

Beyond Individual Accountability: (Re-)Asserting Democratic Control of AI

Daniel J. Bogiatzis-Gibbons

danjamesgibbons@gmail.com

Birkbeck College, University of London

London, UK

ABSTRACT

AI control mechanisms like accountability procedures or technical standards are usually subpolitical: decisions are primarily debated and made within circumscribed subsystems of experts or interest groups, like the professional community of data scientists. However, AI systems are more deeply intertwined with a wider sense of politics than these mechanisms contemplate. In Winner's dual senses, they are incidentally political as they settle disputes within political communities through their design, invention, and arrangement, and inherently political as they reciprocally contribute to and are sustained by patterning of economic, social, and political orders. This work, therefore, draws upon political theory to argue for democratically controlled AI beyond individual notions of accountability. In its weaker form, it demands substantive, rule-bound oversight of state actors' use of AI systems, seeking to remedy historical tendencies toward extra-legal surveillance and strengthen accountability beyond individuals. Conversely, the stronger form advocates for comprehensive democratic control over all facets of AI, even by questioning the permissibility of AI within particular socio-economic spheres, as these systems are becoming fundamental parts of our collective life. I sketch the necessary institutional frameworks to operationalize these two forms of democratic control: first, for the "weak" form through the concept of a "control" power separate from the executive from Sun Yat-Sen's political thought, and second, participatory institutions such as citizens' assemblies. Finally, I discuss actions data scientists can take without legal frameworks for control: furthering new social imaginaries of AI that foreground the possibility of control and involving affected communities in decision-making around AI systems. The concept of democratic control is then both a measuring stick for existing standards and legislation and a clarion call for future advocacy.

CCS CONCEPTS

• **Social and professional topics** → **Government technology policy; Governmental surveillance;** • **Computing methodologies** → **Artificial intelligence; Machine learning.**

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

FAccT '24, June 03–06, 2024, Rio de Janeiro, Brazil

© 2024 Copyright held by the owner/author(s). Publication rights licensed to ACM.

ACM ISBN 979-8-4007-0450-5/24/06

<https://doi.org/10.1145/3630106.3658541>

KEYWORDS

AI, Democratic Control, Accountability, Political Theory, Bureaucracy, Technology Policy

ACM Reference Format:

Daniel J. Bogiatzis-Gibbons. 2024. Beyond Individual Accountability: (Re-)Asserting Democratic Control of AI. In *The 2024 ACM Conference on Fairness, Accountability, and Transparency (FAccT '24)*, June 03–06, 2024, Rio de Janeiro, Brazil. ACM, New York, NY, USA, 11 pages. <https://doi.org/10.1145/3630106.3658541>

1 INTRODUCTION

Given the encroachment of AI systems into new domains of social, economic, and political life, there is a need for effective mechanisms to control their design, deployment, and use. Otherwise, software code, along with its embedded norms, acts as a form of social regulatory tool rather than being subordinate to democratic politics [39]. This is occurring everywhere from Kentucky to Telangana, where error-prone AI has caused significant issues with falsely flagging fraud within social welfare programmes [25, 63]. Existing proposals to bring AI back under social control have, for example, focused on responsibility for private vendors of AI systems [15] or individual decision-makers [59], or alternatively broadening out participation in AI development [7]. There are also many suggested or actual governance mechanisms to secure fairness, transparency, and accountability through technical standards, ethical codes of conduct, corporate controls through contracts and licensing, and most recently, the draft EU *AI Act* [26]. Individual and organizational accountability is, of course, essential – it incentivizes better outcomes and ensures, at the very least, that there is an agent to pin injustices on, drawing a distinct line between blameworthy mistakes (i.e. those due to human agency and choices) and moral bad luck. Likewise, participation is a laudable goal and fits with democratic principles around the moral requirement for consent in the processes that affect you, but it can be co-opted to fit corporate ends [7].

However, these proposals do not go far enough – I argue that AI needs to be brought under democratic control to broaden out value-laden discussions and decisions on fairness, accountability, and transparency from the venues of Beck's "subpolitics": politics within a subsystem of society. These include standards-setting bodies or amongst data scientists, academics, professionalized civil society groups, and even FAccT itself. These "subpolitical" domains are where differences of opinions are often solved using set methodologies, research questions, or standards of evidence, bound together by a delimited range of future employment opportunities [4]. But given their typical circumscription by qualifications or narrow kinds of expertise, they necessarily exclude a multiplicity of voices and

range of lived experience, try as they might. Further, most initiatives do not directly address decisions that are required to give the people a genuine voice regarding the deployment, development, and use of AI. For instance, they all exclude considering whether AI should be deployed in a given domain, like child protection, beyond very narrow categories of banned systems in the *AI Act*. I shall discuss "weak AI" systems in this paper, though the arguments are if anything, strengthened and the need for control accelerated in time by the potential presence of "strong AI" that might permanently preclude the possibility of control.

I first show that AI systems are both incidentally and inherently political and, therefore, proper objects of democratic control. I then turn to a "weak" notion of control, notably control over only government use of AI systems by a state's citizens through a transparent, rule-bound electoral order [53]. I show why even this notion is often not adhered to due to extra-legal surveillance, and I propose regulatory structures inspired by the idea of a separate "control" arm of the government [76]. I then substantiate a "strong" notion of control, drawing on proposals for democracy in our broad social and economic life, here because AI systems are becoming part of the fabric of our shared existence [30] and technology should be returned to its correct place as a set of means for achieving human flourishing rather than as an end-in-itself [71, 73]. Taken together, democratic control ought then to include regulation of consequential AI technologies and the ways those technologies are developed, deployed, used, and embedded in physical hardware. Throughout, I sketch how these notions might be implemented (the "how" of control), considering both institutional design and alternatives that data scientists might pursue in the face of political inaction. I will demonstrate, indeed, that neither current regulations nor accountability proposals meet either of the normative thresholds for democratic control, which the remainder of this paper will flesh out using the tools of political theory.

This paper adds to a growing literature on the democratization of AI, for example, through calls for citizen governance of Big Tech platforms [3] and algorithms broadly [10], for improving the democratic quality of administrative and executive elements of decision-making on AI [32], and finally calls for democratizing notions of algorithmic fairness [75]. I analyze what meaningful democratic control might mean in the context of the ways in which AI is truly political rather than just socio-technical systems. This differs from past scholarship as it focuses on whether "the people" writ large meaningfully control AI systems rather than just whether AI developers or users are held accountable and how we might assess and strengthen the extent of such control. In terms of relevance for policymakers and practitioners, there are three clear contributions:

- (1) democratic control serves as a valuable set of benchmarks for evaluating existing standards and legislation while concurrently serving as a compelling call for future advocacy,
- (2) that in the absence of meaningfully imposed control, practitioners should seek to avoid projects that have the potential to actively diminish control and
- (3) AI systems should be thought of as explicitly political, and that regulation, therefore, needs to enshrine meaningful input into where and if AI systems are used, not just enumerate

technical safeguard measures or devolve power to standards-setting bodies.

This paper is intended to spark debate on democratic control issues and interest within FAcCT about political theory, and as such I will not discuss in great detail the important related work done on how to run citizens' assemblies or other tools of democratic participation.

2 ARE AI SYSTEMS PROPER OBJECTS OF DEMOCRATIC CONTROL?

I argue here that we ought to consider AI systems worth controlling by democratic states because, in Langdon Winner's two senses, they are both incidentally and inherently political.

