Please cite the published version in *Philosophy & Technology*: https://doi.org/10.1007/s13347-024-00792-2

"Democratizing AI" and the Concern of Algorithmic Injustice

Ting-an Lin
University of Connecticut; Stanford University
tinganlin@uconn.edu

Abstract

The call to make artificial intelligence (AI) more democratic, or to "democratize AI," is sometimes framed as a promising response for mitigating algorithmic injustice or making AI more aligned with social justice. However, the notion of "democratizing AI" is elusive, as the phrase has been associated with multiple meanings and practices, and the extent to which it may help mitigate algorithmic injustice is still underexplored. In this paper, based on a socio-technical understanding of algorithmic injustice, I examine three notable notions of democratizing AI and their associated measures—democratizing AI use, democratizing AI development, and democratizing AI governance—regarding their respective prospects and limits in response to algorithmic injustice. My examinations reveal that while some versions of democratizing AI bear the prospect of mitigating the concern of algorithmic injustice, others are somewhat limited and might even function to perpetuate unjust power hierarchies. This analysis thus urges a more fine-grained discussion on how to democratize AI and suggests that closer scrutiny of the power dynamics embedded in the socio-technical structure can help guide such explorations.

Keywords
Democratizing AI; Algorithmic injustice; Socio-technical structure; Structural injustice

1. Introduction

With the profound impacts that artificial intelligence (AI) [1] imposes on the contemporary world, concerns regarding various forms of algorithmic injustice have been raised (Noble 2018; Benjamin 2019; Mohamed, Png, and Isaac 2020; Birhane 2021; Kalluri 2020; Crawford 2021). In response, the idea of making AI more democratic, or to "democratize AI," is sometimes framed as a promising route to mitigate algorithmic injustice. The call to democratize AI is not only proposed by academic researchers (Rahwan 2018; Zimmermann, Di Rossa, and Kim 2020; Wong 2020; Birhane et al. 2022) but also endorsed by people in the tech industries. Many big tech companies have popularized the slogan "democratizing AI" and have talked about their commitment to it (Seger et al. 2023). Despite the popularity of democratizing AI, some

[1] Since the term "artificial intelligence (AI)" was coined in the 1950s, it has been used in various ways, and the technologies underlying systems called AI might also differ. The recent surge of AI has mostly to do with machine learning, which is a type of technology that trains computer systems with significant quantities of data so that they can "learn" or recognize patterns in the existing datasets and then apply them in new cases. Since the focus of this paper is on the recent wave surrounding democratizing AI, I use "AI systems" to refer to machine learning systems unless otherwise noted. Some of the implications of this paper might be extended to AI systems that are designed based on different technologies, although closer examinations are required. While the recent considerations are mainly about the machine learning systems, as I'll discuss in further detail when examining different meanings of "democratizing AI," "AI" here might further mean different things—it is sometimes understood as a tool, as an area of technological innovation, or as a site under governance.

questions remain underexplored. First, what do people mean when they use the term "democratizing AI"? As some scholars have noted, the use of "democratizing AI" is elusive, as it has been associated with multiple meanings and practices with potentially conflicting goals (Rubeis, Dubbala, and Metzler 2022; Seger et al. 2023). Without proper clarification, the discussions surrounding democratizing AI would tend to talk past each other, thereby blocking effective explorations. Second, while democratizing AI is often treated as a desirable trend, the extent to which it helps mitigate algorithmic injustice remains underexplored.[2] How do different notions and practices associated with "democratizing AI" respond to the concern of algorithmic injustice? Are they all suitable, or do some of the notions or practices bear more prospects than others? Would some forms of democratizing AI perpetuate algorithmic injustice?

The overarching goal of this paper is to disambiguate the idea of "democratizing AI" and to examine the implications that "democratizing AI"—when understood and implemented in different ways—has on algorithmic injustice. While some critical reflections on "democratizing AI" have been raised recently (Sætra, Borgebund, and Coeckelbergh 2022; Himmelreich 2022; Seger et al. 2023; Noorman and Swierstra 2023), a systemic examination of these different forms of "democratizing AI" through the lens of algorithmic injustice—which is one of the key motivations for democratizing AI—does not exist. In a time when the slogan of "democratizing AI" is used by different actors for different purposes, it is of timely urgency to pay closer scrutiny to the implications that different proposals of democratizing AI have on shaping the power dynamics embedded in the socio-technical structure. By centering the examinations around algorithmic injustice, this paper aims to shed some light on envisioning a more just path to democratize AI.

To set the ground for the analysis, in Section 2, building on Iris M. Young's theory on structural injustice, I present a socio-technical understanding of algorithmic injustice. According to this understanding, algorithmic injustice is a case of structural injustice that exists when the socio-technical structure shaped by AI systemically exposes large groups of people to undeserved burdens while conferring unearned benefits to others, thereby exacerbating unjustified power hierarchies or imbalances between people along various axes of social categories. Understanding algorithmic injustice in this way, I argue, helps illuminate why democratizing AI might seem a promising response to it. In Sections 3 to 5, I examine how recent proposals on democratizing AI respond to algorithmic injustice. To do so, I lay out three notable notions of democratizing AI that have been used widely: *democratizing AI use* (Section 3), or making AI systems accessible for more users, for example, by reducing the costs of using AI systems; *democratizing AI development* (Section 4), or getting more people involved in the AI development process, such as by lowering technical entry barriers or participatory design; and *democratizing AI governance* (Section 5), or making AI a domain under the democratic governance through, for example, existing democratic institutions in the government sector, direct participation, or representative deliberation. Although these

[2] One notable exception is Himmelreich (2022), who critically examined the proposal to govern AI via direct and broad participation and argued that this proposal is not the right kind of response to algorithmic injustice. However, as the following discussions will illuminate, Himmelreich's analysis only concerns one notion of democratizing AI (what I refer to as *democratizing AI governance*) and focuses on one specific form of measure associated with it. More details will be discussed in Section 5.

three notions and the associated measures discussed in this paper are not an exhaustive list,[3] they cover a good range of notable ideas and can serve as a helpful starting point for the analysis.

In Section 6, I reflect on the analysis and discuss the practical implications of approaching democratizing AI. The analysis from the previous sections reveals that there is no simple yes or no answer to whether "democratizing AI" can mitigate algorithmic injustice; instead, how the detailed notions and practices are laid out makes huge differences. For each of the three notions examined, some versions bear better prospects of mitigating the concern of algorithmic injustice, while others are somewhat limited and might even function to perpetuate problematic power hierarchies. Without a proper distinction between different ways of democratizing AI, we face a danger that I refer to as "democracy washing," which occurs when the label "democratizing AI" is used overly generally and may function to block needed scrutiny of the associated practices and their impacts. I end by discussing how paying closer attention to the power dynamics embedded in the socio-technical structure can help avoid democracy washing and refine approaches to democratizing AI.

Some clarifications are needed before moving on. First, there is a difference between *democratizing AI* and *AI for democracy*—the former is about incorporating the insights or values of democracy into the domain(s) surrounding AI, and the latter is about using AI to support or strengthen democracy as a system of government. Both issues are important and have received much attention in both academic and public discussions.[4] Other questions might concern the relationship between these concepts, and we could perceive that these two efforts might be mutually reinforcing. However, given its limited scope, this paper focuses only on an analysis of democratizing AI.

Second, while this paper focuses on whether and how democratizing AI responds to algorithmic injustice, this does not mean to suggest that addressing algorithmic injustice is the only purpose that democratizing AI aims for or should aim for. Even if the following analysis finds democratizing AI unsatisfactory in mitigating algorithmic injustice, it does not suggest a total dismissal of the movement to democratize AI, nor does it rule out the possibility of justifying democratizing AI via other grounds, such as accelerating technological innovation, ensuring legitimacy, and increasing public trust. Nonetheless, given the influential role that AI has in shaping the socio-technical structure, extra caution should be raised to the potential impacts of democratizing AI on the power dynamics embedded in such a structure.

---

[3] Seger et al. (2023) provide a helpful taxonomy of "democratizing AI," including democratizing use, development, profits, and governance. In this paper, I'll focus my discussions on the three most notable notions: democratizing use, development, and governance, and my usage of these three terms will be more or less similar to their terminologies.

[4] Since this paper focuses on democratizing AI, I briefly mention some examples of how AI for democracy appears in the academic and public realm here. For philosophical discussions on whether and how AI may improve democracy, see, for example, Himmelriech (2021), Mikhaylovskaya (2024), and Landemore (2024). An example of AI for democracy being of concern in the public realm is when the Joe Biden administration launched the International Grand Challenges on "democracy-affirming technologies." Here, the goal was to develop technologies that would better support the democratic system. For more details, see <https://www.whitehouse.gov/ostp/news-updates/2021/12/08/white-house-announces-launch-of-the-international-grand-challenges-on-democracy-affirming-technologies-for-the-summit-for-democracy/>.

2. "Democratizing AI": a tentative response to algorithmic injustice?

