Beyond Participatory AI

Jonne Maas, Aarón Moreno Inglés
Delft University of Technology
j.j.c.maas@tudelft.nl, a.morenoingles@tudelft.nl

Abstract

The 'participatory turn' in AI design has received much attention in the literature. In this paper, we provide various arguments and proposals to move the discussion of participatory AI beyond its current state and towards stakeholder empowerment. The participatory AI literature points to Arnstein's understanding of 'citizen power' as the right approach to participation. Although we agree with this general idea, we argue that there is a lack of depth in analysing the legal, economic, and political arrangements required for a genuine redistribution of power to prioritize AI stakeholders. We highlight two domains on which the current discourse on participatory AI needs to expand. These are (1) the legal-institutional background that could provide 'participation teeth' for stakeholder empowerment and (2) the political economy of AI production that fosters such power asymmetries between AI developers and other stakeholders. We conclude by offering ways forward to explore alternative legal arrangements and ownership models for participatory AI.

1 Introduction

In the past few years, the literature on AI Ethics has seen an increased interest in democratizing AI, often through a participatory approach to AI design. Different forms of participatory mechanisms have emerged to align AI systems with stakeholder values more effectively and to empower stakeholders by allowing them to resist and contest the diverse socio-technical risks and injustices posed by AI-powered technologies. These risks have already been conceptualized and discussed, including questions of racism and bias (Akter 2021; Gebru 2020; Noble 2018), privacy breaches (Bartneck et al. 2021), and mass surveillance trends (Feldstein 2019), amongst other issues. Multiple scholars have engaged with the 'participatory turn' in AI to address these risks (Delgado et al. 2023; Birhane et al. 2022; Donia and Shaw 2021; Costanza-Chock 2020; Denton et al. 2020; Cammaerts and Mansell 2020; Rahwan 2018). These scholars build on participatory design schools such as the Scandinavian tradition and Value-Sensitive Design that have placed democratic participation in the design process at the centre of the discussion as the key to addressing technological risks and stakeholder empowerment.

Participatory AI perspectives have also faced certain criticisms. Sloane et al. (2022) denounce participation-washing and "exploitative and extractive forms of community involvement" in participatory design practice for machine learning systems. Birhane et al. (2022, 3) warn of the risk of reproducing colonial dynamics through the legitimation of unfair power structures "under the veneer of participation", in the same way how former colonizers "co-opt[ed] rules and authority structures, [...] turning participation in governance as a form of colonial power." Himmelreich (2022) objects to participatory AI design on the assumption that calls for democratizing AI are based on weak normative grounds and therefore do not justify the costs associated with participation. The general line of reasoning these critics object to is that (1) democracy (i.e., citizen empowerment) involves citizen participation, and (2) therefore democratic AI (i.e., stakeholder empowerment) requires *any form of* stakeholder participation in AI.

Our aim in this paper is three-fold. First, we argue that although the general participatory AI literature points to stakeholder empowerment as the *right* way of participation, the analysis of the legal, economic, and political arrangements proves too superficial to establish a redistribution of power that genuinely prioritizes stakeholders. Second, we want to justify calls for participatory AI despite the costs. We do so by objecting to Himmelreich's argument that participatory AI rests on normative weak grounds. Third, we highlight two domains the current discourse on participatory AI has insufficiently explored. These are (1) the legal-institutional background that could provide what we call 'participation teeth' for stakeholder empowerment and (2) the political economy of AI production that fosters such power asymmetries between AI developers and other stakeholders.

Based on our analysis, we agree with the critics that participatory AI is, as the literature currently stands, flawed. As numerous scholars in political theory have acknowledged, citizen participation is not a sufficient proxy for citizen

empowerment (Young 1990; Cohen 2009). Democracy requires citizen freedom and autonomy, which requires effective control over democratic processes (Pettit 2012). Moreover, such freedom and autonomy are entangled with the political economy (Gould 1990; Sen 2001). We conclude, however, that we should not be 'against' participatory AI, as Himmelreich claims, but we should move *beyond* the current participatory AI discourse. To address the legal-institutional background, novel regulatory processes and external auditing could help to mitigate unjust power asymmetries. To address the political economy of AI production, we argue in favour of exploring alternative ownership models of the AI design and production processes.

The rest of the paper is structured as follows. Section 2 provides a brief overview of the state of the discussion around participatory AI. In Section 3, we narrow in on Himmelreich's critique that participatory AI is too costly to justify, and we explain that, from a socio-technical perspective, there are good reasons to value participatory AI despite the costs it brings. In Section 4, we present our two criticisms that our characterization of participatory AI fails to address. Section 5 discusses how we can move beyond participatory AI to address our concerns through better regulation and analysing alternative AI ownership models.

2 Participatory AI State of the Discussion: Intrinsic Goals, Co-optation, and "Citizen Power"

Participatory AI is a much-debated topic, partly because researchers have conceptualized it in various ways. The overarching discourse suggests that participatory AI implies the involvement of affected stakeholders during the development and deployment process of AI systems. In this section, we provide an outlook on the state of the discussion.

The work by Delgado et al. (2023) serves as the stepping stone to understanding the diversity among theories of participatory AI. To assess the different perspectives, they propose a distinction between four "dimensions of participation", these are the *goals* ("why is participation needed"), the *scope* ("what is on the table"), the *participants* ("who is involved"), and the *methods* ("what form does participation take") (Delgado et al. 2023, 3-4). These four dimensions indicate the different grounds for consensus and disagreement within the participatory AI literature. They also do an important job of highlighting the multiplicity of schools of thought involved in participatory AI research, including "user-centered design, service design, participatory design, co-design, and value-sensitive design, as well as participatory action research, participatory democracy (including deliberation theory), social choice theory, and mechanism design" (Delgado et al. 2023, 2).

