeoff is linkability: contributions made under the same pseudonym can be connected to each other within this system. If you prefer the least linkability, choose "anonymous" instead of "pseudonym" for public attribution.

11.17 What should I avoid sharing?

Do not submit private, confidential, or regulated information about yourself or others. Avoid personal identifiers, financial information, and any content you do not have the right to share under your chosen license. If a conversation requires private details, do not share it. Edit the content to remove sensitive parts or submit a synthetic example instead.

11.18 How do I set or change my settings?

In OpenWebUI, open Controls (Functions), enable or disable sharing, choose your default license and AI preference signals, and set your attribution. You can also enable automatic feedback to streamline contributions. These settings apply to future shares and are restated on the confirmation screen for each contribution.

11.19 How do I report a problem or request removal?

Open an issue or comment on the relevant pull request with a clear description, or contact support through the chat app. For urgent privacy concerns, say that in the subject so we can prioritize review and quarantine where needed.

If you have questions not covered here, let us know in the chat or open a discussion on the dataset repository. Your feedback shapes how the flywheel evolves.

12 Code

For reference, this chapter contains key snippets of code from the Flywheel Action. (The full code is tracked in the `src` directory, but is duplicated here in case readers are curious.)

Part III

Updating this book

13 Contributing

Thanks for your interest in improving this book and its supporting materials. This chapter explains how to propose changes, what to edit, and how to preview your work locally.

13.1 How You Can Help

- Fix typos, grammar, and clarity in chapters.
- Suggest new sections, examples, figures, or diagrams.
- Improve citations and references.
- Review structure, flow, and consistency across chapters.

13.2 Project Structure

- `book/`: Quarto book source (`.qmd`), bibliography files, and build outputs.
  - `_quarto.yml`: Book configuration (title, chapters, output).
  - `*.qmd`: Chapter files.
  - `references.bib`, `references_blogs.bib`: Citation databases.
  - `docs/` and `_book/`: Built outputs (HTML site, PDFs). You do not need to update these for small edits.
- `readme.md`: Project overview.

13.3 Getting Set Up (Optional but Recommended)

You can submit small edits directly via the GitHub web UI.

At this stage, contributors are not required to build the book for each contribution (PR). If activity scales up, this may change.

For local previews: - Install Quarto CLI: https://quarto.org - For PDF builds: install a LaTeX distribution (e.g., TinyTeX) if you plan to render PDFs.

13.4 Editing Content

- Work in `book/` for the main text.
- Use `.qmd` (Quarto Markdown). Keep front matter minimal unless needed.
- Follow existing naming patterns for files (e.g., `01a_intro.qmd`, `2a_mvp.qmd`).
- If you add a new chapter, also add it to `book/_quarto.yml` under the appropriate `part` and `chapters` list.

13.4.1 Style Guidelines (First Draft)

- Audience: practitioners, policymakers, interested public, and academics; aim for clear, direct language.
- Voice: concise, neutral, plain language; prefer active voice.
- Consistency: American English; sentence case for headings unless a proper noun.
- Structure: short paragraphs, descriptive headings, and lists when helpful.
- Figures: place images in `book/` (you may create `book/images/`) and reference with relative paths. Provide alt text.
- Tables/Callouts (sparingly): use standard Markdown tables and Quarto callouts when helpful.
- Acronyms: spell out on first use; keep consistent across chapters.

13.4.2 Citations & References

- Use Quarto/Pandoc citation syntax in text, e.g., `[@smith2020]` or `[@smith2020, p. 10]`.
- Add scholarly works to `book/references.bib` and blog/web articles to `book/references_blogs.bib`.
  - (We may break these up a bit more as references accumulate as well!)
- Prefer stable identifiers (DOIs, permalinks). Include `url` and `urldate` for web sources when possible.
- Keep references tidy: one entry per work; avoid duplicates.

13.4.3 Cross-References

- Use Quarto cross-refs for figures, tables, and sections when needed, e.g., `(@fig:example)`.
- Keep labels descriptive and unique within the book.

13.5 Previewing Changes Locally

From the repository root:

1) Preview while you edit (auto-reloads):

```
cd book
quarto preview
```

2) Render the full book:

```
cd book
quarto render
```

Outputs go to `book/docs/` (HTML site) and may include a PDF if configured.

Notes on build outputs: - You do not need to commit `book/docs/` or `book/_book/` for small text changes. Maintainers will rebuild as needed. - If you do render locally, avoid large diffs to generated files in your PR unless explicitly requested.

13.6 Making a Pull Request

- Create a feature branch from `main`. Suggested naming: `topic/short-summary` (e.g., `typo/fix-intro-typos`).
- Keep PRs small and focused; open an issue first for large structural changes.
- Include a brief summary of what changed and why.
- If adding a new chapter/section, mention where you placed it and why.
- If you touched citations, mention any new entries added to the `.bib` files.

PR checklist (recommended): - Content: clear, consistent, and scoped to one concern. - If possible, try to Build: book renders locally (optional for small fixes). - Links: relative links resolve; images render; alt text provided. - References: citations compile; bib entries added/updated as needed.

13.7 Reviews & Merging

- Maintainers aim to review within 1–2 weeks.
- We may suggest edits for tone, structure, or scope.
- For substantial changes, we may ask you to split PRs or iterate.

13.8 Code of Conduct

Please be respectful and constructive. Disagreement is fine; personal attacks are not. Maintainers may moderate discussions and contributions to keep the project welcoming and productive.

13.9 License & Attribution

- Unless otherwise noted, contributions will be incorporated under the repository's license (if updated/added). If you require a different licensing arrangement, open an issue before submitting a PR.
- If you adapt text from openly licensed sources, include proper attribution in comments or commit messages and ensure license compatibility.

13.10 Questions

Open an issue with your question or proposal. If you're unsure where to start, consider opening a small PR with suggested changes.