They are incidentally political as their invention, design, and arrangement give new opportunities to settle political issues within and across communities [72]. This can be direct; for example, even error-ridden AI can further austerity politics in social welfare systems because of the difficulty of appealing automated decisions [25]. More broadly, AI systems can constrain what can be reasoned about and consider how citizenship rights and economic opportunities are conditioned based on coming to "know" some households or individuals as riskier than others [11, 58]. AI systems also deepen the characterizing and othering of racialized individuals, as part of the "New Jim Code" [6], for instance, through the selective use of predictive systems according to racial stereotypes [1]. There is a particular future need to understand how fairness problems exist in the Global South [67], but given social contexts of ethnic, caste, and class discrimination across a range of countries, such American findings are likely to be replicated elsewhere. These processes can affect either distributive justice or "who gets what" through the material allocation of resources, or procedural justice, which is "having adequate and fair processes". However, holding the designers of systems accountable might be insufficient if the harm results indirectly through AI systems, diminishing the role of human expertise in irreversible ways. In Kentucky, even when an error-prone automated fraud detection system was withdrawn, the deskilling and automation of social work remained [25].

AI systems are also inherently political; that is, they both encourage the existence of and thrive in particular kinds of social, economic, environmental, and political orders [72]. AI-driven news algorithms can undermine conditions for self-rule in democracies by controlling the information environments within which citizens make decisions [35]. Large technology corporations may come to control who knows what and how that knowledge is distributed [78]. The use of AI systems by states in surveilling and making life-changing decisions about individuals in, for instance, criminal justice [2, 25, 77] and child protection contexts [21, 25, 36] can serve to centralize and often outsource levers of control of often vulnerable populations. As a particular problem for notions of accountability, the action of the AI system in aggregate might be at fault rather than the decisions of a single decision-maker, which can (at least in principle) be appealed. However, AI systems beyond states can also have reciprocal relationships with particular patterning of social orders. Training Large Language Models (LLMs) requires extensive computational power and so is more compatible with centralized, oligopolistic market structures [12].

Through their material embodiment in hardware, LLMs are also part of extractivist supply chains that contribute to climate change and ecosystem degradation [5, 14].

This essay adopts, in general, the democratic principle that those who participate in, use, or are affected by an activity ought to control it, sometimes called the "deliberative turn" in democratic theory. That is in recognition that humans are neither entirely atomistic nor entirely constituted into some collective whole like the nation or humanity, but rather exist in social and economic relationships to each other of varying intensity and causal connection [30]. Dryzek (a leading scholar of deliberative democracy), I think is right to suggest that defensible theories of democracy ought to be critical of established power and traditions (to be robust to that power or traditions solidifying), pluralistic (in recognition that voting is not the only way to exercise power nor a strong means of control), and dynamic to openness (given that circumstances change, such as the existence of certain classes of AI systems) [20].

However, there are four principal limitations to this principle that ought to be respected. In a departure from deliberative democrats, I claim any reasonable answer must respect the fact that states are the primary tool for achieving the conditions for democratic decision-making, including solidarity and establishing clear links between representatives and constituents [61], and are pragmatically the primary mechanism for enforcing laws. Indeed, nation-states are simply very unlikely to cede voting rights to non-citizens, and while theoretical proposals might be aspirational, they should not rely on very unlikely occurrences. Of course, in tension with this, the influence of AI, both technologically and as a powerful organizing idea, is inherently transnational, so there will need to be accommodations made for that (see Section 4.3). The number of decision-making bodies that result from choices of "the people" ought to be limited, given that democratic decision-making can be highly costly [32] and that the concept of an "AI system" is broad enough to encompass thousands of algorithms, many of them relatively unimportant to the social fabric even collectively. Finally, precisely delimiting exactly who is affected by any given activity and how much they are affected is difficult to discern and might rapidly change, leading to ever-changing notions of "the people". It is worth noting that these issues are not problems that are solvable in a static way but rather tensions that need to be dynamically managed, as I will discuss in the rest of this paper.

Still, some might object to the idea that AI systems ought to be controlled at all, and I will briefly address three objections. The first possible objection is that democratic control over such systems might stifle innovation, either harming economic growth or even the potential evolution of the human species according to the transhumanists (see e.g. [27]). In part, this can be resolved by clarifying the bounds of what control might mean (see Section 4.2); for now, I would say it importantly does not mean the public needs to be a guiding hand for every technical decision. More broadly, states have been able to balance the need for technological progress with human flourishing, and ought we not conceive of and debate at least some limits in an age of high technology [73]? The second possible objection is that new forms of democratic control would either detract from existing sources of legitimate democratic power (courts, parliaments, civil service bodies, executive Presidents, etc.) or simply not have the knowledge to exercise the control in a non-empty

way [32]. These problems are surmountable, as deliberative bodies often owe their legitimacy to a statutory mandate, have time- or purpose-limitations, and can have expert testimony (albeit this also requires the ability to have competing experts, or else the experts can capture "deliberative" democratic institutions). Finally, it might be viewed as overstepping the boundaries of legitimate state power, especially by political philosophers such as Nozick, who would centre property rights and simply have states be "nightwatchmen" who enforce such property rights [46]. The claim that private property should be absolute as it comes ultimately from individual labour is especially untenable for many AI systems. That is because AI systems: i) often include technologies that are jointly produced (i.e. not just by their owners), ii) often include open-source elements, ii) often include technologies ultimately derived from public subsidy, and finally, iv) are sometimes trained based on unjustly acquired historical datasets [25]. But more broadly, such theories ignore the ways in which human life is interconnected and not fully able to be captured by market dynamics, which is brought into particularly sharp relief by the kinds of effects of AI systems I have just discussed.

I now move on to argue for two thresholds for democratic control: a more minimal sense of control of state use of AI and a more maximalist sense of control of all AI systems.

3 "WEAK" DEMOCRATIC CONTROL

3.1 Why "weak" control?

In this first "weak" sense, I refer to the minimal width to that boundary of what and where it is controlled, notably the *object* of control is the direct development, deployment, or use of AI systems by nation-states and that is *conducted* through democratic procedures voted on by their citizens. To combat governments' tendencies towards extra-legal surveillance and using technological choices to implement political choices tacitly, I identify three essential characteristics of "weak" democratic control: 1) the use of AI systems by governments must be transparent, 2) according to validly enacted laws of which the electorate has at least influence through fair elections, and 3) it should be supervised by an authority independent of the executive.

The first tendency I seek to solve then is that of executive arms to conduct extra-legal surveillance of the citizenry, which has been present for most of post-war computing history [8, 56], but might be supercharged through facial recognition technology and AI-led signals processing. For instance, it is notable that both the Brazilian General Data Protection Law (LGPD) and the EU General Data Protection Regulation contain carve-outs for national and public security, national defence, and criminal prosecution, which alter the level of protections afforded in case of government access to their citizens' data [31]. Surveillance, if taken far enough, can undermine the idea of self-rule, the normative idea that governments ought to be chosen by those they serve, which most democratic theories view as needing to consist of more than just competition by political parties for power through elections [35]. Indeed, if in a two-party democracy, both parties permit or encourage extra-legal surveillance of a viewpoint or racial minority, then elections could be technically "free" while depriving the people of real choice.

The second tendency touched on in the previous section is to use ostensibly innocuous technological choices to implement significant political decisions through administrative processes. AI then can be a vehicle for the "new technocracy", which is less explicitly anti-democratic than previous iterations of technocracy, instead promoting the idea that technocrats merely help to achieve unambiguously good ends or the purposes of democratically-elected governments [24]. In the AI context, this involves embedding notions of risk through explicit quantification via predictive models and performance management via benchmarking human decision-making against algorithms. However, these are not politically neutral notions – they may serve to identify some families or even whole types of family structures as inherently risky, creating a self-sustaining cycle of interactions with government services [25], or delegitimize arguments for increased funding for support services.

