Eva Erman* and Markus Furendal

# The Global Governance of Artificial Intelligence: Some Normative Concerns

https://doi.org/10.1515/mopp-2020-0046
Published online January 20, 2022

**Abstract:** The creation of increasingly complex artificial intelligence (AI) systems raises urgent questions about their ethical and social impact on society. Since this impact ultimately depends on political decisions about normative issues, political philosophers can make valuable contributions by addressing such questions. Currently, AI development and application are to a large extent regulated through non-binding ethics guidelines penned by transnational entities. Assuming that the global governance of AI should be at least minimally democratic and fair, this paper sets out three desiderata that an account should satisfy when theorizing about what this means. We argue, first, that an analysis of democratic values, political entities and decision-making should be done in a holistic way; second, that fairness is not only about how AI systems treat individuals, but also about how the benefits and burdens of transformative AI are distributed; and finally, that justice requires that governance mechanisms are not limited to AI technology, but are incorporated into a range of basic institutions. Thus, rather than offering a substantive theory of democratic and fair AI governance, our contribution is metatheoretical: we propose a theoretical framework that sets up certain normative boundary conditions for a satisfactory account.

**Keywords:** artificial intelligence, governance of AI, ethics of AI, democracy, fairness

The rapidly increasing capacity of artificial intelligence (AI) to perform tasks previously performed by humans raises concerns and questions about how this will reshape society and social dynamics. Some researchers envision a future utopia where dreary jobs have been automated and human capacities have been enhanced, leaving us free to explore new avenues of the human experience. Others

***Corresponding author: Eva Erman**, Department of Political Science, Stockholm University, Universitetsvägen 10 F, SE-10691, Stockholm, Sweden, E-mail: eva.erman@statsvet.su.se. https://orcid.org/0000-0001-7096-9157
**Markus Furendal,** Department of Political Science, Stockholm University, Universitetsvägen 10 F, SE-10691, Stockholm, Sweden, E-mail: markus.furendal@statsvet.su.se. https://orcid.org/0000-0002-2378-750X

caution against the ways in which intelligent machines might exacerbate existing tensions and conflicts in society, and how the invention of superintelligent systems might ultimately render people politically and morally superfluous. While some believe that AI development is so rapid that these transformations will happen within a decade or two, others remind us of the way similar optimism about AI historically has turned out to be overly optimistic.

Regardless of whether we are utopians or dystopians, and optimistic or sceptical about the pace of change, we should recognize the growing importance of studying the social and ethical impact of transformative AI. This undoubtedly requires an interdisciplinary effort, and recent years have seen important contributions from philosophically inclined engineers as well as technologically inclined philosophers, and there is by now also increased interest in the social sciences and the humanities more generally. Yet, many of the central questions raised by the advent of AI technology are inherently normative and concern values and principles of political morality, which suggests that it is essential that political philosophers turn their attention to the topic.[1]

This paper is one of the first attempts to map some of the fundamental normative concerns that can be fruitfully addressed by political philosophers. A central assumption is that the impact of AI technology will be shaped by how the development and application of AI is governed, and that such governance ideally should be democratically legitimate and fair, as well as global in scope. Rather than offering a substantive proposal for how to achieve this, we survey existing ethics guidelines on AI, in light of which we specify three desiderata that political theorists should satisfy when engaging with the issue of a democratic and fair global governance of AI.

Informed by democratic theory, we argue that a reasonable account should adopt what we call a 'holistic' approach, which includes an analysis of the relationship between core democratic values – such as procedural fairness, transparency, accountability and responsibility – as part of a package, rather than as separate values, an analysis of the normative status of the different agents and institutions involved, and an analysis of different kinds of decision-making. Informed by justice theory, we further argue that a satisfactory account should not limit its focus to the fairness of procedures such as AI-assisted decision-making. Although it is clearly unfair to let potentially biased AI algorithms make politically and socially important decisions, an equally important but somewhat overlooked

1 The term 'political morality' is typically used to refer to values, normative principles and ideals regulating and structuring the political domain, such as social justice, political autonomy, democracy, liberty and political legitimacy. It is sometimes contrasted with morality proper, which typically include questions about fundamental justice, moral rightness, freedom and so on.

question is how fairly the benefits and burdens of transformative AI are distributed among groups in society. Finally, we contend that the fair global governance of AI will not only require the design of institutions regulating AI, but also the reform of many existing economic, political and legal institutions in global governance.

The justification for focusing on democracy and justice in analysing AI governance is not that we assume a necessary internal connection between the two, as this will depend on what substantive theory of democracy and justice one defends. Rather, it rests on the basic normative premise that acceptable or good global AI governance should be both minimally democratic and just. Hence, our desiderata do not rule out reasons for focusing on additional normative ideals. Moreover, another reason is that democracy and justice – and values connected to them – are among the most frequently mentioned in existing ethics guidelines on AI.

The structure of the paper is straightforward. The first section focuses on which subjects are to be governed in the governance of AI (1). In Sections 2 and 3, respectively, we demonstrate the importance of fulfilling one desideratum with regard to democratic governance (2) and two desiderata with regard to fair governance (3). The final section concludes (4).

# 1 What are the Subjects to be Governed in the Governance of AI?

The technological advancements of machine learning and neural networks have recently produced results that both catch the public's eye and influence their lives.[2] The recently revealed AI tool GPT-3, for instance, has been trained on 570 GB of text collected from the internet and can produce writing that is remarkably difficult to distinguish from human-authored text (Marr 2020). Similarly, Google's

**2** Since our argument is concerned with the social and ethical impact of increasingly capable AI systems, we will in general not be interested in the underlying details of the technology, and our analysis does not only apply to machine learning AI systems, but to AI systems defined more broadly. For our purposes, we believe it is sufficient to say that such systems are 'intelligent' when they are able to produce results that would require what we would call intelligence in humans. It is not essential, however, that they replicate or emulate human thinking while doing so (see Dignum 2019, p. 9f.) This broad definition of AI allows us to conduct an analysis that is *not* restricted to cases of AI-assisted decision-making, although some of our examples and much of the recent public discussion have focused on this particular application of AI technology. Given the uncertainty around what kinds of AI systems will be developed and implemented throughout society, we believe it is necessary to theorize the rules, processes, laws and policies that regulate all relevant actors that develop and deploy AI technology, broadly conceived.

AI-based assistant now mimics human speech well enough to place reservations by phone calls (Solon 2018). In a parallel development, public debate has been informed by a number of influential publications regarding potential economic effects of AI technology and what the moral aspects of relying on intelligent machines are (Frey 2019; Russell 2019). As is often the case when a new social phenomenon emerges, scholars have asked how it can be understood within the framework of their discipline, and how the framework might need to be reformed to make better sense of the new phenomenon. This is crucial, since *how* we conceive of a problem determines which answers are imaginable, and the framework of a discipline can sometimes restrict our imagination.