14 Guidelines for Using AI

This project is about AI, so using AI tools is welcome throughout the workflow. However, humans are ultimately accountable for what ships. Use AI to accelerate, but keep humans responsible for correctness, coherence, and integrity.

14.1 Writing with AI

- Human sign-off: read every line before it becomes part of the book. Treat AI output as a draft (sometimes, a VERY EARLY draft) expect to edit and fact-check.
- Iteration over one-shots: current models rarely produce publication-ready text in one pass, so plan to iterate and refine.
- Maintain voice: ensure tone, terminology, and structure match the book's style and audience.
  - Note that the "voice" here is still up for debate :)
- Track changes (recommended): summarize AI-assisted edits in PR descriptions; include key prompts when they inform substantive choices.

14.2 Citations with AI

- Verify metadata: for each citation suggested or formatted by AI, confirm author(s), title, venue, year, DOI/URL, and key fields in `references.bib` or `references_blogs.bib`.
- Check links: ensure all links resolve and are appropriate for the cited claim; add `urldate` for web sources.
- Prefer stable identifiers: use DOIs and permalinks when available; avoid ephemeral URLs.
- Be wary of hallucinated citations: do not add references unless you can independently verify they exist and match the claim.

14.3 Code with AI

- Write tests: include unit tests or executable examples that validate behavior for any non-trivial code contributed to this repo.
- Run when possible: execute tests locally when feasible; include brief notes or logs of results in the PR description.
- Keep records: note model name/version, prompts, and any constraints if they materially affect the code outcome.
- Security and privacy: do not paste secrets or sensitive data into AI tools; scrub logs and examples of credentials.

14.4 General Practices

- Attribution: when AI assistance is significant, disclose it briefly (e.g., "sections drafted with AI, edited by ").
- Accessibility: ensure AI-generated figures, tables, and examples include alt text and are accessible.
- Consistency: align terminology and formatting with existing chapters; prefer incremental, focused edits.

15 Future Directions

This section sketches near-term areas to extend the book and its accompanying materials. Please open issues or PRs to propose additions.

Here are just a few ideas:

15.1 Content Expansions

- Additional case studies: public-sector deployments, civic tech pilots, domain-specific flywheels (health, education, climate).
- Deeper methods chapters: data governance patterns, preference signals, evaluation protocols, and reproducibility practices.
- Community processes: contributor pathways, governance models, and decision-making templates.

15.2 Tooling & Infrastructure

- Consent and data rights flows: prototypes and reusable modules for informed consent, redaction, and data lifecycle controls.
- Evaluation harness: standardized benchmarks, metrics, and reporting for public AI systems; guidance on dataset release strategies.
- Legal/policy pack: practical checklists, model terms, and jurisdictional notes for public AI programs.

15.3 Roadmap & Calls for Contribution

- Identify gaps: add issues tagging "help wanted" for chapters, examples, and figures.
- Build examples: minimal, runnable demos tied to book sections.
- Iterate: refine structure in response to reader feedback and real deployments.

References

Acemoglu, Daron, and Simon Johnson. 2025. "Power and Progress: Our Thousand-Year Struggle over Technology and Prosperity." *Perspectives on Science and Christian Faith*. https://api.semanticscholar.org/CorpusID:265119352.

Anthropic. n.d. "HH-RLHF Dataset." https://github.com/anthropics/hh-rlhf.

Apache Software Foundation. n.d. "Apache Parquet Project." https://parquet.apache.org/.

Ardila, Rosana, Megan Branson, Kelly Davis, Michael Henretty, Michael Kohler, Josh Meyer, Reuben Morais, Lindsay Saunders, Francis M Tyers, and Gregor Weber. 2019. "Common Voice: A Massively-Multilingual Speech Corpus." *arXiv Preprint arXiv:1912.06670*.

Arnold, Eckhart. 2014. "What's Wrong with Social Simulations?" *The Monist* 97: 359–77. https://api.semanticscholar.org/CorpusID:67844223.

arXiv.org. n.d.a. "arXiv API User's Manual." https://info.arxiv.org/help/api/user-manual.html.

———. n.d.b. "arXiv Bulk Data Access." https://info.arxiv.org/help/bulk_data.html.

———. n.d.c. "arXiv OAI-PMH Interface." https://info.arxiv.org/help/oa/index.html.

Aryabumi, Viraat, Yixuan Su, Raymond Ma, Adrien Morisot, Ivan Zhang, Acyr Locatelli, Marzieh Fadaee, Ahmet Üstün, and Sara Hooker. 2024. "To Code, or Not to Code? Exploring Impact of Code in Pre-Training." *arXiv Preprint arXiv:2408.10914*.

Barocas, Solon, and Andrew D. Selbst. 2016. "Big Data's Disparate Impact." *California Law Review* 104 (3): 671–732.

Batty, Michael, and Paul M. Torrens. 2001. "Modeling Complexity : The Limits to Prediction." *Cybergeo: European Journal of Geography*. https://api.semanticscholar.org/CorpusID:102344300.

Baumgartner, Jason, Savvas Zannettou, Brian Keegan, Megan Squire, and Jeremy Blackburn. 2020. "The Pushshift Reddit Dataset" 14: 830–39.

Belkin, Mikhail, Daniel Hsu, Siyuan Ma, and Soumik Mandal. 2019. "Reconciling Modern Machine-Learning Practice and the Classical Bias–Variance Trade-Off." *Proceedings of the National Academy of Sciences* 116 (32): 15849–54. https://doi.org/10.1073/pnas.1903070116.

Bender, Emily M., Timnit Gebru, Angelina McMillan-Major, and Shmargaret Shmitchell. 2021. "On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?" In *Proceedings of the ACM Conference on Fairness, Accountability, and Transparency (FAccT)*, 610–23.

BigCode Project. n.d.a. "BigCode Project Documentation." https://www.bigcode-project.org/docs/about/the-stack/.