3.2 Transparency

As a necessary condition to solve these tendencies, I argue that AI systems used by states must be transparent: at a minimum, citizens must know about their existence and capabilities, as well as other relevant information, to ensure they are answerable to the people. Disclosure could be handled through a registry of AI systems, such as those mandated by Chile [49], Connecticut [62], and the cities of Helsinki [48] and Amsterdam [47].

Why transparency? There are at least three major justifications for transparency in democracies, as summarised in [34]. First, to hold officials and governments accountable, citizens need to know enough salient facts to exercise their rights through the courts or the ballot box. While it may be possible for individuals to make use of this information directly, the volume and complexity of such disclosures often mean they are ineffective without "monitory democracy", a sufficiently rich set of civil society organizations to digest and consider the implications for citizens and help them hold government to account [37]. Second, states must foster open and candid communication for democracy to thrive more broadly, as a culture of secrecy might decay the legitimacy of the entire democratic order. This is particularly concerning given the low level of trust in AI systems [29] and the fact that users tend to continue to rely on technological systems without trusting them [17]. Finally, there is an argument from the right to know how decisions about one or one's fellow citizens are made, as that forms part of the legitimacy of democratic decision-making. This is especially concerning as AI explainability techniques as currently envisioned may not even provide the meaningful information on the logic of processing required by the EU's General Data Protection Regulation (GDPR) [22], let alone sufficiently rich information to be confident in an automated or AI-assisted decision.

What does meaningful transparency consist of? One aspect is clearly the AI system's existence, alongside truthful disclosure of its intended purpose/s, the provenance of the datasets it was built from (including how representative these are of the subject population), and its broad capabilities. These disclosures should be made with potential users in mind, including through understanding their contexts of use and the need to have relevant information to contest specific decisions or seek redress [45]. It would need to go beyond existing Freedom of Information laws in that this

disclosure should be proactive; otherwise, governments have clear incentives to hide potentially unpopular uses of AI systems. The purpose disclosure aspect is critical to highlight so that it is possible to contest whether a system is undergoing scope creep or could have other potential uses. By purpose, I mean something broader than what is called "intended uses" on model cards [43] because it refers to why a system was commissioned and deployed, not just where and how it is intended to be used. However, even this kind of disclosure does not reveal how effective a system is, how it is being deployed, or the effects of deploying the system on communities. To understand the impacts of AI systems, making transparency effective may necessitate imposing a legal requirement to generate new information, such as monitoring and evaluation of systems. A final aspect is understanding how predictions work and become embedded in decision-making, which is essential even with simpler matching algorithms [41]. Otherwise, for example, an algorithm that produces apparently racially fair predictions might be deployed unfairly and disproportionately, resulting in unfair outcomes. In general, transparency might even include the ability for outside organizations to audit the system, although this needs to be balanced against compromising system integrity.

That is not to say that transparency is an absolute requirement. It is unlikely to require the revelation of individual data (except to the affected individual/s themselves) due to the right to privacy. Further, national security interests might agitate for temporal restrictions on disclosing information to subjects on surveillance systems being used on them, though these are open to abuse and would need strong safeguards. That said, there is no reasonable democratic justification for "deep secrets" [54] like the non-disclosure of entire systems. First, while non-disclosure of a system might appear reasonable in the context of intelligence gathering, in many scenarios, foreign states are more likely to have the tools to become aware of a country's intelligence-gathering AI systems than that country's citizens. Second, history provides reasons to doubt the effectiveness of even the best procedural safeguards, though at least knowing the existence of a system means some possibility of redress. Third, non-disclosure of a system forecloses the ability to debate whether the entire exercise is legitimate and deprives the citizenry of the ability to ensure whether an AI system is being operated within the bounds provided by law.

Therefore, transparency of AI systems is an essential necessary (if not sufficient) condition for weak democratic control and one that requires at the least knowledge of those systems, though arguably significantly more information than that. However, how should that control then be exercised?

3.3 Rule-bound, Electoral Orders and Supervision

I claim two significant constraints need to be placed on governments' use of AI systems: the need to confine AI to a set of rules decided with at least the causal influence of elections and the need to divorce the power to enforce those rules from the executive.

In general, a rule-bound order is, following Holmes, one where, to a reasonable approximation, the government has to act according to democratically promulgated laws and equal treatment according to law [33]. In the context of AI, I claim that this requires AI systems

to be brought under the control of specific legislation, granting the power to use predictive technology in a specific instance or range of instances. That is, it should not be treated like merely purchasing word-processing software, which is incidental to a public body's mandate but rather only permissible with an express mandate. Naturally, there is a need for some proportionality to the impact and kind of AI encompassed by the need for specific authorizing legislation, but this is to ensure against the use of AI as a means of covertly achieving political goals without enacting new policies or laws.

How, then, ought those rules to be decided and enforced? One narrow possibility is regular, free elections ensure that there is at least some connection between the preferences of the people and the actions taken by their representatives. In this minimal model, control of AI might be exercised solely through elected officials, as might commonly be done by an elected President or subordinate officers appointed by that President, with oversight provided by parliamentary committees. On one level, this is a necessary part of democratic control, which, paired with transparency and the need for specific rules would allow some form of control over AI systems.

However, I claim that at least in enforcing rules, this should be separated as much as is legally feasible from the executive to ensure that the executive branch does not use AI without authorization or unfairly deploy AI. Sun Yat-Sen [76] developed an analogous idea of a "control" power, separate from the legislative and executive branches, which has its historical origins in the need to govern through extensive and geographically diffuse bureaucracies during imperial China. This control power is held in Sun's five-power model (adopted partially by Taiwan [9]) as a separate arm of government called the Control Yuan that can impeach and censure officials, conduct audits, and impose corrective measures. Sun was concerned primarily by the ineffectiveness of investing the censure and impeachment powers in the legislature because they might either be overused if the President and legislative majority came from different parties or else not utilized at all. The capacity for corrective measures might be especially important in an AI context, as it was intended to go beyond holding individuals to account to be able to change aspects of a policy held inconsistent with certain basic tenets of liberty. For AI, one might imagine a body being able to cease the application, restrict the deployment, or alter the development of particular AI systems within fixed parameters.

Now, major constitutional reform is too high an ask for democratic control of AI for states following the three-power model. However, an important point is that control of AI, given its relationship to the possibility of AI supercharging executive overreach, should be insulated from the control of that same executive power. Sun suggests that this control power should be elected [76] (a feature not maintained in Taiwan [9]), which could be one mechanism by which a regulatory agency could be insulated from an executive. In many cases, a more potentially legally feasible option might be an agency with a board appointed by the legislative arm of government, removable only for good cause, and which might be mandated to include meaningful representation from a cross-section of stakeholders.

3.4 The (in)sufficiency of current regulations

One use for such an ideal of democratic control is critically evaluating current regulatory frameworks and approaches, which I will briefly illustrate.

