



The Political Justification of Global Governance in the Age of Artificial Intelligence

Dr. Urmila Jaiswal

Guest Faculty

Government Girls PG College, Satna, MP, India

Abstract: *Although the concept of “AI governance” is frequently used in the debate, it is still rather undertheorized. Often it seems to refer to the mechanisms and structures needed to avoid “bad” outcomes and achieve “good” outcomes with regard to the ethical problems artificial intelligence is thought to actualize. In this article we argue that, although this outcome-focused view captures one important aspect of “good governance,” its emphasis on effects runs the risk of overlooking important procedural aspects of good AI governance. One of the most important properties of good AI governance is political legitimacy. Starting out from the assumptions that AI governance should be seen as global in scope and that political legitimacy requires at least a democratic minimum, this article has a twofold aim: to develop a theoretical framework for theorizing the political legitimacy of global AI governance, and to demonstrate how it can be used as a compass for critically assessing the legitimacy of actual instances of global AI governance. Elaborating on a distinction between “governance by AI” and “governance of AI” in relation to different kinds of authority and different kinds of decision-making leads us to the conclusions that much of the existing global AI governance lacks important properties necessary for political legitimacy, and that political legitimacy would be negatively impacted if we handed over certain forms of decision-making to artificial intelligence systems.*

Keywords: Political Justification, Global Governance, Age of Artificial Intelligence

I. INTRODUCTION

The concept of “governance” in discussions around AI is currently a term much too broad for its own good. It is used to refer to everything from the plethora of “ethics guidelines” for AI development written by state- and non-state actors (Jobin et al., 2019), to the presence of human oversight in automated processes (AIHLEG, 2019: 16), and hypothetical international laws for preventing undesirable “race dynamics” among superpowers developing AI (Dafoe, 2018: 43–47). This imprecision is not surprising. First, there are a number of ways of defining AI, and the scope of what counts as AI governance hence depends on how AI is delineated, to begin with. Second, both AI and its regulation are rapidly changing phenomena, and technological breakthroughs and suggestions for how to govern AI are constantly presented not only in academic journals, but also in blog posts, podcasts, and non-peer reviewed papers posted online. Contributors to the AI governance debate are not only academics but also nongovernmental organizations (NGOs) and tech companies with a potential interest in shaping the field in accordance with their interests. Indeed, some have suggested that the vagueness of the concept of “governance” lends it flexibility, which in turn explains why it is so popular among scholars and policymakers (Peters, 2012: 19). We will argue that there is a risk, however, that the way in which AI governance is currently conceived creates a blind spot for the distinctively political nature of the relevant questions about the goals of AI development and deployment, and what it would take for the answers to be democratically legitimate. The simplest way to begin staking out what currently constitutes “AI governance” research, at least research relating to normative issues, is to conceive of it as a subdomain of the wider field of “AI ethics.” This field, in turn, could plausibly be defined by analogy to how other fields in applied ethics are delineated: just like “bioethics” is concerned with ethical issues arising in light of advances in the life sciences, AI ethics is the field engaged in answering ethical questions that arise in light of actual and conceivable AI systems.³ Several recent prominent handbooks and academic overviews of the ethics of AI, for instance, introduce the field as concerned with the ethical implications and use of AI applications, and define it by giving examples of issues it covers, including questions around AI-powered autonomous weapons, criminal sentencing, automated work processes, autonomous



