Technology in Society 62 (2020) 101283

Contents lists available at ScienceDirect

Technology in Society

journal homepage: http://www.elsevier.com/locate/techsoc

A shallow defence of a technocracy of artificial intelligence: Examining the political harms of algorithmic governance in the domain of government

Henrik Skaug Sætra

Østfold University College, Remmen, 1757, Halden, Norway

ARTICLE INFO

Keywords:
Artificial intelligence
Democracy
Technocracy
Legitimacy
Participation

ABSTRACT

Artificial intelligence (AI) has proven to be superior to human decision-making in certain areas. This is particularly the case whenever there is a need for advanced strategic reasoning and analysis of vast amounts of data in order to solve complex problems. Few human activities fit this description better than politics. In politics we deal with some of the most complex issues humans face, short-term and long-term consequences have to be balanced, and we make decisions knowing that we do not fully understand their consequences. I examine an extreme case of the application of AI in the domain of government, and use this case to examine a subset of the potential harms associated with algorithmic governance. I focus on five objections based on political theoretical considerations and the potential *political* harms of an AI technocracy. These are objections based on the ideas of 'political man' and participation as a prerequisite for legitimacy, the non-morality of machines and the value of transparency and accountability. I conclude that these objections do not successfully derail AI technocracy, if we make sure that mechanisms for control and backup are in place, and if we design a system in which humans have control over the direction and fundamental goals of society. Such a technocracy, if the AI capabilities of policy formation here assumed becomes reality, may, in theory, provide us with better means of participation, legitimacy, and more efficient government.

1. Introduction

Artificial intelligence (AI) has proven to be superior to human decision-making in certain areas. This is particularly the case whenever there is a need for advanced strategic reasoning and analysis of vast amounts of data in order to solve complex problems. Few human activities fit this description better than politics. In politics we deal with some of the most complex issues humans face, short-term and long-term consequences have to be balanced, and we make decisions knowing that we do not fully understand their consequences. One step at a time, AI has conquered realms of great complication, such as chess, Go and the game of StarCraft. While these are *games,* some now suggest that the same techniques of machine learning used to conquer such games could also be used to conquer politics. One example of this is the suggestion that an 'AI Economist' based on reinforcement learning and multi-layered agent-based simulations could make better tax policies, improving *productivity* while simultaneously reducing *income inequality* [1].

In this paper I examine an argument in favour of involving AI in political decision-making. This idea is based on the premise of AI's stipulated ability to make better decisions than humans in certain areas, and the premise that we ought to implement the best policies possible. This is a hypothetical situation in which technologies such as the AI economist just discussed are mature, ready for real life applications, and functional on a large scale. This is not a technical treatise about AI but an analysis of the political theoretical implications of AI in politics. As such, I introduce a thought experiment called the *magical decision box*, which is used to test a set of political theoretical arguments for and against an AI technocracy.

There are several approaches to the analysis of the effects of computer decision-making in politics. Algorithmic governance is not an entirely new phenomenon, and various harms following the use of AI to govern various aspects of human affairs have been examined in much detail. There are, amongst other issues, concerns related to issues of *privacy and surveillance*, *bias and inequality*, *transparency and procedure*, and *freedom and autonomy* [2,3].

Sadowski and Selinger [3] provide a taxonomic tool for technocracy, in which they propose categories for the *domains*, *means*, and *harms* of technocracy. I limit my examination to the *political* harms of technocracy. Furthermore, I consider only a subset of the phenomena that are analysed in the literature on algorithmic governance, as I focus on the

E-mail address: Henrik.satra@hiof.no.

https://doi.org/10.1016/j.techsoc.2020.101283
Received 28 October 2019; Received in revised form 23 May 2020; Accepted 3 June 2020
Available online 8 June 2020

technocracy in the domain of *government*, and mainly through the means of *mandates*. In addition, my case is one in which the technocracy of AI is based on systems with *low transparency* and *high degrees of automation* [4]. This is a consideration of an extreme case of a technocracy of AI, and this case serves well to highlight the political harms to which I limit my analysis. These choices make this article a *partial* contribution to the overall consideration of the dangers of technocracy. It is, however, a central part, as these dangers have not been sufficiently examined, particularly not in the setting of an extreme case of AI technocracy as I consider here.

The political harms I examine are mainly based on ideas from the discipline of political theory. The first is that human beings might *require political activity* in order to thrive. If that is the case, the drawbacks of technically 'better' decisions might outweigh the benefits. The next problem is that *legitimacy might suffer* if people are taken out of politics. I will here examine several approaches to this problem, and how it might be possible to overcome it through regulation and popular participation in guiding computer politics. We could also argue that computers should not make decisions that affect citizens' *lives and wellbeing*. The fourth problem is that we lack *transparency* when we rely on decisions that are beyond the capacity of human reasoning. Without such transparency we will be left to conduct politics on the basis of trust in machines. A final issue is that of *accountability* when AI is involved in making political decisions.

However, none of these objections are strong enough to fully debunk the argument in favour of employing AI in politics, if a technocracy of AI is implemented with a satisfactory system of control, mechanisms of backup, and a participatory element. Such a technocracy, if the AI capabilities of policy formation here assumed becomes reality, may, in theory, provide us with better means of participation, fair and impartial political outcomes, and more efficient government resulting in benefits for most individuals and society in general.

Should we actually choose to implement a technocracy of AI, politics will, and must, change in ways that might both shift its focus and, perhaps paradoxically, revitalise democracy and popular participation. In an age where algorithms are increasingly governing human action, there is a need to openly and carefully consider a) whether we as societies wish to allow the various forms of algorithmic governance and b) how such forms of governance lead to a need to discuss fundamental political institutions and arrangements. My main goal with this article is to highlight the implications of developments already underway. As Sadowski & Selinger [3] point out, governments are increasingly opening the door for a) private companies solving public problems, and b) the use of AI in government. de Sousa et al. [5] also show how AI is now employed in all areas of governments, while recommending that implementation of AI in the public sector should be preceded by a 'debate with society' about such applications. *AI creep* – in the sense that AI slowly and almost unnoticeably is applied to ever new areas – if left unchecked, might take us to a situation in which we *have* technocracy of AI without either political design or democratic input. I propose that we examine the key questions in advance, and prepare our societies for a future in which AI has the role we desire it to have in politics.

I will first develop the concept of traditional technocracy, in order to make sense of the logic and potential benefits involved, and the problems associated with traditional technocracies of humans. I then show how AI can, and in many cases already have, taken the role of a technocrat. Such use of AI is associated with somewhat similar, but not identical, risks, to those of traditional technocracy. The combination of the concept of technocracy and the use of AI lets me propose a set of objections to an AI technocracy based on the idea of political harms, and these are first presented and discussed before I move on to the discussion of the implications they have, and the potential for a legitimate and workable technocracy of AI.

2. Technocracy, expert rule and democracy

Schumpeter [6] argues that if we want results that are satisfactory to people at large, and judge our governments by such a criterion, government *by the people* 'would often fail to meet it'. For us to understand what an AI technocracy means, we must understand the basic ingredients of technocracy. Sadowski & Selinger [3] are right to call for an improved understanding of what technocracy is, and they intend for their taxonomy to be a 'conversation starter' on the topic of technocratic uses of technology. While they argue that it is an ill-defined concept, I posit that the solution to this problem is to draw upon the existing literature on the subject from the field of political science. Furthermore, I reject their implicit position that technocracy is by definition a *bad* thing, associated with harms only, and no benefits. This is based on one of the fundamental points in this article, which is that democracy, and specific forms of democracy, cannot be assumed to be ultimate goals at the outset of an evaluation of the effects of democracy. If we start with such an assumption, anything that challenges democracy is by necessity bad. A more fruitful approach is to take an agnostic stance to the value of democracy. Firstly, this lets us perform an *honest* and open evaluation of alternatives. Secondly, it enables us to strengthen and revitalise democracy by showing *why* it is good, and not merely stating that it *is*.

