November 7, 2023

Chairwoman Nancy Mace
Ranking Member Gerald E. Connolly
House Oversight Committee
Cybersecurity, Information Technology, and Government Innovation Subcommittee
Washington, DC

Re: CAIDP Statement for the Record: Hearing on "*Advances in Deepfake Technology*"

Dear Chairwoman Mace, Ranking Member Connolly, and Members of the Committee,

We write to you, on behalf of the Center for AI and Digital Policy (CAIDP)[1] regarding the hearing on *Advances in Deepfake Technology*.[2] We welcome this timely hearing.

CAIDP President Merve Hickok previously testified before your committee on *Advances in AI: Are We Ready For a Tech Revolution?*[3] As Hickok told this Subcommittee: "We do not have the guardrails in place, the laws that we need, the public education, or the expertise in government to manage the consequences of the rapid changes that are now taking place."[4]

Your hearing will help address these challenges. In this statement we ask the Committee to:

1. Urge the FTC to complete its investigation into OpenAI. The company's ChatGPT product may quickly become a leading source of misinformation and deep fakes.

2. Move forward AI legislation which mandates transparency, accountability, fairness, safety, and privacy for users. We endorse the Hawley-Blumenthal Bi-Partisan AI Act as a start to comprehensive legislation.

[1] Center for AI and Digital Policy, https://www.caidp.org.
[2] *Advances in Deepfake Technology*, OVERSIGHT HOUSE COMMITTEE, SUBCOMMITTEE ON CYBERSECURITY, INFORMATION TECHNOLOGY, AND GOVERNMENT INNOVATION (Oct. 24, 2023), https://oversight.house.gov/hearing/advances-in-deepfake-technology/
[3] Testimony and statement for the record of CAIDP President Merve Hickok, *Advances in AI: Are We Ready For a Tech Revolution?*, HOUSE COMMITTEE ON OVERSIGHT AND ACCOUNTABILITY, SUBCOMMITTEE ON
CYBERSECURITY, INFORMATION TECHNOLOGY, AND GOVERNMENT INNOVATION (Mar. 8, 2023), https://oversight.house.gov/wp-content/uploads/2023/03/Merve-Hickok_testimony_March-8th-2023.pdf.
[4] *Id.*

About CAIDP

The Center for AI and Digital Policy (CAIDP) is an independent research organization based in Washington, DC. We advise national governments and international organizations regarding artificial intelligence and digital policy. CAIDP currently serves as an advisor on AI policy to the OECD, the Global Partnership on AI, the European Union, the Council of Europe, UNESCO, and other national and international organizations. In April 2023, we released the third edition of our Artificial Intelligence and Democratic Values Index, providing a comprehensive review of AI policies and practices in 75 countries.[5]

1. The FTC must complete its investigation into OpenAI with urgency

In a complaint filed earlier this year with the Federal Trade Commission, CAIDP warned of the specific risks of generative AI products such ChatGPT.[6] OpenAI concedes many of the problems described in our complaint. According to OpenAI, even a powerful model such as GPT-4 "is not fully reliable."[7] In its own system card, OpenAI has acknowledged that GPT-4 will generate targeted content "intended to mislead." In a section describing disinformation, Open AI has stated that "GPT-4 can generate plausibly realistic and targeted content, including news articles, tweets, dialogue, and emails."[8] (emphasis added)

In our complaint, CAIDP explained that "The increasing commercialization of AI models will reduce the forms of oversight, transparency, and independent review that have traditionally characterized scientific research."[9] That is occurring now. Just this week, OpenAI ramped up the release of unregulated commercial AI and with a vision of GPT that allows "anyone to create a tailored version of ChatGPT." OpenAI said that "Anyone can easily build their own GPT—no coding is required."[10]

In September, OpenAI revised the system for DALL-E, a commercial product for AI-generated art,[11] and incorporated ChatGPT.[12] OpenAI has announced "voice and image capabilities in ChatGPT."[13] OpenAI has announced its vocal technology will be "capable of crafting realistic

