First Monday
PEER-REVIEWED JOURNAL ON THE INTERNET

Participation versus scale: Tensions in the practical demands on participatory AI

by Meg Young, Upol Ehsan, Ranjit Singh, Emnet Tafesse, Michele Gilman, Christina Harrington, and Jacob Metcalf

Abstract

Ongoing calls from academic and civil society groups and regulatory demands for the central role of affected communities in development, evaluation, and deployment of artificial intelligence systems have created the conditions for an incipient "participatory turn" in AI. This turn encompasses a wide number of approaches — from legal requirements for consultation with civil society groups and community input in impact assessments, to methods for inclusive data labeling and co-design. However, more work remains in adapting the methods of participation to the scale of commercial AI. In this paper, we highlight the *tensions* between the localized engagement of community-based participatory methods, and the globalized operation of commercial AI systems. Namely, the scales of commercial AI and participatory methods tend to differ along the fault lines of (1) centralized to distributed development; (2) calculable to self-identified publics; and (3) instrumental to intrinsic perceptions of the value of public input. However, a close look at these differences in scale demonstrates that these tensions are not irresolvable but contingent. We note that beyond its reference to the size of any given system, scale serves as a measure of the *infrastructural investments needed to extend a system across contexts*. To scale for a more participatory AI, we argue that these same tensions become opportunities for intervention by offering case studies that illustrate how infrastructural investments have supported participation in AI design and governance. Just as scaling commercial AI has required significant investments, we argue that scaling participation accordingly will require the creation of infrastructure dedicated to the practical dimension of achieving the participatory tradition's commitment to shifting power.

Contents

I. Introduction

Many methods to increase fairness of algorithmic systems do little on their own to shift power from those who design and deploy them. Indeed, the power and the capacity to shape AI systems is largely vested in the occupational class of "the coding elite" — the politically and economically privileged developers employed by technology firms and the academy (Burrell and Fourcade, 2021). However, there is now a clear movement towards new models of AI design and deployment within scholarship on AI equity, accountability, and transparency that emphasizes the participation of people who use or are affected by AI/ML systems. This paper interrogates how this "participatory turn in AI" (Birhane, *et al.*, 2022; Delgado, *et al.*, 2023) intersects with the scale of globalized AI systems.

This move within scholarship raises many questions: Whose preferences are used to set the parameters of algorithmic systems (Lee, *et al.*, 2020)? Which datasets should developers use (Paullada, *et al.*, 2021) and how are these datasets created (Denton, *et al.*, 2020)? What assumptions about affected communities underlie the training, testing, and use of algorithmic systems (Scheuerman, *et al.*, 2020; Stark, 2018)? How does the system shift or consolidate political power (Kalluri, 2020)? Who sets the conditions of measuring and assessing an algorithmic system's impact (Metcalf, *et al.*, 2021; Moss, *et al.*, 2021)? Broadly, these questions point towards an expansion of the scope of who participates in the design and governance of algorithmic systems, or as Robinson (2022) puts it, "share moral burdens across a community — rather than delegating to technicians."

Amid these open questions about the nature of participation, we note an emerging trend among developers and practitioners to invest in diverse forms of participation as aspects of their 'AI safety' research and governance priorities, particularly with respect to large language models (LLMs) and their public-facing applications. The domain of AI safety is poorly bounded at present — it encompasses both near-term risks of harm to existing people, and very-long-term 'existential' risks to a hypothetical future. What ties those time scales together is the challenge of coding 'intelligent' probabilistic machines to reliably behave in a manner that is aligned to human values and interests, also known as 'the alignment problem' (Gabriel, 2020; Leike, *et al.*, 2022). In the AI safety domain, participatory methods have emerged as a possible answer to one factor of the alignment problem: to whose values and interests should the machine be aligned, and how should those parameters be elicited from those people?

Typically the approaches favored by major industry actors have treated participation as a route for writing or refining rulesets that will guide the behavior of products that developers have already made. Researchers at DeepMind have proposed the SocioTEchnical Language agent Alignment (STELA) method, which uses intensive, small focus-groups consisting of under-represented groups for "community-based rule elicitation" (Bergman, *et al.*, 2024). Inspired in part by the Design from the Margins methodology (Rigot, 2022; Harrington, 2020), STELA seeks to establish a general ruleset by inviting a small number of people from "differently situated" groups historically excluded from social power. For example, they found that starting from the perspective of marginalized groups elicited a distinct set of fairness principles. In contrast, researchers from Anthropic have focused on a participatory governance model they call "constitutional AI" that would limit the behavior of machines in the way a constitution sets limits on the operation of a government (Bai, *et al.*, 2022; Roose, 2023; Anthropic, 2023a, 2023b). This solicits crowdsourced input from a much larger and broader set of people to provide feedback on a proposed constitution derived from global human rights documents, to inform the reinforcement learning stage of model development [1]. OpenAI has used a grant-making endeavor to identify methods for "democratic inputs to AI," funding a variety of experimental approaches that translated public or lay-expert inputs into ruleset outputs in the form of 'values cards' that developers can deploy (Eloundou and Lee, 2024). Bridging these different approaches from major developers of AI is the search for a set of rules that can be coded into their centralized, powerful models. Public participation serves two purposes in these cases: (1) affording their governance practices a degree of political legitimacy, and (2) further refining their technical safety measures as code. What this type of public participation is not designed to do is grant the public any ability to contest, reject, alter, or assent to the development or deployment of any given system.

In contrast to this approach, the 'participatory turn' in critical AI ethics and fairness literatures is an emergent move towards governance and design practices that enrolls users, affected peoples, or the general public into decision-making processes over how AI systems should be designed and governed. The purpose of public participation is contested, but is often framed as a way to infuse technological systems with public accountability — especially given their widespread societal impacts and disproportionate harms on marginalized people (Birhane, *et al.*, 2022). To date, there is also little agreement on what methods constitute participatory AI; it has yet to have any clear effects of shifting power toward people and communities who are affected by these systems (Delgado, *et al.*, 2023; Groves, *et al.*, 2023). Vital to these efforts is the heuristic that the measure of meaningful participation is the degree to which it localizes decision making power within affected communities (Le Dantec and Fox, 2015; Arnstein, 1969; Harrington, *et al.*, 2019). Importantly, this orientation often anchors participatory methods within specific neighborhoods, organizations, and community groups — it operates on a localized context.

Compare this orientation to the globalized scale at which commercial AI operates today: Twitter/X feeds stretch across national and language contexts, and ChatGPT interacts with users from anywhere with an Internet connection and AI systems for sensitive use cases like hiring and criminal justice are sold across many jurisdictions. Scale is not the same as *size* or *bigness*, even though it is colloquially understood as such. Previous work from the sociology of science reminds us that *scale is the enabling condition for changes in size*. It is the assertion — or imposition — of similarity across many different technological, social, geographic and ecological contexts that are otherwise distinct and diverse (Gillespie, 2020; Ribes, 2014; Tsing, 2000). Put another way, scale is *not localized*. While there is certainly *bigness* in the ideology and practices of AI development and deployment — *e.g.*, large language models with ever more parameters, global race on foundation models and silicon chips, rapid user growth and adoption — it is the *non-locality* and *isomorphism* of scale that enables such growth in size.

In this paper, we argue that this tension in scales of operationalizing commercial AI and participatory design and research can be engaged through understanding scale as a condition of possibility for changing size (rather than size itself), and directing resources to *scaling for participation*. The tension manifests in methodological challenges for AI's participatory turn, such as: How must an impacted community for the purpose of commercial AI design and governance be constituted, especially when it is used across diverse contexts? Who speaks on behalf of a community and represents their concerns, when the relevant "community" for a system consists of many different localities? When is a new participatory engagement necessary, and when are the results of past participatory design and research efforts adequate for determining the design and governance of future technologies? These challenges are salient to both research and policy efforts in sustaining the participatory turn in AI.

In our close reading of scale and participation in Sections II, III, and IV, we find that the primary concern of the participatory tradition is not localism itself, but self-determination and shifting power. Scale need not be a barrier to these goals, even though in practice it often is. As currently operationalized by the industry, scaling is an aggregation of power: it asserts that human lives can be rendered on a continuous computable spectrum by data scientists in pursuit of solving objective functions chosen by industry interests (Moss, *et al.*, 2021). In contrast, when done well, participation is oriented towards disaggregation of power: it acknowledges differences in people's lives and communities, and demands that AI system designers treat people as individuals who are capable of self-governance. Thus, we argue that the participatory turn in AI can only be sustained when this tension is resolved by making additional infrastructural investments in: (1) ways for communities to collectively identify problems with and act on systems across contexts; (2) formal means for communities to designate representatives — or for self-appointed representatives to become more publicly visible and accountable — and (3) technical resources and conventions developed by and for an affected community that can be shared with other impacted communities. We argue that these investments would better support the disaggregation of decision making power in and require investments commensurate with the scale of the development of commercial AI.

II. Scale as social infrastructure

Scale — as a question of how coordinated macro-scale phenomena emerge from many micro-scale systems — has been a topic of consistent inquiry for social theory. Rather than simply implying *bigness*, scalability of any system (whether technical, political, or economic) often presupposes that one system could be applied across different geographies, times, cultures, and contexts *without reconfiguring*, in pursuit of efficiency and profits. Summerson Carr and Lempert (2016) described scale as an "effect of people's conceptual and practical labor," that institutionalizes a particular way of seeing and organizing the world. These context-independent ways of seeing and organizing the world highlight how the work of managing scale is co-extensive with exercising power. Scale is the sociotechnical achievement of asserting a sameness or isometry across contexts where otherwise diverse forms of life, interests, and values are already in place. Tsing (2012, 2005) provides a powerful example of the relationship between scale and power by noting how colonial plantation economies are a prototypical example of scalability: they transform global landscapes through the imposition of sameness across vast contexts, in order to maintain the logic of an (artificial) commodity market.

This points to the insight that scale is made, not found. Scale is achieved through infrastructural investments that make isometry possible (Tsing, 2000). It refers in part to the techniques and tools that actors use — surveys and descriptive statistics, meetings of key stakeholders, and standards to regulate work practices — to represent and cope with the size of an enterprise. Although scale is neither novel to (nor inherent in) digital technologies, these tools and techniques are imbricated in the ways in which scale has become a central organizing principle and a business strategy for software development (Cutler, 2018). They are responsible in part for the rapid propagation of AI systems across social and economic life (Seaver, 2021; Hanna and Park, 2020; Bondi, 2000). For example, social networking platforms seek to provide all users fundamentally the same experience even as their user base grows across the globe to billions, all without breaking under the pressure of ever-increasing computing demands. This capability does not emerge by itself or by accident, but instead is always a partial and temporary achievement based on investment in data centers, software designed to facilitate scale, standardized user interfaces, publicly funded and/or open technical protocols, and moderation policies and practices that comply with multiple jurisdictions — all of which, in concert, enable the scaling of the service from (relatively) local to global while remaining essentially the *same*.

