Center for AI and Digital Policy

Comments of the
CENTER FOR AI AND DIGITAL POLICY (CAIDP)
to the
PRESIDENT'S COUNCIL OF ADVISORS ON SCIENCE AND TECHNOLOGY (PCAST) WORKING GROUP
on the
GENERATIVE ARTIFICIAL INTELLIGENCE (AI)

On behalf of the Center for AI and Digital Policy (CAIDP), we write to provide a detailed response to the specific questions raised in the White House notice "PCAST Working Group on Generative AI Invites Public Input."[1] We provided a preliminary response for your public session,[2] and established a webpage to track the work of the Working Group.[3] In our response below we reiterate our position in our preliminary response and further address the specific questions posed in the public notice. Our key recommendations are as follows:

1. *Ensure* the development of human-centered and trustworthy Artificial Intelligence based on fundamental rights, democratic values, and the rule of law

2. *Establish* guardrails for AI based on transparency, contestability, traceability, robustness, safety, security, and accountability. Companies should not release AI products that are not safe.

3. *Implement* the AI Bill of Rights, the OECD AI Principles, the UNESCO Recommendations on AI Ethics, and the Universal Guidelines for AI (UGAI).

*About CAIDP*

The Center for AI and Digital Policy (CAIDP) is an independent, non-profit organization that advises national governments and international organizations on artificial intelligence (AI) and digital policy, based in Washington, DC. CAIDP currently serves as an advisor on AI policy to the OECD, the Global Partnership on AI, the Council of Europe, the European Union, UNESCO,

[1] PCAST Working Group on Generative AI Invites Public Input, (May 13, 2023), https://www.whitehouse.gov/pcast/briefing-room/2023/05/13/pcast-working-group-on-generative-ai-invites-public-input/
[2] CAIDP, *Statement, PCAST*, https://www.caidp.org/app/download/8458230763/CAIDP-Statement-PCAST-AI-05142023.pdf
[3] CAIDP, *Resources, PCAST*, https://www.caidp.org/resources/pcast/

and other international and national organizations. We work with more than 600 AI policy experts in over 80 countries.

CAIDP supports AI policies that advance democratic values and promote broad social inclusion based on fundamental rights, democratic institutions, and the rule of law.[4] In April 2023, we released the third edition of our *Artificial Intelligence and Democratic Values* Index,[5] providing a comprehensive review of AI policies and practices in 75 countries. In our evaluation of the United States, we concluded that:

> The U.S. lacks a unified national policy on AI. The United States has endorsed the OECD/G20 AI Principles. . . . The overall U.S. policy-making process remains opaque and the Federal Trade Commission has failed to act on several pending complaints concerning the deployment of AI techniques in the commercial sector. But the administration has launched new initiatives and encouraged the OSTP, NIST, and other agencies to gather public input. The recent release of the Blueprint for an AI Bill of Rights by the OSTP represents a significant step forward in the adoption of a National AI Policy and in the U.S.'s commitment to implement the OECD AI Principles. . . . The absence of a legal framework to implement AI safeguards and a federal agency to safeguard privacy also raises concerns about the ability of the U.S. to monitor AI practices.[6]

CAIDP has endorsed the AI Bill of Rights[7] with specific recommendations in support of implementing this framework.[8] We acknowledge that it is a commitment to affirmatively advance civil rights, equal opportunity, and racial justice and to protect personal data from misuse by AI-powered algorithms.[9]

We also call to attention that the United States has endorsed the OECD AI Principles[10]. According to the OECD AI Principle on Accountability (1.5): "*AI actors should be accountable for the proper functioning of AI systems and for the respect of the above principles, based on their roles, the context, and consistent with the state of art."* (emphasis added).

---

[4] CAIDP Statements, https://www.caidp.org/statements/
[5] CAIDP, *Artificial Intelligence and Democratic Values* (2023), https://www.caidp.org/reports/aidv-2022/
[6] Id. at 1085.
[7] CAIDP, *Support the OSTP AI Bill of Rights*, https://www.caidp.org/statements/ostp/
[8] Lorraine Kisselburgh and Marc Rotenberg, *Next Steps on the AI Bill Of Rights*, Washington Spectator (Nov. 2021), https://washingtonspectator.org/author/lorraine-marc/; CAIDP, Public Voice, https://www.caidp.org/public-voice/
[9] Id. at ii.
[10] *U.S. Joins with OECD in Adopting Global AI Principles,* NTIA (May 22, 2019), https://www.ntia.doc.gov/blog/2019/us-joins-oecd-adopting-global-ai-principles

The Universal Guidelines for Artificial Intelligence ("UGAI") is a framework for AI governance based on the protection of human rights and was adopted in 2018 by the International Conference on Data Protection and Privacy Commissioners. The UGAI has been endorsed by more than 300 experts and 70 organizations in 40 countries. According to the UGAI Assessment and Accountability Obligation, "*An AI system should be deployed only after an adequate evaluation of its purpose and objectives, its benefits, as well as its risks. Institutions must be responsible for decisions made by an AI system.*"[11]

The PCAST Working Group has issued a public notice to invite input on how to identify and promote the beneficial deployment of generative AI, and on how best to mitigate risks. We support the initiative and appreciate the opportunity to provide comments.