From our perspective, the best dimension to synthesize the state of the discussion in participatory AI literature lies in the study of goals pursued by proponents of participatory AI design. We can distinguish between two trends in participatory AI, depending on the goal of the participatory mechanisms. We define these as *instrumental* vs. *intrinsic* goals. Participatory AI projects that pursue an instrumental goal do so to achieve more useful, better-functioning AI systems. Think of a system that reflects outcomes that better represent stakeholders' values (Muller and Liao 2017; Gerdes 2022) or outcomes that increase "profit or legitimacy for the designers" (Delgado et al. 2023, 3-4) through participatory mechanisms. Although we believe the instrumental aim is justified, in this paper, we are primarily interested in the intrinsic value of participatory AI.

For proponents of participatory AI pursuing intrinsic goals, "participation is important because inclusion of the people who may be impacted by the technology [...] is simply the right thing to do" (Delgado et al. 2023, 4). In this sense, stakeholder empowerment entails *citizen* empowerment. The intrinsic value of participation in AI design lies in realizing certain political values, such as equality and freedom. Here, participation aims to empower stakeholders by including them in the design process (cf. Greenbaum 1993; Schuler and Namioka 1993). The need for empowerment is two-fold. First, AI systems affect our lived experiences in multiple ways. Therefore, people subjected to these systems should have a say in supporting self-determination and autonomy. This conception of participatory design represents calls to democratize technology more generally (Sclove 1995). In addition, participatory design can address structural power asymmetries between those that develop and deploy these systems and those affected (see e.g., Birhane et al. 2022; Sloane et al. 2022; Zimmermann, Di Rosa, and Kim 2020; Kalluri 2020; Costanza-Chock 2020; Gould 2019).

One of the main risks that the proponents of participatory AI for its intrinsic value have faced is institutional co-optation. Under the flags of democracy and participation, state actors and corporations aim to benefit from a general legitimation of their investments in AI technologies. Birhane et al. (2022, 6) state that corporate actors can act under the header of participation but use "the results of participatory efforts towards corporate benefits." As a result, these actors can "capitalize on such efforts and build products that maximize profits, with little benefit to communities remains open." Such potential misuse leads to participation-washing, where the ideal of empowerment masks private gain. As the authors elaborate, this form of participation-washing fits into a broader trajectory of using participation to legitimize corporate practices stemming from the Western colonial legacy (Birhane et al. 2022, 6).

To avoid co-optation and foster the pursuit of participatory AI, Sherry Arnstein's (1969) ladder of participation has

become very effective. The ladder metaphorically models citizens' involvement in decision-making processes into different ladder rungs, each representing a degree of participation. At the lowest rung (non-participation), citizens have no influence or voice in decision-making. They are mere spectators, educated or manipulated into supporting the decisions but excluded from the actual design process. As the rungs are climbed, a level of tokenism is reached, in which participation serves a symbolic purpose. Citizens may be consulted, but their inputs hold little weight—just like receiving a token gesture with no genuine impact. At the highest level, Arnstein explores what she calls 'citizen power', representing different stages of meaningful participation. Citizens can collaborate with authorities, sharing the agenda-setting and final decision-making power; they can have substantial influence by having institutional authorities delegate their decision-making power to them, or, at the highest rung of the ladder, they could actively shape decisions by controlling the totality of the decision-making process. The higher rungs of the ladder represent what Young (1990) has described as a just and healthy democratic order.

In the context of AI, an understanding of 'citizen power' is more nuanced. Since AI systems imply transnational networks and data streams, we do not refer to citizens of a particular region or state but to various stakeholders affected by the deployment of AI systems. Which stakeholders are affected depends on the context of deployment and the reach of the AI system. AI systems deployed in narrow contexts, such as the medical field, will demand less for achieving stakeholder empowerment than AI systems deployed in broader contexts, such as public governance, where all citizens of a society are affected, which in turn demands less than AI systems with a global reach such as ChatGPT.

Arnstein's ladder of participation has seen support in the participatory AI literature (e.g., Birhane et al. 2022), and we agree with the general idea that a genuine commitment to empowerment requires providing stakeholders with actual decision-making power. However, although we can point to Arnstein's 'citizen power' as the *right* way of participation, we argue that the participatory AI discourse lacks depth in the analysis of the legal, social, and political arrangements to achieve a real redistribution of power that prioritizes stakeholders. Avoiding participation-washing and moving towards real control would require, for instance, the possibility for stakeholders to refuse the development of the AI system in the first place, which is often not part of the options stakeholders can have when asked to participate in AI design (Delgado et al. 2023). Challenging the way AI systems are produced and deployed requires challenging the legal apparatus and the economic monopolies that uphold power asymmetries between the developers of AI systems and the rest of the stakeholders in the first place, a point of discussion that is not often present in the participatory AI literature.

The key takeaway of this section is that the intrinsic value of participatory AI is only realized at the highest levels of Arnstein's ladder, through citizen power (which participatory AI translates to *stakeholder* power). However, this empowerment process requires rethinking the institutional arrangements that foster unequal power relations between AI developers and other stakeholders. Like any democratic endeavour, such genuine commitment to stakeholder empowerment is resource-intensive and costly. In the next section, we argue that participatory AI design is not unjustified despite these costs.

3 Justifying the Costs of Participatory AI

One of the most prominent criticisms of participatory AI design is found in the paper 'Against "Democratizing AI"' by Johannes Himmelreich (2022). Unlike the title suggests, Himmelreich does not oppose 'democratic AI'. He endorses a deliberative approach to AI governance and argues against claims that more participation is necessarily better. We agree with many of his objections, such as the moral desirability and practical feasibility of calls to participatory AI where *everyone* should be involved (cf. Zimmermann, Di Rosa, and Kim 2020). However, while we do not argue that *more* participation is necessarily better, we do hold that *some* participation is better (and necessary) for stakeholder empowerment, provided it is done in the right way (i.e., according to Arnstein's 'citizen power'). We support a more representative version of participatory AI design, where not everybody is involved, but representative groups are. We thus still categorize ourselves as falling within the group that endorses participatory AI, and we wish to provide some pushback to some of Himmelreich's claims to justify why some participation is necessary for stakeholder empowerment.

