

WIP: Collaborative Design of Contribution Tracking Systems for Decentralized Organizations

Nicholas Vincent
nickvincent@u.northwestern.edu
Northwestern University
USA, Illinois, Evanston

Christine Vandevoorde
christine@govrn.io
Govrn
USA

ABSTRACT

In this short position paper, we discuss how researchers might support “collaboratively designed contribution tracking systems” (e.g., systems to record the contributions of DAO members or contributors to open source projects) and highlight how such systems are well-suited to navigating trade-offs in terms whether contributions are recorded passively (i.e. via logging) or actively (i.e. via forms) and whether contributors have “use-awareness”, i.e. do they know how their contributions will or might be used. We propose a set of opinionated defaults for kick-starting bottoms-up collaborative design of contribution record keeping, and discuss our work in progress aimed at answering empirical questions in this space.

1 INTRODUCTION

There are many contexts in which a group of people (e.g., members of a DAO, or “Decentralized Autonomous Organization”) would want to collaboratively design and use a system for keeping track of contributions to some shared cause. Such a system would allow the people who make contributions to play a role in deciding on a record schema, the design of record creation interfaces (e.g., logging forms), and even how records will be used (e.g., for analysis, for compensation, or for producing user-specific recommendations). To provide a very concrete example, a team working on a software project may want to record contributions to the project in a structured fashion, and all the team members might have opinions – that vary by member, and even change over time – about how those logs should be structured. Or, a group may wish to share their reviews of some class of products to engage in collaborative filtering [5]. In such cases, the group must design an interface for record creation and make decisions about what things should be recorded and how those records should be classified. Furthermore, in many contexts, different contributors will have different opinions about how specific records should be classified or tagged.

Users of collaboratively maintained contribution tracking systems (or record keeping systems more generally) must navigate trade-offs in terms of how data collection occurs. We are especially interested in two specific design dimensions: active vs. passive data collection and use-aware vs. use-unaware data collection. By selecting a set of “initial conditions” for a particular contribution tracking system, we argue it may be possible to make these trade-offs highly visible to contributors, and ultimately support bottoms-up governance of such systems. We also discuss ongoing research to answer empirical questions about how people navigate these trade-offs.

We present a proposal for initializing a collaboratively designed contribution tracker that includes feedback loops so contributors can weigh in on record schema, how contribution records are used, and how preferences are aggregated (where preferences can pertain

to the record schema, record usage, the ranking of contributions, and even preference aggregation itself). We also discuss our near-term research plans to study a prototype version of such a system that focuses on collectively “cultivating” record schema and using contribution records to perform value assessment. We hope our arguments for this approach foster discussion in the emerging decentralized governance research community about this research direction.

Different record keeping systems vary in how much they emphasize passive data contributions – passively generated as people go about their regular activities like committing code or consuming content – and active data contributions – people take explicit actions to generate records, potentially with the intention of impacting other people and downstream systems and processes. We can also think of the active vs. passive distinction as an *implicit* vs. *explicit* distinction, drawing on the classic dichotomy of implicit feedback for recommender systems (i.e. “did the user watch a movie”) and explicit feedback (i.e., “what star rating did the user voluntarily give the movie”). See e.g. [9] for an early discussion and Chapter 5 of [7] for very recent discussion.

The case for distinguishing between use-aware and use-unaware data collection arises from recent scholarship that emphasizes the idea that many kinds of data are produced via under-discussed “data labor” by the public [1, 10] and that the public may have an untapped source of leverage called “data leverage” [16]. In general, active data collection methods are more likely to give contributors use-awareness (e.g., a movie rating system that communicates to users that any movie ratings they provide will be used to improve movie search for other users), but it is not guaranteed that active data collection always fosters use-awareness, or vice versa – we discuss some counterexamples below (see e.g. Fig 2). Furthermore, use-awareness can be thought of as subject to change over time, as new uses of a given set of records can always arise in the future (a prime example being the development of large language models that use code from GitHub [3]). One way contribution records might be used in the DAO space is to determine how much compensation contributors receive; in this context, use-awareness may be high stakes.