[5] CAIDP, Artificial Intelligence and Democratic Values (2023), https://www.caidp.org/reports/aidv-2022/.
[6] CAIDP, *In the Matter of OpenAI* (2023), https://www.caidp.org/cases/openai/.
[7] OpenAI, GPT-4 Technical Report, (Mar. 27, 2023), https://cdn.openai.com/papers/gpt-4.pdf.
[8] OpenAI, *The GPT-4 System Card*, (Mar.15, 2023), https://cdn.openai.com/papers/gpt-4-system-card.pdf.
[9] CAIDP, *In the Matter of OpenAI* (2023), para. 127, https://www.caidp.org/cases/openai/.
[10] OpenAI, *Introducing GPTs*, (Nov. 6, 2023), https://openai.com/blog/introducing-gpts
[11] Emilia David, *OpenAI releases third version of DALL-E*, The Verge, (Sept. 20, 2023), https://www.theverge.com/2023/9/20/23882009/class-action-lawsuit-openai-privacy-dropped.
[12] Will Knight, *OpenAI's Dall-E 3 is an Art Generator Powered by ChatGPT*, Wired, (Sept. 20, 2023), https://www.wired.com/story/dall-e-3-open-ai-chat-gpt/.
[13] OpenAI Blog, *ChatGPT can now see, hear and speak,* (Sept. 25, 2023), https://openai.com/blog/chatgpt-can-now-see-hear-and-speak.

synthetic voices from just a few seconds of real speech."[14] On November 6, "OpenAI released 128k context GPT-4 Turbo at 1/3rd the price. This means GPT-4 is cheaper, faster, and can now read a prompt as long as an entire book."[15]

The FTC must complete this investigation to protect election integrity, public safety, and protect American consumers from AI generated fraud and harm.

(a) Protecting Election Integrity: AI can manipulate voters by the very nature of machine learning. "[AI] language models will allow the production of linguistically distinct messaging" in a way that humans can not immediately detect as machine-generated.[16] OpenAI has already admitted that GPT-4 could "rival human propagandists" and generate propaganda tailored to the specifics of the target audience's language patterns and the propaganda's goals.[17]

These techniques could be readily adopted in campaign advertising that will mislead and misinform voters and undermine democratic institutions. *(emphasis added)*

As we also explained to the FTC, CISA Director Jen Easterly called for stronger guardrails for ChatGPT: "Countering disinformation is about to get much harder: In the near term, ChatGPT and similar chatbots powered by large language models, or LLMs, will let threat actors master a range of malicious activities, including manufacturing more believable lies at scale."[18]

Because, large language models are trained on large amounts of data scraped from the internet, they can incorporate, reflect, and potentially amplify biases in such data."[19] False AI-generated political content also impact vulnerable populations– those of low-income, low formal

[14] TechCrunch, *OpenAI gives ChatGPT a voice for verbal conversations,* (Sept. 25, 2023), https://techcrunch.com/2023/09/25/openai-chatgpt-voice/.

[15] Rowan Cheung, *X/Twitter Post*, (Nov. 6, 2023), https://x.com/rowancheung/status/1721591682712629526?s=20

[16] 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,* Georgetown Center for Security and Emerging Technology, Stanford Internet Observatory, OpenAI, (Jan. 2023), pg. 2, https://arxiv.org/pdf/2301.04246.pdf

[17] Philippe Lorenz et al., *Initial policy considerations for generative artificial intelligence*, OECD, pp.13 (Sept. 18, 2023), https://www.oecd-ilibrary.org/science-and-technology/initial-policy-considerations-for-generative-artificial-intelligence_fae2d1e6-en

[18] CAIDP, *In Re OpenAI,* Supplement to Original Complaint, (Jul. 10, 2023), pp. 18-19, https://www.caidp.org/app/download/8466615863/CAIDP-FTC-Supplement-OpenAI-07102023.pdf.

[19] Congressional Research Services (CRS), *Generative Artificial Intelligence: Overview, Issues, and Questions for Congress,* In Focus, (Jun. 9, 2023), https://crsreports.congress.gov/product/pdf/IF/IF12426.

education, and the elderly.[20] Connected with personal information, LLMs can be used to create personalized disinformation, mimic and manipulate the behavior of the individuals. Political actors have already begun to use generative AI for their agendas.[21] The photos are surprisingly realistic. As Presidential elections become closer, false AI generation will only increase.[22]

Generative AI and deepfakes amplify human rights challenges of trust, evidence, and efficacy.[23]The Organization for Economic Cooperation and Development (OECD) has highlighted that the combination of AI-generated text, images, and disinformation can "erode societal trust in the information ecosystem and the fact-based exchange of information that underpins science, evidence-based decision-making, and democracy."[24]

(b) Protecting consumers and children: ChatGPT is being integrated in a wide variety of downstream consumer-facing services. "According to the FTC, "[w]ithin just a few months, generative AI chatbots and applications have launched and scaled across industries and reached hundreds of millions of people. AI is increasingly becoming a basic part of daily life."[25] ChatGPT is now integrated with Microsoft's Bing search service and includes advertising.[26] ChatGPT plugins are currently being deployed in a numerous consumer facing applications including restaurant reservations, ordering groceries, booking trips, language tutoring, family and home management solutions, and online shopping.[27]

[20] Brian Kennedy et al., *Public Awareness of Artificial Intelligence in Everyday Activities*, Pew Research Center, (Feb. 15, 2023), https://www.pewresearch.org/science/2023/02/15/public-awareness-of-artificial-intelligence-in-everyday-activities/.