Himmelreich rejects the calls for participatory AI design on different grounds, supporting the notion that democracy as an idea "is not well-equipped to afford a proper response to problems of injustice and oppression" (Himmelreich 2022, 1342). One of his arguments is that calls for participatory AI have weak grounds. The idea of participatory AI is that it serves as a general legitimation mechanism for using AI. Legitimation via participation, however, requires justification due to the high costs of participatory democracy. Amongst others, these costs include the fact that participation is generally economically expensive and time-consuming and may lead to worse epistemic results. One way to justify these costs is when there is a presence of coercive power, pervasive impacts, and involvement in cooperative systems. However, Himmelreich argues that in the case of AI these conditions for democratic legitimization do not hold. AI systems do not create new forms of coercion and are only coercive in already coercive sectors (e.g., public

governance, judicial); do not have pervasive impacts (only the underlying societal structures do); and do not extend or create involvement in cooperative systems. Participatory AI, Himmelreich concludes, thus rests on weak normative grounds and does not justify the costs of participation.

While we agree with Himmelreich that participation costs are high, we believe these costs are worthwhile for climbing the ladder of participation. We specifically reject his argument based on *coercion* and *pervasive impacts*. First, we disagree that AI does not coerce. Himmelreich does not specify what he means with coercion, but it involves curtailing natural liberties and restricting individual autonomy. A clear example in the context of the state is that a state can force its citizens to perform certain actions, such as paying taxes, which infringes on citizens' natural rights (in a Lockean sense of property rights). AI systems coerce differently.

They do not coerce us into performing a certain action (e.g., paying taxes); rather, they coerce us in that we—citizens in modern, digital societies—are *forced* to be subjected to these systems. AI systems have become so fundamentally entangled with our society that we cannot function as we do today without these systems. Core societal sectors such as the medical, financial, and judicial have increasingly adapted to and incorporated AI systems. If we focus specifically on one AI application, we can agree with Himmelreich that 'no new coercion' is created. However, broadening our scope to society as a whole, modern societies are structured so that we cannot opt out of AI anymore, as these systems have become deeply engrained in core societal sectors. Taking on this broader perspective, we can say that AI coerces citizens in modern digital societies in a different, broader sense.

This broader understanding of coercion is found along the lines of Gabriel (2022), who argues that AI forms part of the background structure of society. It is also what Sclove implies with his assumption that "technologies profoundly affect and partly constitute those circumstances" (Himmelreich 2022, 1334). Himmelreich (2022, fn. 17) explicitly rejects Sclove's assumption as a reason to democratize technology. However, this rejection misinterprets the meta-analysis Sclove provides on technological influence. Sclove writes:

> Technologies function politically and culturally as social structures by coercing physical compliance; prompting subconscious compliance; constituting systems of social relations; establishing opportunities and constraints for action and self-realization; affecting nonusers; shaping communication; psychological development; and culture generally; and constituting the much of the world within which our lives unfold. (Sclove 1995, 16-17)

Rejecting the assumption that technologies profoundly affect and partly constitute people's lives disregards the fact that we do not live in a social world but in a socio-*technical* world. Although AI 'coercion' in our interpretation of Gabriel and Sclove differs from the more common understanding of coercion in forcing someone to perform a certain action, Himmelreich's claim that AI does not coerce thus does not hold from this broader perspective.

Second, and relatedly, given the interconnectedness between society and technology, we cannot attribute the pervasive impact only to *societal structures*. In doing so, Himmelreich treats AI as a computational artefact and assumes we can isolate the technological artefact in question. However, as many authors have convincingly argued, we cannot separate the computational artefact of AI from its social context, including its developers and deployers (Johnson and Verdicchio 2024; Noorman and Swierstra 2023; Kudina and Verbeek 2019). A socio-technical understanding of AI recognizes that values become necessarily embedded within a system during design and development, embedding and reinforcing these values in society. For instance, defining a 'fair' system matters for how it affects specific groups. Given that these systems are inevitably part of society, to dismiss participatory AI based on the resource intensity of democracy and the lack of a well-grounded foundation overlooks these socio-technical aspects. Thus, we find that calls to democratize AI design through participation are not as unfounded or weakly motivated as Himmelreich suggests.

Although we are sceptical of some of Himmelreich's points to reject participatory AI, we do support many of his claims that much of the debate remains on a superficial level. Aligned with his claim that "[t]he call to "democratize AI" should not emphasize only participation" (Himmelreich 2022, 1344), we contend significant concerns remain with the current discourse on participatory AI that ultimately hinder the possibility of stakeholder power *à la* Arnstein. First, it primarily focuses on what stakeholders can contribute and how that empowers them. Not only should we ask what a stakeholder can contribute, but also *on whose terms* the stakeholder can contribute. Second, the debate overlooks how political-economic structures uphold relations of power between developers and stakeholders. The current discourse then remains limited as it overlooks the broader institutional and economic structures that inform AI development. In the following section, we address these concerns.

4 Moving beyond Participatory AI: Two Arguments

In this section, we present two relevant arguments towards participatory AI. While some criticism of participatory AI has been closely or loosely related to our criticisms, we focus on two points missing from the debate. First, we argue that the participatory AI discourse should consider the arbitrariness of the power relations between developers and

those affected. Second, we argue that participatory AI should consider a political-economic perspective incorporating power relations beyond the developers and end-users.

4.1 The Current Participatory AI Discourse Insufficiently Considers Arbitrary Power Relations between Developers and Affected Stakeholders

Our first argument concerns the power dichotomy between, on the one side, the developers and deployers of an AI system and, on the other side, the people affected by the system. We follow Arnstein in the call that participation–and hence participatory design–does not necessarily empower the participating agent. However, rather than focusing on ineffective participation not empowering an affected stakeholder, we focus on what this lack of empowerment means for the participatory agents' relation with the developers and deployers of AI systems.

