PARTNERSHIP ON AI

The Partnership on AI Response to the White House Office of Science and Technology Policy and National Science Foundation Request for Information: National Artificial Intelligence Research Resource

October 1, 2021

The Partnership on AI (PAI) is a non-profit partnership of academic, civil society, industry, and media organizations creating solutions so that AI advances positive outcomes for people and society. The Partnership on AI studies and formulates sociotechnical approaches to the responsible development of AI technologies to advance the public's understanding of AI and to serve as an open platform for discussion and engagement about AI and its influences on people and society. Today, PAI convenes nearly 100 partner organizations from around the world to be a uniting force for the responsible development and fielding of AI technologies. Partnership on AI staff composed this response based on some of PAI's recent work, much of it inspired and informed collectively by consulting over several years with our international group of multi-stakeholder Partner organizations. The information in this document is provided by PAI and is not intended to reflect the view of any particular Partner organization of PAI.

PAI develops tools, recommendations, and other resources by inviting diverse voices from across the AI community and beyond to share insights that can be synthesized into actionable guidance. We then work to promote adoption in practice, inform public policy, and advance public understanding. Through dialogue, research, and education, PAI is addressing some of the most important and difficult questions concerning the future of AI. Currently, PAI and its Partners work toward the responsible development of AI technologies in four Program areas: (1) AI & Media Integrity, (2) AI, Labor and the Economy, (3) Fairness, Transparency and Accountability, and (4) Safety Critical AI.

PAI is pleased to submit this response to the RFI for the National AI Research Resource on the specific questions referenced below. We will highlight PAI's work and publicly available resources on 1) Demographic Data and Algorithmic Bias 2) Responsible Publication Norms 3) Transparency through Documentation 4) Diversity, Equity, and Inclusion in AI and 5) Inclusion and Access to AI R&D through Multi-stakeholder Partnerships.

Question 1D and 1E:

> "What options should the Task Force consider" on (1D) "including provision of curated data sets" and (1E) "[a]n assessment of, and recommended solutions to, barriers to the dissemination and use of high-quality government data sets as part of the National Artificial Intelligence Research Resource"

Demographic Data and Algorithmic Bias

The provision of curated data sets presents a challenge, particularly in collecting and using demographic data in service of detecting algorithmic bias with its many legal and ethical implications (*Demographic Data Convening*, PAI January 2020). A lack of clarity as to the acceptable uses for demographic data is cited frequently by PAI Partners as a barrier to addressing algorithmic bias in practice. In order for data sets to be curated with an eye towards positively impacting historically disenfranchised groups, there must be awareness of what those groups are. This requires datasets to be disaggregated by race, gender, etc. in order to assess discrimination and inequality.

However, the inclusion of sensitive data is a fraught practice for datasets, especially with regard to categories that could be used for discrimination, such as sexuality, political affiliation, and immigration status (*Don't Overlook the Role of Demographic Data*, PAI Blog, April 2020). Ensuring categories are representative of the populations in question requires an understanding of the constantly shifting nature of identity and ongoing engagement with marginalized populations. For example, the inclusion of the Asian Americans and Pacific Islanders (AAPI) category on US Census forms was a large bureaucratic struggle. While, at the time, it enabled representation of the AAPI community, over time many individuals categorized as AAPI have come to feel misrepresented. Disaggregated datasets should be managed by independent third parties, specifically ones that represent the at-risk groups reflected in the data (*Knowing the Risks: A Necessary Step to Using Demographic Data for Algorithmic Fairness*, PAI Blog, September 2021).

The National AI Research Resource should encourage stakeholders developing and using demographic datasets to:

- Curate Datasets with support of community-based organizations that have trust and experience with the groups and communities in question.

- Ensure the need for demographic data to assess anti-discrimination, fairness, and inequality does not infringe on privacy rights or increase the undue surveillance of protected classes, vulnerable populations, or marginalized groups. (*"What We Can't Measure, We Can't Understand": Challenges to Demographic Data Procurement in the Pursuit of Fairness*, January 2021).

- Engage with the tensions around demographic data usage in AI and align with emerging work on the importance of equitable data as seen, for example, in the Executive Order On Advancing Racial Equality And Support for Underserved Communities Through The Federal Government (*Section 9*).

Question 3:

> "How can the NAIRR and its components reinforce principles of ethical and responsible research and development of AI, such as those concerning issues of racial and gender equity, fairness, bias, civil rights, transparency, and accountability?"