———. n.d.b. "The Stack Dataset on Hugging Face." https://huggingface.co/datasets/bigcode/the-stack/tree/main.

———. n.d.c. "The Stack V2 Dataset on Hugging Face." https://huggingface.co/datasets/bigcode/the-stack-v2.

Blodgett, Su Lin, Solon Barocas, Hal Daumé III, and Hanna Wallach. 2020. "Language (Technology) Is Power: A Critical Survey of 'Bias' in NLP." In *Proceedings of ACL*, 5454–76.

Buolamwini, Joy, and Timnit Gebru. 2018. "Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification." In *Proceedings of the Conference on Fairness, Accountability and Transparency (FAT*)*, 77–91.

Carlini, Nicholas, Matthew Jagielski, Christopher A. Choquette-Choo, Daniel Paleka, Will Pearce, Hyrum Anderson, Andreas Terzis, Kurt Thomas, and Florian Tramèr. 2024. "Poisoning Web-Scale Training Datasets Is Practical." https://arxiv.org/abs/2302.10149.

Carlini, Nicholas, Chang Liu, Úlfar Erlingsson, Jernej Kos, and Dawn Song. 2019. "The Secret Sharer: Measuring Unintended Memorization in Neural Networks." In *Proceedings of USENIX Security Symposium*.

Carlini, Nicholas, Florian Tramer, Eric Wallace, Matthew Jagielski, Ariel Herbert-Voss, Katherine Lee, Adam Roberts, et al. 2021. "Extracting Training Data from Large Language Models." In *Proceedings of USENIX Security Symposium*.

Common Crawl. n.d.a. "Common Crawl – Get Started." https://commoncrawl.org/get-started.

———. n.d.b. "Web Archiving File Formats Explained." https://commoncrawl.org/blog/web-archiving-file-formats-explained.

Crawford, Kate, and Trevor Paglen. 2019. "Excavating AI: The Politics of Images in Machine Learning Training Sets." https://www.excavating.ai/.

Creative Commons. 2023. "Understanding CC Licenses and Generative AI." https://creativecommons.org/2023/08/18/understanding-cc-licenses-and-generative-ai/.

"Cybernetics." 2025. *Wikipedia*. https://en.wikipedia.org/w/index.php?title=Cybernetics&oldid=1300921342.

Databricks. n.d. "Databricks Dolly Repository." https://github.com/databrickslabs/dolly.

Deckelmann, Selena. 2023. "Wikipedia's Value in the Age of Generative AI." *Wikimedia Foundation*. https://wikimediafoundation.org/news/2023/07/12/wikipedias-value-in-the-age-of-generative-ai/.

École Normale Supérieure. n.d. "HowTo100M Project." https://www.di.ens.fr/willow/research/howto100m/.

European Union. 2016. "General Data Protection Regulation (EU) 2016/679." https://eur-lex.europa.eu/eli/reg/2016/679/oj.

———. 2024. "Artificial Intelligence Act." https://eur-lex.europa.eu/.

Federal Trade Commission. 2013. "Children's Online Privacy Protection Rule (COPPA) — 16 CFR Part 312." https://www.ftc.gov/legal-library/browse/rules/childrens-online-privacy-protection-rule-coppa.

Fernandez, Raul Castro. 2023. "Data-Sharing Markets: Model, Protocol, and Algorithms to Incentivize the Formation of Data-Sharing Consortia." *Proceedings of the ACM on*

Management of Data 1: 1–25. https://api.semanticscholar.org/CorpusID:259213174.

Gao, Leo, Stella Biderman, Sid Black, Laurence Golding, Travis Hoppe, Charles Foster, Jason Phang, et al. 2021. "The Pile: An 800GB Dataset of Diverse Text for Language Modeling." *CoRR* abs/2101.00027. https://arxiv.org/abs/2101.00027.

Gebru, Timnit, Jamie Morgenstern, Briana Vecchione, Jennifer Wortman Vaughan, Hanna Wallach, Hal Daumé III, and Kate Crawford. 2018. "Datasheets for Datasets." In *arXiv:1803.09010*.

Grother, Patrick, Mei Ngan, and Kayee Hanaoka. 2019. "Face Recognition Vendor Test (FRVT) Part 3: Demographic Effects." NISTIR 8280. NIST. https://doi.org/10.6028/NIST.IR.8280.

gururise. n.d. "Alpaca Data Cleaned Repository." https://github.com/gururise/AlpacaDataCleaned.

Hendrycks, Dan. n.d. "Competition Math Dataset on Hugging Face." https://huggingface.co/datasets/hendrycks/competition_math.

Hendrycks, Dan, Collin Burns, Saurav Kadavath, Akul Arora, Steven Basart, Eric Tang, Dawn Song, and Jacob Steinhardt. 2021. "Measuring Mathematical Problem Solving with the MATH Dataset." https://arxiv.org/abs/2103.03874.

Hestness, Joel, Sharan Narang, Newsha Ardalani, Gregory Diamos, Heewoo Jun, Hassan Kianinejad, Md Patwary, Mostofa Ali, Yang Yang, and Yanqi Zhou. 2017. "Deep Learning Scaling Is Predictable, Empirically." *arXiv Preprint arXiv:1712.00409*.

Holland, Sarah, Ahmed Hosny, Sarah Newman, Joshua Joseph, and Kasia Chmielinski. 2018. "The Dataset Nutrition Label: A Framework to Drive Higher Data Quality Standards." https://arxiv.org/abs/1805.03677.

Hubinger, Evan, Carson Denison, Jesse Mu, Mike Lambert, Meg Tong, Monte MacDiarmid, Tamera Lanham, et al. 2024. "Sleeper Agents: Training Deceptive LLMs That Persist Through Safety Training." https://arxiv.org/abs/2401.05566.