First, let us clarify the power dichotomy, which we can understand in at least two ways. Some scholars focus on the fact that the groups facing the negative consequences of AI systems are already marginalized groups. Moreover, these groups facing the negative consequences differ from those shaping these systems, which typically belong to better-off demographics (Eubanks 2018; Noble 2018; Benjamin 2019; Kalluri 2020). Therefore, this understanding of the power relation is specifically concerned with existing power asymmetries in society that reflect marginalized vs. non-marginalized groups (male/female, white/black, etc.).

Another way to understand the power dichotomy is that these systems exercise power over the people subjected to the system. For instance, decision-support systems affect whether someone is allocated social benefits, and recommender systems show what people see. In this way, these systems exercise social power over the people subjected to the systems. This power relation extends to the developers and deployers via the AI system because of the sociotechnical aspect of these systems, such as value-embedding during design (Maas 2023). Therefore, this understanding of the power relation is specifically concerned with the relation between the developing and deploying agents vs. those affected by the system. These people constitute the 'relevant stakeholders'. We are specifically interested in this second relation of power.

This second relation of power resembles *governing* power, where 'to govern' refers to implementing and enforcing the norms of the relevant institutions (Lazar 2024). Regarding such governing power, we must ask whether the power is exercised in a legitimate way and by the right authority. This raises the questions of (1) what makes exercises of power legitimate and (2) what constitutes the right or 'proper' authority? (Lazar 2024) One approach to justify governing power is through direct democracy. Zimmermann, Di Rosa, and Kim (2020) provide clear support for this approach. They write that "[r]ather than allowing tech practitioners to navigate the ethics of AI by themselves, we the public should be included in decisions about whether and how AI will be deployed and to what ends." The authors thus specifically highlight the need for participatory approaches to AI design. However, although we believe there is value in public control, the current calls for participatory AI design are insufficient to effectively address the power dichotomy between the 'shapers' and the 'affected'.

Our reason is rooted in the *arbitrariness* of the power relation, or what is known in political philosophy as *domination*. Being subject to arbitrary power means being at the whim of someone else's will (Pettit 1997; Lovett 2010). Think of a benevolent dictator. Even though the dictator treats their subjects well, they still remain a dictator. In other words, subjects living in a dictatorship have no guarantee or control about how the dictator will treat them. The dictator may be benevolent now, but this does not provide the subjects with a guarantee that they will be treated well in the future. This lack of guarantee shows the moral wrong of domination: it leaves people vulnerable to abuse (Pettit 1997). Thus, to meaningfully address the power dichotomy between the 'shapers' and the 'affected', this power relation must be non-arbitrary.

Participatory AI design draws heavily on stakeholder-friendly approaches such as the historical Scandinavian tradition of Participatory Design (Robertson and Simonsen 2012) and Value-Sensitive Design (VSD) (Friedman et al. 2013). PD originated as a way to emancipate the people working with new technological innovations, and VSD was primarily a way to improve technological innovation. Both approaches aim to achieve this by providing stakeholders with a meaningful say by means of stakeholder inclusion during the design process. However, these approaches have a well-known challenge of accounting for power (Bratteteig and Wagner 2012; Friedman et al., 2021). Particularly in VSD, this challenge includes (1) which stakeholders are 'relevant' and should be invited, (2) how developers should include stakeholder input, and (3) how stakeholders can ensure their input is considered appropriately. These three questions raise highly complex questions, such as what defines a 'relevant' stakeholder, who gets to define what is 'relevant', and who should oversee the final decision. As the key challenge of accounting for power is prominent in the VSD literature, these questions have yet to be answered satisfactorily. Given that decisions regarding stakeholder inclusion rest on the shoulders of the designers themselves, VSD is thus prone to arbitrary power relations. When participatory AI design draws on existing design approaches that have such arbitrary relations of power, participatory AI design inherits these limitations.

The problem with participatory AI design resembles a familiar issue with voluntary ethics guidelines. Voluntary commitments to ethics guidelines are notorious for being

ineffective (Hagendorff 2020) and lacking "teeth" (Rességuier and Rodrigues 2020). As LaCroix and Mohseni (2022) explain, voluntary ethics guides create a 'tragedy of the commons' style of social dilemma, where it is always in an individual party's best interest to not participate, even though everyone is better off cooperating.

Similarly, suppose I ask you to participate but can decide at any moment to exclude you or choose which of your input suggestions is sufficiently relevant to consider. In that case, the participation lacks "teeth" and may even become a form of participation-washing. It may look good on paper but does little to equalize power asymmetries between shapers and affected or hold accountable the superior power of the shapers of AI systems. Indeed, the distribution of power remains the same as without the participation. After all, what can people who participate do to ensure their feedback is considered? And what can people who were not invited to participate do to be invited? Just like voluntary ethics guidelines are insufficient, so too for 'direct' or 'participatory' AI design, where the power of participation lies in the hands of the developers. In other words, as long as the power of participation remains in the hands of the developers, this power is arbitrary.

While comparing AI developers to dictators may seem extreme to some, there is an important analogy to be made. A dictator treats subjects well *at the dictator's whim*, and developers have the people participate *at the developers' whim* if participation remains voluntary and up to the developers/company's decision. If one believes the relationship between a dictator and his subjects is morally objectionable, this would hold–at least in a less extreme sense–for the relationship between developers and the people affected if the people's say depends on the arbitrary will of the developers. While critics of participatory AI design worry about the potential to repeat colonialist tendencies (Birhane et al. 2022) or exploit stakeholders (Sloane et al. 2022), there seems to be little attention to these existing concerns of participatory design approaches that already face objections regarding the voluntariness or arbitrariness of stakeholder inclusion. Himmelreich (2022, 1344) states that we need to address the "shortcomings of existing democratic institutions." Integrating non-voluntary commitments to participation in institutional legal arrangements is one way of doing so.