There is by now plenty of philosophical and social scientific studies of some of the AI technology that is closest to being widely adopted, such as automated vehicles and weapons systems (Bonnefon, Shariff, and Rahwan 2016; Sparrow 2007). A substantial amount of attention has also been directed towards speculation about the potential capacity of future AI technology. We have in mind, for instance, the questions raised by the hypothetical invention of artificial general intelligence (AGI), and the existential risk to humanity posed by developing AI that is much smarter than humans, possessing so-called superintelligence. Although they were largely overlooked when the current wave of AI progress started, these issues now receive plenty of attention, and are sometimes categorized into a field called *AI safety* (Amodei et al. 2016; Bostrom 2014; Dafoe 2018, pp. 25–33; Russell 2019; see also Livingston and Risse 2019). Similarly, there is an emerging field of theorists and empirical researchers interested in the issue of *governance of AI* asking, roughly, 'how humanity can best navigate the transition to advanced AI systems' (Dafoe 2018, p. 5).

This field is in its infancy and much work is still largely concerned with untangling the different questions and issues at stake, and categorizing them according to their relative urgency and relevance. We are witnessing, as it were, the process of constructing a discipline dedicated to finding solutions to problems the technology raises. One contribution we wish to make to this process is to stress that there should be more precision with regard to what governance means in this context, and with regard to what kind of subject is to be governed.

We believe the frequent lack of precision in the emerging literature means that AI governance can be spelled out in two very different ways. The first picks up on the issue of AGI and superintelligence and is primarily focused on how the development of AI technology ought to be governed so as to minimize the risk that an AGI or superintelligence could threaten human society. Understood as a technical issue, this requires engineers to come up with ways of making sure that AGI systems are ethically aligned with the values we espouse, such that they have our best interest in mind (Gabriel 2020; see Powers and Ganascia 2020, pp. 35ff.).

Understood as a political issue, it might require institutions that centrally monitor AI research in order to prevent misaligned systems from ever being launched (Dafoe 2018, p. 50). On this first view of governance, it is ultimately the hypothetical AI agent itself that is the subject to be governed. Designing and reforming political institutions is instrumentally important to the extent that it serves the prioritized goal of AI safety. The alternative and comparatively neglected view of AI governance that we will focus on in this paper instead centres around the question of how political institutions influence the development and deployment of AI technology, and how they ought to be arranged so as to realize a broader array of goals and ideals. The question is not how to rein in a potential superintelligent AI system, so to speak, but the more mundane one of how to govern the global actors involved in developing and deploying AI technology in general.[3]

The term 'global governance' typically signifies the inclusion of multiple actors, agents and institutions in the coordination of collective action at the global level. As we will conceive it, the overall aim of global governance is to provide global public goods and avoid global public bads. Many global public goods are increasingly provided with assistance from AI systems, and AI technology could, in a sense, be a public good itself. Hence, by 'global governance of AI' we allude to the rules, processes and decision procedures established by governments, international and intergovernmental organizations, non-state and private actors to regulate the development and deployment of those systems.[4] The aim of this paper is to bring up and analyse a number of desiderata that ought to be considered for this regulatory structure to become (more) democratic and just.

The importance of this kind of governance is obvious, given that technologies do not come into existence on their own and that their effects are not isolated from the rest of society. On the contrary, human decisions and social dynamics always shape what kind of technology is being developed and how it is adopted in society. Hence, it is meaningful to ask questions about the normatively significant effects related to a particular technology and the way in which normative decisions about the development of AI technology create such effects. AI technology is currently regulated primarily in two ways. First, it is indirectly covered by existing legal

**3** There is much more to say about the way in which the notion of the 'governance of AI' is spelled out. For the purposes of this paper, our point is simply that even if it is crucial that a potential superintelligent AI system will act in our best interests, we must not overlook the myriad ways in which other AI-based technology already affects our lives. Focusing only on controlling a prospective Frankensteinian monster risks shifting resources away from critically analyzing the ways in which the technology in Frankenstein's lab already shapes society.

**4** Note that this is different from the increasingly common practice of governance *by* AI, for example when existing institutions adopt AI technologies as part of their governance mechanisms, such as when public authorities adopt automated decision-making.

frameworks, policies and institutions. A newly invented AI-based diagnostic software, for instance, will be covered by the same laws that regulate other medical devices, such as the EU's Medical Devices Directive (Schönberger 2019). Similarly, as we will discuss in Section 2, the EU's GDPR rules provide its citizens some protection from being subject to automated decision-making by AI software. In addition, we are currently seeing some efforts from policy-makers at writing regulations specific to AI technology, such as the recently suggested rules and actions from the European Commission (2021) specifically aimed at 'high-risk' AI systems. Second, such directives are complemented by a kind of soft-law approach, consisting of the vast number of ethics codes and guidelines that have been penned in the private and public sectors in recent years, where AI developers voluntarily commit to upholding particular values and principles in their work. Interestingly, this is an essentially *global* governance, since the vast majority of these documents are authored by international NGOs, multinational corporations, and transnational and international organizations. This makes sense, since the development and deployment of AI technology happens internationally, with little regard to traditional jurisdictional boundaries. Although the drafting of new legal regulations suggest that we are moving into a new phase of AI governance, the soft-law approach of ethics guidelines is arguably still an important governance mechanism for the development and deployment of AI technology. Whether this approach is effective or not is ultimately an empirical question. Regardless, given that these documents play an important role in the current global governance of AI, we believe that central insights can be gleaned by analysing the attention awarded to different values and principles in them. Moreover, our analysis will illustrate that the governance of AI cannot be made more democratic or fair simply by appealing to key values and principles that are associated with democracy and justice, respectively, as expressed in the ethics documents, but that future political philosophical research into the impact of AI, and future developments of the governance of AI, should satisfy the three desiderata that we specify and defend here.

In sum, while existing research often focuses on interesting but arguably still hypothetical scenarios around AI safety, the question of how central political institutions do and should influence the development and deployment of AI technology has been relatively neglected. We believe that much can be won by political philosophers joining the discussion and connecting the dots in a fruitful way. In the subsequent two sections, we draw on insights from democratic theory and justice theory, respectively, to propose three desiderata that a successful account of the democratic and fair global governance of AI should fulfill.