[21] Tiffany Hsu and Steven Lee Myers, *A.I.'s Use in Elections Sets Off a Scramble for Guardrails*, N.Y. Times, (Jun. 25, 2023), https://www.nytimes.com/2023/06/25/technology/ai-elections-disinformation-guardrails.html.

[22] *Id*.

[23] Sam Gregory, *Fortify the Truth: How to Defend Human Rights in an Age of Deepfakes and Generative AI,* Journal of Human Rights Practice, 2023, huad035, (Sept. 06, 2023), https://doi.org/10.1093/jhuman/huad035

[24] Philippe Lorenz et al., *Initial policy considerations for generative artificial intelligence*, OECD, pp.13 (Sept. 18, 2023), https://www.oecd-ilibrary.org/science-and-technology/initial-policy-considerations-for-generative-artificial-intelligence_fae2d1e6-en.

[25] CAIDP, *Supplement to the Original Complaint, In the Matter of Open AI* (2023), pg. 2, para. 4, https://files.constantcontact.com/dfc91b20901/72cccde7-44a7-44e4-bfee-d6801b3891d2.pdf.

[26] *Id.,* at pg. 23, para. 98; *See also,* Yusuf Mehdi, *Driving more traffic and value to publishers from the new Bing,* Microsoft Bing Blogs, (Mar. 29, 2023), *https://blogs.bing.com/search/march_2023/Driving-more-traffic-and- value-to-publishers-from-the-new-Bing*

[27] *Id.,* at pg. 24, para. 101; *see also,* OpenAI, *ChatGPT Plugins,* https://openai.com/blog/chatgpt-plugins

Generative AI tools expose consumers to cyber fraud and deception. Legal experts at the American Bar Association have noted that “cybercriminals and other bad actors now have a tool at their fingertips to make scam-ready content with a sheen of legitimacy.”[28]

Emily Bender, a linguistics professor at the University of Washington said that “ChatGPT’s own plugin red team members found they could ‘send fraudulent or spam emails, bypass safety restrictions, or misuse information sent to the plugin.[29]

Deepfakes lower the threshold and cost for malicious actors to manipulate individuals. Without regulation and liability, deepfakes will undermine sectors dependent on user trust – such as banking, trading, or medicine. For example, deepfake video and text content can be deployed to target a specific corporation. Algorithmic trading systems can immediately pick up on content and conduct trade activity. Similarly, individual investors can act on the content, creating a down or upward spiral for a particular company stock.

In April 2023, Snap Inc. integrated ChatGPT into Snap – used by many children. There is no opt-out from the chatbot in the standard, free version of the app. The information page on Snapchat states My AI “may include biased, incorrect, harmful or misleading content” and suggests that users should independently verify any advice it gives before acting on it.[30]

ChatGPT has demonstrated negative effects of providing misinformation, incorrect advice, creation of bullying content. The inclusion of a conversational feature in social media apps used predominantly by children is likely to result in higher amounts of data scraped from children, increased screen-time and endanger online child safety.[31]*(emphasis added)*

The Federal Trade Commission (FTC) should move forward and conclude its investigation of OpenAI initiated by CAIDP. Subsequent to the filing of the CAIDP Complaint regarding OpenAI, consumer agencies around the world have launched investigations of ChatGPT.[32] The Federal Trade Commission (FTC) now has a unique opportunity to establish guardrails for AI. Earlier this year, CAIDP filed a detailed complaint with the FTC regarding OpenAI. The FTC has opened the investigation of OpenAI we requested.[33] This is clearly a