Against the backdrop of various ethical and societal concerns surrounding AI, "democratizing AI" has become a popular path endorsed by both academics and people in the tech industries. However, when reflecting on the theoretical relations between democracy and injustice, democracy might not always seem to be the most effective response to injustice. What is algorithmic injustice? Why might "democratizing AI" be considered a promising path in response to algorithmic injustice? Before examining how and whether various measures surrounding democratizing AI respond to concerns of algorithmic injustice, clarifications on these issues are needed. In Section 2.1, building on Iris M. Young's theory on structural injustice, I argue that algorithmic injustice can be understood as a case of structural injustice that exists when the socio-technical structure shaped by AI systemically exposes large groups of people to undeserved burdens while conferring unearned benefits to others, thereby exacerbating unjustified power hierarchies or imbalances between people along various axes of social categories. In Section 2.2, I discuss why the rough idea of making AI more democratic seems to be an appealing response to algorithmic injustice, at least when understood as a case of structural injustice.

2.1 Algorithmic injustice: A socio-technical understanding

Social structures can be understood as background conditions that enable and restrict the options and opportunities for individuals. Through the complicated interactions between *norms* (e.g., institutionalized laws and social norms), *schemas* (e.g., associated symbols, meanings, and values), and *distributions of resources* (access to technologies), social structures are constantly formed and shaped and continue to influence individuals' options and opportunities (Young 2011; Haslanger 2016). While all kinds of social structures have this kind of enabling and restricting feature, as Young (2011, 52) put it, *structural injustice* exists when "social processes put large groups of persons under systematic threat of domination or deprivation of the means to develop and exercise their capacities, at the same time that these processes enable others to dominate or to have a wide range of opportunities for developing and exercising capacities available to them." In other words, structural injustice exists when the impact of social structure puts large groups of people under systemic undeserved burdens (e.g., domination, oppression, exploitation, etc.) while conferring unearned benefits (e.g., power and resources) to others. In essence, structural injustice concerns the unjustified power hierarchies or imbalances between people along various axes of social categories (e.g., race, gender, class, etc.).[5]

Building on Young's idea of social structure, I suggest that we can characterize the background conditions of the current world as a *socio-technical structure* where many domains are shaped by AI. By this, I mean that the design, development, and deployment of AI have been closely interacting with other elements of social structure (e.g., norms, schemas, and distribution of resources) and exert profound impacts

[5] Young's discussions of structural injustice focus on domination and oppression as the key burdens associated with structural injustice; however, many scholars have expanded the usage of structural injustice to analyze other forms of associated burdens, such as exploitation (McKeown 2016), alienation (Lu 2017), health disparities (Chung 2021), algorithmic bias (Lin and Chen 2022), and violence (Lin 2024). I follow this trend to use structural injustice as a broader category to capture various forms of unjustifiable power imbalances.

on the resulting socio-technical structure.[6] Such interactions between AI and other elements of social structure are bidirectional, and I have used AI in the healthcare domain as an example to illustrate this phenomenon (Lin and Chen 2022). On the one hand, existing norms, schemas, and distribution of resources influence the kinds of AI systems that are designed, how AI systems are deployed, and who is involved in the decision-making process. For example, the availability of funding, digitized medical data, and the interests of decision-makers hugely impact the kinds of healthcare issues regarded as suitable for developing AI systems. The distribution of medical data influences the performance of developed AI systems and their performance in different populations. How engineers select reference standards (namely, the criteria to be used as a proxy for "ground truth") for validating the performance of AI systems is also shaped by the norms of general practices (e.g., treating doctors' diagnoses as the ground truth). On the other hand, various AI systems have been incorporated into the workflow of medical practices and used to assist doctors' diagnoses, make suggestions on the allocation of material resources, and provide services that allow general users to better track their health conditions. In this way, these developed AI systems, together with the values embedded in the process of AI design, development, and deployment, continue to interact with the existing elements of social structure to shape the healthcare resources and services that are made available to people.

From this perspective, algorithmic injustice can be understood as a case of structural injustice, which exists when the resulting socio-technical structure (formed through the close interactions between AI and other elements of social structure) systemically exposes large groups of people to undeserved burdens while conferring unearned benefits to others, thereby exacerbating unjustified power hierarchies or imbalances between people along various axes of social categories. Indeed, as Young (2011, 62) pointed out, structural injustice is often an unintended consequence. Masses of individuals contribute to large-scale social processes by simply enacting their own projects, believing that they are merely following the rules and trying to accomplish legitimate goals. For example, engineers think they are simply using the "best" datasets available to them, hospitals are merely incorporating AI systems into their practices, and doctors are simply following the suggestions of the algorithms. However, the accumulated outcomes of these interactions may sometimes be unjust. An algorithm widely used in US hospitals systematically distributes fewer medical resources to Black patients (Obermeyer et al. 2019), several diagnosis tools perform much better on White men than on other demographic groups (Shankar et al. 2017; Adamson and Smith 2018; Kaushal, Altman, and Langlotz 2020), and the healthcare needs of people with disabilities remain under-addressed (Smith and Smith 2021). In this way, the design, development, and deployment of AI in healthcare exacerbate existing health disparities among people situated in different social positions.

2.2. The "democratizing" turn in response to algorithmic injustice

In this section, I argue that understanding algorithmic injustice as a case of structural injustice helps illuminate some changes in the discussions surrounding algorithmic injustice. It should be clarified that my claim is not that such an understanding of algorithmic injustice is always explicitly endorsed; rather, this understanding provides helpful explanations for several developments in the field. It explains why the attention has expanded from "biased algorithms" to broader ethical and social issues surrounding AI,

[6] One related discussion that characterizes the basic structure of society as a socio-technical system can be found in (Gabriel 2022).

clarifies the goal of addressing these issues, and sheds some light on why the appeal of making AI more democratic might seem promising.

First, understanding algorithmic injustice as structural injustice helps expand the attention from biased outputs of AI systems to a wider range of ways through which AI systems influence the socio-technical structure and exacerbate unjust power hierarchies and imbalances. In the beginning, under the label of "algorithmic fairness," the discussions in the field have been focused on ensuring parity of some statistical measures of AI systems and debated around which statistical measures are the most appropriate one(s) (Binns 2020; Narayanan, n.d.). Several critics have then pointed out that such an overemphasis on the analysis of the performance of isolated AI systems is problematic and urged to recognize that these systems are embedded in the broader social structure (Hoffmann 2019; Le Bui and Noble 2020). Building on these reflections, calls have been raised to shift attention from addressing biases represented in isolated AI systems to the large-scale influence on power asymmetries that AI contributes to (Kalluri 2020; Mohamed, Png, and Isaac 2020; Crawford 2021; Birhane 2021).

Given that the idea of structural injustice concerns unjustified power imbalances between people of different social categories, the structural-injustice understanding of algorithmic injustice expands the attention to broader aspects surrounding AI. Under this understanding, algorithmic injustice not only concerns how the AI systems distribute resources in an unfair way but also how different groups of people are represented through AI systems and whether they are recognized by the AI systems.
In essence, structural injustice concerns the unjustified power hierarchies or imbalances between people along various axes of social categories. Examples include when women of color are largely misclassified by facial analysis algorithms (Buolamwini and Gebru 2018), when search engines associate people of color with derogatory images (Noble 2018), and when generative AI systems produce text and pictures that amplify racial and gender stereotype (Bianchi et al. 2023), which all contribute to the reproduction of intersectional oppressive systems. Furthermore, algorithmic injustice concerns not only the hierarchical power imbalances represented in the results produced by AI systems but also those embedded throughout the process of designing, developing, and deploying AI systems. Examples include how the development of AI exploits the labor of workers and produces environmental harm while conferring economic benefits to big tech companies (Crawford 2021), how AI systems enable the government to conduct surveillance on people across many domains for central control (Zuboff 2019), and how the decision-making power surrounding AI is clustered into the hands of the privileged few (Gebru 2020; Le Bui and Noble 2020). Overall, the socio-technical understanding of algorithmic injustice presents a more comprehensive ground to critically examine the power dynamics embedded and reproduced through the interactions between AI and other elements of the structure.

Second, understanding algorithmic injustice as a structural injustice makes it salient that it is a socio-political problem that cannot be merely solved through technical approaches but would require some shared efforts among multiple participants of the socio-technical structure to bring about suitable change. According to Young's (2011, 104) *social connection model* (SCM), ordinary individuals "bear responsibility for structural injustice because they contribute by their actions to the processes that produce unjust outcomes." In other words, all agents whose actions contribute to the reproduction of the social structure are "connected" with the social structure and thereby bear responsibility (Young called it *political responsibility*) for the related structural injustice. She further suggested that such a political responsibility is a *shared responsibility* that can only be discharged through collective actions among different actors connected with the social structure (2011, 109).

Third, the structural-injustice understanding of algorithmic injustice also reconceptualizes the goal of addressing these issues. Instead of merely aiming to "de-bias" or design "fair" AI systems, scholars have argued that the goal of addressing algorithmic injustice should be more substantive, focusing not only on avoiding algorithms from exacerbating power imbalances but also considering how algorithms could be used to foster social justice (Davis, Williams, and Yang 2021; Green 2022; Lin and Chen 2022), which could be understood as transforming the socio-technical structure to contain less unjust power hierarchies in the future. These suggestions echo Young's idea about the goal of addressing structural injustice, which is not backward-looking, or about attribution of guilt or fault, but rather primarily forward-looking. As Young put it, "Being responsible in relation to structural injustice means that one has an obligation to join with others who share that responsibility in order to transform the structural processes to make their outcomes less unjust" (2011, 96). All agents who contributed to the (re)production of the social structure need to join the collective actions with others to transform the social structure to make it less unjust in the future.