**CAIDP-Specific Responses to Questions in Public Notice**

1. ***In an era in which convincing images, audio, and text can be generated with ease on a massive scale, how can we ensure reliable access to verifiable, trustworthy information? How can we be certain that a particular piece of media is genuinely from the claimed source?***

We need to establish governance mechanisms and accountability systems rooted in law to assign the responsibilities of the originators/developers of generative AI systems and to provide rights to those who are impacted by outcomes of AI systems.

First and foremost, developers of large language models must adhere to transparency and accountability practices in data collection and construction of datasets. As Bender, Gebru, and McMillan-Major explained:

> As a part of careful data collection practices, researchers must adopt frameworks to describe the uses for which their models are suited and benchmark evaluations for a variety of conditions. This involves providing thorough documentation on the data used in model building, including the motivations underlying data selection and collection processes. This documentation should reflect and indicate researchers' goals, values, and motivations in assembling data and creating a given model.[12]

[11] Public Voice, *Universal Guidelines for Artificial Intelligence*, Guideline 5, https://thepublicvoice.org/ai-universal-guidelines/

[12] Emily M. Bender, Timnit Gebru, Angelina McMillan-Major, Margaret Mitchell, On *the Dangers of Stochastic Parrots: Can Language Models Be Too Big, FAccT '21:*

AI tools can be deployed to help monitor news accuracy. We need to encourage investment and innovation in advanced AI models that can distinguish between human and AI-generated content. This is akin to building an 'AI immune system', which can detect and flag synthetic media by identifying subtle inconsistencies or anomalies that are characteristic of machine-generated outputs.[13] Additionally, leveraging blockchain technology may help by providing an immutable, distributed record of content creation, modifications, and distributions, which allows for source verification and traceability of digital artifacts,[14] though blockchain models also introduce privacy and data protection concerns that should be addressed prior to deployment.

There must be regulations governing the use of generative AI. Any commercial application using generative AI systems should transparently disclose the use of such technologies.[15] This would help mitigate risks of misinformation or impersonation.

***2. How can we best deal with the use of AI by malicious actors to manipulate the beliefs and understanding of citizens?***

We reiterate our recommendation for federal law governing AI that would set standards of liability and accountability of actors in the AI life cycle and supplement existing legislation. We believe that the key to effective AI accountability is to allocate rights and responsibilities for AI developers and users. Federal AI legislation based on the established governance frameworks outlined above, should include:

1. Identification of high-risk systems as those systems that adversely impact fundamental rights and/or civil liberties
2. Mandatory ex-ante human rights impact assessments
3. Third-party/Independent Certification, audit requirements should be required prior to deployment and during the life cycle of AI systems to ensure they remain robust, secure and safe
4. Disclosure requirements for public and private entities deploying AI systems

---

*Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency* (*March 202). Pages 610, 615 (Stochastic Parrots"),* https://doi.org/10.1145/3442188.3445922

[13] Melissa Heikklä, MIT Technology Review, *How to spot AI-generated text*, https://www.technologyreview.com/2022/12/19/1065596/how-to-spot-ai-generated-text/ (December 2022)

[14] Kathryn Harrison and Amelia Leopold, *How Blockchain Can Help Combat Disinformation*, https://hbr.org/2021/07/how-blockchain-can-help-combat-disinformation (July 19, 2021)

[15] UNESCO, *Recommendation on the Ethics of Artificial Intelligence*, (November 23, 2021) https://en.unesco.org/about-us/legal-affairs/recommendation-ethics-artificial-intelligence

5. Complaint and redress procedures should be established for impacted individuals or groups to challenge AI systems

We recommend that developers of foundation models as well as downstream users be held responsible for ensuring safety by design and implementing specific safeguards on transparency of data practices and disclosure on AI-generated content.

The Federal Trade Commission (FTC) has a unique opportunity at this time to curtail the malicious use of AI. CAIDP has filed an extensive complaint with the FTC. concerning OpenAI's business practices. We assert that the company has violated Section 5 of the FTC Act as well as the guidance that the FTC has announced for AI products.[16] The CAIDP complaint provides an immediate opportunity for the FTC to "tackle harms posed by the development and use of specific types of AI systems, such as large language models," as the PCAST public notice proposes for generative AI. The FTC has commenced an investigation and their findings will be integral to understanding agency capability to establish necessary guardrails for generative AI.[17]

PCAST can aslo take guidance from the UNESCO Recommendation on the Ethics of Artificial Intelligence[18] on increasing media and public literacy:

> The development of AI technologies necessitates a commensurate increase in data, media, and information literacy as well as access to independent, pluralistic, trusted sources of information, including as part of efforts to mitigate risks of misinformation, disinformation and hate speech, and harm caused through the misuse of personal data.