Hwang, Sohyeon, Priyanka Nanayakkara, and Yan Shvartzshnaider. 2025. "Trust and Friction: Negotiating How Information Flows Through Decentralized Social Media." *arXiv Preprint arXiv:2503.02150*.

Illinois General Assembly. 2008. "Biometric Information Privacy Act (BIPA), 740 ILCS 14." https://www.ilga.gov/legislation/ilcs/ilcs3.asp?ActID=3004.

International Internet Preservation Consortium. 2017. "The WARC Format 1.1." https://iipc.github.io/warc-specifications/specifications/warc-format/warc-1.1/.

ISO/IEC 23894:2023 Information Technology—Artificial Intelligence—Risk Management. 2023. ISO/IEC.

Jackson, Brandon, B Cavello, Flynn Devine, Nick Garcia, Samuel J. Klein, Alex Krasodomski, Joshua Tan, and Eleanor Tursman. 2024. "Public AI: Infrastructure for the Common Good." Public AI Network. https://doi.org/10.5281/zenodo.13914560.

Jo, Emily, and Timnit Gebru. 2020. "Lessons from Archives: Strategies for Collecting Sociocultural Data in Machine Learning." In *Proceedings of FAccT*, 306–16.

Johnson, Isaac, Lucie-Aimée Kaffee, and Miriam Redi. 2024. "Wikimedia Data for AI: A Review of Wikimedia Datasets for NLP Tasks and AI-Assisted Editing." *arXiv Preprint arXiv:2410.08918*.

jsonlines.org. n.d. "JSON Lines Specification." https://jsonlines.org/.
Kollock, Peter. 1998. "Social Dilemmas: The Anatomy of Cooperation." *Annual Review of Sociology* 24 (1): 183–214. https://doi.org/10.1146/annurev.soc.24.1.183.
LAION. 2022a. "LAION-5B: A New Era of Open Large-Scale Multi-Modal Datasets." https://laion.ai/laion-5b-a-new-era-of-open-large-scale-multi-modal-datasets/.
———. 2022b. "Releasing Re-LAION-5B." https://laion.ai/blog/relaion-5b/.
Library of Congress. n.d. "WARC, Web ARChive File Format." https://www.loc.gov/preservation/digital/formats/fdd/fdd000236.shtml.
Liu, Jason. 2024. "Data Flywheel Go Brrr: Using Your Users to Build Better Products - Jason Liu." https://jxnl.co/writing/2024/03/28/data-flywheel/.
Liu, Jiacheng, Taylor Blanton, Yanai Elazar, Sewon Min, YenSung Chen, Arnavi Chheda-Kothary, Huy Tran, et al. 2025. "OLMoTrace: Tracing Language Model Outputs Back to Trillions of Training Tokens." *arXiv Preprint arXiv:2504.07096*.
Marda, Nik, Jasmine Sun, and Mark Surman. 2024. "Public AI: Making AI Work for Everyone, by Everyone." Mozilla. https://assets. mofoprod. net/network/documents/Public_AI_Mozilla. pdf.
Marwell, Gerald, and Pamela Oliver. 1993. *The Critical Mass in Collective Action*. Cambridge University Press.
McCallister, Erika, Tim Grance, and Karen Scarfone. 2010. "Guide to Protecting the Confidentiality of Personally Identifiable Information (PII)." SP 800-122. NIST.
McDonald, Nora, Benjamin Mako Hill, Rachel Greenstadt, and Andrea Forte. 2019. "Privacy, Anonymity, and Perceived Risk in Open Collaboration: A Study of Service Providers." In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, 1–12.
Meta Stack Exchange. n.d. "Why Is the Stack Exchange Data Dump Only Available in XML?" https://meta.stackexchange.com/questions/267329/why-is-the-stack-exchange-data-dump-only-available-in-xml-file-format.
Miech, Antoine, Dimitri Zhukov, Jean-Baptiste Alayrac, Makarand Tapaswi, Ivan Laptev, and Josef Sivic. 2019. "HowTo100M: Learning a Text-Video Embedding by Watching Hundred Million Narrated Video Clips." *CoRR* abs/1906.03327. http://arxiv.org/abs/1906.03327.
Mitchell, Margaret, Simone Wu, Andrew Zaldivar, Parker Barnes, Lucy Vasserman, Ben Hutchinson, Elena Spitzer, Inioluwa Deborah Raji, and Timnit Gebru. 2019. "Model Cards for Model Reporting." In *Proceedings of the ACM Conference on Fairness, Accountability, and Transparency (FAccT)*, 220–29.
Murphy, Kevin P. 2022. *Probabilistic Machine Learning: An Introduction*. MIT Press. http://probml.github.io/book1.
Narayanan, Arvind, and Vitaly Shmatikov. 2008. "Robust de-Anonymization of Large Sparse Datasets." In *Proceedings of the IEEE Symposium on Security and Privacy*, 111–25.
ndjson. n.d. "NDJSON Specification." https://github.com/ndjson/ndjson-spec.
NISO. 2024. "ANSI/NISO Z39.96-2024, JATS: Journal Article Tag Suite." https://www.niso.org/publications/z3996-2024-jats.
Nissenbaum, Helen. 2004. "Privacy as Contextual Integrity." *Washington Law Review* 79 (1): 119–57.
NIST. 2023. "Artificial Intelligence Risk Management Framework (AI RMF 1.0)." NIST AI