Thus, the scalability of AI systems is the result of massive investments in building social and material infrastructure around living with data and computation including prior telecommunications infrastructure like fiber optic cables (Burrington, 2016), data centers (Hu, 2015), Internet protocols (Edwards, 1998), standards bodies (Mathew, 2014), Internet service providers, natural resources (Crawford, 2021), crowdsourced labor (Gray and Suri, 2019), computer science researchers (Abbate, 2000), defense funding (Edwards, 2010; Mathew 2014), hardware and software development firms (Plantin, *et al.*, 2018), consumer technologies, and more. The political economy of big tech emerges from a particular, historically specific approach to scale that demands that its products and business models become relentlessly devoid of a sense of place, time, context, and community. An immediate consequence of such demands is that commercial actors seek the quickest path to a return on their investment and are thus incentivized to deploy a single model or system, rather than build a diversity of systems that suit varying needs or allow localized communities to demand that systems interact with their communities differently. The ability of global digital systems to operate "at scale" is thus a type of *social infrastructure*: it is an ordering of social relations and background technologies that makes a diversity of scaled systems possible (Star, 1999; Star and Bowker, 2010).

The present-day fact of the deployment of commercial AI over global geographies presents a unique challenge to the application of the participatory traditions in HCI and shared governance described below. As we will detail, participatory governance and design methods converge around a commitment to *locality*; simply put, people are not the same everywhere. A global computational system covers many differing communities that should participate in its design and governance. The size of a single commercial system encompasses a plentitude of small communities and heterogenous needs. Including more communities in participatory methods requires more resources and time, and has practical consequences for whether and how the work of participation is accomplished. However, by distinguishing between scale and size, it becomes possible to ask a slightly different question: how is the size of systems *achieved*? Could it be achieved in a manner with different openings for shared design priorities and governance? Is *dislocation* a necessary artifact of scale — that is, would the participation of one group always be applied outwards to others who were not involved — or could different infrastructures of scalability work to empower affected communities to express a diversity of preferences and interests? What other social infrastructures might be compatible with AI systems and more localized and/or participatory governance?

III. Challenges of organizing participation in AI design and governance

The core commitment of the participatory turn in AI is centering the most impacted people and communities in the design and governance of such systems, and making the design and internal policies and practices of AI systems porous to external input and/or control. Multiple parties have called for such participation as a model for making AI systems more just and equitable (Delgado, *et al.*, 2023). For certain applications — such as cataloging languages underrepresented in AI applications or establishing data rights for indigenous populations — participatory research practices are essential to effectively contextualize and realize the value of the sociocultural data collected (Birhane, *et al.*, 2022; Jo and Gebru, 2020).

There is a growing body of scholarship on AI applications that invoke participation as a central value and methodological commitment, including, but not limited to: citizen juries to integrate the public's values in national responsible innovation agendas (Balaram, *et al.*, 2018); participatory problem formulation to ensure that AI applications are designed for "deep problem solving" that addresses stakeholders' real needs (Martin, *et al.*, 2020); mobilizing the goals and organizing capacity of social and labor movements to govern AI for social good (Bloch-Webha, 2022); integrating the perspectives of refugees and disaster survivors in digital humanitarianism (Berditchevskaia, *et al.*, 2021); engaging with worker perspectives on the design of worker well being measurement and management systems (Lee, *et al.*, 2021); inclusion of advocacy groups in the construction of AI oversight toolkits (Katell, *et al.*, 2020); and voluntary and consented benchmarking models such as Meta's Casual Conversations (Hazirbas, *et al.*, 2022) and Alphabet's Monk Skin Tone scale (Doshee, 2022).

Under increasing public pressure, tech companies have also sought out civil society feedback and participatory engagements. Two recent studies surveying participatory AI practices in commercial AI labs found some common themes that demonstrate the challenges of meaningful participation. Groves, *et al.* (2023) interviewed practitioners at commercial AI labs responsible for engaging the public in participatory AI efforts. They found that even inside of companies doing this work there is little agreement about standards for and the purpose of satisfactory participation, leading to a diversity of *ad hoc* methods within and across companies and being subsumed under risk management frameworks. They claimed that characteristics that would seem to motivate participatory AI adoption — concerns about consumer exploitation, pursuit of profit motive, etc. — actually become barriers to its adoption inside of cautious tech companies unfamiliar with public engagement. Delgado, *et al.* (2023) similarly found that there was no consensus about the how and why of participatory AI among practitioners. They noted, "most current participatory AI efforts consult stakeholders for input on discrete implementation parameters, rather than empower them to make key AI design decisions." They identified a tension between the practitioners' original motivations for participatory work and the practical constraints of the work such as time and resources. This tension results in a shift toward proxies for participation that achieve the appearance of *consultation* without the *empowerment* of affected communities implied by rigorous participatory design or governance practices.

Platform companies have increasingly engaged in attempts at participatory governance not directly related to AI applications, by soliciting both user and expert feedback, partly for practical reasons of scaling their operations and partly in response to intense public backlash (Klonick, 2020; Dvoskin, 2023) and shareholder activism (Bass, 2021; Rydzak, 2022). Caplan (2023, 2022) has described this emergent approach as "networked governance," which spans such diverse efforts as crowdsourced content flagging, consultation with academic experts and civil society organizations, user councils, and trust and safety or content moderation policy committees. Networked governance infrastructures and institutions represent "efforts to distribute the responsibility for content policy-making and enforcement," both to counter claims of centralized power, and to address resource and functional needs required to operate at scale (Caplan, 2023; see also Gillespie, 2018). As Gillespie (2020) pointed out, debates on resources required for content moderation often invoke the "problems of scale" as a "discursive justification" for policy decisions such as investing more heavily in quasi-automated content moderation systems.

The ongoing troubles of tech companies with achieving meaningful participation has engendered concerns that invoking any intervention as "participatory" can imply the consent and collaboration of affected communities in a way that places them at risk of co-optation by more powerful actors (Ahmed, 2020). Sloane, *et al.* (2022) warn of the harms of extractive and exploitative forms of participatory practice in data science — a phenomenon called participation washing. They argue that many participatory consultations run the risk of failing in the face of the impetus toward "context-independent scalability" in developing AI; the key to avoiding these harms is for such participatory consultations to be more *context specific* and about the lived experiences of the consulted communities with particular systems.

A similar set of challenges can be observed in ongoing efforts of legislators in proposing community participation and consultation in various degrees as a regulatory obligation. While municipal-scale participation defines concrete participation mechanisms (Gilman, 2022), national efforts remain underspecified. One global comparison of national strategies for AI governance concluded that they often referred to the value of public participation, but "were usually abstract and consistently overshadowed by other roles, values and policy concerns" (Wilson, 2022). For example, the EU's Digital Services Act distributes content moderation responsibilities to "trusted flagger" organizations tasked with identifying illegal material (Caplan, 2023; Morar and Santos, 2022; European Parliament, 2022), and Article 35 of the General Data Protection Regulation (Council of the European Union, 2016), which governs data collection and processing in the EU, requires that "[w]here appropriate, [data controllers] shall seek the views of data subjects or their representatives" (Council of the European Union, 2016).

In the United States, the Senate has considered — but not passed — some version of an Algorithmic Accountability Act since 2019, specifying that developers must "meaningfully consult" with "representatives of and advocates for impacted groups, civil society and advocates, and technology experts" as often as necessary to conduct algorithmic impact assessments (Wyden, 2022). In light of congressional inaction, the Executive Branch has stressed the importance of public participation in technological regimes. The 2023 Executive Order on Safe, Secure, and Trustworthy Artificial Intelligence (AI), calls for a coordinated, federal government-wide approach toward AI governance that requires "engagement with affected communities" (*Federal Register*, 2023b). For instance, to prevent unlawful discrimination in federal programs using AI, the Order mandates that federal agencies "consider opportunities to increase coordination, communication, and engagement about AI as appropriate with community-based organizations; civil-rights and civil-liberties organizations; academic institutions; industry; State, local, Tribal, and territorial governments; and other stakeholders." This emphasis on public participation echoes similar executive branch statements regarding AI in the White House's Blueprint for an AI Bill of Rights (2022) and the National Institute for Standards and Technology's "AI risk management framework" (National Institute of Standards and Technology, 2023), which are consistent with the Biden White House's push for greater public participation across the regulatory state. Along similar lines, the Executive Order on Modernizing Regulatory Review (*Federal Register*, 2023a) requires that all federal agencies' "regulatory actions ... be informed by input from interested or affected communities." Executive orders, however, are not enforceable as law, and can be overturned by subsequent administrations.

All told, these legislative and regulatory moves around participation thus far illustrate the challenges of participation at scale — local or municipal governments have historically been the most rigorous to date

(Gilman, 2022), whereas the national scale can often be vague or underspecified. Meanwhile, technology companies have not invested in infrastructures for local input and governance in international contexts, which has often resulted in harm. Today we have access to abundant knowledge and know-how to scale up technical infrastructures but less about the social infrastructures that would be necessary to organize meaningful public participation to meet this need. In the next section, we illustrate how understanding the challenge of infrastructuring for meaningful participation requires a deeper understanding of the participatory design tradition and its commitments.

IV. Community-based methods to organize participation

There is a substantial history of research and implementation of participation in technology design, particularly in the field of human computer interaction (HCI) and the approach of participatory action research. A close look at lessons drawn from the participatory tradition demonstrates that while participatory methods tend to act on local context, their primary commitment is not to localism but to shifting power. Participatory methods tend to act on local context, but we can draw lessons from this lineage of participatory methods for modern technologies of scale.

