GETTING-Plurality
Research Network

Date: July 7, 2023
Docket ID: OSTP-TECH-2023-0007
From: GETTING-Plurality Research Network, Edmond & Lily Safra Center for Ethics, Harvard University
Submitted via regulations.gov

GETTING-Plurality Research Network Comment on the White House OSTP National Priorities for Artificial Intelligence: Bolstering Democracy and Civic Participation

GETTING-Plurality is a multi-disciplinary research network at the Edmond & Lily Safra Center for Ethics at Harvard University led by Professor Danielle Allen and Professor Allison Stanger. The network pursues foundational analysis and theory, field-building, and policy development to foresee and mitigate potential harms to democracy and to strengthen the public benefit and democracy-supportive effects flowing from technology innovation. More information available at: https://gettingplurality.org/

Recent advances in AI have unleashed technologies capable of derailing democratic processes and undermining democratic institutions. Regulation, government-funded research, and pro-democracy reform have become ever more urgent. The development and deployment of AI models presents significant threats to bedrock democratic principles like non-domination, representation, self-determination, and political equality. Alternatively, these technologies offer opportunities for smarter deliberative systems and for more efficient and productive engagement between citizens and their governments. Specifically, emerging AI models present both opportunities and threats for our **(1) information ecosystems, (2) decision-making systems, and (3) responsive government and healthy democratic communities**. Our aim is twofold: first, to harness these opportunities for a flourishing democracy, and second, to strengthen our democratic institutions against the threats posed by AI.

How can AI be used to strengthen civic engagement and improve interactions between people and their government?

Regulators and government agencies can leverage Large Language Models (LLMs) to build healthy digital public spaces and to improve access to civic information and government services.

Information Ecosystems

- AI can be used to automate certain steps of the journalistic process, including routine data analysis and drafting preliminary reports of local events and recaps, allowing journalists to

direct their attention to other essential aspects of investigative reporting.

Decision-Making Systems

AI has already proven useful at aggregating and synthesizing the results of online deliberative processes to enhance citizen participation in government. The federal government should harness these opportunities to strengthen participatory decision-making systems beset by polarization and deadlock.

- Online deliberative platforms like Polis can be leveraged at the local level to collect and synthesize citizen input and debate on policy questions. In Taiwan, for example, citizens and government officials have found success debating and coming to agreements on hard-to-resolve policy questions through the use of online deliberative platforms like vTaiwan.
- To alleviate the already present harms of excessive polarization and online radicalization—harms which will only grow more worrisome as LLMs enable users to rapidly generate toxic content at near-zero marginal cost—the federal government can invest in research and enforce regulations to harness the power of bridging-based content moderation and recommendation systems. Such systems leverage recent advances in machine learning to design civic technology platforms, social media content recommendation schemes, and models of human-moderated group discussion which build trust and agreement between previously divided groups.[1]

Responsive Government and Healthy Democratic Communities

- Since LLMs can process, summarize, and interpret unstructured text data, they have the potential to make language-based systems technologically interoperable, allowing data sharing and faster interaction between agencies and between government and members of private or civil society.
  - This poses opportunities and risks. While LLMs may enable vastly more efficient collaboration between government agencies, their tendency to generate false information (hallucinate) means that no system interoperability should be left entirely to LLMs.
  - The interoperability of language-based systems might prove particularly useful for information and data sharing between the federal government and state and local governments in areas like public health, where the absence of national data standards presents challenges to inter-jurisdictional collaboration.[2]
- Currently, citizens seeking support or information from federal agencies have to navigate somewhat complex websites, call busy phone lines, or visit in person offices.

---

1 Ovadya, Aviv. "Bridging-Based Ranking: How Platform Recommendation Systems Might Reduce Division and Strengthen Democracy." Belfer Center for Science and International Affairs. 2022. https://www.belfercenter.org/publication/bridging-based-ranking; Ovadya, Aviv, and Luke Thorburn. "Bridging Systems: Open Problems for Countering Destructive Divisiveness across Ranking, Recommenders, and Governance." 2023. https://arxiv.org/abs/2301.09976.