vehicles and sex robots, and what the moral status of AI agents are (Bostrom and Yudkowsky, 2014; Dubber et al., 2020; Liao, 2020; Müller, 2020). Furthermore, most would recognize that AI development is not a deterministic process, but rather a practice that can be steered toward or away from certain outcomes. Often using lofty and imprecise terms, it is commonplace to state that “society” or “humanity” can “guide” AI development toward certain goals. One major obstacle to this, it is assumed but rarely explained with analytical rigor, is how to make sure that AI systems can be made to “align” with “human values,” often meaning roughly that the effects they produce should be in line with what “we” wish to see (Gabriel and Ghazavi, 2022). We suggest that the currently most common way of understanding what “AI governance” is, follows from combining the characterization of AI ethics as constituted by the list of questions and issues that AI actualizes, with the assumption that we can collectively influence the development of AI. Hence, “AI governance” often refers to the mechanisms and structures needed to avoid “bad” outcomes and achieve “good” outcomes with regard to the listed issues, for example, autonomous vehicles, automated decision-making, and so forth. For instance, one influential definition states that “the field of AI governance studies how humanity can best navigate the transition to advanced AI systems, focusing on the political, economic, military, governance, and ethical dimensions” (Dafoe, 2018: 5). Similarly, two recent special issues on the topic were introduced by stating, respectively, that “Understanding and managing the risks posed by AI is crucial to realise the benefits of the technology” (Taeihagh, 2021: 143), and that “Ethical governance is needed in order to develop standards and processes that allow us to transparently and robustly assure the safety of ethical autonomous systems and hence build public trust and confidence” (Winfield et al., 2019: 510, cf. Gasser and Almeida (2017), Cath (2018: 2), and Jelinek et al. (2021: 141)). Relatedly, although many governance efforts are currently happening at the national or subnational level, the transnational nature of externalities created by AI technology suggests that at least some governance mechanisms will have to be global in scope. And although much of the AI development is done by a few companies, primarily in the US and China, they are large, transnational firms, employing people across the world, and implementing the technology in the global market. Since the issues and potential problems identified in AI ethics arise regardless of national boundaries, their solutions require governance mechanisms that are similarly global (cf. Cihon et al., 2020; Erdelyi and Goldsmith, 2018). We believe this is a straightforward way of understanding AI governance, and that it is only natural that a young and interdisciplinary field of study has coalesced like this. Indeed, we find it difficult to think of more fruitful ways of defining the field. It is crucial, however, to recognize that this is a largely outcome-focused view, in the sense that (global) AI governance is defined as whatever helps us solve issues in AI ethics. On this view, the various different international AI standards, company guidelines, or codes of conduct for AI developers, policy frameworks, national laws, and international policies around AI are all forms of governance because they are potential ways of achieving “good” and avoiding “bad” outcomes to the issues AI raises. This focus on outcomes tinges the way we identify what is at stake, however, directing our attention toward certain issues and away from others. Specifically, we believe it raises three sets of worries about how AI governance research is conducted, and how global AI governance itself is set up.

II. THE POLITICAL LEGITIMACY OF AI GOVERNANCE: A THEORETICAL FRAMEWORK

While the prevailing outcome-oriented view of AI governance stresses one important aspect of “good governance,” since a focus on solving specific issues in AI ethics to avoid bad outcomes and achieving good ones is crucial for improving the whole global governance structure of AI, measuring good governance by the effects of governance mechanisms runs the risk of overlooking important procedural aspects of good AI governance. One of the most important properties of good governance from the perspective of normative political theory is political legitimacy. In this section, we develop a theoretical framework for analyzing the political legitimacy of actual and hypothetical global AI governance institutions. Indeed, since political legitimacy is a heavily disputed notion in both the empirical and normative literature alluding to everything from effectiveness to justice the presumption that global governance must at least be minimally democratic to be legitimate is controversial. However, we assume for the sake of argument that it is sufficiently undisputed in the normative theoretical literature, at least among many political philosophers. As stated in the introduction, our aim is not to defend a substantive first-order theory of global political legitimacy, but rather to spell out and defend some normative boundary conditions that any satisfactory account of the political legitimacy of AI governance must respect. Specifically, we will elaborate on the distinction between “governance of AI” and “governance by AI” in relation to different kinds of authority and different kinds of decision-making, and argue that a



satisfactory account of the legitimacy of global AI governance must respect that governance of AI and governance by AI have a specific normative relationship and raise different normative demands. Among other things, it will be argued that governance by AI cannot be democratic by itself, since it needs delegation from governance of AI, which in turn must be authorized through a democratic process. With regard to the notion of “minimally democratic,” this is supposed to be interpreted in gradual (more or less legitimate) rather than binary (threshold for legitimacy) terms. Whether a threshold should be defended is the task for a substantial theory to answer. It refers instead to the idea that we are limited to specifying certain normative boundary conditions for global political legitimacy of AI, to which a substantial account may add further conditions. The argument is thus made on a metanormative and metatheoretical level, which entails a higher level of abstraction and a higher level of generality. Whether a substantial theory will defend a threshold notion and specify where that is located, will depend on whether the theory aims at being ideal or non-ideal in character, and on what feasibility constraints are presumed, and so on—questions we do not take a stand on here.