100-1. National Institute of Standards; Technology; https://www.nist.gov/ai.
NLM. n.d. "Journal Article Tag Suite." https://jats.nlm.nih.gov/.
Obermeyer, Ziad, Brian Powers, Christine Vogeli, and Sendhil Mullainathan. 2019. "Dissecting Racial Bias in an Algorithm Used to Manage the Health of Populations." *Science* 366 (6464): 447–53.
OpenAI. 2022. "Introducing Whisper." https://openai.com/index/whisper/.
———. n.d.a. "Grade-School Math (GSM8K) Repository." https://github.com/openai/grade-school-math.
———. n.d.b. "GSM8K Hugging Face Dataset Card." https://huggingface.co/datasets/openai/gsm8k.
———. n.d.c. "OpenAI API Reference – Chat." https://platform.openai.com/docs/api-reference/chat.
OpenAssistant. n.d. "OpenAssistant OASST1 Dataset Card." https://huggingface.co/datasets/OpenAssistant/oasst1.
OWASP. 2023. "OWASP Top 10 for Large Language Model Applications." https://owasp.org/www-project-top-10-for-large-language-model-applications/.
Project Gutenberg. n.d.a. "Project Gutenberg File Formats." https://www.gutenberg.org/help/file_formats.html.
———. n.d.b. "Project Gutenberg Offline Catalogs and Feeds." https://www.gutenberg.org/ebooks/offline_catalogs.html.
Pushshift. n.d. "Pushshift.io." https://pushshift.io/.
Radford, Alec, Jong Wook Kim, Tao Xu, Greg Brockman, Christine McLeavey, and Ilya Sutskever. 2022. "Robust Speech Recognition via Large-Scale Weak Supervision." https://arxiv.org/abs/2212.04356.
Rafailov, Rafael, Archit Sharma, Eric Mitchell, Stefano Ermon, Christopher D. Manning, and Chelsea Finn. 2024. "Direct Preference Optimization: Your Language Model Is Secretly a Reward Model." https://arxiv.org/abs/2305.18290.
Raji, Inioluwa Deborah, Indra Elizabeth Kumar, Aaron Horowitz, and Andrew D. Selbst. 2022. "The Fallacy of AI Functionality." *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency*. https://api.semanticscholar.org/CorpusID:249872658.
Rakova, Bogdana, Renee Shelby, and Megan Ma. 2023. "Terms-We-Serve-with: Five Dimensions for Anticipating and Repairing Algorithmic Harm." *Big Data & Society* 10 (2): 20539517231211553.
Reddit. n.d. "Reddit API Documentation." https://www.reddit.com/dev/api/.
Reddit Help. n.d. "Reddit Data API Wiki." https://support.reddithelp.com/hc/en-us/articles/16160319875092-Reddit-Data-API-Wiki.
Roche, Adam, and Yali Sassoon. 2024. "What Is a Data Flywheel? A Guide to Sustainable Business Growth." *Snowplow Blog*. https://snowplow.io/blog/what-is-a-data-flywheel.
"Rosenbach v. Six Flags Entertainment Corp." 2019. 2019 IL 123186, Supreme Court of Illinois.
Selbst, Andrew D., Danah Boyd, Suresh Venkatasubramanian Friedler, Suresh Venkatasubramanian, and Janet Vertesi. 2019. "Fairness and Abstraction in Sociotechnical Systems." In *Proceedings of the ACM Conference on Fairness, Accountability, and Transparency*

(FAccT), 59–68.

Shankar, Shreya. 2024. "Data Flywheels for LLM Applications." *Shreya Shankar's Blog*. https://www.sh-reya.com/blog/ai-engineering-flywheel/.

Shelby, Renee, Shalaleh Rismani, Kathryn Henne, AJung Moon, Negar Rostamzadeh, Paul Nicholas, N'Mah Yilla-Akbari, et al. 2023. "Sociotechnical Harms of Algorithmic Systems: Scoping a Taxonomy for Harm Reduction." In *Proceedings of the 2023 AAAI/ACM Conference on AI, Ethics, and Society*, 723–41. AIES '23. New York, NY, USA: Association for Computing Machinery. https://doi.org/10.1145/3600211.3604673.

Shen, Judy Hanwen, Inioluwa Deborah Raji, and Irene Y Chen. 2024. "The Data Addition Dilemma." *arXiv Preprint arXiv:2408.04154*. https://arxiv.org/abs/2408.04154.

Sorscher, Ben, Robert Geirhos, Shashank Shekhar, Surya Ganguli, and Ari Morcos. 2022. "Beyond Neural Scaling Laws: Beating Power Law Scaling via Data Pruning." *Advances in Neural Information Processing Systems* 35: 19523–36.

Stack Exchange. n.d. "Stack Exchange Data Explorer Help." https://data.stackexchange.com/help.

Stanford CRFM. 2023. "Alpaca: A Strong, Replicable Instruction-Following Model." https://crfm.stanford.edu/2023/03/13/alpaca.html.

Sweeney, Latanya. 2000. "Simple Demographics Often Identify People Uniquely." *Carnegie Mellon University, Data Privacy Working Paper*.

Tan, Joshua, Nicholas Vincent, Katherine Elkins, and Magnus Sahlgren. 2025. "If Open Source Is to Win, It Must Go Public." *arXiv Preprint arXiv:2507.09296*.

Tatsu Lab. n.d. "Stanford Alpaca GitHub Repository." https://github.com/tatsu-lab/stanford_alpaca.

TensorFlow. n.d. "TFRecord and Tf.train.example Tutorial." https://www.tensorflow.org/tutorials/load_data/tfrecord.

TensorFlow Datasets. n.d.a. "C4 Dataset in TensorFlow Datasets." https://www.tensorflow.org/datasets/catalog/c4.

———. n.d.b. "C4 Generator Code." https://github.com/tensorflow/datasets/blob/master/tensorflow_datasets/text/c4.py.

Tran, Chau, Kaylea Champion, Andrea Forte, Benjamin Mako Hill, and Rachel Greenstadt. 2020. "Are Anonymity-Seekers Just Like Everybody Else? An Analysis of Contributions to Wikipedia from Tor." In *2020 IEEE Symposium on Security and Privacy (SP)*, 186–202. IEEE.

U.S. Copyright Office. 2024. "Copyright and Artificial Intelligence: Policy Studies and Guidance." https://copyright.gov/ai/.