We are interested in two research questions about these dimensions of contribution records. First, very few systems let data creators have a say in how the active vs. passive data collection trade-off is navigated. When given the chance, how will different groups of contributors collectively navigate trading off between active and passive data collection? More specifically, (a) will contributors change the data schema, and (b) will these changes generally make data entry more active (e.g., by adding new dimensions to records) or more passive (e.g., by replacing active record creation forms with

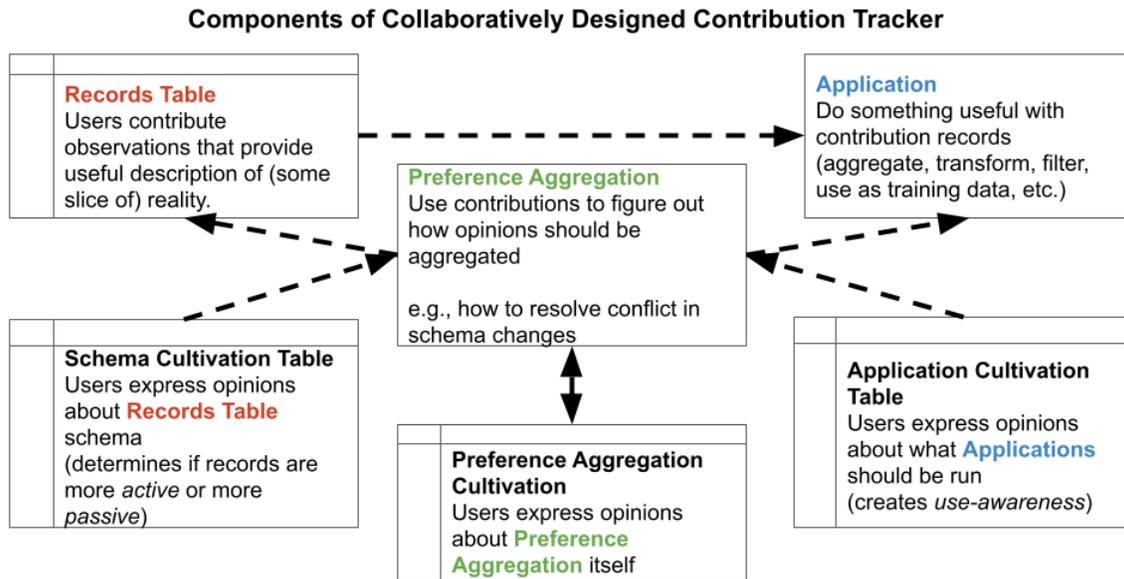


Figure 1: Visual summary of the components of a starting point / opinionated default for a collaboratively designed contribution tracker.

passive forms of contribution recording like scraping other data sources)?

Second, if data creators are given more use-awareness about the record keeping system (i.e., told how data contributions can affect downstream processes or told how their data contributions will be valued), will they change their data contribution patterns? We can examine this question both quantitatively – by examining observed and self-reported changes in data contribution behavior – and qualitatively, by talking directly to contributors about their experiences with such a “use-aware” system.

By examining the impact of these two dimensions, we can provide insights that are broadly useful for a large class of data-driven systems that includes both “governance” systems and more traditional “AI” applications like recommendation and classification.

2 RELATED WORK

This work builds on a variety of recent initiatives to implement on decentralized governance.

A primary motivating use case for contribution tracking is identifying and measuring flows of value. Our proposed work draws heavily on progress from authors of the SourceCred tool, who have tested techniques for thinking of contributions in terms of flows of value along a contribution graph [8]. Our work also relates to research efforts that have studied blockchain governance in practice used ethnographic techniques [11].

Our work is additionally highly related to the concepts of “modular politics” [12] and “peer governance” [4]. Notably, we similarly take a bottoms-up approach, and aim to create an iterative process that allows for “dissensus” [2]. While we have not yet directly connected our planned experiments with the open standard of Modular

Politics or a software infrastructure like PolicyKit [17], this is a ripe area for future work. More generally, we see this work as aligned with efforts to experiment with online governance [15]. We plan to mix in the wild experiments (specifically, with active DAOs) with simulation experiments in the spirit of early collective action work [6].

