Center for AI and Digital Policy

AI-RFI@nitrd.gov

Comments of the

THE CENTER FOR AI AND DIGITAL POLICY

To the Office of Science and Technology Policy, on behalf of the National Science and Technology Council's (NSTC) Select Committee on Artificial Intelligence (Select Committee), the NSTC Machine Learning and AI Subcommittee (MLAI-SC), the National AI Initiative Office (NAIIO), and the Networking and Information Technology Research and Development (NITRD) National Coordination Office (NCO), on the

National Artificial Intelligence Research and Development Strategic Plan

March 4, 2022

On behalf of the Center for AI and Digital Policy (CAIDP), we write in response to the RFI request on the National Artificial Intelligence Research and Development Strategic Plan (the "AI Strategic Plan").[1]

CAIDP is an independent non-profit organization that advises national governments and international organizations on artificial intelligence (AI) and digital policy. We work with more than 100 AI policy experts in almost 40 countries. In February 2022, we released the second edition of our report, *Artificial Intelligence and Democratic Values Index*[2], providing a comprehensive review of the AI policies and practices in 50 countries. Using a methodology to assess AI policies against democratic values and human rights, the Index includes detailed narrative reports, quantitative assessments, and ratings and rankings across a dozen metrics to measure progress towards human-centric and trustworthy AI values. The CAIDP currently serves as an advisor on AI policy to the OECD, the Global Partnership on AI, the Council of Europe, the European Union, and other international and national organizations.

We strongly support OSTP's proposals to update the AI Strategic Plan and appreciate the opportunity to provides comments. Our comments focus on:

Strategy 3: *Understand and address the ethical, legal, and societal implications of AI*;
Strategy 4: *Ensure the safety and security of AI systems*;

[1] Office of Science and Technology Policy, *RFI request on the National Artificial Intelligence Research and Development Strategic Plan* (Feb. 22, 2022) ("OSTP RFI on AI Strategic Plan"), https://www.federalregister.gov/documents/2022/02/02/2022-02161/request-for-information-to-the-update-of-the-national-artificial-intelligence-research-and

[2] CAIDP, *Artificial Intelligence and Democratic Values Index* (2022), https://www.caidp.org/reports/aidv-2021/

Strategy 6: *Measure and evaluate AI technologies through standards and benchmarks*; and Strategy 7: *Better understand the national AI R&D workforce needs*.

CAIDP has already endorsed the AI Bill of Rights,[3] one of the OSTP's six policy priorities, and made specific recommendations for that initiative.[4] We recommended a small number of clear, powerful principles and unnecessary qualifiers, loopholes, and exceptions. We suggested building on prior AI policy initiatives such as the OECD AI Principles and the Universal Guidelines for AI (UGAI).[5] In October 2018, over 250 organizations and experts, representing more than 30 countries and including the American Association for the Advancement of Science, endorsed the UGAI.[6] The Universal Guidelines for AI are intended to maximize the benefits of AI, to minimize the risk, and to ensure the protection of human rights. UGAI, already widely endorsed by the AI community, provides a good starting point but there is more to do.

Regarding the AI Bill of Rights, CAIDP also urges proceeding on a bipartisan basis. Eliminating bias, promoting fairness, ensuring accountability, and transparency for AI-based systems could also help align the political parties behind a common national purpose.

We also call your attention to the 2022 G7 Leader statement endorsing "Human-Centric AI", calling for "robust transparency" to oppose algorithmic bias.[7] This is a powerful statement from world leaders to address a problem that OSTP has identified as one of the great challenges in the AI field. The G7 leaders, including the United States, also committed to working together for a "values-driven digital ecosystem for the common good that enhances prosperity in a way that is sustainable, inclusive, transparent and human-centric." They called for a "human-centric approach to artificial intelligence," building on the work of the Global Partnership for Artificial Intelligence (GPAI) advanced by the Canadian and French G7 Presidencies in 2018 and 2019 and looking forward to the GPAI Summit in Paris in November 2021.