Participatory design in the field of human computer interaction (HCI) enrolls users in the design process with the aim of better supporting their interests and goals (Carroll and Rosson, 2013; Muller and Druin, 2012). Many accounts locate its origins in Scandinavian countries in the 1970s. One seminal project, a collaboration between researchers at a Norwegian computing institute and a union of iron and metal workers (Nygaard and Bergo, 1975) emerged from an effort to defy the ways that datafication was being used unilaterally by employers to rationalize working conditions that harmed workers. The research team worked closely with union workers to understand the effects of computerizing the workplace and to define their own tactics and goals in response. In the successive decades, multiple technology designers have sought input from users in the design process. Participation falls on a spectrum from eliciting one-way feedback from users to a more longitudinal, intensive collaboration between users and designers to “co-design” technologies (Sanders and Stappers, 2008). However, there continue to be ongoing contests over what is considered “participation” (Graham and Rouncefield, 2008; Krüger, *et al.*, 2019); the question of what “counts” as participation is not only methodological, but political (Balka, 2010). Some participatory design work is critiqued for furthering commercial interests (Irani, *et al.*, 2010; Bratteteig and Wagner, 2016; Balka, 2010) rather than pursuing means for design to reflect the culture and values of users (Friedman, *et al.*, 2017; Abokhodair and Vieweg, 2016). In line with this assessment, design methods with a strong commitment to democratic processes and shifting power (Corbett and Loukissas, 2019; Le Dantec and Fox, 2015; DiSalvo, *et al.*, 2008) are primarily practiced at a local level with community groups, neighborhoods, and in shared spaces. For instance, Christopher Le Dantec and Sarah Fox (2015) and Carl DiSalvo, *et al.* (2008) reflect on the value of participatory design as tied to its enactment at a local level that requires deep collaboration to sustain relationships.

In the context of sustaining relationships, participatory design in HCI is also influenced by participatory action research, which is an approach to community-based research that follows the lead of relationships with organizers and organizations affected by a given issue to achieve social change. It is practiced at the scale of communities, defined either geographically or by shared identities (Israel, *et al.*, 1998). Its value commitments were formalized in the late 1970s via a set of principles to: (1) assert the need for community self-determination; (2) foster partnership across all phases of a project; and (3) collaboratively define and solve problems with and for communities (McTaggart, 1991). It aims to restructure decision-making power between researchers and participants, and to support community mobilization (Gaventa, 1988). While its complex history defies an easy origin story, this tradition draws inspiration from the work of Paulo Freire’s (1978) pedagogy of the oppressed, feminist epistemology and praxis, and grassroots civil rights and workers movements across the U.S., Latin America, and South Asia.

The influence of participatory action research on HCI has brought attention to the contours of race and power in participatory design as a vector for unintentional harm (Harrington, *et al.*, 2020, 2019). Scholars have examined power relations in HCI participatory design workshops and found a negligence of community strengths, lack of historical context, and silencing of participants' experiences in the rooms where "engagements" happen. These scholars aim to extend the participatory action research tradition (Fox, *et al.*, 2017; Hayes, 2011) with an explicit focus on race and marginalized communities in prioritizing community self-determination and collaborative problem-solving. Participatory methods only serve democratic ends when community members are considered decision-makers and when they are present and actively involved when project decisions are made.

As HCI draws more deeply from the participatory action tradition, methods are usually applied at the scale of small groups in community centers or collaboratively facilitated by community members themselves (Harrington, *et al.*, 2019; Dillahunt and Veinot, 2018). Harrington, *et al.*, (2019) emphasized the importance of engaging communities as rooted within a specific locality: "We define community as settings where individuals share a geographic proximity in which they work and live, in addition to sharing access to resources. As a result of this proximity, these individuals often share similar societal relations, common environmental challenges, and barriers" (Harrington, *et al.*, 2019). At larger scales, a commitment to individuals' voices and contributions is more challenging to implement in the face of time and resource constraints. Resource-intensive participation also sits in tension with constraints introduced by incentive structures and funding schemes (Young, *et al.*, 2021).

This brief account has only focused on recent community-based interventions in the field of HCI. There is a long and complex history of research on participatory design within computational and social sciences that has explored a multiplicity of methodological approaches and grappled with the problem of what constitutes meaningful participation. Different traditions, applications, and contexts of operationalizing participation have mobilized the concept in diverging ways. Fields with deeper experience in public engagement like urban planning have articulated the instabilities in the meaning of participation as a matter of degree. Along these lines, foundational work by Sherry Arnstein (1969) on the "ladder of citizen participation" asserts a spectrum of citizen involvement in decision-making from non-participation and tokenism to partnership and citizen control. The work of building towards meaningful participation is not a matter of finding the "right" community to engage with or articulating the "true" representation of affected communities. It is rather a matter of grappling with equity in exchange, balance of power relations, and practically achieving community self-determination. In the next section, we dive deeper into the tensions of scaling meaningful participation in the context of commercial AI.

V. How social infrastructure for participation can contend with key tensions

The tensions between the localized focus of the participatory tradition and the international scale of commercial AI raise challenges for pragmatic application of known methods, and require both nuanced negotiation with community self-identification and value commitments that support equity and agency. These tensions are neither resolvable with 'solutions,' nor indicative of incommensurabilities that cannot be addressed. We bring attention to these tensions not to eliminate them — that would be impossible — but to point to the social infrastructures that could sustain a mode of participatory engagement which redistributes power toward the people most affected by AI. To achieve this end, we argue that the participatory tradition remains a valuable resource for commercial AI governance, but will require commensurate forms of time and resource investment in novel infrastructures, social structures, and legal regimes. We identify three tensions in the pursuit of practical application of participatory methods. The first is pragmatic; machine learning models are a difficult design material for engaging with participatory methods. The second relates to divergences in identity categories; identities that are salient to participatory research tradition *vis-à-vis* the operation of technical systems. Finally, the third focuses on the value placed on participation; differences in approaches that rely on placing intrinsic vs. instrumental value on participation in developing AI systems.

Tension 1: Materiality of AI is a constraint for participation

AI is “a new and difficult design material” (Dove, *et al.*, 2017) — it can be slippery and unpredictable (Yang, *et al.*, 2020; Ehsan, *et al.*, 2023). The differences in materiality of AI when compared with traditional design materials is a major reason why user-experience (UX) design innovations in AI are challenging (Dove, *et al.*, 2017). In traditional design, the substrate (design material) is stable and allows designers to anticipate how it will react to user actions. However, designers are not afforded the same luxury with AI — its stochasticity and non-deterministic nature mean that there is no guarantee that if we press the same button twice, the results will necessarily be the same (Dove, *et al.*, 2017; Ehsan, *et al.*, 2023). For example, if you prompt ChatGPT twice with the same prompt, there is no guarantee it will produce the same outputs. Moreover, *what* AI does — its properties — are often emergent and not obvious during use (unlike traditional software engineering where the outputs are more predictable). When designing with AI, it can be hard to ‘hold on’ to the design material, which makes the design practice challenging.

When the design activity is *participatory*, this challenge is compounded. Whereas many participatory design methods in HCI are geared toward engagement with a stable interface, many machine learning models operate behind the scenes of an interface mediating the relationship between data subjects and the model. Conversely, much of the design work of AI systems is tied to decisions that are disconnected from the user-facing interface — from problem formulation, to dataset collection and curation, to model selection and building, which occur well prior to the creation of the application layer that users interact with. In this sense, AI models cannot be treated as a monolith; they evolve through the different lifecycle stages (Suresh and Guttag, 2021) and at a distance from the users. Therefore, at each stage the needs of participation (*who* participates on *what* with *whom*) can be different; at each stage of the development process what participation means varies because of the scope of the work and who the relevant stakeholders are (Ehsan, *et al.*, 2021).

Tension 2: The question of identity is salient to the problem of who can/should participate

The second tension relates to diverging perspectives about who should be invited to participate based on what attributes of a person’s identity are salient to making decisions on a technology at hand. Many participatory methods are tied to the location where a design artifact is deployed, and thus the focus can be on residents of an effected neighborhood or city, or members of a particular community center, school, house of worship, or workplace. At these localized scales, people’s self-identity is immediately relevant, including their intersectional identities along different race, ethnic, gender, religious, or class identities, or shared lived experiences and geographic proximity. Many current applications of AI consider more abstracted identity categories to be relevant, such as facial structure or behavioral advertising labels. Moreover, sometimes labels imposed by a system are wholly incommensurable with the ways a person may self-identify, inflicting a form of data violence (Hoffmann, 2021), such as the limitation to binary gender or the lack of interracial demographic categories. This tension between bottom-up and top-down approaches is a constitutive challenge for any participatory engagement.

Foundational work in AI ethics documents cases in which commercial AI fails along racialized, gendered, and intersectional lines. In particular, Buolamwini and Gebru (2018) found steep performance disparities in facial recognition technology between lighter- and darker-skinned people, particularly darker-skinned women. Even as recent efforts have improved facial recognition model performance on people with darker skin tones, both the problem as diagnosed by Buolamwini and Gebru — and technology firms’ responses — have primarily been based on the reductive attribute of skin color as the salient data category for identity. This focus on calculable phenotype is a common failure mode of other AI systems, where the machine acts on features that can be mathematically rendered, but not on socially or culturally relevant features that cannot be calculated or lack data (Stark, 2018). For example, computer vision technologies may perceive members of the African diaspora as ‘the same’ because of their calculable skin tone (*i.e.*, their skin tone can be rendered as the same number on a standardized scale), but entirely miss socially-situated differences between diasporic communities (*e.g.*, African-American vs Afro-Caribbean) that may have distinct preferences and interests about how to relate to public safety and surveillance that uses computer vision technologies. Industry attempts to repair AI fairness across social difference through participation has thus far been focused on addressing benchmarking tools, such as Google's investment in a more complex skin tone metric for digital photography (Doshee, 2022; Google Responsible AI, 2022), rather than creation of sustained engagements with communities that would illuminate localized differences. This has implications for determining 'who belongs' in AI design efforts: is it the people that the computer groups together according to its own parameters and capabilities, or is it the group that self-identifies according to its own criteria?

This tension becomes readily apparent when attempting to foster engagements with technology design and governance engagements even at the local level. An advocate or community-based participatory design advocate does not ask 'who has which skin tone on this benchmark dataset?' Instead, they would ask about social relationships, such as 'who does this community recognize as a leader or advocate, where does this community gather, who is currently laboring in the interests of this group?' Indeed, there can be hundreds of different ethnic, racial, religious, and other identity groups that are present in one jurisdiction and salient to a person's experience of a given system that may not be readily calculable or remain invisible in available datasets (Onouha, 2018). This does not indicate that participation at scale is impossible, but it does demand that moves between scales of participation are done carefully.

Tension 3: From participation as a means to an end to a commitment to valuing community control

The third tension lies between approaches that instrumentalize participation to improve a technology for profit and those that place intrinsic value on participation less for its outcome, and more for recognizing individual dignity, promoting democracy, and recognizing self-determination. In instrumental approaches to public participation, an engagement is seen as a means-to-an-end, rather than a means to achieve justice (Sloane, *et al.*, 2022). Methods that rely on public participation instrumentally may focus on its usefulness for improving products; these include data labor to improve systems; which Sloane, *et al.* (2022) stated "upholds and improves [machine learning] systems and therefore is valuable for the owners of the[se] systems." It can also include approaches like red-teaming, which is a means to identify ways that a system fails in order to improve a product (Friedler *et al.*, 2023). Some aspects of co-design may also have instrumental ends, as co-creating projects with communities they serve supports a project's ultimate responsiveness to needs on the ground, uptake, and sustainability. Participatory efforts have been criticized as having other instrumental ends aside from improvements to systems; for instance, a participatory effort intended to support trust in institutions may also be instrumentally valuable as a means for institutions to continue the status quo by creating an outlet for disagreement (Ahmed, 2020).