3 PROBLEM DEFINITION

We consider the broad category of problems in which a group of people want to work collaboratively on some kind of record keeping task where each record corresponds to a unit of “contribution”; each individual contributor has some opinions about (1) the record schema, (2) the design of the record creation interface, and (3) how the records should be used. To keep the discussion of these questions tethered to specific examples, we will refer to our example use-cases from above: a team working on a software project that keeps “contribution logs” (i.e., something like logs kept for a software development shop to bill clients) and a group of friends who want to run their own collaborative filtering movie recommender system. We might further imagine each of these groups is a DAO, though our main ideas can apply to non-DAO groups as well, especially those that exhibit networked, distributed, or consortium structures.

We assume that contributors want to structure their contribution records in some fashion (i.e., enforce a data schema), but that each group and each context might be better suited for different kinds of structure. For instance, perhaps the software team wants each contribution log entry to be labeled with a set of categories (but each member has a different idea about the ideal number of categories, and whether or not certain kinds of contributions could be accounted for via post-hoc analysis). Similarly, perhaps the friends

trying to bootstrap a recommender system want to debate the merits of thumbs-up/thumbs-down style ratings versus 5-star ratings vs. “score out of 100” ratings, or are interested in rating items along many dimensions (e.g., a 1-10 funniness score, a 1-10 drama score).

Importantly, changes to these schema may fundamentally change how data contributions are collected, i.e. what is the “input vocabulary” and what do “input forms” look like. On the highly unstructured end, perhaps records are completely unstructured text entries of different lengths or images of different sizes. On the highly structured end, perhaps records are generated via a survey form with only yes/no questions. Middle ground examples might include text entries of limited length (e.g., tweets) or something like browsing history.

We take the stance a key goal of *collective maintenance and cultivation* of the system is that the opinions of each data contributor should be taken into account. We are particularly interested in how contributors collectively reach consensus (or not [2]) about updating individual proposed changes to the canonical record schema.

3.1 Active vs. Passive Data Collection and Use-awareness of Data Collection

Different choices regarding the design of the input form, and thus the input vocabulary, will correspond to different levels of how active/explicit certain records are. If records require a contributor to take an explicit action (e.g., fill out an entry box, select from a drop-down, etc.) this is a form of active data creation. If records are generated passively (e.g., someone uploads their browsing history which was generated passively by their browser software), we can say that the resulting records will lean towards being more passive/implicit.

The input form will also impact the dimensionality of the records, i.e. how many columns does a record have (and therefore how many questions/entry boxes are in the form). Some record keeping tasks require higher dimensionality data. For instance, a form that asks users for a single 5-star rating is entirely active, and a form that asks for a 1-10 funniness score and a 1-10 action score is also entirely active, but the second form has twice the dimensionality (two variables on {1, 10} vs just a single variable on {1, 10}).

A related but separate question we can ask about a particular records system is whether a person creating a record knows how that record will be used. We can call this property “use-awareness”. Use-awareness can be very broad, because many kinds of records can have many uses (text data is a helpful example here, especially given a recent trend of using language models to solve a huge variety of different tasks [14]). We can think of broad use-awareness as whether or not a given data creator knows at a high-level what kind of application their records will be used for. We can think of a more specific use-awareness as whether a given data creator can reason about the specific impact of their contributions on downstream models, using techniques from the area of *data valuation* [13]. Higher use-awareness amongst contributors could in turn lead to more opinions being expressed about how the records should be structured.

We hypothesize that use-awareness and activeness are related in the following way: in general, passive data collection has the tendency to lower use-awareness, and use-awareness tends to make

data collection more active. This relationship between data collection dimensions is visualized in Fig. 2.

However, in theory these two dimensions are not perfectly coupled, as it is possible to have passive data collection with high use-awareness (all git contributions are constantly analyzed, but everyone knows this is happening), and to have very active data collection with low use-awareness (everyone takes a super laborious survey at the end of every work day, but nobody really knows what it’s used for). These counterexamples are also included in Fig. 2.

Consider the following example of how use-awareness may impact data creator behavior: a store installs a camera to observe customer behavior, and will use the footage to rearrange products on the shelves based on how people recorded on video act. At first, the camera is hidden. Data creation is entirely passive and creators cannot have any use-awareness. If one person notices the camera, and yells loudly, other customers may start acting differently (e.g., avoiding the camera), with the intention of impacting how the data will be used.