---

[3] The White House, *Join the Effort to Create A Bill of Rights for an Automated Society* (Nov. 10, 2021), https://www.whitehouse.gov/ostp/news-updates/2021/11/10/join-the-effort-to-create-a-bill-of-rights-for-an-automated-society/

[4] 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/

[5] *OECD AI Principles* (2019), https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449; The Public Voice, *Universal Guidelines for Artificial Intelligence* (2018) ("Universal Guidelines for AI"), https://thepublicvoice.org/AI-universal-guidelines/

[6] The Public Voice, *Universal Guidelines for Artificial Intelligence* – Endorsements (2018) https://thepublicvoice.org/AI-universal-guidelines/endorsement/

[7] *G7 Leaders Endorse Human-Centric AI, Call Out Bias,* (June 13, 2021), https://www.whitehouse.gov/briefing-room/statements-releases/2021/06/13/carbis-bay-g7-summit-communique/; see also *G7 Leaders Endorse Human-Centric AI, Call Out Bias,* CAIDP Update 2.24 (June 14, 2021), https://www.caidp.org/app/download/8326521963/CAIDP-Update-2.24.pdf

We write now to renew CAIDP's earlier recommendations, encourage the adoption of the AI Bill of Rights, and make additional proposals to advance the goals set out in the AI Strategic Plan.

Review/Recommendations for *Strategy 3: Understand and Address the Ethical, Legal, and Societal Implications of AI*

We applaud the goal of addressing the ethical, legal, and societal implications in AI. We further support the emphasis on fairness, transparency, and accountability as foundational values in designing ethical AI systems.

The Universal Guidelines for AI emphasize similar points. The Fairness Obligation (UGAI-4) states that institutions must ensure that AI systems do not reflect unfair bias or make impermissible discriminatory decisions. The Fairness Obligation recognizes that all automated systems make decisions that reflect bias, but such decisions should not be normatively unfair or impermissible. There is no simple answer to the question on what is unfair or impermissible. The evaluation often depends on context, but the fairness obligation makes clear that an assessment of objective outcomes alone is not sufficient to evaluate a system. Normative consequences must be assessed, including those that preexist or may be amplified by an AI system. As OSTP Director Alondra Nelson has explained, the OSTP should be "open about the history of science and technology's flaws and failures."[8] The consequences of the deployment of technology must be assessed with an understanding of the past, and a future lens that protects human dignity and civil rights.

Strategy 3 could be further strengthened to incorporate *considerations related to sustainability, and environmental issues.*

Problem: Greater emphasis on research of societal, ethical implications of AI-related to sustainability required.

The National AI R&D Strategic Plan implements the National AI Initiative (NAII) Act of 2020.[9] This includes action to: "support research and other activities on ethical, legal, environmental, safety, security, bias, and other appropriate societal issues related to artificial intelligence." The OSTP AI Strategic Plan calls attention to "societal issues such as equity and

[8] Khari Johnson, *Alondra Nelson wants to make science and tech more just*, Wired (June 29, 2021), https://www.wired.com/story/alondra-nelson-make-science-tech-more-just/

[9] House of Representatives, National Defense Authorization Act for Fiscal Year 2021 (2020), https://www.congress.gov/116/crpt/hrpt617/CRPT-116hrpt617.pdf#page=1210, 1210

climate change."[10] Moreover, Director Nelson has highlighted "groundbreaking clean energy investments" among six policy priorities for the agency.[11]

The need to focus on environmental issues for AI is timely.[12] The UNESCO Recommendation on the Ethics of AI focuses specifically on Protecting the Environment.[13] As the UNESCO Recommendation states:

