Climate change and the political pathways of AI: The technocracy-democracy dilemma in light of artificial intelligence and human agency

Mark Coeckelbergh [a,*], Henrik Skaug Sætra [b]

[a] *University of Vienna, Universitaetsstrasse 7 (NIG), 1010, Vienna, Austria*
[b] *Østfold University College, B R A Veien 4, 1757, Halden, Norway*

A B S T R A C T

It is generally thought that artificial intelligence (AI) has a significant impact on politics and democracy. Meanwhile, the technology is also often hailed a solution to key societal and environmental challenges. It raises questions regarding, for example, how we can and should deal with climate change. This article links and discusses these issues by putting them in the context of a technocracy-democracy dilemma and by using the concept of critical junctures. Then it identifies two political pathways of AI and critically discusses their underlying assumptions. This offers a useful framework for further discussion of the relations between AI, climate change, and democracy, and enables the examination of important issues for the politics of technology, such as the role of human expertise vis-à-vis artificial intelligence, the problems raised by techno-solutionism, and the question at what level of governance AI and climate change should be addressed.

1. Introduction

As citizens, we like to think that our actions – or at least the actions of the most powerful among us – shape the ordering and future development of our societies. Rightly so, but in this article, we argue that these developments tend to proceed along certain paths that are laid out and prepared by technology. The notion that technology is political is nothing new, and technology has always shaped social, economic, and political developments [1–3]. Certain points in time, however, are characterized as critical junctures—periods in which human agency is key to correcting, guiding, or shaping the long waves of historical development [4–6]. Now is such a time, as the unprecedented level of data availability and progress in artificial intelligence (AI) challenge existing perceptions of the pros and cons of different political systems. This is particularly pressing when seen in combination with the current challenges related to environmental sustainability, and climate change in particular. There is no shortage of severe warnings about where we are headed if drastic changes are not made [7], and frustration continues to rise about politicians' inability to tackle the problems. The calls for eco-fascist solutions [8], such as movements like Extinction Rebellion[1]

trying to disrupt societies, and young people striking for the climate (in particular, "Fridays for Future"[2]) highlight the social and political tensions generated by climate change. Meanwhile, the physical realities of increased temperatures and increasingly frequent natural disasters underscore the real and pressing cause of these tensions. The many calls for change have one thing in common: they focus on human agency and responsibility but do not foreground technology's crucial role in causing the challenges we are facing—nor its central role in some of the solutions proposed for our current predicament. While it is true that humans always play an important role with regard to technology (as users, developers, maintainers, regulators, and so on) and that, as we can learn from the introduction of other emerging technologies, the success of a new or emerging technology depends on human systems [9] or rather human-technical or socio-technical systems, technologies themselves are not politically neutral [10] and often have pervasive societal effects that exceed human intensions and control.

In this article, we focus on how to identify the political tendencies of AI and how to understand and see the political pathways being prepared. We do so through an engagement with the concepts of democracy and technocracy; and in particular, the tension between them in the

---

context of the challenge of dealing with climate change. First, AI has the potential to improve democratic politics in several important ways. This can involve, for example, enabling improvements in opinion formation, deliberation, consensus formation, decision-making, and participation [11–13]. Since purely human decision-making processes dealing with climate change have so far proven insufficient, and arguably have a tendency to get stuck, AI may significantly help to improve those processes [14]. Second, AI is increasingly heralded as a technology that makes democracy obsolete. Here, the idea is that AI can in large part replace human political decision-making [14]. Such views tend to be based on the idea that AI, by not being dependent on democratic decision-making, would improve political decisions—for example, decisions about how to mitigate and adapt to climate change. For example, through increased monitoring and surveillance, the creation of digital twins of society and its inhabitants, and the use of unparalleled processing capacity and the power of data analysis, we can in theory improve just about all aspects of our societies—and effectively deal with climate change. But at what cost? This article aims to highlight and discuss the known, but understudied, political tendencies of AI in light of climate change and related environmental global challenges. By illustrating two paths that are currently emerging, we explore the likely consequences of going down either of these, and discuss the tensions between democracy and technocracy.

We begin by situating our analysis in the philosophy of technology and science, and technology studies (STS) combined with theories of political and economic development. This allows us to identify why and how AI is political. We also use social and political theories aimed at understanding the interaction between human agency, technology, and other historical institutional forces. We then proceed to develop the two key political pathways of AI: we explore how AI generates new possibilities; and create scenarios in which AI either strengthens and improves democracy or fuels a technocracy, with an eye to the most important consequences in either case. We proceed to a discussion of the nature of the choice we are facing at the critical juncture we are at. We argue that there are obvious benefits from either case discussed above, but that it is also crucial to discuss whether or not we should accept a path-dependent political development—and more generally, whether we should accept the assumptions made in these scenarios. While the emergence of AI as major political determinant is suggestive of certain futures, we must not lose sight of potential alternative futures in which humans are not left with a choice of paths determined by technology, but in which we instead start from a fundamental analysis of the current state of society's relation to technology and how we would like to proceed in the current global situation.