3.2 Key Hypotheses of Ongoing Work

We make the following hypotheses, which motivate our experiments in this space:

- Different contributors will have different opinions about data schema, and trying to aggregate these opinions in some manner can lead to better record keeping systems (though we need not always seek consensus [2]).
- Increasing use-awareness will generally result in systems that better serve the needs of contributors, unless this use-awareness creates excessively large costs in terms of attention.
- In most cases, active data collection will lead to better use awareness, though this can be disrupted by power dynamics (e.g. an employer who hides the purpose of labeling tasks assigned to employees). Use-awareness may also drive interest in active data creation and lead users to express more opinions about record schema.
- If communities can make their own choices about these parameters, and adjust as needed over time, record keeping systems will work better for them!

Below, after outlining what collaboratively design contribution tracking systems that emphasize different aspect of data collections might look like, we briefly discuss ongoing work to test these key hypotheses.

4 PROPOSED STARTING POINT FOR DESIGN

Here, we propose a set of starting components to (1) help data creators collectively decide on a schema for the records they contribute, (2) help data creators determine how active or passive data collection should be, and eventually (3) test the impact of use-awareness on data creator behavior and record usefulness. Because our goal is to make this system collaboratively designed, any choices the research team makes are only a starting point: users should be able to change almost all aspects of the contribution tracker. In our early experiments, we will keep most of these proposed components

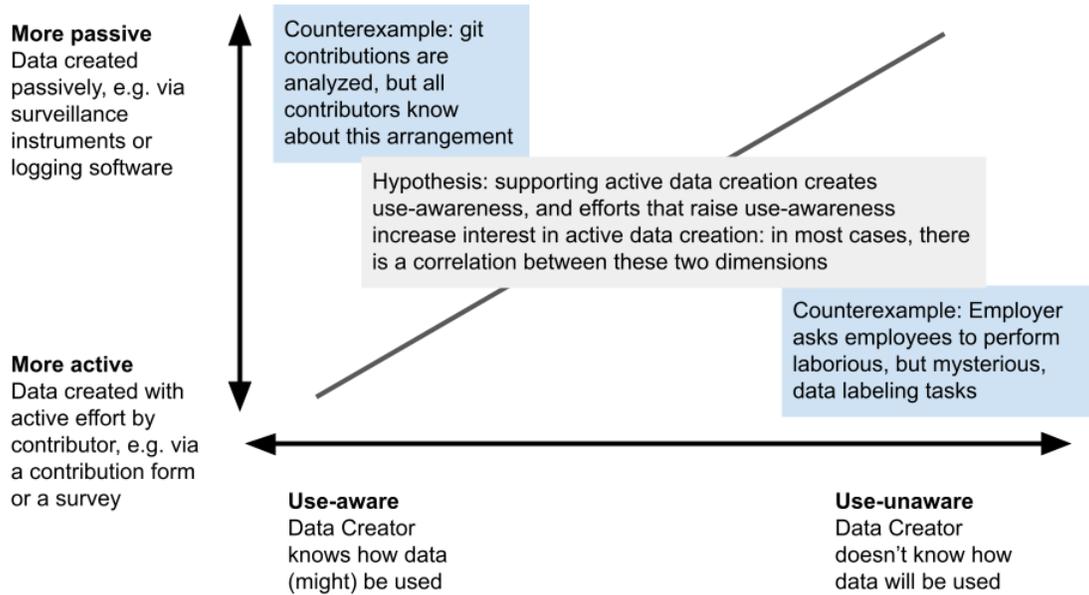


Figure 2: Visual summary of our hypothesized relationship between activeness of data contribution and use-awareness.

fixed but aim to incrementally relax the “top down” constraints on design.

The relationship between components is visually described above in Fig. 1.

4.1 Records and Schema Cultivation

First, there will be one form that stores contribution records in a so-called “Records” data table. This aptly-named table is designed to hold the primary records of interest. In our running examples, a prototypical record might tell us something like “Alice fixed Bug 123” or “Bob watched Star Wars”.