> The Recommendation emphasises that AI actors should favour data, energy and resource-efficient AI methods that will help ensure that AI becomes a more prominent tool in the fight against climate change and on tackling environmental issues. The Recommendation asks governments to assess the direct and indirect environmental impact throughout the AI system life cycle. This includes its carbon footprint, energy consumption and the environmental impact of raw material extraction for supporting the manufacturing of AI technologies. It also aims at reducing the environmental impact of AI systems and data infrastructures. It incentivizes governments to invest in green tech, and if there are disproportionate negative impact of AI systems on the environment, the Recommendation instruct that they should not be used.[14]

AI should also be aligned with the United Nations Sustainable Development Goals[15] including cross-cutting environmental issues, as additionally emphasized by the OECD AI Principles, which have been endorsed by the United States.[16]

As it stands, Strategy 3 says little about environmental impact and sustainability. The Strategy should be revised to consider the carbon footprint of AI, modeling and data infrastructure, environmental degradation, and waste concerns.

Recommendation 1: CAIDP recommends an interdisciplinary perspective in developing, designing, and managing AI, specifically including environmental and climate research perspectives. The call for multidisciplinary perspectives lacks environmental science, ecosystem and resource management, as well as social science. OSTP Deputy Director Dr. Jane Lubchenco

[10] OSTP RFI on AI Strategic Plan.
[11] OSTP, The Director's Office (2022), https://www.whitehouse.gov/ostp/directors-office/
[12] Intergovernmental Panel on Climate Change (IPCC), Sixth Assessment Report (Feb. 28, 2022), https://www.ipcc.ch/assessment-report/ar6/
[13] *UNESCO Recommendation on the Ethics of AI* (2021), https://unesdoc.unesco.org/ark:/48223/pf0000377897
[14] UNESCO, *UNESCO member states adopt the first ever global agreement on the Ethics of Artificial Intelligence* (Nov. 25, 2021), https://en.unesco.org/news/unesco-member-states-adopt-first-ever-global-agreement-ethics-artificial-intelligence
[15] *United Nations Sustainable Development Goals* (2015) https://sdgs.un.org/goals
[16] *OECD AI Principles* (2019), https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449

made this point at the recent White House Climate Roundtable.[17] While it is a positive step to call for the inclusion of interdisciplinary perspectives including engineering and "other disciplines,"[18] there is a clear need to address crucial AI environmental, energy, and equity impacts with expertise from the physical and social sciences.

Recommendation 2: CAIDP recommends making environmental impact a focus area for Strategy 3. Specifically, AI Sustainability and AI Development should be incorporated in the "Building ethical AI" and "Designing architectures for ethical AI" subheadings of Strategy 3.

In this regard, a focus on environmental sustainability can promote a well-being approach to human dignity and quality of life. Research has shown that AI-enabled systems require exponentially rising computing power. This increase in computing power requires substantial energy consumption, generating a huge carbon footprint and upending the green effects of digitalization. This problem has raised additional ethical concerns, as well as the well-being of the planet and thus humans.[19] To address this concern, more research should be focused on reducing AI energy consumption, environmental degradation, mineral extraction, and waste. Researchers are developing AI system for training and running certain neural networks that reduce the carbon emissions.[20]

Under this framework, the priority becomes the development of more efficient computing systems that as a goal will not damage the environment,[21] given that human well-being is dependent on ecological well-being. As such, it is of paramount importance to build efficient hardware and AI-based algorithms that require less energy to ensure improved computational efficiency and a smaller carbon footprint. This sets up the critical need to support AI governance frameworks that require the implementation of standards and independent oversight over carbon accounting. Furthermore, this framework would increase the demand for the inclusion of other disciplines like environmental science, geology, oceanography, planetary science, astrobiology, etc.