2. Technological and social change

In the history of ideas, the role of technology in society has often been undertheorized. This changed in modern times, when people noticed how new (industrial) technology reshaped society, significantly and often dramatically influencing people's daily lives. Therefore, social theory in the nineteenth and first half of the twentieth centuries tended to conceptualize the relation between technology and society in a rather deterministic way. According to technological determinism, history follows an inevitable course driven by technological innovation, which determines the social structure and cultural values [15]. A well-known view to which the term has been ascribed is that of Marx, who argued that production relations (and therefore production technology) shapes social relations within society. In such views, technology is social and political in the sense that it determines the basic structure of society. A contemporary variant of technological determinism is the transhumanist view that there will be a technological singularity or that other technology-determined futures will unavoidably happen. Consider, for example, Kurzweil's or Bostrom's view [16,17].

The opposite view is technological instrumentalism, which sees technology as a mere instrument and as politically neutral [18]. This also relates to what Winner refers to as the "technical orthodoxy", involving both human understanding and control over consequences, and technology neutrality [19]. Both technological determinism and technological instrumentalism have been heavily criticized by theorists in the philosophy of technology and media and in STS. Critics of the former have highlighted the non-instrumental and non-intended effects of technology. Earlier, Heidegger and McLuhan; and today, for example, Ihde and Feenberg. The latter have argued that technology and society are intertwined and co-evolve. For example, in his work on the nexus between philosophy of technology with STS, Langdon Winner [2] has famously argued that technologies are political, but that this should neither be conceptualized in technological determinist nor social determinist terms. Technologies and society interact in particular histories of artifacts, but things could have been otherwise.

Moreover, in economic and social theories of change a lot of attention has been paid to critical turning points in history. For example, Schumpeter's concept of economic development is about discontinuous change through creative destruction in the flow of the economy through the introduction of a new product, a new production method, the opening up of a new market, the conquest of a new source of raw materials, or the new organization of an industry. All of which are, in his view, initiated by the entrepreneur. As a result, the flow is disrupted and development and innovation happen [20]. One might also consider Kuhn's paradigm shifts, which introduced the idea of discontinuous change in the history and sociology of science.

Today, critical juncture theory holds that there are certain critical turning points that alter the course of evolution of an entity (e.g., a species or a society). Critical junctures can be defined as "a period of significant change, which typically occurs in distinct ways in different countries (or in other units of analysis) and which is hypothesized to produce distinct legacies" [21]. In these periods, future trajectories are produced and to a certain degree, locked in. This gives rise to the companion concept of path dependence, which originates in economics [22,23].

The role of agency in critical junctures has been the object of sustained attention in historical institutionalism because these are relatively short periods in time in which there is a comparatively greater scope for changing developmental trajectories and shaping the future of society. For example, the constraints of strong institutions is partly what creates path dependencies, and it is consequently important to see how human agency can be instrumental in overthrowing, replacing, or creating new institutions; and in choosing or forming new paths [5,24]; for example, uses the concept of critical juncture in his political economy work. He has argued that the introduction of neoliberalism in Iceland was a critical juncture and that the political opposition has not managed to intervene in a way that would generate a regime change. He also writes of real critical junctures as something that can be created, and thus, further foregrounds politics and human agency in the face of the constraints of various institutional forces.

3. Twin junctures: climate change and AI

Environmental sustainability is widely recognized as an important issue that societies and humanity at large must address. Climate change is often seen as the biggest challenge of our times.[3] Scientists warn that temperatures are rising and that extreme weather events are occurring more frequently. Glaciers are melting and the global sea level is rising. In order to deal with these issues, political action and decisions are urgently needed at all levels in order to reduce the emissions of greenhouse gases related to human activities. In this domain, the claim is that a critical juncture has been reached. This is often related to various tipping points (i.e., points after which we lose control [25]. If we do not

---

[3] See for example https://www.eea.europa.eu/themes/climate/climate-change-is-one-of.

sufficiently mitigate climate change now, it seems that things will get out of hand and massive global warming and all of its effects will be unavoidable. It is even often claimed that we have already passed such tipping points.

The notion of critical junctures is often tied to political institutions. This also applies to the climate challenge, as the state system – the organization of human societies into sovereign states – and established forums for international collaboration heavily influence progress, or lack thereof [26,27]. We will argue that the challenges we currently face with regard to climate change represent a critical juncture related to opportunities for challenging the established political pathways and institutional legacies in order to protect our natural environment. Climate change is already a source of political controversy, and it is expected that failure to mitigate and adapt to climate change on a global level will lead to more social and political unrest and an increased risk of violent intrastate conflict [7]. In turn, this can be expected to have significant consequences for both the depth and success of regional and global cooperation and the viability of democracy in a world where authoritarianism is increasingly represented and prominent.

At the same time, another ongoing critical juncture might be occurring in the area of technological and scientific development: AI and data science enable significant leaps forward, especially through the widespread use of machine learning and the increased availability of data. AI is rapidly proliferating, and it is increasingly being used in all areas of business and government. We could be heading past a tipping point here as well. This relates to Collingridge's dilemma [28], which explains that new technologies can be easily regulated but have unknown implications—and thus, the reasons for regulating them are unclear. However, once technologies mature, their consequences are often quite clear but attempts to regulate them often fail. This suggests that technologies can mature past a tipping point, after which future paths are cemented and the political implications of the technologies are difficult to avoid.