## 2 Democracy and the Global Governance of AI

Recent years' breakthroughs in AI technology have sparked both a lively public debate about the normative aspects of AI, and the emerging computer science subdiscipline concerned with fairness, accountability and transparency. Importantly, it has also generated an extensive number of ethics guidance documents. Since 2016, over 80 such documents have been produced in the public and private sectors – mainly by intergovernmental agencies, international NGOs and multinational corporations – consisting of strategies, principles, codes and guidelines intended to be used in decision-making (Schiff et al. 2020). In a global governance context, several influential policy documents have recently been produced, such as the *OECD Principles on Artificial Intelligence* (OECD 2019) and the *Ethics Guidelines for Trustworthy AI* (EU's High-Level Expert Group on AI), and as we stressed above, these documents in effect make up an important part of the current global governance dealing with the development and deployment of AI.

Comprehensive studies of these ethics documents expose a number of common values and principles, several of which are deeply connected both to democracy, treated in this section, and to justice, treated in the subsequent section. We survey these studies for two reasons: first, for diagnostic purposes, to get an overall picture of which values and principles are regarded as important by key actors and institutions dealing with AI; and second, for illustrative purposes, illuminating our own desiderata in relation to real world cases. Hence, our claim is not that existing ethics documents promote or prevent the democratic governance of AI. Rather, we use them as a starting point to develop our theoretical framework consisting of certain normative boundary conditions specifying what is required for the global governance of AI to be minimally democratic, and how this entails a specific (and different) view of the key principles and values associated with democracy that are stressed in the ethics documents.

With regard to democracy, three of the most frequently occurring principles are not only central in the ethics documents but are also often discussed in democratic theory. The most prevalent of these focuses on the value of *transparency*, which is accentuated in almost all of the over 80 studied ethics documents. While transparency is related to a wide range of aspects, such as the explainability of AI as well as the minimization of its harm, many documents from intergovernmental and non-governmental organizations tie transparency to participation, dialogue and principles of democracy, which emphasize mediation and interaction with stakeholders and the public (Jobin, Ienca, and Vayena 2019, pp. 391f.). The assumption often seems to be that transparency and openness promote freedom and autonomy by not reducing knowledge and options for citizens. In other places,

transparency is seen not so much as an ethical principle but as a second-order condition to realize other ethical principles (Larsson 2020, p. 12).[5]

The second most common principle in the ethics documents focuses on justice understood as *fairness*, which is also seen as closely tied to democracy, especially in documents produced by the public sector. While justice involves many aspects (more on this in Section 3), fairness in this context is expressed as equity and as impartial, non-biased and non-discriminatory procedures. Many sources also focus on fairness as inclusion and equality, for example, through the involvement of civil society and other relevant stakeholders in a dialogical manner. Another aspect concerns fair access to AI as well as the need to obtain correct, complete and diverse data in order to avoid the way in which AI-assisted decision-making risks reinforcing existing biases (Hagendorff 2020, p. 102; Jobin, Ienca, and Vayena 2019, p. 394).

A third key principle focuses on *accountability* and *responsibility*, which are also essential properties of a well-functioning democracy (Larsson 2020; Schiff et al. 2020). What is commonly called 'responsible AI' refers to clarifying which actors – such as AI designers, institutions and corporations – are responsible and accountable for the actions and decisions of AI systems. There is also disagreement on whether humans are the only actors who should ultimately be responsible and accountable for technological products or whether AI should be so in a human-like manner (Jobin, Ienca, and Vayena 2019, pp. 394f.; Schiff et al. 2020).

Arguably, it is not at all strange that the ethics guidelines often focus on values and principles intimately tied to democracy, given that one of the major normative concerns with regard to the global governance of AI is the lack of democratic control over its development and application. The strategies, principles and codes expressed in the ethics documents – produced by both public and private actors in the global domain and intended to be used in decision-making on AI – are a testimony of this concern. At the same time, serious criticism has been raised against the whole industry of ethics guidelines. Above all, critics have argued that rather than leading to an ethical governance of AI, it leads to 'ethics washing' by directing focus away from strong enforcement mechanisms, such as a binding legal framework, towards weak mechanisms, such as self-assigned commitments. It is claimed that this is in fact why ethics documents are so appealing to AI institutions and companies, in particular in a global context (Bietti 2020; Hagendorff 2020).

Now, it is an empirical question whether and to what extent voluntary commitments to ethics guidelines prevent the development of stronger mechanisms or whether they could be advanced in tandem. From a normative point of view, however, it is important to acknowledge that the AI guidelines produced by private

5 For a discussion of transparency in AI, see Larsson and Heintz 2020.

actors resemble those of 'corporate social responsibility', that is, a kind of private business self-regulation striving towards the realization of societal goals of an ethical and humanitarian nature. But since AI systems today are used in most areas of societal importance, such as police, health, education and mobility, there is an increasing demand for systematically grounded ethics guidelines in the public sector, both in terms of soft law and policy instruments and in terms of legally binding law. A normatively robust approach to global AI governance is imperative because even if successful, the 'trust me' kind of corporate self-regulation suggested by private actors has major normative and practical limitations (Whittaker et al. 2018, pp. 30ff.).

Addressing these concerns, we argue in this section that a reasonable account of the democratic global governance of AI should satisfy a desideratum in the form of a holistic approach consisting of three aspects: an analysis of the relationship between key values (and principles), an analysis of the normative status of the different entities (agents and institutions) involved, and an analysis of different kinds of decision-making by those entities. In our view, a response to all three is vital for developing a satisfactory account. Importantly, though, our analysis is not a criticism of existing ethics guidelines, but rather an attempt to specify what conditions an acceptable account of democratic AI governance should fulfill. To what extent current ethics documents would supplement, support or hinder a democratization depends in part on a number of empirical factors that will not be investigated here. Our analysis can, however, help identify the normative status of such documents with regard to how these documents came about, who authored them, who sanctioned them, who endorsed them and on what grounds. Thus, the proposed desideratum should be understood in ideal-theoretical terms, by which we mean that the holistic approach offers a general normative framework for the analysis of democracy in AI governance, rather than non-ideal principles intended to be directly applied to current states of affairs. Indeed, due to the complexity of global governance structures, assessing their democratic credentials is far from a straightforward task, not least since it depends in large part on assumptions about feasibility that are tied to proposed normative principles. However, our hope is that this more abstract framework is still useful in guiding normative theorizing about and development of global AI governance by specifying conditions that minimally must be met for an account of democratic governance to be plausible.

## 2.1 How Democratic Values and Principles are Knit Together

Let us start with the first aspect of the holistic approach, namely, an analysis of the relationship between values (and principles) from a democratic standpoint. As

noted above, transparency, fairness, accountability and responsibility are all essential properties of democracy. In the ethics documents – as well as in the debate about the shape of AI governance – these tend to be articulated as *separate* values specified in principles, which are then operationalized mathematically and implemented through technical solutions (Hagendorff 2020, p. 103). Were we to implement such policies, however, we would run the risk of committing ourselves to the underlying assumption that the more each of these values is realized, the more democratization takes place. Let us call this the 'additive view' of democracy. The term 'additive' refers to the idea that strengthening a number of core 'democratic' values (i.e. values that we associate with democracy) leads to more democracy (Erman 2010).