---

[17] OSTP, *Readout of White House Climate Science Roundtable on Countering "Delayism" and Communicating the Urgency of Climate Action* (Feb. 25, 2022), https://www.whitehouse.gov/ostp/news-updates/2022/02/25/readout-of-white-house-climate-science-roundtable-on-countering-delayism-and-communicating-the-urgency-of-climate-action/

[18] House of Representatives, *National Defense Authorization Act for Fiscal Year 2021* (2020) https://www.congress.gov/116/crpt/hrpt617/CRPT-116hrpt617.pdf#page=1210, 1210

[19] Emily M. Bender, Timnit Gebru, Angelina McMillan-Major, and Shmargaret Shmitchell, *On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?* In Conference on Fairness, Accountability, and Transparency (FAccT '21),(March 3–10, 2021), https://doi.org/10.1145/3442188.3445922

[20] H. Cai, et al., *Once-For-All: Train One Network and Specialize it for Efficient Development*, published as a conference paper at ICLR 2020, https://arxiv.org/abs/1908.09791.

[21] A. Gupta, *The Imperative for Sustainable AI Systems* (Sept. 18, 2021), https://thegradient.pub/sustainable-ai/.

These recommendations address the questions raised in the subheading "What uses of AI might be considered unethical?" In our view, issues of sustainability and the significant environmental impacts of AI systems (such as energy consumption, extraction of rare minerals, and pollution) should be a required dimension of AI development. Inclusion of the language above will help mitigate this concern.

Review/Recommendations for *Strategy 4: Ensure the Safety and Security of AI Systems*

Strategy 4 of the 2019 National AI R&D Strategic Plan updates the 2016 plan by focusing on the rapid growth in AI security and safety and stresses the need for creating robust and trustworthy AI systems.

We call your attention to two fundamental obligations for AI systems set out in the Universal Guidelines for AI, salient in ensuring safety and security: Obligations of Accountability (UGAI-5) and Public Safety (UGAI-8).[22] The obligation to be accountable for AI systems speaks to the ongoing need for assessment of the risks during the design, development, and implementation of systems. Developing standard risk analysis tools for AI systems must include assessment of risks at all levels, and defined context-specific benchmarks to indicate when a system is ready for deployment. It's essential that investments in ethics and social science research address questions responsibility and accountability. The institutions, the designers, and the operators of AI systems retain responsibility for the consequences of AI systems. As the Universal Guidelines for AI further state:

> Safety and security are fundamental concerns of autonomous systems – including autonomous vehicles, weapons, and device control – and risk minimization is a core element of design. Less certain, however, is how to determine and set standards for levels of autonomy across broad applications, and understanding levels of autonomy (and the correlated level of human control) is an interdisciplinary research challenge. The UGAI underscores the obligation of institutions to assess public safety risks that arise from the deployment of AI systems, and implement safety controls.[23]

While we agree that trustworthy AI is "a critical issue that requires Federal Government R&D investments, along with collaborative efforts among government, industry, academia, and civil society,"[24] *independent oversight, international cooperation, clear definitions, and system resilience* are necessary to achieve this goal. The three recommendations provided here are imperative to meet the goals set out in the OSTP AI Strategic Plan; most notably, the promise to

[22] Universal Guidelines for AI.
[23] Ibid
[24] OSTP, *The National AI R&D Strategic Plan: 2019 Update* (June 2019), https://www.nitrd.gov/pubs/National-AI-RD-Strategy-2019.pdf, 24

"build a society where everyone can live with equal dignity and hope and opportunity, as well as equal safety and security."[25]

Problem 1: The need for standardization and independent oversight.

Recommendation 1: New technologies such as AI pose new challenges for privacy, dignity, autonomy, and equality. Metrics for explainability, interpretability, and transparency should be established to protect fundamental rights, human well-being, and to increase public trust.[26] These metrics alongside Privacy Enhancing Technologies would help protect privacy.[27] Additionally, standardized metrics for explainable, interpretable, and transparent systems will increase users' trust in these systems. After standardization, an independent audit –for which its methodologies also require standardization– and the resulting evaluation must confirm the system performs as intended to be certified.

Problem 2: The need for international cooperation.

