PUBLIC CITIZEN
1600 20th Street, NW • Washington, D.C. 20009 • 202/588-1000 • www.citizen.org

Public Citizen Comments to the
President's Council of Advisors on Science and Technology
Working Group on Generative Artificial Intelligence
August 1, 2023

Founded in 1971, Public Citizen is a national public interest organization with more than 500,000 members and supporters. We have worked throughout our more than 50-year history to strengthen the nation's democracy and are pleased to offer these comments in response to the working group's crucial investigation into the intersection of generative artificial intelligence (AI) and democracy. These comments respond to each of the questions posed by the working group.

The rapid development of generative AI technology poses very serious threats to democratic integrity and maintenance of social trust.

Public Citizen believes that these threats – not to mention other extraordinary risks – are sufficient to merit an extended pause on further development and deployment of generative AI technologies. Society needs time to digest what these technologies portend and policymakers need time to craft appropriate controls and guardrails. The social price of delaying AI development and the benefits it may offer does not compare to the grave risks, many of which may not be remediable.

That said, we recognize that a moratorium, while justified, is unlikely to be agreed upon; and even if a moratorium were agreed upon, it would be necessary to recommend specific policies to put in place in advance of lifting the moratorium. Our comments here are presented against that backdrop.

In these comments, we briefly sketch several sets of concerns about AI, democracy and the destructive of trust. Then we offer a range of policy suggestions. Our overarching policy recommendation is: **All AI-generated content –videos, imagery, audio, text -- should be labeled as such.**

Overview of concerns

Deepfakes and elections: The most acute concern about AI, democracy and social trust is the use of deepfakes to persuade voters that a candidate for office did or said something that they in fact did not do or say. Political actors in the United States and overseas are starting to employ deepfake technology and there is every reason to suspect they will continue to do so, as the technology improves and so long as there is no legal proscription against employing deepfakes.[1] It is easy to imagine a blockbuster deepfake video being released shortly before an election, go "viral" on social media, with no ability for voters to determine that its claims are fraudulent, and determine the outcome of an election.

Disinformation and Misinformation: The ability of generative AI to produce nonrepetitive text almost instantaneously in response to prompts threatens to supercharge disinformation efforts. This may take the form of more frequent and higher quality disinformation posts on social media. It may also involve more sophisticated and elaborate disinformation efforts, such as the creation of websites pushing misinformation – a laborious task for humans, but not for generative AI tools. It is entirely possible that generative AI will enable the creation of vastly more disinformation of higher quality and in more sophisticated formats than anything society has previously experienced.[2]

Voter manipulation: As various generative AI tools improve and are connected to vast databases about individuals' preferences, interests, beliefs, purchasing practices, race, gender, religion, geography and much more, candidates and political actors will be able to deliver individualized, customized messages – even by human-appearing avatars that are refined to meet individual preferences – that may potentially be contradictory to other individualized, customized messages delivered to other individual voters. This portends a kind of manipulation that would upend and displace deliberative democracy.[3]

Undermining of social trust: The specific issues iterated above are particular manifestations of a broader threat posed by generative AI: the destruction of social trust. Deepfakes, dishonest text delivered and dressed up through sophisticated systems, and customized, individualized manipulation are issues that threaten to become pervasive – not just in electoral contexts, but in everyday activities, informal social interactions and commercial marketing and transactions. This malicious use of AI-generated media will foreseeably create very serious social harms, including attack porn, bullying, reputation destruction due to personal grievance, blackmail and commercial manipulation. Beyond these harms, pervasive use of AI-generated media in this context will erode social trust – the ability of people to believe what they see, hear and read online, including that they are dealing with actual humans. As a result, people will not just be fooled by fake information, they will refuse to believe true information, dismissing what is authentic as synthetic. The possible destruction of social trust threatens the very fabric of society, including our ability to carry out commercial transactions. It may make deliberative democracy impossible, or at minimum far more challenged; the very likely consequence is a severe

[1] See Tiffany Hsu and Steven Lee Myers, "A.I.'s Use in Elections Sets Off a Scramble for Guardrails," New York Times, June 25, 2023, https://www.nytimes.com/2023/06/25/technology/ai-elections-disinformation-guardrails.html; James Vincent, "DeSantis attack ad uses fake AI images of Trump embracing Fauci," The Verge, June 8, 2023, https://www.theverge.com/2023/6/8/23753626/deepfake-political-attack-ad-ron-desantis-donald-trump-anthony-fauci; https://www.politico.com/news/2023/07/17/desantis-pac-ai-generated-trump-in-ad-00106695; Ruth Michaelson, "Turkish presidential candidate quits race after release of alleged sex tape," The Guardian, May 11, 2023, https://www.theguardian.com/world/2023/may/11/muharrem-ince-turkish-presidential-candidate-withdraws-alleged-sex-tape.

