Artificial Intelligence and Democracy

REIMAGINING DEMOCRACY FOR AI

Aviv Ovadya

Aviv Ovadya *is a research fellow at newDemocracy; an affiliate at the Centre for the Governance of AI, the Berkman Klein Center at Harvard, and the Safra Center for Ethics's GETTING-Plurality research network; and a founder of the AI & Democracy Foundation. His work focuses primarily on ensuring that the governance of AI can keep up with the rate of AI advances.*

The foundations of democracies and our international order rest on implicit assumptions that AI advances are rapidly disrupting, as machines become increasingly capable of simulating the actions and outputs of humans. This is especially true when it comes to generative AI and foundation models for general-purpose AI. The ramifications are global and range from deepfakes and electoral manipulation in democracies, to panopticon surveillance across autocracies, to cataclysmic catastrophes such as AI-assisted biological attacks from nonstate actors.[1] All this is on top of the significant challenges posed by existing AI systems.

If we continue on our current course, advances in AI may take us down one of two possible paths toward a dystopian future: that of *autocratic centralization,* where powerful corporations or authoritarian countries unilaterally control extraordinarily powerful AI systems, or of *ungovernable decentralization,* where everyone has unrestricted access to those incredibly powerful systems and, because there are no guardrails, can use them to cause massive, irreversible harm.

I advocate a third path—that of combined *democratic centralization* and *democratic decentralization*—and accelerating investment in the democratic infrastructure needed to make such a path viable. There are many things we must do to safeguard the world's existing democracies from the possible dangers posed by AI. Here, I focus on four interconnected innovations—representative deliberations, AI-augmentation, democracy-as-a-service, and platform democracy—that would enable

Journal of Democracy Volume 34, Number 4 October 2023

democracies to better confront the challenges arising from AI and ensure a positive future. These innovations are not themselves solutions to the challenges around AI, but they can provide a foundation that may enable us to act at the necessary speed and scale.

Representative deliberations. This innovation comes from modern applied deliberative democracy (building on ideas from ancient Athens) and aims to make the process of devising and deciding on policy solutions democratically representative. This is achieved by creating a representative microcosm (or minipublic) of the populace being governed and giving it the time, information, and resources to deliberate wisely. Deliberators are selected from the population through what is known as sortition or a democratic lottery—such that every person has a roughly equal chance of being selected.[2] Thus, far fewer people (anywhere from forty to eight-hundred, depending) are involved than are in a referendum or election. Limiting the number of participants makes it possible for conveners to invest more resources per deliberator, so that those chosen can be compensated for their time (often forty hours or more) spent grappling with the issue, in facilitated dialogue with each other, experts, and stakeholders.[3]

In a high-quality deliberative-democracy process, sortition (which removes many of the perverse incentives of electoral politics) is coupled with significant investment to ensure that those selected 1) *can participate*, by providing appropriate compensation, childcare, eldercare, and the like, which helps to reduce self-selection; 2) *have sufficient context,* by providing briefing materials about the decision at hand and access to experts and stakeholders; and 3) *can deliberate effectively,* through facilitated discussions and activities that ultimately result in decisions.

The democratic legitimacy of such processes comes from the representative makeup of the assemblies—far more representative than one finds in a standard elected body. Moreover, the best representative deliberations effectively communicate the "deliberative journey" to the rest of the concerned population through mass media. In this way, the broader public can see people similar to themselves learning about the issues, learning from one another, and coming to a set of conclusions that might initially have been counterintuitive. The best processes also include a mechanism for collecting public feedback and opinions, which are then shared with the deliberators along with the more traditional multistakeholder and expert input. This approach of bringing the entire population along on the deliberative journey, what I call *parascaling*, is particularly helpful for maintaining democratic legitimacy.

Representative deliberations have already been used hundreds of times—by governments around the world at every level, from small towns all the way up to the EU and an UN-endorsed global pilot. Sometimes called citizen assemblies, citizen juries, citizen panels, or deliber-

ative polls (albeit with significant differences across different approaches[4]), representative deliberations are usually convened by a government to answer a specific question, often one that involves difficult tradeoffs or value dilemmas, for example: "How can we lower climate emissions to 40 percent of our 1990 level?" or "Should we continue building nuclear power plants?"

The key ingredient that modern representative deliberations provide, at least in theory, is the ability to provide informed policy responses to any targeted question, with democratic legitimacy, for any population. Whereas many busy voters may need to cast their ballots on gut instinct, participants in representative deliberations are given the compensated time and resources to make decisions based on extensive information and deliberation—perhaps making the process more robust to AI-augmented advertising and manipulation. Representative deliberations also have an advantage over solely multistakeholder processes, because a representative body can act as a "democratic adjudicator," thereby democratically weighting the voices of the different stakeholders.