A primary locus of scholarship on the intrinsic value of public participation of public participation is in community-based participatory research and participatory action research, which both focus on shifting power as the primary value of public participation (Arnstein, 1969; Hayes, 2011; Harrington, *et al.*, 2019). Even as these approaches seek to support a person's self-determination, they recognize the challenges of achieving genuine co-creation with affected communities. This commitment requires communities to be involved from the beginning of a project, a factor that itself often relies on the existence of established relationships with community groups that are maintained over time (Martin, *et al.*, 2020). As Martin, *et al.* (2020) described the healthcare risk assessment system that disparately failed black communities: "If the problem formulation step had facilitated the equitable participation of African American community members who have lived experience within the US healthcare system, it is likely this undesired outcome could have been averted" [2]. Building these relationships at scale is difficult. It requires investments in identifying and learning about communities, engaging in culturally sensitive outreach, and supporting community participants with the training, financial, and other logistical resources to participate meaningfully, and recognizing the strengths of the community in relationship with the developers, rather than focusing solely on its deficits to improve (Harrington, *et al.*, 2019; Gilman, 2023, 2022).

VI. How participation can build infrastructure for scalar action

The tensions outlined in the previous section raise a deeply consequential question for participatory methods: What might it mean to infrastructure participation for scale, and how is this infrastructure practically achieved? In this section, we examine three examples by which community-based participation accreted in infrastructures that allowed for distribution of decision-making and knowledge-making power to protect people's rights across diverse contexts. Indeed, we find that infrastructuring for new scales of participation create conditions for advancing equity goals. These outcomes are possible precisely because of the way scale applies the same outcomes isometrically across contexts. Each of these cases demonstrate how attention to the local can be leveraged to create isometric change in larger contexts. Notably, the last two cases are attempts by communities to use data that is usually used against them; these cases underscore how participatory methods require more than just eliciting input from a broader number of people, but require more substantive shifts in power responsive to that input. These examples demonstrate how investments in public processes can support the efforts to shift power and decision-making to those impacted by widely-used AI systems. We consider such efforts as exemplary instances of infrastructuring for participation at scale.

Case 1: Surveillance ordinance as a models for localized governance of systems

First, we consider the initiative, passed in some form in many cities across the United States, to advocate for model surveillance legislation at the municipal level. In 2017, Seattle passed the first Community Control of Police Surveillance (CCOPS) law, aiming to provide community oversight of law enforcement surveillance technologies. Similar laws have since been enacted by over 20 U.S. cities, motivated in large part by concerns about over-policing of Black and brown communities and revelations of secret government surveillance programs (Young, *et al.*, 2019). CCOPS laws generally require city council approval before a police agency can acquire and deploy surveillance technologies, such as predictive policing tools, social media tracking, facial recognition technology, and automated license plate readers (Marlow, 2021).

Under the Seattle CCOPS statute, each city department must prepare a surveillance impact report (SIR) for each technology it is using or considering adopting (Young, *et al.*, 2019). The SIR must describe the technology along with its potential impacts on civil rights, the benefits to the public from the technology's adoption, and a description of public engagement efforts — including public presentations, structured discussions, and public feedback received. The ordinance specifically states that "[t]he community meeting or meetings should be accessible, be noticed in multiple languages, be held in communities impacted by the proposed acquisition, and collect information about potential disparate impacts on disadvantaged groups." A "Community Surveillance Working Group" is then assigned to evaluate the SIR, along with any public comments, then prepare a privacy and civil liberties impact assessment. The surveillance technology must ultimately be approved by the city council and remains subject to an annual equity impact assessment that considers public complaints or concerns.

In Seattle, as elsewhere, the bill's outcomes in achieving meaningful public participation have been mixed (Southerland, 2023; Degroff and Cahn, 2021; Young, *et al.*, 2019). Outcomes include greater transparency about police use of surveillance technologies, which in turn, has led to outright bans or limits on certain technologies. For instance, Seattle, along with San Francisco, Oakland, and Berkeley banned facial recognition technology altogether (Southerland, 2023). However, the CCOPS laws have limitations. The laws do not require the City Council to abide by public input; police responses can be opaque and misleading; city officials have mixed definitional understandings of what AI systems are covered by the law; it has been challenging to engage affected community members; and volunteer community members can lack the resources and capacity to engage in the lengthy and complex review process (Southerland, 2023; Degroff and Cahn, 2021; Young, *et al.*, 2019). Further, at times, the police have ignored the legal requirements entirely, with no accountability (Southerland, 2023) The most pointed critique of the CCOPS approach is that it legitimizes the use of police surveillance technology (Southerland, 2023).

The evolving experience with CCOPS laws provides insights on the issue of scale and participation. The CCOPS statutes navigate the **first tension of materiality** by recognizing that AI tools are malleable and change over time and thus require constant monitoring. For instance, the Seattle statute requires a new round of city council approval when any technology is materially updated, including a "change the purpose or manner in which a surveillance technology may be used." Here, the Seattle statute contends with the **second tension of identity** by requiring that the Community Surveillance Working Group have at least five members from "equity-focused organizations" that are focused on groups historically subject to disproportionate surveillance. The working group also assists the city council in ensuring that "members of vulnerable communities have the opportunity to provide input and feedback" on surveillance technologies throughout the SIR approval process. This codifies the participatory goal of representativeness, although it has been hard to achieve in practice and it is resource intensive to reach impacted communities and provide logistical, financial, and culturally-appropriate support for community members (Gilman, 2022). With respect to the **third tension of valuing participation**, the CCOPS statute indicates some commitments to the intrinsic value of public participation, even as the law does not provide a direct conduit for public input to tech developers; rather, public participation in the CCOPS framework impacts municipal procurement decisions. It has led to effective grassroots organizing and bans of certain technologies deemed harmful by the public. Still, it does not shift decision-making power to the public; they remain in an advisory capacity only. There is no evidence yet that this law has changed the actual design of the technologies at issue, although it might do so as the model of participation that it imbues scales outward.

Case 2: Worker's Algorithm Observatory

Our second case focuses on participatory methods used by the Workers' Algorithm Observatory (WAO), led by Dan Calacci, Samantha Dalal, and Danny Spitzberg across multiple academic institutions and the Mozilla Foundation (Calacci, *et al.*, forthcoming). For gig and platform workers, the algorithmic systems underlying the allocation of work, payment, and other job outcomes are determinative — even constitutive — of a person's working conditions. These systems are also opaque to workers: the double bind of intellectual property barriers and machine learning models' complexity make it more challenging for workers to understand the terms under which they work, and even harder to respond or challenge them. The WAO project directly draws inspiration from the history of Marxist workers' inquiry (Calacci, *et al.*, forthcoming). Marx conceptualized this model as an epistemological challenge to existing modes of social inquiry that relied on the perspective and values of the bourgeoisie to produce numerous national economic studies about the working class. In workers' inquiries, workers are directly surveyed by fellow organizers as experts about their own working and living conditions (McAllister, 2022; Mohandesi and Haider, 2013; Woodcock, 2014). Workers' inquiries served a dual purpose for socialist organizers: (1) to produce knowledge, and (2) to scale their organizing efforts by creating a sense of solidarity among readers across different industries and localities.

Similarly, the WAO project is an intervention to provide platform workers more transparency into their working conditions by reverse engineering answers to key questions about the workings of platforms. The first step in this work is to identify workers' key questions. The WAO team has forged relationships with rideshare driver organizers in two cities who have collaboratively defined high-priority questions and through content analysis of questions posted in forums where platform workers interact. In their efforts to answer these questions, the WAO team created a tool by which drivers could consent to sharing app data directly with the research team. Since gig platforms do not provide an API for collecting data on trips, the tool serves as a resource for a third-party adversarial audit. Furthermore, it does not require participation from a large number of platform workers. A dozen consenting drivers sharing their data with WAO over time provided data on thousands trips. Using initial results from analyzing this data, WAO has demonstrated that rideshare companies have increased their cut of payment from workers over time and showed how pay varies with respect to different types of trip.

Responding to the top-down administration of the gig work that limits the power of individuals to influence platform decision making and the choices available to them in finding work, WAO is an example of 'infrastructuring' participation. The app that workers can install on their phones, the team's work to produce findings and actionable insight for the questions workers indicated were important to them, and the work to share these insights back with rideshare driver organizers and their communities together makes up the sociotechnical data infrastructure produced by WAO. The findings of the WAO are likely to be coextensive to the degree that the operations of the platform app itself are applied isometrically across contexts.

The WAO navigates the **first tension of materiality** by using participation not to co-design the system but to define its impacts. WAO renders the harms that communities face into their primary focus and material of collaborative definition and investigation. These harms are both specific and concrete. Just as the workers' inquiry tradition supports the notion that workers are the authority on their own experiences, rideshare drivers in WAO are experts on the impacts that AI systems have on their lives without needing technical expertise in machine learning or AI itself. The WAO addresses the **second tension of identity** between the question of *who* should participate and on what grounds through both direct and indirect participation: self-identified workers describe their concerns on public forums that WAO takes as input into its analysis. Workers also collaborate with organizers and the WAO team directly to identify high-priority questions. In this way, the act of worker's inquiry itself creates the set of participating workers, and the rationale by which they represent others. WAO does not resolve important questions about who is qualified to represent or speak on behalf of workers; however, the use of broader public forums as an input in their process arguably increases its legitimacy as infrastructure for participation at a broader scale. With respect to the **third tension of valuing participation**, WAO's work is guided by the intrinsic value of participation in workers' inquiry: workers deserve to know and understand their working conditions. However, it is also invested in creating instrumental value for organizers, firms, and the wider public by fostering a public conversation on the working conditions created by gig platforms. While no formal mechanism exists for bringing the findings back to developers to change the platform, the findings do create a launching point for organic and indirect change, such as via worker campaigns to make demands of companies like Lyft and Uber.