III. GLOBAL POLITICAL LEGITIMACY: CONCEPTUAL AND NORMATIVE ASSUMPTIONS

As noted in the previous section, a large part of AI governance is inherently global in scope and it is indeed a subset of global governance more generally. This suggests that we must apply a conceptual apparatus that is fitting for a global context to address the question of political legitimacy. For this purpose, the existing literature on AI governance would benefit from importing insights from the research on global governance conducted in political science and political theory. This section hence elaborates some general conceptual and normative assumptions made in the article about how political legitimacy and democracy must be understood in a global, as opposed to domestic, context.

Today, it is widely acknowledged among political scientists and normative theorists that political legitimacy is a desirable quality of global governance institutions more generally, and not only those regulating AI. Hence, even though there is little agreement on the specific regulative content of principles of political legitimacy, there is consensus on the value of political legitimacy in global governance. To theorize the political legitimacy of global governance, however, means to identify certain features of global political legitimacy which distinguishes it from political legitimacy as it has been applied traditionally. In the philosophical literature, political legitimacy is typically described as a virtue of political arrangements and the rules (laws) that are made within them. It refers to the justification of coercive power, political power or political authority, and usually signify the right to rule, and on most accounts also entail political obligations (Buchanan, 2002, 2004; Christiano, 1996; Wellman, 1996). Moreover, it is commonly assumed that principles of political legitimacy are supposed to regulate the relationship between what we may broadly call “rule-takers” and “rule-makers,” that is, between political entities—agents and institutions—that make, apply and enforce rules and the subjects to whom these rules apply (Buchanan, 2002; Erman, 2020; Buchanan and Keohane, 2006; Valentini, 2012). This concept of political legitimacy has been developed for a domestic context, however, and is therefore not fully fitting for the global domain and for the specific circumstances of global politics, such as the emergence of new forms of rule, new actors, and the exercise of unchecked powers (Bohman, 2004). To make the concept useful for normative purposes in the global domain, and allow it to successfully classify objects in that realm, two modifications are called for, which are acknowledged by key theorists in the literature on global political legitimacy and which still preserves the core meaning of political legitimacy as it has traditionally been understood.

In sum, on the proposed conceptual framework, the function of principles of global political legitimacy is to regulate the relationship between political entities and those over whom they exercise political power to determine under what conditions they have the right to make rules, that is, to make political decisions regarding laws and policies, and where benefits and costs are attached to, for example, compliance and noncompliance, noninterference and interference (Erman, 2020). Moving from conceptual to normative presumptions, we assume in this article that global governance arrangements around AI, but also more generally, must at least be minimally democratic in order to fulfill the requirements of global political legitimacy. It is possible, of course, to claim that the promise or threat of AI is so great that it is justified to give up democratic control over AI development in favor of rule by purported experts. We will not engage with this view, and this is not the place to defend our assumption about political legitimacy requiring a minimum of democracy. We believe that the latter is a sufficiently shared view among political philosophers, harmonizing with a wide range of theories of political legitimacy, since it is only expressed here as one condition of legitimacy, not necessarily a sufficient one. This means that it is compatible with accounts requiring that further



conditions are met, such as in terms of efficiency and effectiveness, distributive justice, or the like. Moreover, since we do not pursue a substantive argument offering a specific theory of the political legitimacy of global AI governance, but rather aim to defend certain normative conditions that should be met when theorizing such a theory, it is important to adopt a broad definition of democracy, to make it compatible with all main conceptions in democratic theory, ranging from voting-centered views to views focusing on deliberation and civic engagement. Democracy is here seen as a normative ideal rather than simply a decision method, which means that the principles offered as a minimal democratic threshold for global political legitimacy are seen as part of such an ideal. To treat democracy as a decision method is rather uninteresting in the present context, since that would entail that the answer to what a democratic threshold would be is determined solely by the normative ideal that motivated the choice of democracy (as method) to begin with. Instead, our broad understanding of democracy as an ideal alludes to “the rule by the people,” which is seen as a particular form of political self-determination or self-rule. On this view, a political entity (e.g. system, polity or institution) is democratic if, and only if, those who are affected by its decisions have an opportunity to participate in their making as equals (Erman, 2020). With these general conceptual and normative assumptions on the table concerning global political legitimacy, let us move over to the specific task of spelling out a number normative boundary conditions that a satisfactory account of the political legitimacy of AI governance must respect.