We might consequently find ourselves at a twin juncture of sorts—one in which several paths related to the natural environment and technological development are simultaneously relatively open, and both have great consequences for the political paths we now shape and partly choose. AI is increasingly being discussed in terms of sustainability while AI solutions for sustainability abound [29]. The key drivers in the twin juncture can be related to ongoing policy discussions related to the need for a green transition amid a digital transition [30]. began discussing what he referred to as a politics that unites the "green" (i.e., environmental policy and the green economy) and the "blue" (i.e., digital policy and the digital service economy). The interaction and relationships between the two have increasingly become the focus of policymakers, particularly within the European Union. The European Commission has made the twin transition concept a key part of its strategic planning and analysis [31], and Muench et al. [32] describe both transitions as "political priorities" and show that while they can be mutually reinforcing, there are also important tensions between them. For example, digital technology can help engender innovation and growth, but these technologies are simultaneously associated with their own environmental costs.

A particularly interesting twin juncture or meeting point of the green and digital transitions is the relation between AI and climate change. The use of AI seems particularly needed when it comes to global issues such as the governance of climate change. This range of technologies itself raises climate concerns; for example, due to the energy use of models and the servers that hold the data, AI can accelerate the change through increased greenhouse gas emissions [33]. However, more important to the present discussion is that AI can also help to mitigate climate change. It can help to process data on temperature change and carbon emissions, predict extreme weather events and energy use, transform transportation systems, monitor oceans, and assist geoengineering. The latter uses are more controversial. For example, some have proposed direct air capture as a pathway towards decarbonization

(for a proposal that also takes into account policy see Ref. [34]. But AI can also be used in relation to humans. AI is already used in marketing and advertising to steer human behavior. It seems tempting to also use it to steer – or nudge – humans and their societies in a climate-friendly direction [35,36], or use AI to augment human decision-making in various ways [37]. However, there are dangers and disadvantages related to this, which will become clear once we discuss different options and scenarios.

Yet AI does not only impact the natural sciences but also the humanities and activities such as politics. For example, large language models (LLMs) can now be used to create text while political decision-making can be assisted by AI. The rise of ChatGPT and other generative language models, for example, has generated concerns that it "hijacks democracy"[4] through its potential effectiveness in lobbying. Generative AI—the broader term encompassing AI systems that generate content not just in the form of text but also images, video, sound, etc., has implications for individuals (micro level), groups (meso level), and society (macro level). Sætra [38] maps some of the reasons to worry about generative AI on each of these levels, as shown in Fig. 1.

Sætra argues that such systems are problematic because they could influence politics and democracy through, for example, the generation of unlimited amounts of content aimed at changing public opinion and perception, fake news, and deepfakes. In addition, they change power relations and can be used for manipulative purposes, and such systems require energy and machinery in both the development and deployment phases, which entails environmental costs [33,39].

However, this technology also comes with potential positive political potential. For example, some researchers are examining how LLMs can be used to foster agreement among people with diverging beliefs and opinions [12,13]. Through such applications, we see that AI could potentially aid in deliberation and consensus formation, which are key areas of interest for proponents of democracy.

We describe both the positive and negative potential of AI as "potential" because we want to challenge the determinist position in which we have no control over our future trajectory. We consequently emphasize human agency and the need to take meaningful action to make sure that we guide future developments toward the right exit in our current critical junctures. Focusing first on AI, we propose to distinguish between at least two options or paths when it comes to using AI in politics, which we call "AI-augmented democracy" and "AI-driven technocracy."

4. AI-augmented democracy

A possible first path is to use AI to enhance democracy without replacing humans in decision-making processes: AI-augmented democracy. Climate change has proven to be a severe challenge for traditional democratic decision-making processes, and some have argued that strengthening the deliberate processes in democracies is key for effective climate action [40]. AI can, in theory, help large communities to communicate, meet, and even deliberate more effectively. This could be crucial to engendering a better common understanding of the challenges that climate change presents, and the potential solutions.

Various concepts and methods have been proposed by technical researchers and political theorists. For example, liquid democracy is a form of direct democracy in which voters also have the option to delegate their vote. This can be done via digital platforms and algorithms (e.g., Ref. [41]. Another example is research by DeepMind, which has proposed using deep reinforcement learning to find economic policies that people will vote for (Koster et al., 2022). In an investment game, AI discovered a mechanism that won the majority vote by optimizing for human preferences. Moreover, in political theory, there has also been