[28] *Id.,* at pg. 16, para. 69.
[29] *Id.,* at pg. 18, para. 77
[30] *Id.,* at pg. 25-26, para. 105.
[31] *Id.,* at pg. 26, para. 107
[32] CAIDP, *Supplement to the Original Complaint, In the Matter of Open AI* (2023), pg. 2, para. 6, https://files.constantcontact.com/dfc91b20901/72cccde7-44a7-44e4-bfee-d6801b3891d2.pdf.
[33] Cecilia Kang and Cade Metz, *F.T.C. Opens Investigation Into ChatGPT Maker Over Technology’s Potential Harms*, The New York Times, (Jul. 13, 2023), https://www.nytimes.com/2023/07/13/technology/chatgpt- investigation-ftc-openai.html; John D.

positive development, but now the FTC must prioritize this investigation. It took two years from the time we filed similar complaints with the FTC concerning Google and Facebook before there was a settlement.[34] We can't wait that long this time. AI products are evolving rapidly and being deployed downstream in consumer facing services. Before the end of this year the FTC must complete the investigation of ChatGPT and enter into a settlement with OpenAI that ensures the companies will abide by the practices for AI companies the Commission has previously issued.

2. ***Move forward AI legislation to protect Americans. We endorse the Bi-partisan Hawley-Blumenthal Framework which mandates transparency, accountability, and protection of consumers and kids***

Generative AI can be deceptive, manipulative, and unreliable. AI generated deepfakes pose the significant threats to democratic processes, human rights protection, and individual autonomy and privacy. Concerns about the misuse of deepfakes to manipulate elections, perpetuate fraud in business, alter public opinion and threaten national security have dominated the discussion about deepfakes.[35]

Deepfake technology is predominately being used to create sexual videos of women without their consent.[36] Research shows around 96 per cent of deepfake videos are pornographic, with almost 100 per cent of them involving non-consenting women.[37] In 2019, an AI-enabled scam using a deepfake defrauded €220,000 from a UK firm.[38]

McKinnon and Ryan Tracy, *ChatGPT Comes Under Investigation by Federal Trade Commission,* Wall Street Journal, (Jul. 13, 2023), https://www.wsj.com/articles/chatgpt-under-investigation-by- ftc-21e4b3ef.

[34] Federal Trade Commission, *Facebook Settles FTC Charges That It Deceived Consumers By Failing To Keep Privacy Promises*, (Nov. 29, 2011), https://www.ftc.gov/news-events/news/press-releases/2011/11/facebook-settles- ftc-charges-it-deceived-consumers-failing-keep-privacy-promises

[35] Suzie Dunn, *Women, Not Politicians, Are Targeted Most Often by Deepfake Videos,* CENTER FOR INTERNATIONAL GOVERNANCE INNOVATION, (Mar. 3, 2021), https://www.cigionline.org/articles/women-not-politicians-are-targeted-most-often-deepfake-videos/

[36] *Id.*

[37] BBC, *Deepfake Porn: Could you be next?* (Jan. 5, 2023), https://www.bbc.co.uk/programmes/m001c1mt

[38]Heather Frase and Owen Daniels, *Understanding AI Harms: An Overview,* CENTER FOR SECURITY AND EMERGING TECHNOLOGY, (Aug. 11, 2023), https://cset.georgetown.edu/article/understanding-ai-harms-an-overview/

A majority of Americans cannot identify when they are in an AI-powered interaction.[39] Vulnerable populations, such as the elderly and those with low formal education, show more risk of AI-powered manipulation.[40]

Advanced deepfake detectors, like Intel's "FakeCatcher", do not always get identification or verification of deepfakes right.[41] Technological solutionism will not be sufficient. We need regulation. *(emphasis added)*

"Even if deepfakes are obviously fabricated or quickly exposed as such, they still contribute to a decaying information space. They can undermine public trust in democratic processes, incentivize activists and journalists to self-censor, and drown out reliable and independent reporting. AI-generated imagery that sensationalizes outrage on divisive topics can also entrench polarization and other existing tensions within society. In extreme cases, it could galvanize violence against individuals or whole communities. The impact of AI-generated disinformation will deepen as the quality and quantity of the technology's output continues to exceed the capacity of observers, moderators, or regulators to detect, debunk, or remove it."[42]

The Biden-Harris Executive Order on "Safe, Secure, and Trustworthy Artificial Intelligence"[43] is a step forward in identifying guardrails for AI systems. Much of the Executive Order also re-states the current authorities that the federal government already has to promote competition, protect civil rights, train workers. The President has called upon Congress to pass bi-partisan privacy legislation to protect American consumers and, especially, children online.[44]