To begin with, what is the situation where transparency is required for meaningful control of government AI systems? Even the Helsinki [48] and Amsterdam registers [47] contain relatively scant information, missing for many systems, even basic quantification of accuracy or fairness. While the registers often disclose the purpose for which an AI system has been deployed, this information is not generally accompanied by consideration of unintended future uses or proof of the purpose being the one stated. The draft EU *AI Act* text does represent an important step towards meaningful transparency [26], requiring for certain high-risk systems transparency in a centralized open database many of the disclosure aspects discussed previously. Further, it requires that high-risk AI systems be designed and developed to ensure sufficient transparency, with an explicit link to the possibility of users and providers of those systems comprehending how they function. However, these may preempt more exacting national regulations on algorithmic transparency, appear to omit some questionable biometric practices, and do not apply to other "routine" uses of AI [68]. The situation elsewhere is much worse – for example, the UK White Paper contains no proposals for a register of government (or any) AI systems [19], and the transparency provision of Biden's Executive Order [55] only applies to specific AI use cases commissioned as a result of the *Advancing American AI Act*. While Freedom of Information (FOI) requests in some jurisdictions may encourage disclosure of some requests, these require at least the knowledge that a particular system exists and careful navigation of often complex legislative frameworks.

In terms of rules and regulatory independence, while the EU *AI Act* would ban specific applications of AI, it would allow many routine applications of AI by states to proceed with minimal control and might foreclose the possibility of more restrictive regulation by member states [68]. In other jurisdictions, governments may continue to rely on arguments around incidental powers to treat many applications of AI as part and parcel of everyday governance. The regulators that are the competent national bodies for enforcing the EU Act are also not required to be separate from the executive, unlike data protection bodies as mandated by the GDPR [68]. In the US, regulatory independence from the executive is arguably a less legally tractable problem because of Supreme Court decisions such as *Seila Law* [50].

Further, all of these regulations fail to allow for decisions on whether certain forms or purposes for AI should be subsidized or prioritized over others, or most broadly, whether we should consider predictive optimization presumptively illegitimate [25, 70]. Instead, they prioritize the notion of largely unchecked innovation and economic growth, even at the expense of the ability of democracies to rein in that innovation or its effects later. Even if accountability mechanisms like the necessity for a human-in-the-loop for all decisions were entrenched, this would not be sufficient for weak control. It would evade more diffuse, aggregate harms such as deskilling helping professions or changing the tone of political argumentation towards technocratic solutions to social problems.

Overall, "weak" democratic control of AI is indeed a high bar: it requires much more than any current system likely provides and should serve as a benchmark to compare with future regulations of governmental AI systems. However, there are reasons to doubt that even this "weak" notion is sufficiently broad, so now I discuss the "strong" forms of democratic control, which extend beyond the use of AI by states.

4 "STRONG" DEMOCRATIC CONTROL

4.1 Why "strong" control?

The notion of "strong" control involves widening the sense of democracy from "weak" control in two senses: the power over AI systems to include ostensibly private systems and the "demos" or sovereign people to include mechanisms for control by those affected by particular AI systems. Strong control constitutes a family of notions encompassing diverse potential institutional structures and acceptable breadth, but I delineate the key considerations here. Why, then, should we adopt this more controversial and broader notion of control?

At its core, to elaborate on Section 2, I adopt the argument that freedom should be seen as the exercise of human agency within the scope of developing ourselves within the context of social relationships, including through relationships of reciprocity [30]. That freedom to develop oneself within a social context on joint projects, taken together with the notion of equality, argues for representation in decision-making concerning human activities that one participates in, is affected by, or uses. Intuitively, then, limitations on the development or use of technology ought to be most justified when they are likely to impede the conditions for the self-development of persons, either by depriving them of a safe physical environment or the opportunity to participate in a political or social one as equals. Those limitations ought to be imposed by a suitably bounded collection of voting persons who represent the interests of those participating in, using, or who are affected by that technology. I also adopt the argument that we should put technology back in its place as a set of means rather than an end in itself, and therefore, the people ought to have a say over the purposes for which that technology is built and used. Why? As previously stated, our technological choices, rather than the mere fact of the existence of a particular technology, structure entire forms of life, rather than merely being "impacts" or "side effects" of some autonomous process [71, 73]. It may be that in doing so, a suitable collective chooses to allow a reasonable level of risk in order to allow for forms of pro-social innovation or simply because developing more advanced AI systems could be seen as a form of individual self-expression, a species of self-development, but this should not be an unthinking default.

AI systems then ought to be controlled because they fundamentally re-order the pace, qualitative type, and scope of economic and political power throughout our collective life and over the lives of specific affected groups. For instance, Amazon's tracking systems enable increasingly fine-grained control of human bodies in factory work. However, in doing so, they also isolate workers, discouraging [16] labour organizing, reinforcing though not predetermining an atomized, micro-targeted, and physically dangerous patterning of labour relations [44, 66]. One could instead imagine tracking

items rather than people or recalculating time rates to incorporate buffers for more humane working conditions [16]. What kind of power should be asserted to remake our way of life in this more human-centric way, and who exactly are "the people"?

4.2 Power

In terms of power, I claim that, in proportion to the power a system holds over our collective life, the purposes for which AI systems are used should be the central focus of democratic control as well as high-level constraints on their development and deployment, and then there should be specific, even on-site, controls for especially impactful systems. This should be, through some legally appropriate mechanism, binding. Otherwise, participation becomes either a normatively thin notion with little chance of achieving justice [32], or, at worst, a cynical way of disposing of citizen concerns about technology. Any control exercised here is not absolute: technological change proceeds in multiple directions that can be impossible for any individual or even whole institutions to anticipate, and sometimes control might need to be instituted retroactively. However, the impossibility of complete control is not a strong argument for choosing not to assert even partial control.

To structure further discussion of this issue, there are five senses for "control of technology": technical control (operation in practice like directing a robot), tactical control (who, how, and where technology may be used in a given place, i.e. a factory or a benefits office), design control (formal and technical constraints on development, for instance accuracy metrics), regulatory control (permitted and prohibited uses), and production control (what is allowed to be produced) [30]. While the constellation of elements that constitute strong control should vary with context, most reasonable regimes for strong control would involve regulatory and sometimes tactical control, but generally not direct technical, design, or production control. This likely would nest most of the control elements discussed in the "weak control" section, and indeed, the argument for transparency is likely stronger given the general lack of justifications for partial secrecy derived from national security considerations.

The critical element of control for AI systems broadly is regulatory control, such as input over whether AI is safe for child protection or incentivizing language models that do not produce harmful carbon emissions or create the risk of torrents of fake news. Regulatory control is crucial as production control is typically challenging to exercise over many AI systems. This is because the barriers to entry over building many classes of models are low, typically requiring just coding skills, a laptop, and an Internet connection. That is not to say these are not real barriers to participation in building AI; merely that it is a much lower bar than technologies the people can more easily impose production control upon, like vaccines, nuclear power, or foodstuffs. On a more localized level, tactical and regulatory control should enable the enactment of Davis' proposal for Amazon warehouses [16]. For instance, this could occur through adopting broad rules about tracking technologies on a societal level, supplemented by worker representation at a factory level to supervise tactical control mechanisms. Regulatory and tactical control may not be sufficient for some questions. The people may then exercise indirect control over design by choosing

to embed particular values in equalities legislation that might only be met by specific fairness metrics, as is arguably already true of EU anti-discrimination law [69]. Similarly, they might even exercise production control in limited circumstances, especially when AI is embodied in consequential and dangerous hardware, with, say, autonomous weapons systems.