The companies that deploy Generative AI systems must establish fact-checking mechanisms and promote the use of trusted sources. To augment the trustworthy design of generative AI, there should be a collaboration with media organizations and companies to establish rigorous fact-checking processes, ensuring the dissemination of accurate information and comprehensive AI education and media literacy programs as recognized by UNESCO. In addition to accountability practices, the public must be able to recognize and evaluate AI-generated content effectively, fostering resilience against manipulation.

***3. What technologies, policies, and infrastructure can be developed to detect and counter AI-generated disinformation?***

[16] CAIDP, In the Matter of OpenAI, https://www.caidp.org/cases/openai/
[17] CNN, *FTC is ChatGPT-maker OpenAI for potential harm to consumers,* (July 13, 2023) https://www.cnn.com/2023/07/13/tech/ftc-openai-investigation/index.html
[18] UNESCO, *Recommendation on the Ethics of Artificial Intelligence,* (November 23, 2021), https://en.unesco.org/about-us/legal-affairs/recommendation-ethics-artificial-intelligence

The UNESCO Recommendation on AI ethics recommends ethical governance and stewardship for AI. UNESCO specifically requires Member States to implement a mechanism that requires AI actors to ensure that training data sets do not foster cultural, economic, or social inequalities, prejudice, the spreading of disinformation and misinformation, and disruption of freedom of expression and access to information. The United States recently rejoined UNESCO, citing specifically the need to carry forward work at UNESCO on AI ethics.[19]

AI-based detection of disinformation remains largely limited to hybrid, human–machine approaches in which human fact-checkers identify a piece of disinformation, and only thereafter is an AI model used to detect variations (i.e., identical or similar posts) of disinformation.[20] This leaves the risk of reactive remediation, and therefore, may require creating an AI-based counter-disinformation framework[21] which requires addressing challenges through regulatory, technology-oriented, and capacity-building measures.

In line with the UNESCO Recommendation on AI ethics, we propose the development of AI models that anayze patterns, context, and inconsistencies within content. These models will proactively identify AI-generated disinformation by detecting anomalies, biased narratives, or misinformation campaigns. There must be comprehensive media literacy programs to educate citizens about the risks of AI-generated disinformation, and accompanying regulations that address the responsible use of AI and hold malicious actors accountable for AI-generated disinformation and impose penalties for the deliberate dissemination of false or misleading information.

The four key priorities for an AI-based counter-disinformation framework include: (1) strengthening the ability to recognize contextual nuance and adapt to novel disinformation, (2) assessing the impact on human rights and fostering societal resilience through digital literacy, (3) building organizational capacity for AI adoption, and (4) prohibiting the use of AI systems that generate disinformation or lack meaningful human control.

4. ***How can we ensure that the engagement of the public with elected representatives—a cornerstone of democracy—is not drowned out by AI-generated noise?***

[19] UNESCO, *The United States of America announces its intention to rejoin UNESCO in July*, Press Release, June 12, 2023, https://www.unesco.org/en/articles/united-states-america-announces-its-intention-rejoin-unesco-july

[20] Linda Slapakova, *Towards an AI-Based Counter-Disinformation Framework*, RAND, (March 29, 2021), https://www.rand.org/blog/2021/03/towards-an-ai-based-counter-disinformation-framework.html

[21] Id.

The Public Broadcasting Service has reported that "The implications for the 2024 campaigns and elections are as large as they are troubling: Generative AI can not only rapidly produce targeted campaign emails, texts or videos, it also could be used to mislead voters, impersonate candidates and undermine elections on a scale and at a speed not yet seen."[22] "Public and private organizations must be obligated to disclose whenever content has been generated by generative AI, when that content may have an effect on decisions affecting con- sumers, consumer rights more broadly, or democratic processes."[23]

To ensure that the public continues to engage with elected officials and hold them accountable, companies that deploy AI systems that disseminate political information must meet essential transparency requirements. Akin to "nutrition labels", viewers should be able to see disclosures that outline information sources and fact-checking procedures. Social media companies must be responsible for implementing labeling mechanisms to ensure users can differentiate human-generated content from AI generated content. With such labeling mechanisms, users can see disclosures/warnings about content from political figures allowing constituents to follow the public messaging of their leaders.[24]

However, labels provide only partial transparency and partial accountability for AI systems. We favor robust mechanisms for algorithmic transparency that provide access to the logic, factors, and data that provide the basis for the outputs.[25]

Other strategies include watermarks for deepfakes, pass restrictions on the data collection practices that enable harmful political microtargeting and digital advertising, and invest in national media literacy education. While none of these have been established as fool-proof yet, now is the time to accelerate measures towards accountability and disclosure systems that can counter disinformation.