From this perspective, we can see why the call to "democratize AI" would gain traction. The rough themes that are usually associated with democracy, such as power decentralization or having a space for more diverse groups to interact with each other, seem to sketch some promising directions that are in line with the structural responses to algorithmic injustice, which warns against technocentric responses and urges the participation and involvement of more diverse groups of people in shared efforts. For example, after a critical analysis of the limitations of the technical approach to algorithmic injustice, Zimmermann et al. (2020) stated that "We need greater democratic oversight of AI not just from developers and designers, but from all members of society." Similarly, Wong (2020, 226) argued that the development of fair algorithms is not a technical task but "a *political* question that should be resolved *politically*" and suggested that "one promising way forward is through democratic communication." As Gebru (2020, 264) put it, "the design of ethical AI starts from whom is given a seat at the table." Many scholars thus urged the participation of more stakeholders in the decision-making process of AI design (Jo and Gebru 2020; Costanza-Chock 2020; Birhane 2021). At the same time, the slogans of "democratizing AI" or "making AI more democratic" are also embraced by people in the tech industries. Many big tech companies, including Google (Hasbe and Lippert 2020), Microsoft (Microsoft News Center 2016), and Meta (Meta AI 2022), all talk about democratizing AI, presenting it as a desirable goal or even a commitment for the company; although it is less clear what's the motivation behind such a trend in the industry.

Despite the popularity surrounding "democratizing AI," scholars have pointed out that this term has been used in different ways, making the notion elusive (Rubeis, Dubbala, and Metzler 2022; Sætra, Borgebund, and Coeckelbergh 2022; Seger et al. 2023). Furthermore, even though the concern of algorithmic injustice is one key background concern that animates the call to democratize AI, there has not been much detailed analysis of how these different notions and associated measures may address the concern of algorithmic injustice. In the following sections, I analyze three notable uses of democratizing AI: democratizing AI use (Section 3), democratizing AI development (Section 4), and democratizing AI governance (Section 5) in terms of how they respond to algorithmic injustice (as a case of structural injustice). For each of the three notions, I explain the rationales behind it, highlight popular practices that have been proposed, and discuss their respective prospects and limits in responding to the concerns raised by algorithmic injustice.

3. Democratizing AI use: Prospects and limits in mitigating algorithmic injustice

The first notion of democratizing AI, *democratizing AI use*, is about making AI accessible to a wider range of people that goes beyond AI specialists. For example, Microsoft has indicated its commitment to democratizing AI—"to take it from the ivory towers and make it accessible to all" (Microsoft News Center 2016). Here, AI is viewed as a powerful tool, and the motivation behind democratizing AI use is that such a powerful tool should not be only used or accessed by a small group of people who are already privileged; rather, it should be made more widely available, allowing more people (in terms of number and diversity) to directly use AI or enjoy the benefits of AI-assisted services. To do so, the popular practices surrounding democratizing AI use include reducing the costs of using AI systems or providing services to allow users to incorporate AI systems into their workflows (Seger et al. 2023).

Democratizing AI use bears some promise in mitigating existing power hierarchies in the socio-technical structure. Importantly, if we understand AI as a powerful tool that can equip users with a better capacity to complete various tasks, then it seems reasonable that such a powerful tool should not be clustered into the hands of a small group of people who are already privileged, as such a tendency to confer benefits to them would exacerbate existing power imbalances in the society. Such an idea can also be seen in the efforts to democratize the use of some other forms of technology, from computers and the internet to 3D printing. After all, as the term "digital divide" indicates, there is unequal access to modern information and communications technology between people with different demographic and socio-economic characteristics, including race, gender, income, education level, geographic location, etc. (Ragnedda and Muschert 2013). As emerging digital technologies are being incorporated into various aspects of the contemporary world, the capacity to navigate through them is crucial, and limited access to digital technologies disadvantages groups of people in their opportunities in education, employment, and political participation, among other domains. With AI's fast-growing impact on various domains of society, it is sensible to ensure that people in marginalized positions have access to useful services that AI can provide to avoid widening the digital divide.

In addition to reducing digital divides, increasing access to AI tools has the potential to reduce other forms of social inequalities when it is used to provide benefits or resources to marginalized groups. For example, consider the phenomena of existing health disparities in which medical resources have been distributed unevenly across various social categories (Nelson 2002; Manuel 2018). With AI-assisted tools for diagnosis becoming cheaper and available to more people, AI has the potential to provide medical services for those who are currently underserved. For example, Google has developed an AI system that helps healthcare workers perform large-scale screening of diabetic retinopathy, an eye condition that may cause vision loss in people with diabetes. The system has been deployed in India, where there is a drastic shortage of eye doctors compared with the populations with diabetes (Gulshan et al. 2016). This explains why, in the healthcare domain, several scholars have proposed democratizing AI use by emphasizing the need to enable more medical practitioners, not just those who are in large medical institutions, to use AI tools that bear the potential to bring higher quality and more equitable healthcare services to currently underserved populations (Allen et al. 2019).

However, democratizing AI use is also accompanied by the serious dangers of replicating or even exacerbating existing structural injustices. After all, while the reasoning behind promoting broader access to AI has been focused on the potential benefits that AI as a tool might bring to the world, the potential harms that AI tools can bring about should not be overlooked. This concerns is referred to as *dual-use research of concern*, namely that the same technology has the potential to be used for both good and bad purposes (Selgelid 2013, p.4). Indeed, several instances of the dual-use nature of AI systems have been reported (Koplin 2023). Deepfake, which was originally designed to equip the user to do video editing, has been used to impose new forms of objectification and domination on women (Rini and Cohen 2022). ChatGPT, a text-generation tool that can be used for educational purposes, has also been used to create propaganda and disinformation (Deng and Lin 2022). These examples of the malevolent use of AI manifest the fact that intentions for good purposes in designing AI systems are not sufficient to prevent the potential negative impacts it might have. Making AI systems widely accessible increases the risk of using them in negative ways, many of which might further perpetuate existing injustices in the world.

Furthermore, the use of AI systems may perpetuate existing power imbalances, even without malicious users. Ample evidence indicates that AI systems tend to produce unfair predictions and amplify stereotypes, and when these AI systems are widely used as suggested by democratizing AI, the unjust impacts on the oppressed would be further exacerbated. This can be observed when an AI system that systematically judges that Black patients do not need as much healthcare support is broadly used in many hospitals in the US, thereby exacerbating existing health disparities (Obermeyer et al. 2019). Similarly, consider the case of Google's search engine, which associates women of color with degrading images and descriptions (Noble 2018) or the case in which many easily accessible text-to-image generative AI systems produce images with racial and sexual stereotypes (Bianchi et al. 2023). With more users accessing Google's search engine or generative AI models, the unjust impact of these distorted stereotypes is more difficult to mitigate.

The reflections above reveal that there is no straightforward relationship between access to AI tools and algorithmic injustice. Allowing more people to use AI tools bears the potential of reducing digital divides and bringing valuable resources and benefits to marginalized populations; however, wider access to AI systems also increases the danger of broadcasting problematic impacts, which often function to strengthen existing inequalities between social groups. Instead of treating more access to AI systems as always better, context-sensitive analysis is required for more fine-grained examinations of the pros and cons of increasing access to AI systems, deciding on the scope of access, and exploring alternative measures to realize the desired goals (more to be discussed in Section 6).

4. Democratizing AI development: Prospects and limits in mitigating algorithmic injustice

The second notion of democratizing AI, *democratizing AI development*, is about getting a wider range of people involved in the process of AI development. In contrast to democratizing AI use, in which AI is considered an end product, in democratizing AI development, AI is understood as an innovative area, and the focus is on who is involved in the process of designing and developing these AI systems. Democratizing AI development suggests that the opportunity to be part of the process of AI design and development should not be clustered in the hands of small populations but rather be distributed to larger and more diverse populations.

Democratizing AI development has the potential to shape our existing socio-technical structure and make it more just. First, democratizing AI development targets skewed power distributions in terms of the power to design and develop AI systems, which have huge impacts on the contemporary world. Currently, the majority of the power to develop AI systems is clustered in the industry (AI Index Steering Committee 2023). With AI systems bringing about enormous influence on the socio-technical structure, the striking accountable power that is clustered in the hands of small groups of AI developers raises serious concerns about domination (Maas 2023). As Christine T. Wolf, a member of IBM Research, put it, "Democratizing AI calls for wider participation in the everyday work practices of AI, motivated by the concern that the ability to engineer and put to use AI technologies should not rest in the hands of a privileged few" (Wolf 2020, 15). From this perspective, democratizing AI development holds the promise of reducing unjustified power imbalances, thereby mitigating the potential threats of domination.

Second, democratizing AI development has the potential to make the produced AI systems better accommodate the needs, interests, and perspectives of diverse populations. Currently, many widely disseminated AI tools are developed by big tech companies in Global North, where there is a dearth of demographic diversity (Neely, Sheehan, and Williams 2023); such a lack of diversity in the development team might have blind spots that omit some essential needs or perspectives. For example, Apple launched a health-tracking app described as a "comprehensive health-tracking app," but the app lacked the capacity to track the menstrual cycle; hence, the product failed to recognize large groups of biological women's needs and resulted in a form of exclusion (Eveleth 2014). If more women were included in the design team or involved in the design process, there would have been a great chance of catching such blind spots before the app's launch. In addition to gender, multiple examples have revealed that AI systems often fail to accommodate the perspectives of people in marginalized positions, such as people with disabilities (Smith and Smith 2021) and gender-non-confirmative groups (Keyes 2018; Hamidi, Scheuerman, and Branham 2018; Costanza-Chock 2018), which further strengthen the existing oppressive systems against them. From this perspective, involving more diverse groups of people—especially those who are currently in more marginalized social positions or bearing perspectives that are currently underrepresented—in the design process seems to be a reasonable move to reduce these blind spots with oppressive consequences.