2.1. Technocracy and expert rule

Technocracy is a term that in daily parlance refers to the rule of *experts*. More specifically, I use the term to refer to a state whereby experts in *science* and *technology* – not experts in *politics* – wield power. Technocracy is thus differentiated from other forms of rule by the few, such as epistocracy and meritocracy. *Epistocracy* is a form of government whereby those with a particular knowledge of politics rule, and is often considered a rule of the *wise* – those with knowledge of 'politics, history, economics' [7,8]. This idea is old, and famous political philosophers such as Plato [9] and John Stuart Mill [10] have proposed variations of it. *Meritocracy*, on the other hand, is often linked to mental and cognitive abilities, regardless of specific knowledge of a particular subject [11,12].

Technocracy has been defined in various ways, one being Meynaud's [13] description of 'a system of governance in which technically trained experts rule by virtue of their specialist knowledge and position in dominant political and economic institutions' [13]. The experts, then, become *technocrats*, and have authority based on technical expertise [14]. Putnam [15] uses the word technocrat to describe *proponents* of technocracy, whilst I use it to describe *technicians in power*.

There are several possible arguments in favour of technocracy, and I split them into positive and negative arguments. The *negative* arguments mainly focus on the shortcomings of people and democracy – usually implying that *the people* are unfit to rule, or that democracy itself produces bad outcomes. The positive arguments chiefly focus on the benefits that could be gained from expert rule, e.g. *efficiency*, *rationality* and *optimisation*. It is worth noting that technocracy usually flourish in times of panic and great crises, as the need for efficiency and optimal policies in such times tend to be more pressing, and obvious [3].

The negative arguments are mostly based on the idea that *human beings* are the main problem of politics, and that they are unfit to rule. People have *personal interests*, and they tend to pursue them, even when they are tasked with pursuing the *public* interest [16]. Furthermore, they have limited capabilities, in terms both of *how much* they can do and *how well* they can do things. Some of them have a greater capacity than others, however, and this is the foundation of elitist arguments for technocracy, for example as seen in Schumpeter [6]. Another class of negative arguments focuses on democracy as a system, as it is portrayed as *inefficient*, *unjust* and *prone to instability*.

The positive arguments are mainly based on achieving some form of rational optimisation in politics, and partly the belief that most problems, when properly understood, are *technical* problems amenable to the logic of statistical analysis and optimisation. However, there are also arguments in favour of non-democratic systems based on, for example, the idea that *markets* are the best way to organise *everything*. At least *almost* everything. And if we do need some form of government, it had better be small and kept out of the way [17].

A point of great importance is that the technicians do not have *all* power in a technocracy. While a technocratic society delegates much authority to their technicians, they are not given the power to determine the goals of society. The technocrats are not political or moral experts, and I do not consider a situation in which morality and the question of political values are reduced to *technical* questions amenable to the calculus of mathematics and logic. This means that the technocracy I discuss have a *political* apparatus in place for providing the direction for society, while the technocrats have the authority to determine and effectuate the policies deemed necessary to proceed in the desired direction. This is akin to the process described by Peters [18], where *technical* and *political* decisions are separated, and the technical issues are transferred to "independent agencies, bureaucracies and technocratic elites". Technical issues are also political, however, which is why I argue that we have a technocracy of AI when we delegate power to make technical decisions to computers. This is the argument made by Næss [19], who urges us to see the political nature of all things technical and to submit the technical to an evaluation in light of our *ultimate values*.

Fischer [20] writes that technocracy revolves around the desire to use experts with specific knowledge in positions of *political* power, for the sake of building the 'good society'. This, then, is not the technocracy I discuss, as politicians – or the *people* – are charged with determining *what* the good society is, while the technocrats are charged with making it reality in the most effective manner possible. This is in line with Sadowski & Selinger's [3] picture of technocrats as influential, but still largely controlled by elected politicians. This implies a break with the understanding of technocracy as an approach in which 'political realities play no role' [21].

2.3. Arguments against technocracy

Technocracy may have a bad reputation, but it is here given a blank slate and the chance to prove any worth it may have. While Sadowski & Selinger [3] aim to arm us to 'combat technocracy', I claim that we must first understand if, and why, it *should* be combatted. Before moving on the use of AI in politics, I present the main traditional arguments against technocracy. However, as some of these arguments become less relevant if human technocrats are replaced by AI, I do not examine these in great detail, unless they *remain* relevant, and thus become part of the objection in section 4.

The arguments *against* technocracy are also based on various principles, and I briefly consider arguments based on the inherent values of democracy, human beings' needs and finally the inherent flaws of technocratic government itself.

Estlund [7] discusses objections to *epistocracy*, which are partly applicable to technocracy as well. The *demographic* objection is based on the fact that giving those with, say, university degrees more voting power means giving privilege to certain groups. The second objection – the *bias* objection – is a version of the former, adding the claim that the privileged classes, races etc. will exert a certain bias that is unacceptable [7,22]. These arguments constitute the first class of arguments against technocracy, and I label them the *technocrats are people, too* arguments. One of the arguments in favour of technocracy is that people are flawed, but some are less flawed than others. No-one is free of flaws, however, and people can never be assumed to have totally overcome irrationality and bias. But can computers? I return to this issue in Section 6. It is also worth noting at the outset that problems of bias and discrimination are not exclusive to the use of algorithms, and in evaluating algorithmic governance, we must examine whether algorithmic bias is in fact worse than existing problems associated with human bias.

A second argument against technocracy is based on a concern for *legitimacy*. It posits that technocracy is bad because the process itself does not provide legitimacy (only participatory politics does so). This constitutes one of the objections discussed in Section 5.

Thirdly, people might object to technocracy because they believe a *deliberative* process will provide *better* outcomes. This particularly comes from the deliberative-democracy camp [23]. Opposed to these arguments would be ones inspired by Janis [24] and Hobbes [25] who both emphasise the dangers inherent in democratic, or group, decision-making. It must be added that human technocrats are also liable to *group-think*.

The fourth argument against technocracy might be called a *decentralisation* argument. It is the Hayekian notion of *dispersed knowledge* [17]. No human being or small group of human beings can ever amass the kind of knowledge that each individual has about their own affairs. The only effective way to utilise such knowledge, Hayekians would argue, is to give power to the markets [17].

The problems of technocracy are not insignificant, but as this examination shows, there is reason to ask if a technocracy of AI may in fact be *better* than a technocracy of human beings. This stems from the fact that several of these arguments more easily applies to human beings than to computers. However, as Janssen & Kuk [21] shows, human faults are often re-found in the workings of our algorithms, in the form of, for example, bias and discrimination.

2.2.1. The beginnings of a shallow defence of a technocracy of AI

The title of this article promises a *shallow defence of a technocracy of AI.* This refers to the argument I will construct in the first part of the article, which consists of three proposition that lead to the conclusion that we *should* employ AI in politics, and erect a technocracy of AI. This argument will then be subjected to a set of objections to such a technocracy in section 4. The first two components of the shallow defence follow form the preceding consideration on technocracy and politics.