U.S. Department of Education. 1974. "Family Educational Rights and Privacy Act (FERPA)." https://www2.ed.gov/policy/gen/guid/fpco/ferpa/index.html.

U.S. Department of Health and Human Services. 2000. "HIPAA Privacy Rule — 45 CFR Parts 160 and 164." https://www.hhs.gov/hipaa/for-professionals/privacy/index.html.

Vincent, Nicholas, David Bau, Sarah Schwettmann, and Joshua Tan. 2023. "An Alternative to Regulation: The Case for Public AI." *arXiv Preprint arXiv:2311.11350*.

Vincent, Nicholas, Mark Surman, and Jake Hirsch-Allen. 2025. "Canada as a Champion for Public AI: Data, Compute and Open Source Infrastructure for Economic Growth and

Inclusive Innovation."
Weidinger, Laura, John Mellor, et al. 2021. "Ethical and Social Risks of Harm from Language Models." *arXiv Preprint arXiv:2112.04359*.
Wikimedia Meta-Wiki. n.d. "Wikipedia Data Dumps – Dump Format." https://meta.wikimedia.org/wiki/Data_dumps/Dump_format.
Wikipedia. n.d. "Wikipedia Database Download." https://en.wikipedia.org/wiki/Wikipedia:Database_download.
Wolpert, David H, and William G Macready. 2002. "No Free Lunch Theorems for Optimization." *IEEE Transactions on Evolutionary Computation* 1 (1): 67–82.

Part IV

Appendices

16 Appendix 1: LLM Data Schemas

Here, we describe many variants of LLM data. This will be relevant for when we extend the flywheel to include more types of data, and especially shift towards promoting the sharing (via opt-in flywheels, but also via new market mechanisms) of richer "content data".

16.1 Open Web / Crawls

- **WARC/WAT/WET**
  - *WARC* (container for HTTP request/response records) — spec & overview: (International Internet Preservation Consortium 2017; Library of Congress, n.d.)
  - *WAT* (JSON metadata extracted from WARC) and *WET* (plain text extracted from HTML) — Common Crawl guides (Common Crawl, n.d.a, n.d.b)
- **C4 (Colossal Clean Crawled Corpus)** — TFDS catalog & generator code (TensorFlow Datasets, n.d.a, n.d.b)
- **The Pile** (22-source, mixed corpus) — paper (Gao et al. 2021)

16.2 Encyclopedic / Books

- **Wikipedia XML dumps** (page/revision XML; SQL tables for links) — (Wikimedia Meta-Wiki, n.d.; Wikipedia, n.d.)
- **Project Gutenberg**
  - *Books*: plain text/HTML master formats; ePub/MOBI derived (Project Gutenberg, n.d.a)
  - *Catalog schema*: daily RDF/XML (also CSV) for metadata; offline catalogs (Project Gutenberg, n.d.b)

16.3 Scientific / Legal

- **arXiv** (Atom/OAI-PMH metadata; bulk & API) — (arXiv.org, n.d.c, n.d.a, n.d.b)
- **JATS XML** (journal article tag suite) — (NISO 2024; NLM, n.d.)

16.4 Code

- **BigCode — The Stack / The Stack v2** (source files + license/provenance metadata; dedup variants) — (BigCode Project, n.d.b, n.d.c, n.d.a; **stack_paper?**)

16.5 Forums / Q&A / Social

- **Stack Exchange dumps** (XML: Posts, Users, Comments, Votes, etc.) — (Meta Stack Exchange, n.d.; Stack Exchange, n.d.)
- **Reddit**
  - *API JSON* schema — official docs (Reddit, n.d.; Reddit Help, n.d.)
  - *Pushshift* (historical dumps; research dataset) — (Pushshift, n.d.; Baumgartner et al. 2020)

16.6 Instruction / Conversations (Post-training SFT)

- **OpenAI-style chat schema** (role-tagged: `system|user|assistant`, plus tool calls) — (OpenAI, n.d.c)
- **Alpaca** (JSON prompts/instructions/outputs) — (Stanford CRFM 2023; Tatsu Lab, n.d.; gururise, n.d.)
- **Databricks Dolly-15k** (human-written instruction/response pairs) — (Databricks, n.d.)

- **OpenAssistant OASST1** (message-tree conversations with roles) — (OpenAssistant, n.d.)

16.7 Preference / Feedback (RLHF & DPO)

- **HH-RLHF** (Anthropic helpful/harmless, JSONL pairs: `chosen` vs `rejected`) — (Anthropic, n.d.)
- **DPO format** (prompt + preferred vs dispreferred response) — (Rafailov et al. 2024)

16.8 Multimodal (for VLMs/ASR)

- **LAION-5B / Re-LAION-5B** (image–text pairs with CLIP scores; links) — (LAION 2022a, 2022b)
- **Whisper** (weakly-supervised ASR; audio → text pairs) — (Radford et al. 2022; OpenAI 2022)
- **HowTo100M** (YouTube instructional video clips + narrations) — (École Normale Supérieure, n.d.; Miech et al. 2019)

16.9 Math Reasoning (often for post-training/eval)

- **GSM8K** (grade-school word problems; JSON) — (OpenAI, n.d.a, n.d.b)
- **MATH** (competition problems with step-by-step solutions) — (Hendrycks et al. 2021; Hendrycks, n.d.)

16.10 Common Storage Containers

- **JSON Lines / NDJSON** — (jsonlines.org, n.d.; ndjson, n.d.)
- **TFRecord** — (TensorFlow, n.d.)
- **Apache Parquet** — (Apache Software Foundation, n.d.)

17 Appendix 2 — Preference Signals for AI Data Use (CC signals + IETF AI Preferences)

This appendix provides a brief description of, a links to, information on emerging "AI Preference Signaling" from Creative Commons and the IETF (other initatives and orgs may be added as well).

