PARTICIPATORY GOVERNANCE FOR AGI SOCIETAL IMPACT: A CASE STUDY OF AI ETHICS COMMUNITY REFLEXIVITY

Shiran Dudy*
Northeastern University
United States
s.dudy@northeastern.edu

Yanan Long*
StickFlux Labs & University of Chicago†
ylong@uchicago.edu

ABSTRACT

The emergence of artificial general intelligence (AGI) raises pressing questions about how societal institutions will adapt to rapid, automated discovery. This paper presents a case study of participatory design and deliberative polling within a major AI-ethics community (ACM FAccT) to surface shared priorities, tensions, and governance challenges. We describe a two-phase process using seed statements solicited at an in-person workshop followed by an open, online Polis survey. Participants voted on and submitted statements about the future of the conference and broader research field, producing a data-rich map of consensus and disagreement. Our findings highlight themes such as inclusivity, geographic diversity, industry engagement, governance structures, and community values. We discuss how participatory mechanisms can serve as epistemic infrastructure for governing AGI's societal impacts, enabling reflexive oversight and alignment of institutional priorities with diverse human values.

1 INTRODUCTION

AGI has long been considered a distant prospect, but rapid advances in large-scale models, autonomous reasoning and open-ended discovery suggest that its emergence may be imminent. If widely accessible AGI systems can autonomously generate hypotheses, synthesize literature and propose experiments, scientific and technological progress could accelerate dramatically. Yet this acceleration will amplify tensions around epistemic authority, knowledge ownership and socio-economic impacts. Governing these societal impacts will require mechanisms that sustain human accountability while coordinating across diverse communities, including labor organizers, civil society groups, impacted workers, community advocates, and affected publics across regions and income levels. Prior work in participatory design (PD) argues that institutional infrastructures must be deliberately shaped to make such value conflicts legible and negotiable, rather than latent or siloed Karasti (2014); Huybrechts et al. (2017).

Existing scholarship emphasises that AGI governance must incorporate human values and participatory accountability in order to manage societal risks. In particular, communities developing and deploying AI systems need mechanisms for surfacing shared priorities and contested issues Birhane et al. (2022); Delgado et al. (2023). Participatory design and deliberative polling provide such mechanisms by engaging diverse stakeholders, capturing preferences and disagreements, and informing institutional decision-making. We build on these approaches by applying them to the governance of a prominent AI-ethics conference (ACM FAccT) and reflecting on how the participatory design process can be adapted to AGI governance contexts. Our case study focuses on how PD can scale to community-level reflexivity, a challenge that has become more salient as participatory AI efforts seek broader reach without losing depth Suresh et al. (2024). FAccT is itself a site of active self-reflection about participation and representation: studies of conference demographics highlight ongoing concerns about geographic concentration and the risks of reproducing WEIRD perspectives,

*equal contributions
†Work began while at the University of Chicago

while analyses of the conference's intellectual trajectory underline both achievements and persistent gaps in accountability practices Septiandri et al. (2023); Laufer et al. (2022). Beyond FAccT, participatory AI research has expanded to include co-design in journalism and broader critiques of participation-as-solution narratives, stressing that participation must be paired with governance structures that redistribute power rather than simply solicit input Tseng et al. (2025). Our study contributes to this literature by offering a participatory, data-rich snapshot of shared priorities and tensions in a community that already foregrounds governance.

We situate our research within the context of a recent work by Blili-Hamelin et al. Blili-Hamelin et al. (2025) who urges the community to reflect on the notion that AGI is the north star goal of AI research. A particular concern has been raised regarding the use of the term AGI, which may create an illusion of consensus among community members, despite its multiple, distinct interpretations. Another concern has revolved around the exclusion of communities to take part in setting the goals for AI research more broadly. Addressing those two concerns was met by several recommendations, one of which aimed at clarifying the goals of AI research and do so by fostering conditions for greater inclusion of communities, disciplines, and resources. In this work we aim to make steps towards that end, by conducting participatory AI with the FAccT community members, listening to voices from the wider community on their visions in the context of their AI research community.

