Power to the People? Opportunities and Challenges for Participatory AI

Abeba Birhane
Mozilla Foundation & University College Dublin
Ireland
abeba.birhane@ucdconnect.ie

William Isaac
DeepMind
UK
williamis@deepmind.com

Vinodkumar Prabhakaran
Google
USA
vinodkpg@google.com

Mark Díaz
Google
USA
markdiaz@google.com

Madeleine Clare Elish
Google
USA
mcelish@google.com

Iason Gabriel
DeepMind
UK
iason@deepmind.com

Shakir Mohamed
DeepMind
UK
shakir@deepmind.com

ABSTRACT

Participatory approaches to artificial intelligence (AI) and machine learning (ML) are gaining momentum: the increased attention comes partly with the view that participation opens the gateway to an inclusive, equitable, robust, responsible and trustworthy AI. Among other benefits, participatory approaches are essential to understanding and adequately representing the needs, desires and perspectives of historically marginalized communities. However, there currently exists lack of clarity on what meaningful participation *entails* and what it is *expected* to do. In this paper we first review participatory approaches as situated in historical contexts as well as participatory methods and practices within the AI and ML pipeline. We then introduce three case studies in participatory AI. Participation holds the potential for beneficial, emancipatory and empowering technology design, development and deployment while also being at risk for concerns such as cooptation and conflation with other activities. We lay out these limitations and concerns and argue that as participatory AI/ML becomes in vogue, a contextual and nuanced understanding of the term as well as consideration of who the primary beneficiaries of participatory activities ought to be constitute crucial factors to realizing the benefits and opportunities that participation brings.

CCS CONCEPTS

• **Human-centered computing** → *Interaction design theory, concepts and paradigms.*

EAAMO '22, October 6–9, 2022, Arlington, VA, USA

ACM ISBN 978-1-4503-9477-2/22/10.
https://doi.org/10.1145/3551624.3555290

KEYWORDS

Participatory AI, Machine Learning, Power, Justice

ACM Reference Format:
Abeba Birhane, William Isaac, Vinodkumar Prabhakaran, Mark Díaz, Madeleine Clare Elish, Iason Gabriel, and Shakir Mohamed. 2022. Power to the People? Opportunities and Challenges for Participatory AI. In *Equity and Access in Algorithms, Mechanisms, and Optimization (EAAMO '22), October 6–9, 2022, Arlington, VA, USA*. ACM, New York, NY, USA, 8 pages. https://doi.org/10.1145/3551624.3555290

1 INTRODUCTION

Artificial Intelligence (AI) has taken a 'participatory turn', with the reasoning that *participation* provides a means to incorporate wider publics into the development and deployment of AI systems. The greater attention to participatory methods, participatory design, and the emerging imaginary of participatory AI, follows as a response to the changing attitudes towards AI's role in our societies, in light of the documented harms that have emerged in the areas of security, justice, employment, and healthcare, among others [2, 18, 39, 54, 62, 79, 81, 82]. The field of artificial intelligence is faced with the need to evolve its development practices—characterized currently as technically-focused, representationally imbalanced, and non-participatory—if it is to meet the optimistic vision of AI intended to deeply support human agency and enhance prosperity. Participation has a vital role to play in aligning AI towards prosperity, especially of the most marginalized, but requires a deeper interrogation of its scope and limitations, uses and misuses, and the place of participation within the broader AI development ecosystem.

A growing body of work has shown the different roles and formats that participation can take in the development of AI, including: new approaches to technical development in NLP in healthcare [25, 63], in the development of alternative design toolkits and processes [44, 56], and methods that range from structured interviews to citizens juries [5]. In these cases, participation is meant to move beyond individual opinion to center the values of inclusion, plurality, collective safety, and ownership, subsequently shifting the relationship from one of designer-and-user to one of co-designers and co-creators. Participation is expected to lead to systems of self-determination and community empowerment. Yet, caution has been raised about the possibility of 'participation-washing' [75], where efforts are mischaracterized under the banner of participation, are weakly-executed, or co-opt the voice of participants to achieve predetermined aims.

In this paper, we advance a view that participation should continue to grow and be refined as a key component of the AI development and deployment lifecycle as a means to empower communities, especially those at the margins of society that AI often disproportionately and negatively impacts. To achieve this through Participatory AI, greater clarity is needed on *what* participation is, *who* it is supposed to serve, *how* it can be used in the specific context of AI, and *how* it is related to the mechanisms and approaches already available. Our paper makes three contributions towards developing this deeper understanding of participatory AI. Firstly, we situate the participatory process within its broader genealogical development. We develop this historical hindsight in section 2, considering histories of participatory development as well as its colonial inheritances and its newer digital forms. Secondly, we present three case studies in section 3 taken from selected participatory projects to further concretize various forms of existing participatory work. we reframe participation by introducing a characterisation that allows the multiple forms of participation seen in practice to be compared in Appendix B. We then describe potential limitations and concerns of participation in section 4 and conclude in Section 5.

2 GENEALOGY OF PARTICIPATION

> "*'je participe, tu participes, il participe, nous participons, vous participez … ils profitent'* (in English: *'I participate, you participate, he participates, we participate, you participate … they profit'*)". [1]

This quote appears widely in works related to participation [3, 16, 46]—it poetically captures the cycles of enthusiasm and use of participation and participatory methods as a remedy for many problems of social life, but ending with a sense of disenchantment, exploitation, and with asymmetrical power dynamics in society left unchanged. We see the same enthusiasm for participation in AI at present, which renews this quote's relevance for the analysis of participatory approaches in AI. The quote points to the many historical lessons upon which new participatory undertakings can draw from, which is often absent in AI research; it also serves as a warning of one type of participatory outcome to avoid. To draw on this experience, in this section, we begin with a summary of participation's historical roots, look specifically at participation for technological development, and then review the current landscape of participatory AI.