Moreover, diversifying AI development teams may further play a crucial role in shaping the power relations of existing socio-technical structures. As feminist standpoint theorists argue, one's knowledge about the world is hugely influenced by one's social position, and being situated in more oppressed and marginalized positions equips one with valuable epistemic resources in challenging injustices embedded in existing social structures (Harding 1992; Collins 2002; Wylie 2003). Conducting inquiries from the lived experiences of the marginalized, feminist standpoint theorists suggest, would help reveal the existing power relations and raise questions that may bring about social change. Following this reasoning, involving marginalized populations and ensuring that their perspectives are given proper weight in the process of AI design (e.g., in deciding the kinds of issues worth developing AI systems to address and in evaluating whether the designed AI systems succeed in accommodating their needs and interests) is a crucial step toward more just design.

Although democratizing AI development bears some potential in mitigating algorithmic injustice at the conceptual level, an assessment of the measures that have been adopted reveals some concerns and limits. Below, based on the aspects of AI development and the target populations to get involved, I differentiate and briefly examine two categories of practices to democratize AI development: (1) lowering technical entry barriers and (2) participatory design.

The first category of practices focuses on the technical aspects of AI development. It aims to allow a wider range of people (in terms of both number and diversity) to develop their AI systems by lowering technical entry barriers, such as making models and datasets more accessible. One main reason for the inaccessibility of current AI development, which is mostly done in the industry and by big tech companies, has to do with the fact that AI development is resource-intensive (requiring huge amounts of data, computing power, and money) and the current situation that industry actors have much greater amounts of resources compared with those in other sectors. To change the situation, some initiatives have been taken to make these resources more accessible, such as sharing pre-trained AI models or datasets. For example, in 2022, Meta shared its newly developed large language model, Open Pretrained Transformer (OPT-175B), with research communities (people in academia, civil society, government, and some industry research organizations), aiming to enable a "much broader segment of the AI community" to conduct the related research (Meta AI 2022).

Although involving more diverse populations (especially those currently in marginalized positions) in the domain of AI development is laudable, I want to raise some caution about its limitations. First, similar to the dual-use concerns raised regarding democratizing AI use, making it easier for more people to develop AI systems might also increase the danger of bringing about broader unjust impacts through AI systems, either in the cases of malicious designers or merely negligent designers.[7] Second, such a practice of lowering the technical entry barriers and allowing more AI systems to be produced would contribute to the already surging environmental and climate costs of AI (Strubell, Ganesh, and McCallum 2020; AI Now Institute 2023; Nordgren 2022). It is well documented that negative environmental impacts tend to be disproportionately imposed on already marginalized groups (Gochfeld and Burger 2011), such as racial and ethnic minorities. In this way, despite good intentions, lowering technical entry barriers might have an unjust side effect to exacerbate these unjust impacts.

Another concern should be raised regarding the limited structural change that this move can bring about. When the underlying infrastructure—namely those pre-trained AI models and datasets—remain the same, allowing more diverse people to develop AI systems on top of it might not bring much different results. Furthermore, when the same set of pre-trained models and datasets are used by many developers and turn into different guises, this might contribute to algorithmic monoculture, which occurs when multiple decision-makers rely on the same or similar set of algorithms (Kleinberg and Raghavan 2021), thus contributing to the phenomena of outcome homogenization, where individuals or groups experience the same outcomes repeatedly across different AI systems (Bommasani et al. 2022). This is especially worrisome when AI systems developed on top of the same set of infrastructures repeatedly assign undesirable outcomes to the same groups of individuals, thereby contributing to systemic exclusion and reinforcing unjustifiable hierarchies. Without some closer scrutiny of the values embedded in the models and datasets, attracting more people in the domain of AI development would only tend to produce AI products that replicate the problematic outcomes.

[7] Thanks to an anonymous reviewer for raising this concern.

By contrast, under the label "participatory AI," the second category of initiatives recognizes that there are many non-technical aspects throughout the AI development process (e.g., deciding what kind of AI systems should be designed, whose interests should be taken into account, and how to make trade-offs between different values when in conflict) and aims to involve more non-tech practitioners (such as the stakeholders and end-users) in the process. Scholars have argued that participatory approaches can increase the representation of the marginalized in datasets (Jo and Gebru 2020) as well as provide the means for marginalized groups to challenge power asymmetries (Costanza-Chock 2020). Some examples of positive outcomes of participatory design have also been recorded, such as optimizing worker well-being and building machine translation tools for low-resourced languages (Birhane et al. 2022). Thus, participatory design provides some direction to reflect on the value embedded in the development of AI systems and shows some promise in bringing more substantive structural change.

Even though the ideal cases of participation bear great potential in shaping a more just socio-technical structure through the design of AI, real-world constraints often prevent them from happening. Interviews with AI researchers and practitioners reveal that "they felt caught between an idealized, ambitious vision of stakeholder empowerment and the practical constraints of time and resources" (Delgado et al. 2021). Another survey identified practical barriers to embedding participatory approaches in AI labs in industries, including resource intensity, atomization, lack of clear and shared understanding of practices, etc., which together "result in a piecemeal approach to participation that confers no decision-making power to participants and has little ongoing impact for AI labs" (Groves et al. 2023). Further, caution has been raised regarding "participation washing," meaning that efforts are mischaracterized under the label of participation and are used to achieve other goals or even to legitimize injustices (Sloane et al. 2020; Birhane et al. 2022). As the label of participation is used in a wide range of ways, including some forms of practices that are tokenizing or even exploitative, there is a danger that companies might use participatory design as a smokescreen without making real efforts to change the power dynamics embedded in the AI development process.

Overall, the reflections on democratizing AI development reveal that while this idea is promising in contributing to algorithmic justice at the conceptual level, closer examinations of the measures and real-world constraints in practice reveal mixed results. Practices that focus on reducing the technical entry barriers would allow more people to contribute to the technological development of AI; however, this move would also tend to contribute to environmental costs and algorithmic monoculture, which both have negative implications for already marginalized populations. By contrast, participatory designs pay attention to the values embedded in the process of AI development and work to include more diverse stakeholders in the process, which might bring about more fundamental structural change. However, it should be noted that the meaningful participation and engagement of diverse populations takes great effort; in reality, many participatory design initiatives fall short of bringing in the desired goals and might even risk becoming tokenizing or exploitative.

5. Democratizing AI governance: Prospects and limits in mitigating algorithmic injustice

The third notion of democratizing AI, *democratizing AI governance*, shifts attention to the governance of AI, and it suggests that many AI-related issues should be a domain governed through some democratic practices, namely, practices of collective decision-making under which people are treated as equals. Compared with the previous two notions, AI here is understood even more broadly as a socio-technical domain that raises various questions for decision-making, including both *ex post* regulations on a certain developed AI system (e.g., Who should have access to use a certain AI system? Who should be held accountable when an AI system reinforces racist stereotypes?) and *ex ante* ones (e.g., Which AI systems should be developed? How can a trade-off between values, such as accuracy and fairness, be made? Should AI development be slowed down?). The discussions on AI governance concern who should be making these decisions and through what kinds of processes, and the idea of democratizing AI governance suggests that these decisions should be made by general members of society through democratic mechanisms and processes, namely methods of collective decision-making that can be characterized as treating the people as equals, are the routes to take. For example, as Zimmermann et al. (2020) stated, "[D]eveloping and deploying weak AI involves making consequential choices—choices that demand greater democratic oversight not just from AI developers and designers, but from all members of society." Along a similar reasoning line, Wong (2020) argued that deciding what it means to be algorithmic fairness includes a crucial political dimension and should appeal to democratic mechanisms.

Currently, the power of decision-making surrounding AI is highly clustered in the hands of tech companies that develop AI systems. Many tech companies have proposed some forms of AI ethics principles and have suggested that they act accordingly. Even though some of these internal regulations are needed, they are insufficient. Notably, concerns have been raised regarding the quality of these principles, revealing that they tend to be abstract, lack theoretical ground, and fail to provide concrete guidance for practice (Hagendorff 2020; Floridi 2019). This situation thus raises the concern that these companies are merely virtue-signaling but are not taking the potential risks that AI may impose on society seriously. Moreover, as compliance with AI ethics principles is based on voluntary grounds and lacks external enforcement, the general public lacks mechanisms to hold these companies accountable when things go wrong.

The call to democratize AI governance aims to rectify the current situation, and it bears the prospect of mitigating algorithmic injustice to some extent. Introducing democratic mechanisms into the governance surrounding AI enables more checks and balances and avoids the unilateral power of decision-making (Seger et al. 2023). According to the all-affected principle, those who will be affected by certain decisions should have some power to influence decision-making (Goodin 2007). As AI bears great power in shaping the fundamental socio-technical structure in modern society and even the global world, the reasoning thus suggests that *almost everyone* should be given some power to influence decision-making. While it is extremely difficult to involve everyone who is affected by AI in the collective decision-making process, the all-affected principle might still work as an aspiration for exploring practical measures that allow more people to offer their opinions. From this perspective, democratic processes through which members of society can, in certain forms, participate in the decision-making process of AI regulations seem attractive.

