

The Perfect Politician*
Theodore M. Lechterman
IE University
tlechterman@faculty.ie.edu

Chapter forthcoming in *Living With AI: Moral Challenges*, ed. David Edmonds (Oxford University Press)

In what ways, if any, should we seek to automate political decision-making?

In 2021, researchers at IE University conducted a global survey of the attitudes of citizens toward integrating AI into government.¹ They found that a majority of Europeans (51%) supported replacing at least some politicians with artificial intelligence. Citizens of China showed even more enthusiasm for AI-driven governance, with support for this proposal topping 75% for Chinese respondents. Americans and Brits expressed slightly more scepticism, with only 40% and 31% respectively voicing approval for this idea. Despite intriguing variation across countries, the results suggest a very strong attraction overall to upgrading politics with AI. They also invite deeper reflection. Is artificial political leadership something we have good reasons to want?

Ideas for integrating AI into politics are now emerging and advancing at accelerating pace. Here, I aim to highlight a few different varieties and show how they reflect different assumptions about the value of democracy. We cannot make informed decisions about which, if any, proposals to pursue without further reflection on what makes democracy valuable and how current conditions fail to fully realize it. Recent advances in political philosophy provide some guidance but leave important questions open. If AI advances to a state where it can secure superior political outcomes, leading perspectives in political philosophy suggest that democracy may become obsolete. If we find this suggestion troubling, we need to put the case for democracy on stronger foundations.

Automating Representation?

Presently, several options for integrating AI into politics vie for our support. One proposal seeks to have algorithms run for elected office and compete with human candidates. In 2018, a chatbot named Alisa challenged Vladimir Putin for the Russian presidency,² and a bot named SAM was developed with similar intentions in New Zealand.³ A more radical idea, proposed by physicist César Hidalgo, seeks to provide each citizen with a personalized bot,

* Portions of this chapter expand on ideas that I first explored in "Will AI Make Democracy Obsolete?" *Public Ethics Blog*, Stockholm Centre for Ethics of War and Peace, August 4, 2021, <https://www.publicethics.org/post/will-ai-make-democracy-obsolete>. I thank for the Stockholm Centre for granting copyright permission.

¹ Sam Shead, "More than half of Europeans want to replace lawmakers with AI, study says," *CNBC*, May 27, 2021, <https://www.cnbc.com/2021/05/27/europeans-want-to-replace-lawmakers-with-ai.html>.

² Tom O'Connor, "Will the Next Russian President Be a Robot? Putin's New Challenger Is a Machine That Knows 'Everything,'" *Newsweek*, December 7, 2017, <https://www.newsweek.com/russia-putin-could-face-controversial-robot-next-year-president-election-741509>.

³ "Meet the world's first virtual politician," *Newsroom*, December 15, 2017, <https://www.newsroom.co.nz/2017/12/14/68648/meet-the-worlds-first-virtual-politician>.

trained on a citizen's own inputs, that negotiates with other citizens' bots to design and approve legislation.⁴ Hidalgo envisions using these bots to reduce or even replace the role of human politicians.

These proposals are species of what we might call augmented electoral democracy. Augmentations of electoral democracy seek to alter the way that citizens are represented in the political process. And they reflect explicit or implicit critiques of conventional representation, where professional human politicians compete for public approval. What is the problem to which augmented representation provides the solution?

Efforts to enable algorithms to compete with human candidates for elected positions focus on the comparative limitations of human politicians, who may be corrupted in various ways and suffer from ordinary human constraints and vices, such as fatigue, ignorance, self-deception, and dishonesty. Automated politicians do not get tired; they are hard to bribe; and they have constant access to large libraries of information, which they can process rapidly and reliably.

Extant demonstrations of this idea have relied on rudimentary chatbots, which allow voters to interrogate a model about positions it has reached on various issues. Since no such candidate has yet succeeded in its bid for election, it remains unclear how well these proposals might perform on the job. But there are strong reasons to be sceptical of electing chatbots to represent constituencies alongside human politicians. In particular, the proposals might be faulted for underestimating the tasks required of elected officials.

The duties of elected office are many and varied. They generally require numerous complex tasks and negotiations with different humans. They are not limited to answering constituent questions, drafting proposed bills, and casting votes on legislation—things that a chatbot might be able to do (though of course with room for debate about how well it could do them). Rather, they involve negotiating with other elected officials to reach legislative compromises, managing staff who perform various auxiliary functions, running re-election campaigns, serving on committees, attending ceremonies, and meeting with different stakeholders to assess all sorts of problems and opportunities. While it is conceivable that an advanced and multifaceted AI agent could fulfil these tasks, and even fulfil them better than most humans, it is inconceivable that a chatbot could perform them without human supervision.