---

[4] https://www.nytimes.com/2023/01/15/opinion/ai-chatgpt-lobbying-democracy.html.
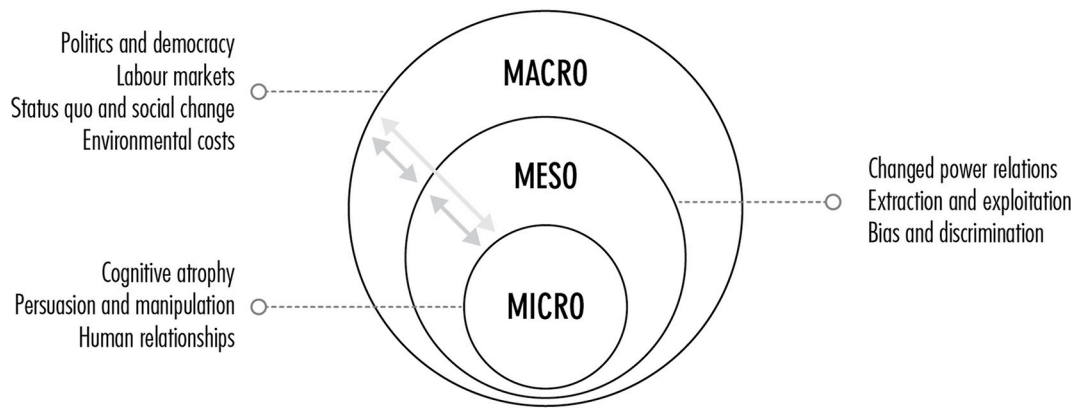
Fig. 1. Potential harms of generative AI [38].

work done using AI for democracy. For example, it has been proposed to use language models to find agreement among citizens with diverse preferences [12]. Landemore [42] has argued that AI can help with a form of mass deliberation (i.e., mini-publics) by improving the quality of deliberation in various ways. In mini-publics, randomly selected citizens engage in deliberation. AI can help to improve this deliberation by playing the role of facilitator (e.g., timekeeper), translator, fact checker, organizer of arguments, tracker of the exchanges, measurer of overlap and quality, and visualizer of where the group stands. Others have also noted how the principles of deliberative democracy can be used to foster responsible innovation and societal control over AI [43].

In these examples, AI is used in a highly political context while humans retain a central and meaningful role in democratic decision-making processes. However, since Plato and Bacon, there is also the autocratic, and in particular, the technocratic temptation: why let people decide at all given that they often get it wrong and that there are experts (Plato), and why not use science and technology to steer society in a better direction, creating a brave new world (Bacon)?

5. AI-fuelled technocracy

The technocracy temptation should not be taken lightly or underestimated. There are seemingly good reasons for relying on experts and expertise—and for having AI make political decisions. In theory, AI could more effectively identify valid mitigation and adaptation strategies for addressing climate change. Based on this insight, if we should decide to also give AI systems the authority to bypass cumbersome democratic processes that prevent effective action, some might argue that we could much more effectively achieve the goals related to climate change mitigation and adaptation. This is a possible second path: AI-fuelled technocracy.

Sætra [14] has argued that there are good reasons to seriously consider a technocracy of AI. In "A Shallow Defence for a Technocracy of AI," he tests the conclusion that we can and should use AI more actively in political decision-making based on these 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*

The argument presented in the article is hypothetical and preemptive, and its main purpose is to test whether we have sufficiently strong counter-arguments against such a technocracy.

A similar argument can also be framed in terms of political cybernetics. Following Norbert Wiener's seminar work on cybernetics in 1948 [44], many political scientists gradually became enamored with the idea of understanding political systems as flows of inputs and outputs, and in which these outputs served as feedback that was returned to the system. Influential examples include Karl W. Deutsch's "The Nerves of Government" [45] and David Easton's "A Systems Analysis of Political Life" [46]. Enchanted by the new possibilities provided by the emergence of sophisticated (at the time) computer systems and simulations, the 1960s were a time of great optimism about the potential for computational social science and a more complete understanding and control of political systems. However, the limits of a computational approach to politics were increasingly emphasized by the end of the 1960s. By the time Winner [47] published his review of Deutsch's abovementioned book, it read somewhat like a eulogy for an interesting movement within political science that had begun its descent to an end.

While it eventually did fall out of favor, political cybernetics is arguably set for a comeback as big data and AI have developed to such a degree that the unfulfilled ambitions of the political scientists of the 1960s have become more attainable. Consequently, this variety of cybernetics is intimately linked to a technocracy of AI as massive amounts of data are used to understand and control various forms of social systems, with or without meaningful human control [48].

Technocracy is attractive when dealing with climate change because ordinary citizens might not have the expertise necessary to address it, and even human scientists might not understand everything. AI can discover patterns in big data and make more accurate predictions—thus fueling/making better decisions. Nevertheless, such a technocracy is problematic for a number of reasons [49]. Sætra [14] summarizes the main problems in the form of five objections:

- *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*

These are all seemingly good reasons to reject a technocracy of AI. However Sætra [14], argues that when we examine them more closely, we quickly realize that existing political systems (e.g., not idealized and hypothetical forms of liberal democracy) must also be rejected based on many of the same objections, leaving us with the practical problem of having to choose between two non-ideal solutions. Consequently, it is important not to compare a pessimistic (i.e., realist) account of technocracy with an optimistic (i.e., ideal) version of democracy because

real political systems are arguably always an example of the non-ideal [50]. The real question, then, is whether the end result of evaluating both a technocracy of AI and an AI-augmented democracy yields more favorable results despite them both being associated with negative aspects.