In order to appropriately theorize democratic AI governance, we contend that this additive view should be rejected, since it does not appreciate the extent to which these values are tied together when regarded as parts of democracy. Above all, the additive view cannot account for the fact that democracy, 'the rule by the people', includes two sides that are intimately connected. First, there is an *access* side of political power, where those affected (e.g. by being subjected or by being the relevant stakeholder) should have a say in the decision-making according to some normative standard. Second, there is an *exercise* side of political power, where those very decisions should apply in appropriate ways according to some normative standard (Erman 2019, p. 129). These two aspects are typically analogous to the terms 'input legitimacy' and 'output legitimacy' often used in discussions about democracy in the general political science literature (see Scharpf 1999). Whereas the access side of political power concerns the participation, inclusiveness, deliberation, procedural fairness and political equality of the decision-making, generating input legitimacy of political authority, the exercise side concerns the effectiveness and performance of that authority, generating output legitimacy.

Crucially, however, input and output legitimacy are *knit* together such that we cannot compensate the lack of input legitimacy with more output legitimacy,[6] and thus draw conclusions about increased democracy by only looking at strengthened output values, such as effectiveness and performance, as would be allowed by the additive view. Rather, the normative ideal of democracy incorporates an essential *connection* between input and output legitimacy, which constitutes a two-way relation between those who make decisions and those to whom these decisions apply. In other words, principles of democratic legitimacy regulate the *relationship* between rule-makers and rule-takers, responding to the question of who exercises

**6** Even if some would argue that it may compensate to some degree (e.g. Scharpf 1999), this could only be justified above a certain threshold of input legitimacy.

power over whom. Hence, in the present context, it is *not* the case that any agents are supposed to be accountable and responsible (e.g. AI designers or multinational corporations) any more than it is the case that any agents are supposed to have the opportunity to participate in fair and impartial decision procedures as equals. Strengthening each of these values does not necessarily entail more democratic AI governance; it depends on how principles promoting fairness, equality, accountability, responsibility and transparency connect to each other. Moreover, as will be discussed next, it also depends on which entities (agents and institutions) are involved in the governance and what kind of decisions they make.

Needless to say, the answer to the question of what a democratic global governance of AI might look like depends heavily on what we mean by democracy. Since our contribution is metatheoretical, in the sense that we do not aim to develop any substantive theory but focus on desiderata for a good account, it is important to adopt a broad definition to make it consistent with all main conceptions in democratic theory, ranging from models based on voting and electoral representation to models based on civic engagement and deliberation. On this definition, 'a political system is democratic if, and only if, those to whom its decisions apply have an opportunity to participate in their making as equals' (Valentini 2014, p. 791).

## 2.2 The Normative Status of Entities and Different Kinds of Decision-Making

As mentioned earlier, a holistic approach focuses not only on how democratic values and principles are tied together, but also on the normative status of the entities (agents or institutions) that formulate and fortify these values and principles in the form of documented standards and regulations, as well as on the different kinds of decision-making through which this is done. We unfold these two aspects employing two distinctions: authorized and mandated entities, and coercive and non-coercive decision-making.

In global governance, there is a myriad of guidelines, standards, policy documents and laws regulating the AI space, pronouncing principles for the protection and promotion of fairness – understood as inclusion and equality and as impartial procedures – as well as of transparency, accountability and responsibility. But the democratic quality of this regulatory structure will very much depend on what kinds of authority lend it support and on what grounds. According to the broad understanding of democracy presumed here, the access side of political power requires that those who are affected by the regulatory structure (e.g. by being subjected or by being the relevant stakeholder) have influence in the decision-making about its

basic form and content. What we mean by *authorized* entities (agents and institutions) are precisely those entities that are approved through such a democratic procedure. Typical examples of authorized entities are parliaments, like nation-state parliaments or the EU parliament. The use of the term 'authorized entities', however, is meant 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.

Indeed, having influence in decision-making means different things depending on which substantive democratic model is defended, but typically it means at least having a formalized 'say' in the decision process, in the form of equal voting rights.[7] For sure, most theorists – not least proponents of participatory democracy and deliberative democracy – would add a number of conditions to fulfill the requirement of 'influence', such as having access to deliberative fora that feed into the decision process (Habermas 1996). But for our metatheoretical purposes, we may leave this an open question. Apart from authorized entities, what we here call *mandated* entities are entities that have been delegated political power by authorized entities, typical examples being executive bodies and administrative agents and institutions.

Intimately connected to the distinction between authorized and mandated entities is the distinction between *coercive* and *non-coercive* decision-making, which in practice typically alludes to *law* and *policy-making*.[8] The latter distinction is often overlooked in the theoretical literature on democracy, which is unfortunate since there are differences between law and policy that are of normative significance from a democratic standpoint. In general terms, laws are both more formal and more fundamental than policies, constituting a system of rules that sets out procedures, principles and standards that mandate, proscribe or permit certain relationships between people and institutions. Policies typically consist of statements setting out certain procedural or substantive goals of what should be

**7** Of course, some political philosophers question that electoral vote could count as influence other than randomly (and rarely) when our vote happens to count such that it 'tips the scale', as it were (Kolodny 2014). But from a democratic point of view, this is the wrong way of understanding authorization. It is not primarily a causal or an epistemic notion, but a normative one, which must be understood at the appropriate level of abstraction: one is the co-author of the system of rules to which one is supposed to comply. Authorization in this sense should be understood as mediated authorship (indeed, we are rarely the sole authors of things that we do in life), through institutions, representation and social norms.

**8** Importantly, though, it is the *coercive* property that is of normative importance. In a global governance context, for example, we might have laws that are non-coercive and thus more like policies, such as global administrative law.

achieved in the near or remote future. Importantly, though, they 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. In sum, law-making is specific in the sense that laws generally *coerce* persons by mandating or proscribing certain relationships between people and institutions, backed up by force or the threat of sanctions (Erman 2020).

On the basic normative level, these theoretical building blocks are tied together in two ways. First, there is a link between authorized and mandated entities that is essential for democracy, since it establishes a justificatory hierarchy, where the former has supremacy over the latter, lending the latter legitimacy through delegation. Second, authorized entities are law-making entities, whereas mandated entities are policy-making entities.