[1]From a French poster by Atelier Populaire, May 1968. V&A Museum collections accession number E.784-2003

2.1 Historical Participation

Over the last century, participatory approaches have risen globally to the fore across all sectors, including in international development, healthcare provision, decision-making about science, democracy, the environment, and social movements for self-determination, among others [14, 16, 21, 29, 37, 46]. This rise is driven by the multitude of benefits associated with participation. Participatory approaches, by engaging citizens in scientific, democratic or environmental decision-making, for example, enables these processes to become transparent, accountable, and responsive to the needs of those who participate. Participatory methods also establish a distinct opportunity to learn from and include knowledge created by the people directly affected by existing or new undertakings. When such collective knowledge is missing, its absence leads to failure leaving projects to be based solely upon technocratic or elite perspectives [28, 59, 77]. Moreover, at its best, participation leads to individual and collective empowerment as well as social and structural change via the cultivation of new skills, social capital, networks, and self-determination among those who contribute. This has the potential to make a sustained positive impact to the welfare and benefit of communities over time [27, 73].

The desire to unlock these benefits through novel forms of organization played a central role in the development of participatory approaches to research and decision-making, a trend that is most often traced back to the work of Scandinavian researchers in the 1970s and 80s [4, 16]. The 'Scandinavian approach' to participation is concerned primarily with the creation of 'workplace democracy' understood as a system of structured consultation and dialogue between workers and employees with the aim of giving workers greater control over wages and the allocation of tasks. Building upon this idea, participatory approaches have been used to countenance different kinds of response to the challenges posed to workers by technological innovation. As examples, the *Scandanavian Collective Resource Approach* helps workers actively manage processes of technological adoption by promoting knowledge-sharing about new technologies, improving the ability of unions to negotiate collectively with employers, and identify mutually beneficial trajectories [4, 47]. The *British Socio-Technical Systems Approach* to participation was developed to promote the notion of a *systems science* where new technologies, the workers (with emphasis on their psychological and physical well-being), and the environment that they are embedded in, are held to be an interactive part of a larger system that needs to be collectively managed. The latter school of thought set out to promote workers' autonomy through their active participation in the design of socio-technical systems as a whole [4, 26].

The power of participation has led to a proliferation of approaches, including the principle of maximum feasible participation [61], and one of the most regarded uses in the process of participatory budgeting [13]. Today there are numerous tools and processes available for anyone to build on the established practices across a range of participatory methods, whether they include Delphi interviews, citizen's juries, role-playing exercises and scenarios, workshops, force field analyses, or visual imagery (like those in the participatory development toolkit [45]), alongside the professionalization of participation through organisations like the International Association for Public Participation.

Despite these advancements, historical analysis of the roots of participation reveal some of its failings and shortcomings. Long before calls for participation in the workplace, the notion of participation played a central role in the administration of the British empire. The "Dual Mandate in British Tropical Africa" colonial system of the 1920s was rooted in the co-optation of traditional rules and authority structures [51]. By establishing a hierarchical division of power that was enforced using the authority of local rules and chiefs, colonial projects claimed their legitimacy under the veneer of participation. They mandated local people to abide by colonial rules, turning participation in governance into a form of colonial power.

The risk that participation could simply mask uneven power relations, making it easier to perpetuate a dynamic that is fundamentally extractive remains a major concern. One of the most astute critique in this vein was raised by Arnstein [3] that established the now widespread image of the ladder of participation that provided a linear model with which to characterise participation, from extractive to empowering. Although the ladder was powerful, the key critique Arnstein [3] raised was one of power. And this critique has been extended further, labelling the fervour for participation in all sectors as a form tyranny, pushing out other approaches for equity that would be more appropriate while using participatory methods to facilitate an illegitimate and unjust exercise of power. As Arnstein [3] writes: "... participation without redistribution of power is an empty and frustrating process for the powerless. It allows the powerholders to claim that all sides were considered, but makes it possible for only some of those sides to benefit. It maintains the status quo."

2.2 Participation for Technological Innovation

Of relevance to machine learning research, are the specific roles that calls for participation have played in the context of computing and technological innovation. In the U.S., participatory design was widely adopted by large corporations creating office technologies in the 1970s and 80s [4]. The key idea was to increase the quality of products, and to reduce time and costs, by bridging the gap between those doing the design (removed from day to day use) and those that were designed for by involving end-users in the process. Participation in this sense was primarily conceived of as technical work done by participants for the sake of economic efficiency; an aspiration that fit well with larger organizational goals. While this approach yielded some benefits for consumers via an improved product, the value of participation was limited: it did not need to benefit those engaged in co-design and was very unlikely to empower them.

One salient question centers upon whether participation in the technology development pipeline necessarily requires those involved to be actively engaged in the process. For those who focus on participation in the form of activism, movement-building and community initiatives, active engagement is essential [16]. Yet, others define participation more widely [7], so that it encompasses types of incidental participation that arises by simply being part of an environment or process that involves others. This phenomenon is increasingly pronounced in digital media environments, e.g., having one's data harvested and used in AI development by virtue of "participating" in social media, sometimes referred to as mediated participation [46], participation through the medium of technology. Yet this type of "passive" participation has increasingly been linked to privacy violations and surveillance [80, 83]. Furthermore, to support the ideal of a "participatory condition" of society and technological development requires a degree of agency and intentionality [7, 46, 80]. To participate, requires knowing that one is participating.

2.3 The Emergence of Participatory AI

While the participatory mechanisms have served as a constant backdrop for the development of modern technologies, it's emergence within the context of artificial intelligence (AI) and machine learning based applications specifically have been relatively recent. Given its origins as a more speculative academic and industrial technological endeavor, the initial cycles of AI research largely missed the prior waves of participatory research that other technologies of comparable vintage (e.g. personal computing, internet networking, computer software). However, the shift away from logic-based AI systems towards more data-driven paradigms such Deep Learning [48] as well as new infrastructure for capturing and leveraging human-generated data (e.g. Amazon Mturk) prompted greater demand for "non-expert" participation in the construction of AI systems.