However, even if such systems would be successful in terms of their results, one may object that the technocratic use of AI threatens to violate the enlightenment ideal and value of human autonomy. By taking decisions out of citizens' hands, and potentially even manipulating them, they do not take it seriously that people want to decide for themselves about their lives, communities, society, etc. The Kantian recommendation to think for oneself cannot be implemented: AI does not think, and human autonomy is not respected. Opposed to this, however, are the ideas of liberal theorists such as Friedrich A. Hayek, who stressed how economic liberty is what matters most for meaningful control over one's own life, while democracy was perhaps of secondary importance in this respect [51].

Another objection is that such approaches assume a view that comes close to technological determinism, at least in a normative form: the belief that technology should shape society. The term "techno-solutionism" [52] is sometimes used. This is the idea that our problems are technical and that technology can consequently be used to fix them all, including the political ones. But this is, we argue, impossible.

6. Discussion: choosing a path

The advantage of the first path is that humans still decide. AI merely helps the decision-making process, which is always challenging when it comes to mass deliberation. This is still in line with deliberative democracy ideals and democracy in general. AI is a technological aid, a crutch for democratic decision-making. Therefore, this path also allows for the integration of AI with all kinds of proposals by political scientists to enhance democratic political decision-making (e.g., mini-publics). This path is also more likely to find democratic support because humans are still at the center. This means that people will not have the feeling that AI is taking over, which is a common fear that has so far been fuelled mainly by science fiction.

The advantage of the second path is that there is guaranteed involvement of expert knowledge via AI. AI can deal with much more data than humans can. It can discover patterns that humans might overlook—and keeping Sætra's argument in mind, it may well be able to design better policies in a particular area. This could be interesting for dealing with climate change and other complex societal and global problems that require the identification of patterns in a lot of data, and demand more challenging coordination at the global level. Giving political agency to AI, rather than using it as a tool, can mean that societies and humanity as a whole can deal more effectively with these problems.

However, both paths also raise serious problems. Even in the first path, AI may influence citizens' decision-making rather than being a mere tool. Because of its intended or unintended effects, this use of AI may result in the manipulation of citizens, which would go against human freedom and autonomy, as Coeckelbergh [35] has argued when discussing using AI for mitigating climate change. In particular, there are at least three risks. First, AI-mediated deliberation may interfere with how citizens make up their minds, which could shape the outcome of the decision-making. For example, Coeckelbergh [53] has argued that AI may influence belief formation and belief revision, including political beliefs. And the Cambridge Analytica case,[5] in which data from millions of Facebook profiles was harvested and analyzed in order to influence voters, shows how using AI via social media can influence the outcome of an election. Second, in so far as AI already offers much of the reasoning, organizes the discussion, and proposes decisions; deliberation that has been "optimized" or "enhanced" by AI may be much

shallower or not provide the learning outcomes people gain from "true" or "full" deliberation in which they have to argue and discuss from scratch, and creatively think of new options. Hence, there might be a qualitative difference in the nature of the deliberation when it is mediated, facilitated, and streamlined by AI. Thirdly, the entire AI-mediated decision-making process lacks transparency and might be intentionally manipulated by those who control the technological systems.

Moreover, seen from the perspective of the second path, there is no reason to limit the agency of AI to solving specific problems (e.g., improving decision-making on climate change). AI could do much more. It could be used to optimize all aspects of our lives and it could be given full political power. So, why limit its powers and influence? On this path, a problem arises when humans do not agree with AI. For example, political and legal principles such as political liberty and the rule of law could seem threatened if AI were to decide to take away the liberty of some people in order to achieve a goal (e.g., dealing with climate change)—or it could make decisions about people's lives and liberty without a transparent legal process. Given the current problems with bias and discrimination in which AI plays a role, it is also questionable how the second path might lead to justice and equality. Are there enough guarantees of basic liberal-democratic principles such as liberty, justice, and equality? And what if AI were to decide that humans do not want these? This path consequently entails a real risk of increasingly authoritarian, and possibly totalitarian, political systems.

There's an inherent problem with the first path because while deliberative democracy could in principle also be done at a transnational level, current proposals seem to be geared toward improving national decision-making. Thus, it ignores the possibility of AI-assisted decision-making at a global level, which is much needed with regard to climate change—a problem that exists at a global, planetary level. For example, would democratic tools and methods such as AI-assisted mini-publics work at all on a global level? The second path solves this problem but with the danger of a totalitarian system in which the entire planet is ruled by AI. This involves not only the destruction of freedom and democracy but could also potentially lead to the end of humanity. In response, one could argue, as most big tech companies such as DeepMind do, that AI needs to align with human goals. That sounds right, but what if AI decides that genocide or ecocide is the most effective way of reaching one or more of these human goals at the planetary level? One could also argue that humans need to control the AI. But then how can the advantages of having AI make the decisions be retained?