In sum, the power relations between the developers and deployers on the one hand and the people affected on the other depend on institutional arrangements. We must seriously consider how these institutional arrangements are shaped to see *on whose terms* participation happens. We need participation teeth if we are to address arbitrary power relations successfully. In Section 5.1, we discuss how we can address arbitrariness. But first, we discuss our second critique of current participatory AI discourse.

4.2 The Current Participatory AI Discourse Does Not Sufficiently Address the Political-Economic Mechanisms That Uphold Power Asymmetries

The first argument that has been introduced deals with the power dichotomy between, on the one hand, the developers and deployers of an AI system and, on the other hand, the people affected by the system. Our second argument is that the participatory AI discourse lacks a wider political-economic focus exploring the social structures that uphold these power asymmetries. Without a political-economic outlook addressing how such power is distributed and reproduced, we cannot effectively ensure stakeholder empowerment.

As we have explained in our first criticism of the participatory AI discourse, power relations do not exist in a vacuum; they depend on checks and control mechanisms, such as by means of legal institutional embedding. Similarly, these power relations are shaped and upheld by social norms and economic structures, such as markets, that systematically reproduce arbitrary power (Haugaard and Pettit 2017; Gädeke 2020). In this sense, AI developers hold power over AI users because they take up a specific role in organizing the "production, distribution, and consumption of resources," referred to as the *political economy* (Mosco 2009; Wasko 2004). Thus, this power is rooted in economic relations, and an in-depth analysis of such power relations in the AI development process calls for studying the political economy of AI. There have been growing interdisciplinary efforts to set out a political economy of AI within the discussions of digital capitalism (Srnicek 2018; Zuboff 2019; Dyer-Witheford et al. 2019; Fuchs 2019; Luitse and Denkena 2021). We argue that the participatory AI literature has been disconnected from these research lines and does not sufficiently consider a political-economic outlook, which is necessary to deal with power asymmetries and ensure stakeholder empowerment.

There are two Marxian concepts through which power asymmetries become explicit, which can help us explore the political economy of AI. First, there are *relations of production*, a term that responds to the different ways people are organized for the production of goods. Second, there are the *means of production*, which relates to the ownership and control of the productive assets of society, allowing the production of such goods (Marx 1859).

On the one hand, power becomes explicit in the relations of production, which entail "the dynamics of the labour market, particularly in the process of workers entering and exiting firms, and in the operation of those firms themselves" (Bryan 2020, 120). These relations of production include arbitrary power exerted in the workplace and workers' lack of freedom to change or quit their jobs. Søren Mau (2023) refers to this as the economic power of capital or 'mute compulsion', the social domination that regulates the conditions through which capital is reproduced.

In the context of artificial intelligence, the analysis of the relations of production should be twofold. First, it should critically look at the companies and state actors involved in the extraction of lithium and other minerals (necessary for AI systems to function) and the exploitation to which its workers have been subject, following extractivist and neocolonial logics (Crawford 2021; Birhane et al. 2022). Concerns about dangerous manufacturing and resource-intensive logistics should also be considered here (Dauvergne 2022). There are doubts about how effectively workers can opt out of the production of AI technologies at different points of the production process. Second, it should reflect on the material conditions of workers training the AI models (Perrigo 2023), as well as data extraction concerns, especially in the context of large language models. Such a reflection on the material conditions of workers would also impact workers in other sectors, such as people dedicated to artistic work. While participatory AI proponents might consider workers affected by AI technologies that companies deploy, they often fail to amplify the scope to consider the conditions and desires of workers beyond those directly involved in producing AI technologies.

On the other hand, relating to the 'means of production', within a capitalist economic system, the different structural relations of domination are "rooted in the distribution of control over the means of production," which is "generally granted to the owners of productive assets" (Bryan 2020, 128). In the case of AI technologies, it becomes difficult to define the boundaries of what counts as the means of production because of the complex ecosystem of organizations in which it emerges. Data, for example, could be a productive asset, as well as the networks of underwater cables, GPUs, or other computing systems, depending on how it is framed. Different analyses have highlighted that the key means of production of AI corporations is the massive computing power necessary for it to work, as this is what allows these companies to drive competition (Srnicek 2018; Dyer-Witheford et al. 2019; Luitse and Denkena 2021; Varoufakis 2024). The ownership of this variety of assets is a determinant of understanding the power configuration in the AI design process.

In the context of AI, most of the decision-making power of the design and production of global AI is in the hands of a small corporate oligarchy. As Muldoon (2022, 78) comments, the "once disruptive and dynamic start-ups have grown into powerful monopolies able to use their resources to lobby regulators and implement laws to suit their own needs." In terms of total investment in AI systems deployment and research, companies like Google, Amazon, and Microsoft are way at the forefront, overtaking venture capital firms. In 2023, the deals amount "to two-thirds of the $27bn raised by fledgling AI companies" (Hammond 2023). Rethinking the distribution of the means of production thus entails challenging the accumulation of power of big-tech corporations.

To sum up, our second argument on why the participatory AI discourse fails to address power asymmetries lies in its very limited analysis of the political economy of AI, i.e., the ways AI systems are produced, distributed, and used are a byproduct of the capitalist economic infrastructure. Power emerges from this very core: the relations of production and the ownership and control of the means of production.

5 Suggestions and Ways Forward

Based on our previous criticisms, we propose two preliminary ways to move beyond participatory AI. First, there is a need to realize 'participation teeth' that address concerns regarding voluntary commitments to stakeholder inclusion. Second, we explore alternative AI ownership models to address the political economy of AI better.

5.1 Realizing 'Participation Teeth'

To successfully climb the ladder of participation and effectively empower affected stakeholders of AI systems, we need to reduce the arbitrariness of the power relations between a system's developers and its stakeholders. Doing so requires addressing the fact that decisions regarding stakeholder inclusions are in the hands of the developers.