Responsible Publication Norms

A key development in reinforcing ethical and responsible research concerns establishing norms around the responsible publication of AI research. Over the past two years, PAI has conducted research and hosted multistakeholder convenings with the AI community to explore how advances in AI research can be disseminated in a responsible manner given their potential for misuse. Other research communities have established norms and procedures for publishing high-risk research such as bioethics and cybersecurity. However, the AI community has had less time to develop similar practices. The use or misuse of advancements in AI systems can lead to potential accidents, unintended consequences, inappropriate applications, and malicious uses but also contribute to potential systemic harms. AI can be used in biometric or facial recognition technology deployed in surveillance applications or bias in datasets and algorithms that are amplified when misapplied in different contexts like criminal justice or hiring.

In May 2021, PAI published a white paper titled, *Managing the Risks of AI Research: Six Recommendations for Responsible Publication*, containing recommendations for the AI research community on responsible research and publication practices for anticipating downstream consequences. This report synthesizes insights gathered from our research and convenings, with over 30 individuals from PAI's partner organizations and other stakeholders contributing to the paper. Shortly after its release, *Nature Machine Intelligence* published an editorial endorsing the white paper's recommendations.

The key takeaway from PAI's paper and surrounding engagement on this issue is that the AI community must build a responsible research culture that anticipates potential downstream consequences of AI research and mitigates risks. While our recommendations have primarily focused on interventions at the point of publication, there are other important, earlier stages in the research pipeline including conception, funding, access to computing resources/infrastructure, where actors can effect change to contribute to a responsible research culture.

The National AI Research Resource should consider the following three recommendations to reinforce responsible publication norms and foster a responsible research culture that includes a consideration of ethical and societal consequences by AI researchers.

- Disclose in publications

AI researchers must be encouraged to report the level of contribution their paper is making -- is it an incremental improvement or an entirely new technique -- and disclose the motivation behind the research so those evaluating can better understand its potential impact and develop mitigating strategies. AI researchers must also be encouraged to report the computation used in research projects to better understand considerations around reproducibility, and downstream consequences such as environmental impact.

- Normalize discussion about the downstream consequences of research

  Researchers should be invited to reflect on the ethical and societal consequences of their research relative to the level of advances to the field provided by their work. For incremental advances, a short statement citing work that discusses the consequences of similar research in more detail may be sufficient. For more significant advances, a more substantial discussion is warranted.The goal of this exercise is to not only inform those evaluating the research including publication venues about potential negative impacts but to also encourage researchers to thoughtfully consider all the ways their research could be harmful, examine potential second-or third-order effects, and consider mitigation.

- Review potential downstream consequences earlier in the research pipeline

  Research teams should be encouraged to build opportunities to consider societal impacts earlier during the research process including when initially formulating research ideas. In cases where the research could have high-stakes applications, or is likely to be a significant technical advancement, a more thorough review of societal impacts will be necessary. This exercise can take many different shapes including holding discussions with researchers, hiring cross-disciplinary experts, getting inputs from underrepresented inputs etc. While these processes lack the rigor of formal mechanisms like university Institutional Review Boards, they can still go a long way towards mitigating negative impacts and can be less burdensome for researchers involved.

Transparency Through Documentation

The Partnership on AI is working towards establishing new norms on transparency through documentation with our project on Annotation and Benchmarking on Understanding and Transparency of Machine learning Lifecycles (ABOUT ML). The project identifies best practices for documenting and characterizing key components and phases throughout the ML system lifecycle from design to deployment, including annotations of data, algorithms, performance, and maintenance requirements (*How ABOUT ML Taps Collective Wisdom*, PAI Blog September 2019). A goal of documentation for system deployment is to write down the socially salient

aspects of performance, including fairness, robustness, explicability, and other topics. Relevant and difficult-to-answer questions include what tests, monitoring, and evaluation have been done, and how does monitoring relate to social outcomes (*Operationalizing AI Ethics Through Documentation: ABOUT ML in 2021 and Beyond*, PAI Blog, April 2021).

A core tenet of the project holds that documentation is important to consider as both an institutional process and an artifact because many teams and individuals have to incorporate completing and updating such an artifact into their work in order for it to be useful. ABOUT ML's goal is not only to recommend what information should go into documentation for all ML systems but also to recommend how organizations can effectively reshape their processes to enable the reliable completion and maintenance of documentation in an ongoing manner (*Section 1.1.2* ABOUT ML Reference Document).

The Partnership on AI researchers and Partners, along with other stakeholders and public commenters, have identified a need to create documentation for internal accountability. The need for accountability motivates organizations to invest in and build the internal processes and infrastructure to implement and scale the creation of documentation artifacts (*Section 2.2.1* ABOUT ML Reference Document).