[2] Josh A. Goldstein, Girish Sastry, Micah Musser, Renée DiResta, Matthew Gentzel, and Katerina Sedova, "Generative Language Models and Automated Influence Operations: Emerging Threats and Potential Mitigations," January 2023, https://arxiv.org/pdf/2301.04246.pdf,

[3] Archon Fung, Lawrence Lessig, "How AI Could Take Over Elections—And Undermine Democracy," Scientific American, June 7, 2023, https://www.scientificamerican.com/article/how-ai-could-take-over-elections-and-undermine-democracy.

intensification of political tribalism, as people choose to believe what reinforces their existing paradigm and to dismiss what contradicts it – a problem the nation already confronts but which likely would become orders of magnitude worse. In this way, all misleading or fraudulent AI-generated media – not just overtly political media – threatens democracy.

Recommendations on remedies

Even this brief review of AI threats to democracy and maintenance of social trust makes it clear that no individual remedy is adequate to the challenges faced. We believe the recommendations offered here should be helpful and many are urgently needed, but they are surely only part of the necessary solution set.

Our overarching recommendation is that all AI-generated content – videos, imagery, audio, text -- should be labeled as such. This standard should be adopted in law and embraced by all generative AI corporations as an industry norm. Such a rule will inevitably face enforcement challenges and be circumvented. But while enforcement is a real problem that will need to be addressed eventually, the immediate imperative is establishing legal requirements and social norms. Most people and businesses will respect the law and established norms. But if there are no clear rules of the road, we should anticipate chaotic consequences. More specific recommendations follow:

1. Ban all deepfakes in electoral politics. Public Citizen has petitioned the Federal Election Commission (FEC) to ban deepfakes to the extent possible under its existing "fraudulent misrepresentation" (52 U.S.C. §30124) authority.[4] The FEC and similarly empowered state election agencies should immediately employ the authority they have to ban or limit deepfakes. However, the FEC's fraudulent misrepresentation authority does not extend to independent committees (e.g., Super PACs) or outside organizations (e.g., 501(c)(4) social welfare organizations or trade associations). Achieving a complete ban on deepfakes in politics will require federal legislation. We believe this is single most urgent policy reform needed to address generative AI threats to democracy.

2. Ban all deepfakes, with a prominent disclosure that media content is AI-generated exempting it from the category of deepfake. Even beyond the election context, there are almost no conceivable legitimate purposes for deepfakes, which definitionally are intended to deceive, can do immense personal harm, and threaten to undermine social trust. Prominent disclosures will cure most of these problems and should be mandated by law and social norm. Exceptions can be created as needed for parody or other limited cases.

3. Require disclosure of all AI-generated content, including text. While there are already many valid uses for AI-generated content and may be many more as generative AI evolves and

[44] As we note in our petition: 1) the prohibition on fraudulent misrepresentation does not apply generally to the use of artificial intelligence in campaign communications, but only to deepfakes or similar communications; 2) the prohibition on fraudulent misrepresentation would not apply to cases of parody, where an opposing candidate is shown doing or saying something they did not, but where the purpose and effect is not to deceive voters and, therefore, where there is no fraud; and 3) the prohibition on fraudulent misrepresentation would not apply in cases where there is a sufficiently prominent disclosure that the image, audio or video was generated by artificial intelligence and portrays fictitious statements and actions; the fact of a sufficiently prominent disclosure would eliminate the element of deception and fraud. Public Citizen petition to the Federal Election Commission, July 13, 2023, https://www.citizen.org/article/second-submission-petition-for-rulemaking-to-clarify-that-the-law-against-fraudulent-misrepresentation-applies-to-deceptive-ai-campaign-communications/.

gains wider deployment, the spread of AI-generated content masquerading as authentic or authored/created by a human risks a massive increase in disinformation and misinformation and undermining of social trust. Disclosure can offset many of these risks and legislation should require disclosure in all cases, with limited exceptions. Additionally, AI corporations should label each output as "AI-generated." That consistent practice would help establish a universal social norm that AI-generated content should always be disclosed as such. Again, enforcement issues deserve attention, but enforcement challenges should not delay or deter practices that can immediately establish social norms in this early period of widespread generative AI use.