---

[22] PBS, *AI-generated disinformation poses threat of misleading voters in 2024 election,* (May 14, 2023), https://www.pbs.org/newshour/politics/ai-generated-disinformation-poses-threat-of- misleading-voters-in-2024-election

[23] Norwegian Consumer Council, Ghost in the machine – Addressing the consumer harms of generative AI (Jun. 2023), https://storage02.forbrukerradet.no/media/2023/06/generative-ai-rapport-2023.pdf

[24] Emily Saltz, Tommy Shane, Victoria Kwan, Claire Leibowicz, Claire Wardle, *It matters how platforms label manipulated media. Here are 12 principles designers should follow,* First Draft News, https://firstdraftnews.org/articles/it-matters-how-platforms-label-manipulated-media-here-are-12-principles-designers-should-follow/.

[25] UNESCO, Privacy expert argues "algorithmic transparency" is crucial for online freedoms at UNESCO knowledge café (Dec. 4, 2015) ("At the intersection of law and technology – knowledge of the algorithm is a fundamental right, a human right."), https://www.unesco.org/en/articles/privacy-expert-argues-algorithmic-transparency-crucial-online-freedoms-unesco-knowledge-cafe; Marc Rotenberg, *Artificial Intelligence and the Right to Algorithmic Transparency*, in The Cambridge Handbook of Information Technology, Life Sciences and Human Rights (Cambridge 2022)

Technology can also be leveraged as a tool to create new digital forms of constituent outreach through mobile apps and AI-powered services. However, this would need to be supplemented with electoral and campaign reforms that ensure elected representatives take responsibility for using generative AI tools.

5. ***How can we help everyone, including our scientific, political, industrial, and educational leaders, develop the skills needed to identify AI-generated misinformation, impersonation, and manipulation?***

We need to strengthen and adapt our laws to address the novel challenges posed by AI. Regulations should mandate the explicit disclosure when content is AI-generated, to ensure transparency.[26] Further, laws must be designed to penalize malicious uses of AI, thereby discouraging acts of misinformation and impersonation.

Partnerships between the public, private, and civil society sectors are crucial to harness the expertise available in all domains.[27] Private sector technology firms possess technical know-how and resources essential for developing and implementing AI detection tools. On the other hand, public entities can provide necessary regulatory oversight and ensure that the deployed tools are fair, and equitable, and do not infringe upon civil liberties. A collaborative dialogue between these sectors can result in effective strategies that both curtail the misuse of AI and uphold democratic values.

Education, based on the need to assess the impact of AI systems, must be institutionalized at all levels. AI literacy can be incorporated into school curricula, enabling future generations to understand and engage with AI critically.[28] Specialized training modules or workshops for leaders across sectors can equip them with the necessary skills to identify AI-generated content. Higher education institutions should encourage research on the societal impacts of AI, fostering a culture of understanding and vigilance towards AI's potential misuse.

[26] GovTrack, *AI Disclosure Act*, (June 27, 2023), https://govtrackinsider.com/ai-disclosure-act-would-require-all-ai-content-to-say-disclaimer-this-output-has-been-generated-9d9ff7993a03

[27] The White House, *National AI Research and Development Strategic Plan 2023 Update*, (May 2023), https://www.whitehouse.gov/wp-content/uploads/2023/05/National-Artificial-Intelligence-Research-and-Development-Strategic-Plan-2023-Update.pdf

[28] Melissa Heikkilä, *AI literacy might be ChatGPT's biggest lesson for schools*, MIT Technology Review, (April 12, 2023), https://www.technologyreview.com/2023/04/12/1071397/ai-literacy-might-be-chatgpts-biggest-lesson-for-schools/

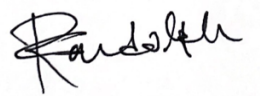
We are proposing an AI literacy strategy that emphasizes critical thinking, the ability to interrogate AI-generated outputs, and maintaining human control over AI systems. This is almost the precise opposite of replacing teachers with AI systems, as some have recommended.

We welcome this initiative from PCAST to draw public comment for addressing the critical risks to democracy posed by generative AI. Thank you for your consideration of our recommendations. We would welcome the opportunity to discuss this further.

Sincerely,

Marc Rotenberg
CAIDP Executive Director

Merve Hickok
CAIDP President



Christabel Randolph
CAIDP Law Fellow

Desmond Israel
CAIDP Research Assistant

Sunny Gandhi
CAIDP Research Assistant

Sneha Revanur
CAIDP Research Assistant

Center for AI and Digital Policy