Recommendation 2: AI standards should be produced and harmonized at the international level (with primary locus being in intergovernmental fora and global standards bodies with strong NGO presence) to ensure common ground around security, safety, and system resilience. This determination should be made by diverse groups with a variety of expertise.[28] The process of developing standards should not be dominated or led by industry groups - the voices and concerns of civil society and affected communities should be effectively represented. Standard-setting activities should protect fundamental rights.[29] CAIDP recommends that these organizations publish annual reports that describe specific steps taken to ensure broad-based participation in the development of technical standards as well as the consideration of fundamental rights.[30]

Problem 3: The need for clear definitions and system resilience.

---

[25] The White House, *A New Chapter for the White House Office of Science and Technology Policy* (Feb. 17, 2022), https://www.whitehouse.gov/ostp/news-updates/2022/02/17/a-new-chapter-for-the-white-house-office-of-science-and-technology-policy/

[26] NIST, *U.S. Leadership in AI: A Plan for Federal Engagement for* Standard (July 2, 2019) (draft for public comment), https://www.nist.gov/system/files/documents/2019/07/02/plan_for_ai_standards_publicreview_2july2019.pdf

[27] The White House, *US and UK to Partner on Prize Challenges to Advance Privacy-Enhancing* Technologies (Dec. 8, 2021), https://www.whitehouse.gov/ostp/news-updates/2021/12/08/us-and-uk-to-partner-on-a-prize-challenges-to-advance-privacy-enhancing-technologies/

[28] *CEN-CENELEC response to the EC white Paper on AI*, Version 2020-06, https://www.cencenelec.eu/media/CEN-CENELEC/Areas%20of%20Work/CEN%20sectors/Digital%20Society/Emerging%20technologies/cen-clc_ai_fg_white-paper-response_final-version_june-2020.pdf

[29] EU-US Trade and Technology Council, *Inaugural Joint Statement* (Sept. 29, 2021), https://ec.europa.eu/commission/presscorner/detail/en/STATEMENT_21_4951

[30] *CAIDP Statement to European Commission on Proposed AI Act* (July 2021). https://www.caidp.org/statements/

Recommendation 3: Standardization, independent audit, system certification, and determination of international common ground depend on three foundational requirements: (1) consensus-based provision of precise definitions and terminology of technical terms (e.g. AI, automation, explainability, interpretability, transparency) for standardization and determination of international common ground;[31] (2) continuation of system updating to include (a) new data resulting from a data-centric strategy for system integrity and thus model improvement as data evolves,[32] and (b) new core AI functionalities resulting from rapid AI advances to maintain system resilience against adverse conditions like cybersecurity risks;[33] and (3) consideration of practices of inclusive design for AI systems.[34]

Review/Recommendations for *Strategy 6: Measure and Evaluate AI Technologies through Standards and Benchmarks*

Strategy 6 establishes that "standards, benchmarks, testbeds, and their adoption by the AI community are essential for guiding and promoting R&D of AI technologies."[35] This section also identifies developing a broad spectrum of AI standards, establishing AI technology benchmarks, increasing the availability of AI testbeds, and engaging the AI community in standards and benchmarks as areas for improvement.

We call your attention to the UGAI principles standards and benchmarks, Assessment and Accountability (UGAI-5) and Accuracy, Reliability, and Validity (UGAI-6).[36] Assessment determines whether an AI system should be established. AI systems should be deployed only after an adequate assessment of its purpose, objectives, risks, and benefits. Imperatively, such assessments must include a review of individual, societal, economic, political, and technological impacts, and a determination can be made that risks have been minimized and will be managed. Individual level risk assessments might include a fundamental rights impact assessment; societal level risk assessments might involve public health or economic impact assessments. If an assessment reveals substantial risks, especially to public safety and cybersecurity, then the project