One significant adoption of non-expert participation was in the construction of large scale benchmark datasets such as ImageNet [22], where the research team utilized over 49,000 workers from Amazon's mechanical turk (Mturk) platform across 167 countries [23] to perform image recognition tasks to filter and validate the roughly 14 million images in the dataset. Despite this effort being quite broad in its "participation", the highly variable ethical and documentation standards [24, 34] for data enrichment or moderation tasks means that these contributors often fail to be discussed when publishing the final artefacts or protected by conventional research participant standards (e.g. Beneficence, Respect for Persons, Justice). Other area has been in the form of content moderation, where non-expert participants are used to review misinformation or graphic media to prohibit the display of harmful content on internet platforms (e.g YouTube, Facebook, Twitter) but also serve as labelled training data for machine learning classifiers deployed to expedite policy enforcement. The proliferation of machine learning across multiple industries has further ingrained and expanded the general data enrichment and moderation paradigm, but the abuses and concentration of these forms of "Ghost Work" in low income countries and peoples have also been extensively documented in recent years [36, 42, 60, 69].

In parallel to the expansion of data enrichment and moderation, problematic applications of machine learning tools in high stakes domains such as criminal justice [9, 52, 68], healthcare [6, 8], and hiring [1, 66] have prompted both researchers, civil society, and regulators to increasingly urge greater use participatory methods to mitigate sociotechnical risks [72] not addressed by algorithmic adjustments or transformations. The recent upswing in participatory tools have varied in approach and point in the machine learning product lifecycle, including: auditing methods for machine learning based interventions [43, 70], public consultation methods such as citizen juries [5, 78] or joint problem formulation [56], information disclosures such as model cards [41, 58] or datasheets [33, 40], and artefact co-development [38, 49, 74]. A central tension of this this recent wave of "participatory" is whether these mechanisms should merely serve to aid in the refinement of relevant machine learning sytem or rather emphasize lived experience as a critical form of knowledge and employ experiential learning as a force for community empowerment and advance algorithmic equity [30, 44] or ensure wider humanitarian or societal benefits [10, 12]. The heavy influence of industry stakeholders calling for greater participation without resolving these tensions has led to concerns of 'participation-washing' and calls for a greater need to focus on broader social structures and uneven power asymmetries [15, 75], as well as the limits of participation in specific applications, such as healthcare [25].

While the advancement of an emergent subfield of "Participatory AI" has its own critical questions and tensions which are important to further contexualize, the field needs to continue to reflect both its instrumental and broader purposes. The sections below focus exploring three specific areas:

(1) Standards: Despite many activities applying the label of participatory, there are yet no clear consensus on what minimum set of standards or dimensions one should use to assess or evaluate a given potential participatory mechanism. Though not an exhaustive list, attributes such as the degree of **Reciprocity**, **Reflexivity**, and **Empowerment**, as well as the **Duration** of a task are applicable and salient considerations for all participatory mechanisms. Please see the list of questions in Appendix A that further aid the reflexive process for those embarking on participatory activities.
(2) Goals: There is no single unified vision of what Participatory AI tools are intend to achieve, but rather a bundle of overlapping goals and motivations. These include *algorithmic performance improvements, process improvements, and collective exploration.* This is further explored in the Appendix B. While each of these objectives are participatory to some degree, the composition of the stakeholders and relative degree of influence in ultimately shaping the development and impact of a given machine learning system vary significantly. Thus, researchers and developers must ensure that the forms of participatory mechanisms utilized align with the downstream technical, ethical and sociotechnical risks.
(3) Limitations: Invoking both lessons from history and contemporary cases, we will discuss some emerging limitations of utilizing participatory methods as a means of mitigating technical, ethical and sociotechnical risks. These include concerns of participatory mechanisms serving as a viable substitute for legitimate forms of democratic governance and regulation, co-optation of mechanisms in asymmetrical stakeholder settings, and conflation with other social goals such as inclusion or equity. See Section 4 for more.

Below, we present three "sites" or case studies of Participatory AI across the machine learning lifecycle to explore the substantive areas outlined above. The decision to utilize a case study-based approach is intentional, aiming to provide a deeper understanding of the substantive questions in the context of existing or recent cases. Our hope is that this approach will lend a greater appreciation for the nuance and complexity implementing participatory mechanisms in this setting often presents to all the relevant stakeholders. Each case begins with a background description before presenting a contextual analysis. Through an investigation of existing participatory practices, we aim to offer a wider lens on the use of participatory mechanisms in the AI/ML context, where those goals can be attained through participatory approaches, and a clear understanding of potential limitations such that future efforts can hopefully fully realize the impact of meaningfully incorporating participation in AI/ML development and use.

3 THREE CASE STUDIES IN PARTICIPATORY AI

Participatory activities, process, and projects exist in a variety of forms. Within the *AI for social good* field, for example, participatory activities are evoked as a means to improve AI systems that impact communities where, ideally, impacted groups take part as stakeholders through participatory design and implementation [12]. Participation has also been instrumental in designing and building algorithms that serve communities for purposes such as operating on-demand food donation transportation service [49] as well as for building tools, for instance, (Turkopticon) that allow low wage workers — Amazon Mechanical Turkers, in this case — to evaluate their relationships with employers and support one another [42]. Similarly, algorithmic tools that optimize worker well-being through participatory design and algorithmic tool building has been put forward by Lee et al. [50]. In collecting, documenting, and archiving, sociocultural data needed to train and validate machine learning systems, participatory approaches can be critical for representing the "non-elites, the grassroots, the marginalized" in datasets, as outlined by Jo and Gebru [43]. For justice oriented design of technology, participatory approaches provide the means for marginalized communities to organize grassroots movements, challenge structural inequalities and uneven power dynamics allowing communities to build the kind of technology that benefits such communities [18]. Abstaining from participation can also be a form of participation in another related practice, as shown in Waycott et al. [81], who analysed older adults who refused to participate in technological intervention evaluation. Participatory projects, process, and objectives, therefore, are diverse and multifaceted. Below, we present three case studies that illustrate three various models of participation.