## 2.3 A Holistic Approach to Global AI Governance

With the three aspects of the holistic approach on the table, let us apply this analytical framework to the AI context. It gives us the tools needed to assess the *democratic* credentials of different agents and institutions, different guidelines, regulations and documents, and different principles articulating democratic values in the AI space.[9]

To begin with, there is a multitude of legal regulations already in place today, which are general but apply also to AI technology, ranging from human rights and liability law to antidiscrimination laws and road traffic regulations. The EU's GDPR law, for instance, has a large effect on AI since it regulates the sizable volumes of digital data on which AI depends and articulates several rights of importance, such as the right against automated decision-making, the right to explanation and the right to erasure and to data portability. Adopting the holistic approach to assess the democratic credentials of GDPR, we see that since it was introduced by the European Union (in 2018) it was legislated by an authorized entity, as those affected (citizens of EU member states by proxy) had influence in the decision-making, thus securing values such as inclusion, equal participation and procedural fairness.

**9** Unless, of course, we reject the idea of democracy as expressed in our broad definition. One could, for instance, instead regard democracy simply as a decision method; then the defence of that method would be justified with reference to the normative ideal (e.g. utilitarianism) that motivated that choice in the first place. But such a view would be highly controversial for anyone who cherishes democracy as a normative ideal, and would not be very interesting in the discussion about democratic AI governance.

Another example is anti-discrimination laws. Discrimination is prohibited in many constitutions and treaties, including the United Nations Declaration of Human Rights and the European Convention on Human Rights. The latter contains EU law against direct discrimination (i.e. discrimination on the basis of race, sex, religion, color, political or other opinions) as well as indirect discrimination (i.e. discrimination occurring when a seemingly neutral practice places persons of, for example, ethnic or racial origin at a particular disadvantage). Both are used to combat discriminatory AI decisions, in particular protecting against indirect (and thus unintentional) discrimination that is more likely to occur in this context (Zuiderveen Borgesius 2020, p. 19).

These are examples of general laws that also apply to AI. But several challenges remain to strengthening the democratic quality of AI space, so specific laws might be needed. For instance, U.S. senators recently introduced a bill specifically designed to require law enforcement agencies to seek a court order before trying to access personal data from third-parties like private AI-centred companies (Robertson 2021). Moreover, some experts have claimed that the GDPR is too ambiguous to establish a strong protection for individuals and does not in fact provide a 'right to explanation' of automated decision-making (Wachter, Mittelstadt, and Floridi 2017). Moreover, since the GDPR only applies to decisions that are based *exclusively* on automated processing (unless one has given explicit consent), many kinds of automated decisions are out of reach for the GDPR's rules.[10] Also anti-discrimination laws face challenges. Existing laws protect against discrimination on particular grounds, which means that AI-assisted decision-making could potentially reinforce existing biases by sorting people into newly invented classes that merely correlate with the protected grounds or by accepting too many false positives or negatives in its classifications. This suggests, at the very least, that the introduction of AI-assisted decision-making warrants an oversight of existing non-discrimination law or an effort to ensure that intellectual property rights do not make oversight impossible (Zuiderveen Borgesius 2020).

If we move beyond authorized entities and law-making and take a look at mandated entities, their democratic credentials varies considerably. In fact, many policy documents on AI are instituted by entities without delegated power from authorized entities. Compare, for example, the processes by which two particularly influential ethics codes were developed. The so-called Asilomar AI principles were reached through a deliberative process among participants in the Beneficial AI conference in Asilomar, California in 2017 and have since been signed by almost

**10** See Article 21 and 22 and Recitals (71) and (72) of the GDPR.

6000 AI researchers and others (Future of Life Institute 2017). The EU Commission's Ethics Guidelines for Trustworthy AI, in turn, were developed by the independent High-Level Expert Group on AI and finalized, after a round of open consultation, in 2020. Both documents were developed in a transparent way by experts in AI technology, including researchers, representatives from different organizations, and business leaders, and both were driven by the same overall normative aim and without profit. Yet, unlike the authors of the Asilomar principles, the EU expert group were appointed by institutions that are, ultimately, sanctioned by the EU. On the suggested holistic view, this difference is normatively significant. In the EU case, the justificatory link between authorized and mandated entities is established, but in the case of the Asilomar principles the authors are, as it were, self-appointed. Furthermore, the access side (input legitimacy) and exercise side (output legitimacy) of political power are knit together in the EU case, generating both chains of democratic legitimacy and of accountability. Worth noting is that this difference cannot be accounted for by the additive view, which does not have the resources to distinguish the two cases in this respect, since it focuses on the promotion of separate values. So even if the additive view may be used to draw conclusions about the transparency of the process of reaching the Asilomar principles – for example, arguing that more transparency means improved AI governance in *that* respect – it cannot be used to draw conclusions about, for example, accountability without coupling these output aspects to the democratic values of the input side. In addition, this shows that the governance of AI development and deployment cannot be made more democratically legitimate simply by appealing to values and principles central to democracy, but that the procedures used to establish governance mechanisms also matter.

The global governance of AI is currently characterized both by non-binding ethics guidelines and by binding laws. The former can often complement the latter, not least since rapidly changing AI technology is difficult to pin down and legislation always risks lagging behind. Yet, the illustration above shows how a holistic analysis suggests that ethics guidelines can only be part of a governance structure that purports to be democratic. To get deeper knowledge of how to best democratize global AI governance, more research on the whole regulatory structure is called for. We might, for instance, need both new general laws that apply to AI as well as more authoritative policy instruments (e.g. ethics guidelines) from mandated entities. Regardless of which democratic model is favoured, we have argued here that upholding the justificatory link between authorized and mandated entities as well as the tie between input and output legitimacy is important to be able to assess the democratic credentials of decisions, agents and institutions in the AI space.

# 3 Justice and the Global Governance of AI

If democracy responds to the question of who exercises power over whom, one could say that justice responds to the question of who owes what to whom. In contrast with moral theory in general, justice is typically thought to be concerned with the moral quality of basic institutions rather than individual actions. Even if there is disagreement on how to specify this in detail, there is broad agreement on the general characterization that principles of justice establish when institutions give their subjects what they are entitled to – that is, when they respect their rights (Erman 2016, p. 37; Valentini 2012, p. 595).

A majority of the ethics guidelines reviewed recognize the effects of AI in terms of justice, and, as we noted above, 'justice' and 'fairness' are the most commonly mentioned principles after 'transparency' (Jobin, Ienca, and Vayena 2019, p. 394). Similarly, the idea of non-maleficence – that AI should not cause foreseeable or unintentional harm – figures in more than half of the documents, as does the less frequent, mirroring principle of beneficence. For instance, in line with Google's old motto of 'Don't be evil', the company's first principle guiding their AI applications is that AI should 'be socially beneficial', and the company promises not to design or deploy AI in 'technologies that cause or are likely to cause overall harm' (Google 2021).[11]

This section attempts to specify further what justice and fairness could be taken to mean, here, with the aim of demonstrating that a proper account of a just global governance of AI should satisfy two desiderata. The first concerns the way that justice and fairness are not only procedural ideals but also help us evaluate the distribution of benefits and burdens, and the way in which we need both aspects in order to analyse the relevant concerns raised by AI technology. The second concerns the way in which the wide impact of the technology entails that a fair global governance of AI requires not simply new regulation of tech companies, but a comprehensive review of many economic, legal and political institutions as well.