There are worries that this power might not be possible for the people to exercise because, for example, certain groups might be less willing or able to participate, or even that the people generally might never have a "sophisticated" understanding of AI systems [74]. However, the people need not choose specific model accuracy metrics or training sets; they ought instead to be involved in questions around what values a system embodies, review claims from experts around the impacts of systems, and make binding rules on the control of such systems. They could be aided by an expert secretariat, with appropriate representation from a variety of disciplines, in order to ensure that technocrats do not amass power over such bodies by determining the course of the debate, as Wong et al. worry [74]. Alternatively, experts could be subject to adversarial questioning, as proposed by Pamuk [52], to engage with the values and interests guiding research. The willingness argument is more difficult to overcome and is subject to more local nuances about the wariness of formalized or state institutions. However, if sincere efforts are made to give power, not just a voice, to underrepresented groups, and there is investment in their participation, then some of those issues might be overcome. Further, there is a long history of participatory democracy being either proposed by political theorists (see e.g. [20], [57], [38]) and a particular interest in their use for climate issues (see e.g. [20], [23]), which can be drawn upon in thinking about the design of such institutions.

4.3 People and Participation

That said, even if it is possible to delimit what kind of power should be exercised, who ought to exercise it?

I now give one answer to how these conditions might be satisfied: an interlocking system of "vertical" domestic whole-of-society citizens' assemblies on the broad purposes and allowable uses of AI, international fora for dispute resolution, and citizens' assemblies on "horizontal" issues like the use of AI in children's social care or policing. I then briefly consider the primary status quo (standards-setting).

At a macro-level, there ought to be citizen-wide participation in selecting the general purposes for which AI should be permitted to be deployed, along with a general selection of values and norms it ought to embody. This aligns with a "condominium model" of democracy, where there is an entrenchment of agreed-upon aims and purposes for a democracy [53]. The "condominium model" contrasts a more minimal form where parties contest and non-violently transfer power through elections, which can lead to relatively weak forms of popular control over power. It ensures that democracy is normatively "thick", embodying a shared set of values, rather than merely a "thin" procedure for law-bound resolution of particular kinds of political disputes [32]. I prefer to leave the shape of this body open to further debate. However, I have in mind a citizens' assembly that might promulgate initial rules, followed by a successor body that would meet at regular intervals to consider

if technological developments necessitated reopening debate. One reason for not making debates over values continuous is that it would otherwise compete with national parliaments for legitimacy.

Given the international character of AI systems, this would require transnational fora. As these are likely to become difficult to make legally binding on states, they are not perfect substitutes for a cosmopolitan, international democracy. Further, the various national bodies on AI might produce conflicting visions of deploying these technologies, as governments have so far. However, it is worth recognizing the obvious truth that an international democracy is not likely either on the short timescale required to solve the problems already caused by AI, or indeed at all (especially given the continuance of non-democracies, the idea of nationalism and of states that do not exercise effective control of their territories). National bodies at least provide more meaningful control than presently exists, and explicit transnational fora could pressure states into compromise. National bodies also might be able to resolve issues with third-party states by, for instance, refusing to allow AI systems that contribute to climate change excessively or are only possible with exploitative mining practices.

On a "horizontal" level, descending from the high-level debate on values, there ought to be bodies tasked with deciding on the acceptability and rules around AI in specific high-importance sectors (ideally with the coverage chosen by the macro-level body), such as social care or news media within a given state. One might object that who is affected changes too often, but there are often intuitive answers, which may need contesting over time as technology and society evolve. For example, in aged care, this might include representatives for seniors, those who run care homes, the staff of those homes, and at-home carers. It may also include some representation for specialized AI companies that provide tools for aged care. To be fair, this process might require some oversight to ensure that new groups can contest their exclusion from particular "horizontal" bodies.

I prefer to leave how these bodies would be chosen and voting rules relatively open, but they ought to balance how much different groups are affected and the size of different groups. In terms of group size, it is important to consider whether this originates from historically discriminatory processes of narrowing citizenship [40]. It may also be necessary to balance the potentially more inclusive idea that those in these bodies would be randomly selected from the included groups with the need for more substantive representation by chosen individuals who might be seen to represent the group's views better. One final issue here is that some groups, especially children or specific groups of vulnerable adults, might be unable to fully participate in this kind of decision-making due to impaired capacity to do so. Any such body ought to carefully weigh having these groups participate as much as possible and having "virtual" representation through advocacy groups with their own agendas, interests, and values.

The setting of standards for AI is particularly central to the EU *AI Act*, with a presumption of conformity for systems certified as meeting set standards [26]. It is also a feature promoted by the UK White Paper [19] and US Executive Order [55]. However, standards-setting, while an important part potentially of the design process, is, as stated in the introduction, a creature of subpolitics [4]. It cannot meet the normative standard for strong control because it

does not attempt to have a democratically legitimate franchise, and it leaves value-laden choices open to experts and interest groups with their own narrow sets of interests. Indeed, standards-setting has been shown to, while appearing politically neutral, be riven by distributional conflicts between groups of companies [60] and involvement in international trade disputes [42].

Therefore, while delimiting "the people" is fraught practically and normatively, an interlocking series of citizen-led participatory institutions that derive power from nation-states, I suggest, is one route forward. Given the urgency of controlling AI systems, it seeks to trade off the feasibility of proposed changes within current political and economic orders with the need to have meaningful participation and power for affected persons over the influence of AI systems on their lives.

5 WHAT SHOULD DATA SCIENTISTS DO IN THE ABSENCE OF POLITICAL ACTION?

Securing either the weak or strong notions of democratic control that I have set out would, nonetheless, require substantial legal and political reforms. In their absence, I suggest two imperfect substitutes: first, data scientists should aim to approximate control by encouraging informal participation of those affected by their AI systems and refraining from building systems that impede control, and second, they should contribute to more accurate social imaginaries of AI in order to help create the conditions for democratic control.

5.1 Informal Control

One stop-gap measure if control is not yet imposed through the establishment of formal institutions is to approximate what that control might otherwise have looked like voluntarily, though such participatory AI initiatives often fall short of meaningful inclusion [7]. However, what would this need to be meaningful, and where might it obstruct material and legally binding democratic control? On one level, a partial alternative to participatory democracy is more informal involvement of affected stakeholders in developing and deploying AI systems, for example, through consultation, advisory councils, or informal incorporation of public values into decisions about the intended purposes of particular AI systems. More broadly, technology companies and data scientists ought to refrain from designing AI in domains that meet with widespread public disapproval (if this occurs) or are especially likely to have harmful, non-intended uses. That would be especially true of AI systems that might harm the social conditions that sustain democracy by, for instance, depriving persons of agency over the content they read [35, 78].

However, in taking these actions, data scientists should be mindful of structuring and communicating about these forms of "approximating control" in ways that show they are steps towards democratic control rather than replacements for control. There is a clear and present danger in creating the appearance of democratic control through the invocation of constitutional metaphors. A notable example of this is Meta's use of quasi-democratic language in setting up its Oversight Board by using the phrase "people" instead of customers in its bylaws (invoking "We, the People") or references to the board as a "supreme court". At the same time, the amount

of actual democratic participation in that body is minimal, and it lacks democratic legitimation. The problem is that this may stymie actual mechanisms for democratic control, as the argument would be that they are redundant and merely duplicate existing oversight [13]. That is particularly worrying because historical examples of where informal regulatory mechanisms have eventually become binding on participants occurred in localized situations with relatively equal participants (vs. transnational AI with especially in the LLM field few prominent players), similarities of interests (vs. highly divergent economic and social interests with AI systems), credible commitments (vs. voluntary, non-binding commitments by tech companies), and the availability of mutual monitoring (vs. low transparency of AI) [51].

Therefore, in the absence of legally binding democratic control, one substitute is approximating the regulations that a hypothetical control mechanism would impose. However, that should not occur through the appearance of democratic control but rather meaningful attempts at eliciting the participation of affected communities and reflectively considering what technologies ought to be built.