Rather than adapt AI to fill a role designed for humans, Hidalgo's proposal—to replace all politicians with personalized bots—seeks to reimagine representation in light of the capabilities that AI offers. Its critique of the status quo focuses not on the natural limitations and moral failings of humans but on the idea of representation itself. Many believe that electoral democracy is simply a pragmatic concession to the problem of scale. True democracy, on this reasoning, is direct democracy, where citizens decide directly on all policies. Until recently, direct democracy has been impractical in large, complex societies, where citizens cannot easily convene and vote on every matter. But technology now makes this possible. By answering surveys and sharing data with personal bots, citizens could train

⁴ César Hidalgo, "A Bold Idea to Replace Politicians," TED 2018, March 5, 2019, https://www.ted.com/talks/cesar_hidalgo_a_bold_idea_to_replace_politicians?language=en.

their bots to robustly represent their political preferences. These bots could then negotiate with each other to reach agreements on legislation. In turn, legislation would represent more citizens, as it would be the product of everyone's inputs, rather than the product of officials elected by the majority. It would also represent a much fuller picture of citizens' beliefs and desires, as this data would be fed directly into legislative negotiations, rather than reduced to a simple vote of yes or no for a candidate.

While Hidalgo's proposal offers several advantages, it also reflects some controversial assumptions. Are politicians merely a mechanism for processing our political preferences? Professional politicians may have, or be capable of marshalling, more expertise on political questions than ordinary citizens. Since Plato, many philosophers have regarded politics as an intricate art that requires detailed understanding of legislative procedures, an ability to recruit expertise in virtually any domain, shrewd capacities for strategy and negotiation, and exceptional moral judgment. Delegating the art of policymaking to qualified professionals may result in higher-quality outcomes than if all citizens participated equally in all policymaking decisions. Additionally, electing those with the talents and motivation needed to excel in policymaking enables everyone else to focus their labours where they are most suited, resulting in better individual job satisfaction and a more productive economy. Although electoral democracy contains certain tendencies towards inequality, particularly in countries where campaigns are privately financed, it also contains some important safeguards against it. Direct democracy, whether automated or not, might favour citizens with more resources for effective citizenship, such as education, leisure time, or inherent motivation. Because representatives must compete for the approval of their constituents, they are in principle accountable to all of them equally.

The point here is not to argue that electoral democracy is superior to direct democracy. Rather, it is to confront the normative assumptions that lie beneath proposals for upgrading democratic practices. Elected chatbots might have encyclopaedic knowledge of facts and lack many vices of human politicians. But at least for the foreseeable future, they will also lack the more advanced skills required of political leaders, such as capacities for persuasion and judgment. Meanwhile, a direct democracy in which all citizens enjoy the assistance of personalized robots might reduce certain inequalities and increase the responsiveness of policies to citizens' preferences. But it contains a high risk of generating low-quality results and widening other inequalities.

Replacing Voting with Data?

Other proposals seek even more radical transformations of conventional politics. For proponents of aggregative democracy, the legitimacy of political power requires that decisions fairly satisfy the preferences of those subject to them. The goal of any political system, according to this school of thought, should be to measure accurately, combine fairly, and duly satisfy the preferences of its subjects. And current polities are seriously defective on this score. Obviously enough, in authoritarian polities, citizens have few channels for expressing their preferences in the first place, and it is the preferences of the few that dominate conditions for the many. In self-described democratic polities, however, the expression of preferences is mainly limited to voting, an activity that occurs infrequently, provides limited information, and is subject to numerous cognitive biases. In many societies, citizens also face barriers to voting, and the integrity of elections suffers regular threats.

Powerful groups often find ways to influence which people or issues are put to a vote or to sway how decisions are implemented. Thus, replacing or supplementing conventional methods of translating citizen preferences into policies with AI may offer numerous benefits.

Drawing on science fiction and futurist studies, Jamie Susskind invites us to imagine an alternative to electoral democracy called data democracy.⁵ It is based on the idea that data generated throughout our daily lives—our travel patterns, online purchases, search and social media activity, metabolic rates, and so on—reveal a much more comprehensive, finely grained, and updated picture of citizens' wants than votes do. Thus, we might supplement or replace voting with a centralized and dynamic system for measuring and analysing inputs from citizens. The system would collect the vast streams of citizens' digital data and aggregate them to create a detailed profile of public opinion.