Key links:

- "CC Signals: A New Social Contract for the Age of AI"
- "CC Signals Implementation"
- "creativecommons/cc-signals"
- https://www.ietf.org/archive/id/draft-ietf-aipref-vocab-02.html

What CC signals are: A Creative Commons framework for *reciprocal* AI reuse: content stewards can allow specific machine uses if certain conditions are met (e.g., credit, contributions, openness). Overview & implementation notes.

- **Four proposed CC signals (v0.1)**
  - **Credit (cc-cr)** — cite the dataset/collection; RAG-style outputs should link back when feasible.
  - **Credit + Direct Contribution (cc-cr-dc)** — proportional financial/in-kind support.
  - **Credit + Ecosystem Contribution (cc-cr-ec)** — contribute to broader commons.
  - **Credit + Open (cc-cr-op)** — release model/code/data to keep the chain open. Source (draft repo & posts).
- **IETF AI Preferences (aipref) — the transport & vocabulary**
  - **Vocabulary**: a machine-readable set of *categories* (e.g., `ai-use`, `train-genai`) and *preferences* (`y` = grant, `n` = deny) with **exceptions**. Drafts.
  - **Attachment**: how to convey these preferences via **HTTP `Content-Usage` header** and **robots.txt** extensions. Drafts.
  - **Structured Fields**: uses RFC-standardized HTTP structured field values.

  - **Robots Exclusion Protocol** baseline.

- **Putting them together (content-usage expression)**
  - Shape:

    `<category>=<y|n>;exceptions=<cc-signal>`

    Example in **robots.txt** (allow everything, but *AI use denied unless Credit*):

    ```
    User-Agent: *
    Content-Usage: ai-use=n;exceptions=cc-cr
    Allow: /
    ```

    Example **HTTP header** (deny *gen-AI training* unless *Credit + Ecosystem*):

    ```
    Content-Usage: train-genai=n;exceptions=cc-cr-ec
    ```

    (Syntax and examples from CC & IETF drafts.)

- **Operational notes (for this repo's flywheel)**
  - **Per-record fields** to store: `license` (CC0/CC-BY/CC-BY-SA) and `ai_pref` (IETF aipref value + optional CC signal), plus optional `attribution` handle. (Aligns with CC write-ups & IETF drafts.)
  - **Placement**:
    * *Location-based* signals via **robots.txt** for site/paths.
    * *Unit-based* signals via **HTTP** `Content-Usage` on dataset files and API responses.
  - **Interoperability expectations**: signals are normative *preferences*; adherence relies on ecosystem norms (similar to robots.txt & CC license culture).

18 Appendix 3: LLM Policy Docs

This Appendix contains a list of links to various live Terms of Service, Privacy Policy, and related docs for major LLM providers, including both private players like frontier labs and public AI-adjacent actors like AI2, Mozilla.

- AI2: ToS | Privacy
- Mozilla Common Voice: ToS | Privacy
- OpenAI: Terms of Use
- Anthropic
- QwenChat

19 Diffable Terms

Can we use diffs to show people the exact difference between two contribution-focused projects?

Here we provide an example of a Flywheel ToS that is meant to use terms that are very close to Mozilla Common Voice, such that someone who previosuly contributed to CV (or is familiar with Mozilla) might actually read the *diff* to help them understand the implications of data sharing.

Here is the ToS for Mozilla Common Voice

Common Voice Legal Terms

Effective November 4, 2024

Through Common Voice, you can donate your voice, written sentences, and the other resources we need to build an open-source voice database that anyone can use to make innovative voice recognition apps for devices and the web.

You may only participate in Common Voice if you agree to these Common Voice Legal Terms (the "Terms"). 1. Eligibility

Common Voice is open to anyone over the age of 19. If you are 19 or under, you must have your parent or guardian's consent and they must supervise your participation in Common Voice.

Common Voice is part of the Mozilla Community. As a result, if you chose to participate, you agree to follow the Mozilla Community Participation Guidelines. 2. Your Contributions

We make Mozilla's Common Voice dataset available under the Creative Commons CC0 public domain dedication. That means it's public and we've waived all copyrights to the extent we can under the law. If you participate in Common Voice, we require that you do the same. You agree that Mozilla may offer all of the contributions you make available to Common Voice, including text, recordings, validations, and feedback (the "Contributions") to the public under the CC0 public domain dedication.

In order to participate in Common Voice, Mozilla also requires that you make three assurances:

First, that your Contributions are entirely your own creation.

Second, that your Contributions do not infringe on any third parties' rights.

Third, that your Contributions comply with Mozilla's Acceptable Use Policy.

If you cannot make these assurances, you may not participate in Common Voice.

In addition, if you participate in the Common Voices Spontaneous Speech project, you agree not to include any sensitive or personal data about yourself and others in the recordings you submit. 3. Your Account

You do not have to create an account to participate in Common Voice.

To participate in some additional features in Common Voice (such as adding small or bulk sentence submissions, reviewing sentences, or adding or transcribing prompts), you will need to create a Mozilla account. You can create a Mozilla account here. Your use of Mozilla accounts is governed by the Mozilla Accounts Terms of Service and Mozilla Accounts Privacy Notice. If you create an account, Mozilla will ask for your email address and a username of your choice. Optionally you may also provide an avatar and certain demographic data through the Common Voice Platform. Demographic data helps us and other researchers improve and create speech-to-text technology and tools.

If you participate in the Alpha or Beta testing for the Common Voice Spontaneous Speech project, you must provide your email address so that we can send you a link to log in to participate in the Common Voice Spontaneous Speech project.

When you provide information for your Mozilla account in connection with Common Voice, you give Mozilla all permissions necessary to: keep track of information about your Contributions, associate those Contributions with your account, email, username, and the demographic information you provide, publish your Contributions publicly along with any demographic information, and publish metrics about your Contributions (such as number of recordings and languages) along with your username on the leaderboard.

Mozilla will not publicly post or publish your email address.