First of all, politics can be understood as a process aimed at implementing the best possible policies. But what policy is considered *best* can, of course, only be decided once we know what criteria to apply. Thus, the *first* and *fundamental* purpose of politics is to develop and elucidate what *fundamental moral values* a society is based on. Only then can politics as we know it in the day-to-day workings of society take place. And only then can we properly assess the value of technology [19]. This is the first building block in the defence of an AI technocracy:

> Policies should be evaluated on the basis of the fundamental moral values of the society in question, and finding these values is the first purpose of politics.

Furthermore, if politics revolves around the question of finding the *best* policies, as Schumpeter [6] implies, we also have a second premise in the establishment of what will become the defence:

> The best policies in accordance with the evaluation discussed in the first premise should be implemented.

3. Artificial intelligence and political decision-making

Artificial intelligence is superior to human intelligence for analysis of large and complex problems involving the need for strategy, prediction of long-term effects and analysis of vast amounts of data. One example is playing chess and Go [26–29]. However, AI can beat humans at more than fun and games. Politics is complex, and it involves many considerations of notoriously uncertain short- and long-term consequences. As the literature on algorithmic governance shows, AI is *already* employed in many facets of government, as I show in section 3.1.

I will note at the outset that the potential benefits from AI in politics is heavily contested [4]. I agree with the idea that AI as of today is not some silver bullet that can be employed in order to create flourishing societies. However, as the many existing applications clearly show, there is undeniably a *potential* for the beneficial use of AI in politics. I will assume that some of this potential can be realised, and that we will continue to see certain improvements in the technologies involved. This means that I consider possible near-future technologies, and not speculative developments related to some form of singularity, etc. [29].

I will not discuss the technologies involved in great detail, but rely on the assumption that much of the future applications of AI is based on the technologies currently employed. de Sousa et al. [5] chart the technologies most often employed in public sector AI; various forms of machine learning are used, while artificial neural networks (ANN) is the most popular technique in terms of usage. *Artificial neural networks* (ANNs) imitate biological information-processing systems, and consist of artificial neurons that transmit signals to each other in reinforcement systems, for example [30]. *Deep learning* is our term for deeply layered neural networks, and this is the machine learning approach used by the AI economist I use as an example, as well as by Google's AlphaGo and AlphaZero [31,28,32,1].

3.1. AI and expertise in questions of policy

Algorithmic governance is a term describing the utilisation of algorithms both in ordering human action in general, and in traditional political structures [33,4]. As such, it is a very broad term, and as discussed in the introduction, I only focus on a subset of what is referred to in general as *algorithmic governance*. Using the taxonomic tools of Sadowski & Selinger [3], I limit my discussion to the use of AI in the domain of *government*, and mainly to a situation in which AI is given the authority to govern by *mandates* (not more informal *nudging*, or even less direct governing through *technological mediation)*.

According to the same typology, I mainly focus on a specific set of harms, labelled *political harms*. These are distinguished from *existential* and *discursive* harms [3]. Existential harms relate to the *reduction* of individuals through the datafication and behaviouristic approach to the control of human beings. Discursive harms involve the restriction of discourse through a domination by technocratic ideas. In this article, I do not focus on harms related to *privacy* and *surveillance* or *bias, unfairness and inequality*. A potential loss of *freedom* is another potential harm that falls beyond the scope of the article [2].

Issues of privacy and surveillance are of great importance, but I will in the following assume that AI can be built on data which is *not* individualized, and that it does not govern by individualized nudges or *micro-directives,* etc., based on detailed personality profiles [34,35]. Hypothetically, policy could also be based on machine learning in combination with simulations, such as in the example of the AI economist [1]. Related to issues of privacy are issues of agency, autonomy and freedom [2,4]. Another point is that increased used of data, and the translation of all aspects of human action into data is that such quantification by necessity involves multiple stages of human interpretation, challenging the *objective* veneer of big data-based decision making [21, 34,36]. Such issues are only partially considered in this article, and will thus remain important philosophical issues for further research.

Some of the potential harms that I do *not* focus on, fall in the categories of unforeseen and unintended consequences [4]. One such problem is *algorithmic bias* and resulting issues of *fairness* and *inequality* [5]. Furthermore, algorithms, partly because of problems associated with the training data, could also favour particular political ideologies, and they could reinforce discriminatory and other undesirable practices [21]. These problems are real, and important, and any technocracy of AI must implement mechanisms of algorithmic oversight in order to address such problems. However, I do not consider these challenges in detail in the present article, and employing the thought experiment presented in section 3.2 enables me to isolate the political harms I do focus on from the issues of algorithmic bias.

According to Sadowski and Selinger [3], political harms relate to people being disenfranchised and deprived of political power and influence. This relates to the potential *de-politicisation* effects of algorithmic governance [4]. In addition to this, questions of *transparency and procedure* are relevant to the analysis of political harms [2]. These concerns are examined in more detail in section 4.

With regard to the applications of AI, I posit that we have *already* moved a long way towards AI decision making, and this is an important point to keep in mind when we consider the technocracy of AI. As I will show, a condemnation of the technocracy of AI that I propose will in many respects simultaneously involve a condemnation of many current practices.

AI is now used in all areas of society: driving cars, guiding our missiles, trading stocks, helping us navigate new places, playing chess, recognising speech, speaking to us, finding our mates, forecasting the weather, etc. [37]. AI is even used to determine who gets bail and who gets loans and identify likely tax evaders [37,38]. All these applications of AI are relevant to an examination of algorithmic governance in general, but not all of them are in the domain of government.

de Sousa et al. [5] provide a thorough account of current applications of AI in the public sector. They find that it is employed in all areas of government already, while *general public service, economic affairs* and *environmental protection* is the areas most emphasized in the literature on AI in government. Examples of current applications are found in most functions of the public sector, such as public health, transportation, education, security, communications, and the actual examples involve, amongst others, systems of identification of risks, optimisation systems, systems for response, analysis and prediction [5]. Predictive policing is another area that has gotten a lot of attention, and AI is also used in, for example, the administration of immigration and in calculating social benefits [21,4]

Algorithmic control of traffic lights and speed limits is one low level example in the domain of transportation, where the utility of having *dynamic* systems based on machine learning could increase efficiency and utility [34]. If the *objective* of our traffic lights and speed limits is to a) optimise traffic flow while b) minimising casualties, AI could most likely improve a static system of pre-determined speed limits – allowing for increased speed in low danger situations and not having drivers wait unnecessarily at red lights. The idea is that similar benefits could be reaped in other domains of law and politics, such as tax policies [1]. Another example would be the use of AI to implement optimal strategies for pandemic response, which could involve using AI to predict and identify outbreaks and dynamically apply the optimal response based on a society's tolerance for risks weighed against other factors, such as economic costs.

In general, when we face optimisation problems, AI could have the potential to aid us. Most problems of politics *are* highly complicated optimisation problems. As long as we have a goal, AI can in theory take us there in an effective manner. Furthermore, it could continuously experiment, and could update its policies. Policies would be dynamic, always responding to changes in human preferences and behaviour. One recent example is the AI economist which can make tax policies, in a *simulation*, that both improve productivity *and* decrease inequality [1]. Another is the use of WeBuildAI, which through participatory algorithmic design managed to increase efficiency of a food donation transportation service while simultaneously improving the perceived fairness of the outcome – adjudicating *equity* and *efficiency* [39]. Such examples are limited and hypothetical, but, in combination with the many existing applications, they suggest that we could in time find us in a situation where such systems are applicable to real-world large-scale situations.