2 Iyamu, Ihoghosa, Oralia Gómez-Ramírez, Alice XT Xu, et al. "Challenges in the development of digital public health interventions and mapped solutions." Digital Health. 2022. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9152201/.

Recommendations:

1. NIST should fund interdisciplinary research into bridging-based systems, and the FCC should develop guidelines, best-practice recommendations, and incentives for social media platforms to build and utilize bridging-based content recommender systems.
2. Agencies using participatory decision-making models or public input processes in the federal government should develop pilot programs to integrate AI deliberative tools and can offer funding and support to incentivize local jurisdictions to do the same. Such programs should not abandon human oversight of such deliberative processes, and AI tools should be used only to supplement human-directed sense-making and preference-aggregation processes, not to entirely replace them.
3. In order to harness the opportunities of interoperability, the federal government can pilot programs authorizing the use of LLMs to support federal agency operations, and ought to require a 'human in the loop' in cases where LLMs are used to perform agency work.
4. Agency websites can use LLMs to develop chatbots to direct individuals to the information, services and requisite forms to meet their needs.
5. The federal government should require developers of new AI systems to meet standards of data portability and develop protocols for companies to share their training data and procedures with government and university researchers interested in the effects and deficiencies of such models.

What are the key challenges posed to democracy by AI systems? How should the United States address the challenges that AI-generated content poses to the information ecosystem, education, electoral process, participatory policymaking, and other key aspects of democracy?

The federal government should aim to protect against the threats posed by LLMs to our information ecosystems, our electoral processes, and our other democratic institutions by investing in tools to increase trust in public information, enhancing opportunities for participatory policymaking, and adopting a set of renovations that strengthen the democratic foundations of our government.

Information Ecosystems

The federal government should aim to set standards and invest in technical solutions to combat threats related to digital (1) misinformation due to hallucination and accidentally generated false content or (2) disinformation due to the intentional use of LLMs for manipulation of the information commons.

- Challenge: Hallucination, misinformation and manipulative content
  - Prior to LLMs, infrastructural risks and content-based risks were separate. Facebook and other social media platforms posed infrastructural risks to our information ecosystem through their control of ranking algorithms, news feeds, and, ultimately, digital public spheres. Apart from these infrastructural powers held by platform companies, particular bad actors posed content-based risks through the production of misinformation and

manipulative content.[3] LLMs as they currently exist pose both kinds of risks:
    - LLMs are currently prone to hallucination, and are easily used to produce misinformation. Thus, they can be used to intentionally or accidentally undermine individual or group claims to fair and accurate information and representation.[4]
    - The current training paradigm to prevent LLMs from generating toxic or dangerous content (Reinforcement Learning with Human Feedback), places significant power in the hands of human "trainers" and does not offer full-spectrum protection against all kinds of misinformation or fraud schemes.[5]
- Challenge: Pollution of the digital commons
    - LLMs threaten to pollute information ecosystems with a barrage of newly generated content, decreasing the chances of commonly shared information and information sources.

Decision-Making Systems

In addition to instituting guardrails over information ecosystem threats, the federal government should aim to establish protocols for decision-making processes (aligned with democratic values).

- Challenge: American democracy aims to combine both majoritarian decision-making processes and strong protections for minoritarian interests and concerns. The use of AI systems like Polis to aggregate and synthesize the results of deliberative participatory systems might risk amplification of majoritarian views without a sufficient balance of individual, idiosyncratic, and yet nevertheless important concerns.
- Challenge: AI-generated lobbying materials and AI-generated law, rules, and policy.
    - LLMs have performed well on economics exams, Bar exams, strategy tests and games, and other tests relevant to policy making and lobbying processes. They may become convenient tools for lobbying firms and special interests looking to supercharge their influence in politics.[6]

Responsive Government and Healthy Democratic Communities

In addition to all of these AI-specific risks, there will surely be unforeseen risks and challenges presented by the use of LLMs across contexts. In order to make sure that our democratic institutions can withstand these potential threats, the federal government should adopt several broad-based reforms aimed at a healthier democracy.