3.1 Case 1: Machine translation for African languages

Description. Over 400 participants from more than 20 countries have been self-organizing through an online community. Some of the projects that have emerged from this community focus on the inclusion of traditionally omitted "low-resourced" African languages within the broader machine translation (MT) research agenda [63]. The projects sought to ensure that MT work for low-resourced languages is driven by communities who speak those languages.

There were no prerequisites placed on participation or fixed roles assigned to participants but rather the roles emerged organically and participatory input was sought at every level of the process from language speaking to model building and participants moved fluidly between different roles. The research process was defined collaboratively and iteratively. Meeting agendas were public and democratically voted on. Language data was crowd-sourced, annotated, evaluated, analyzed and interpreted by participants (from participants). The specific project on MT for African languages produced 46 benchmarks for machine translation from English to 39 African languages and from 3 different languages into English [63].

Analysis. The MT project by the Masakhane NLP community illustrates a grassroots (or bottom up) attempt at using participatory mechanism to build new systems improve the underlying performance of existing NLP systems through the inclusion of traditionally under-resourced African languages. The Masakhane MT project sought to increase the degree of empowerment for the stakeholders involved in the project. In this context, **Empowerment** reflects the degree of impact the participants have in shaping the participatory relationship, and ultimately the project or product. An empowering process is one that is often **Reciprocal**: it is bi-directional, emergent, and dynamic, and one where core decisions or artefacts are informed by active participation rather than one based on the idea of placating a community or notions of saviourism. For example, in this case, the idea is to not only crowd-source activities such as crowd-sourcing of language data, participant evaluation of model output, and production of benchmarks but also to create and foster avenues and forums to veto or voice objections.

Having said that, although the project is built on the idea that MT for low-resourced languages should be done by the language speakers themselves, for language speakers, based on community values and interests, it is still possible to see how the research, datasets and tools may be co-opted and monopolized by commercial actors to improve products or models without supporting the broader grassroots effort or the community's interests or needs. As a result the primary beneficiaries of participatory data sourcing may not be speakers of 'low-resourced' languages but actors with access to such sufficient data and compute resources, thus gaining financial benefits, control and legitimacy off of such participatory efforts.

3.2 Case 2: Fighting for Māori data rights

Description. Through participatory initiatives that took place over 10 days in 2018 as part of the *Te Hiku* NLP project, the Māori community in New Zealand both recorded and annotated 300 hours of audio data of the *Te Reo Māori* language [35]. This is enough data to build tools such as spell-checkers, grammar assistants, speech recognition, and speech-to-text technology. However, although the data originated from the Māori speakers across New Zealand and was annotated and cleaned by the Māori community itself, Western based data sharing/open data initiatives meant that the Māori community had to explicitly prevent corporate entities from getting hold of the dataset. The community thus established the Māori Data Sovereignty Protocols [67] in order to take control of their data and technology. Sharing their data, the Māori argued, is to invite commercial actors to shape the future of their language through tools developed by those without connection to the language. By not sharing their data, the Māori argue they are able to maintaining their autonomy and right to self-determination. They insist that, if any technology is to be built using such community sourced data, it must directly and primarily benefit the Māori people. Accordingly, such technology needs to be built by the Māori community itself since they hold the expert knowledge and experience of the language.

Analysis. The Māori case study is an illuminating example that brings together participatory mechanisms as means for methodological innovation while offering reciprocity to the relevant stakeholders. It is a process that prioritizes the *net* benefit of participants, especially those disproportionately impacted by oppressive social structures, who often carry the burdens of negative impacts of technology [11, 20, 57, 60] and reflecting a fair or proportionate return for the value of the participatory engagement. This is of particularly importance when seeking to utilize participatory mechanisms to achieve methodological innovation, or where the process yields unique insights that can inform new or innovative technological artefacts (as opposed to a means to achieve a particular pre-determined technical objective).

Because the data originates from the language speakers themselves and is annotated and cleaned by the Māori community, existing laws around data sovereignty [67] often require that those communities are key decision makers for any downstream applications. In this case, the Māori are committed to the view that any project created using Māori data must directly benefit and needs to be carried out by the Māori themselves. This high degree of reciprocity between stakeholders lead to a case where the needs, goals, benefits and interests of the community is central to participatory mechanism itself. This case study goes further than others by providing avenues for foregrounding reciprocity and refusal (when they are not aligned with the participants values, interests and benefits).

3.3 Case 3: Participatory dataset documentation

Description. A team of researchers put forward participatory methods and processes for dataset documentation — *The Data Card Playbook* — which they view as the route to creating responsible AI systems [2]. According to the team, the Playbook can play a central role in improving dataset quality, validity and reproducibility, all critical aspects of better performing, more accurate, and transparent AI systems. The Playbook comprises of three components – Essentials, Module one, and Module two – all activities supplemented by ready to download templates and instructions. These participatory activities cover guidance ranging from tracking progress, identifying stakeholders, characterizing target audiences for the Data Card, to evaluate and fine-tune documentation, all presented and organized in a detailed and systematic way. The Playbook aims to make datasets accessible to a wider set of stakeholders. The Playbook is presented as a people-centered approach to dataset documentation, subsequently, with the aim of informing and making AI products and research more transparent.