## 3.1 Procedural and Distributive Justice and Fairness

One indicative observation of how the concepts of justice are understood in the ethics guidelines is the attention given to the problem of algorithmic bias, that is, the way in which decision-making algorithms can reflect and exacerbate

11 For instance, Google has vowed not to participate in developing AI used in weapons systems (Wakabayashi and Metz 2018).

existing injustices. Essentially, machine-learning algorithms acquire their skills by studying earlier decisions made by humans, and many systems used for decision-making are simply mathematical models based on how we have acted. It is now widely recognized that any biases that have shaped human decision-making therefore risk being magnified and reinforced by the move to automated decision-making: if historical employment decisions have been biased in favour of men, for instance, an AI algorithm trained on these decisions will likely suggest that a company continue hiring mostly men (O'Neil 2016, ch. 6). The tendency to address this problem in terms of fairness suggests that the solution would require the same kind of impartial and non-biased decision-making procedures that we discussed in the analysis of democracy above. This fits well with the view that principles of justice establish when institutions give their subjects what they are entitled to. For instance, one of the things that seems to be unjust about biased AI-assisted decision-making is that it frustrates people's legitimate expectation that like cases should be treated alike and attaches too much weight to statistical correlations drawn from individuals with similar characteristics. A second reason, however, is that equal treatment might also reinforce existing inequalities, unless institutions instead treat individuals differently in order to achieve equality of outcome. In both cases, the focus is on fairness as a *procedural* ideal specifying how institutions ought to treat individuals.[12]

We believe it is crucial, however, to recall that justice and fairness can also be spelled out as *distributive* ideals, in the sense that they help determine what constitutes a fair distribution of a particular set of benefits and burdens. This is essential, since the general adoption of AI technology will have both beneficial and potentially harmful *indirect* effects on different individuals. Fairness spelled out as an ideal only pertaining to procedures cannot address this and neither can the principles of beneficence and non-maleficence, since it is uninformative to simply state – as the ethics documents often do – that bad effects should be minimized and that AI should be 'socially beneficial'. The more pressing question is rather how to think about cases where the benefits and burdens of AI are unequally distributed between different groups and when the realization of some benefits are conditional on creating burdens. Pursuing this question requires engaging with the rich literature of distributive justice that already exists in political philosophy. Two examples illustrate this point.

**12** Algorithmic bias is indeed an important topic, but we cannot here discuss when and why discrimination is wrong. Our point, above, is that fairness understood procedurally requires treating like cases alike but that, because of historical injustice and existing inequalities in real-world societies, this does not necessarily entail that institutions should treat all individuals the same (see Ronzoni 2009).

One of the most important ways in which the current governance of AI – or lack or lack thereof – has distributive upshots is the fact that the technology is in high demand and can yield profits to those who develop and own it. At the same time, the potential harmful effects that AI can have and the mitigation of that risk is largely shared by society as a whole. One way to conceptualize this is to say that currently the profits of AI technology are privatized, while the negative externalities of the technology are socialized (see Korinek and Stiglitz 2017).

For instance, the profits generated by AI technology are to a large extent made possible by free access to the personal data that is constantly created by people's online presence, and made useful to algorithms by precariously employed so-called 'click workers' who manually sort through and arrange the raw data, often in the Global South (see Hagendorff 2020, 105f.). Specifically, this data is crucial when training machine learning algorithms, such as the text-producing GPT-3 technology mentioned above, and the availability of this data is generally held as one of the preconditions for the current AI boom. Given this, it is not implausible to turn to theories of distributive justice to assess whether such data should be seen as a shared resource, much like other kinds of public commons, and whether profits that were created using the data should be privatized to the extent they currently are. Spiekermann et al. (2020), for instance, have recently defended a progressive data tax on companies that use the data to make a profit, by reference to the way it would counteract the negative externality of increasing societal economic inequality that free access to data leads to.

Similarly, many worry about the fact that AI technology soon will be able to replace many of the tasks that so far only humans have been able to perform and that this is likely to cause many jobs to disappear. So far, however, the most influential academic analyses of the phenomenon and possible political responses have been penned by economists (Acemoglu and Restrepo 2018; Frey 2019) or technologists (Brynjolfsson and McAfee 2014), and the issue has yet to be discussed extensively by political philosophers. It is ultimately an empirical question whether this will bring about mass unemployment or if, as in previous technological shifts, innovations will have an offsetting effect of not only destroying but also creating new jobs. On the other hand, even if new jobs are created, they might be much less satisfying. Alternatively, those who have been displaced by machines might not be qualified for them. While other disciplines have focused on the macroeconomic and political instability technology might bring about, there are many prior normative questions left for political philosophers, such as whether people have a right to a meaningful job, and what kinds of compensation claims, if any, are justified (see Danaher 2019). In sum, there is much to suggest that unless regulated, the benefits of AI-driven automation – such as new and cheaper goods and services – will fall on certain groups, while the negative externalities

created – such as unemployment or reduced job satisfaction – will be carried by other groups who also gain less from technological shift itself.

Since these pressing problems fall outside of the purview of fairness understood procedurally, this illustrates the desideratum of allowing justice and fairness to be interpreted in non-procedural terms. Without questioning the tendency among scholars to research the ills of and solutions to algorithmic bias, our point is simply that AI technology also raises fairness issues concerning the more indirect benefits and burdens it creates. As the next section will argue, this in turn explains the desideratum of adopting a comprehensive view of what institutional reform might be necessary to fairly distribute the benefits and burdens of AI technology.

## 3.2 The Impact of AI on Existing Institutions

The lesson from the examples above is not that the existence of negative externalities requires AI development to be stopped. Rather, it shows that AI governance should be attentive to how the profits and externalities are distributed among those whose interests are significantly affected. Yet, since AI technology is likely to have extensive effects throughout society, we will also need to think through how the technology will change and possibly undermine the way many other existing institutions work, and consequently how they should be reformed or reinvented in order to counter these effects. Two examples illustrate this point.