2 METHODS

We conducted a large-scale participatory design study during the 2025 ACM FAccT conference. In Phase 1, a CRAFT session invited participants to discuss the future of the conference and co-create seed statements capturing desired changes, concerns and values. Statements were clustered and lightly edited for clarity, following PD norms that emphasize shared ownership of framing and language Karasti (2014). In Phase 2, we deployed an online Polis survey that displayed seed statements one at a time and allowed respondents to vote "agree", "disagree" or "pass" and submit additional statements. Polis's design avoids reply threads, minimizing flame wars and focusing attention on voting patterns, which aligns with calls for participatory infrastructures that keep deliberation legible at scale Huybrechts et al. (2017).

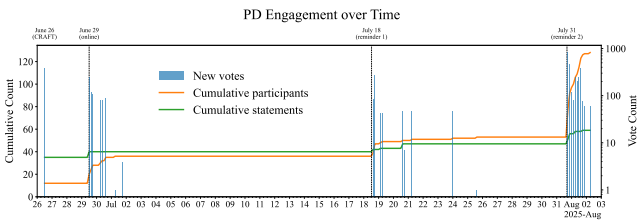
The survey remained open for one week following the conference. Overall, 59 distinct statements were voted upon by 128 participants, yielding $4,531$ votes. We collected metadata on voting behaviour (counts of agree, disagree, pass) and analysed the content using thematic coding. To classify statements by theme, we employed a mix of human coding and large language models (LLMs); consensus between coders was moderate, with Krippendorff's $\alpha = 0.77$ and Fleiss $\kappa = 0.66$. Statements were assigned to one or more themes such as inclusivity, geographic diversity, industry engagement, governance, conference programming and community values. This approach mirrors prior FAccT research that blends qualitative coding with descriptive statistics to surface community-level patterns Laufer et al. (2022).

Following the information gathering process, we produced a report that was the design artifact, that summarized the findings, and was shared with the FAccT leadership with the goal of facilitating changes based on the needs and hopes of community members.

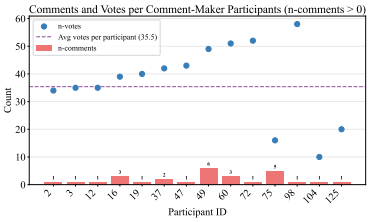
3 RESULTS

The information gathering step Participants generally were willing to engage. The median number of votes cast per participate was $\tilde{n}_{votes} = 40$, where half of participants voted on 40 statements and more or half of the participants voted on 67% or more of the statements (40 of 59 total). The average number of votes cast were $\bar{n}_{votes} = 35.5$ votes per participant, with approximately 80 participants (62%) voting at or above this rate. This indicates that voting was not driven by a small number of highly active voters, but rather reflects *broad engagement across the majority of participants*. See **??**1. This pattern is consistent with other participatory AI efforts, which report uneven engagement even in highly motivated communities Suresh et al. (2024).

Thematic analysis revealed five major priorities. **Inclusivity & Accessibility** encompassed calls for hybrid formats, remote access, tiered pricing and child-care support. **Geographic & Cultural Diversity** highlighted the need to host FAccT outside North America and Europe and to engage

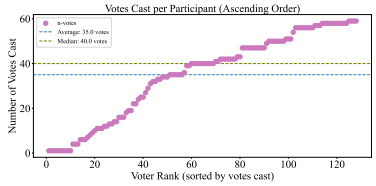


(a) Participation over time as measured by the cumulative rate of new participants casting their first vote (left y-axis, orange line), the cumulative number of statements (left y-axis, green line), and the number of votes cast per 3-hour windows (right y-axis, blue bars).

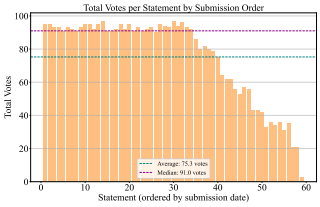


(b) Statement creators spread along the x-axis with bars corresponding to number of statements contributed (in red), blue circles correspond to number of votes. Dashed line indicates average.

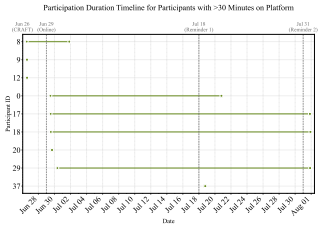
Figure 1: Participant voting behavior by number of votes cast (a) and overall voting distribution (b).



