On the Normative Importance of the Distinction Between ‘Governance of AI’ and ‘Governance by AI’

By Eva Erman and Markus Furendal - 19 April 2023

Eva Erman and Markus Furendal urge researchers and the public to better think through the role of politics in the age of AI. This is the second post in a new EGG commentary series exploring how AI’s development is affecting economic, social and political decision-making around the world.

In an era where increasingly complex and capable artificial intelligence (AI) systems are unveiled at a steady pace, the effects that AI technology may have on economic, social, and political issues become increasingly clear. While many tasks in blue-collar jobs have already been automated, for instance, knowledge workers have generally been considered to perform creative tasks that machines are unable to recreate. Yet, recent advances in ‘generative AI’ technologies that instantly create text or images based on short prompts, have left illustrators, writers and office workers reconsidering their job security.

Although new technologies such as AI are sometimes thought of as self-propelling forces, their social and normative implications are not predetermined but rather

a result of political decisions and dynamics. In light of the anticipated and actual social impact of AI technology, calls for AI governance are thus more common than ever. Yet, even though the term AI governance is widely adopted, it is still largely undertheorized, and frequently used to describe a variety of distinctive phenomena and ideas.[i] Our aim in this short text, which draws on arguments we have presented at lengthelsewhere, is to introduce a helpful distinction which may reduce the risk of misunderstanding, and enable researchers and the public to better think through the role of politics in the age of AI.

Governance of AI and governance by AI

In the public and academic debates on the social impact of AI, the term AI governance is often used to refer to two phenomena that we suggest are in fact distinct: the phenomenon of 'governance of AI' and the phenomenon of 'governance by AI'. The former term refers to the kinds of emerging governance structures at various levels of policy-making that regulate and steer AI development and deployment. The most relevant example is perhaps the EU's so-called 'AI Act', which some expect will move from the draft stage into a binding regulation before the end of 2023. The latter, by contrast, describes the phenomenon of institutions implementing AI systems into their existing governance mechanisms. Many public authorities, for instance, already rely on AI systems to process data, automate decision-making, and detect suspected fraud. When public agencies do this, they govern by, or at least with the help of, AI.[ii]

Governance of AI includes hard law such as the coming EU AI Act, but also countless efforts best described as soft law approaches. These include recommendations, standards, ethical guidelines and declarations, codes of conduct and similar instruments developed by AI companies, NGOs, international organizations, or other actors in the AI space.[iii]Given that the soft law approach is agile and that its instruments can be adopted even when there is little international cooperation and agreement, soft law makes up a substantial share of global AI governance, and many expect it to remain the dominant approach.

Governance by AI could perhaps also be described as more or less soft, depending on its character and effects. On this view, we are in a sense softly governed by the recommendation algorithms or customer service chatbots that we encounter in our daily lives, whose guidance we are ultimately at liberty to turn down. But we are also

governed in another, more consequential way by, for instance, private insurance companies that calculate algorithmically derived risk profiles, and public authorities that employ automated decision-making about crucial issues like access to welfare benefits. In many cases, decision-making is supported by – rather than outsourced to – machines, such that there is still a human in the loop, who formally makes the decision recommended by AI technology. Some suggest that we could go further, however, and altogether hand over decision-making to machines. Optimists suggest that the AI-driven data analytics tool could collect citizens' views and thereby extend and equalize political influence. The developers of the AI chatbot 'Politician Sam', for instance, claim that it can analyze social media to accurately capture the political views of voters, and promise that it can thereby deliver 'true representation', 'active engagement', and 'better policy'.

Although these two notions risk being conflated by the widespread use of the monolithic term 'AI governance', we argue that it is important to keep them distinct, not least if we consider AI governance in relation to key normative ideals, such as democracy. The reason becomes apparent once we ask what it means for AI governance to be politically legitimate. Elsewhere, we have developed an account of the political legitimacy of AI governance, which attaches significance not only to the outcomes of, but also the procedures in, governance. Applying this account to actual AI governance suggests that both governance of AI and governance by AI can be politically legitimate under certain circumstances, but that these circumstances differ.

The political legitimacy of AI governance

We argue that the governance of AI is not necessarily politically legitimate when and because it produces 'good' outcomes, i.e. realizing the benefits and avoiding the risks of AI development. It also matters how we have come up with such a list of benefits and risks, and the goals of AI governance more broadly. Specifically, this process has to live up to some minimum threshold of democracy, where those who are affected by the decisions have an opportunity to participate in their making as equals.