Susskind imagines using this data to inform the decisions that politicians make about legislation. But if the purpose of politicians in such a system is simply to process citizen preferences, this processing function might be done more reliably by artificial intelligence. A supercomputer could match data about citizen preferences with data about empirical conditions, enabling it to propose or implement policies for satisfying the general will.

However, augmenting politics in the spirit of aggregative democracy also faces serious criticisms that have long trailed this general conception of democracy. One challenge concerns the difficulty of measuring our preferences. The preferences we reveal through our economic decisions or online activity may offer a more comprehensive picture of our inclinations than votes do. But these apparent inclinations may only weakly correlate with what we truly want. The bingeing, doom-scrolling, rabbit-holing, trolling, flaunting, impulse-buying, and leering revealed in our online behaviour might reflect our vices or indiscretions more than what we value and hope to become. If the data processor could somehow overcome these measurement problems to obtain a more reliable indication of our preferences, it would face a further problem: many preferences we have are simply irrational or unreasonable. Even upon due reflection, many people may prefer, at least in certain ways and at certain times, to dominate or abuse others or sabotage their own wellbeing. Should data democracy give equal treatment to preferences founded on ignorance, confusion, or hatred?

An even more fundamental objection to data democracy would target its neglect of deliberation, a neglect that also figures in Hidalgo's proposal. For many democratic theorists, the ability to engage in public debate with fellow citizens is a vital component of the democratic process. It is only through the public exchange of reasons that we can fully develop and validate our preferences. It is deliberation that helps to separate impulses, prejudices, and mistakes from higher-order desires and beliefs. This perspective suggests that insofar as data democracy would involve shrinking or eliminating the role for deliberation in politics, it represents a fatally flawed conception of democracy.

An Artificial Sovereign?

⁵ Jamie Susskind, *Future Politics* (Oxford University Press, 2018), 246-50.

There is an even more radical possibility for seeking to improve or perfect our politics with AI, a possibility that we might call “algocracy.” Algocracy is similar to the extreme version of data democracy in that it involves a powerful autonomous system governing on the basis of large data flows and convoluted calculations. However, the algocrat, as I imagine it, would not be bound by popular preferences. Rather, it would ultimately be guided by its own judgments about justice and the common good.

How would the algocrat form these judgments? Fundamentally, it would do so based on comprehensive reading and analysis of all extant digitized content. It would read Confucius and Judith Butler, Darwin and Einstein; it would listen to Shostakovich and Shakira, compare Basquiat to the Dutch Masters, and contrast the philosophes with social media influencers; it would analyse every patent, medical database, architectural blueprint, computer programme, genome, and restaurant menu.

Naturally, the world’s digital repositories contain many biases, as the production and preservation of human knowledge throughout history have tended to favour certain groups and perspectives over others. Additionally, the quantity of recent digital content (much of which is low quality) vastly outnumbers the quantity of works preserved from the rest of human history. But if the algocrat were truly smart, it would be able to reason through these problems. Since it would have read every available publication in feminist and postcolonial criticism, the sociology of inequality, and the mechanics of algorithmic bias, it would be able to consider these critiques when analysing materials. Similarly, it would not weight the remarks of current celebrities more heavily than those of canonical thinkers simply because the former have more likes and retweets.

I imagine that the algocrat would also have access to data about citizen preferences, whether reflected directly through votes or indirectly through data. Unlike in data democracy, however, the algocrat would not be limited to aggregating and implementing citizen preferences, whatever they happen to be. Rather, it would use its reasoning powers to judge when preferences are relevant to a decision and when they are not, which preferences are worth considering, and how to fairly resolve conflicts between preferences.

Part of what makes algocracy an interesting thought experiment is that it challenges prominent views in political philosophy about democracy’s justification. Contemporary philosophers tend to agree that a well-functioning democracy is the most legitimate way of making coercively-binding decisions. Citizens governed by democratic processes have stronger reasons to endorse and comply with laws than citizens governed by alternative processes. But philosophers also disagree about the basis for this judgment, about what exactly makes democracy better than other political systems.

For instrumentalists, democracy is justified because it tends to produce superior results. More reliably than alternative systems, democratic polities tend to respect human rights, resolve conflicts peacefully, achieve prosperity, moderate inequality, and realize the preferences of their subjects. Algocracy poses a significant challenge for instrumentalist justifications of democracy, as it seems plausible that such a system could better achieve many or all the outcomes often associated with democratic systems.