IV. GOVERNANCE BY AI AND GOVERNANCE OF AI

In the literature on AI governance, focus is typically directed either at “governance by AI” or at “governance of AI.” As mentioned in the introduction, our task to spell out and defend some basic normative conditions that any satisfactory account of political legitimacy should respect, is pursued through an elaboration of this distinction. Although the distinction is rarely explicitly noted, we suggest that “governance by AI” is a suitable term to describe the phenomenon of existing governance structures adopting AI technologies as part of their governance mechanisms, such as when public authorities adopt AI-based automated decision-making. “Governance of AI,” on the contrary, is a suitable term for describing the governance structures that regulate and steer AI development and deployment itself, such as the current move from the European Commission to draft legislation banning certain kinds of AI applications. A concern about governance by AI is a concern about a subset of the issues identified in AI ethics, namely those AI-related phenomena that have to do with political rather than merely moral matters, for instance, regarding the responsibility for the exercise of public authority. There is a clear tendency in institutions engaging in public decision-making, for instance, to partially or completely automate decision-making, or rely on AI technology in other ways (van Noordt and Misuraca, 2022). The key political–philosophical issue raised by this is how the special characteristics of AI technology affects existing governance structures or enable new ones. Many worry, for instance, that the opacity of Machine Learning algorithms undermines the distinctively public character of decision-making that is necessary for democratic legitimacy, or that AI-based facial recognition could allow states to set up massive surveillance systems without the authority of the public (Liao, 2020: 17; Powers and Ganascia, 2020: 38). Governance of AI refers, we suggest, to the actual or possible governing of the policy area of AI development and deployment. The issue here is not how AI technology affects existing governance, but rather what types of governance mechanisms and what actual policies would steer the development and deployment of AI technology in the direction deemed desirable, and the characteristics of these processes. For instance, Yeung et al. (2020) have offered a human rights-based framework for the design, implementation, and legal oversight of AI technology, and Jelinek et al. (2021) have recently suggested that the G20 should create a “coordinating committee” to deal with risks created by AI technology. The distinction between governance by AI and governance of AI is rarely mentioned in the literature, and above all, it is to our knowledge not problematized from the viewpoint of AI ethics.¹¹ This is unfortunate, since it risks hiding the fact that the debate around AI governance incorporates two sets of largely different phenomena, whose relationship ought to be theorized. The normative considerations regarding global political legitimacy that we have just presented suggest that governance by and governance of AI are, in fact, related. Assessing the political legitimacy of existing and conceivable AI governance will hence require particular attention being paid to what is governed and how it is governed. For the latter purpose, we argue that the distinction has important normative implications from the standpoint of democracy and thus for global political legitimacy.