Case 3: Invisible Institute — Citizens Police Data Project

Finally, we consider the participatory model used by the Chicago-based investigative journalism organization, The Invisible Institute, to cultivate community collaboration to examine 27,000 complaints against the Chicago Police Department from between 2011–2015. Through a lawsuit, a team of civil rights lawyers and advocates fought for access to the Chicago Police Department's complaint records filed by members of the public (Kalven, 2023; Reynolds-Tyler, *et al.*, 2023; Thomas, *et al.*, 2023). Two court victories led police misconduct complaint records to become public. The Invisible Institute worked with these records to recover the stories buried in the unstructured dataset, further obscured by the police's narrow categorization schema. In the Beneath the Surface project, Trina Reynolds-Tyler, Tarak Shah, their collaborators and community members worked together in 2021 to recruit and train volunteers to survey community members about their experiences with police violence — often resulting in highly personal conversations about painful events (Shah and Reynolds-Tyler, 2023). The narrative qualitative data collected by Invisible Institute's volunteers were used as the foundation for the labeling work. The volunteers read the stories buried within public complaints against the police department to recover lost stories of gender-based violence and police misconduct and re-compile them into a novel dataset. They recovered over 4,000 accounts of gender-based violence, which were then used to train a machine learning tool to analyze the rest of the unstructured dataset.

Members of the community determined the parameters by which data about their lives were to be represented in the dataset and resulting tools. The Invisible Institute's work to convene localized conversations and reasoning about categories of abuse and harm allowed each case to be seen from a new 'scale,' as members of a class or category. All of the outputs from these conversations — the relationships, discussion spaces, new categories, labeled data, and data visualizations — are simultaneously a socio-technical achievement of participation and an enabling condition for advocates, journalists, lawyers, and community members to assert power at greater scales for police accountability. The sociotechnical infrastructure needed to analyze this data included methods to elicit a reflexive understanding of experiences with police violence among community members. In this sense, it was collective labeling of this data that *constituted the community* which was otherwise not visible to itself: those who shared experiences of police violence in Chicago.

With respect to the **first tension of materiality**, the Invisible Institute coded police records themselves in order to produce the data that was used to train their machine learning system using data labels that were inductively defined by community-identified harms. With respect to the **second tension of identity**, the project originated with strong ties to place-based community and especially the communities of color who most experience police violence. The Beneath the Surface project contends with **the third tension of valuing participation** primarily by promoting the intrinsic value in people claiming ownership in something that was done to them — specifically, harms done *to* them — untethered from the emergent outcomes of this process of owning an experience and reliving its trauma.

VII. Conclusion

The hope for a redistribution of power through participation in technology development and governance depends in large part on the epistemic authority of 'situated standpoints' (Haraway, 1988; Corple, *et al.*, 2020; Ehsan, *et al.*, 2023). All people have limited and imperfect knowledge of the world, which can only be achieved from within a contextual 'standpoint'; in many cases, those who occupy a 'standpoint' that is excluded from social and political privilege may have a better understanding of how social and political systems actually operate. For example, the parent who has to negotiate their way through a child welfare bureaucracy may not know how deep learning techniques work, but they likely have a better understanding than any member of the coding elite of how to navigate the consequences and exigencies of a social benefits distribution algorithm (Eubanks, 2018). The moral case for participatory AI is that those who are closest to the consequences of technology have authoritative and unique knowledge of how sociotechnical systems function, and *should* have substantial power to determine how these systems are designed and operate.

In contrast, the epistemic authority of the machine learning industry derives from a directly opposed source: the non-situatedness of modeling. People occupy situated standpoints, models do not. Machine learning is a practice of abstracting away from the type of contexts in which people actually live — a machine learning model is a mathematical description of how to efficiently achieve a desired outcome that is derived from historical data about how that task was carried out many, many times previously. While every piece of training data was collected in a context, the purpose of the abstraction is to find a rule that can be applied generally without reference to context. That model is therefore also abstracted away from core social concepts like 'fairness' and 'justice' (Selbst, *et al.*, 2019). Modeling is definitionally the creation of an abstraction, and the economic value of that abstraction typically increases as it is repurposed outside the context of its collection. It takes a lot of labor and energy to create a machine learning model, so economic efficiency demands it be put to ever-more general purposes. Machine learning models are therefore part of the infrastructure of scale because *models are inherently isometric*, and are now a core aspect of how the technology industry centralizes power. *Modeling imposes sameness across contexts*.

The tensions between scale and participation are therefore not at root about differences in *size*, but about differences in who has authority, how that authority is achieved, and what can be accomplished with the power that authority brings. Size is certainly a constraint — particularly around the resources necessary to achieve meaningful participation — but it is not inherently incommensurable with meaningful participation. Once we recognize scale as a sociotechnical achievement that enables changes in size, then the central question about the success of participatory AI for redistributing power becomes what methods and infrastructures bridge the epistemic authority of participation with the epistemic authority of algorithmic modeling. How can a community meaningfully participate in building datasets that are meant to be an abstract representation? And how can communities meaningfully engage with models based on their datasets that produce insights about their bodies of knowledge and lived experiences? Participatory AI-based tools not only are built on data that represents bodies of knowledge and lived experiences of communities, but also must produce insights that are meaningful for communities and promote their interests. In conclusion, we offer a broad orientation for the ongoing efforts to think through and practically achieve this turn.

About the authors

Meg Young is a Participatory Methods Research Fellow at the Data & Society Research Institute.
E-mail: meg [at] datasociety [dot] net

Upol Ehsan is a doctoral candidate in the School of Interactive Computing at Georgia Tech.
E-mail: ehsanu [at] gatech [dot] edu

Ranjit Singh is a senior researcher at Data & Society.
E-mail: ranjit [at] datasociety [dot] net

Emnet Tafesse is a research analyst with the AI on the Ground Initiative at Data & Society.
E-mail: emnet [at] datasociety [dot] net

Michele Gilman is the Venable Professor of Law at the University of Baltimore School of Law.
E-mail: mgilman [at] ubalt [dot] edu

Christina Harrington is is the Director of the Equity and Health Innovations Design Research Lab at Carnegie Mellon University.
E-mail: charring [at] andrew [dot] cmu [dot] edu

Jacob Metcalf is a researcher at Data & Society, where he leads the AI on the Ground Initiative.
E-mail: jake [dot] metcalf [at] datasociety [dot] net

Acknowledgements

Research for this paper was funded in part by the Siegel Family Endowment and the Open Society Foundation.

Notes

1. Anthropic's constitutional AI is similar to polling (Collective Intelligence Project, 2023) and citizen assemblies (Ovadya, 2023; Seger, *et al.*, 2023) proposed by other researchers to accomplish the same task.

2. Martin, *et al.*, 2020, p. 2.

References

Janet Abbate, 2000. *Inventing the Internet*. Cambridge, Mass.: MIT Press.

Norah Abokhodair and Sarah Vieweg, 2016. "Privacy & social media in the context of the Arab Gulf," *DIS "16: Proceedings of the 2016 ACM Conference on Designing Interactive Systems*, pp. 672–683. doi: https://doi.org/10.1145/2901790.2901873, accessed 10 March 2024.

Alex Ahmed, 2020. "We will not be pacified through participation," *Tech Otherwise* (16 October), at https://techotherwise.pubpub.org/pub/qanxilt9/release/1, accessed 27 September 2022.

Anthropic, 2023a. "Collective constitutional AI: Aligning a language model with public input" (17 October), at https://www.anthropic.com/news/collective-constitutional-ai-aligning-a-language-model-with-public-input, accessed 5 February 2024.

Anthropic, 2023b. "Claude's constitution" (9 May), at https://www.anthropic.com/news/claudes-constitution, accessed 5 February 2024.

Sherry R. Arnstein, 1969. "A ladder of citizen participation," *Journal of the American Institute of Planners*, volume 35, number 4, pp. 216–224.
doi: https://doi.org/10.1080/01944366908977225, accessed 10 March 2024.

Yuntao Bai, Saurav Kadavath, Sandipan Kundu, Amanda Askell, Jackson Kernion, Andy Jones, Anna Chen, Anna Goldie, Azalia Mirhoseini, Cameron McKinnon, Carol Chen, Catherine Olsson, Christopher Olah, Danny Hernandez, Dawn Drain, Deep Ganguli, Dustin Li, Eli Tran-Johnson, Ethan Perez, Jamie Kerr, Jared Mueller, Jeffrey Ladish, Joshua Landau, Kamal Ndousse, Kamile Lukosuite, Liane Lovitt, Michael Sellitto, Nelson Elhage, Nicholas Schiefer, Noemi Mercado, Nova DasSarma, Robert Lasenby, Robin Larson, Sam Ringer, Scott Johnston, Shauna Kravec, Sheer El Showk, Stanislav Fort, Tamera Lanham, Timothy Telleen-Lawton, Tom Conerly, Tom Henighan, Tristan Hume, Samuel R. Bowman, Zac Hatfield-Dodds, Ben Mann, Dario Amodei, Nicholas Joseph, Sam McCandlish, Tom Brown, and Jared Kaplan, 2022. "Constitutional AI: Harmlessness from AI feedback," *arXiv*:2212.08073 (15 December).
doi: https://doi.org/10.48550/arXiv.2212.08073, accessed 10 March 2024.

Brhmie Balaram, Tony Greenham, and Jasmine Leonard, 2018. "Artificial intelligence: Real public engagement," *Royal Society for the encouragement of Arts, Manufactures and Commerce* (30 May), at https://www.thersa.org/reports/artificial-intelligence-real-public-engagement, accessed 10 March 2024.

Ellen Balka, 2010. "Broadening discussion about participatory design: A reply to Kyng," *Scandinavian Journal of Information Systems*, volume 22, number 1, article 7, pp. 77–84, and at at http://aisel.aisnet.org/sjis/vol22/iss1/7, accessed 10 March 2024.

Dina Bass, 2021. "Microsoft agrees to investor push for human rights review of government deals," *Bloomberg* (13 October), at https://www.bloomberg.com/news/articles/2021-10-13/microsoft-agrees-to-investor-push-for-human-rights-review-of-government-deals, accessed 10 March 2024.

Aleks Berditchevskaia, Eirini Malliaraki, and Kathy Peach, 2021. "Participatory AI for humanitarian innovation," Nesta (15 September), at https://www.nesta.org.uk/report/participatory-ai-humanitarian-innovation-briefing-paper/, accessed 10 March 2024.

Stevie Bergman, Nahema Marchal, John Mellor, Shakir Mohamed, Iason Gabriel and William Isaac, 2024. "STELA: A community-centred approach to value elicitation for AI alignment," *Scientific Reports*, volume 14, number 1 (19 March), article number 6616.
doi: https://doi.org/10.1038/s41598-024-56648-4, accessed 10 March 2024.