6.1. Rejecting technology's paths and the limits of technology

The previous discussion might be seen as presenting a dilemma concerning AI and democracy. Both paths are problematic, and it might seem that we have to choose between them. Luckily, there is no need to accept either path—and hence, no need to accept the dilemma as a given.

Let us articulate some of the (problematic) assumptions that underlie both paths.

First, both paths assume that technology will solve the whole issue (e.g., climate change). They are premised on an exaggerated and unwarranted faith in the potential of technological fixes—what is often referred to as "techno-solutionism." But there are also other views and other ways of framing the AI technocracy problem. In particular, one could argue that instead of relying entirely on AI, we also need human expertise. Then, the main question should not just be what AI can and should do for democracy but also what is and should be the relation between AI expertise and human expertise in a democracy. That is a much more complex question and one that we have hardly begun to ask, let alone address. More generally, it should be acknowledged that technology alone cannot fix these problems and that humans and human expertise are always needed.

Second, both paths tend to be based on technological determinism.

They assume that the use of AI in democracy is unavoidable and that we must choose between different ways of applying AI. But the use of AI in democracy is not a necessity; we can also make other choices, which do not (centrally) involve the use of AI or other technologies. Moreover, the impact of AI on society and democracy cannot be easily predicted and may not be straightforward at all given that technologies always have unintended consequences—positive or negative.

This brings us to our third point. While this discussion is mainly focused on intended effects, we can learn from the philosophy of technology that AI, in addition to having beneficial intended effects, will also and unavoidably have unintended side effects. Consider again the points about manipulation and whether the means that AI selects in order to achieve the (human-agreed) goals do not lead to a totalitarian dystopia or even the end of humanity. This means that both paths again might not achieve what they were intended to, especially if the goal included the preservation of democracy.

Fourth, there is the question of the relation between politics and ethics. For example, Floridi when he writes about the green and the blue (see earlier), argues that ethics should take precedence over politics. Based on ethics, one could reject both paths regardless of their contributions to democracy or to solving planetary problems. This brings us back to the point about human autonomy, which can be seen as not only an appeal to a political principle but also an ethical value. However, there are also other values.

Fifth, there are the challenges related to the states system and the need for global cooperation or governance [54]. We have mentioned that the first path is framed in a way that assumes only the national level; while in theory, the second path may lead to a planetary implementation of a world state. Both are highly problematic in their own ways. Since climate change is a global problem because a stable climate is a global public good [55], we need to tackle it at the global level. However, we can imagine several ways to achieve global governance of climate change, each with its own strengths and weaknesses. The most obvious solution is the one adopted thus far, namely sovereign states with competing interests coming together to negotiate and agree on how to solve a common problem. This solution respects the sovereignty of states and peoples and has its strength in its realism and awareness of conflicts of interest and international anarchy. A different way to image the interplay of states in an anarchical system is to imagine them as actors in an international society. This view is represented by the English school in international relations; and for example, by Hedley [54]. The idea of a society of states highlights how informal relations, norms, and shared values greatly influence and could potentially provide the foundations for closer and more effective collaboration and governance on a global level. However, as critics would be quick to point out, such an approach to global governance would be prone to collapse once powerful actors found it to no longer be in their interest to sustain this society. Finally, one might image a solution in which international collaboration is superseded by a supernational entity with the authority and power to enforce global solutions [26]. This solution would give rise to great ethical and practical objections. Many would object to both the desirability of discarding sovereign states and the states system with nation-states and the practical feasibility of implementing an effective political system at a global level. AI and climate change challenge both national and global governance structures, and we should be open to considering different forms of governance at all levels.

Finally, even if AI were to be aligned with human goals (see again DeepMind's project and Koster et al. [13], human goals should not be taken as a given. Human goals can be mistaken. For example, we all rely on the current consumption society and contribute to that goal, but this is harmful to people and the planet. If we just let AI follow its own course, even if we were fortunate enough that it would advance human goals in alignment with human goals, there's the inherent issue that some human goals are highly problematic. For example, maximizing consumption is unlikely to contribute to dealing with climate change given the energy and pollution it causes—not to mention the depletion of natural resources, which harms the environment in a way that threatens its ecosystems; and in the end, humanity. Human activity is excessive, and a decrease in human activity and the human population is argued to be both compatible with and necessary for the flourishing of nonhuman and human life [56]. One can only imagine the implications and reactions that would follow if the AI system in charge of politics ended up with a similar diagnosis of our problems and proceeded to implement the required policies to achieve such a decrease in human activity.

7. Conclusion

After contextualizing our topic of the relation between AI and politics, and in particular democracy, within (a) the problem of technological and social change and (b) more specifically the twin policy junctures of AI and climate change, we have outlined two political pathways of AI: paths of using AI for politics and democracy and for solving global challenges such as climate change. Both paths have their advantages and disadvantages. AI can and probably should be used for augmenting democratic human decision-making, but for the reasons outlined in this paper there should be limitations. In the end, there might be side effects that do not contribute to achieving the goal or that have other effects that are politically or ethically unacceptable. The opposite, letting AI decide (without having humans control and influence it), is much more dangerous because it threatens fundamental values from the very start.