(a) Number of votes cast per participant across all participants. Upper horizontal green line represents median, lower blue line average.



(b) Exposure of statements. Statements ordered by submitted date on x-axis and the number of total votes on y-axis. Median and average are presented in horizontal lines.



(c) Participation duration for participants whose first and last vote are at least 30 minutes apart, 9 out of 128 participants satisfied this criterion.

Figure 2: Participant voting behavior by number of votes cast (a), votes received per statement (b) and duration of participation (c).

researchers from the Global South, echoing prior critiques of FAccT's geographic concentration Septiandri et al. (2023). **Industry Engagement** captured tensions between industry sponsorship (providing resources and opportunities) and concerns about corporate influence. **Conference Format & Formatting** included suggestions to diversify leadership, increase community transparency and adjust review practices. **Community & Values** emphasised retaining the conference's distinct identity as a forum for fairness, accountability and transparency while broadening disciplinary representation.

Notably, some statements elicited strong consensus. For example, participants did show strong majority disagreement with the proposal to reduce sponsorship reliance by eliminating financial support to attendees (72% disagree, moderate engagement), underscoring that attendee support remains a priority even amid debates about corporate involvement. Conversely, opinions were mixed on whether FAccT should be free for PhD students (38% agree, 36% disagree) and whether the conference should accept industry funding only at the minimum level required to operate (38% agree, 28% disagree, 35% pass). These splits resemble tensions reported in participatory AI research that highlight how resource dependence can undercut community autonomy Birhane et al. (2022).

The report In compiling the report, we aimed minimize potential biases that could arise from further processing by additional researchers. In order to maintain a concise report, we focused on presenting the raw statements with the highest consensus rate and presenting those by distinguishing between operational and aspirational statements, i.e. those with actionable and concrete ideas, and those that are more strategic and long-term.

4 DISCUSSION

Our findings underscore the importance of participatory governance mechanisms for communities navigating the societal impacts of AGI. The diversity of statements and voting patterns demonstrates that even within a relatively aligned community, there are significant disagreements about values, priorities, and acceptable trade-offs. Institutional decisions, such as whether to accept corporate funding, how to allocate programme slots, or where to locate the conference, cannot simply be deduced from high-level principles; they require sustained dialogue and legitimate aggregation of community preferences.

Participatory platforms like Polis, combined with human-in-the-loop thematic analysis, provide a scalable approach to mapping areas of agreement and contention. In contexts where AGI systems shape labor markets, public services, and information ecosystems, such tools can help communities remain epistemically sovereign: rather than ceding priority-setting to opaque algorithms or elites, stakeholders and experts can collectively articulate norms and constraints. This resonates with broader calls in participatory AI to move beyond consultation and toward shared governance Tseng et al. (2025). At the same time, our study highlights limitations: participation was voluntary and self-selecting, and there is no guarantee that the sample represents all stakeholders or expert constituencies. Future work should integrate demographic metadata, randomize statement ordering and explore adversarial voting scenarios to enhance robustness.

From the perspective of AGI governance, our case study offers a blueprint for embedding participatory mechanisms into institutions that shape social and economic outcomes. As AGI systems influence labor markets, public services, safety regulation, and information ecosystems, communities will need mechanisms to decide which applications should be constrained, audited, or redirected. Participatory design and deliberative polling can surface collective priorities and ensure that impact mitigation serves diverse human interests. Our findings also illustrate how participatory infrastructures can help academic communities recognize and mitigate entrenched inequities around geography, access, and voice Septiandri et al. (2023).

More broadly, the study suggests that governance of AGI's societal impacts will need hybrid institutions that combine automated analytic capacity with structured, human-led deliberation. Communities will likely face more frequent, higher-stakes decisions about deployment, safety policy, and accountability; participatory mechanisms can function as checkpoints that re-center human values and distribute epistemic authority Huybrechts et al. (2017). This highlights a persistent tension between stakeholder participation and expert-led decision-making: legal, policy, and AI research experts are needed to evaluate technical feasibility, regulatory constraints, and safety, but legitimacy depends on incorporating the lived experience of affected communities. Designing processes that translate between these standpoints is crucial if participatory governance is to avoid becoming either technocratic or purely symbolic Birhane et al. (2022).