As we noted above, the quality of training data is crucial for the success of developing machine learning innovations, and some of the best available data is currently held by government bodies. Allowing private AI companies to use health care data, for instance, can potentially enable the development of entirely new ways of diagnosing and treating many illnesses. But as we argued before, regulatory frameworks ultimately decide who owns this data and, by extension, who gets to reap the profits from the potentially lucrative end product it enables. Currently, the answer often seems to be that the private company may sell access to the product back to those whose data enabled it.[13] Whether this is considered just will depend on whether we think of data as a collectively owned resource, but also whether we believe that the premiums entrepreneurs are awarded for refining such resources should be considered just (see Nozick 1974). Examples like this will only become increasingly common in the future, and our point is that they cannot be

13 For instance, the Google-owned company DeepMind were granted access to eye scans from one of the leading eye hospitals in the U.K. and developed an algorithm capable of identifying patients with a serious condition to the same extent as experts (de Fauw et al. 2018). DeepMind now owns the technology and the profits it can yield (Edmonds, Cook, and Corbett 2020).

addressed merely by regulating what AI developers may or may not do. Rather, an analysis of justice in AI governance needs to include other legal and political institutions, such as those regulating intellectual property and public data use. Relating back to the analysis of democracy above, one possible implication of this is that, at the very least, there needs to be institutions that allow democratic deliberation and decision-making regarding the continuation of this practice.

Second, if there is a massive shift toward AI-based automation, this will arguably generate stress on central welfare state institutions. Not only could the share of citizens needing economic assistance grow larger than the institutions were designed to handle, labour-replacing technology will also tend to reinforce the current trend of profits going to capital rather than labour (Karabarbounis and Neiman 2014). Currently, low capital taxes are often motivated by claiming that they make not only owners of capital but everyone better off by helping to attract investments and stimulating growth. Assuming that this is true, the widely discussed difference principle defended by Rawls (1999) could, for instance, be interpreted as saying that the tax cuts are compatible with justice, even though they benefit some more than others. Yet, since the promise of AI-driven automation creates additional incentives for companies to replace human workers with machines (i.e. capital), this eventually risks becoming an existential threat to the income tax base that funds much of the welfare state. Simply put, if there are too few employed people that can be taxed and capital taxes are too low, it is difficult to raise the resources needed for welfare state institutions. This has led to a revived interest in the idea of a universal basic income as a way of both distributing resources fairly and maintaining consumer demand (Brynjolfsson and McAfee 2014). Yet, such proposals must not overlook the issue of how to collect the necessary resources to begin with, in light of how global AI-driven automation can undermine the relevant institutions. This illustrates some of the ways in which we need to consider how to transform the way we currently fund institutions, and what would be acceptable replacements if they turned out not to be viable in the long run.

One possible institutional reform to ensure access to the benefits of AI innovations and thereby enable redistributive policies is a radical idea defended in the debate on the development of new drugs. Some have suggested that the fundamental problem with how drugs are currently developed is the reliance on a system of patents. Patents essentially grant the right to a temporary monopoly on the use of a new technical development to the patent holder, and leave decisions about research and development to be dictated by the financial incentives created by consumer demand. This leads pharmaceutical companies to care less about certain diseases, and it restricts access to potentially life-saving medicines to those with the means to pay for them. Activists and scholars (e.g. Stiglitz 2006) have

suggested that an attractive way of complementing the patent system would be to create a fund that awards prizes to those who succeed in making certain developments. These prizes would cover the costs involved in developing new drugs, thus replacing one of the main justificatory mechanisms of the patent system. Importing this idea into the AI technology case makes sense, since both drugs and AI systems are characterized by high development costs, while being comparatively cheap to run or replicate. Awarding prizes instead of patents would ideally make sure that advancements in AI technology are widely shared, and it could prevent companies from building up large patent portfolios in order to charge excessive prices or capture market shares. This would not only help to address the profit-sharing issues related to data use mentioned above, but also enable a degree of democratic control over what kinds of AI technology to research and develop, and what jobs we would like to see automated or not.[14]

In sum, the way in which reforms like this one promise to address some of the central issues regarding fairness and justice illustrates the second desideratum relating to these ideals. It is plausible to assume that the profits of AI innovation may be differently distributed if people who have generated the necessary data have a say in how it is being used. Similarly, a shift towards a prize system could potentially both counteract unequal wealth accumulation and promote democratic control over the development of AI. The conclusion at the end of Section 2 – that our focus should be on the whole regulatory structure and not only laws and guidelines applying to AI directly – follows from the analysis of this section as well: we should indeed resist thinking of the global governance of AI as merely about properly regulating what AI developers may or may not do. The social and ethical impact of AI is so wide that a successful account of the fair global governance of AI must instead adopt a comprehensive approach to reforming and reinventing a wide range of political, economic and legal institutions.

# 4 Conclusions

The conceptual and normative tools of political philosophy can clearly be of use in addressing many of the normative questions raised by the increasing use of AI technology in society. We have called for a more careful analysis with regard to how to understand AI governance and the subjects that ought to be governed.

**14** Another benefit of this proposal is that the AI development community is, arguably, characterized by an ethos of cooperation and sharing, as evidenced by the fact that much development is open-sourced, that is, that many of the essential tools and software may be used for free. For an overview of the intellectual property policies in the AI industry, see Calvin and Leung (2020).

Under the assumption that the global governance of AI should ideally be democratic and fair, we have suggested three desiderata that a successful account should satisfy. First, democracy should be theorized in a holistic way, recognizing how different core values interact and are tied together, the relationship between authorized and mandated entities, and the normative difference between law and policy-making. Second, it should address not only how fairly the technology treats individuals directly, but also the indirect distributive effects created. Finally, addressing these problems requires not only new institutions to regulate AI but a comprehensive review of many political and economic institutions. Thus, rather than offering a first-order theory of democratic and fair AI governance, our contribution is best described as methodological and metatheoretical in the sense that it offers a theoretical framework through a number of desiderata, which sets up certain normative boundary conditions for a satisfactory account.

**Acknowledgments:** We owe special thanks to Kim Angell and Theodore M. Lechterman for comments on earlier drafts of this article. For valuable feedback we also wish to thank the participants at the Nordic Network in Political Theory, the workshop "Ethics of AI Use in the Public Sector" at the Royal Institute of Technology, and the panel "Governing AI" at the Mancept Workshops in Political Theory. In addition, we thank the editor and anonymous referees of *Moral Philosophy & Politics*.

**Research funding:** We are grateful to the Marianne and Marcus Wallenberg Foundation (MMW 2020.0044) and the Swedish Research Council (VR 2018-01549) for generously funding our research.

# References

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.

Amodei, D., C. Olah, J. Steinhardt, P. Christiano, J. Schulman, and D. Mané. 2016. "Concrete Problems in AI Safety." arXiv:1606.06565 [cs].

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.

Bonnefon, J., A. Shariff, and I. Rahwan. 2016. "The Social Dilemma of Autonomous Vehicles." *Science* 352 (6293): 1573–6.

Bostrom, N. 2014. *Superintelligence: Paths, Dangers, Strategies*. Oxford: Oxford University Press.

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.

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).

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).