[2]https://pair-code.github.io/datacardsplaybook/

Analysis. This case study encapsulates the kind of participatory activities that support participation as a form of algorithmic performance and/or dataset quality improvement. An indirect benefit of this approach is that mechanisms designed to explore the space of a given artefact or process inevitably offer a potential for **Reflexivity**, critical evaluation and meaningful feedback. Reflexivity as part of a participatory process is a critical element for improving trust between stakeholders and conveying a sense legitimacy of the ultimate artefact.

However, because the central drive for these specific participatory practices are motivated by objectives such as dataset quality improvement, the participants are assigned pre-defined roles and very clear tasks. Dataset quality and transparent dataset documentation indeed impact the performance and accuracy of AI systems, which can all play a role in the larger goal of fair, inclusive, and just AI systems. Nonetheless, this form of participation focuses on fine-grained activities that come with pre-defined goals and objectives means that there is little room (if at all any) for co-exploring, co-creating, and/or negotiating the larger objectives, reflections, and implication of AI systems. There is no guarantee that an AI system that is created using improved and better datasets with the help of the Data Card Playbook cannot be used in applications that harm, disenfranchise, and unjustly target the most marginalised in society. Computer vision systems that are used to target refugees or marginalized communities in any society, for example, result in a net harm to the targeted regardless of their improved performance. Participation for algorithmic performance improvement is not necessarily equipped to deal with such concerns.

4 LIMITATIONS AND CONCERNS

Like any method, participation has limitations, and we briefly explore these here, and also refer to the large body of work in these topics [7, 16, 17, 46, 53, 55]. Effective participation should serve specific purposes and should not be conflated with other tasks and activities, such as consultation, inclusion, or labour. Moreover, participation cannot be expected to provide a solution for all concerns, and is not a solution for all problems. When used in considered ways, participation can be an important tool in the responsible development of AI. We consider here the role of participation in relation to democracy, its conflation with other activities, concerns on cooptation of participatory activities, the challenges of measuring the effectiveness of participatory efforts, and the challenges of balancing expertise and incentives.

Democratic governance. In democratic societies, it is useful to think of democracy as an apparatus that responds to the right of citizens to determine the shape of practices that govern their lives. Consequently, participation is not the best mechanism for decisions/values/norms that are better decided and codified by democratic institutions, governance and laws. In a democratic system, participants are rights-holders: the people to whom decision-makers are accountable, and the body in which authority is ultimately vested. This distinction is important when an undertaking involves matters of significant public concern or modalities of operation, such as the coercive enforcement of law, that require stronger forms of validation and legitimacy [32]. Participatory activities convened by private actors or parallel institutions, cannot stand in for democratic politics, and participatory AI should not aspire to do so or be perceived to meet this function.

Conflation with other activities. By acknowledging participation's limitations, we can refine what it does and does not entail. As one example, inclusion is often conflated with participation [65]. Being included might have practical consequences on the ability of people and groups to engage in participatory processes. But inclusion is not necessarily participation, since any individual can be included in a group, yet not participate, e.g., by never voting, writing, or acting. Inclusion is then related, but in some ways different from participation, and needs its own attention, which also depends on an understanding of any systemic and social barriers in place (e.g., racism, patriarchy, wealth exclusion). Attempts to include can also be exclusionary. When we invite people to participate it is never everyone: some people are always excluded. Typically those excluded are the very worst-off, those with low literacy, those who do not have the time to seek out participatory opportunities, are not members of the right networks, etc [76]. At other times exclusion is needed for safe, open participatory action. And the purposeful abstention, collective refusal, dissent, or silent protest are themselves forms of participation (e.g., as illustrated by the Maori data rights case study).

Cooptation. Concerns remain of participation becoming a mechanism of cooptation. Specific concerns are raised through current economic and capitalist models that seek to dissolve social relations and behaviours into commodities that are then open to monetization [19, 83]. The history of colonial tactics showed how traditional participatory structures were co-opted to claim legitimacy. The case study on machine translation for African languages raises related concerns, where the efforts of grassroots participatory actions, and their data-sharing initiatives, leaves opens the door for cooptation, where corporate actors can use the results of participatory efforts towards corporate benefits. The potential for corporate actors to capitalize on such efforts and build products that maximize profits, with little benefit to communities remains open. In such circumstances, not only are those that participated disempowered, but corporations then emerge as the legitimate arbiters of African languages and subsequently language technology.

Effectiveness and Measurement. One core concern with participatory methods is that it is difficult to measure and provide attribution to the positive benefits of participation. The type of participation described in Appendix B are likely to result in benefits (though real) that are gradual and intangible, e.g., a sense of empowerment, knowledge transfer, creation of social capital, and social reform. In particular, participation that enables in-depth understanding and iterative co-learning can defy quantification and measurement. Investing in such types of participation may appear wasteful when outcomes are measured using blunt instruments such as cost-benefit analysis, and it could instead be the limitations of the metrics we use to evaluate participatory approaches that are an obstacle to the effective use of participatory approaches. The problem of measurement of impact and a general monitoring, evaluation and learning (MEL) framework is generally difficult, so also points to areas for further research to effectively make participatory methods part of regular practice [31].

Expertise and Incentives. One aim of participatory methods is to spread knowledge about technical systems and their impacts.

This involves the knowledge of technical experts, but also the local knowledge embedded in the lived-experience of communities. There is an epistemic burden on all stakeholders in the participatory process to gather enough information to reason, questions, act or decide [64, 71]. The need then to always learn and gather information requires participatory approaches that occur at different time frames, over various duration, and with different groups. Participation necessitates an assessment of the incentives involved, which can become distorted by epistemic burden, fundamentally affecting the participatory process. Put simply, participatory methods cannot rely on simplified assumptions about the reasons people have for engaging in a participatory process. This returns to the need to challenge uneven power distributions and oppressive social structures, as well as the ways that 'community' itself can hide power dynamics.