AI augmentation. Although representative deliberation has many valuable attributes, running such processes across many languages and cultures is incredibly challenging. Broader public involvement beyond the core deliberators (and the input they receive) can increase the legitimacy and quality of a deliberation's decisions. This is where the second key innovation, AI augmentation, comes in. Although it makes sense to be cautious about applying AI to govern AI, it might also be necessary if governance is to keep up with the technology.

Among the most obvious opportunities are the significant advances in translation and real-time interpretation. As AI systems are increasingly able to understand context-specific slang or behavior, and thus can explain cultural differences, they can further enhance representative deliberations. Additionally, there has already been considerable innovation in technologically augmented deliberative and democratic processes. Collective-response systems, such as Polis and Remesh, can use more established kinds of AI to help identify points of consensus.[5] Such systems can also be run in a sequence to approximate some kinds of deliberation at a massive scale, as a complement to more traditional deliberative processes.

Large language models such as GPT-4 can also enhance deliberative processes by helping to generate new points of potential consensus within a group based on its members' prior perspectives, and by understanding and explaining the perspectives of the stakeholder groups.[6] Finally, one of the biggest obstacles to running deliberative processes is the cost and training of expert facilitators, and there is thus significant investment now in substituting some of the roles of human facilitators with AI systems.[7] More generally, AI systems could be extremely help-

ful in organizing and summarizing information both from experts and others, particularly if hallucination problems (when an AI system generates false information) are resolved.

This is just a small taste of the potential innovations being explored for supporting either the collective-intelligence component of deliberation or the representative-legitimacy component of the decision-making. The combination of representative deliberations and AI-augmentation can potentially enable robust deliberative and participatory democratic processes at every scale.

Democracy-as-a-service. Beyond improving democratic processes themselves, there is also the question of who should run them and why. In a democracy-as-a-service model, third-party organizations run participatory and deliberative processes *for* governments, thereby allowing processes, tools, and best practices to spread and be built on more rapidly. For example, organizations such as newDemocracy, Missions Publiques, Deliberativa, and Healthy Democracy run representative deliberations for local, national, and supranational (the EU, for example) governments around the world. In the process, these organizations build expertise at conducting deliberations and helping governments adhere to best practices for implementing the results. The Democracy R&D network,[8] an international network of organizations aiming to improve democracy, accelerates the development of representative deliberations by enabling knowledge-sharing across organizations and incorporating researchers interested in improving those processes. In many ways, democracy-as-a-service is not particularly new—consider, for example, the companies developing voting machines. Yet we should not underestimate how much innovation is spurred on by having nimble third-party organizations develop and pilot novel end-to-end democratic processes.

Platform democracy and AI democracy. Over the past two decades, tech companies including Meta, Google, and Apple have become wealthier than many governments, and their platforms and products shape and constrain the actions of billions of people. Although for-profit corporations' having so much power (including over political processes) is deeply problematic, there can also be benefits to limiting how much power governments can exert over these corporations, given the perverse incentives for political leaders (both elected and autocratic) to maintain their own power.

This dilemma has generally been considered unsolvable, leading me to put forward the approach of "platform democracy"—colloquially defined as "governing platforms democratically."[9] This calls for the corporations running platforms to work with democracy-as-a-service providers to run representative and AI-augmented deliberations for developing their policies and governance structures. I worked with Twitter to set

up the pilot for such a process, but that was stalled by the platform's acquisition in 2022. Meta, however, ran platform-democracy pilots of increasing scale, culminating most recently in a representative deliberation across 32 countries in 19 languages to guide policies with difficult tradeoffs between privacy and security.[10] While the process was far from perfect, including in its democratic bona fides, it was a valuable first step in testing democratic alternatives and showing that they can work if adequately resourced and empowered.

Similarly, "AI democracy" has begun building momentum. Perhaps inspired by what Meta has done, nearly all the organizations developing the most advanced AI systems have begun exploring how they themselves might incorporate democratic processes. For example, OpenAI, the company behind ChatGPT, has launched a grant program to fund ten projects exploring democratic systems that the company might use for AI alignment and governance.[11] (Disclosure: I am an advisor to this program.)

Applying Democratic Innovations

The OpenAI grant shows how many of these innovations, beyond their general benefits for democracy, are also directly relevant to addressing significant challenges posed by AI advances.

AI alignment. "AI democracy" built upon augmented representative deliberations can help in developing the principles for *aligning* AI systems—that is, ensuring that an AI system operates according to a set of principles, such as "do no harm." Significant technical challenges must be overcome in order to implement alignment: figuring out how to design and train a system so that it sticks to a set of principles, including common-sense principles that might not have been explicitly or perfectly specified, and deciding what those principles should be aligned to and how to balance tradeoffs appropriately.[12] While some of the decisions about such principles may be delegated to the direct user of an AI system, there will always be some base set of values that is encoded by default—and which may be required in order to limit severely harmful activity. Currently it is primarily the AI companies themselves, at least those based outside of authoritarian states, that are deciding what generative and general-purpose AI systems should align to.