You can choose not to appear on the leaderboards. If you do, Mozilla will not publish data about your recordings in association with your username. However, Mozilla will still make your text and recordings publicly available as part of Common Voice and will include information about your recordings in the overall metrics that are available publicly.

4. Communications

If you subscribe to receive our newsletters or register for a Mozilla account in connection with Common Voice, you may receive emails from us in connection with your account. 5. Disclaimers

By participating in Common Voice, you agree that Mozilla will not be liable in any way for any inability to use Common Voice or for any claim arising out of these terms. Mozilla specifically disclaims the following:

Indirect, special, incidental, consequential, or exemplary damages, including without limitation direct or indirect damages for loss of goodwill, work stoppage, lost profits, loss of data, or computer malfunction.

Any liability of Mozilla under these Terms is limited to $500.

You agree to indemnify and hold Mozilla harmless for any liability or claim that results from your participation in Common Voice.

Mozilla provides Common Voice "as is." Mozilla specifically disclaims any legal guarantees or warranties such as "merchantability," "fitness for a particular purpose," "non-infringement," and warranties arising out of a course of dealing, usage or trade. 6. Notices of Infringement

If you think something in Common Voice infringes your copyright or trademark rights, please see our policy for how to report infringement. 7. Updates

Every once in a while, Mozilla may decide to update these Terms. We will post the updated Terms online.

If you continue to use Common Voice after we post updated Terms, you agree to accept that this constitutes your acceptance of such changes. We will post an effective date at the top of this page to make it clear when we made our most recent update. 8. Termination

Mozilla can suspend or end anyone's access to Common Voice at any time for any reason. If we decide to suspend or end your access, we will try to notify you by the email address associated with your account or the next time you attempt to access Common Voice.

The Contributions you submit to Mozilla will remain publicly available as part of Common Voice, even if we terminate or suspend your access. 9. Governing Law

California law applies to these Terms. These Terms are the entire agreement between you and Mozilla regarding Common Voice.

Here is a directly comparable ToS for a Public AI Data Flywheel (PAIDF)

Opt-in Data Flywheel Legal Terms (CV-style, with license choice + external auth)

Effective [DATE]

Through the Opt-in Data Flywheel, you can donate your chats, chat feedback, and the other resources we need to build an AI dataset that anyone who meets certain criteria can use to advance public-interest AI evaluation, research, and development.

You may only participate in the Public AI Data Flywheel (PAIDF) if you agree to these Public AI Data Flywheel Terms (the "Terms"). 1. Eligibility

The PAIDF is open to anyone over the age of 19. If you are 19 or under, you must have your parent or guardian's consent and they must supervise your participation in the PAIDF.

The PAIDF is part of the public AI network community. If you choose to participate, you agree to follow our Community Participation Guidelines, based on guidelines from Mozilla. 2. Your Contributions

License selection (per item). For each contribution you submit to the PAIDF (including text, recordings, validations, and feedback, collectively, the "Contributions"), you must select a license from the following: CC0-1.0, CC BY 4.0, or CC BY-SA 4.0. You agree that we may make your Contributions publicly available under the license you select for that item.

AI preference signals (per item). You may attach an AI preference signal to a Contribution. We will store, transmit, and display that signal with the Contribution and document how our systems interpret such signals. We cannot guarantee that downstream users or providers will honor such signals.

Assurances. To participate, you must assure that:

```
your Contributions are entirely your own creation;

your Contributions do not infringe on any third parties' rights; and

your Contributions comply with Our Acceptable Use Policy.
```

If you cannot make these assurances, you may not participate in the PAIDF.

In addition, you agree not to include any sensitive or personal data about yourself or others in the submissions you provide. 3. Your Account

No Flywheel accounts. You do not need a native Flywheel account to participate.

External authentication (optional). You may authenticate using an external account, currently Hugging Face or an Open WebUI instance. Your use of those external services is governed by their respective terms and privacy notices. When you authenticate, we may receive limited identifiers those providers share with us (for example, a username or user ID).

Public attribution and metrics. We may associate your Contributions with the public handle you select (including a pseudonym) and may publish metrics about your Contributions (such as number of submissions and languages) along with that handle on a leaderboard. We will not publicly post or publish your email address.

4. Communications

If you interact with us using an external account in connection with the Opt-in Data Flywheel, you may receive emails or messages from us in connection with your participation. 5. Disclaimers

By participating in the PAIDF, you agree that we will not be liable in any way for any inability to use the PAIDF or for any claim arising out of these Terms. We specifically disclaim the following:

Indirect, special, incidental, consequential, or exemplary damages, including without limitation direct or indirect damages for loss of goodwill, work stoppage, lost profits, loss of data, or computer malfunction.

Any liability under these Terms is limited to $500.

You agree to indemnify and hold us harmless for any liability or claim that results from your participation in the PAIDF.

We provide the Opt-in Data Flywheel "as is." We specifically disclaim any legal guarantees or warranties such as "merchantability," "fitness for a particular purpose," "non-infringement," and warranties arising out of a course of dealing, usage or trade. 6. Notices of Infringement

If you think something in the PAIDF infringes your copyright or trademark rights, please see our policy for how to report infringement. 7. Updates

Every once in a while, we may decide to update these Terms. We will post the updated Terms online.

If you continue to use or participate in the PAIDF after we post updated Terms, you agree to accept that this constitutes your acceptance of such changes. We will post an effective date at the top of this page to make it clear when we made our most recent update. 8. Termination

We can suspend or end anyone's access to the PAIDF at any time for any reason. If we decide to suspend or end your access, we will try to notify you by the email address associated with your external account (if shared with us) or the next time you attempt to access the PAIDF.

The Contributions you submit will remain publicly available as part of the PAIDF under the license you selected for each item, even if we terminate or suspend your access. 9. Governing Law

California law applies to these Terms. These Terms are the entire agreement between you and us regarding the Opt-in Data Flywheel.