- Challenge: Existing incentive structures make our democracy particularly vulnerable to the harmful effects of polluted information ecosystems, the influence of algorithmic bias in

---

3 Simons, Josh. *Algorithms for the People: Democracy in the Age of AI.* Princeton University Press, 2023.

4 Risse, Mathias. *Political Theory of the Digital Age.* Cambridge University Press, 2023.

5 Liu, Gabrielle Kaili-May. "Perspectives on the Social Impacts of Reinforcement Learning with Human Feedback" 2023. https://arxiv.org/abs/2303.02891

6 Sanders, Nathan E., and Bruce Schneier. "How AI Could Write Our Laws." MIT Technology Review. 2023. https://www.technologyreview.com/2023/03/14/1069717/how-ai-could-write-our-laws/.

decision-making processes, and other well-known challenges posed by the use of AI systems. [7]

- Currently, zero-sum opposition campaigns incentivize negative campaigning, large districts increase the distance between Americans and their representatives, and party-based political systems supercharge polarization and enable slim pluralities to govern broadly.[8]

Recommendations:
On addressing challenges posed by AI–

1. Invest (perhaps through NIST) in research on contextualization engines, provenance, digital identity and AI content watermarking.
   a. Contextualization engines can be developed and deployed to help citizens identify mis/disinformation and understand when and where they may be the targets of potential scams. They offer users the opportunity to easily seek additional context behind claims they encounter online. Research has shown that such easily-accessible contextualization engines contribute to media literacy and could reduce individual and systemic effects of misinformation or manipulation campaigns.[9]
   b. Research on data transparency, provenance, AI watermarking, and tools for establishing non-transferable digital identity proves promising for protecting against AI-generated misinformation and for protecting against preference manipulation.[10]
3. Establish clear standards of liability and responsibility (perhaps through the DOJ) for dangerous and untruthful content produced using LLMs to clarify the standing of LLMs with respect to Section 230 and the 1st Amendment.
4. Consider establishing data transparency, provenance, and mandatory watermarking standards for at AI Labs developing LLMs.
5. Establish standards for clear consumer-facing warnings about LLMs' tendencies to hallucinate; develop liability standards for hallucination.
6. Limit the use of LLMs in high-risk contexts such as campaign materials, or other participatory contexts prone to manipulation.
7. Invest in supporting local and national information ecosystems that private social media

---

7 Brangham, William, and Sarah Clune Hartman, "Security expert [Bruce Schneier] warns of AI tools' potential threat to democracy." PBS. 2023. https://www.pbs.org/newshour/show/security-expert-warns-of-ai-tools-potential-threat-to-democracy.

8 Allen, Danielle, Stephen B. Heintz, Eric P. Liu, "Project: Commission on the Practice of Democratic Citizenship." https://www.amacad.org/project/practice-democratic-citizenship; Commission on the Practice of Democratic Citizenship, "Our Common Purpose." *American Academy of Arts and Sciences.* 2020. https://www.amacad.org/ourcommonpurpose/report.

9 Ovadya, Aviv. "Contextualization Engines." Cybersecure Policy Exchange. 2021. https://www.cybersecurepolicy.ca/policy-brief-contextualization-engines.

10 Kirchenbauer, John, Jonas Geiping, et al. "A Watermark for Large Language Models" 2023. https://arxiv.org/abs/2301.10226; Kirchenbauer, John, Jonas Geiping, et al. "On the Reliability of Watermarks for Large Language Models." 2023. https://arxiv.org/abs/2306.04634; "Project Origin: Protecting Trusted Media." https://www.originproject.info/.

platforms have supplanted, including local journalism and publicly-supported news and programming. Consider funding these investments with a tax on digital advertising.
8. The Congress, White House, and Federal Agencies should establish clear rules related to AI-generated content and lobbying material and AI-generated language in bills, agency policies, or other administrative or official documents.