Abeba Birhane, William Isaac, Vinodkumar Prabhakaran, Mark Díaz, Madeleine Clare Elish, Iason Gabriel, and Shakir Mohamed, 2022. "Power to the people? Opportunities and challenges for participatory AI," *EAAMO '22: Proceedings of the Second ACM Conference on Equity and Access in Algorithms, Mechanisms, and Optimization*, article number 6, pp. 1–8.
doi: https://doi.org/10.1145/3551624.3555290, accessed 10 March 2024.

Hannah Bloch-Wehba, 2022. "Algorithmic governance from the bottom up," *Brigham Young University Law Review*, volume 48, number 1, article number 6, pp. 69–136, and at https://digitalcommons.law.byu.edu/lawreview/vol48/iss1/6, accessed 10 March 2024.

André B. Bondi, 2000. "Characteristics of scalability and their impact on performance," *WOSP '00: Proceedings of the Second International Workshop on Software and Performance*, pp. 195–203.
doi: https://doi.org/10.1145/350391.350432, accessed 10 March 2024.

Tone Bratteteig and Ina Wagner, 2016. "What is a participatory design result?" *PDC '16: Proceedings of the 14th Participatory Design Conference: Full papers*, volume 1, pp. 141–150.

doi: https://doi.org/10.1145/2940299.2940316, accessed 10 March 2024.

Joy Buolamwini and Timnit Gebru, 2018. “Gender shades: Intersectional accuracy disparities in commercial gender classification,” *Proceedings of Machine Learning Research*, pp. 77–91, and at https://proceedings.mlr.press/v81/buolamwini18a.html, accessed 10 March 2024.

Jenna Burrell and Marion Fourcade, 2021. “The society of algorithms,” *Annual Review of Sociology*, volume 47, pp. 213–237.
doi: https://doi.org/10.1146/annurev-soc-090820-020800, accessed 10 March 2024.

Ingrid Burrington, 2016. *Networks of New York: An illustrated field guide to urban internet infrastructure*. Brooklyn, N.Y.: Melville House.

Dan Calacci, Samantha Dalal, Danny Spitzberg, and Andrés Monroy-Hernandez, forthcoming. “FairFare: Towards crowdsourced algorithmic inquiry in the gig economy.”

Robyn Caplan, 2023. “Network platform governance: The construction of the democratic platform,” *International Journal of Communication*, volume 17, pp. 3,451–3,472, and at https://ijoc.org/index.php/ijoc/article/view/20035, accessed 10 March 2024.

Robyn Caplan, 2022. “Networked governance,” *Yale-Wikimedia Initiative on Intermediaries & Information* (29 August), at https://law.yale.edu/sites/default/files/area/center/isp/documents/networkedgovernance_ispessayseries_aug2022.pdf, accessed 10 March 2024.

John M. Carroll and Mary Beth Rosson, 2013. “Wild at home: The neighborhood as a living laboratory for HCI,” *ACM Transactions on Computer-Human Interaction (TOCHI)*, volume 20, number 3, article number 16, pp. 1–28.
doi: https://doi.org/10.1145/2491500.2491504, accessed 10 March 2024.

Collective Intelligence Project, 2023. “We should all get to decide what to do about AI” (23 May), at https://cip.org/blog/alignment, accessed 9 November 2023.

Eric Corbett and Yanni Loukissas, 2019. “Engaging gentrification as a social justice issue in HCI,” *CHI ’19: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. paper number 280, pp. 1–16.
doi: https://doi.org/10.1145/3290605.3300510, accessed 10 March 2024.

Danielle J. Corple and Jasmine R. Linabary, 2020. “From data points to people: Feminist situated ethics in online big data research,” *International Journal of Social Research Methodology*, volume 23, number 2, pp. 155–168.
doi: https://doi.org/10.1080/13645579.2019.1649832, accessed 10 March 2024.

Council of the European Union, 2016. “Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation),” at https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:32016R0679&from=EN#d1e3546-1-1, accessed 10 March 2024.

Kate Crawford, 2021. *The atlas of AI: Power, politics, and the planetary costs of artificial intelligence*. New Haven, Conn.: Yale University Press.
doi: https://doi.org/10.2307/j.ctv1ghv45t, accessed 10 March 2024.

Kim-Mai Cutler, 2018. “The unicorn hunters,” *Logic(s) Magazine*, number 4, at https://logicmag.io/scale/the-unicorn-hunters/, accessed 9 November 2023.

Stevie Degroff and Albert Fox Cahn, 2021. "New CCOPS on the beat: An early assessment of community control of police surveillance laws," *Surveillance Tech Oversight Project (STOP)*, at https://www.stopspying.org/ccops, accessed 10 March 2024.

Fernando Delgado, Stephen Yang, Michael Madaio, and Qian Yang, 2023. "The participatory turn in AI design: Theoretical foundations and the current state of practice," *EAAMO '23: Proceedings of the 3rd ACM Conference on Equity and Access in Algorithms, Mechanisms, and Optimization*, article number 37, pp. 1–23.
doi: https://doi.org/10.1145/3617694.3623261, accessed 10 March 2024.

Emily Denton, Alex Hanna, Razvan Amironesei, Andrew Smart, Hilary Nicole, and Morgan Klaus Scheuerman, 2020. "Bringing the people back in: Contesting benchmark machine learning datasets," *arXiv*:2007.07399 (14 July).
doi: https://doi.org/10.48550/arXiv.2007.07399, accessed 10 March 2024.

Tawanna R. Dillahunt and Tiffany C. Veinot, 2018. "Getting there: Barriers and facilitators to transportation access in underserved communities," *ACM Transactions on Computer-Human Interaction* volume 25, number 5, article number 29, pp. 1–39.
doi: https://doi.org/10.1145/3233985, accessed 10 March 2024.

Carl DiSalvo, Illah Nourbakhsh, David Holstius, Ayça Akin, and Marti Louw, 2008. "The Neighborhood Networks project: A case study of critical engagement and creative expression through participatory design," *PDC '08: Proceedings of the Tenth Anniversary Conference on Participatory Design 2008*, pp. 41–50.

Tulsi Doshee, 2022. "Improving skin tone representation across Google" (11 May), at https://blog.google/products/search/monk- skin-tone-scale/, accessed 10 March 2024.

Graham Dove, Kim Halskov, Jodi Forlizzi, and John Zimmerman, 2017. "UX design innovation: Challenges for working with machine learning as a design material," *CHI '17: Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, pp. 278–288.
doi: https://doi.org/10.1145/3025453.3025739, accessed 10 March 2024.

Brenda Dvoskin, 2023. "Expertise and participation in the Facebook Oversight Board: From reason to will," *Telecommunications Policy*, volume 47, number 5, 102463.
doi: https://doi.org/10.1016/j.telpol.2022.102463, accessed 10 March 2024.

Paul N. Edwards, 2010. ,"Some say the Internet should never have happened," In: W. Russell Neuman (editor). *Media, technology and society: Theories of media evolution*. Ann Arbor: University of Michigan Press, pp. 141–160.

Paul N. Edwards, 1998. "Y2K: Millennial reflections on computers as infrastructure," *History and Technology*, volume 15, numbers 1–2, pp. 7–29.
doi: https://doi.org/10.1080/07341519808581939, accessed 10 March 2024.

Upol Ehsan, Koustuv Saha, Munmun De Choudhury, and Mark O. Riedl, 2023. "Charting the sociotechnical gap in explainable AI: A framework to address the gap in XAI," *Proceedings of the ACM on Human-Computer Interaction*, volume 7, number CSCW1, article number 34, pp. 1–32.
doi: https://doi.org/10.1145/3579467, accessed 10 March 2024.

Tyna Eloundou and Teddy Lee, 2024. "Democratic inputs to AI grant program: Lessons learned and implementation plans," *OpenAI* (16 January), at https://openai.com/blog/democratic-inputs-to-ai-grant-program-update, accessed 8 February 2024.

Virginia Eubanks, 2018. *Automating inequality: How high-tech tools profile, police, and punish the poor*. New York: St. Martin's Press.

European Parliament, 2022. "Regulation (EU) 2022/2065 of the European Parliament and of the Council of 19 October 2022 on a Single Market For Digital Services and amending Directive 2000/31/EC (Digital Services Act) (Text with EEA relevance)," at http://data.europa.eu/eli/reg/2022/2065/oj/eng, accessed 10 March 2024.

Federal Register, 2023a. "Safe, secure, and trustworthy development and use of artificial intelligence" (1 November), at https://www.federalregister.gov/documents/2023/11/01/2023-24283/safe-secure-and-trustworthy-development-and-use-of-artificial-intelligence, accessed 5 February 2024.

Federal Register, 2023b. "Modernizing regulatory review" (11 April), at https://www.federalregister.gov/documents/2023/04/11/2023-07760/modernizing-regulatory-review, accessed 5 February 2024.

Sarah Fox, Jill Dimond, Lilly Irani, Tad Hirsch, Michael Muller, and Shaowen Bardzell, 2017. "Social justice and design: Power and oppression in collaborative systems," *CSCW '17 Companion: Companion of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*, pp. 117–122. doi: https://doi.org/10.1145/3022198.3022201, accessed 10 March 2024.

Paulo Freire, 1978. "Pedagogy of the oppressed," In: John Beck, Chris Jenks, Nell Keddie, and Michael F.D. Young (editors). *Toward a sociology of education*. New York: Routledge, pp. 374–386. doi: https://doi.org/10.4324/9780429339530, accessed 10 March 2024.

Sorelle Friedler, Ranjit Singh, Borhane Blili-Hamelin, Jacob Metcalf, and Brian J. Chen, 2023. "AI red-teaming is not a one-stop solution to AI harms: Recommendations for using red-teaming for AI accountability," *Data & Society* (25 October), at https://datasociety.net/library/ai-red-teaming-is-not-a-one-stop-solution-to-ai-harms-recommendations-for-using-red-teaming-for-ai-accountability/, accessed 10 March 2024.

Batya Friedman, David G. Hendry, and Alan Borning, 2017. "A survey of value sensitive design methods," *Foundations and Trends in Human–Computer Interaction*, volume 11, number 2, pp. 63–125. doi: https://doi.org/10.1561/1100000015, accessed 10 March 2024.

Iason Gabriel, 2020. "Artificial intelligence, values, and alignment," *Minds and Machines*, volume 30, number 3, pp. 411–437. doi: https://doi.org/10.1007/s11023-020-09539-2, accessed 10 March 2024.

John Gaventa, 1988. "Participatory research in North America," *Convergence*, volume 21, numbers 2–3, pp. 19–48.

Tarleton Gillespie, 2020. "Content moderation, AI, and the question of scale," *Big Data & Society* (21 August). doi: https://doi.org/10.1177/2053951720943234, accessed 10 March 2024.