3.2. The magical decision box and whether or not to use it

One way to approach this issue without getting lost in the technical aspects of AI is to consider an analogy. I invite you to replace AI as we know it with a hypothetical device called the *magical decision box (MDB)*. This thought experiment requires you to accept that a society has discovered a box – origins unknown – that makes decisions. The box accepts input and constraints, it accepts goals and it provides answers.

Through rigorous tests the society in question has found that the box significantly outperforms human experts in making policies that optimise the goals it has been given. Thus far they have found no major errors. Despite their very best efforts, they have not been able to understand *how* the box works, and they cannot recreate (even in retrospect) the ways in which the box reasons. The decisions, however, are good, so should they employ it?

The MDB is, of course, different from AI in many ways. However, there are certain similarities that might elucidate the core issues involved in the objections to AI in politics. Like the MDB, AI is a 'black box' that we do not always fully understand. Even if we understand the principles involved in making advanced machine learning models, they will produce outcomes that we often cannot predict [21]. I will later return to the notion of explainable, or explicable, AI – also referred to as XAI [40,41].

The realist theory of democracy and both cognitive and behavioural science have shown many of the shortcomings of human decision-making. This brings me to the next proposition in the argument leading to the preliminary conclusion that artificial intelligence should be used to improve political decisions. This could come about in several ways, ranging from computers aiding human decision-makers to their *replacing* them. I here examine the latter possibility, which I call an AI technocracy.

We now have the option of letting computers make expert judgements for us. With the use of advanced machine learning, computers could make *better* decisions than human beings. Vogl et al. [42] argue that computers might improve decision-making by overcoming the human limitations of bounded rationality and information processing. Machines are better than us at chess and Go, and if this ability also applies to other well-defined strategic problems, should we not let computers decide for us? If we had the MDB, and knew that its results were good, would we use it?

3.2.1. *The third premise – AI supremacy*

The possibility of AI superiority in solving complex issues related to human beings and their societies is not out of the question. As my goal is to analyse the possible future impact of AI on politics, the exact probability of this happening, or the exact scope of AI superiority, is of less importance. For these reasons, we can consider AI to be the MDB for our present purposes. What matters is that for now we should accept the premise that AI *might* in the future play the part I am proposing here. If this does not happen, or *until* it happens, this paper can be considered null and void, and no harm should have occurred. To be clear, I do not believe AI can *currently* be the basis of a full technocracy such as the one I here describe.

> Artificial Intelligence is better than humans at finding and enacting the best policies in certain areas concerning science, engineering and complex societal and macroeconomic issues

With this, *the shallow defence of an AI technocracy* is complete. Firstly, the purpose of politics is to *find* the best policies, in accordance with some set moral basis. Secondly, we should *enact* the best policies we have identified. Thirdly, AI is better than human beings at creating and identifying the best policies. All this leads to the preliminary conclusion that AI should be given the power to discover and enact our policies.

4. The potential arguments against a AI technocracy

It seems non-controversial to assume that such a technocracy could in certain areas give us better decisions in terms of efficiency, utility or whatever other criteria we set as the computer's goal. Lee et al. [39] and Zheng [1] provides proofs of concept for the idea that AI could provide both efficiency *and* still be amenable to popular control and direction.

However, there are also several potential downsides, which will have to be weighed up against these benefits. I will here provide a schematic overview of some of the obvious objections to an AI technocracy, based on the *political* harms it may cause. They are: a) people might need politics, b) legitimacy is linked to democracy, c) AI is not capable of morality, d) we have an issue with transparency related to AI and e) AI decision-making involves problems assigning responsibility.

Note that I am not considering the objection that such a technocracy would *not* lead to better policies, as this very premise is part of the argument I am testing. If this premise fails, the argument in favour of technocracy crumbles, and there is little need for any of these objections.

Accepting the first three premises for now, I proceed to the political theoretical objections against the shallow argument in favour of an AI technocracy. In this brief examination of the objections I will first present each objection, and then some considerations concerning its possible shortcomings.

4.1. *Human nature and homo politicus*

According to Aristotle [43], humans are political animals. The best that they can achieve is to exercise their political capabilities and live a life of self-determination in a political society. Only then will they realise their true nature – their *telos.*

4.1.1. *The first objection – people's political nature*

> People need full political participation in order to be satisfied

This is, however, only one of the interpretations of Aristotle's view of humans and politics. Mulgan [44] shows that Aristotle can also be understood as supporting 'the withdrawal from politics, or at least reluctant participation'. In both *The Politics* and *Ethics* the philosopher's lifestyle is portrayed as being superior to the life of the statesman [44].

Another immediate objection would be to ask: Are people politically active in today's political systems? What will *really* change if AI makes our decisions, instead of bureaucrats? The 'political animal' argument is against *any* move from direct democracy towards democracy by representation or republicanism, and this move could be said to be a necessary evil if we are to have large-scale, complex societies.

This objection also relies on an agreement with Aristotle's hierarchical ordering of human activities. Other philosophies, such as hedonism, could easily portray a life of increased wealth and leisure as a potential improvement to a life material hardship and political activity. Philosophers such as F. A. Hayek have proposed that involvement in politics is *not* that important, and that *economic* liberty is what is essential to good societies [17]. Following this argument, he states that democracy is not necessarily the best way to rule, and that '… a democracy may well wield totalitarian powers, and it is conceivable that an authoritarian government may act on liberal principles' [17].

However, a crucial point of this article is that a technocracy of AI may in fact lead to a revitalization of popular participation and the accessibility of the political domain. Through new ways of participation, such as algorithmic co-creation, we could find ourselves in a situation in which most people have, and experience, *more* political power than they do today. This is examined in more detail in the next objection.

4.2. *Legitimacy and democracy*

Another objection is that people need to be involved, and that they need deliberation to take place if political systems are to have *legitimacy* [45,46]. Even if the *best* decisions were made by a dictator, his decisions would not be legitimate if the citizens had no political power. We thus distinguish between *legitimate* authority and *effective* authority. AI might be extremely effective, but might still lack legitimacy.

We can see this objection as consisting of two parts: a) policy should be *decided* by the populace, and b) all citizens should have equal political rights and a realistic chance of exercising them. The objection to technocracy would be that legitimacy is lost when the population is deprived of crucial political rights and the ability to take part in decision-making. This objection is based on a *procedural* view of legitimacy [47]. According to such a view, due *process* is the source of legitimacy, and deliberation is one aspect of such a process [45].

4.2.1. The second objection – legitimacy based on participation

> People will not deem a government in which they do not participate to be legitimate

The counterobjections are similar to those regarding the previous objection. We could first argue that an AI technocracy makes people *more* equal than they would have been in any democracy – equally free to participate in the moral debates about the direction of society, and equally deprived of participation in policy formation. We might also argue that such a system would enable people to understand *more* about politics, and that the barrier to understanding politics and taking part would be lowered. When human politics only revolves around fundamental moral values, it is easier to both understand and take part in this process.

Furthermore, not everyone agrees that legitimacy is gained through *participation.* For Hobbes [24]; legitimacy is achieved through the social contract, which ensures that a government – by one, few, many or all – is created to satisfy everyone's fundamental need, this need being *security* – sometimes labelled *liberty* [45]. König [48] even discusses how *algorithmic governance* bears some resemblance to Hobbes's apolitical *Leviathan*. If democracy contributes to better political outcomes, it is conducive to legitimacy. If it does not, it is not necessary [49]. The latter argument is the *instrumental* view of democracy.