**V. NORMATIVE BOUNDARY CONDITIONS FOR THE LEGITIMACY OF GLOBAL AI GOVERNANCE**

Large parts of the current global governance of AI consists of numerous standards, guidelines, policy frameworks, laws, as well as international policies, and political entities rely on governance by AI in more and more areas. In order for global AI governance to be legitimate, we will suggest, it needs to follow the principles that regulate the relationship between the political entities involved and those over whom they exercise political power. These principles determine under what conditions the governing entities have the right to make political decisions. Recall that according to the general understanding of democracy assumed here, the political entities (agents and institutions) making up the regulatory AI structure are democratic if, and only if, those who are affected by the decisions have an opportunity to participate in their making as equals. Given that we assume that global political legitimacy requires a democratic minimum, the democratic quality of this regulatory structure will depend on (a) what kinds of authority lend it support and on what grounds, and (b) what kinds of decision-making that are conducted. This section develops this argument in order to support the wider claim that an account of the legitimacy of global AI governance must respect that governance of AI and governance by AI have a specific normative relationship and raise different normative demands. Broadly speaking, there are two kinds of authority involved in AI governance, which have fundamentally different normative status from a democratic standpoint: what we here call “authorized entities” and “mandated entities,” respectively. What we mean by authorized entities are precisely those entities which are approved directly through such a democratic procedure and thus have the right to make political decisions (“right to rule”). Examples of authorized entities are nation-state parliaments and the EU parliament. Three clarificatory remarks are in order. First, we aim to capture in more abstract terms the normative status of those entities without any necessary ties to the current statist framework, since this opens up more space for future institutional arrangements in global governance with other properties and forms (Erman and Furendal, 2022). Second, we stay neutral with regard to which kinds of agents have the opportunity to participate as equals in the procedure and thus authorize an entity, since the normative boundary conditions we spell out here are meant to capture in more general terms what a theory of global political legitimacy should respect. Finally, democratic authorization could take numerous different forms, depending on which democratic model is favored. The relevant agents could, for instance, be democratic individuals, which is typical in the domestic context, or democratic states, which is typical in many global settings. The conditions also stay neutral vis-a-vis different formal decision rules because, while we are used to individual-majoritarian rule in domestic contexts, a variety of voting rules are applied in global governance arrangements (majoritarian, weighted, unanimity, and so on). For example, while the World Bank applies a weighted procedure, the World Trade Organization (WTO) exhibits majoritarian “one-state, one-vote” decision rules. Apart from authorized entities, what we mean by mandated entities are those entities that have been delegated political power by authorized entities, typical examples being executive bodies and administrative agents and institutions (Erman, 2020). Democratic delegation from authorized to mandated entities could take numerous forms, depending on the model of democracy we are assuming. It could also happen for a variety of reasons, including when, and because, delegation allows for higher-quality, decentralized, or more far-reaching governance, or because it is dictated by a constitution aimed at balancing powers. As we will explicate below, when authorized entities maintain a possibility of revoking the delegated power from the mandated entities, both constitute parts of a “legitimacy chain” that must not be broken. These two different kinds of authority are in turn tied to different kinds of decision-making, from the standpoint of democracy. To explain why, we must return to the idea of democracy as an ideal of political self-determination or self-rule. The normative core of this ideal is the idea that a person is only obliged to comply with the rules she has had the opportunity to authorize. Thus, democracy as a form of political self-determination has to do with political “authorship,” according to which an agent may only be rightfully coerced by decision-making insofar as she has taken part in its authorization. This means that there is a fundamental difference between coercive decision-making and non-coercive decision-making from a democratic standpoint. In practice, coercive decision-making usually comes in the form of lawmaking, whereas non-coercive decision-making typically comes in the form of policymaking. Here we use “coercion” broadly, to make our argument neutral vis-a-vis specific theories of democracy and compatible with the permissive understanding of global political legitimacy sketched earlier. It refers not only to a physical aspect, which is often stressed in the literature involving force, sanctions, or threat of disciplinary action but also an authoritative aspect, which has to do with being subjected to authoritative commands. In both cases, benefits and costs are usually tied to compliance and noncompliance. Typically, laws are more fundamental and formal than policies,