Tarleton Gillespie. 2018. *Custodians of the Internet: Platforms, content moderation, and the hidden decisions that shape social media*. New Haven, Conn.: Yale University Press. doi: https://doi.org/10.12987/9780300235029, accessed 10 March 2024.

Michele Gilman, 2023. "Democratizing AI: Principles for meaningful public participation," *Data & Society*, at https://datasociety.net/wp-content/uploads/2023/09/DS_Democratizing-AI-Public-Participation-Brief_9.2023.pdf, accessed 9 February 2024.

Michele Estrin Gilman, 2022. "Beyond window dressing: Public participation for marginalized communities in the datafied society," *Fordham Law Review*, volume 91, number 2, article number 7, pp. 503–555, and at https://ir.lawnet.fordham.edu/flr/vol91/iss2/7, accessed 10 March 2024.

Gillian R. Hayes, 2011. "The relationship of action research to human-computer interaction," *ACM Transactions on Computer-Human Interaction*, volume 18, number 3, article number 15, pp. 1–20. doi: https://doi.org/10.1145/1993060.1993065, accessed 10 March 2024.

Google Responsible AI, 2022. "Skin tone research @ Google AI | Developing the Monk Skin Tone Scale," at https://skintone.google/the-scale, accessed 6 February 2023.

Connor Graham and Mark Rouncefield, 2008. "Probes and participation," *PDC '08: Proceedings of the Tenth Anniversary Conference on Participatory Design 2008*, pp. 194–197.

Mary L. Gray and Siddharth Suri, 2019. *Ghost work: How to stop Silicon Valley from building a new global underclass*. Boston, Mass.: Houghton Mifflin Harcourt.

Lara Groves, Aidan Peppin, Andrew Strait, and Jenny Brennan, 2023. "Going public: The role of public participation approaches in commercial AI labs," *FAccT '23: Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency*, pp. 1,162–1,173. doi: https://doi.org/10.1145/3593013.3594071, accessed 17 October 2023.

Alex Hanna and Tina M Park, 2020. "Against scale: Provocations and resistances to scale thinking," *arXiv*:2010.08850 (17 October). doi: https://doi.org/10.48550/arXiv.2010.08850, accessed 10 March 2024.

Donna Haraway, 1988. "Situated knowledges: The science question in feminism and the privilege of partial perspective," *Feminist Studies*, volume 14, number 3, pp. 575–599.. doi: https://doi.org/10.2307/3178066, accessed 10 March 2024.

Christina N. Harrington, 2020. "The forgotten margins: What is community-based participatory health design telling us?" *Interactions*, volume 27, number 3, pp. 24–29. doi: https://doi.org/10.1145/3386381, accessed 10 March 2024.

Christina Harrington, Yolanda Rankin, Jasmine Jones, Robin Brewer, Sheena Erete, Tawanna Dillahunt, and Quincy Brown, 2020. "A call to action for the ACM," *Interactions* (22 June), at https://interactions.acm.org/blog/view/a-call-to-action-for-the-acm, accessed 10 March 2024.

Christina Harrington, Sheena Erete, and Anne Marie Piper, 2019. "Deconstructing community-based collaborative design: Towards more equitable participatory design engagements," *Proceedings of the ACM on Human-Computer Interaction*, volume 3, number CSCW, article number 216, pp. 1–25. doi: https://doi.org/10.1145/3359318, accessed 17 October 2023.

Gillian R. Hayes, 2011. "The relationship of action research to human-computer interaction," *ACM Transactions on Computer-Human Interaction*, volume 18, number 3, article number 15, pp. 1–20. doi: https://doi.org/10.1145/1993060.1993065, accessed 10 March 2024.

Caner Hazirbas, Joanna Bitton, Brian Dolhansky, Jacqueline Pan, Albert Gordo, and Cristian Canton Ferrer, 2022. "Towards measuring fairness in AI: The casual conversations dataset," *IEEE Transactions on Biometrics, Behavior, and Identity Science*, volume 4, number 3, pp. 324–332. doi: https://doi.org/10.1109/TBIOM.2021.3132237, accessed 10 March 2024.

Anna Lauren Hoffmann, 2021. "Terms of inclusion: Data, discourse, violence," *New Media & Society*, volume 23, number 12, pp. 3,539–3,556. doi: https://doi.org/10.1177/1461444820958725, accessed 10 March 2024.

Tung-Hui Hu, 2015. *A prehistory of the cloud*. Cambridge, Mass.: MIT Press. doi: https://doi.org/10.7551/mitpress/9780262029513.001.0001, accessed 10 March 2024.

Lilly Irani, Janet Vertesi, Paul Dourish, Kavita Philip, and Rebecca E. Grinter, 2010. "Postcolonial

computing: A lens on design and development," *CHI '10: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 1,311–1,320.
doi: https://doi.org/10.1145/1753326.1753522, accessed 10 March 2024.

Barbara A. Israel, Amy J. Schulz, Edith A. Parker, and Adam B. Becker, 1998. "Review of community-based research: Assessing partnership approaches to improve public health," *Annual Review of Public Health*, volume 19, pp. 173–202.
doi: https://doi.org/10.1146/annurev.publhealth.19.1.173, accessed 10 March 2024.

Eun Seo Jo and Timnit Gebru, 2020. "Lessons from archives: Strategies for collecting sociocultural data in machine learning," *FAT* '20: Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, pp. 306–316.
doi: https://doi.org/10.1145/3351095.3372829, accessed 10 March 2024.

Pratyusha Kalluri, 2020. "Don't ask if artificial intelligence is good or fair, ask how it shifts power," *Nature*, volume 583, number 7815 (7 July), p. 169.
doi: https://doi.org/10.1038/d41586-020-02003-2, accessed 10 March 2024.

Jamie Kalven, 2023. "Why does the Chicago Police Department tolerate abusive racists in its ranks?" *The Intercept* (9 December), at https://theintercept.com/2023/12/09/chicago-police-department-racism-civilian-complaints/, accessed 18 December 2023.

Michael Katell, Meg Young, Dharma Dailey, Bernease Herman, Vivian Guetler, Aaron Tam, Corinne Bintz, Daniella Raz, and P.M. Krafft, 2020. "Toward situated interventions for algorithmic equity: Lessons from the field," *FAT* '20: Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, pp. 45–55.
doi: https://doi.org/10.1145/3351095.3372874, accessed 10 March 2024.

Kate Klonick, 2020. "The Facebook Oversight Board: Creating an independent institution to adjudicate online free expression," *Yale Law Journal*, volume 129, number 8, 2,418–2,499, and at https://www.yalelawjournal.org/feature/the-facebook-oversight-board, accessed 10 March 2024.

Max Krüger, Ana Bustamante Duarte, Anne Weibert, Konstantin Aal, Reem Talhouk, and Oussama Metatla, 2019. "What is participation? Emerging challenges for participatory design in globalized conditions," *Interactions*, volume 26, number 3, pp. 50–54.
doi: https://doi.org/10.1145/3319376, accessed 10 March 2024.

Christopher A. Le Dantec and Sarah Fox, 2015. "Strangers at the gate: Gaining access, building rapport, and co-constructing community-based research," *CSCW '15: Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*, pp. 1,348–1,358.
doi: https://doi.org/10.1145/2675133.2675147, accessed 10 March 2024.

Jennifer Lee, Meg Young, P.M. Krafft, and Michael A Katell, 2020. "Power and technology: Who gets to make the decisions?" *Interactions*, volume 28, number 1, pp. 38–46.
doi: https://doi.org/10.1145/3442420, accessed 10 March 2024.

Jan Leike, John Schulman and Jeffrey Wu, 2022. "Our approach to alignment research," *OpenAI* (24 August), at https://openai.com/blog/our-approach-to-alignment-research, accessed 5 February 2024.

Min Kyung Lee, Ishan Nigam, Angie Zhang, Joel Afriyie, Zhizhen Qin, and Sicun Gao, 2021. "Participatory algorithmic management: Elicitation methods for worker well-being models," *AIES '21: Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society*, pp. 715–726.
doi: https://doi.org/10.1145/3461702.3462628, accessed 10 March 2024.

Chad Marlow, 2021. "The people, not the police, should decide if and how surveillance technologies are used in their communities," *ACLU* (25 May), at https://www.aclu.org/news/criminal-law-reform/the-people-not-the-police-should-decide-if-and-how-surveillance-technologies-are-used-in-their-communities, accessed 10 March 2024.

Donald Martin, Jr., Vinodkumar Prabhakaran, Jill Kuhlberg, Andrew Smart, and William S. Isaac, 2020. "Participatory problem formulation for fairer machine learning through community based system dynamics," *arXiv*:2005.07572 (15 May).
doi: https://doi.org/10.48550/arXiv.2005.07572, accessed 10 March 2024.

Ashwin Jacob Mathew, 2014. "Where in the world is the Internet? Locating political power in Internet infrastructure," Ph.D. dissertation, University of California, Berkeley, at https://escholarship.org/uc/item/13m8k8ns, accessed 10 March 2024.

Clark McAllister, 2022. *Karl Marx's workers' inquiry: International history, reception, and responses*. *Notes from Below*, at https://notesfrombelow.org/media/other/karl-marxs-workers-inquiry-STANDARD-Jul2022.pdf, accessed 10 March 2024.

Robin McTaggart, 1991. "Principles for participatory action research," *Adult Education Quarterly*, volume 41, number 3, pp. 168–187.
doi: https://doi.org/10.1177/0001848191041003003, accessed 10 March 2024.

Jacob Metcalf, Emanuel Moss, Elizabeth Anne Watkins, Ranjit Singh, and Madeleine Clare Elish, 2021. "Algorithmic impact assessments and accountability: The co-construction of impacts," *FAccT '21: Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, pp. 735–746.
doi: https://doi.org/10.1145/3442188.3445935, accessed 12 March 2021.

Salar Mohandesi and Asad Haider, 2013. "Workers' inquiry: A genealogy," *Viewpoint Magazine* (27 September), at https://viewpointmag.com/2013/09/27/workers-inquiry-a-genealogy/, accessed 20 October 2023.

David Morar and Bruna Santos, 2022. "Is the DSA a new dawn of legislating platform governance globally?" *Lawfare Institute* (30 November), at https://www.lawfaremedia.org/article/dsa-new-dawn-legislating-platform-governance-globally, accessed 10 March 2024.

Emanuel Moss, Elizabeth Anne Watkins, Ranjit Singh, Madeleine Clare Elish, and Jacob Metcalf, 2021. "Assembling accountability: Algorithmic impact assessment for the public interest," *Data & Society Research Institute* (29 June), at https://datasociety.net/library/assembling-accountability-algorithmic-impact-assessment-for-the-public-interest/, accessed 10 March 2024.