One way of addressing the shortcomings of existing democratic institutions is by means of regulation. The AI Act is the most advanced regulatory initiative for AI that has been proposed (European Parliament 2024). The AI Act is anticipated to have wide-reaching consequences, including those beyond AI systems that affect European citizens. The EU provides a large consumer-base for many AI systems, which will likely have at least a limited 'Brussels effect' (where non-EU players come to conform to the EU regulation despite not being part of the EU) (Siegmann and Anderljung 2022).

However, the AI Act needs to be more specific on stakeholder inclusion. For instance, it states that "[w]hen identifying the most appropriate risk management measures, the provider should document and explain the choices made and, when relevant, involve experts and external stakeholders" (recital 65), or "[w]here appropriate, to collect relevant information necessary to perform the impact assessment, deployers of high-risk AI system […] could involve relevant stakeholders" (recital 96). These statements leave much room for interpretation and voluntary inclusion to the provider and deployer of the AI system, potentially rendering stakeholder inclusion relatively meaningless. A potential solution to mitigating arbitrary power relations could be, for instance, to make it mandatory to include stakeholders in the design process.

There are many obstacles to incorporating participation into legislation in a way that forces developers of AI systems to include stakeholders. Should the law specify a specific number of stakeholders to include? Should it already indicate which stakeholder groups are relevant to include depending on which AI application? Both seem infeasible and undesirable. The context of development and deployment is highly relevant. Consider the medical field. Different countries, even different hospitals in the same country, operate under different assumptions regarding diagnosing, treatment, and more. How should a general AI Act address these differences? Moreover, it is an obstacle to controlling how stakeholder input can best be considered. Do the developers write a reflection report after their discussions with stakeholders? In the case of designing an AI used for diagnosing in healthcare, it seems unlikely we can identify a set of measures that fits different cultural and context-specific preferences. So, how can we go about forcing AI developers to include stakeholders?

First, it is important to recognize that it is unnecessary to state precisely which stakeholders must be invited and which not. What can be incorporated, however, is that an independent party (either a body of members or one individual) is mandatorily (i.e., by law) included in the design process with special training in stakeholder inclusion. This person can suggest which stakeholders are relevant to include and how many and serve as a control mechanism to ensure that the developers take stakeholder input into consideration.

Second, people must know whether they can be included and how their input will be included. Developers must disclose the information discussed and the input received from the stakeholders and show how they incorporated it. The involved stakeholders must have the opportunity to contest if they disagree about how the developers considered their input. However, the most important thing at these stages is that this is independently checked (the independent party that supports the developing team with stakeholder inclusion could undertake this check).

5.2 Exploring Alternative AI Ownership Models

To successfully challenge the political-economic structures that reproduce power asymmetries, we need to rethink how the productive assets related to AI are distributed and owned. This is not an easy task, as it is difficult to determine the scope; where AI starts and finishes is not clear-cut. If we aim for a small scope, we are at risk that our participatory proposals will not be sufficient to address power asymmetries, and Arnstein's 'citizen power' will not be meaningfully achieved. If we aim for a global scope, we encounter that redesigning the whole AI production cycle would be tied to wider democratic endeavours and political processes. However, we can start by providing different proposals that highlight the value of rethinking ownership regarding AI.

The literature about *platform cooperativism* already provides a good starting point. Coined by Trebor Scholz (2016), this term encompasses proposals surrounding "new ownership models for the Internet" to establish democratic governance. It starts with multi-stakeholder and worker-owned cooperatives in digital labour platforms. In his work, Scholz calls for reproducing the technical capabilities of platforms that Big Tech corporations typically own towards ownership systems rooted in solidarity and the redistribution of benefits. In this sense, platform cooperativism "technological, cultural, political, and social changes" beyond the mere change of legal ownership structures, but a wider reframing of the meaning of innovation and efficiency "with an eye on benefitting all" (Scholz 2016, 14).

On similar grounds, there are other pluralistic approaches to the social ownership of AI, like James Muldoon's proposal of *platform socialism*. This term describes "multiple and overlapping associations" that promote collective ownership models of digital assets "from mutual societies to platform cooperatives, data trusts and international social networks" (Muldoon 2022, 6). This "complex ecology of organizations" could only be supported by a radical rethink of the scale in which the public could meaningfully own them (Muldoon 2022, 6). Muldoon suggests four levels of subsidiarity to work with: *local, regional, national,* and *global*. Ownership of digital assets would then ideally be managed at the smallest possible level to achieve more direct democratic control. Although a clear roadmap about how to achieve platform socialism still needs to be outlined, this could serve as a good solution towards the previously mentioned issue of scaling democratic governance.

Other ownership alternatives try to benefit from the affordances that certain technologies could provide. Such is the case of Decentralized Autonomous Organizations (DAOs) (Spelliscy et al. 2024), which use blockchain systems to design technical solutions that improve cooperative governance. Amongst the features they work on achieving are an effort to establish effective voting systems for the stakeholders, increase member engagement and accountability, improve the predictability of compensations, and better organizational transparency, amongst others.

6 Conclusion

In this paper, we have provided various arguments and proposals to go beyond the current state of the discussion in participatory AI. First, we have argued that the general participatory AI literature needs to consider the broader institutional and socio-economic context as a necessary foundation to establish the *right* way of participation that leads to Arnstein's 'citizen power'. Second, we have argued that

participatory AI has a normative justification, particularly when considering a socio-technical perspective on AI systems. Third, we have indicated that the current state of participatory AI discourse needs to address power asymmetries between developers and stakeholders more extensively. We have highlighted two domains that need to be explored more in the current discourse for participatory AI. These are (1) the legal-institutional background that could provide 'participation teeth' for stakeholder empowerment and (2) the political economy of AI production that fosters such power asymmetries. While we have provided initial ways forward to explore alternative legal arrangements and ownership models for participatory AI, these lines of research require further investigation.