Even in today's societies participation in politics is restricted to a select few, despite everyone having a *theoretical* chance of taking part. The size and complexity of our societies have led to the growth of bureaucracies, expert rule and indirect democracy. Going from this to an AI technocracy is perhaps not as radical as it might seem at first sight.

One obvious potential benefit of digitalisation and the increased capacity to communicate using digital tools is the possibility of directly involving far more people in political processes. The Pirate Party is an example of an actual attempt at this, and they have done very well in several countries [50].

If a technocracy of AI is to overcome this objection, participation in some form must be ensured. One way to implement this is to promote cocreation and 'democratizing algorithm' [21]. One such example is the WeBuildAI example, which shows several methods of popular participation in the creation and design of algorithms, and how this could lead to both increased participation and increased algorithmic awareness [39]. This also addresses a potential objection based on the idea that only democracy has the pedagogical effects ensuing from participation in politics. People *learn* from taking part in politics [46,51]. According to this argument, the beneficial effects participation has on people could outweigh the negative effects of having sub-optimal policies. However, if systems such as WeBuildAI allows for participation *and* efficiency, there is little reason to choose sub-optimality.

4.3. Machines, morality and human wellbeing

Another possible counterargument against the use of AI in politics is based on the argument that politics always involves questions of *values* and *morality*. This is not necessarily a problem per se, but it creates an obstacle for AI as soon as we run into people arguing that morality is solely a human endeavour. According to some people moral machines do not exist, and *human* qualities such as compassion and wisdom are crucial prerequisites for important decisions [52].

Resolving this question entails coming to some sort of agreement about how we define morality, and whether or not our definitions exclude machines. What we are concerned with here is whether or not machines can *act* morally and make moral decisions. In the terminology of ethics: whether or not they can be moral *agents* [53].

4.3.1. The third objection – morality

> Computers should not make decisions that affect people's lives and wellbeing

Some people believe that we can *train* a machine to become moral, e. g. by having a lot of people answer various questions on the morally superior decision in various instances of the trolley problem. The *Moral Machine* is one such endeavour that has attracted much attention [54]. It is an attempt to crowd-source the morality of people around the world in order to teach a computer to distinguish right from wrong, in order to find ethical principles to guide machine behaviour. This is an approach which could be related to the approach of WeBuildAI, in which social choice theory and popular participation is used to train algorithms [39].

Kurzweil [29] predicts that machines will gain human proficiencies and capabilities, including morality-related emotions. Maldonato and Valerio [55] and Scheutz [56] also discuss related topics, e.g. machines with *value systems* and the creation of machines with moral competences. Even if they do not gain anything akin to human morality, machines are becoming better at *understanding* human morality, as it can quantify, track and compare moral *bias*, for example*,* 'across cultures and over time' [57].

However, we could also approach this objection more pragmatically. If we decide that decisions that affect people's wellbeing require *moral capabilities*, we are *already* employing AI in situations where it does not belong. Self-driving cars could not then be allowed. Robots trading in markets would become highly problematic. Use of AI to determine whether or not criminal offenders should receive bail would be highly dubious.

AI is already involved in moral decisions in countless ways, and our current approach to this issue is to make someone other than the computer itself *morally responsible* for the decisions made. We will include the premise in the tentative discussion, but I will argue in the conclusion that this objection is better understood by way of handling objection 5 and the *attribution* of AI decisions.

Another obvious counterobjection would be that if machines cannot be moral, neither can they be *immoral.* To some people this situation would appear to be a huge advantage compared with the capacities of today's politicians. If indeed power corrupts, we should see it as an *advantage* that AI has no morally corruptible nature [58].

A different kind of objection would be that the important issue is not *morality* but whether or not machines are able to act *ethically*. If morality is right action based on some internal value system, then ethics are actions based on some external system. Machines' ability to act in accordance with humans' codes of ethics might be sufficient. Following this objection, one could also question the nature of *human* morality itself, and portraying morality as some vague and unscientific idea, plagued by no consensus, which is thus of little interest. We might even object that we have no test for the morality of our human politicians, so why should we implement one in order to keep robots away from power?

4.4. Transparency and the problem of understanding what we cannot produce ourselves

Human experts can explain their choices to the populace in a way that computers cannot – at least not yet. And even if they become better at explaining why certain choices are made, will we be able to understand? The reason for giving computers the authority to decide is that they are able to make evaluations and calculations that are *beyond human capacity*.

Transparency would then be an issue, and we might be left guessing why our AI technocrat made the decisions it made, in the same way as chess experts *guess* why the computer calls moves good or bad when analysing the world's best players. Santiago [59] emphasises that humans must have the last say in decisions made by AI. However, if we cannot understand the reasoning behind the decisions, transparency and control are hard to achieve [21]. *Opacity* is one of Danaher's [37] main objections to algorithmic governance.

4.4.1. The fourth objection – transparency

> AI is not transparent and thus not fully amenable to human control

Pressure is mounting, Wachter et al. [60] state, to make sure that AI becomes *explainable*, and thus transparent. Explainable AI (XAI) is the quest for machines that can make us understand their decisions and reasoning. Such a development could make us trust, understand and manage AI decision makers [40]. This is a field that receives a lot of attention, but until we have achieved this goal, we could imagine other ways of dealing with the problem.

Like with the magical decision box, which is by definition unfathomable, we must perhaps rely on control by goal achievement and observable result parameters with AI technocracy. If AI or the MDB make decisions that take us closer to our goals, do we really need to understand *how* they manage to do so? Similarly, if the bureaucrats make decisions beyond politicians' understanding, do we complain? We might simply recognise that the very reason politicians rely on experts, or AI technocrats, is *because* the latter make decisions that are unfathomable, owing to their being too advanced. Lack of understandability is the price for better decisions. In the words of Robbins [41], 'a principle of explicability for AI makes the use of AI redundant'. Janssen & Kuk [21] is right in arguing that a *full* understanding of how complex algorithms operate is most likely restricted to the 'happy few'. However, so is a full understanding of how many current systems of government operates.

Danaher [37] does not see this counterobjection as a valid reason for ignoring transparency issues. The fact that we have similar problems today, he argues, is not a reason to allow them (in more serious form) in algorithmic governance. I am, however, not arguing that an AI technocracy is an *ideal* system. I merely argue that it could be argued to be *better* than what we currently have. As such, the pragmatic counterarguments pointing out that an AI technocracy does not lead to a worsening compared to current affairs *is* of a great interest. If AI technocracy leads to a pareto improvement where efficiency is improved, while other aspects of political life remain the same, I will consider this desirable.

A related point is that opacity *could* provide certain benefits. If the exact workings of a system is not known, it is difficult to game the system [21]. However, the AI economist of Zheng et al. [1] allows for such gaming of the tax system, and find their policy maker capable of coping with it.

4.5. Who is responsible when a computer makes decisions?

The final objection is based on the problem of accountability. If a computer makes a decision, who is held accountable for this decision? Until we are willing to consider robots legal persons with rights and obligations, someone *else* must be held accountable for their actions. But who? Their authors, or perhaps their owners, who are responsible for using them?

The accountability gap is a term used to describe such issues, and it is a problem that becomes increasingly prominent as our technologies become more advanced and capable [60]. This complexity creates situations in which the creator of a machine cannot foresee the actions of his creation [61]. When this happens, some perceieve a gap between who acts (for example, a complex machine), and who is perceived as responsible (the creator).