[31] Krafft, P. M., Meg Young, Michael Katell, Karen Huang, and Ghislain Bugingo. "Defining AI in policy versus practice" In *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society (2020)*, pp. 72-78. 2020.
[32] Gerdes, Anne. "A participatory data-centric approach to AI Ethics by Design." *Applied Artificial Intelligence* (2021): 1-19.
[33] Eigner, Oliver, Sebastian Eresheim, Peter Kieseberg, Lukas Daniel Klausner, Martin Pirker, Torsten Priebe, Simon Tjoa, Fiammetta Marulli, and Francesco Mercaldo. "Towards Resilient Artificial Intelligence: Survey and Research Issues." In *2021 IEEE International Conference on Cyber Security and Resilience (CSR)*, pp. 536-542. IEEE, 2021.
[34] Berkman Klein Center, AI and Inclusive Design, https://aiandinclusion.org
[35] OSTP, *The National AI R&D Strategic Plan: 2019 Update* 33 (June 2019), https://www.nitrd.gov/pubs/National-AI-RD-Strategy-2019.pdf
[36] Universal Guidelines for AI

should not move forward. Accountability for the outcomes and consequences of AI systems lies with the institutions. As the UGAI states:

> Institutions have the obligation to ensure the accuracy, reliability, and validity of AI systems. Benchmarks should be developed against which these standards can be measured. For example, standards should demonstrate that the AI system has been tested for reliability and external validity (i.e., is valid within the population and application context in which it will be deployed). If developed using value-sensitive design, and trained on datasets that are appropriate for a specific user population, AI algorithms and technologies embedded within those contexts will reflect its values, and perform reliably. For example, systems modeled on a dataset of young adults from the United States is likely not to have validity if deployed in a population of aging seniors in Africa because of demographic, cultural, and biological differences.[37]

We encourage adoption of UGAI key principles and make additional proposals. We recommend improving standards and benchmarks by *adding social impact of technology as a separate standard, creating standards that can adapt and keep pace with the speed of technological evolution, and increasing engagement with a diverse community of AI stakeholders*.

Problem 1: Are these benchmarks technical, social impact, or both?

As the report mentions, NIST has planned to develop a broad spectrum of AI standards which include software engineering, performance, metrics, safety, usability, interoperability, security, privacy, traceability, and domain, not including societal impacts.

Recommendation 1: Social impact of technology needs to be added as a separate standard. Most AI-based solutions directly or indirectly affect society disproportionately, therefore, Social Impact Assessment based on model risk framework: to define social impact level for different AI systems based on domain and grading technologies and impact level [38] [39] should be added as a standard. There should be AI risk level-based standards which include high, medium, and low risk AI standards.[40] For creating benchmarks and standards, it is also essential to understand and

[37] Ibid
[38] OECD, *Policy Brief on Social Impact Measurement for Social Enterprises* (2015), https://www.oecd.org/social/PB-SIM-Web_FINAL.pdf
[39] Floridi L, Cowls J, King TC, Taddeo M. *How to Design AI for Social Good: Seven Essential Factors*. Sci Eng Ethics. 2020 Jun;26(3):1771-1796. doi: 10.1007/s11948-020-00213-5. Epub 2020 Apr 3. PMID: 32246245; PMCID: PMC7286860.
[40] European Commission, *Regulatory framework proposal on artificial intelligence*, https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai

determine the levels of risk of AI systems—without understanding the impact of the AI system on human rights, there is little evidence and knowledge for detecting the risk level. We recommend including the Human Rights, Democracy, and Rule of Law Impact Assessment (HUDERIA) proposed by the Council of Europe Ad Hoc Committee on Artificial Intelligence (CAHAI) as a possible benchmark for determining the risk level associated with AI systems.[41]

Problem 2: Lack of standards that can adapt and keep pace with technological evolution.

The strategy states that "standards must be hastened to keep pace with the rapidly evolving capabilities and expanding domains of AI applications."[42] Two key aspects that could be considered when developing standards are developing processes for creating standards faster and ensuring that standards are evaluated and changed in line with the pace of technology.