Acknowledgements

JM's contribution to this work was partially supported by the EU H2020 ICT48 project „Humane AI Net" under contract # 952026. The support is gratefully acknowledged. We also thank Filippo Santoni de Sio, Jeroen van den Hoven, and three anonymous reviewers for their insightful comments on earlier drafts of this paper.

References

Akter, S.; McCarthy, G.; Sajib, S.; Michael, K.; Dwivedi, Y. K.; D'Ambra, J.; and Shen, K. N. 2021. Algorithmic bias in data-driven innovation in the age of AI. *International Journal of Information Management,* 60, doi.org/10.1016/j.ijinfomgt.2021.102387.

Angwin, J.; Larson, J.; Mattu, S.; and Kirchner, L. 2022. Machine bias. In *Ethics of data and analytics*, edited by K. Martin, 254-64. Auerbach Publications.

Arnstein, S. R. 1969. A ladder of citizen participation. *Journal of the American Institute of planners,* 35(4): 216-224.

Bartneck, C.; Lütge, C.; Wagner, A.; and Welsh, S. 2021. Privacy issues of AI. In *An introduction to ethics in robotics and AI*, edited by C. Bartneck; C. Lütge; A. Wagner; S. Welsh, 61-70. Springer.

Birhane, A.; Isaac, W.; Prabhakaran, V.; Diaz, M.; Elish, M. C.; Gabriel, I.; and Mohamed, S. 2022. Power to the People? Opportunities and Challenges for Participatory AI. In Proceedings of the 2nd ACM Conference on Equity and Access in Algorithms, Mechanisms, and Optimization (EAAMO '22). Association for Computing Machinery, New York, NY, USA, Article 6, 1–8. doi.org/10.1145/3551624.3555290.

Bratteteig, T.; and Wagner, I. 2012. Disentangling power and decision-making in participatory design. In Proceedings of the 12th Participatory Design Conference: Research Papers - Volume 1 (PDC '12). Association for Computing Machinery, New York, NY, USA, 41–50. doi.org/10.1145/2347635.2347642.

Bryan, A. 2020. The economic implications of Republican political thought. PhD dissertation, School of Law, King's College London, London, UK.

Cammaerts, B.; and Mansell, R. 2020. Digital platform policy and regulation: Toward a radical democratic turn. *International Journal of Communication,* 14: 135-54.

Cohen, J. 2009. Reflections on deliberative democracy. In *Contemporary Debates in Political Philosophy,* edited by T. Christiano and J. Christman, 217–35. Blackwell Publishing Ltd.

Costanza-Chock, S. 2020. *Design justice: Community-led practices to build the worlds we need*. The MIT Press.

Dauvergne, P. 2022. Is artificial intelligence greening global supply chains? Exposing the political economy of environmental costs. *Rev Int Polit Econ* 29(3): 696–718. doi.org/10.1080/09692290.2020.1814381.

Delgado, F.; Yang, S.; Madaio, M.; and Yang, Q. 2023. The Participatory Turn in AI Design: Theoretical Foundations and the Current State of Practice. In Proceedings of the 3rd ACM Conference on Equity and Access in Algorithms, Mechanisms, and Optimization (EAAMO '23). Association for Computing Machinery, New York, NY, USA, Article 37, 1–23. doi.org/10.1145/3617694.3623261

Denton, E.; Hanna, A.; Amironesei, R.; Smart, A.; Nicole, H.; and Scheuerman, M. K. 2020. Bringing the people back in: Contesting benchmark machine learning datasets. arXiv:2007.07399.

Donia, J.; and Shaw, J. A. 2021. Co-design and ethical artificial intelligence for health: An agenda for critical research and practice. *Big Data & Society,* 8(2). doi.org/10.1177/20539517211065248.

Dyer-Witheford, N.; Kjøsen, A. M.; and Steinhoff, J. 2019. *Inhuman Power: Artificial Intelligence and the future of capitalism*. Pluto Press.

European Parliament. 2024. *Artificial Intelligence Act*. https://artificialintelligenceact.eu/the-act. Accessed 13-05-2024.

Feldstein, S. 2019. The global expansion of AI surveillance. Technical Report. Washington DC: Carnegie Endowment for International Peace.

Friedman, B.; Kahn, P. H.; Borning, A.; and Huldtgren, A. 2013. Value sensitive design and information systems. In *Early engagement and new technologies: Opening up the laboratory*, edited by N. Doorn, D. Schuurbiers, I. van de Poel, M. E. Gorman, 55-95. Springer Dordrecht. doi.org/10.1007/978-94-007-7844-3.

Friedman, B.; Harbers, M.; Hendry, D. G.; van den Hoven, J.; Jonker, C.; and Logler, N. 2021. Eight grand challenges for value sensitive design from the 2016 Lorentz workshop. *Ethics and Information Technology*, 23: 5-16.

Fuchs, C. 2019. *Rereading Marx in the age of digital capitalism.* Pluto Press.

Gabriel, I. 2022. Toward a theory of justice for artificial intelligence. *Daedalus*, 151(2): 218-231.

Gebru, T. 2020. Race and gender. In *The Oxford handbook of ethics of AI*, edited by M. D. Dubber; F. Pasquale; S. Das, 251-69. Oxford University Press. doi.org/10.1093/oxfordhb/9780190067397.001.0001.

Gerdes, A. 2022. A participatory data-centric approach to AI ethics by design. *Applied Artificial Intelligence,* 36(1), 2009222.

Gould, C. C. 1990. *Rethinking democracy: Freedom and social co-operation in politics, economy, and society*. Cambridge University Press.

Gould, C. C. 2019. How democracy can inform consent: cases of the Internet and bioethics. *Journal of Applied Philosophy*, *36*(2), 173-191.

Greenbaum, J. 1993. A design of one's own: towards participatory design in the United States. In *Participatory Design*, edited by D. Schuler and A. Namioka. CRC Press.

Hagendorff, T. 2020. The ethics of AI ethics: An evaluation of guidelines. *Minds and machines*, 30(1): 99-120. doi.org/10.1007/s11023-020-09517-8.