4. The United States government should commit not to use deepfakes or generative AI to influence other countries' elections, politics or military operations. This is ethically imperative and necessary to have credibility in demanding other countries' not influence U.S. society. The U.S. Department of Defense is currently seeking a contractor to develop deepfake tools for battlefield deployment.[5] It should withdraw that specific proposal request and renounce the use of deepfakes and deployment of AI media in other countries.[6]

5. Negotiate international agreements to manage AI and democracy threats, starting with a ban on governmental dissemination of deepfakes and anonymous, unlabeled AI-generated content in other nations. Managing the threats from AI will plainly require international agreements, because the technology is not going to be restricted to the United States. One of the simplest and more urgent agreements would prohibit governments and affiliated enterprises from spreading deepfakes and misleading AI-generated content in other nations.

6. Encourage private corporate actors to consense around a system to track media content's provenance through an open standard. We recognize there are technological challenges to developing a provenance standard, as well as potentially competing models and corporate interests. But if companies are going to release generative AI technologies with the ability to create synthetic media that appears authentic, they must accept the obligation to arrive at provenance standard imminently. A provenance standard – a complement to, not a substitute for, a prominent labeling requirement – would enable individuals to discover the true source of media (and any alterations), mitigate the various harms identified here, reduce (though not eliminate) enforcement challenges, and make it easier for platforms to identify AI-generated content in order to treat it appropriately.

7. Media and platform companies should be encouraged not to transmit undisclosed generative AI content – with a priority on agreeing to prohibit the posting and sharing of deepfakes. Most of the concerns raised here about generative AI and democracy are contingent on widespread dissemination of AI-generated content. For now, at least, whether content is widely disseminated depends on whether the major media and platform corporations facilitate or permit it. Blocking strategies will depend on developing technology to identify AI-generated content – which can be assisted by the AI companies (in many cases, one and the same as the platforms) – but media and platform corporations should adopt blocking policies even in advance of effective technological solutions, in order to establish social norms.

[5] Sam Biddle, "U.S. Special Forces Want to Use Deepfakes for Psy-Ops," The Intercept, March 6, 2023, https://theintercept.com/2023/03/06/pentagon-socom-deepfake-propaganda.

[6] Public Citizen letter to Secretary Lloyd Austin and Secretary Anthony Blinken, May 18, 2023, https://www.citizen.org/article/u-s-government-should-reject-deepfakes-in-foreign-affairs.

8. Adopt effective privacy protections and ban surveillance advertising. The concern about manipulative, customized content imagines a build out of current advertising and communications practices, based on pervasive surveillance of people online and aggregation of their detailed interests and histories. Effective protections against privacy and prohibitions on surveillance marketing strategies would work to curtail manipulative, customized misinformation

9. Online platform corporations should prohibit personalized advertising based on generative AI tools that customize messages to individuals. As a general matter, such AI-generated customized marketing is poised to overwhelm consumers and users' natural defenses and skepticism, and invites corporations and others to employ AI "representatives" who deliver contradictory messages to individuals. In the civic space, in particular, such individualized and customized messaging would pave the way for systemic manipulations. Platforms should refuse to let this occur.

10. Establish licensing standards for large language models and other generative AI technologies. A license system for corporations or organizations that wish to operate and deploy such powerful technologies would enable the imposition of upstream controls that could mitigate many of the threats to democracy.

11. Do not impose barriers to public commenting in rulemaking and other contexts. There is a legitimate worry about how generative AI may be used to overwhelm public commenting processes, an irony at the moment that the Biden administration is taking bold steps to open up and make the process more inclusive.[7] Public Citizen can state from experience that even steps that make commenting modestly more difficult dramatically reduce civic engagement; and we firmly believe that there is importance in broad engagement to signal support or opposition to proposals, including through form comments. Before moving to more burdensome approaches – many of which may prove a greater barrier to real people than bad actors empowered by AI – we recommend establishing a legally binding principle that comments may only be submitted by real people and requiring disclosure of any AI-generated text or content.

12. Be wary of encouraging skepticism about online media as a strategy to mitigate risks. While all persons, including especially children, will benefit from being taught or encouraged to use critical thinking and media literacy skills, at this time we believe encouraging the general public to search out AI-generated manipulation would be a wrong turn. The fear is that in trying to protect people from AI-generated manipulation, such an effort would further the eroding of social trust. Moreover, as AI technologies improve, it is highly likely that the broad public will not have the skills to identify distinguish AI-generated content from authentic content. By contrast, if there is disclosure, people will be able to evaluate appropriately what they see, hear and read.

Thank you for the opportunity to submit these comments.

Robert Weissman,
President, Public Citizen,
rweissman@citizen.org.

[7] Office of Information and Regulatory Affairs, "Broadening Public Engagement in the Federal Regulatory Process," July 19, 2023, https://www.whitehouse.gov/omb/information-regulatory-affairs/broadening-public-engagement-in-the-federal-regulatory-process.