Recommendation 2: OSTP works with public partners, including the IEEE and ISO, towards the development of processes and procedures that will help to reduce the time to create AI standards. Currently, there are multiple AI standards, benchmarks, and policy frameworks developed by private organizations, NGOs and international organizations such as OECD. We recommend partnering with organizations which have inclusive participation mechanisms to create and change standards faster in line with the pace of technology.

Problem 3: More effort needs to be given towards ensuring engagement with a diverse community of AI stakeholders.

NIST AI standardization activities include engagement with standards organizations and plans to engage with AI communities that have diverse backgrounds made up of users, industry, academia, and government.

Recommendation 3: OSTP must encourage the engagement of diverse communities of AI by: proactively identifying local and marginal communities and indigenous groups; and including diverse stakeholders from domain experts, academic, private, government, and social sectors, including representatives from different sized organizations; to ensure fairness and prevent bias in development of standard and benchmarks. Standard-setting activities must have safeguards in place to balance the industry interests with societal and fundamental rights concerns. If necessary, special funding and membership paths should be available for marginal communities.

[41] Council of Europe AD HOC Committee on Artificial Intelligence (CAHAI), *Human Rights, Democracy and Rule of Law Impact Assessment of AI systems* (Mar. 11, 2021), https://rm.coe.int/cahai-pdg-2021-02-subworkinggroup1-ai-impact-assessment-v1-2769-4229-7/1680a1bd2d

[42] OSTP, *The National AI R&D Strategic Plan: 2019 Update* (2019), https://www.nitrd.gov/pubs/National-AI-RD-Strategy-2019.pdf, 33

Review/Recommendations for *Strategy 7: Better Understand the National AI R&D Workforce Needs*

As the speed and scale of AI technologies expand both domestically and globally, much care must be given to ensure the proper recruitment and retention of AI researchers and practitioners to ensure a viable pool of talent for tomorrow's intellectual demands. Strategy 7 states, "It is critical to maintain a robust academic research ecosystem in AI that, in collaboration with industry R&D, can continue to deliver tremendous dividends by advancing national health, prosperity, and welfare, and securing the national defense."[43] As CAIDP stated in our June 2021 statement to U.S. financial agencies, we believe AI systems should be designed in a way that respects the rule of law, human rights, democratic values, and diversity, and they should include appropriate safeguards – for example, enabling human intervention where necessary – to ensure a fair and just society."[44] Further, the UGAI indicated several key principles that are critical considerations for workplace need strategies including and relevant here, including the right to transparency (UGAI-1) and human determination (UGAI-2), and obligations of fairness (UGAI-4).[45] Using that perspective, we sense that Strategy 7 could be enhanced by *ensuring interdisciplinary education, reforming fellowships and educational programs, and ensuring the inclusion of diverse voices among educators and students.*

Also, CAIDP recommends following best practices for AI procurement decisions and adopting talent management strategies that are needed to support oversight mechanisms to ensure the protection of human rights and wellbeing of citizens and accountability, transparency and fairness of AI systems. The responsible use of emerging technologies can be supported by public policy improvements to government procurement processes. Through reform of the guidelines by which the government exercises its purchasing power, governments can fulfill their leadership role in steering technology policy. As explained in the World Economic Forum's Unlocking Public Sector AI through Government Procurement initiative AI Procurement in a Box report:

> Government procurement officials cannot be expected to have the most up-to-date knowledge in every highly specialized field. To safeguard the responsible future use of AI technologies, a multistakeholder effort with cross-sector participation and interdisciplinary expertise is required to create authoritative guidelines. The procedural norms are even more urgent now. What information should be recorded

[43] OSTP, The National AI R&D Strategic Plan: 2019 Update (Jun 2019), https://www.nitrd.gov/pubs/National-AI-RD-Strategy-2019.pdf, 37
[44] CAIDP, *CAIDP Statement to U.S. Financial Agencies on Use of AI by Financial Institutions* (June 30, 2021), https://www.caidp.org/statements/
[45] The Public Voice, *Universal Guidelines for Artificial Intelligence: Endorsement* (2018), https://thepublicvoice.org/AI-universal-guidelines/endorsement/

and how explanations need to be documented is what lays the foundation for fairness and impartiality in the administrative process. To preserve due process and predictability, a coalition can help ensure that the right questions are asked.[46]

Problem 1: Multidisciplinary teams composed of stove-piped experts are insufficient.