If entries into the Records table can be entirely automated (e.g., a record is automatically generated when viewing a video, work logs are generated as coders make git commits), the system moves towards passive data contribution. The maximally passive record keeping system has no forms at all; all contribution records are generated “naturally” (i.e. through surveillance infrastructure, or more charitably through “workflow logging tools”). The less that surveillance/logging is possible, the more active effort each entry in the Records table will require. In our prototypes, however, it will be up to contributors to decide that they want to make certain records passive.

There will be a second form that enters records into what we call the Schema Cultivation table. We can also think of this as a list of “maintenance requests”. Each record in the Schema Cultivation table represents a contributor expressing an opinion about how Records should be structured. The Schema Cultivation table provides a bottoms-up way for contributors to update data schema and determine how active or passive data contributions should be.

Including something like a Schema Cultivation table also highlights that negotiating these design decisions requires some degree

of active/explicit data contribution, because Schema Cultivation entries are themselves actively collected data. It also introduces the idea that even if data contributors are use-aware, they might not be aware that they are able to propose changes to the record data schema itself, so this should be transparently communicated.

Here, a contributor might express that they think it is wasteful to manually fill out a form every time they commit code – this information could be scraped from commit history. Or they might express they want to change the movie rating schema from a up-vote/downvote to a 5-star approach, or a score out of 100 approach.

4.2 Applications

Next, the system should include a component that keeps track of how contribution records will be used, which we will call the Application table. Each entry describes a process for analyzing or using some Records.

The Application table also has a corresponding Cultivation table, so contributors can weigh in on how their contribution records should be used. An entry might express that a user is unhappy with their contributions being used as training data for machine learning, or conversely a user expressing that they think their contributions are not being used enough. More formally, these records would consist of a description of an application and whether a particular user approves or disapproves of that application.

4.3 Preference Aggregation and Implementation

Finally, the links between “cultivation” tables and actual tables must be handled via some mechanism for preference aggregation

and some mechanism to implement changes based on preference aggregation.

For instance, a very simple high trust default preference aggregation approach might be: whenever anyone expresses an opinion via an entry to the Schema Cultivation table, implement that schema change immediately. A low trust version might, only implement if every contributor submits an identical Schema Cultivation request. A middle ground might involve some kind of majority vote or weighted average, which would require a definition of who is an active user so that a majority or average can be identified.

Finally, in the full version of this architecture, users should also have the option to contribute their opinions about how preference aggregation should work. These opinions will be processed by the existing preference aggregation mechanism, so there are likely major concerns with teams getting stuck in undesirable equilibrium. This is a concern we can study with simulation experiments.

4.4 Summary of Architecture from Contributor Perspective

Here, what summarize what our proposed architecture would like like from a contributor perspective.

A contributor can make five distinct kinds of contributions:

- Submit “active” contribution records
- Generate “passive” records through my activity
- Express opinions about record schema (contribute a record to the Schema Cultivation table)
- Express opinions about how preferences should be aggregated (contribute a record to the Preference Aggregation Cultivation table)
- Express opinions about how records should be used in downstream applications (contribute a record to the Application Cultivation table)

4.5 Opinionated Defaults to Get the System Running

If we assume we will use the above proposed architecture, then the key challenge of the research team in this context becomes *selecting opinionated defaults* for system parameters. The opinionated defaults will serve as a set of initial conditions, and the crowd will adjust these parameters from their initial states over a designated epoch.

As an example, the research team might act in a top-down fashion to design two forms – a Records entry form and Schema Cultivation form – and select a default Preference Aggregation approach to kick things off. The research team must also select how frequently Preference Aggregation will be run.

For instance, an opinionated default that could be sufficient to get this kind of system running may look like:

- (1) The Records form consists of an open text input form for describing a contribution and a drop-down menu with contribution categories.
- (2) The two Applications are to display a dashboard summarizing weekly contributions and to provide a weekly “contribution score” for each user.
- (3) The Preference Aggregation Mechanism is: if more than 50% of active contributors express the same opinion, implement that opinion.
- (4) The Preference Aggregation Mechanism is run every 2 weeks.