Alignment concerns have become even more salient as people have started experimenting with more autonomous systems built upon generative AI models. As systems transition from performing distinct, clearly defined tasks such as translation or image recognition to autonomously accomplishing complex objectives that require interacting with the environment, such as driving or medical research, the potential risks of misaligned systems increase.

Unfortunately, if unsurprisingly, differences of perspective around such values appear to be exacerbating mistrust and geopolitical risk, as AI organizations and governments with differing values race to ensure that the most powerful systems are aligned with *their* values—and one casualty of this race is likely to be critical guardrails. Representative deliberation can help to address these challenges by providing a broadly acceptable mechanism for navigating across those competing values, democracy-as-a-service enables corporations to convene such deliberations while staying at arm's-length, and AI-augmentation may even enable such processes to be feasible globally.

Global agreements. To further address these risks and challenges of powerful AI systems, we are likely to need some form of globally agreed-upon policies around the development, deployment, and distribution of such systems—for example, mandating that AI systems should be trained and aligned not to support the development of chemical and biological weapons.[13] This may sound straightforward, but it brings up a thorny issue related to open-source AI systems. Open-source systems reduce centralized corporate control of AI systems and make research easier. Unfortunately, it might not be possible to prevent people from "retraining" an open-source AI system to overcome its alignment guardrails—this has already been done with some of the most powerful open-source models. And it is impossible to "unrelease" an open system once it has been shared publicly, which means that a single actor could irreversibly impact the entire planet. Some argue that if the risks of such open releases are significant enough, we might need a global prohibition on the development or open distribution of certain types of AI systems.

There is currently significant disagreement about how to navigate such dilemmas, and meaningful consensus is exceedingly difficult to achieve, due to challenges including the speed of change; uncertainty and disagreement around the degree and direction of AI impacts; distrust among key actors; ease of replication; and the lack of a broadly trusted process for weighting conflicting ethical obligations. The same democratic innovations may be invaluable here also, providing a complement to more traditional geopolitical negotiations.

Implementing the new processes. There are a variety of ways to implement augmented democratic processes, including: 1) convening on an *ad-hoc basis*, whenever a question needs to be answered; 2) convening on a *regular basis*—for example, once a month or quarter—and potentially settling on the exact question just before a deliberation begins, thus removing the delay between identifying a question and running a deliberative process; 3) convening on a *continuous basis*, which could be particularly helpful when addressing value alignment; 4) running a set of interacting processes that feed into one another, like the Ostbel-

gien Model does, where one smaller representative deliberative body does the agenda-setting, determining what questions are most important, and then a larger representative body deliberates on those questions; the interacting assemblies can also serve as checks on each other.

AI advances are driving us toward a dystopian future of autocratic centralization and ungovernable decentralization. Still, we have an opportunity to aim instead for democratic centralization and democratic decentralization.

For example, a company such as OpenAI or Google, or a consortium working across such companies, might convene AI-augmented representative deliberations once every quarter to address the issue of value alignment and update their core "AI constitutions" accordingly. Similarly, for global agreements, the United Nations or a consortium of multistakeholder or multilateral actors could act as an official convener, raising questions around complicated issues such as biosecurity to identify points of global agreement. Challenges would, of course, remain—not least being the problem of getting authoritarian regimes to participate, although there is some evidence that countries such as China might be open to such processes, as they do not directly threaten political leadership (as long China retains control on how the proceedings and results are broadcast within the country). In fact, parts of China have run representative deliberations in the past, and Chinese citizens participated as part of the global climate assembly pilot.[14]

"Deliberative Infrastructure" Before It's Too Late

In the last nine months, we have gone from having almost no recognition of the necessity to think about democratic innovations to seeing almost every major AI company begin to explore how best to incorporate aspects of deliberative democracy into their work. I have also started to see interest from people in government and international organizations such as the UN. But if we are to have any chance of running deliberative processes in time to handle the urgent issues around AI at a global level, we need to massively improve every aspect of the representative deliberative process—enabling sortition in regions where it is currently difficult, developing the organizational expertise to run deliberations that will sometimes span the globe, and applying the latest technologies to enhance and reimagine these processes.