To illustrate this point, consider the process of developing the EU's AI Act, on the one hand, and the AI-ethical work inside an AI-developing company, on the other. The EU's efforts in AI governance seeks to promote "trustworthy AI", an ideal which

presupposes respect for human autonomy, prevention of harm, fairness and explicability. Similarly, the large company Microsoft is committed to promoting what it calls "responsible AI", which is assumed to include values like fairness, reliability, and privacy, as well as inclusiveness, transparency, and accountability. At the face of it, it is very difficult to tell these somewhat vague ideals apart, and it is an empirical question which of the two efforts will ultimately be most significant. The EU is hoping for a 'Brussels effect', where legislative action in Europe sets the standard for the rest of the world. On the other hand, Microsoft is such a dominant player in the AI sector that their internal guidelines might very well be more consequential for the future of AI development.

Our point, however, is that aside from the actual effects of AI governance, it also matters whether there is a 'chain of legitimacy' between those who make decisions, and those who are affected by them. From this perspective, the key difference is that the EU's rules are legitimate, since they can ultimately be traced back to EU citizens, while private companies like Microsoft exercise authority that lacks this kind of legitimacy. This conclusion follows from an argumentthat we have made elsewhere regarding the global governance of AI, which we will summarize here.

Currently, most attempts to steer AI development toward certain outcomes are initiated either by private actors such as Microsoft, or by public entities. The latter can be better understood by distinguishing between two ways in which citizens can give public institutions the right to rule. First, authorized entities have been granted power by citizens through a direct authorization. You authorize your nation-state's parliament, for instance, when you go to the ballot box and elect a representative. In the AI space, authorized entities set up the legal structure for the societal goals and overall aims of AI development and deployment as well as the basic form of the main institutions of the AI space, through coercive decision-making. These institutions have the right to rule because they have been established through a democratic procedure in which those affected by AI (in one form or the other) have had an opportunity to participate as equals in shaping the "control of the agenda" concerning AI.

Mandated entities, by contrast, have been delegated political power not from citizens directly, but from authorized entities. They make non-coercive administrative decisions, work out policies, and so forth. You are governed by a mandated entity when you, for instance, follow rules set out by executive bodies or interact with public

administrative agents when applying for benefits or permits. A mandated entity can also further delegate authority to another mandated specialized entity. It may be appropriate to do so if, for instance, it enables higher-quality, decentralized and specialized governance.

Our conclusion about the difference between the EU's and Microsoft's AI governance follows from the assumption that democracy presupposes that instances of authorization and delegation constitute a legitimacy chain between those affected and the decision-making entities. The governance of AI is hence legitimate only when there is such a legitimacychain. Even though the EU is often accused of having a 'democratic deficit', there is nevertheless a formal democratic connection between individual EU citizens and the institutions in which this law is taking shape. The AI Act was first proposed by the European Commission, whose legitimacy can be traced back to the citizens of EU through a chain of authorization and delegation. By contrast, the soft law approach spearheaded by non-authorized, non-mandated tech companies like Microsoft lack this kind of legitimacy. Regardless of how laudable their aims are and how efficient a soft law approach is, there would thus be a legitimacy deficit if these initiatives were the only kind of governance of AI.

Legitimate governance by AI?

The distinction between governance of and governance by AI is also significant because it helps us understand what is going on in cases like the AI-Politician Sam. AI systems are often described as having superhuman capacities to gather, analyze and summarize data. In general, decisions are handed over to AI systems precisely because we believe they are better than humans at identifying the right option. We suspect that some will think that it makes sense to hand over many of the decisions that are currently made in a democratic fashion to AI systems, if it would lead to better outcomes (whatever that is taken to mean). If that happens, then it appears that an AI system would be an authorized entity, wielding legitimate authority over its human subjects.

In our view, however, governance by AI systems in this stronger sense cannot be legitimate. This is because democracy is not merely a decision-making method to reach good decisions, but also an ideal of self-determination, according to which those who

are supposed to comply with the rules have had the opportunity to authorize them by participating in their making as equals. There is no principled reason why mandated entities such as public administrations cannot legitimately engage in governance by AI, such as AI-assisted decision-making in relation to a predetermined set of issues within an already established legal framework, like when to grant or deny applications for welfare support.[iv]And AI systems could perhaps even become a kind of mandated entities, if authorized entities delegate some power to them. Human decision-makers in a parliament could, for instance, rely on AI-based technology to make more informed and thus better decisions.[v]

Yet, since the aims and goals of political communities ought to be deliberated and decided upon collectively, by the people bound by rules and regulation, we find it difficult to defend the claim that an AI system could also be the ultimate source of political legitimacy, i.e., that it could be an authorized entity. On our view, even if such a hypothetical AI agent would provide better decisions, handing over authority to it would negatively impact political legitimacy as we have conceptualized it.[vi] Given the speed at which AI systems are currently developing, however, we believe this is a key issue for future research.