For a stronger, more resilient democracy–
9. Study state and local implementation of ranked choice voting, which incentivizes less negative campaigning, coalition formation, and decreases election costs
10. Increase the size of the House of Representatives (the lower the ratio between representative and constituents, the less vast are local news deserts and the less powerful is misinformation/disinformation).
11. Use AI tools and deliberative tools to improve efficiency of functioning of a larger legislative body.
12. Replace party primaries with all-comers preliminaries so bots and bad actors empowered to make more toxic content with LLMs can't capture the whole system by capturing a civil society organization (e.g. a party) that houses only a minority portion of the population.

What steps can the United States take to ensure that all individuals are equipped to interact with AI systems in their professional, personal, and civic lives?

The federal government should aim to support broader civic and digital education and should ensure that particular uses of AI technologies in public administration are both explainable and justifiable.

Information Ecosystems
- New fields such as prompt-engineering have emerged from a growing realization that specific expertise is required to most effectively use LLMs for desired ends.[11]
  - LLMs are most effective and useful when users are sufficiently educated about their internal operations and training procedures and are aware of the specific limitations of a given system. Some systems, for example, might be particularly well-suited to summon real-time data, while others might be better suited to longer written outputs or perhaps coding.

Decision-Making Procedures
- Challenge: The so-called black box problem poses a threat to the legitimacy of decision-making procedures. In contexts where AI tools (either predictive algorithms or LLMs) are used to make decisions or inform, or otherwise support policymakers or administrators as they make decisions, citizens ought to be able to request detailed, understandable explanations and

11 Popli, Nik. "AI Prompt Engineer Job" Time Magazine. 2023. https://time.com/6272103/ai-prompt-engineer-job/.

justifications. Whereas explanations might offer insight into the technical process by which a given system generates an output, justifications offer the reasoning behind the use of a given system.[12] In order to furnish and publicly debate the merits of explanations and justifications of the functionality and uses of AI systems for legitimate democratic decision-making procedures, users and stakeholders will require increased digital literacy and civic-mindedness.

Responsive Government and Healthy Democratic Communities

- Perhaps the most important way American institutions can support the effective and safe use of LLMs is through reforms meant to strengthen mutual trust in and knowledge of our democratic institutions.
  - To call upon experts for guidance through the use of LLMs without sacrificing trust in the efficacy and responsiveness of our institutions, we require broader civic education, an understanding of fellow citizens as "civic friends."[13]

Recommendations:

1. As part of the regulatory process, require firms to develop educational materials on the specific strengths and weaknesses of particular AI models as they are released (perhaps following pre-release audits and/or reporting) and develop materials on prompt-engineering best practices; these materials might be modeled on the educational materials that accompany the release of new pharmaceuticals.
2. Require companies developing and deploying LLMs to post visible consumer-facing information on their interface so users are aware they are interacting with AI systems or generated content, with clear information about their particular system's hallucination prevalence and risk. Consider this analogous to FDA reporting requirements for food and drug allergens and side effects.
3. Pass the Civics Secures Democracy Act, and include funding for civic education providers, and framework developers (for instance, the Educating for American Democracy Roadmap), to integrate education about AI and democracy.

Thank you for this opportunity to comment on this important topic. The GETTING-Plurality Research Network welcomes any further discussion and can be reached at contact@gettingplurality.org

EDMOND & LILY SAFRA Center for Ethics

JUSTICE, HEALTH & DEMOCRACY IMPACT INITIATIVE

12 Gillis, Talia B., and Josh Simons. "Explanation < Justification: GDPR and the Perils of Privacy." Pennsylvania Journal of Law and Innovation. 2019. https://www.law.upenn.edu/live/files/9790-gillis-and-simons-explanation-lt-justification.

13 Allen, Danielle. *Talking to Strangers.* University of Chicago Press, 2004.