Another attractiveness of democratizing AI governance in reducing algorithmic injustice can be seen in considering the epistemic value of democracy, according to which democratic decision-making is generally more reliable than alternative methods to produce correct outcomes (Anderson 2006; Landemore 2020). Many questions regarding AI governance are complicated, and by drawing on the cognitive diversity of the people, there is a good chance that cognitively diverse groups can come up with more comprehensive principles that concern various forms of risks that AI might impose on society and different groups of people compared with those currently proposed by tech companies.

There is no doubt that “democratizing AI governance” is a general term, and, like the notions examined above, there might be various different ways to envision its implementations. Below, I briefly examine three popular types of strategies for democratizing AI governance that have been proposed: (1) through existing democratic institutions in the government sector, (2) through direct participation, and (3) through representative deliberation.

The first approach to implementing democratic governance on AI appeals to existing democratic institutions and mechanisms, such as legislation and regulation standards, for deciding on and enacting the relevant regulations. Although it is somewhat slow, some state and global governing bodies are taking initiatives in this regard. For example, the European Union (EU) has been working on the AI Act, targeted for implementation in the next few years.[8] The government of the United States (US) released the blueprint for an AI Bill of Rights at the end of 2022[9] and issued the Executive Order on Safety, Secure, and Trustworthy Artificial Intelligence at the end of 2023.[10] At the international level, the Global Partnership on Artificial Intelligence (GPAI)[11], hosted by the Organization for Economic Co-operation and Development (OECD), seeks to promote intergovernmental cooperation in advancing the appropriate use of AI, and it currently has nearly 30 member countries. With some of the legally binding regulations put in place, there is hope that instances of algorithmic injustice will be reduced. As various legal regulations are still in the process of enacting them, more examinations should follow.

However, relying primarily on the state and global government to enact regulations on AI does not always seem desirable. Notably, fast-emerging AI technologies pose challenges for lawmakers and policymakers. To enact suitable and effective regulations, these stakeholders need to have a good understanding of the nature of the new technologies, but this criterion is hard to meet, given the fast-paced innovation and the lack of expertise. Furthermore, while democratic representation through elections in an ideal world tends to reflect the perspectives of those who are represented, in our non-ideal world, concerns should be raised regarding how existing power dynamics might influence the real result. For example, research has revealed that US policy faces the serious concern of regulatory capture; namely, the regulator is co-opted to prioritize the interests of minor constituencies, such as industry and those with the highest income, but is not responsive to the interests of the working classes and middle classes (Gilens 2012; Hacker and Pierson 2010). There is thus a serious concern about how effective this approach may be in mitigating the concern of algorithmic injustice and ensuring that the interests of the marginalized are respected.

The second proposal for democratizing AI governance suggests that AI governance should appeal to novel or reformed existing institutions, which operate through more direct, bottom-up, and inclusive participation of multi-stakeholders. For example, in considering suitable responses to algorithmic injustice, Zimmerman et al. (2020) indicated that “Broaching questions of algorithmic justice via the democratic process would give members of communities most impacted by algorithmic bias more direct democratic power over crucial decisions concerning weak AI—not merely after its deployment, but also at the design stage.” They used San Francisco’s Stop Secret Surveillance Ordinance as an example of incorporating

[8] For more information on the AI Act proposed by the EU, see <https://artificialintelligenceact.eu/>
[9] See <https://www.whitehouse.gov/ostp/ai-bill-of-rights/> for more details.
[10] See <https://www.whitehouse.gov/briefing-room/presidential-actions/2023/10/30/executive-order-on-the-safe-secure-and-trustworthy-development-and-use-of-artificial-intelligence/> for more details.
[11] See <https://gpai.ai/> for more details.

bottom-up democratic procedures into the governance of AI. Considering various cases of algorithmic bias, the ordinance banned the use of facial recognition tools in policing and stated, "Decisions regarding if and how surveillance technologies should be funded, acquired, or used, and whether data from such technologies should be shared, should be made only after meaningful public input has been solicited and given significant weight."[12]

The general spirit of appealing to procedures that are more direct, bottom-up, and inclusive seems to be pointing in the right direction in responding to the concern of power asymmetries; nonetheless, the emphasis on broad and direct participation of the general public has attracted some concerns. Himmelreich (2022) argued that the proposal of governing AI through direct and broad democratic participation is both theoretical and practically insufficient in mitigating algorithmic injustice. There, Himmelreich based his argument mainly on *minimalist* and *aggregative* conceptions of democracy: The former sees democracy as mostly about competitive elections, and the latter understands democracy as mostly about a process of fair aggregation. If we understand democratizing AI governance in these terms, then it means that in making decisions about conflicts of values or interests, we should appeal to processes such as voting to make the final call. As these two conceptions of democracy are largely procedural, only concerned with formal equality and agnostic about moral views, Himmelreich (2022) argued that they are theoretically insufficient in addressing algorithmic injustice. Worse, the majority rule that is often used in these models might function to exclude the voices of the marginalized and, in this way, further support existing power asymmetries and exacerbate algorithmic injustice.

Recently, there has been a third type of proposal for democratizing AI governance that emphasizes representative deliberation, especially through some forms of political mechanisms named *deliberative mini-publics*. Deliberative mini-publics are a kind of innovative mechanism for implementing democratic deliberation, where randomly selected populations are convened to learn about related information, deliberate with other participants on issues of public concern, and derive results that can be used as guidance for public decisions (Dahl 2008; Escobar and Elstub 2017).[13] Over the past few years, deliberative mini-publics have been used by different sectors to solicit democratic input for AI governance. For example, in 2019, the National Institute for Health Research in the UK convened two citizens' juries (a form of deliberative mini-publics) to deliberate on how to make a trade-off between the accuracy and explainability of AI systems (van der Veer et al. 2021). In 2020, the Ada Lovelace Institute held the Citizens' Biometrics Council to explore people's attitudes toward the use of biometrics technologies, such as facial recognition and digital fingerprinting. More recently, at the end of 2022, in collaboration with Stanford's Deliberative Democracy Lab, Meta launched a series of Community Forums following the design of deliberative polling (another form of deliberative mini-publics) to get feedback from diverse groups of people worldwide regarding the regulation of metaverse and generative AI (Clegg and Global Affairs 2023).

[12] See <https://www.eff.org/document/stop-secret-surveillance-ordinance-05062019> for the full document.

[13] Depending on the forms of mini-publics, the results produced might be different. For example, the result derived from deliberative polls (which incorporate larger groups of participants and fewer meetings) would be the survey of the opinions of the represented after deliberation, while in citizens' juries or citizens' assemblies (which utilize relatively smaller groups but a more extended series of meetings), the results would tend to be more comprehensive reports or recommendations.

As a form of innovative mechanism, deliberative mini-publics provide a promising way to implement the spirit of direct democracy while presenting some structures to allow smaller but representative groups to engage in more in-depth deliberation. With the guidance of moderators in the deliberative process, deliberative mini-publics show signs of reducing the impact of power imbalances in the process (Siu 2017). Moreover, studies have provided support for their potential to overcome individual biases and produce more comprehensive solutions to social injustices (Karpowitz, Raphael, and Hammond 2009; Luskin et al. 2014; Fishkin 2009; Grönlund, Setälä, and Herne 2010; Niemeyer 2011), which reveals the potential for appealing to deliberative mini-publics as a platform for coordinating collective efforts to address structural injustices, including but going beyond algorithmic injustice.[14] Still, the reflections from a socio-technical perspective suggest that continuous attention should be paid to the details of implementing deliberative mini-publics. For example, at what stages of AI development should mini-publics be used to intervene? Who should decide what kinds of questions mini-publics should discuss? Furthermore, questions should be raised regarding how the outcome of deliberative mini-publics will be used in practice: Will the government and tech companies follow the suggestions delivered by the mini-publics, or will they be treated as mere recommendations and followed only when the results go along with their own agendas?

Overall, the reflections on democratizing AI governance reveal that while appealing to democratic processes seems desirable in reducing the power concentration of unilateral decision-making, many democratic mechanisms themselves still face the challenges of power imbalances, and it is not clear that models of democracy are all well-equipped in response. For example, the electoral representative democracy that exists in many contemporary worlds often leads to regulatory capture, serving the interests of the powerful rather than the public. The aggregative model of democracy, which regards democracy as largely a formal procedure of aggregation, tends to endorse majority rule, which risks marginalizing the perspectives of minorities. The appeal to some innovative mechanisms, such as deliberative mini-publics, while showing promise, also raises questions for examination.