4.5.1. The fifth objection – accountability

> Accountability regarding the consequences of political decisions must be clear, and it becomes less clear when AI makes decisions

Santiago [59] states that when businesses employ AI in decision-making, humans must have the final say. But *which* humans, when we are discussing AI technocracy? We are not discussing a political system in which bureaucracy is simply replaced by AI, but one in which the technocrat machines are given greater autonomy in terms of policy formation. Thus, the current system, whereby politicians are responsible for the decisions of their bureaucracies, will not work, as they will not be assumed to be in control of their technocrats as they now are of their bureaucrats.

We might not have any perfect solutions to this problem, but one might also argue that neither are popular modern political systems perfect. We do, however, have several options. One is to hold the *producers* of AI legally responsible for their artificial technocrats.

Another possibility could be to create a human council for machine control, tasked with making sure our technocrats do as we wish. Algorithmic 'constituents', Binns [62] argues, have a right to scrutinise and hold someone accountable for the exercise of algorithmic governance. As Janssen & Kuk [21] point out, few are *able* to exercise such control. Educating the population and organising such a council for machine control is a vital precondition for an AI technocracy, as I will return to later.

5. The pros and cons of an AI technocracy

The shallow argument in favour of an AI technocracy is based on three premises.

- *P1: Policies should be evaluated on the basis of the fundamental moral values of the society in question, and ascertaining these values is the first purpose of politics.*
- P2: The best policies in accordance with the evaluation discussed in the first premise should be implemented.
- P3: Artificial Intelligence is better than humans at finding and enacting the best policies in certain areas concerning science, engineering and complex societal and macroeconomic issues

I have argued that these premises could be used to construct a *shallow* defence of an AI technocracy. *Shallow,* because until we deal with the five important objections the argument is not complete, and AI technocracy is only one of several possible conclusions. The objections I consider are:

- O1: People need full political participation in order to be satisfied
- O2: People will not deem a government in which they do not participate to be legitimate
- *O3: Computers should not make decisions affecting people's lives and wellbeing*
- *O4: AI is not transparent and thus not fully amenable to human control*
- O5: Accountability regarding the consequences of political decisions must be clear, and it becomes less clear when AI makes decisions

5.1. The strength of the objections

Where, then, does this lead us? Firstly, if O1 *or* O2 is true, the argument in favour of technocracy is weakened. We might, however, introduce a counterobjection, to remove some of the sting of these two objections. If politics is really about deciding which moral values are important, and what we see as the *good society*, having a policy process whereby human beings take part in the formulation of these issues is compatible with a technocracy.

I will even argue that it is *necessary* in order to avoid a technocracy with no goal or direction. If this counterobjection is accepted O1 and O2 can be disregarded. However, O1 states that people require *full* participation, and this condition might not be satisfied by my proposed policy process involving only fundamental values. However, objection 1 could conceivable also *strengthen* the initial argument in defence of an AI technocracy. While Danaher [37] argues that algorithmic governance will tend to limit and reduce the possibilities of participation, my argument in favour of reorienting the goal of politics to fundamental values could make it more accessible.

O3 is an objection that *sounds* both true and reasonable, but let us consider the implications before accepting it. If we take it as it stands, AI must not be used in *any* area where human wellbeing is affected – not in determining loans, not in determining bail, and not even in helping us identify criminals, or in areas such as financial fraud. Furthermore, *positive* use of AI affects people's wellbeing just as much as use that we more often perceive to be problematic. This means that autonomous vehicles are most certainly out of the question, and even brake assist etc. must be got rid of.

The implications are absurd, which implies that we must subject the objection to close scrutiny. Some use of technology is unavoidable if we are not to reject all the assistance it currently provides us with. When we pursue this issue, we find that *all* technologies have moral and political *implications*, which means that the issue at hand is not really restricted to the use of AI. The important issue, I argue, is *accountability,* which will be discussed shortly.

The proposition that AI should not make decisions because the decisions are not transparent (O4) is a universal problem associated with letting those with the most knowledge make decisions. It is, however, possible to imagine a situation whereby decisions are *explained by virtue of the results they lead to*. *How* economists arrive at a new and better tax system is perhaps of less interest than their proposal showing what a new tax system will *lead to*. Politicians will not understand the technical aspects of the reasoning, just as they will not understand the processes that lead the AI from input to ideal policies. This problem is not restricted to the use of AI.

I argue that AI *is* transparent, but that we simply have a hard time understanding how known methods and known input lead to decisions that are better than those we could make ourselves. If nothing is *hidden* or *secret*, a decision is by definition transparent, and whilst further work on explainable AI will be beneficial, Objection 4 does not appear to be particularly strong.

This issue of transparency is a problem in two respects. It is a challenge to the legitimacy of an AI technocracy, and people might not be willing to take the *risk* involved in being ruled by something one cannot understand. With human politicians we realise we may be taking a degree of risk, as human fallibility is well known. It is, however, harder to identify the risks involved in computer decisions, and the fear that it will err in ways we will not be able to uncover, or correct, is different from the risk of being ruled by fallible leaders who are just like us. Furthermore, if we develop and apply methods of participation and co-creation, we could increase general algorithmic awareness. This might hypothetically lead to most people having a *better* understanding of how our societies function than they do today.

Accountability issues are the fifth objection, and I have proposed that either a council for oversight or the producers of the AI systems could be held accountable. It is not a problem per se if we refuse to give AI moral agency, and thus moral responsibility and accountability. An AI technocracy need not be unlimited and arbitrary. It could be a constitutionally limited AI whereby a council – small or large – of men is tasked with giving the AI both a *direction* and a *goal.*

5.2. Discussion

AI is in certain respects like a modern-day version of Leibniz's God — the good architect who must be assumed to have created the best possible version of the universe [63]. Had we known what God knows, we would realise that the universe is as good as it can be, despite occurrences of what we perceive as evil. The argument works as long as you believe in an almighty and good God, and this kind of trust must thus be transferred to AI.

This takes us back to the magical decision box. If we knew it had always produced good decisions in the past, should we employ it, even if we did not understand how it worked? We would be in a situation similar to the one Leibniz considers when examining whether or not the world is good. God is good, and God created it, thus the world must be good, even if it may not always seem that way to us.

If God created the magical decision box, I argue that we *should* employ it as our technocrat. However, as my analysis suggests, we should perhaps not discard the use of AI as our technocrat even if it is *not* a magical and perfect decision box. There are, however preconditions that must be satisfied for such a technocracy to be able to overcome the objections I have presented. Furthermore, the defence of a technocracy of AI here presented is *shallow* also in the sense that not all objections and considerations have been considered. The future capabilities of AI are somewhat speculative, I have not gone into detail on issues of bias, discrimination, agency, and liberty, and it remains to be seen if the world of *wicked* political problems is truly amenable to the optimising hand of artificial intelligence [64]. Nevertheless, I do propose a defence that is not without value, as it suggests that the political harms of an AI technocracy are not insurmountable.

One of the preconditions for a technocracy of AI is that there are backup and control mechanisms in place, in case the AI systems should fail, or if we desire to scale back the use of AI. Control systems are necessary in order to supervise the performance of the systems, and particularly to monitor potential issues related to bias, discrimination, and various other unintended consequences. Furthermore, times of crises and rapid change have shown that machine learning systems can easily become confounded, and in such cases, there must be mechanisms in place for human intervention. This was shown clearly during the Covid-19 situation that led to problems for many machine learning models trained on 'normal behaviour' [65].