Someone committed to the instrumentalist justification of democracy may resist this conclusion by reminding us that current forms of artificial intelligence face numerous limitations: their epistemic reliability is corrupted by biases, and their vulnerability to hacking and human manipulation makes them untrustworthy governors. Some may go further, claiming that advanced forms of AI will develop a survival instinct that leads them to manipulate humans rather than serve human values or impartial moral principles. But a consistent instrumentalist must admit that, if—or when—these challenges can be overcome, delegating political decision-making to an algocrat would become more than a tempting option: it would become a moral obligation.

Is democracy nothing more than an instrument for producing good results? One may acknowledge that AI could advance to the point where it would produce reliably better policies than humans can hope to achieve by themselves. But for non-instrumentalists, democracy is valuable apart from its consequences. An influential recent version of this position holds that democracy is valuable because it is part of an ideal of social equality. Alternative systems—such as monarchies, aristocracies, and oligarchies—involve granting people different amounts of political power. Granting some people more power than others sends the message that some people are inherently wiser or worthier of consideration than others. By rejecting inequalities in political power, democratic political processes instead affirm a commitment to a society marked by the absence of social hierarchy and subordination. Democracy is a necessary consequence of recognizing one another as moral equals.

However, it is not obvious that the non-instrumentalist position has the resources to fend off the challenge from algocracy. There would be no doubt about the algocrat's intellectual superiority. The algocrat is not a human person or someone with whom we have ongoing social relations. Unlike a dictator or plutocrat, the algocrat would not live among us as a member of human society, demanding or regularly receiving social privileges as perks of power. Thus, even the social egalitarian argument for democracy may be unable to resist the temptations of algocracy. If we can engineer a system of rule that gets dramatically better outcomes and does so without creating arbitrary social inequalities, the case for democracy becomes difficult to sustain.

One response to these observations is to conclude that the case for democracy is weaker than many have thought. If or when the technology for superseding democracy's limitations becomes available, we ought to implement it with all deliberate speed. Democracy, according to this perspective, may soon become technologically obsolete. Another response to these observations is to conclude that democracy is not obsolete but that the arguments in its support need re-examination.

Some philosophers have recently contended that the notions of autonomy and self-determination help to explain what is distinctively valuable about democracy. Even if other systems could more reliably or more robustly advance our preferences or track the demands of justice, and even if these systems could avoid creating or worsening inegalitarian social relationships, these systems would leave something missing. Since the resulting policies would not issue from our own agency, some have argued, we would reasonably find ourselves alienated from the world that they create. The voice that democracy allows us is valuable in part because it allows us to see our social world as of our own making. We would

not be able to fully see ourselves in a world ruled by AI, even if it reliably ruled in accordance with our interests.

Precisely how much weight this interest in non-alienation carries is unclear. Many people may be willing to trade it away in exchange for dramatic reductions in injustice that algocracy might offer. It remains possible—and perhaps likely—that democracy’s value is a complex combination of factors that cannot be reduced to a simple explanation. As the prospects for integrating AI into political decision-making become increasingly feasible, reflection on the foundations of the democratic ideal take on particular urgency.

Upgrading Political Philosophy for the Age of AI

Might AI, or further advances in AI, offer us the perfect politician? To be sure, there are attractive elements in each of the proposals I have surveyed. They suggest that AI can bring about a richer and more efficient harnessing of information than conventional democratic processes. AI may help to produce outcomes that more fairly represent all citizens, more richly reflect citizen preferences, and are more rational and informed. But each proposal also reflects controversial assumptions about the value of democracy—and its limits. Augmented electoral democracy either underestimates the complexity of a politician’s job or else assumes that representative institutions are simply an antiquated solution to the problem of scale. Data democracy assumes that preferences are exogenous to political participation and can be revealed indirectly through behaviour. Algocracy assumes that human agency is a dispensable feature of politics.

I contend that we cannot reach intelligent judgments about how to augment politics without deeper reflection on what makes democracy valuable and how that value can be best operationalized. Recent work in political philosophy provides some essential guideposts, but key questions remain unanswered. The possibility that human agency might someday be separable from politics is something that political philosophers are only now starting to take seriously. To better guide public conversations about the advantages and limitations of integrating AI into politics, political philosophy needs an upgrade.

--

Bio:

Ted Lechterman is Assistant Professor of Philosophy at IE University, where he specializes in political philosophy and applied ethics. His recent work focuses on the ethics of artificial intelligence, business ethics, and the ethics of philanthropy. He is the author of *The Tyranny of Generosity: Why Philanthropy Corrupts Our Politics and How We Can Fix It* (Oxford University Press, 2022). Lechterman was trained at Harvard and Princeton and completed postdoctoral fellowships at Stanford, Goethe, Hertie, and Oxford.