6. Avoid "democracay washing": reflections and practical implications for democratizing AI

Even though democratizing AI is often framed as a desirable response to mitigate algorithmic injustice, my analysis from Sections 3 to 5 on three forms of democratizing AI—democratizing AI use, democratizing AI development, and democratizing AI governance—presents a more nuanced picture. Some versions of each of the three notions examined bear prospects in mitigating the concern of algorithmic injustice, while others are either limited in bringing positive change or may even function to perpetuate various unjustified power hierarchies in the current socio-technical structure. This analysis does not mean to dismiss the movement surrounding democratizing AI but rather aims to urge a more fine-grained discussion on democratizing AI. By laying out different notions and measures of democratizing AI and considering their potential impacts on the socio-technical structure, we can have more constructive discussions on identifying the conditions under which democratizing AI would be desirable and explore suitable (both existing and potential) measures to achieve such a goal. To this end, I want to highlight two practical implications. First, I raise caution against a danger I will refer to as "democracy washing," namely when the label "democratizing AI" is used in an overly general manner and may function to block needed

[14] I provided a more general discussion on the moral ground and practical feasibility of using deliberative mini-publics to address structural injustice in (reference redacted).

scrutiny on the associated practices and their impacts. Second, I suggest paying closer attention to the power dynamics embedded in the socio-technical structure to help guide the explorations surrounding democratizing AI.

First, as the label "democratizing AI" has been popularized by many big tech companies and used to describe a variety of activities, concern should be raised regarding whether this slogan has functioned to wash away closer scrutiny of detailed practices. As mentioned in Section 4, in the discussions on participatory design, scholars have been using "participation washing" to warn against the trend of using participation in an overly general way without differentiating efforts that really bring about the positive transformation of power dynamics (such as community empowerment or self-determination) from those that fail to manifest such spirits (Sloane et al. 2020; Birhane et al. 2022). Relatedly, concerns have also been raised regarding "ethics washing" when tech companies publish ethics guidelines and talk about their commitment to ethics in order to wash away concerns about their practices (Wagner 2018; van Maanen 2022). Along similar lines, the analysis presented in Sections 3 to 5 reveals that the discussions surrounding democratizing AI might fall into similar traps, which I will refer to as "democracy washing." Notably, the label "democratizing AI," similar to "participation," has been used in an overly general way to characterize various measures but fails to distinguish their crucial differences in terms of their impacts on the socio-technical structure. Further, while tech companies are not the only sector that talks about democratizing AI, by describing many of their actions under this slogan, there is a tendency to portray what they are doing as absolutely laudable and thereby block closer scrutiny of the detailed practices and their impacts. With these two aspects of democracy washing go hand in hand, there is the danger that some practices that perpetuate unjust power hierarchies and exacerbate algorithmic injustice would be implemented under the seemingly desirable guise of "democratizing AI."

To avoid the danger of democracy washing in the pursuit of democratizing AI, I suggest that more attention should be paid to the power dynamics embedded in the socio-technical structure. Considering the overuse of "democratizing AI," some people might question whether some ways of talking about the democratization of AI should be regarded as illegitimate. For example, Seger et al. (2023) suggest that "'AI Democratization" is a (mostly) unfortunate term" (720) as in many cases the label is used (including democratizing AI use and democratizing AI development) as "almost synonymously with 'increasing accessibility'" (720). As a result, they suggest to preserve the term "democratizing AI" to refers to the democratizing AI governance while using labels such as "broad accessibility" in other cases.

While I'm sympathetic to the observation about the overuse of this label, I want to suggest something a bit different.[15] While the multiple notions of democratizing AI that have been proposed under the current movement have made the usage elusive, one valuable aspect is that they together point out a variety of different ways that practices surrounding AI could shape the socio-technical structure, such as through its use, its development, and its governance. Like many other issues of structural injustice, there are many components involved in contributing to the maintenance of algorithmic injustice, and the measures to combat algorithmic injustice and pursue algorithmic justice would be multifaceted. From this perspective, I want to suggest that it is permissive to keep the labels of democratizing AI for these different notions (as long as making it clear which notion is referred to). On the other hand, the more important lesson from the structural-injustice analysis, I suggest, is to ensure that consideration of how democratizing

[15] Thanks to anonymous reviewers for pushing me to clarify this point.

AI would shape the power dynamics embedded in the socio-technical structure is made at the forefront. In other words, we should raise the question: *How might social practices surrounding democratizing AI—whether AI is regarded as a tool, as an area of technological innovation, or as a site under governance—help cultivate a more just socio-technical structure where the unjustified power imbalances are mitigated?*

Below, I discuss how approaching democratizing AI in this way can help guide the movement by refining suitable notions and practices; I will examine democratizing AI use, development, and governance in order. First, when thinking about democratizing AI use, instead of simply suggesting that more users are always better, case-by-case examinations of the scope of access to AI systems with accompanying strategies on malicious use and unintentional misuse are needed (Solaiman 2023). Furthermore, in cases where increasing access to AI tools is beneficial for mitigating unjustified power imbalances, attention should be paid to the different challenges that potential users might face. Here, I would like to propose a distinction between *formal access* versus *substantive access* to AI tools, where the former means that there are few or no external restrictions (e.g., cost, permission requirement) that prevent one from accessing using AI tools, while the latter means that people have sufficient capacities to access and use AI tools. Currently, the dominant measures taken to democratize AI use focus on increasing formal access to AI, such as reducing the cost of using or running AI tools. While the removal of external restrictions on AI systems is crucial, it is insufficient to achieve the benefits of democratizing AI use, such as reducing digital divides. For example, people who bear a more skeptical attitude toward new technologies or those who lack digital literacy might still not use AI systems proactively, even if those tools and services are freely available. In those scenarios, efforts should also be made to address the barriers to substantive access, such as providing education to foster AI literacy and equip people to use and critically evaluate AI (Long and Magerko 2020).

Then, in thinking about democratizing AI development, in addition to including more diverse people in the AI development process—either by contributing to the technical or non-technical aspects of AI development—efforts should also be made to ensure that diverse perspectives are properly represented and taken into account. This suggestion reiterates the need to distinguish between different forms of participation (Arnstein 1969; Sloane et al. 2020; Birhane et al. 2022). Additionally, when pre-trained models or components of models (such as datasets) on which the broader community can build AI tools are made available, caution should be taken to examine the value encoded in these components before using them to build AI systems. For example, many voice datasets owned by big tech companies to build automated speech recognition (ASR) systems underrepresent marginalized groups (e.g., non-English speakers, people of color, etc.), contributing to these ASR systems' permanence biases toward some demographic attributes, including gender, age, accent, and speech rate. In this case, simply making these datasets available for a wider population to develop ASR systems would not be ideal; worse, with more people developing AI systems based on such skewed datasets, the unjust impacts would risk being enlarged. In responding to such a situation, Mozilla's Common Voice project provides a better example. Concerning the lack of representativeness of large voice datasets for building speech recognition systems, Mozilla initiated the Common Voice project—a crowdsourcing project that aims to mobilize people around the world to build a publicly available voice dataset where people can contribute their voices and help validate the voice data (Ardila et al. 2019). Since the project's launch in 2017, its dataset has now included more than a hundred languages, including some low-resource ones. By inviting the general public to participate in this project, Common Voice not only makes the development of speech technology more decentralized but also raises a critical eye on the biases embedded in existing speech recognition systems.

Lastly, the discussions on democratizing AI governance would not be complete without considering the impacts that different democratic conceptions, processes, and mechanisms would have on the socio-technical structure. For example, when adopting aggregative democracy for decision-making, we should be wary of the concern of the majority tyranny and persistent minorities; when appealing to electoral representative democracy, we should pay attention to the potential of regulatory capture or other forms of corruption that prevent the interests of the public from being properly taken into account. One might think that this implication would make the task of democratizing AI governance too daunting; after all, these are problems of the democratic process and not issues specific to the governance of AI. Although I agree that these are not issues unique to the governance of AI, as long as we are serious about putting AI-related decision-making into democratic oversight, it is hard to see how we can make such suggestions responsibly without making efforts to mitigate the associated risks, such as enlarging power hierarchies. Moreover, the fact that these issues are not unique to the governance of AI suggests that explorations on how to democratize AI governance can build on previous studies to refine democratic processes. The recent efforts to adopt models of deliberative mini-publics for AI governance (as described in Section 5), although not without its own limitations, is an example of this sort. From this perspective, the question of how to implement the democratic governance of AI may even serve as a helpful lens for re-envisioning the democratic governance of a variety of public issues.

7. Conclusion

In this paper, I presented a socio-technical understanding of algorithmic injustice and examined how three different notions of "democratizing AI"—democratizing AI use, democratizing AI development, and democratizing AI governance—respond to it. Building on Iris M. Young's theory on social structure, I argued that algorithmic injustice can be understood as a case of structural injustice that exists when the socio-technical structure shaped by AI systemically exposes large groups of people to undeserved burdens while conferring unearned benefits to others, thereby exacerbating unjustified power hierarchies between people along various aces of social categories. From this perspective, algorithmic injustice concerns the unjustified power imbalances that the resulting socio-technical structure imposes on people situated in different social positions, and it calls for collective efforts among participants of the socio-technical structure to address it. Even though democratizing AI is often framed as a desirable response to algorithmic injustice, my analysis reveals there are crucial differences among them. For each of the three notions examined, while some versions bear prospects of mitigating algorithmic injustice, others might have the tendency to perpetuate it. Rather than suggesting a total dismissal of the quest to democratize AI, my analysis suggests that paying closer attention to the power dynamics embedded in the socio-technical structure would help avoid what I call "democracy washing" and guide more constructive explorations on how to democratize AI.

References

Adamson, Adewole S., and Avery Smith. 2018. "Machine Learning and Health Care Disparities in Dermatology." *JAMA Dermatology* 154 (11): 1247–48.