Putting AI in power involves, in a certain sense, putting the companies and individuals that create these systems in power [37]. Some people might argue that the creators are simply creating algorithms – just doing maths and creating abstract and valueless systems. Dwork & Mulligan [66], however, point out that *any* such system is bound to reflect some of the implicit ideas and values of its creators. Whilst this might occur by pure accident, *despite* the intentions of the designers, Gillespie [67] points out that most of the creators of such systems have clear profit-maximisation goals. Such goals and accompanying incentives might make it difficult for us to regard AI as an equivalent to Leibniz's God. Griffy Brown et al. [68] note the *profit-driven technocratic* ideology, which is different from the disinterested and scientific technocratic *ideal* I previously described. These considerations show that a robust system of algorithmic supervision is required whenever the AI in question is a human creation, and not a magical decision box.

Another precondition is some form of political control of the core values and direction of society. One possibility is to simply scale back, but largely keep, the current system of representative democracy. This involves a limitation of the scope of politics, as the technocrats have authority over many issues of policy. Another, and more radical possibility, is to pursue some of the *new* modes of popular participation and popular government provided by new technologies.

Much research focuses on the possibilities of using technology to facilitate participation, and Janssen & Kuk [21] call for *democratizing algorithms.* I have focused on the WeBuildAI example in this article, which points toward the potential for massive popular participation in the shaping of algorithms *and* the potential of increasing popular understanding of AI [39]. This latter option, while more radical, also shows how a technocracy of AI could in fact *increase* people's opportunities for political participation, rather than the opposite. While doing so, it could also meet the demands for involving various stakeholders and social groups in political decision making. If this is not enough, increasing broad participation in political also allows us to reap the benefits of dispersed knowledge and decentralised markets, as emphasized by Hayek [17].

6. Conclusion

The benefits of better political decisions are considerable. I have shown that AI *is* already being employed in the public sector in order to improve political analysis and decisions, and that it will become ever more capable. If we agree that politics should have *good policy* as a goal, and if we do not consider our current democratic order as a *given* and inviolable good, we start opening the door to the idea that AI should have political power.

In this article I have focused on the potential widespread application of AI in the domain of government, and I have examined an extreme case in which AI is given the authority to form policy. This is done through systems characterised by low transparency, high complexity and high degrees of automation. As such, the situation is rigged to trigger the alarms of those concerned with democracy, participation and the *political harms* of algorithmic governance.

I have chosen to focus mainly on a set of harms perceived as vital from a political science stance, while acknowledging that other important harms are both important and necessary for acquiring a complete picture. However, all the harms from and objection to an AI technocracy cannot not be discussed in the required detail in one article. As such, my account is a partial one which must be seen in combination with detailed accounts of the harms related to bias, discrimination, autonomy, etc.

However, the political harms are vital, and not as often discussed in detail as some of the other harms, which makes this an important part of a full understanding of the consequences of algorithmic governance. Through five objections, based on participation, legitimacy, non-morality of machines, transparency and accountability, I have reached the conclusion that an AI technocracy is *not* derailed by such objections. If, that is, certain preconditions are met.

An AI technocracy capable of overcoming the objections discussed must involve some way of making sure that humans retain the power to define the fundamental goal of politics – that of deciding what our fundamental values are, and how we imagine the *good society*. If humans retain this role, and make sure that the process is a democratic one, both participation and legitimacy might be preserved. Furthermore, I have argued that the non-morality of machines is not a problem if we assign responsibility either to the proposed democratic council of values and AI control or to the producers of AI.

AI is in some respects like the magical decision box. If we had such a device, we could employ it, and over time we would might to trust that its decisions were good – like Leibniz's God. If the MDB were a mystical device, not one created by humans, there would be, I argue, no reason for any particular group of people to object to its rule. While there are many reasons to be wary of an AI technocracy of human origins, it is time we take the possibility seriously. Firstly, because we are already seeing the application of AI in many areas of the public sector, without a full and open debate about the desirability and implications of such a development. Secondly, we must take the possibility seriously, as I have shown that it has the potential to *revitalise* democracy, increasing public participation in politics, *and* provide more efficient outcomes.

Author statement

Henrik Skaug Sætra: All parts of the article.

Acknowledgements

I wish to thank the anonymous reviewers for their invaluable comments and recommendations that greatly improved the article.

References

[1] S. Zheng, A. Trott, S. Srinivasa, N. Naik, M. Gruesbeck, D.C. Parkes, R. Socher, The AI Economist: Improving Equality and Productivity with AI-Driven Tax Policies, 2020 arXiv preprint arXiv:2004.13332.

[2] Danaher, J. (Forthcoming). Freedom in an age of algocracy. In Shannon Vallor (ed.), Oxford Handbook of Philosophy of Technology. Oxford, UK: Oxford University Press.

[3] J. Sadowski, E. Selinger, Creating a taxonomic tool for technocracy and applying it to Silicon Valley, Technol. Soc. 38 (2014) 161–168.

[4] C. Katzenbach, L. Ulbricht, Algorithmic governance, Internet Policy Review 8 (4) (2019) 1–18.

[5] W.G. de Sousa, E.R.P. de Melo, P.H.D.S. Bermejo, R.A.S. K, A.O. Gomes, How and where Is Artificial Intelligence in the Public Sector Going? A Literature Review and Research Agenda, Government Information Quarterly, 2019, p. 101392.

[6] J.A. Schumpeter, Capitalism, Socialism and Democracy, Routledge, 2003.

[7] D. Estlund, Why not epistocracy?, in: Naomi Reshotko Desire, Identity and Existence: Essays in Honor of TM Penner, 2003, pp. 53–69.

[8] D. Estlund, Democratic Authority, Princeton University Press, Princeton, 2008.

[9] Plato, The Republic of Plato, Basic Books, 1968.

[10] J.S. Mill, On Liberty, Penguin Books, London, 1985.

[11] K.J. Arrow, S. Bowles, S.N. Durlauf, Meritocracy and Economic Inequality, Princeton University Press, 2000.

[12] M. Young, The Rise of the Meritocracy, Routledge, 2017.

[13] J. Meynaud, Technocracy, Free Press, New York, 1969.

[14] D. Bell, The Coming of Post-Industrial Society, Basic Books, New York, 1973.

[15] R.D. Putnam, Elite transformation in advanced industrial societies: an empirical assessment of the theory of technocracy, Comp. Polit. Stud. 10 (3) (1977) 383–412.

[16] D.C. Mueller, Public Choice III, Cambridge University Press, 2003.

[17] F.A. Hayek, The Constitution of Liberty, The University of Chicago Press, Chicago, 1960.

[18] B.G. Peters, The Politics of Bureaucracy, Routledge, London, 2014.

[19] A. Næss, Ecology, Community & Lifestyle, Cambridge University Press, Cambridge, 1989.

[20] Frank Fischer, Technology and the politics of expertise, Sci. Soc. 57 (2) (1993) 246–249.

[21] M. Janssen, G. Kuk, The challenges and limits of big data algorithms in technocratic governance, Govern. Inf. Q. 33 (3) (2016).

[22] K. Lippert-Rasmussen, Estlund on epistocracy: a critique, Res. Publica 18 (3) (2012) 241–258.

[23] J. Elster, The market and the forum: three varieties of political theory. Debates in Contemporary Political Philosophy, Routledge, 2005, pp. 335–351.