Conclusion

In conclusion, we believe that much could be won by researchers and the public paying closer attention to the ambiguous character of the concept of 'AI governance'. Moreover, both of the phenomena we have described here raise substantial normative and practical questions about the way in which politics and AI technology interact. We have begun to describe here – in much-simplified and broad terms – some of the considerations one should keep in mind when considering the political legitimacy of AI governance. Given the wide-ranging and deep effects that the advent of AI technology is likely to have on societies world-wide, it is crucial to continue to study and develop theories for when and how the governance of AI, as well as governance by AI, live up to the ideals that should characterize people's social and political interactions.

Eva Erman is Professor at the Department of Political Science, Stockholm University. She works in the field of political philosophy, with special interest in democratic theory, critical theory and philosophical methodology. Erman is the author of The Practical Turn in Political Theory (2018) and Human Rights and Democracy: Discourse Theory and Global Rights Institutions (2005), and has published numerous articles in scholarly journals, including British Journal of Political Science, The Journal of Politics, and Political Studies. Furthermore, since 2008, Erman is the founder and Editor-in-Chief of the journal Ethics & Global Politics (Routledge).

Markus Furendal is a postdoctoral researcher at the Department of Political Science, Stockholm University. He focuses on issues concerning the social and ethical impact of AI, distributive justice, and the future of work. His work has been published in journals such as Political Studies, Social Theory & Practice, Philosophy & Technology, and Journal of Applied Philosophy.

Photo by Pavel Danilyuk

Notes

[i] See, for instance, Allan Dafoe, "AI Governance: A Research Agenda". Governance of AI Program, Future of Humanity Institute, University of Oxford, (2018), Anna Jobin, Marcello Ienca, and Effy Vayena, "The Global Landscape of AI Ethics Guidelines," Nature Machine Intelligence 1, no. 9 (September 2019): 389–99, , and cf. B. Guy Peters, "Governance As Political Theory," in The Oxford Handbook of Governance, ed. David Levi-Faur (Oxford: Oxford University Press, 2012)

[ii] We do not claim that this distinction is entirely novel. For instance, researchers have used the term "governance by algorithms" to describe the impact of recommendation algorithms on people's construction of a social order. Natascha Just and Michael Latzer, "Governance by Algorithms: Reality Construction by Algorithmic Selection on the Internet," Media, Culture & Society 39, no. 2 (March 2017): 238–58. See also Kuziemski Maciej, and Misuraca Gianluca. 2020. "AI Governance in the Public Sector: Three Tales from the Frontiers of Automated Decision-Making in

Democratic Settings." Telecommunications Policy 44 (6), and Christian Katzenbach and Lena Ulbricht, "Algorithmic Governance,"Internet Policy Review 8, no. 4 (2019). Further, the term "algocracy" has been used to refer to something akin to what we here label governance by AI, by John Danaher, "The Threat of Algocracy: Reality, Resistance and Accommodation,"Philosophy & Technology 29, no. 3 (2016): 245–68.

[iii] Thilo Hagendorff, "The Ethics of AI Ethics: An Evaluation of Guidelines," Minds and Machines 30, no. 1 (March 2020): 99–120, https://doi.org/10.1007/s11023-020-09517-8; Jobin, Ienca, and Vayena, "The Global Landscape of AI Ethics Guidelines."

[iv] There might, however, be practical reasons to be wary of this, relating to the explainability, accuracy and underlying fairness of the predictions and decisions generated by AI systems. For discussions, see Renée Jorgensen, "Algorithms and the Individual in Criminal Law." Canadian Journal of Philosophy 52, no. 1 (2022), 61-77; Kaun, Anne. 2022. "Suing the Algorithm: The Mundanization of Automated Decision-Making in Public Services through Litigation." Information, Communication & Society 25, no. 14: 2046–62; Kate Vredenburgh, "Fairness." In The Oxford Handbook of AI Governance, edited by Justin B. Bullock, Yu-Che Chen, Johannes Himmelreich, Valerie M. Hudson, Anton Korinek, Matthew M. Young, and Baobao Zhang. Oxford: Oxford University Press, (2022); and Liesbet van Zoonen, "Data Governance and Citizen Participation in the Digital Welfare State." Data & Policy 2 (2020): 10–17.

[v] The problems mentioned in the previous endnote are relevant also in this scenario.

[vi] One objection to our view could say that people can authorize an AI system to make future decisions in their place, just like members of parliament are authorized to make decisions in representative democracies. Yet, this once again implicitly assumes that authorized entities implement an agenda that is already set. Our point above is that the content and shape of the democratic agenda ought to be deliberated and decided upon collectively, by those bound by it. For an alternative line of argument reaching a similar conclusion as ours, see Ludvig Beckman, , Jonas Hultin Rosenberg, and Karim Jebari. "Artificial Intelligence and Democratic Legitimacy. The Problem of Publicity in Public Authority." AI & Society, (2022), who suggest that it is the opacity of machine learning undermines the publicity necessary for the AI to have legitimate authority.