5.2 New AI Imaginaries

In the absence of political leadership, a second core task for data scientists as a profession is contributing to or amplifying more productive imaginaries of AI: that is, the ways we imagine and have expectations around AI systems, how they fit into our social fabric, and the ethical notions and narratives that undergird those expectations [64]. Taylor argues that even the possibility of modern democracy is based around social imaginaries like a public sphere where ideas are debated openly on an equal playing field or a formally equal, self-governing people [64]. Political elites respond to social imaginaries because they grow up embedded in and shaped by those ways of viewing the world and because they co-create the preferences and priorities of the voting public.

Our current social imaginary of AI is dominated by technofuturist ways of thinking, which privilege three sets of claims: that progress is inevitable, that AI is only present as disembodied, masterful algorithms rather than also materially manifested in both production and hardware, and that we should tend to view social problems as engineering problems [14]. The inevitability of AI progress continues a history of perceiving modern technological change as autonomously progressing without reference to or the desirability of any specific set of social purposes, except perhaps continued economic growth and higher productivity [71, 73]. On disembodiment, AI is imagined as producing new, even "alien" insights while depending on data gathered from poorly paid labour for classifying images and image taxonomies that encode eugenicist or racist assumptions [14]. Indeed, even the reasons that AI might outperform humans at certain kinds of predictive tasks – such as potentially being able to see more and a more comprehensive range of cases, and therefore suffer less from inferring from small numbers [65] – are often de-emphasized in favour of explanations around this "alien" quality of machine predictions. Social problems are often portrayed as solely due to a lack of information or human decision-making errors rather than investment in social welfare, allowing for the uncritical adoption of predictive technology [25]. This social imaginary potentially undermines the notion of an equal,

sovereign people by both privileging the labour of some to achieve higher technological progress and by privileging AI as a higher form of intelligence.

Data scientists have a particular duty to use their relative influence over popular discourse to aid in re-centering the feasibility and desirability of democratic control. That is because the feasibility and desirability of controlling any technology is predicated on how the debate around it is framed, particularly its capabilities, form, and the trustworthiness of those who develop and build it. Data scientists matter to the construction of AI imaginaries because of their perceived expertise in building these technologies, and they could contribute to a more reasoned debate on their limits and risks. Further, they have more power to speak on the uses and path of development of AI technologies than most citizens, as news organizations potentially view their perspectives as more important and credible. In doing so, there is the possibility of taking up too much space in public discourse and of, as Benjamin puts it (regarding design studies), diminishing "generic" human activities and speech in the guise of branded or jargonistic methodologies [6]. Further, the relative economic privilege of data scientists might tend to shield them from the real harms of the algorithms they make [18]. One potentially counteracting move is to amplify the voices of those already weaving dissonant narratives, whether through civil society activists or less conventional sources for political arguments, such as science fiction. Therefore, while narratives of AI then might appear to be purely aesthetic or outside hard-nosed political debates, they are worth reflecting on and contributing to the reshaping of as they can constrain or enable whether AI is viewed as controllable and worthy of control.

6 CONCLUSION

Political theory and debates about AI systems have suffered from (relative) mutual neglect, with Gabriel pointing out, for example, that liberal democratic theory often assumes away the particulars of technological systems [28]. However, as I have shown here, both disciplines should be more invested in the other, just as their objects of study are interconnected. Political actors can not just misuse AI systems to settle particular disputes; instead, whole political orders and AI technology can be co-constitutive. In response, I have set out a "weak" form of control of state-used AI in a rules-bound order and a "strong" form of participative democratic control of AI use, deployment, and development throughout societies. I recap three primary lessons for academics, civil society groups, and practitioners.

First, while individual decision-maker accountability is vital, it cannot, without democratic control, address the scope of potential harm that AI might cause. This is partly because accountability occurs too late for those already harmed by a system and does not sufficiently address harms arising from the entire system being deployed, such as reshaping the nature of work. Accountability is also best secured in a democratic, rules-based order, especially by controlling bureaucracies through supervisory powers.

Second, critical studies of AI systems could further consider the mechanisms by which particular objectives will be brought about, be they notions of justice, ensuring individual accountability, or creating transparency. A critical contribution of political theory is

moving between notions of ideal social outcomes and the forms of government that might best bring them about or secure against particular political risks. The two notions of democratic governance here could be used for a more in-depth critique of particular extant governance mechanisms, such as the EU's regulatory framework for AI and its place in the general data protection system and digital regulation.

Third, democracy within normative debates on AI systems, development, and deployment of AI need not always mean representative democracies and their formal regulatory institutions. Political theory has both spurred the development of and been inspired by participatory movements that hold the potential to reshape how AI might be controlled and promote positive social values. I articulated particular duties within a democratic context for data scientists: amplifying or contributing to social imaginaries of AI that foreground the potentiality of democratic control and enabling participation, where feasible, of affected communities in their work.

No conceptual framework or governance mechanism proposed in this paper can offer a panacea for AI's pervasive and growing social, political, and economic influence. However, by constructing and defending two notions of democratic control, I have illustrated the need for more substantive and participatory oversight in the AI landscape. In doing so, I have demonstrated the essential role of political theory in critiquing and making proposals for the control of AI and, more broadly, computing systems.

ACKNOWLEDGMENTS

I acknowledge the support of the UK ESRC in funding my PhD Studentship, and in attending FAccT.

REFERENCES

- [1] Saar Alon-Barkat and Madalina Busuioac. 2022. Human-AI Interactions in Public Sector Decision Making: "Automation Bias" and "Selective Adherence" to Algorithmic Advice. *Journal of Public Administration Research and Theory* 33, 1 (02 2022), 153–169. <https://doi.org/10.1093/jopart/muac007> arXiv:<https://academic.oup.com/jpart/article-pdf/33/1/153/48511387/muac007.pdf>
- [2] Julia Angwin, Jeff Larson, Surya Mattu, and Lauren Kirchner. 2016. *Machine Bias*. Retrieved Dec 29, 2023 from <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>
- [3] Ugur Aytac. 2024. Big Tech, Algorithmic Power, and Democratic Control. *The Journal of Politics* (forthcoming) (2024), 1–25.
- [4] Ulrich Beck. 1997. Subpolitics: Ecology and the Disintegration of Institutional Power. *Organization & Environment* 10, 1 (1997), 52–75.
- [5] Emily M. Bender, Timnit Gebru, Angelina McMillan-Major, and Shmargaret Shmitchell. 2021. On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?. In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency* (Virtual Event, Canada) (FAccT '21). Association for Computing Machinery, New York, NY, USA, 610–623. <https://doi.org/10.1145/3442188.3445922>
- [6] Ruha Benjamin. 2019. *Race After Technology: Abolitionist Tools for the New Jim Code*. Polity, Cambridge, UK.
- [7] Abeba Birhane, William Isaac, Vinodkumar Prabhakaran, Mark Diaz, Madeleine Clare Elish, Jason Gabriel, and Shakir Mohamed. 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* (, Arlington, VA, USA.) (EAAMO '22). Association for Computing Machinery, New York, NY, USA, Article 6, 8 pages. <https://doi.org/10.1145/3551624.3555290>
- [8] David Burnham. 1983. *The Rise of the Computer State*. Random House, London.
- [9] Ernest Caldwell. 2017. Widening the Constitutional Gap in China and Taiwan: History, Reform, and the Transformation of the Control Yuan. *University of Illinois Law Review* 2017, 4 (2017), 101–128.
- [10] Federica Carugati and Margaret Levi. 2021. *A Moral Political Economy*. Cambridge University Press, Cambridge.