We note that in this position paper, we have omitted many of the messy details needed to run the system in practice. Our plan is for our pilot studies, we will use a highly “human-in-the-loop” approach, i.e. the research team will act in cooperation with the team of contributors. For instance, when we “implement an opinion” expressed in the Schema Cultivation table, there will be a process of manual interpretation and implementation in code. In much later stages, this could be semi-automated following the large body of work on smart contracts.

5 PLANNED STUDIES

We are actively working on studies to (1) test our hypotheses in Section 3.2, (2) better understand the impact of different *opinionated defaults* on the outcome of collaboratively designed contribution tracking systems, and (3) more generally test the viability of this design approach. We plan to invite DAOs actively using the Govrn¹ product to participate in these collaborative design pilot studies. (There is currently a waitlist of about 50 DAOs.)

We currently plan to answer these questions by (1) iteratively testing this contribution tracking approach with active DAOs and (2) conducting simulation experiments, including agent-based models, to understand the likely pitfalls with using certain preference aggregation methods.

In our early experiments, we will first focus on studying Schema Cultivation. We will constrain contributor choices regarding Preference Aggregation, Applications, and system parameters like how often the schema is updated (e.g., will suggest an opinionated default of “majority vote is run every two weeks”, and make minor modifications based on DAO-specific concerns).

After conducting small-scale pilot studies that focus on Schema Cultivation, we will focus on contribution value assessment. That is, we observe what happens when participants use the collaborative design approach to decide how different contributions should be assigned relative value.

At each stage of our experiments, we plan to incorporate simulations designed to identify likely failure points and undesirable equilibrium outcomes.

6 DISCUSSION

The collaboratively designed contribution tracker idea opens the door to a number of additional research questions. Here, we briefly discuss several directions that we think may be particularly interesting. We also conclude with a discussion of the implications of this research direction for DAOs in particular.

6.1 Volatility and Equilibria

The proposed architecture necessarily opens the door to highly volatile data schema and record keeping interfaces. It could be the case that Records schema change every other week, potentially

¹<https://linktr.ee/Govrn>

causing data loss and/or making analysis very hard. However, it is possible to select opinionated defaults that make such systems less volatile or impose constraints on the allowable frequency of changes.

Ultimately, in some contexts volatility in the record keeping system may actually more accurately reflect the “collective consensus” (or lack thereof). Or, contributors may end up implementing a steady state where records are relatively unstructured (and then structure can be re-added post-hoc using machine learning and other data processing techniques).

Overall, this is an area where early experiments with active projects may be very informative. As we conduct these experiments, we will look for evidence of volatility in schema, applications, and preference aggregation, and also look for common patterns in terms of what equilibria emerge.

6.2 Connections with Data-dependent Technologies

As noted above, our work is heavily inspired by efforts to measure the value of “data labor” by the public and empower the public to use this value to participate in the governance of data-dependent technologies, i.e. by gaining “data leverage” [16].

Very directly, records in a contribution tracker could be used as training data, so there is great potential to study contribution tracking systems in terms of how different user behaviors impact the performance of downstream data-dependent technologies. Furthermore, tools that communicate to contributors how their choices might affect different technologies could be highly effective in fostering use-awareness, potentially with cascading effects (e.g., contributors gain use-awareness and decide to adjust record schema and/or change how records are used). Testing this will require implementing visualizations or messages that are valid in some sense, i.e. we will need to implement valid “data valuation” techniques [13].

6.3 Implications for DAOs

In our brief proposal above, we have primarily described our proposed design choices in a manner that is agnostic to any particular tech stack, and not necessarily tied to any particular web3 technologies. However, this line of work is particularly relevant to DAOs for two primary reasons: (1) DAOs have an emphasis on collective governance of processes and technology, and (2) DAO member contributions often go unacknowledged, unrecorded, and uncompensated. Thus DAOs are doubly interested in collaboratively designed contribution recording systems because they have implications for contributor compensation and contribution-based decision-making power or governance as well as reputation and credentialing of expertise within the DAO, all in keeping with the DAO ethos of collective design.

Govrn has built a beta protocol and web app for this and has a significant waitlist of interested DAOs and DAO contributors in addition to five currently active DAOs logging contributions with Govrn. These DAOs are recording contributions with opinionated default record schema agreed upon by DAO leaders. The next step is to introduce the collaborative cultivation process outlined above

so that DAO contributors have the ability to cultivate the schema and application of the contribution records.