Recommendation 1: OSTP must work to ensure collaborators have both deep subject matter expertise and interdisciplinary knowledge to readily build more effective connections. Computing is an interdisciplinary field that requires innovative interdisciplinary education. For current and future workforce talent to successfully manage the interdisciplinary nature of computing, education must not just expose learners to different disciplines' knowledge but integrate the disciplines within the learning process. As the demand for multidisciplinary teams grows, such teams can no longer rely exclusively on single subject matter experts.

Problem 2: Increasing educational programs, fellowships, and activities dedicated. exclusively to the quantitative fields of ML is insufficient.

Recommendation 2: Learning from K-12 to postgraduate education needs reforming and resources to sustain. To achieve success with recommendation (1), curricula will need to balance in-depth subject matter with expansive related subject matter coverage, as well as traditional assessments with creative and experiential practice-based learning experiences.

Problem 3: Insufficient care given to ensure effective representation by underrepresented groups.

Recommendation 3: From the instructors, faculty, researchers, designers, developers, project managers and directors to the learners themselves, diverse voices must not only be included at the planning stage of research projects but integrated within the various phases of the AI lifecycle of design, development, and deployment. This necessarily implies respectfully accounting for the different experiences of all through, for example, data acquisition and user experience feedback. Echoing the OSTP recommendation for STEM equity,[47] we support diverse representation at all levels and across all AI disciplines.

[46] World Economic Forum, *AI Procurement in a Box: Project overview* (June 2020), https://www3.weforum.org/docs/WEF_AI_Procurement_in_a_Box_Project_Overview_2020.pdf

[47]The White House, A New Chapter for the White House Office of Science and Technology Policy (Feb. 17, 2022), https://www.whitehouse.gov/ostp/news-updates/2022/02/17/a-new-chapter-for-the-white-house-office-of-science-and-technology-policy/

CAIDP encourages OSTP, the Select Committee, and NAIIO, in consultation with the NSTC Subcommittee on Machine Learning and AI and the NITRD AI R&D Interagency Working Group, to incorporate the recommendations above in the AI Strategic Plan.

Finally, we call attention to the UNESCO Recommendation on the Ethics of Artificial Intelligence.[48] CAIDP contributed to the development of the Recommendation,[49] and expressed "strong support for adoption."[50] The United States is not yet a signatory to the UNESCO AI Recommendation. We strongly encourage OSTP to advance U.S. global leadership in AI ethics and urge the Administration to endorse the UNESCO Recommendation.

Thank you for your consideration of our views.[51] We welcome the opportunity to discuss further.

Marc Rotenberg
CAIDP President

Merve Hickok
Research Director

Karine Caunes
Global Program Director

Jason Johnson
CAIDP Research Fellow

Tamra Moore
CAIDP Research Fellow

Somaieh Nikpoor
CAIDP Research Fellow

[48] UNESCO, Recommendation on the Ethics of Artificial Intelligence (2021), https://en.unesco.org/artificial-intelligence/ethics
[49] CAIDP, CAIDP Update 1.4– "UNESCO Pursues Humanistic Approach for AI" (Aug. 10, 2020), https://www.caidp.org/app/download/8292333863/CAIDP-Update-1.4.pdfs
[50] CAIDP, CAIDP Update 2.41– "UNESCO Pursues Humanistic Approach for AI" (Nov 25, 2021), https://www.caidp.org/
[51] CAIDP acknowledges the significant contributions to this statement of the 2022 CAIDP Research Group, North America Team, including Sharvari Dhote, Kathy Kim, Monica Lopez, Nidhi Sinha, and Narain Yucel.