- [11] Sarah Cohen, Werner Nutt, and Yehoshua Sagie. 2015. The emergence of a finance culture in American households, 1989–2007. *Socio-Economic Review* 13, 2 (2015), 575–601. <https://doi.org/10.1093/ser/mwu0035>
- [12] Competition and Markets Authority. 2023. *AI Foundation Models: Initial Review*. Retrieved Dec 29, 2023 from https://assets.publishing.service.gov.uk/media/64528e622f6220013a6a491/AI_Foundation_Models_-_Initial_review_.pdf
- [13] Josh Cows, Philipp Darius, Dominiquo Santistevan, and Moritz Schramm. 2022. Constitutional metaphors: Facebook’s “supreme court” and the legitimization of platform governance. *New Media & Society* (2022), 1–25. <https://doi.org/10.1177/14614448221085559>
- [14] Kate Crawford. 2021. *Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence*. Yale University Press, New Haven, CT.
- [15] Kate Crawford and Jason Schultz. 2019. AI Systems and State Actors. *Columbia Law Review* 119, 7 (2019), 1941–1972.
- [16] Jenny L Davis. 2023. ‘Affordances’ for Machine Learning. In *Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency* (Chicago, IL, USA) (FAcCT ’23). Association for Computing Machinery, New York, NY, USA, 324–332. <https://doi.org/10.1145/3593013.3594000>
- [17] Trevor Deley and Elizabeth Dubois. 2020. Assessing Trust Versus Reliance for Technology Platforms by Systematic Literature Review. *Social Media + Society* 6, 2 (2020), 2056305120913883. <https://doi.org/10.1177/2056305120913883> arXiv:<https://doi.org/10.1177/2056305120913883>
- [18] Catherine D’Ignazio and Lauren F. Klein. 2023. *Data Feminism*. MIT Press, Boston, MA.
- [19] DIST (The Department of Science, Innovation, and Technology) and The Office for Artificial Intelligence. 2021. A pro-innovation approach to AI regulation. <https://www.gov.uk/government/publications/ai-regulation-a-pro-innovation-approach/white-paper>.
- [20] John S. Dryzek. 2002. *Deliberative Democracy and Beyond: Liberals, Critics, Contestations*. Oxford University Press. <https://doi.org/10.1093/019925043X.001.0001>
- [21] Lynn Eaton. 2017. *Is it right to use AI to identify children at risk of harm?* <https://www.theguardian.com/society/2019/nov/18/child-protection-ai-predict-prevent-risks>
- [22] Lilian Edwards and Michael Veale. 2017. Slave to the Algorithm? Why a ‘Right to an Explanation’ Is Probably Not the Remedy You Are Looking For. *Duke Law & Technology Review* 16, 1 (2017), 18–83.
- [23] Mads Ejlsing, Adam Veng, and Irina Papazu. 2023. Green politics beyond the state: radicalizing the democratic potentials of climate citizens’ assemblies. *Climate Change* 176, 73 (2023). <https://doi.org/10.1007/s10584-023-03550-z>
- [24] Anders Esmark. 2020. *The New Technocracy*. Bristol University Press, Bristol.
- [25] Virginia Eubanks. 2018. *Automating Inequality*. St. Martin’s Press, New York, NY.
- [26] European Commission. 2021. Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts. <https://artificialintelligenceact.eu/the-act/>.
- [27] Steve Fuller and Veronika Lipińska. 2014. *The Proactionary Imperative*. Palgrave Macmillan, London.
- [28] Iason Gabriel. 2022. Toward a Theory of Justice for Artificial Intelligence. *Daedalus* 151, 2 (05 2022), 218–231. https://doi.org/10.1162/daed_a_01911 arXiv:https://direct.mit.edu/daed/article-pdf/151/2/218/2060584/daed_a_01911.pdf
- [29] Nicole Gillespie, Steven Lockey, Caitlin Curtis, Javad Pool, and Ali Akbari. 2023. *Trust in Artificial Intelligence: A global study*. Retrieved Dec 29, 2023 from <https://doi.org/10.14264/00d3c94>
- [30] Carol C. Gould. 1988. *Rethinking Democracy: Freedom and social cooperation in politics, economy, and society*. Cambridge University Press, Cambridge.
- [31] OneTrust Data Guidance. 2020. *Comparing Privacy Laws: GDPR vs. LGPD*. https://www.dataguidance.com/sites/default/files/gdpr_v_lgpd_revised_edition.pdf
- [32] Johannes Himmelreich. 2023. Against “Democratising AI”. *AI and Society* 38 (2023), 1333–1346. <https://doi.org/10.1007/s00146-021-01357-z>
- [33] Stephen Holmes. 2003. Lineages of the Rule of Law. In *Democracy and the Rule of Law*, José María Maravall and Adam Przeworski (Eds.). Oxford University Press, 19–61.
- [34] Christopher Hood. 2006. Transparency in Historical Perspective. In *Transparency: The Key to Better Governance?*, Christopher Hood and David Heald (Eds.). British Academy. <https://doi.org/10.5871/bacad/9780197263839.003.0001>
- [35] Andreas Jungherr. 2023. Artificial Intelligence and Democracy: A Conceptual Framework. *Social Media + Society* 9, 3 (2023), 1–14. <https://doi.org/10.1177/20563051231186353>
- [36] Anna Kawakami, Venkatesh Sivaraman, Hao-Fei Cheng, Logan Stapleton, Yanghui Cheng, Diana Qing, Adam Perer, Zhiwei Steven Wu, Haiyi Zhu, and Kenneth Holstein. 2022. Improving Human-AI Partnerships in Child Welfare: Understanding Worker Practices, Challenges, and Desires for Algorithmic Decision Support. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (New Orleans, LA, USA.) (CHI ’22). Association for Computing Machinery, New York, NY, USA, Article 52, 18 pages. <https://doi.org/10.1145/3491102.3517439>
- [37] John Keane. 2009. *The Life and Death of Democracy*. Simon and Schuster, New York, NY.
- [38] Antonin Lacelle-Webster and Mark E. Warren. 2021. Citizens’ Assemblies and Democracy. <https://doi.org/10.1093/acrefore/9780190228637.013.1975>
- [39] Lawrence Lessig. 1999. The Law of the Horse: What Cyberlaw Might Teach. *Harvard Law Review* 113, 2 (1999), 501–549.
- [40] Annabelle Lever. 2023. Democracy: Should We Replace Elections with Random Selection? *Danish Yearbook of Philosophy* 56, 2 (2023), 136 – 153. <https://doi.org/10.1163/24689300-bja10042>
- [41] Amelie Marian. 2023. Algorithmic Transparency and Accountability through Crowdsourcing: A Study of the NYC School Admission Lottery. In *Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency* (Chicago, IL, USA) (FAcCT ’23). Association for Computing Machinery, New York, NY, USA, 434–443. <https://doi.org/10.1145/3593013.3594009>
- [42] Walter Mattli and Tim Büthe. 2003. Setting International Standards: Technological Rationality or Primacy of Power? *World Politics* 56, 1 (2003), 1–42. <https://doi.org/10.1353/wp.2004.0006>
- [43] Margaret Mitchell, Simone Wu, Andrew Zaldivar, Parker Barnes, Lucy Vasserman, Ben Hutchinson, Elena Spitzer, Inioluwa Deborah Raji, and Timnit Gebru. 2019. Model Cards for Model Reporting. In *Proceedings of the Conference on Fairness, Accountability, and Transparency* (Atlanta, GA, USA) (FAT* ’19). Association for Computing Machinery, New York, NY, USA, 220–229. <https://doi.org/10.1145/3287560.3287596>
- [44] Luke Munn. 2022. *Automation Is a Myth*. Stanford University Press, Stanford, CA.
- [45] Chris Norval, Kristin Cornelius, Jennifer Cobbe, and Jatinder Singh. 2022. Disclosure by Design: Designing Information Disclosures to Support Meaningful Transparency and Accountability. In *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency* (Seoul, Republic of Korea) (FAcCT ’22). Association for Computing Machinery, New York, NY, USA, 679–690. <https://doi.org/10.1145/3531146.3533133>
- [46] Robert Nozick. 1974. *Anarchy, State, and Utopia*. Basic Books, New York City.
- [47] City of Amsterdam. 2020. *AI register*. Retrieved Dec 29, 2023 from <https://algoritmeregister.amsterdam.nl/>
- [48] City of Helsinki. 2020. *AI register*. Retrieved Dec 29, 2023 from <https://ai.hel.fi/>
- [49] Observatory of Public Sector Innovation. 2023. *Chile’s road to algorithmic transparency: Setting new standards in Latin America*. Retrieved April 11 2024 from <https://oecd-opsi.org/blog/chile-algorithmic-transparency/>
- [50] Supreme Court of the United States. 2020. *Seila Law LLC v. Consumer Financial Protection Bureau*. https://www.supremecourt.gov/opinions/19pdf/19-7_n6io.pdf
- [51] Elinor Ostrom. 1990. *Governing the Commons*. Cambridge University Press, Cambridge.
- [52] Zeynep Pamuk. 2021. *Politics and Expertise*. Princeton University Press, Princeton, NJ.
- [53] Philip Pettit. 2008. Three Conceptions of Democratic Control. *Constellations* 15, 1 (2008), 46–55. <https://doi.org/10.1111/j.1467-8675.2008.00473.x>
- [54] David E. Pozen. 2010. Deep Secrecy. *Stanford Law Review* 62, 2 (2010), 257–33.
- [55] President Biden. 2023. Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence. <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/>.
- [56] Joel R. Reidenberg. 2014. The Data Surveillance State in Europe and the United States. *Wake Forest Law Rev* 49 (2014), 583–608.
- [57] Min Reichamps, Julien Vrydagh, and Yanina Welp. 2023. *De Gruyter Handbook of Citizens’ Assemblies*. De Gruyter, Berlin.
- [58] Akos Rona-Tas. 2017. The Off-Label Use of Consumer Credit Ratings. *Historical Social Research* 42, 1 (2017), 52–76.
- [59] Noel Sharkey. 2016. Staying in the loop: human supervisory control of weapons. In *Autonomous Weapons Systems: Law, Ethics, Policy*, Nehal Bhuta, Susanne Beck, Robin Geiss, Hin-Yan Liu, and Claus Kress (Eds.). Cambridge University Press, 23–38. <https://doi.org/10.1017/CBO9781316597873>
- [60] Timothy Simcoe. 2012. Standard Setting Committees: Consensus Governance for Shared Technology Platforms. *American Economic Review* 102, 1 (February 2012), 305–36. <https://doi.org/10.1257/aer.102.1.305>
- [61] Sarah Song. 2012. The boundary problem in democratic theory: why the demos should be bounded by the state. *International Theory* 4, 1 (2012), 39–68. <https://doi.org/10.1017/S1752971911000248>
- [62] State of Connecticut. 2023. Public Act No. 23-16. https://custom.statenet.com/public/resources.cgi?id=ID:bill:CT202300051103&ciq=ncl&client_md=783378b6b29fa0fdab69672348011c6&mode=current_text
- [63] Tapasya, Kumar Sambhav, and Divij Joshi. 2024. *How an algorithm denied food to thousands of poor in India’s Telangana*. Retrieved April 9, 2024 from <https://www.aljazeera.com/economy/2024/1/24/how-an-algorithm-denied-food-to-thousands-of-poor-in-indias-telangana>
- [64] Charles Taylor. 2003. *Modern Social Imaginaries*. Duke University Press Books, Durham, NC.
- [65] Amos Tversky and Daniel Kahneman. 1971. Belief in the Law of Small Numbers. *Psychological Bulletin* 76, 2 (1971), 105–110.