5 CONCLUSION

To characterise AI as participatory is to acknowledge that the communities and publics beyond technical designers have knowledge, expertise and interests that are essential to the development of AI that aims to strengthen justice and prosperity. Participation is then an essential component in achieving these aims, yet hype and misunderstanding of the participation's role risks reducing its effectiveness and the possibility of greater harm and exploitation of participants. This paper contributes towards clarifying the understanding and role of participation in AI, situating participation within its historical development, as central to contending with systems of power, as seeking forms of vibrant participation, and a as set of methods that has limitations and specific uses.

Participation in AI is a broad set of practices, but whether we use participation for algorithmic improvement, methodological innovation, or collective exploration, they can be characterised along axes of empowerment and reflexive assessment, along which we seek to move away from transactional engagements towards forms of vibrant participation that are in constant engagement with their publics and increase community knowledge and empowerment. As the case studies show, there are desirable forms of participation that are already available that we can draw inspiration from. Participation takes many different from across the AI pipeline, but for researchers, a key aim is to build the reflexive process that the probe questions hoped to initiate. New AI research and products will increasingly rely on participation for its claims to safety, legitimacy and responsibility and we hope this paper provides part of the clarity needed to effectively incorporate participatory methods; and where we find participation lacking, we hope to provide a basis for critique, which is itself a powerful form of participation.

ACKNOWLEDGMENTS

We would like to thank Nenad Tomašev, Sean Legassick, and Courtney Biles for the invaluable comments on an earlier version of the paper. We would also like thank the EAAMO reviewers for their feedback.

REFERENCES

[1] Ifeoma Ajunwa. 2019. An Auditing Imperative for Automated Hiring. (2019).
[2] Miguel Arana-Catania, Felix-Anselm Van Lier, Rob Procter, Nataliya Tkachenko, Yulan He, Arkaitz Zubiaga, and Maria Liakata. 2021. Citizen participation and machine learning for a better democracy. *arXiv preprint arXiv:2103.00508* (2021).
[3] Sherry R Arnstein. 1969. A ladder of citizen participation. *Journal of the American Institute of planners* 35, 4 (1969), 216–224.
[4] Peter M Asaro. 2000. Transforming society by transforming technology: the science and politics of participatory design. *Accounting, Management and Information Technologies* 10, 4 (2000), 257–290.
[5] Brhmie Balaram, Tony Greenham, and Jasmine Leonard. 2018. Artificial Intelligence: real public engagement. *London: RSA. https://www. thersa. org/discover/publications-and-articles/reports/artificial-intelligence-realpublic-engagement* (2018).
[6] Imon Banerjee, Ananth Reddy Bhimireddy, John L Burns, Leo Anthony Celi, Li-Ching Chen, Ramon Correa, Natalie Dullerud, Marzyeh Ghassemi, Shih-Cheng Huang, Po-Chih Kuo, et al. 2021. Reading Race: AI Recognises Patient's Racial Identity In Medical Images. *arXiv preprint arXiv:2107.10356* (2021).
[7] Darin Barney, Gabriella Coleman, Christine Ross, Jonathan Sterne, and Tamar Tembeck. 2016. *The participatory condition in the digital age.* Vol. 51. U of Minnesota Press.
[8] Ruha Benjamin. 2019. Assessing risk, automating racism. *Science* 366, 6464 (2019), 421–422.
[9] Lyria Bennett Moses and Janet Chan. 2018. Algorithmic prediction in policing: assumptions, evaluation, and accountability. *Policing and society* 28, 7 (2018), 806–822.
[10] Aleks Berditchevskaia, Eirini Malliaraki, and Kathy Peach. 2020. Participatory AI for humanitarian innovation. (2020).
[11] Abeba Birhane. 2021. Algorithmic injustice: a relational ethics approach. *Patterns* 2, 2 (2021), 100205.
[12] Elizabeth Bondi, Lily Xu, Diana Acosta-Navas, and Jackson A Killian. 2021. Envisioning Communities: A Participatory Approach Towards AI for Social Good. *arXiv preprint arXiv:2105.01774* (2021).
[13] Yves Cabannes. 2004. Participatory budgeting: a significant contribution to participatory democracy. *Environment and urbanization* 16, 1 (2004), 27–46.
[14] Robert Chambers. 1994. The origins and practice of participatory rural appraisal. *World development* 22, 7 (1994), 953–969.
[15] Alan Chan, Chinasa T Okolo, Zachary Terner, and Angelina Wang. 2021. The Limits of Global Inclusion in AI Development. *arXiv preprint arXiv:2102.01265* (2021).
[16] Jason Chilvers and Matthew Kearnes. 2015. *Remaking participation: Science, environment and emergent publics.* Routledge.
[17] Bill Cooke and Uma Kothari. 2001. *Participation: The new tyranny?* Zed books.
[18] Sasha Costanza-Chock. 2020. *Design justice: Community-led practices to build the worlds we need.*
[19] Nick Couldry and Ulises A Mejias. 2019. *The costs of connection.* Stanford University Press.
[20] Nick Couldry and Ulises Ali Mejias. 2021. The decolonial turn in data and technology research: what is at stake and where is it heading? *Information, Communication & Society* (2021), 1–17.
[21] Andy Dearden and Haider Rizvi. 2008. Participatory design and participatory development: a comparative review. (2008).
[22] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. 2009. Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition.* Ieee, 248–255.
[23] Emily Denton, Alex Hanna, Razvan Amironesei, Andrew Smart, and Hilary Nicole. 2021. On the genealogy of machine learning datasets: A critical history of ImageNet. *Big Data & Society* 8, 2 (2021), 20539517211035955.
[24] 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 preprint arXiv:2007.07399* (2020).
[25] Joseph Donia and James A Shaw. 2021. Co-design and ethical artificial intelligence for health: An agenda for critical research and practice. *Big Data & Society* 8, 2 (2021), 20539517211065248.
[26] Pelle Ehn and Morten Kyng. 1987. The collective resource approach to systems design. *Computers and democracy* (1987), 17–57.
[27] Paul Farmer, Fernet Léandre, Joia Mukherjee, Rajesh Gupta, Laura Tarter, and Jim Yong Kim. 2001. Community-based treatment of advanced HIV disease: introducing DOT-HAART (directly observed therapy with highly active antiretroviral therapy). *Bulletin of the World Health Organization* 79, 12 (2001), 1145.
[28] James Ferguson. 1994. *The anti-politics machine:" development," depoliticization, and bureaucratic power in Lesotho.* U of Minnesota Press.
[29] Paolo Freire. 1996. Pedagogy of the oppressed (revised). *New York: Continuum* (1996).
[30] Seth Frey, PM Krafft, and Brian C Keegan. 2019. " This Place Does What It Was Built For" Designing Digital Institutions for Participatory Change. *Proceedings of the ACM on Human-Computer Interaction* 3, CSCW (2019), 1–31.
[31] Iason Gabriel. 2017. Effective altruism and its critics. *Journal of Applied Philosophy* 34, 4 (2017), 457–473.