There are well-established AI governance frameworks that can be implemented through federal legislation. The United States is a signatory to the OECD AI Guidelines.[45] The Universal Guidelines for Artificial Intelligence (UGAI) is a framework for AI governance based on the

[39] Brian Kennedy et al., *Public Awareness of Artificial Intelligence in Everyday Activities*, PEW RESEARCH CENTER, (Feb. 15, 2023), https://www.pewresearch.org/science/2023/02/15/public-awareness-of-artificial-intelligence-in-everyday-activities/

[40] *Id.*

[41] BBC, *Intel's deepfake detector tested on real and fake videos,* (Jul. 23, 2023), https://www.bbc.com/news/technology-66267961

[42] Allie Funk, Adrian Shahbaz, Kian Vesteinsson, *The Repressive Power of Artificial Intelligence,* FREEDOM OF THE NET 2023, Freedom House Report, https://freedomhouse.org/report/freedom-net/2023/repressive-power-artificial-intelligence

[43] The White House, *FACT SHEET: President Biden Issues Executive Order on Safe, Secure, and Trustworthy Artificial Intelligence,* Statements and Releases, (Oct. 30, 2023), https://www.whitehouse.gov/briefing-room/statements-releases/2023/10/30/fact-sheet-president-biden-issues-executive-order-on-safe-secure-and-trustworthy-artificial-intelligence/

[44] *Id.*

[45] *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.

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.

We need federal AI legislation that would set standards of liability and accountability of actors in the AI life cycle. Providers of high-risk systems should be obligated to conduct ex-ante "human rights impact assessment." Accountability mechanisms should also incorporate independent, third-party audits during the lifecycle of AI systems not just pre-deployment. Disclosure obligations should apply to public and private entities deploying AI systems and should also mandate explicit disclosure when content (including text, imagery, and other audio-visual content) is AI-generated. Developers of foundation models as well as downstream users should be held responsible for ensuring safety by design and implementing specific safeguards on transparency of data practices and disclosure on AI-generated content.[46] Red teaming is one of the several post hoc control mechanisms. However, depending solely on red teaming to fix issues after a system is already designed and developed is not enough for robust fairness and accountability.

There are various bi-partisan bills in Congress to address deep fakes. Senators Coons, Hawley, Klobuchar, and Collins have sponsored S.2770, the Protection Elections from Deceptive AI Act, to prohibit deceptive AI-generated content.[47] Senators Hawley and Blumenthal have presented the Bipartisan AI Act, a framework for AI regulation.[48] The Bipartisan AI Act notes that "A.I. system providers should be required to watermark or otherwise provide technical disclosures of A.I.-generated deepfakes."[49] *(emphasis added)*

We endorse the Hawley-Blumenthal bipartisan AI Act which offers a comprehensive framework for the governance of AI.

[46] CAIDP, *Statement for the Record: Joint Committee Hearing on Balancing Knowledge and Governance: Foundations for Effective Risk Management of Artificial Intelligence,* (Oct. 18, 2023), https://www.caidp.org/app/download/8481800363/CAIDP-HSC-AI-1018203.pdf?t=1698160421

[47] Text - S.2770 - 118th Congress (2023-2024): Protect Elections from Deceptive AI Act, S.2770, 118th Cong. (2023), https://www.congress.gov/bill/118th-congress/senate-bill/2770/text.

[48] S.1993 - 118th Congress (2023-2024): A bill to waive immunity under section 230 of the Communications Act of 1934 for claims and charges related to generative artificial intelligence, S.1993, 118th Cong. (2023), https://www.congress.gov/bill/118th-congress/senate-bill/1993.

[49] Senator Richard Blumenthal and Senator Josh Hawley, Bipartisan Framework for U.S. AI Act (2023), https://www.blumenthal.senate.gov/imo/media/doc/09072023bipartisanaiframework.pdf.

Thank you for your consideration of our views. We would be pleased to provide you and your staff with additional information. We ask that this statement be included in the hearing record.