constituting a system of rules that sets out principles, procedures, and standards that proscribe, mandate, or permit certain relationships between people and institutions. Policies usually consist of statements setting out certain procedural and substantive goals of what should be accomplished in the near or remote future. Importantly, however, policies comply with laws and are formulated within a legal framework, even if they may aim to fundamentally change an existing law or identify a new law that is needed (Erman, 2020; Erman and Furendal, 2022). In our view, these two kinds of decision-making (coercive and non-coercive) are connected to the two kinds of authorities discussed above in the following way: it is authorized entities who are the rightful lawmakers (or more accurately, makers of coercive decisions) in the AI space. They may then delegate authority to mandated entities, which become rightful policy-makers (or more accurately, makers of all kinds of non-coercive decisions) in the AI space through this delegation. This means that a satisfactory account of the political legitimacy of AI governance should respect that there is a specific normative relationship between governance of AI and governance by AI from a democratic point of view, and that the two therefore raise different normative demands for global political legitimacy. Recall that governance of AI refers to the existing and conceivable governance structures that regulate and steer the policy domain of AI development and deployment. Governance by AI, by contrast, describes how existing governance structures adopt AI technology in the practice of governing. Since mandated entities gain their legitimacy through delegation from authorized entities, there is a justificatory hierarchy where the latter have supremacy over the former. This, in turn, creates a specific normative relationship between the two, and raises different demands in order for each to be legitimate. The same justificatory hierarchy exists between the governance of AI and governance by AI, where the latter is normatively “parasitical” on the former. Let us explain. Authorized entities in the AI space, would be those entities which have the right to rule because they have been established through a democratic procedure in which those affected by AI (in one form or the other) have had an opportunity to participate as equals in shaping the “control of the agenda” concerning AI, generally seen as a fundamental property of democracy (Dahl, 1989). This is why authorized entities necessarily must engage in governance of AI (rather than governance by AI), since they set up the legal structure for the overall aims and societal goals of AI development and employment as well as the basic form of the main institutions of the AI space, through coercive decision-making. These aims and goals cannot, we will suggest below, be deliberated and legitimately decided upon by an AI system. Mandated entities, on the other hand, are delegated authority to make non-coercive decisions (policies, etc.) applying to AI as a policy area or more generally, and may do so either with or without the use of AI. So, there is no principled reason why mandated agencies cannot legitimately engage in governance by AI, such as AI-assisted decision-making in government agencies, in relation to a predetermined issue or set of issues in a policy area, within an already established legal framework. Decisions around the overall aims and societal goals of AI development, however, can only legitimately be done by authorized entities, not AI systems themselves. It also follows from the normative conditions defended here that authorized and mandated entities can create a legitimacy chain between those affected and the entities that create coercive law and non-coercive policy. This requires, however, that each authorization and delegation fulfills a requirement of accountability. Precisely what this requirement entails and how demanding it is depends on what democratic model is favored. What is argued here, however, is that the normative boundary conditions for global political legitimacy at a minimum require that a property of accountability is constituted by a right to revocation, that is, a right to withdraw authority/political power. In order to secure that the right to revocation can be properly exercised, there is also a need for a degree of transparency with regard to the decision-making process. For without transparency, an authorized or mandated entity could not know whether it needs to withdraw power or not. Moreover, with regard to those affected by the decisions, a normative boundary condition is that they have a right to justification (Forst, 2011), which minimally requires publicity in the decision-making, that is, that they can be offered reasons in the individual case for the decisions made. This entails that if the last mandated entity is an AI, which typically cannot offer reasons for the decision in individual cases, governance by AI must again be assisted by governance of AI through some delegated public official or civil servant who can explain why the decision was X rather than Y in that particular case.

VI. CONCLUSION

We started this article by suggesting that “AI governance” typically refers to the structures and mechanisms needed to avoid bad outcomes and achieve good outcomes with regard to an already identified set of AI-related issues. We then



raised a number of concerns with this predominant outcome-focused view of AI governance. Above all, we expressed the concern that a focus on the effects of governance mechanisms runs the risk of overlooking important procedural aspects of good AI governance. Against the backdrop of this lacuna in current research together with the two presumptions that political legitimacy is one of the most important procedural properties of good governance, and that political legitimacy at least must be minimally democratic we have developed a theoretical framework consisting of a number of normative boundary considerations that any reasonable account of the global political legitimacy of AI governance should respect. This was done by elaborating on the distinction between “governance by AI” and “governance of AI” in relation to different kinds of authority and different kinds of decision-making. Finally, we demonstrated how this framework could be used as a critical compass for assessing the legitimacy of actual instances of global AI governance. Among other things, we pointed out that both the idea of outsourcing democratic decision-making to AI systems, and the soft law documents and ethics guidelines produced by non-mandated private actors in AI governance have deficiencies with regard to political legitimacy.