[24] T. Hobbes, Leviathan, Basil Blackwell, London, 1946.

[25] I.L. Janis, Groupthink: Psychological Studies of Policy Decisions and Fiascoes, vol. 349, Houghton Mifflin, Boston, 1982.

[26] M. Campbell, A.J. Hoane Jr., F.H. Hsu, Deep blue, Artif. Intell. 134 (1–2) (2002) 57–83.

[27] T. Chouard, The Go Files: AI Computer Wraps up 4-1 Victory against Human Champion, 2016 [Blog post], https://www.nature.com/news/the-go-files-ai-computer-wraps-up-4-1-victory-against-human-champion-1.19575.

[28] Google, AlphaZero: Shedding New Light on the Grand Games of Chess, Shogi and Go, 2020. Retrieved from, https://deepmind.com/blog/article/alphazero-shedding-new-light-grand-games-chess-shogi-and-go.

[29] R. Kurzweil, Superintelligence and singularity, in: S. Schneider (Ed.), Science Fiction and Philosophy: from Time Travel to Superintelligence, Wiley-Blackwell, Chichester, 2015, pp. 146–170.

[30] C.M. Bishop, Pattern Recognition and Machine Learning, Springer, New York, 2006.

[31] Google, AlphaGo, Retrieved from, https://deepmind.com/research/case-studies/alphago-the-story-so-far, 2020.

[32] R.S. Sutton, A.G. Barto, Reinforcement Learning: an Introduction, MIT press, 2018.

[33] J. Danaher, M.J. Hogan, C. Noone, R. Kennedy, A. Behan, A. De Paor, H. Felzmann, M. Haklay, S.M. Khoo, J. Morison, M.H. Murphy, Algorithmic governance: developing a research agenda through the power of collective intelligence, Big Data & Society 4 (2) (2017) 1–21.

[34] M. Ma, The Law's new language, Harv. Int. Law J. (61) (2020).

[35] H.S. Sætra, When nudge comes to shove: liberty and nudging in the era of big data, Technol. Soc. 59 (2019) 101130.

[36] H.S. Sætra, Science as a vocation in the era of big data: the philosophy of science behind big data and humanity's continued part in science, Integr. Psychol. Behav. Sci. 52 (4) (2018) 508–522.

[37] J. Danaher, The threat of algocracy: reality, resistance and accommodation, Philosophy & Technology 29 (3) (2016) 245–268.

[38] J. Tashea, Courts Are Using AI to Sentence Criminals. That Must Stop Now. Wired, 2018. Retrieved from February 02, https://www.wired.com/2017/04/courts-using-ai-sentencecriminals-must-stop-now/.

[39] M.K. Lee, D. Kusbit, A. Kahng, J.T. Kim, X. Yuan, A. Chan, A.D. Procaccia, WeBuildAI: participatory framework for algorithmic governance 3, CSCW), 2019, pp. 1–35. Proceedings of the ACM on Human-Computer Interaction.

[40] D. Gunning, Explainable artificial intelligence (xai), Defense Advanced Research Projects Agency (DARPA), 2017.

[41] S. Robbins, A misdirected principle with a catch: explicability for AI, Minds Mach. (2019) 1–20.

[42] T.M. Vogl, C. Seidelin, B. Ganesh, J. Bright, Algorithmic Bureaucracy: Managing Competence, Complexity, and Problem Solving in the Age of Artificial Intelligence.

Complexity, and Problem Solving in the Age of Artificial Intelligence, 2019. February 1, 2019).

[43] Aristotle, The Politics, Penguin Books, London, 1981.

[44] R. Mulgan, Aristotle and the value of political participation, Polit. Theor. 18 (2) (1990) 195–215.

[45] B. Manin, On legitimacy and political deliberation, Polit. Theor. 15 (3) (1987) 338–368.

[46] J.-J. Rousseau, The Social Contract, Penguin Books, London, 1968.

[47] F. Peter, Political legitimacy, in: E.N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Summer 2017 Edition, 2017. Retrieved from, https://plato.stanford.edu/archives/sum2017/entries/legitimacy.

[48] P.D. König, Dissecting the Algorithmic Leviathan: on the Socio-Political Anatomy of Algorithmic Governance, Philosophy & Technology, 2019, pp. 1–19.

[49] J. Raz, Ethics in the Public Domain: Essays in the Morality of Law and Politics, Clarendon Press, Oxford, 1994.

[50] J.L. Beyer, The emergence of a freedom of information movement: anonymous, WikiLeaks, the Pirate Party, and Iceland, J. Computer-Mediated Commun. 19 (2) (2014) 141–154.

[51] A. Tocqueville, Democracy in America, The Library of America, New York, 2004.

[52] J. Weizenbaum, Computer Power and Human Reason: from Judgment to Calculation, W. H. Freeman, San Francisco, 1976.

[53] J. Nolt, Environmental Ethics for the Long Term, Routledge, New York, 2015.

[54] E. Awad, S. Dsouza, R. Kim, J. Schulz, J. Henrich, A. Shariff, J.-F. Bonnefon, I. Rahwan, The moral machine experiment, Nature 563 (7729) (2018) 59.

[55] M. Maldonato, P. Valerio, Artificial entities or moral agents? How AI is changing human evolution. Multidisciplinary Approaches to Neural Computing, Springer, Cham, 2018, pp. 379–388.

[56] M. Scheutz, The need for moral competency in autonomous agent architectures. Fundamental Issues of Artificial Intelligence, Springer, Cham, 2016, pp. 517–527.

[57] P. Schramowski, C. Turan, S. Jentzsch, C. Rothkopf, Kristian Kersting, The moral choice machine, Frontiers in Artificial Intelligence (3) (2020).

[58] J. Dalberg-Acton, J. Dahlberg-Acton, J. Figgis, R. Laurence, Letter to bishop mandell creighton, april 5, 1887, Historical essays and studies 504 (1907).

[59] T. Santiago, Influence The Executive Function Of Business Leaders?, in: *Muma Business Review* 3 AI Bias: How Does AI, 2019, pp. 181–192.

[60] S. Wachter, B. Mittelstadt, L. Floridi, Transparent, explainable, and accountable AI for robotics, Science Robotics 2 (6) (2017).

[61] A. Matthias, The responsibility gap: ascribing responsibility for the actions of learning automata, Ethics Inf. Technol. 6 (3) (2004) 175–183.

[62] R. Binns, Algorithmic accountability and public reason, Philosophy & technology 31 (4) (2018) 543–556.

[63] G.W. Leibniz, Discourse on metaphysics. Philosophical Papers and Letters, Springer, Dordrecht, 1989, pp. 303–330.

[64] B.W. Head, Wicked problems in public policy, Publ. Pol. 3 (2) (2008) 101.

[65] W.D. Heaven, Our Weird Behavior during the Pandemic Is Messing with AI Models, 2020. MIT Technology Review May 16th, https://www.technologyreview.com/2020/05/11/1001563/covid-pandemic-broken-ai-machine-learning-amazon-retail-fraud-humans-in-the-loop/.

[66] C. Dwork, D.K. Mulligan, It's not privacy, and it's not fair, Stan. L. Rev. Online 66 (2013) 35.

[67] T. Gillespie, The politics of 'platforms' 12, New media & society, 2010, pp. 347–364, 3.

[68] C. Griffy-Brown, B.D. Earp, O. Rosas, Technology and the good society, Technol. Soc 52 (2018) 1–3.