AI Index Steering Committee. 2023. "Artificial Intelligence Index Report 2023." Institute for Human-Centered AI, Stanford University. https://aiindex.stanford.edu/report/.

AI Now Institute. 2023. "The Climate Costs of Big Tech." AI Now Institute. April 11, 2023. https://ainowinstitute.org/spotlight/climate.

Allen, Bibb, Sheela Agarwal, Jayashree Kalpathy-Cramer, and Keith Dreyer. 2019. "Democratizing AI." *Journal of the American College of Radiology: JACR* 16 (7): 961–63.

Anderson, Elizabeth. 2006. "The Epistemology of Democracy." *Episteme; Rivista Critica Di Storia Delle Scienze Mediche E Biologiche* 3 (1-2): 8–22.

Ardila, Rosana, Megan Branson, Kelly Davis, Michael Henretty, Michael Kohler, Josh Meyer, Reuben Morais, Lindsay Saunders, Francis M. Tyers, and Gregor Weber. 2019. "Common Voice: A Massively-Multilingual Speech Corpus." *arXiv [cs.CL]*. arXiv. http://arxiv.org/abs/1912.06670.

Arnstein, Sherry R. 1969. "A Ladder of Citizen Participation." *Journal of the American Institute of Planners* 35 (4): 216–24.

Benjamin, Ruha. 2019. *Race After Technology: Abolitionist Tools for the New Jim Code*. John Wiley & Sons.

Bianchi, Federico, Pratyusha Kalluri, Esin Durmus, Faisal Ladhak, Myra Cheng, Debora Nozza, Tatsunori Hashimoto, Dan Jurafsky, James Zou, and Aylin Caliskan. 2023. "Easily Accessible Text-to-Image Generation Amplifies Demographic Stereotypes at Large Scale." In *Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency*, 1493–1504. FAccT '23. New York, NY, USA: Association for Computing Machinery.

Binns, Reuben. 2020. "On the Apparent Conflict between Individual and Group Fairness." In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, 514–24. FAT* '20. New York, NY, USA: Association for Computing Machinery.

Birhane, Abeba. 2021. "Algorithmic Injustice: A Relational Ethics Approach." *Patterns (New York, N.Y.)* 2 (2): 100205.

Birhane, Abeba, William Isaac, Vinodkumar Prabhakaran, Mark Diaz, 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*, 1–8. EAAMO '22 6. New York, NY, USA: Association for Computing Machinery.

Bommasani, Rishi, Kathleen A. Creel, Ananya Kumar, Dan Jurafsky, and Percy Liang. 2022. "Picking on the Same Person: Does Algorithmic Monoculture Lead to Outcome Homogenization?" *arXiv [cs.LG]*. arXiv. http://arxiv.org/abs/2211.13972.

Buolamwini, Joy, and Timnit Gebru. 2018. "Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification." In *Proceedings of the 1st Conference on Fairness, Accountability and Transparency*, edited by Sorelle A. Friedler and Christo Wilson, 81:77–91. Proceedings of Machine Learning Research. PMLR.

Chung, Ryoa. 2021. "Structural Health Vulnerability: Health Inequalities, Structural and Epistemic Injustice." *Journal of Social Philosophy* 52 (2): 201–16.

Clegg, Nick, and Global Affairs. 2023. "Bringing People Together to Inform Decision-Making on Generative AI." Meta. June 22, 2023. https://about.fb.com/news/2023/06/generative-ai-community-forum/.

Collins, Patricia Hill. 2002. *Black Feminist Thought: Knowledge, Consciousness, and the Politics of Empowerment*. Routledge.

Costanza-Chock, Sasha. 2018. "Design Justice, A.i., and Escape from the Matrix of Domination." *Journal of Design and Science*, July. https://doi.org/10.21428/96c8d426.

———. 2020. "Design Justice: Community-Led Practices to Build the Worlds We Need." https://library.oapen.org/handle/20.500.12657/43542.

Crawford, Kate. 2021. *The Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence*. Yale University Press.

Dahl, Robert A. 2008. *Democracy and Its Critics*. Yale University Press.

Davis, Jenny L., Apryl Williams, and Michael W. Yang. 2021. "Algorithmic Reparation." *Big Data & Society* 8 (2): 20539517211044808.

Delgado, Fernando, Stephen Yang, Michael Madaio, and Qian Yang. 2021. "Stakeholder Participation in AI: Beyond 'Add Diverse Stakeholders and Stir.'" *arXiv [cs.AI]*. arXiv. http://arxiv.org/abs/2111.01122.

Deng, Jianyang, and Yijia Lin. 2022. “The Benefits and Challenges of ChatGPT: An Overview.” *Frontiers in Computing and Intelligent Systems* 2 (2): 81–83.

Escobar, Oliver, and Stephen Elstub. 2017. “Forms of Mini-Publics.” newDemocracy Foundation.

Eveleth, Rose. 2014. “How Self-Tracking Apps Exclude Women.” *Atlantic* 15. http://www.leeannhunter.com/digital/wp-content/uploads/2015/08/Eveleth.SelfTrackingApps.pdf.

Fishkin, James. 2009. *When the People Speak: Deliberative Democracy and Public Consultation*. OUP Oxford.

Floridi, Luciano. 2019. “Translating Principles into Practices of Digital Ethics: Five Risks of Being Unethical.” *Philosophy & Technology* 32 (2): 185–93.

Gabriel, Iason. 2022. “Toward a Theory of Justice for Artificial Intelligence.” *Daedalus* 151 (2): 218–31.

Gebru, Timnit. 2020. “Race and Gender.” In *The Oxford Handbook of Ethics of AI*, edited by Markus D. Dubber, Frank Pasquale, and Sunit Das, 253–70. Oxford University Press.

Gilens, Martin. 2012. *Affluence and Influence*. Princeton University Press.

Gochfeld, Michael, and Joanna Burger. 2011. “Disproportionate Exposures in Environmental Justice and Other Populations: The Importance of Outliers.” *American Journal of Public Health* 101 Suppl 1 (Suppl 1): S53–63.

Goodin, Robert E. 2007. “Enfranchising All Affected Interests, and Its Alternatives.” *Philosophy & Public Affairs* 35 (1): 40–68.

Green, Ben. 2022. “Escaping the Impossibility of Fairness: From Formal to Substantive Algorithmic Fairness.” *Philosophy & Technology* 35 (4): 90.

Grönlund, Kimmo, Maija Setälä, and Kaisa Herne. 2010. “Deliberation and Civic Virtue: Lessons from a Citizen Deliberation Experiment.” *European Political Science Review* 2 (1): 95–117.

Groves, Lara, Aidan Peppin, Andrew Strait, and Jenny Brennan. 2023. “Going Public: The Role of Public Participation Approaches in Commercial AI Labs.” In *Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency*, 1162–73. FAccT ’23. New York, NY, USA: Association for Computing Machinery.

Gulshan, Varun, Lily Peng, Marc Coram, Martin C. Stumpe, Derek Wu, Arunachalam Narayanaswamy, Subhashini Venugopalan, et al. 2016. “Development and Validation of a Deep Learning Algorithm for Detection of Diabetic Retinopathy in Retinal Fundus Photographs.” *JAMA: The Journal of the American Medical Association* 316 (22): 2402–10.

Hacker, Jacob S., and Paul Pierson. 2010. *Winner-Take-All Politics: How Washington Made the Rich Richer--and Turned Its Back on the Middle Class*. Simon and Schuster.

Hagendorff, Thilo. 2020. “The Ethics of AI Ethics: An Evaluation of Guidelines.” *Minds and Machines* 30 (1): 99–120.

Hamidi, Foad, Morgan Klaus Scheuerman, and Stacy M. Branham. 2018. “Gender Recognition or Gender Reductionism? The Social Implications of Embedded Gender Recognition Systems.” In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, 1–13. CHI ’18. New York, NY, USA: Association for Computing Machinery.

Harding, Sandra. 1992. “RETHINKING STANDPOINT EPISTEMOLOGY: WHAT IS ‘STRONG OBJECTIVITY?’” *The Centennial Review* 36 (3): 437–70.

Hasbe, Sudhir, and Ryan Lippert. 2020. “Democratization of Machine Learning and Artificial Intelligence with Google Cloud,” November 16, 2020. https://cloud.google.com/blog/products/data-analytics/democratization-of-ml-and-ai-with-google-cloud.

Haslanger, Sally. 2016. “What Is a (social) Structural Explanation?” *Philosophical Studies* 173: 113–30.

Himmelreich, Johannes. 2021. “Should We Automate Democracy?” In . Oxford University Press.

———. 2022. “Against ‘democratizing AI.’” *AI & Society*, January. https://doi.org/10.1007/s00146-021-01357-z.

Hoffmann, Anna Lauren. 2019. “Where Fairness Fails: Data, Algorithms, and the Limits of Antidiscrimination Discourse.” *Information, Communication and Society* 22 (7): 900–915.