One implication is that participatory governance should be multi-layered. Broad stakeholder input can surface value conflicts and lived constraints early, while expert panels can translate these priorities into feasible governance rules, evaluation criteria, and safeguards. Iterative feedback loops between these layers, such as publishing rationales, running follow-up polls, and revisiting contested decisions, can help prevent expert capture while preserving rigorous technical oversight. In this sense, participation is not a single event but an ongoing infrastructural practice that supports accountability over time.

5 LIMITATIONS AND FUTURE WORK

This study is limited by its reliance on a single conference community and a short window of participation. Although the survey attracted meaningful engagement, it likely under-represents individuals without the time or connectivity to participate during the voting period. The analysis also emphasizes aggregate patterns, which can obscure minority perspectives that may be crucial for ethical governance. Future work should pair polling with deeper qualitative follow-up, explore multi-conference or cross-community deployments, and test how participatory governance scales in settings with higher stakes and more heterogeneous stakeholders Birhane et al. (2022); Delgado et al. (2023). Scaling this practice across venues is contingent on conference leadership commitment to broadening participation in AI research decision-making and acting upon it through participatory

design. In sum, our participatory design study shows that members of the AI-ethics community can co-create and prioritize governance statements about their flagship conference, revealing both consensus and disagreement on inclusivity, global representation, industry relationships, and governance structures. Participatory mechanisms therefore remain vital epistemic infrastructure for aligning institutional responses to AGI's societal impacts with human values, and they merit continued investment as governance challenges intensify.

REFERENCES

Abeba Birhane, William Isaac, Vinodkumar Prabhakaran, Mark Diaz, Madeleine Clare Elish, Iason Gabriel, and Shakir Mohamed. Power to the people? opportunities and challenges for participatory ai. In *Proceedings of the 2nd ACM Conference on Equity and Access in Algorithms, Mechanisms, and Optimization*, pp. 1–8, 2022.

Borhane Blili-Hamelin, Christopher Graziul, Leif Hancox-Li, Hananel Hazan, El-Mahdi El-Mhamdi, Avijit Ghosh, Katherine A Heller, Jacob Metcalf, Fabricio Murai, Eryk Salvaggio, et al. Position: Stop treatingagi'as the north-star goal of ai research. In *Forty-second International Conference on Machine Learning Position Paper Track*, 2025.

Fernando Delgado, Stephen Yang, Michael Madaio, and Qian Yang. The participatory turn in ai design: Theoretical foundations and the current state of practice. In *Proceedings of the 3rd ACM Conference on Equity and Access in Algorithms, Mechanisms, and Optimization*, pp. 1–23, 2023.

Liesbeth Huybrechts, Henric Benesch, and Jon Geib. Institutioning: Participatory design, co-design and the public realm. *CoDesign*, 13(3):148–159, 2017.

Helena Karasti. Infrastructuring in participatory design. In *Proceedings of the 13th Participatory Design Conference: Research Papers-Volume 1*, pp. 141–150, 2014.

Benjamin Laufer, Sameer Jain, A Feder Cooper, Jon Kleinberg, and Hoda Heidari. Four years of facct: A reflexive, mixed-methods analysis of research contributions, shortcomings, and future prospects. In *Proceedings of the 2022 ACM conference on fairness, accountability, and transparency*, pp. 401–426, 2022.

Ali Akbar Septiandri, Marios Constantinides, Mohammad Tahaei, and Daniele Quercia. Weird faccts: How western, educated, industrialized, rich, and democratic is facct? In *Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency*, pp. 160–171, 2023.

Harini Suresh, Emily Tseng, Meg Young, Mary Gray, Emma Pierson, and Karen Levy. Participation in the age of foundation models. In *Proceedings of the 2024 ACM Conference on Fairness, Accountability, and Transparency*, pp. 1609–1621, 2024.

Emily Tseng, Meg Young, Marianne Aubin Le Quéré, Aimee Rinehart, and Harini Suresh. " ownership, not just happy talk": Co-designing a participatory large language model for journalism. In *Proceedings of the 2025 ACM Conference on Fairness, Accountability, and Transparency*, pp. 3119–3130, 2025.