There is incredible capacity and momentum in the democratic-innovation ecosystem, but the rate of AI advances is far faster. I have therefore been exploring the possibility of setting up a fund focused on

democratic innovation to accelerate the design, testing, evaluation, and composition of such processes at increasing scale, working in partnership with civil society, academia, AI companies, and multistakeholder and multilateral organizations for implementation. I would like to see governments around the world developing similar focused funds to ensure that we can rapidly build the capacity to run complex end-to-end processes for both alignment and policy. Corporations advancing AI should also signal their willingness to invest in democratic governance and alignment, with funds pre-allocated for running processes that can satisfy particular criteria, whether developed in-house or externally. This would create a market incentive for rapid investment in the development of democratic processes.

Beyond the benefits of aligning and governing artificial intelligence, there are of course other urgent societal challenges where better decisionmaking and coordination would be invaluable. Relative to their potential benefits, augmented representative deliberations provide a significantly underresourced approach to creating the agreement and political will necessary to tackle the most challenging issues of our time.

AI advances are driving us toward a dystopian future of autocratic centralization and ungovernable decentralization. Still, we have an opportunity to aim instead for democratic centralization and democratic decentralization. It is a great gift that the same technology which is so destabilizing may also be harnessed to help overcome the problems it is creating. We should not spurn this gift. As our technological capacities grow, we must ensure that our ability to govern those capacities grows with them. There is a tremendous amount that we need to do right now to address present and significant risks and harms—but there is also little time to waste if we want to be ready to tackle the even more significant crises that are coming.

NOTES

1. Artificial Intelligence and Biosecurity Risk Assessment Act, S2399, 118th Cong. (2023); Emilia Javorsky and Hamza Chaudhry "Convergence: Artificial Intelligence and the New and Old Weapons of Mass Destruction," *Bulletin of the Atomic Scientists,* 18 August 2023, *https://thebulletin.org/2023/08/convergence-artificial-intelligence-and-the-new-and-old-weapons-of-mass-destruction/*.

2. Hélène Landemore, Andrew Sorota, and Audrey Tang, "Why Picking Citizens at Random Could Be the Best Way to Govern the AI Revolution," *Fortune,* 20 June 2023, *https://fortune.com/2023/06/20/why-picking-citizens-at-random-best-way-to-govern-ai-revolution-tech-politics/*.

3. OECD, *Innovative Citizen Participation and New Democratic Institutions: Catching the Deliberative Wave* (Paris: OECD Publishing, 2020), *www.oecd.org/gov/innovative-citizen-participation-and-new-democratic-institutions-339306da-en.htm.*

4. Aviv Ovadya, "Deliberative Polls, Citizen Assemblies, and an Online Deliberation

Platform," *Reimaging Technology,* 10 July 2023, *https://reimagine.aviv.me/p/deliberative-poll-vs-citizen-assembly-meta-pilot*.

5. Aviv Ovadya, "'Generative CI' through Collective Response Systems," 1 February 2023, *https://arxiv.org/abs/2302.00672*.

6. Christopher T. Small et al., "Opportunities and Risks of LLMs for Scalable Deliberation with Polis," 20 June 2023, *http://arxiv.org/abs/2306.11932*; Michiel Bakker et al., "Fine-Tuning Language Models to Find Agreement Among Humans with Diverse Preferences," 36th Conference on Neural Information Processing Systems (NeurIPS 2022), *https://proceedings.neurips.cc/paper_files/paper/2022/file/f978c8f3b5f-399cae464e85f72e28503-Paper-Conference.pdf*; AI Objectives Institute, "Introducing Talk to the City," *https://aiobjectives.org/blog/introducing-talk-to-the-city-our-collective-deliberation-tool*.

7. Ovadya, "Deliberative Polls, Citizen Assemblies, and an Online Deliberation Platform."

8. See *https://democracyrd.org*.

9. Aviv Ovadya, "Towards Platform Democracy: Policymaking Beyond Corporate CEOs and Partisan Pressure," paper, 18 October 2021, *www.belfercenter.org/publication/towards-platform-democracy-policymaking-beyond-corporate-ceos-and-partisan-pressure*.

10. Aviv Ovadya, "Meta Ran a Giant Experiment in Governance. Now It's Turning to AI," *Wired,* 10 July 2023, *www.wired.com/story/meta-ran-a-giant-experiment-in-governance-now-its-turning-to-ai*.

11. Wojciech Zaremba et al., "Democratic Inputs to AI," OpenAI blog, 25 May 2023, *https://openai.com/blog/democratic-inputs-to-ai*.

12. Brian Christian, *The Alignment Problem: Machine Learning and Human Values,* 1st ed. (London: W.W. Norton, 2020).

13. There is a danger here of overreach; global agreements should be developed only for issues that have significant impacts for everyone.

14. James S. Fishkin et al., "Deliberative Democracy in an Unlikely Place: Deliberative Polling in China," *British Journal of Political Science* 40, no. 2 (2010): 435–48.