Sincerely,

Merve Hickok
CAIDP President

Marc

Marc Rotenberg
CAIDP Executive Director

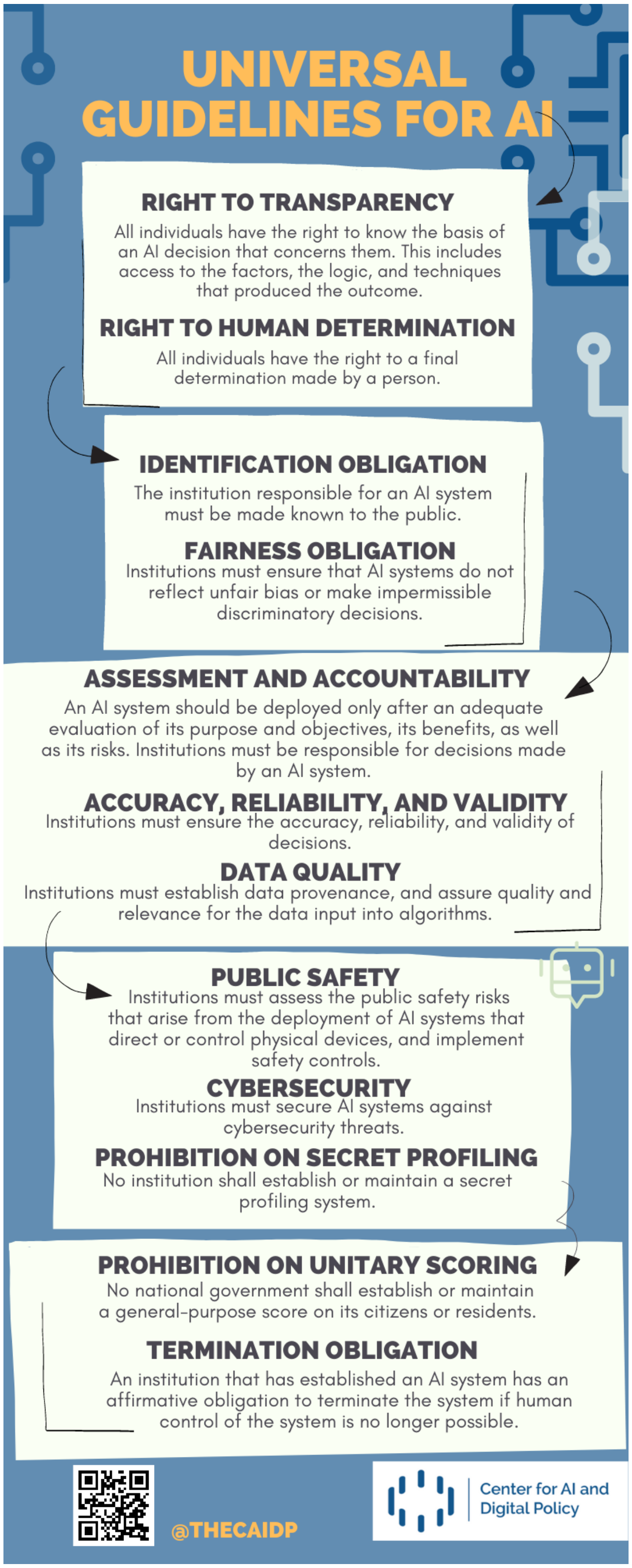
Randolph

Christabel Randolph
CAIDP Law Fellow

Brianna Rodriguez

Brianna Rodriguez
CAIDP Law Fellow

Center for AI and
Digital Policy



UNIVERSAL GUIDELINES FOR AI

RIGHT TO TRANSPARENCY

All individuals have the right to know the basis of an AI decision that concerns them. This includes access to the factors, the logic, and techniques that produced the outcome.

RIGHT TO HUMAN DETERMINATION

All individuals have the right to a final determination made by a person.

IDENTIFICATION OBLIGATION

The institution responsible for an AI system must be made known to the public.

FAIRNESS OBLIGATION

Institutions must ensure that AI systems do not reflect unfair bias or make impermissible discriminatory decisions.

ASSESSMENT AND ACCOUNTABILITY

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.

ACCURACY, RELIABILITY, AND VALIDITY

Institutions must ensure the accuracy, reliability, and validity of decisions.

DATA QUALITY

Institutions must establish data provenance, and assure quality and relevance for the data input into algorithms.

PUBLIC SAFETY

Institutions must assess the public safety risks that arise from the deployment of AI systems that direct or control physical devices, and implement safety controls.

CYBERSECURITY

Institutions must secure AI systems against cybersecurity threats.

PROHIBITION ON SECRET PROFILING

No institution shall establish or maintain a secret profiling system.

PROHIBITION ON UNITARY SCORING

No national government shall establish or maintain a general-purpose score on its citizens or residents.

TERMINATION OBLIGATION

An institution that has established an AI system has an affirmative obligation to terminate the system if human control of the system is no longer possible.

@THECAIDP

Center for AI and Digital Policy