[32] Iason Gabriel. 2022. Toward a Theory of Justice for Artificial Intelligence. *Daedalus* 151, 2 (2022), 218–231.

[33] Timnit Gebru, Jamie Morgenstern, Briana Vecchione, Jennifer Wortman Vaughan, Hanna Wallach, Hal Daumé Iii, and Kate Crawford. 2021. Datasheets for datasets. *Commun. ACM* 64, 12 (2021), 86–92.

[34] R Stuart Geiger, Kevin Yu, Yanlai Yang, Mindy Dai, Jie Qiu, Rebekah Tang, and Jenny Huang. 2020. Garbage in, garbage out? Do machine learning application papers in social computing report where human-labeled training data comes from?. In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*. 325–336.

[35] Tim Graham. 2021. Māori are trying to save their language from Big Tech. https://www.wired.co.uk/article/maori-language-tech

[36] Mary L Gray and Siddharth Suri. 2019. *Ghost work: How to stop Silicon Valley from building a new global underclass.* Eamon Dolan Books.

[37] Christopher Groves. 2017. Remaking participation: Science, environment and emergent publics.

[38] Aaron Halfaker and R Stuart Geiger. 2020. Ores: Lowering barriers with participatory machine learning in wikipedia. *Proceedings of the ACM on Human-Computer Interaction* 4, CSCW2 (2020), 1–37.

[39] Christina N Harrington. 2020. The forgotten margins: what is community-based participatory health design telling us? *Interactions* 27, 3 (2020), 24–29.

[40] Sarah Holland, Ahmed Hosny, Sarah Newman, Joshua Joseph, and Kasia Chmielinski. 2018. The dataset nutrition label: A framework to drive higher data quality standards. *arXiv preprint arXiv:1805.03677* (2018).

[41] Ben Hutchinson, Andrew Smart, Alex Hanna, Emily Denton, Christina Greer, Oddur Kjartansson, Parker Barnes, and Margaret Mitchell. 2021. Towards accountability for machine learning datasets: Practices from software engineering and infrastructure. In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*. 560–575.

[42] Lilly C Irani and M Six Silberman. 2013. Turkopticon: Interrupting worker invisibility in amazon mechanical turk. In *Proceedings of the SIGCHI conference on human factors in computing systems*. 611–620.

[43] Eun Seo Jo and Timnit Gebru. 2020. Lessons from archives: Strategies for collecting sociocultural data in machine learning. In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*. 306–316.

[44] 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. In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency* (Barcelona, Spain) *(FAT* '20)*. Association for Computing Machinery, New York, NY, USA, 45–55.

[45] Christopher M Kelty. 2017. The Participatory Development Toolkit. In *Little Development Devices / Humanitarian Goods*, Vol. 9. Limn.

[46] Christopher M Kelty. 2020. *The participant: A century of participation in four stories.* University of Chicago Press.

[47] Philip Kraft and Jørgen P Bansler. 1994. The collective resource approach: the Scandinavian experience. *Scandinavian Journal of Information Systems* 6, 1 (1994), 4.

[48] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. 2015. Deep learning. *nature* 521, 7553 (2015), 436–444.

[49] Min Kyung Lee, Daniel Kusbit, Anson Kahng, Ji Tae Kim, Xinran Yuan, Allissa Chan, Daniel See, Ritesh Noothigattu, Siheon Lee, Alexandros Psomas, et al. 2019. WeBuildAI: Participatory framework for algorithmic governance. *Proceedings of the ACM on Human-Computer Interaction* 3, CSCW (2019), 1–35.

[50] Min Kyung Lee, Ishan Nigam, Angie Zhang, Joel Afriyie, Zhizhen Qin, and Sicun Gao. [n.d.]. Participatory Algorithmic Management for Worker Well-Being. ([n. d.]).

[51] Lord Frederick JD Lugard. 1922. *The dual mandate in British tropical Africa.* Routledge.

[52] Kristian Lum and William Isaac. 2016. To predict and serve? *Significance* 13, 5 (2016), 14–19.

[53] John Mackenzie, Poh-Ling Tan, Suzanne Hoverman, and Claudia Baldwin. 2012. The value and limitations of participatory action research methodology. *Journal of hydrology* 474 (2012), 11–21.

[54] Michael Majale. 2008. Employment creation through participatory urban planning and slum upgrading: The case of Kitale, Kenya. *Habitat International* 32, 2 (2008), 270–282.

[55] Ghazala Mansuri and Vijayendra Rao. 2012. Localizing development: Does participation work? (2012).

[56] 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 preprint arXiv:2005.07572* (2020).