Hammond, G. 2023. *Big Tech outspends venture capital firms in AI Investment Frenzy*. Financial Times. https://www.ft.com/content/c6b47d24-b435-4f41-b197-2d826cce9532.

Haugaard, M.; and Pettit, P. 2017. A conversation on power and republicanism: an exchange between Mark Haugaard and Philip Pettit. *Journal of Political Power,* 10(1): 25–39. doi.org/10.1080/2158379X.2017.1286083.

Himmelreich, J. 2023. Against "democratizing AI". *AI & SOCIETY*, *38*(4): 1333-1346. doi.org/10.1007/s00146-021-01357-z.

Johnson, D. G.; and Verdicchio, M. 2024. The socio-technical entanglement of AI and values. *AI & SOCIETY*. doi.org/10.1007/s00146-023-01852-5.

Kalluri, P. 2020. Don't ask if artificial intelligence is good or fair, ask how it shifts power. *Nature,* 583(7815): 169-169. doi.org/10.1038/d41586-020-02003-2.

Kudina, O.; and Verbeek, P. P. 2019. Ethics from within: Google Glass, the Collingridge dilemma, and the mediated value of privacy. *Science, Technology, & Human Values*, 44(2): 291-314. doi.org/10.1177/0162243918793711.

LaCroix, T.; & Mohseni, A. 2022. The tragedy of the AI commons. *Synthese,* 200(4). doi.org/10.1007/s11229-022-03763-2.

Lazar, S. 2024. Automatic Authorities: Power and AI. arXiv:2404.05990.

Lovett, F. 2010. *A general theory of domination and justice*. Oxford: Oxford University Press.

Luitse, D.; and Denkena, W. 2021. The great transformer: Examining the role of large language models in the political economy of AI. *Big Data & Society,* 8(2). doi.org/10.1177/20539517211047734.

Maas, J. 2023. Machine learning and power relations. *AI & SOCIETY*, *38*(4): 1493-1500. doi.org/10.1007/s00146-022-01400-7.

Marx, K. 1859. *A Contribution to the Critique of Political Economy.* Translated by S. W. Ryazanskaya. London: Lawrence & Wishart.

Mau, S. 2023. *Mute compulsion: A Marxist theory of the economic power of capital.* Verso Books.

Mosco, V. 2009. *The Political Economy of Communication.* London: Sage.

Muldoon, J. 2022. *Platform socialism: How to reclaim our digital future from big tech*. London: Pluto Press.

Muller, M.; and Liao, Q. V. 2017. Exploring ai ethics and values through participatory design fictions. *Human Computer Interaction Consortium*.

Noble, S. U. 2018. Algorithms of oppression: How search engines reinforce racism. In *Algorithms of oppression.* New York: New York University Press.

Noorman, M.; and Swierstra, T. 2023. Democratizing AI from a socio-technical perspective. *Minds and Machines*. doi.org/10.1007/s11023-023-09651-z.

Perrigo, B. 2023. Exclusive: OpenAI USed Kenyan Workers on Less Than $2 Per Hour to Make ChatGPT Less Toxic. https://time.com/6247678/openai-chatgpt-kenya-workers. Accessed: 2024-02-06.

Pettit, P. 1997. *Republicanism: a theory of freedom and government*. Oxford University Press.

Pettit, P. 2012. *On the people's terms: a republican theory and model of democracy.* Cambridge University Press.

Rahwan, I. 2018. Society-in-the-loop: programming the algorithmic social contract. *Ethics and information technology,* 20(1): 5-14.

Rességuier, A.; and Rodrigues, R. 2020. AI ethics should not remain toothless! A call to bring back the teeth of ethics. *Big Data & Society*, 7(2). 2053951720942541.

Robertson, T., and Simonsen, J. 2012. Challenges and opportunities in contemporary participatory design. *Design Issues*, 28(3): 3-9.

Scholz, T. 2016. Platform cooperativism. Challenging the corporate sharing economy. Technical Report. New York: NY/Rosa Luxemburg Foundation.

Schuler, D.; and Namioka, A. (Eds.). 1993. *Participatory design: Principles and practices*. CRC Press.

Sclove, R. 1995. *Democracy and technology*. Guilford Press.

Sen, A. K. 2001. *Development as freedom*. Oxford University Press.

Siegmann, C.; and Anderljung, M. 2022. The Brussels effect and artificial intelligence: How EU regulation will impact the global AI market. arXiv:2208.12645.

Sloane, M.; Moss, E.; Awomolo, O.; and Forlano, L. 2022. Participation Is Not a Design Fix for Machine Learning. In Proceedings of the 2nd ACM Conference on Equity and Access in Algorithms, Mechanisms, and Optimization (EAAMO '22). Association for Computing Machinery, New York, NY, USA, Article 1, 1–6. doi.org/10.1145/3551624.3555285.

Spelliscy, C., Hubbard, S., Schneider, N., & Vance-Law, S. 2023. Toward equitable ownership and governance in the digital public sphere. Technical Report. Harvard: Harvard Kennedy School/Belfer Center for Science and International Affairs.

Srnicek, N. 2018. Platform monopolies and the political economy of AI. In *Economics for the Many*, edited by J. McDonnell, 153-63. London: Verso, 153–163.

Varoufakis, Y. 2024. *Technofeudalism: What killed capitalism*. Melville House.

Wasko, J. 2004. The Political Economy of Communications. In *The SAGE handbook of media studies,* edited by J. Downing. Sage Publications.

Young, I. M. 1990. *Justice and the Politics of Difference*. Princeton University Press.

Zimmermann, A., Di Rosa, E., and Kim, H. 2020. Technology can't fix algorithmic injustice. Boston Rev.

http://bostonreview.net/science-nature-politics/annette-zimmermann-elena-di-rosa-hochan-kim-technology-cant-fx-algorithmic. Accessed: 2024-03-12.

Zuboff, S. 2019. *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*. New York: Public Affairs.