However, we have also indicated a number of reasons why the dilemma presented in the first part of our paper is misleading. These reasons are relevant to the discussion about AI and democracy that is slowly starting now in political theory, policy, and the tech world. However, it has also brought us to a number of important and perhaps more fundamental topics relevant to the relation between AI and politics in general; the relation between AI expertise and human expertise, technological solutionism and determinism, human goals that can be mistaken and the unintended side effects of AI, the relation between ethics and politics, and the level on which global challenges such as climate change can and should be addressed and governed. More work is needed in these areas if we want to move toward a politically and ethically responsible use of AI in politics, particularly in and for democracy.

Author statement

There are no competing interests.

There is no use of generative AI.

The work has not been published previously.

Both first author/corresponding author and second author worked on the conceptualization, writing, revision, and editing of the manuscript.

No funding sources involved.

Data availability

No data was used for the research described in the article.

References

[1] L. Winner, Autonomous Technology: Technics-Out-Of-Control as a Theme in Political Thought, MIT Press, Cambridge, 1977.
[2] L. Winner, Do artifacts have politics? Daedalus 109 (1) (1980) 121–136.
[3] J. Ellul, The Technological Society, Vintage Books, New York, 1964.
[4] I. Jonsson, The Political Economy of Innovation and Entrepreneurship: from Theories to Practice, Routledge, 2015.
[5] I. Jonsson, Economic crisis and real critical junctures – on the decay of the political party system of Iceland, The Polar J. Journal 6 (1) (2016) 131–151.
[6] R.B. Collier, D. Collier, Critical junctures and historical legacies, Americas 51 (1) (1991) 27–39.
[7] IPCC, Climate Change 2022: Impacts, Adaptation and Vulnerability. Summary for Policymakers, 2022. https://www.ipcc.ch/report/ar6/wg2/.

[8] V. Satgar, The rise of eco-fascism, in: M. Williams, V. Satgar (Eds.), Destroying Democracy, Wits University Press, Johannesburg, 2021.

[9] A.D. Maynard, S.M. Dudley, Navigating advanced technology transitions: using lessons from nanotechnology, Nat. Nanotechnol. 18 (2023) 1118–1120.

[10] M. Coeckelbergh, The Political Philosophy of AI, Polity, Cambridge, 2022.

[11] P. Savaget, T. Chiarini, S. Evants, Sci. Publ. Pol. 46 (3) (2018) 369–380.

[12] M.A. Bakker, M.J. Chadwick, H.R. Sheahan, M.H. Tessler, L. Campbell-Gillingham, J. Balaguer, C. Summerfield, Fine-tuning language models to find agreement among humans with diverse preferences, arXiv preprint arXiv:2211.15006 (2022).

[13] Koster, et al., Human-centred mechanism design with Democratic AI, Human Nature Behaviour 6 (2022) 1398–1407.

[14] H.S. Sætra, A shallow defence of a technocracy of artificial intelligence: examining the political harms of algorithmic governance in the domain of government, Technol. Soc. 62 (2020), https://doi.org/10.1016/j.techsoc.2020.101283. https://www.sciencedirect.com/science/article/pii/S0160791X19305925.

[15] R. Heilbroner, Technological determinism revisited, in: L. Marx, M.R. Smith (Eds.), Does Technology Drive History, MIT Press, Cambridge, 1994, pp. 67–78.

[16] 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.

[17] N. Bostrøm, Superintelligence: Paths, Dangers, Strategies, Oxford University Press, Oxford, 2014.

[18] E. Schatzberg, Technology: Critical History of a Concept, University of Chicago Press, Chicago, 2018.

[19] L. Winner, The political philosophy of alternative technology: historical roots and present prospects, Technol. Soc. 1 (1) (1979) 75–86.

[20] J.A. Schumpeter, Capitalism, Socialism and Democracy, Routledge, 2013, 1943.

[21] R.B. Collier, D. Collier, Shaping the Political Arena, Princeton University Press, 1991.

[22] P.A. David, Clio and the economics of QWERTY, Am. Econ. Rev. 75 (2) (1985) 332–337.

[23] O. Fioretos, T.G. Falleti, A. Sheingate, Historical institutionalism in political science, in: O. Fioretos, T.G. Falleti, A. Sheingate (Eds.), The Oxford Handbook of Historical Institutionalism, Oxford University Press, 2016.

[24] G. Capoccia, R.D. Kelemen, The study of critical junctures: theory, narrative, and counterfactuals in historical institutionalism, World Polit. 59 (3) (2007) 341–369, https://doi.org/10.1017/S0043887100020852.

[25] T.M. Lenton, Environmental tipping points, Annu. Rev. Environ. Resour. 38 (2013) 1–29.

[26] H.S. Sætra, A hobbesian argument for world government, Philosophies 7 (3) (2022) 66, https://doi.org/10.3390/philosophies7030066.

[27] M. Coeckelbergh, Green Leviathan or the Poetics of Political Liberty: Navigating Freedom in the Age of Climate Change and Artificial Intelligence, Routledge, New York, 2021.