Michael J. Muller and Allison Druin, 2012. "Participatory design: The third space in human–computer interaction," In: Julie A. Jacko (editor). *The human–computer interaction handbook: Fundamentals, evolving technologies, and emerging applications*. Third edition. Boca Raton, Fla.: CRC Press, pp. 1,125–1,153.
doi: https://doi.org/10.1201/b11963, accessed 10 March 2024.

National Institute of Standards and Technology, 2023. "AI risk management framework: AI RMF (1.0)," *NIST AI* 100-1.
doi: https://doi.org/10.6028/NIST.AI.100-1, accessed 10 March 2024.

Kristen Nygaard and Olav Terje Bergo, 1975. "The trade unions — New users of research," *Personnel Review*, volume 4, number 2, pp. 5–10.
doi: https://doi.org/10.1108/eb055278, accessed 10 March 2024.

Mimi Onuoha, 2018. "On missing data sets," *International Workshop on Obfuscation*, pp. 38–40, and at http://www.obfuscationworkshop.org/wp-content/uploads/2017/10/obfuscation-workshop-report.pdf#page=38, accessed 5 February 2024.

Aviv Ovadya, 2023. "'Generative CI' through collective response systems," *arXiv*:2302.00672 (1 February). doi: https://doi.org/10.48550/arXiv.2302.00672, accessed 5 February 2024.

Amandalynne Paullada, Inioluwa Deborah Raji, Emily M. Bender, Emily Denton, and Alex Hanna, 2021. "Data and its (dis)contents: A survey of dataset development and use in machine learning research," *Patterns*, volume 2, number 11, 100336.
doi: https://doi.org/10.1016/j.patter.2021.100336, accessed 10 March 2024.

Jean-Christophe Plantin, Carl Lagoze, Paul N. Edwards, and Christian Sandvig, 2018. "Infrastructure studies meet platform studies in the age of Google and Facebook," *New Media & Society*, volume 20, number 1, pp. 293–310.
doi: https://doi.org/10.1177/1461444816661553, accessed 10 March 2024.

rina Reynolds-Tyler, Invisible Institute, and Sarah Conway, 2023. "What happens when your loved one goes missing?" *Chicago Reader* (14 November), at http://chicagoreader.com/news-politics/chicago-police-missing-persons/, accessed 7 February 2024.

David Ribes, 2014. "Ethnography of scaling, or, how to a fit a national research infrastructure in the room," *CSCW '14: Proceedings of the 17th ACM Conference on Computer Supported Cooperative Work & Social Computing*, pp. 158–170.
doi: https://doi.org/10.1145/2531602.2531624, accessed 10 March 2024.

Afsaneh Rigot, 2022. "Design from the margins," *Harvard Kennedy School. Belfer Center for Science and International Affairs* (13 May), at https://www.belfercenter.org/publication/design-margins, accessed 5 February 2024.

David G. Robinson, 2022. *Voices in the code: A story about people, their values, and the algorithm they made*. New York: Russell Sage Foundation.

Kevin Roose, 2023. "What if we could all control A.I.?" *New York Times* (17 October), at https://www.nytimes.com/2023/10/17/technology/ai-chatbot-control.html, accessed 5 February 2024.

Jan Rydzak, 2022. "Dissecting big tech's shareholder showdown," *Tech Policy Press* (16 June), at https://techpolicy.press/dissecting-big-techs-shareholder-showdown/, accessed 10 March 2024.

Elizabeth B.-N. Sanders and Pieter Jan Stappers, 2008. "Co-creation and the new landscapes of design," *CoDesign*, volume 4, number 1, pp. 5–18.
doi: https://doi.org/10.1080/15710880701875068, accessed 10 March 2024.

Morgan Klaus Scheuerman, Kandrea Wade, Caitlin Lustig, and Jed R. Brubaker, 2020. "How we've taught algorithms to see identity: Constructing race and gender in image databases for facial analysis," *Proceedings of the ACM on Human-Computer Interaction*, volume 4, number CSCW1, article number 58, pp. 1–35.
doi: https://doi.org/10.1145/3392866, accessed 10 March 2024.

Nick Seaver, 2021. "Care and scale: Decorrelative ethics in algorithmic recommendation," *Cultural Anthropology*, volume 36, number 3, pp. 509–537.
doi: https://doi.org/10.14506/ca36.3.11, accessed 10 March 2024.

Elizabeth Seger, Aviv Ovadya, Divya Siddarth, Ben Garfinkel, and Allan Dafoe, 2023. "Democratising AI: Multiple meanings, goals, and methods," *AIES '23: Proceedings of the 2023 AAAI/ACM Conference on AI, Ethics, and Society*, pp. 715–722.
doi: https://doi.org/10.1145/3600211.3604693, accessed 9 February 2024.

Andrew D. Selbst, Danah Boyd, Sorelle A. Friedler, Suresh Venkatasubramanian, and Janet Vertesi, 2019.

“Fairness and abstraction in sociotechnical systems,” *FAT* ’19: Proceedings of the Conference on Fairness, Accountability, and Transparency*, pp. 59–68.
doi: https://doi.org/10.1145/3287560.3287598, accessed 7 July 2019.

Tarak Shah and Trina Reynolds-Tyler, 2023. “Keynote: The community built a model: Using participatory AI to analyze Chicago Police data,” at https://www.youtube.com/watch?v=kbPXUq-sVpQ, accessed 7 February 2024.

Mona Sloane, Emanuel Moss, Olaitan Awomolo, and Laura Forlano, 2022. “Participation is not a design fix for machine learning,” *EAAMO ’22: Proceedings of the Second ACM Conference on Equity and Access in Algorithms, Mechanisms, and Optimization*, article number 1, pp. 1–6.
doi: https://doi.org/10.1145/3551624.3555285, accessed 9 February 2024.

Vincent M. Southerland, 2023. “The master’s tool and a mission: Using community control and oversight laws to resist and abolish police surveillance technologies,” *UCLA Law Review*, volume 70, at https://www.uclalawreview.org/the-masters-tools-and-a-mission-using-community-control-and-oversight-laws-to-resist-and-abolish-police-surveillance-technologies/, accessed 10 March 2024.

Susan Leigh Star, 1999. “The ethnography of infrastructure,” *American Behavioral Scientist*, volume 43, number 3, pp. 377–391.
doi: https://doi.org/10.1177/00027649921955326, accessed 9 February 2024.

Susan Leigh Star and Geoffrey C. Bowker, 2010. “How to infrastructure,” In: Leah A. Lievrouw and Sonia Livingstone (editors). *Handbook of new media: Social shaping and social consequences of ICTs*. London: Sage, pp. 230–245.
doi: https://doi.org/10.4135/9781446211304, accessed 17 October 2023.

Luke Stark, 2018. “Algorithmic psychometrics and the scalable subject,” *Social Studies of Science*, volume 48, number 2, pp. 204–231.
doi: https://doi.org/10.1177/0306312718772094, accessed 10 March 2024.

E. Summerson Carr and Michael Lempert (editors), 2016. *Scale: Discourse and dimensions of social life*. Berkeley: University of California Press.
doi: https://doi.org/10.1525/luminos.15, accessed 10 March 2024.

Harini Suresh and John Guttag, 2021. “A framework for understanding sources of harm throughout the machine learning life cycle,” *EAAMO ’21: Proceedings of the First ACM Conference on Equity and Access in Algorithms, Mechanisms, and Optimization*, article number 17, pp. 1–9.
doi: https://doi.org/10.1145/3465416.3483305, accessed 10 March 2024.

Cai Thomas, Lauren Grace Cargo, and Zac Manuel, 2023. “Trina Reynolds Tyler: Bridge builders,” at https://www.youtube.com/watch?v=nzTmVQSpzhY, accessed 7 February 2024.

Anna Lowenhaupt Tsing, 2012. “On nonscalability: The living world is not amenable to precision-nested scales,” *Common Knowledge*, volume 18, number 3, pp. 505–524.
doi: https://doi.org/10.1215/0961754X-1630424, accessed 10 March 2024.

Anna Lowenhaupt Tsing, 2005. *Friction: An ethnography of global connection*. Princeton, N.J.: Princeton University Press.
doi: https://doi.org/10.2307/j.ctt7s1xk, accessed 10 March 2024.

Anna Lowenhaupt Tsing, 2000. “Inside the economy of appearances,” *Public Culture*, volume 12, number 1, pp. 115–144.
doi: https://doi.org/10.1515/9780822383215-010, accessed 10 March 2024.

White House, 2022. “Blueprint for an AI Bill of Rights,” at https://www.whitehouse.gov/ostp/ai-bill-of-rights/, accessed 10 March 2024.

Christopher Wilson, 2022. “Public engagement and AI: A values analysis of national strategies,” *Government Information Quarterly*, volume 39, number 1, 101652.
doi: https://doi.org/10.1016/j.giq.2021.101652, accessed 10 March 2024.

Jamie Woodcock, 2014. “The workers’ inquiry from Trotskyism to Operaismo: A political methodology for investigating the workplace,” *Ephemera*, volume 14, number 3, pp. 493–513, and at https://ephemerajournal.org/contribution/workers%25E2%2580%2599-inquiry-trotskyism-operaismo-political-methodology-investigating-workplace, accessed 10 March 2024.

Ron Wyden, 2022. “U.S. Algorithmic Accountability Act of 2022,” at https://www.congress.gov/bill/117th-congress/senate-bill/3572, accessed 10 March 2024.

Qian Yang, Aaron Steinfeld, Carolyn Rosé, and John Zimmerman, 2020. “Re-examining whether, why, and how human-AI interaction is uniquely difficult to design,” *CHI ’20: Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, pp. 1–13.
doi: https://doi.org/10.1145/3313831.3376301, accessed 10 March 2024.

Meg Young, P.M. Krafft, and Michael A. Katell, 2021. “A call for scholar activism,” at https://people.csail.mit.edu/pkrafft/papers/critplat-acm-interactions-power-tech.pdf, accessed 10 March 2024.

Meg Young, Michael Katell and P.M. Krafft, 2019. “Municipal surveillance regulation and algorithmic accountability,” *Big Data & Society* (25 September).
doi: https://doi.org/10.1177/2053951719868492, accessed 10 March 2024.

Editorial history

Received 21 September 2023; accepted 10 March 2024.

Participation versus scale: Tensions in the practical demands on participatory AI
by Meg Young, Upol Ehsan, Ranjit Singh, Emnet Tafesse, Michele Gilman, Christina Harrington, and Jacob Metcalf.
First Monday, volume 29, number 4 (April 2024).
doi: https://dx.doi.org/10.5210/fm.v29i4.13642