[57] Sabelo Mhlambi. 2020. From rationality to relationality: ubuntu as an ethical and human rights framework for artificial intelligence governance. *Carr Center for Human Rights Policy Discussion Paper Series* 9 (2020).

[58] Margaret Mitchell, Simone Wu, Andrew Zaldivar, Parker Barnes, Lucy Vasserman, Ben Hutchinson, Elena Spitzer, Inioluwa Deborah Raji, and Timnit Gebru. 2019. Model cards for model reporting. In *Proceedings of the conference on fairness, accountability, and transparency*. 220–229.

[59] Timothy Mitchell. 2002. *Rule of experts.* University of California Press.

[60] Shakir Mohamed, Marie-Therese Png, and William Isaac. 2020. Decolonial AI: Decolonial theory as sociotechnical foresight in artificial intelligence. *Philosophy & Technology* 33, 4 (2020), 659–684.

[61] Daniel P Moynihan. 1969. Maximum Feasible Misunderstanding; Community Action in the War on Poverty. (1969).

[62] Michael J Muller and Allison Druin. 2012. Participatory design: The third space in human–computer interaction. In *The Human–Computer Interaction Handbook*. CRC Press, 1125–1153.

[63] Wilhelmina Nekoto, Vukosi Marivate, Tshinondiwa Matsila, Timi Fasubaa, Tajudeen Kolawole, Taiwo Fagbohungbe, Solomon Oluwole Akinola, Shamsuddeen Hassan Muhammad, Salomon Kabongo, Salomey Osei, et al. 2020. Participatory research for low-resourced machine translation: A case study in african languages. *arXiv preprint arXiv:2010.02353* (2020).

[64] Jennifer Pierre, Roderic Crooks, Morgan Currie, Britt Paris, and Irene Pasquetto. 2021. Getting Ourselves Together: Data-centered participatory design research & epistemic burden. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–11.

[65] Kathryn S Quick and Martha S Feldman. 2011. Distinguishing participation and inclusion. *Journal of planning education and research* 31, 3 (2011), 272–290.

[66] Manish Raghavan, Solon Barocas, Jon Kleinberg, and Karen Levy. 2020. Mitigating bias in algorithmic hiring: Evaluating claims and practices. In *Proceedings of the 2020 conference on fairness, accountability, and transparency*. 469–481.

[67] Stephanie Carroll Rainie, Tahu Kukutai, Maggie Walter, Oscar Luis Figueroa-Rodríguez, Jennifer Walker, and Per Axelsson. 2019. Indigenous data sovereignty. (2019).

[68] Rashida Richardson, Jason M Schultz, and Kate Crawford. 2019. Dirty data, bad predictions: How civil rights violations impact police data, predictive policing systems, and justice. *NYUL Rev. Online* 94 (2019), 15.

[69] Sarah T Roberts. 2019. Behind the screen. In *Behind the Screen*. Yale University Press.

[70] Nithya Sambasivan, Shivani Kapania, Hannah Highfill, Diana Akrong, Praveen Paritosh, and Lora M Aroyo. 2021. "Everyone wants to do the model work, not the data work": Data Cascades in High-Stakes AI. In *proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–15.

[71] Scott Scheall and Parker Crutchfield. 2021. The priority of the epistemic. *Episteme* 18, 4 (2021), 726–737.

[72] Andrew D Selbst, Danah Boyd, Sorelle A Friedler, Suresh Venkatasubramanian, and Janet Vertesi. 2019. Fairness and abstraction in sociotechnical systems. In *Proceedings of the conference on fairness, accountability, and transparency*. 59–68.

[73] Amartya Kumar Sen. 2009. *The idea of justice.* Harvard University Press.

[74] Mark P Sendak, William Ratliff, Dina Sarro, Elizabeth Alderton, Joseph Futoma, Michael Gao, Marshall Nichols, Mike Revoir, Faraz Yashar, Corinne Miller, et al. 2020. Real-world integration of a sepsis deep learning technology into routine clinical care: implementation study. *JMIR medical informatics* 8, 7 (2020), e15182.

[75] Mona Sloane, Emanuel Moss, Olaitan Awomolo, and Laura Forlano. 2020. Participation is not a design fix for machine learning. *arXiv preprint arXiv:2007.02423* (2020).

[76] Gayatri Chakravorty Spivak. 2003. Can the subaltern speak? *Die Philosophin* 14, 27 (2003), 42–58.

[77] Kentaro Toyama. 2015. *Geek heresy: Rescuing social change from the cult of technology.* PublicAffairs.

[78] Sabine N van der Veer, Lisa Riste, Sudeh Cheraghi-Sohi, Denham L Phipps, Mary P Tully, Kyle Bozentko, Sarah Atwood, Alex Hubbard, Carl Wiper, Malcolm Oswald, et al. 2021. Trading off accuracy and explainability in AI decision-making: findings from 2 citizens' juries. *Journal of the American Medical Informatics Association* 28, 10 (2021), 2128–2138.

[79] Maja Van der Velden, Christina Mörtberg, et al. 2015. Participatory design and design for values. *Handbook of Ethics, Values, and Technological Design: Sources, Theory, Values and Application Domains* (2015), 41–66.

[80] Carissa Véliz. 2020. *Privacy is power.* Random House Australia.

[81] Jenny Waycott, Frank Vetere, Sonja Pedell, Amee Morgans, Elizabeth Ozanne, and Lars Kulik. 2016. Not for me: Older adults choosing not to participate in a social isolation intervention. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. 745–757.

[82] Susanne Weber, Marian Harbach, and Matthew Smith. 2015. Participatory design for security-related user interfaces. *Proc. USEC* 15 (2015).

[83] Shoshana Zuboff. 2019. *The age of surveillance capitalism: The fight for a human future at the new frontier of power.* Profile books.