Jo, Eun Seo, 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, 306–16. FAT* '20. New York, NY, USA: Association for Computing Machinery.
Kalluri, Pratyusha. 2020. "Don't Ask If Artificial Intelligence Is Good or Fair, Ask How It Shifts Power." *Nature* 583 (7815): 169.
Karpowitz, Christopher F., Chad Raphael, and Allen S. Hammond. 2009. "Deliberative Democracy and Inequality: Two Cheers for Enclave Deliberation among the Disempowered." *Politics & Society* 37 (4): 576–615.
Kaushal, Amit, Russ Altman, and Curt Langlotz. 2020. "Geographic Distribution of US Cohorts Used to Train Deep Learning Algorithms." *JAMA: The Journal of the American Medical Association* 324 (12): 1212–13.
Keyes, Os. 2018. "The Misgendering Machines: Trans/HCI Implications of Automatic Gender Recognition." *Proc. ACM Hum.-Comput. Interact.*, 88, 2 (CSCW): 1–22.
Kleinberg, Jon, and Manish Raghavan. 2021. "Algorithmic Monoculture and Social Welfare." *Proceedings of the National Academy of Sciences of the United States of America* 118 (22). https://doi.org/10.1073/pnas.2018340118.
Koplin, Julian J. 2023. "Dual-Use Implications of AI Text Generation." *Ethics and Information Technology* 25 (2): 32.
Landemore, Hélène. 2020. *Open Democracy: Reinventing Popular Rule for the Twenty-First Century*. Princeton University Press.
———. 2024. "Can Artificial Intelligence Bring Deliberation to the Masses?" In *Conversations in Philosophy, Law, and Politics*, edited by Ruth Chang and Amia Srinivasan. Oxford University Press.
Le Bui, Matthew, and Safiya Umoja Noble. 2020. "We're Missing a Moral Framework of Justice in Artificial Intelligence." *The Oxford Handbook of Ethics of AI*, 163–79.
Lin, Ting-An. 2024. "Sexual Violence and Two Types of Moral Wrongs." *Hypatia: A Journal of Feminist Philosophy*.
Lin, Ting-An, and Po-Hsuan Cameron Chen. 2022. "Artificial Intelligence in a Structurally Unjust Society." *Feminist Philosophy Quarterly* 8 (3/4).
Long, Duri, and Brian Magerko. 2020. "What Is AI Literacy? Competencies and Design Considerations." In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, 1–16. CHI '20. New York, NY, USA: Association for Computing Machinery.
Lu, Catherine. 2017. *Justice and Reconciliation in World Politics*. Cambridge University Press.
Luskin, Robert C., Ian O'Flynn, James S. Fishkin, and David Russell. 2014. "Deliberating across Deep Divides." *Political Studies* 62 (1): 116–35.
Maanen, Gijs van. 2022. "AI Ethics, Ethics Washing, and the Need to Politicize Data Ethics." *Digital Society : Ethics, Socio-Legal and Governance of Digital Technology* 1 (2): 9.
Maas, Jonne. 2023. "Machine Learning and Power Relations." *AI & Society* 38 (4): 1493–1500.
Manuel, Jennifer I. 2018. "Racial/Ethnic and Gender Disparities in Health Care Use and Access." *Health Services Research* 53 (3): 1407–29.
McKeown, Maeve. 2016. "Global Structural Exploitation: Towards an Intersectional Definition." *Global Justice: Theory Practice Rhetoric* 9 (2).
Meta AI. 2022. "Democratizing Access to Large-Scale Language Models with OPT-175B," May 23, 2022. https://ai.meta.com/blog/democratizing-access-to-large-scale-language-models-with-opt-175b/.
Microsoft News Center. 2016. "Democratizing AI: For Every Person and Every Organization," September 26, 2016. https://news.microsoft.com/features/democratizing-ai/.
Mikhaylovskaya, Anna. 2024. "Enhancing Deliberation with Digital Democratic Innovations." *Philosophy & Technology* 37 (1): 3.
Mohamed, Shakir, Marie-Therese Png, and William Isaac. 2020. "Decolonial AI: Decolonial Theory as Sociotechnical Foresight in Artificial Intelligence." *Philosophy & Technology* 33 (4): 659–84.
Narayanan, Arvind. n.d. "Translation Tutorial: 21 Fairness Definitions and Their Politics." Accessed October 3, 2023. https://facctconference.org/static/tutorials/narayanan-21defs18.pdf.
Neely, Megan Tobias, Patrick Sheehan, and Christine L. Williams. 2023. "Social Inequality in High Tech:

How Gender, Race, and Ethnicity Structure the World's Most Powerful Industry." *Annual Review of Sociology* 49 (1): 319–38.
Nelson, Alan. 2002. "Unequal Treatment: Confronting Racial and Ethnic Disparities in Health Care." *Journal of the National Medical Association* 94 (8): 666–68.
Niemeyer, Simon. 2011. "The Emancipatory Effect of Deliberation: Empirical Lessons from Mini-Publics." *Politics & Society* 39 (1): 103–40.
Noble, Safiya Umoja. 2018. *Algorithms of Oppression: How Search Engines Reinforce Racism*. NYU Press.
Noorman, Merel, and Tsjalling Swierstra. 2023. "Democratizing AI from a Sociotechnical Perspective." *Minds and Machines* 33 (4): 563–86.
Nordgren, Anders. 2022. "Artificial Intelligence and Climate Change: Ethical Issues." *Journal of Information, Communication and Ethics in Society* 21 (1): 1–15.
Obermeyer, Ziad, Brian Powers, Christine Vogeli, and Sendhil Mullainathan. 2019. "Dissecting Racial Bias in an Algorithm Used to Manage the Health of Populations." *Science* 366 (6464): 447–53.
Ragnedda, Massimo, and Glenn W. Muschert. 2013. *The Digital Divide: The Internet and Social Inequality in International Perspective*. Routledge.
Rahwan, Iyad. 2018. "Society-in-the-Loop: Programming the Algorithmic Social Contract." *Ethics and Information Technology* 20 (1): 5–14.
Rini, Regina, and Leah Cohen. 2022. "Deepfakes, Deep Harms." *Journal of Ethics and Social Philosophy* 22 (2). https://doi.org/10.26556/jesp.v22i2.1628.
Rubeis, Giovanni, Keerthi Dubbala, and Ingrid Metzler. 2022. "'Democratizing' Artificial Intelligence in Medicine and Healthcare: Mapping the Uses of an Elusive Term." *Frontiers in Genetics* 13. https://doi.org/10.3389/fgene.2022.902542.
Sætra, Henrik Skaug, Harald Borgebund, and Mark Coeckelbergh. 2022. "Avoid Diluting Democracy by Algorithms." *Nature Machine Intelligence* 4 (10): 804–6.
Seger, Elizabeth, Aviv Ovadya, Ben Garfinkel, Divya Siddarth, and Allan Dafoe. 2023. "Democratising AI: Multiple Meanings, Goals, and Methods." In . New York, NY, USA: ACM. https://doi.org/10.1145/3600211.3604693.
Shankar, Shreya, Yoni Halpern, Eric Breck, James Atwood, Jimbo Wilson, and D. Sculley. 2017. "No Classification without Representation: Assessing Geodiversity Issues in Open Data Sets for the Developing World." *arXiv [stat.ML]*. arXiv. http://arxiv.org/abs/1711.08536.
Siu, Alice. 2017. "Deliberation & the Challenge of Inequality." *Daedalus* 146 (3): 119–28.
Sloane, Mona, Emanuel Moss, Olaitan Awomolo, and Laura Forlano. 2020. "Participation Is Not a Design Fix for Machine Learning." *arXiv [cs.CY]*. arXiv. http://arxiv.org/abs/2007.02423.
Smith, Peter, and Laura Smith. 2021. "Artificial Intelligence and Disability: Too Much Promise, yet Too Little Substance?" *AI and Ethics* 1 (1): 81–86.
Solaiman, Irene. 2023. "The Gradient of Generative AI Release: Methods and Considerations." In *Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency*, 111–22. FAccT '23. New York, NY, USA: Association for Computing Machinery.
Strubell, Emma, Ananya Ganesh, and Andrew McCallum. 2020. "Energy and Policy Considerations for Modern Deep Learning Research." *Proceedings of the AAAI Conference on Artificial Intelligence* 34 (09): 13693–96.
Veer, Sabine N. van der, Lisa Riste, Sudeh Cheraghi-Sohi, Denham L. Phipps, Mary P. Tully, Kyle Bozentko, Sarah Atwood, et al. 2021. "Trading off Accuracy and Explainability in AI Decision-Making: Findings from 2 Citizens' Juries." *Journal of the American Medical Informatics Association: JAMIA* 28 (10): 2128–38.
Wagner, Ben. 2018. *Ethics as an Escape from Regulation. From "ethics-Washing" to Ethics-Shopping?* Philipps-Universität Marburg.
Wolf, Christine T. 2020. "Democratizing AI? Experience and Accessibility in the Age of Artificial Intelligence." *XRDS* 26 (4): 12–15.
Wong, Pak-Hang. 2020. "Democratizing Algorithmic Fairness." *Philosophy & Technology* 33 (2): 225–

44.

Wylie, Alison. 2003. "Why Standpoint Matters." In *Science and Other Cultures: Issues in Philosophies of Science and Technology*, edited by Robert Figueroa and Sandra G. Harding, 26–48. Routledge, New York.

Young, Iris Marion. 2011. *Responsibility for Justice*. Oxford University Press.

Zimmermann, Annette, Elena Di Rossa, and Hochan Kim. 2020. "Technology Can't Fix Algorithmic Injustice." *Boston Review*, January 9, 2020. https://www.bostonreview.net/articles/annette-zimmermann-algorithmic-political/.

Zuboff, Shoshana. 2019. *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*. PublicAffairs.