REFERENCES

- [1]. Acemoglu, D., and P. Restrepo. 2018. “The Race Between Man and Machine: Implications of Technology for Growth, Factor Shares, and Employment.” *American Economic Review* 108 (6): 1488–542, <https://doi.org/10.1257/aer.20160696>.
- [2]. Amodei, D., C. Olah, J. Steinhardt, P. Christiano, J. Schulman, and D. Mané. 2016. “Concrete Problems in AI Safety.” arXiv:1606.06565 [cs].
- [3]. Bietti, E. 2020. “From Ethics Washing to Ethics Bashing: A View on Tech Ethics from Within Moral Philosophy.” In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, January 2020, 210–9.
- [4]. Bonnefon, J., A. Shariff, and I. Rahwan. 2016. “The Social Dilemma of Autonomous Vehicles.” *Science* 352 (6293): 1573–6, <https://doi.org/10.1126/science.aaf2654>.
- [5]. Bostrom, N. 2014. *Superintelligence: Paths, Dangers, Strategies*. Oxford: Oxford University Press.
- [6]. Brynjolfsson, E., and A. McAfee. 2014. *The Second Machine Age: Work, Progress, and Prosperity in a Time of Brilliant Technologies*. New York: W. W. Norton & Company.
- [7]. Calvin, N., and J. Leung. 2020. *Who Owns Artificial Intelligence?* Oxford: Future of Humanity Institute. Working paper. https://www.fhi.ox.ac.uk/wp-content/uploads/Patents_-FHI-Working-Paper-Final-.pdf (accessed October 11, 2021).
- [8]. Dafoe, A. 2018. ‘AI Governance: A Research Agenda’. *Governance of AI Program*. Future of Humanity Institute, University of Oxford. www.fhi.ox.ac.uk/govaiagenda (accessed October 11, 2021).
- [9]. Danaher, J. 2019. *Automation and Utopia: Human Flourishing in a World Without Work*. Cambridge: Harvard University Press.10.4159/9780674242203
- [10]. De Fauw, J., J. R. Ledsam, B. Romera-Paredes, S. Nikolov, N. Tomasev, S. Blackwell, H. Askham, X. Glorot, B. O’Donoghue, D. Visentin, and G. van den Driessche. 2018. “Clinically Applicable Deep Learning for Diagnosis and Referral in Retinal Disease.” *Nature Medicine* 24 (9): 1342–50, <https://doi.org/10.1038/s41591-018-0107-6>.
- [11]. Dignum, V. 2019. *Responsible Artificial Intelligence: How to Develop and Use AI in a Responsible Way*. Cham: Springer International Publishing.10.1007/978-3-030-30371-6.
- [12]. Edmonds, D., S. Cook, and J. Corbett 2020. “The NHS, AI and Our Data.” *Analysis*. BBC Radio 4. February 3, 2020. Also available at <https://www.bbc.co.uk/programmes/m000dyc2>.
- [13]. Erman, E. 2010. “Why Adding Democratic Values is Not Enough for Global Democracy.” In *Legitimacy Beyond the Nation-State?* edited by E. Erman, and A. Uhlin. New York: Palgrave Macmillan.
- [14]. Erman, E. 2016. “Global Political Legitimacy beyond Justice and Democracy?” *International Theory* 8 (1): 29–62, <https://doi.org/10.1017/s1752971915000196>.
- [15]. Erman, E. 2019. “Does Global Democracy Require a World State?” *Philosophical Papers* 48 (1): 123–53, <https://doi.org/10.1080/05568641.2019.1588153>.
- [16]. Erman, E. 2020. “A Function-Sensitive Approach to the Political Legitimacy of Global Governance.” *British Journal of Political Science* 50 (3): 1001–24, <https://doi.org/10.1017/s0007123417000850>.



- [17]. European Commission. 2021. *Europe Fit for the Digital Age: Artificial Intelligence*. Press release. April 21, 2021. https://ec.europa.eu/commission/presscorner/detail/en/ip_21_1682 (accessed October 11, 2021).
- [18]. Frey, C. B. 2019. *The Technology Trap: Capital, Labor, and Power in the Age of Automation*. Princeton: Princeton University Press.10.1515/9780691191959
- [19]. Habermas, J. 1996. *Between Facts and Norms: Contributions to a Discourse Theory of Law and Democracy*, trans. W. Rehg. Cambridge: MIT Press.10.7551/mitpress/1564.001.0001
- [20]. Hagendorff, T. 2020. "The Ethics of AI Ethics: An Evaluation of Guidelines." *Minds and Machines* 30 (1): 99–120, <https://doi.org/10.1007/s11023-020-09517-8>.
- [21]. Jobin, A., M. Ienca, and E. Vayena. 2019. "The Global Landscape of AI Ethics Guidelines." *Nature Machine Intelligence* 1 (9): 389–99, <https://doi.org/10.1038/s42256-019-0088-2>.
- [22]. Karabarounis, L., and B. Neiman. 2014. "The Global Decline of the Labor Share." *The Quarterly Journal of Economics* 129 (1): 61–103, <https://doi.org/10.1093/qje/qjt032>.
- [23]. Kolodny, N. 2014. "Rule Over None I: What Justifies Democracy?" *Philosophy & Public Affairs* 42 (3): 195–229, <https://doi.org/10.1111/papa.12035>.