Danaher, J. 2019. *Automation and Utopia: Human Flourishing in a World Without Work*. Cambridge: Harvard University Press.

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.

Dignum, V. 2019. *Responsible Artificial Intelligence: How to Develop and Use AI in a Responsible Way*. Cham: Springer International Publishing.

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.

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.

Erman, E. 2016. "Global Political Legitimacy beyond Justice and Democracy?" *International Theory* 8 (1): 29–62.

Erman, E. 2019. "Does Global Democracy Require a World State?" *Philosophical Papers* 48 (1): 123–53.

Erman, E. 2020. "A Function-Sensitive Approach to the Political Legitimacy of Global Governance." *British Journal of Political Science* 50 (3): 1001–24.

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).

Frey, C. B. 2019. *The Technology Trap: Capital, Labor, and Power in the Age of Automation*. Princeton: Princeton University Press.

Future of Life Institute. 2017. *Asilomar AI Principles*. https://futureoflife.org/ai-principles/ (accessed October 11, 2021).

Gabriel, I. 2020. "Artificial Intelligence, Values and Alignment." ArXiv:2001.09768 [Cs].

Google. 2021. *Our Principles*. Google AI. https://ai.google/principles/ (accessed October 11, 2021).

Habermas, J. 1996. *Between Facts and Norms: Contributions to a Discourse Theory of Law and Democracy*, trans. W. Rehg. Cambridge: MIT Press.

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

High-Level Expert Group on AI. 2019. *Ethics Guidelines for Trustworthy AI*.

Jobin, A., M. Ienca, and E. Vayena. 2019. "The Global Landscape of AI Ethics Guidelines." *Nature Machine Intelligence* 1 (9): 389–99.

Karabarbounis, L., and B. Neiman. 2014. "The Global Decline of the Labor Share." *The Quarterly Journal of Economics* 129 (1): 61–103.

Kolodny, N. 2014. "Rule Over None I: What Justifies Democracy?" *Philosophy & Public Affairs* 42 (3): 195–229.

Korinek, A., and J. Stiglitz. 2017. *Artificial Intelligence and its Implications for Income Distribution and Unemployment*. National Bureau of Economic Research, Inc. NBER Working paper 24174.

Larsson, S., and F. Heintz. 2020. "Transparency in Artificial Intelligence." *Internet Policy Review* 9 (2): 1–16.

Larsson, S. 2020. "On the Governance of Artificial Intelligence Through Ethics Guidelines." *Asian Journal of Law and Society* 7 (3): 437–51.

Livingston, S., and M. Risse. 2019. "The Future Impact of Artificial Intelligence on Humans and Human Rights." *Ethics & International Affairs* 33 (2): 141–58.

Marr, B. 2020. "What is GPT-3 and Why is it Revolutionizing Artificial Intelligence?" *Forbes*. https://www.forbes.com/sites/bernardmarr/2020/10/05/what-is-gpt-3-and-why-is-it-revolutionizing-artificial-intelligence/ (accessed October 08, 2020).

Nozick, R. 1974. *Anarchy, State, and Utopia*. New York: Basic Books.

O'Neil, C. 2016. *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*. New York: Crown Publishers.

OECD. 2019. *OECD Principles on Artificial Intelligence*. Paris: OECD. Technical report.

Powers, T., and J. Ganascia. 2020. "The Ethics of the Ethics of AI: Mapping the Field." In *The Oxford Handbook of Ethics of AI*, edited by M. Dubber, F. Pasquale, and D. Das. Oxford: Oxford University Press.

Rawls, J. 1999. *A Theory of Justice, rev. ed.* Cambridge: Belknap Press of Harvard University Press.

Robertson, A. 2021. "Lawmakers Propose Ban on Police Buying Access to Clearview AI and Other Data Brokers." *The Verge*. https://www.theverge.com/2021/4/21/22395650/wyden-paul-fourth-amendment-is-not-for-sale-act-privacy-data-brokers-clearview-ai (accessed October 11, 2021).

Ronzoni, M. 2009. "The Global Order: A Case of Background Injustice? A Practice-dependent Account." *Philosophy & Public Affairs* 37 (3): 229–56.

Russell, S. 2019. *Human Compatible: Artificial Intelligence and the Problem of Control*. New York: Viking.

Scharpf, F. 1999. *Governing in Europe: Effective and Democratic?* Oxford: Oxford University Press.

Schiff, D., J. Biddle, J. Borenstein, and K. Laas. 2020. "What's Next for AI Ethics, Policy, and Governance? A Global Overview." In *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society*, February 2020, 153–8.

Schönberger, D. 2019. "Artificial Intelligence in Healthcare: A Critical Analysis of the Legal and Ethical Implications." *International Journal of Law and Info Technology* 27 (2): 171–203.

Solon, O. 2018. "Google's Robot Assistant Now Makes Eerily Lifelike Phone Calls for You." *The Guardian*. https://www.theguardian.com/technology/2018/may/08/google-duplex-assistant-phone-calls-robot-human (accessed October 10, 2020).

Sparrow, R. 2007. "Killer Robots." *Journal of Applied Philosophy* 24 (1): 62–77.

Spiekermann, K., A. Slavny, D. V. Axelsen, and H. Lawford-Smith. 2020. "Big Data Justice: A Case for Regulating the Global Information Commons." *The Journal of Politics* 83 (2): 577–88.

Stiglitz, J. 2006. "Give Prizes Not Patents." *New Scientist* 21–21.

Valentini, L. 2012. "Assessing the Global Order: Justice, Legitimacy, or Political Justice?" *Critical Review of International Social and Political Philosophy* 15 (5): 593–612.

Valentini, L. 2014. "No Global Demos, No Global Democracy? A Systematization and Critique." *Perspectives on Politics* 12 (4): 789–807.

Wachter, S., B. Mittelstadt, and L. Floridi. 2017. "Why a Right to Explanation of Automated Decision-Making Does Not Exist in the General Data Protection Regulation." *International Data Privacy Law* 7 (2): 76–99.

Wakabayashi, D., and C. Metz. 2018. *Google Promises Its A.I. Will Not Be Used for Weapons*. The New York Times. June 7 2018. https://www.nytimes.com/2018/06/07/technology/google-artificial-intelligence-weapons.html (accessed October 19, 2020).

Whittaker, M., K. Crawford, R. Dobbe, G. Fried, E. Kaziunas, V. Mathur, S. M. West, R. Richardson, J. Schultz, and O. Schwartz. 2018. *AI Now Report 2018*, 1–62. https://ainowinstitute.org/AI_Now_2018_Report.html (accessed October 11, 2021).

Zuiderveen Borgesius, F. J. 2020. "Strengthening Legal Protection against Discrimination by Algorithms and Artificial Intelligence." *The International Journal of Human Rights* 24 (10): 1572–93.