- [66] Steven P. Vallas, Hannah Johnston, and Yana Mommodova. 2022. Prime Suspect: Mechanisms of Labor Control at Amazon's Warehouses. *Work and Occupations* 49, 4 (2022), 421–456. <https://doi.org/10.1177/0730888422110692>
- [67] Niels van Berkel, Zhanna Sarsenbayeva, and Jorge Goncalves. 2023. The methodology of studying fairness perceptions in Artificial Intelligence: Contrasting CHI and FAccT. *International Journal of Human-Computer Studies* 170 (2023), 102954. <https://doi.org/10.1016/j.ijhcs.2022.102954>
- [68] Michael Veale and Frederik Zuiderveen Borgesius. 2021. Demystifying the Draft EU Artificial Intelligence Act — Analysing the good, the bad, and the unclear elements of the proposed approach. *Computer Law Review International* 22, 4 (2021), 97–112. <https://doi.org/doi:10.9785/crl-2021-220402>
- [69] Sandra Wachter, Brent Mittelstadt, and Chris Russell. 2021. Why fairness cannot be automated: Bridging the gap between EU non-discrimination law and AI. *Computer Law and Security Review* 41 (2021), 105567. <https://doi.org/10.1016/j.clsr.2021.105567>
- [70] Angelina Wang, Sayash Kapoor, Solon Barocas, and Arvind Narayanan. 2023. Against Predictive Optimization: On the Legitimacy of Decision-Making Algorithms That Optimize Predictive Accuracy. In *Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency* (Chicago, IL, USA) (FAccT '23). Association for Computing Machinery, New York, NY, USA, 626. <https://doi.org/10.1145/3593013.3594030>
- [71] Langdon Winner. 1977. *Autonomous Technology: Technics-out-of-Control as a Theme in Political Thought*. MIT Press, Cambridge, MA.
- [72] Langdon Winner. 1980. Do Artifacts Have Politics? *Daedalus* 109, 1 (1980), 121–136.
- [73] Langdon Winner. 2020. *The Whale and The Reactor* (2nd ed.). University of Chicago Press, Chicago, IL.
- [74] Janis Wong, Deborah Morgan, Vincent J. Straub, Youmna Hashem, and Jonathan Bright. 2022. *Key challenges for the participatory governance of AI in public administration*. Retrieved Jan 2, 2024 from <https://www.turing.ac.uk/news/publications/key-challenges-participatory-governance-ai-public-administration>
- [75] Pak-Hang Wong. 2020. Democratizing Algorithmic Fairness. *Philosophy and Technology* 33, 2 (2020), 225–44.
- [76] Sun Yat-Sen. 1928. *San Min Chu I: The Three Principles of the People*. Commercial Press, Shanghai. Translated by Frank W. Price.
- [77] Miri Zilka, Holli Sargeant, and Adrian Weller. 2022. Transparency, Governance and Regulation of Algorithmic Tools Deployed in the Criminal Justice System: A UK Case Study. In *Proceedings of the 2022 AAAI/ACM Conference on AI, Ethics, and Society* (Oxford, United Kingdom) (AES '22). Association for Computing Machinery, New York, NY, USA, 880–889. <https://doi.org/10.1145/3514094.3534200>
- [78] Shoshana Zuboff. 2014. *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*. Profile Books, London.