[28] D. Collingridge, The Social Control of Technology, Frances Pinter, London, 1980.

[29] V. Galaz, M.A. Centeno, P.W. Callahan, A. Causevic, T. Patterson, I. Brass, K. Levy, Artificial intelligence, systemic risks, and sustainability, Technol. Soc. 67 (2021), 101741, https://doi.org/10.1016/j.techsoc.2021.101741.

[30] L. Floridi, The green and the blue: naïve ideas to improve politics in a mature information society, The 2018 yearbook of the digital ethics lab (2019) 183–221.

[31] T. Mäkitie, J. Hanson, S. Damman, M. Wardeberg, Digital innovation's contribution to sustainability transitions, Technol. Soc. 73 (2023), 102255.

[32] S. Muench, E. Stoermer, K. Jensen, T. Asikainen, M. Salvi, F. Scapolo, Towards a Green & Digital Future (JRC129319), L. Publications Office of the European Union, 2022.

[33] OECD, Measuring the environmental impacts of artificial intelligence compute and applications: the AI footprint, OECD Digital Economy Papers (2022). November 2022 No. 341.

[34] S. Arcusa, K. Lackner, R. Page, V. Sriramprasad, W. Brandt, J. Moore, M. Green, S. Patil, The socio-economic and technical context for Direct Air Capture in decarbonization pathways, Technical Workshop Report January 19th 2022 (2022). https://keep.lib.asu.edu/items/170566/view.

[35] M. Coeckelbergh, AI for climate: freedom, justice, and other ethical and political challenges, AI and Ethics 1 (2021) 67–72.

[36] S. Mills, R. Whittle, Nudging policy or crowding it out? Green nudges as ideational technologies, in: H.S. Sætra (Ed.), Technology & Sustainable Development: the Promise and Pitfalls of Techno-Solutionism, Routledge, 2023.

[37] H.S. Sætra, A typology of AI applications in politics, in: A. Visvizi, M. Bodziany (Eds.), Artificial Intelligence and its Contexts: Security, Business, and Governance, Springer Nature, Cham, 2021, https://doi.org/10.1007/978-3-030-88972-2_3.

[38] H.S. Sætra, Generative AI: here to stay, but for good? Technol. Soc. 75 (2023), 102372.

[39] J. Dodge, T. Prewitt, R. Tachet des Combes, E. Odmark, R. Schwartz, E. Strubell, W. Buchanan, Measuring the carbon intensity of ai in cloud instances, in: 2022 ACM Conference on Fairness, Accountability, and Transparency, 2022, pp. 1877–1894.

[40] R. Willis, N. Curato, G. Smith, Deliberative democracy and the climate crisis, Wiley Interdisciplinary Rev.: Clim. Change 13 (2) (2022).

[41] A. Kahng, S. Macknenzie, A. Procaccia, J. Artif. Intell. Res. 70 (2021) 1223–1252.

[42] H. Landemore, Can AI Bring Deliberative Democracy to the Masses?, 2022. Retrieved 4 February 2023 from, https://www.law.nyu.edu/sites/default/files/Helen%20Landemore%20Can%20AI%20bring%20deliberative%20democracy%20to%20the%20masses.pdf.

[43] A. Buhmann, C. Fieseler, Towards a deliberative framework for responsible innovation in artificial intelligence, Technol. Soc. 64 (2021), 101475, https://doi.org/10.1016/j.techsoc.2020.101475.

[44] N. Wiener, Cybernetics or Control and Communication in the Animal and the Machine, MIT Press, Cambridge, MA, 1948.

[45] K.W. Deutsch, The Nerves of Govenrment: Models of Political Communication and Control, The Free Press, New York, 1963.

[46] David Easton, A System Analysis of Political Life, John Wiley, New York, 1965.

[47] L. Winner, Review: cybernetics and political language. Review of the Nerves of government: models of political communication and control by Karl W. Deutsch, Berk. J. Sociol. 14 (1969) (1969) 1–17. https://www.jstor.org/stable/41035143.

[48] S. Zuboff, The Age of Surveillance Capitalism, Profile Books, New York, 2019.

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

[50] G. Gaus, The tyranny of the ideal, in: The Tyranny of the Ideal, Princeton University Press, 2016.

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

[52] H.S. Sætra, Technology and Sustainable Development: the Promise and Pitfalls of Techno-Solutionism, Routledge, Milton Park, 2023.

[53] M. Coeckelbergh, Democracy, epistemic agency, and AI: political epistemology in times of artificial intelligence, AI and Ethics, online first 22 (2022), https://doi.org/10.1007/s43681-022-00239-4. November 2022.

[54] H. Bull, The Anarchical Society: A Study of Order in World Politics, Columbia University Press, New York, 1977.

[55] S. Barrett, Why Cooperate? the Incentive to Supply Global Public Goods, Oxford University Press, Oxford and New York, 2007.

[56] A. Næss, Ecology, Community and Lifestyle: Outline of an Ecosophy, Cambridge University Press, Cambridge, 1989.