7 ACKNOWLEDGMENTS

We are very grateful for feedback from Aaron Soskin, Michael Zargham, Ellie Rennie, Livia De, Alex Keating, and Stefen Deleveau.

REFERENCES

- [1] ARRIETA-IBARRA, I., GOFF, L., JIMÉNEZ-HERNÁNDEZ, D., LANIER, J., AND WEYL, E. G. Should We Treat Data as Labor? Moving beyond “Free”. *AEA Papers and Proceedings* 108 (May 2018), 38–42.
- [2] BREKKE, J. K., BEECROFT, K., AND PICK, F. The Dissensus Protocol: Governing Differences in Online Peer Communities. *Frontiers in Human Dynamics* 3 (2021).
- [3] CHEN, M., TWOREK, J., JUN, H., YUAN, Q., PINTO, H. P. D. O., KAPLAN, J., EDWARDS, H., BURDA, Y., JOSEPH, N., BROCKMAN, G., ET AL. Evaluating large language models trained on code. *arXiv preprint arXiv:2107.03374* (2021).
- [4] DE FILIPPI, P., AND SCHNEIDER, N. Editorial: Peer Governance in Online Communities. *Frontiers in Human Dynamics* 3 (2021).
- [5] KOREN, Y., RENDLE, S., AND BELL, R. Advances in collaborative filtering. *Recommender systems handbook* (2022), 91–142.
- [6] MACY, M. W. Chains of cooperation: Threshold effects in collective action. *American Sociological Review* (1991), 730–747. Publisher: JSTOR.
- [7] MCAULEY, J. *Personalized machine learning*. Cambridge University Press, 2022.
- [8] MIYAZONO, E. SourceCred: An Introduction to Calculating Cred and Grain, 2020.
- [9] OARD, D. W., KIM, J., AND OTHERS. Implicit feedback for recommender systems. In *Proceedings of the AAAI workshop on recommender systems* (1998), vol. 83, pp. 81–83. tex.organization: AAAI.
- [10] POSNER, E. A., AND WEYL, E. G. *Radical Markets: Uprooting Capitalism and Democracy for a Just Society*. Princeton University Press, 2018.
- [11] RENNIE, E., ZARGHAM, M., TAN, J., MILLER, L., ABBOTT, J., NABBEN, K., AND DE FILIPPI, P. Towards a participatory digital ethnography of blockchain governance, Feb. 2022.
- [12] SCHNEIDER, N., DE FILIPPI, P., FREY, S., TAN, J. Z., AND ZHANG, A. X. Modular Politics: Toward a Governance Layer for Online Communities. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW1 (Apr. 2021), 16:1–16:26.
- [13] SIM, R. H. L., XU, X., AND LOW, B. K. H. Data Valuation in Machine Learning: “Ingredients”, Strategies, and Open Challenges. In *Proceedings of the Thirty-First International Joint Conference on Artificial Intelligence* (Vienna, Austria, July 2022), International Joint Conferences on Artificial Intelligence Organization, pp. 5607–5614.
- [14] TAMKIN, A., BRUNDAGE, M., CLARK, J., AND GANGULI, D. Understanding the capabilities, limitations, and societal impact of large language models. *arXiv preprint arXiv:2102.02503* (2021).
- [15] TCHERNICHOVSKI, O., FREY, S., JACOBY, N., AND CONLEY, D. Experimenting With Online Governance. *Frontiers in Human Dynamics* 3 (2021).
- [16] VINCENT, N., LI, H., TILLY, N., CHANCELLOR, S., AND HECHT, B. Data Leverage: A Framework for Empowering the Public in its Relationship with Technology Companies. In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency* (2021), pp. 215–227.
- [17] ZHANG, A. X., HUGH, G., AND BERNSTEIN, M. S. PolicyKit: Building Governance in Online Communities. In *Proceedings of the 33rd Annual ACM Symposium on User Interface Software and Technology*. Association for Computing Machinery, New York, NY, USA, Oct. 2020, pp. 365–378.