A key component of this project is its Steering Committee, comprised of around 30 experts, researchers and practitioners recruited from a diverse set of PAI Partner organizations. The Steering Committee guides the process of updating ABOUT ML resources based on the public comments submitted and new developments in research and practice.

The convenings and conversations inspired by the ABOUT ML project have led experts in the space to converge on the need to modify the sets of questions and information to be shared externally. This need is based on the constraints of what organizations are willing to share and what information external stakeholders require to consider the ML system to be sufficiently transparent. There should be a broad and public conversation between organizations that build ML systems and key external stakeholders — including civil society organizations, policymakers, end users, and non-users impacted by ML systems — to determine what information would be necessary in documentation for external accountability.

Diversity, Equity, and Inclusion in AI

The lack of diversity in the field of AI has been well-documented. Simply put, diverse teams result in better outcomes that can issue-spot and control for many of the challenges AI researchers and developers face. As an industry, AI struggles to both recruit and retain team members from diverse backgrounds, particularly women and minoritized communities. Despite widespread awareness of AI's diversity gap, the crisis continues, and in spite of significant investments to address the issue, there remains a lack of clarity about which

initiatives work best. By investigating the pervasive challenges in ethnic, gender, and cultural diversity in the field of artificial intelligence, PAI's Diversity, Equity, and Inclusion (DEI) Workstream seeks to turn collected insights into actionable resources for those striving to make a more inclusive environment for people working in AI.

We view this work as paramount to the advancement of responsible AI. If we do not work to sufficiently address diversity and inclusion on the teams developing the technology, we risk compounding existing economic and social disparities experienced by women and minoritized individuals and communities. PAI has benefited from a dedicated research fellowship to advance diversity and inclusion in AI. The goal of this work is to learn both from the lived experiences of women and minoritized individuals in the field of AI and from those involved in DEI initiatives at organizations to share knowledge across the field about what the key challenges are and what solutions work.

The National AI Research Resource should consider the following resources to guide the development of transparency through documentation and encourage its community of stakeholders to take concrete steps and commitments to reinforce core ethical principles such as transparency and diversity, equity, and inclusion in the field of AI.

- ABOUT ML Reference Document

  The goal of this Reference Document is to synthesize insights and recommendations from the existing body of literature to begin a public multistakeholder conversation about how to improve ML transparency. By providing a guide for practitioners to start taking transparency seriously, this document serves as a first step. A foundational resource, this living document includes an extensive literature review, suggested documentation sections for datasets, and surfaces current challenges of implementing documentation.

- ABOUT ML Process Guide

  Because documentation is both a process and a set of artifacts, transparency and documentation need to be an explicit part of the discussion at each step of the workflow. The ABOUT ML Process Guide provides suggested documentation questions and considerations for each phase of the ML system lifecycle — from design and setup to observation and maintenance — compiled from the ABOUT ML Reference Document and academic literature.

- Beyond the Pipeline: Addressing Attrition as a Barrier to Diversity in AI

  A forthcoming study conducted in-depth interviews with managers, people working in DEI, and workers who identified as belonging to historically excluded identities and analyzed themes from those interviews to get at the heart of the AI field's attrition

problem. The paper distills these learnings into a set of insights and recommendations that those working in AI organizations can take to improve upon their current DEI practices, beyond implicit bias and diversity training.

Questions 4-6:

"What building blocks already exist for the NAIRR, in terms of government, academic, or private-sector activities, resources, and services? 5. What role should public-private partnerships play in the NAIRR? What exemplars could be used as a model? 6. Where do you see limitations in the ability of the NAIRR to democratize access to AI R&D? And how could these limitations be overcome?"

Inclusion and Access to AI R&D through Multistakeholder Partnerships

Working with its multistakeholder community of partners in academia, civil society, industry and media, PAI is committed to increasing access, inclusion and participation in AI R&D.

One of PAI's Partners, the Tech Policy Lab at the University of Washington, has extensive expertise in applying value-sensitive design approaches to technology policy. In 2019, PAI worked with the Tech Policy Lab to implement their Diverse Voices methodology within PAI's ABOUT ML project. The aim was to solicit views and feedback from communities who are often the least likely to be consulted in the formation of machine learning system documentation practices that may impact them. The insights garnered through this consultation informed the inclusion of a glossary in the ABOUT ML resource library as well as the design and structure of the materials to promote clarity and navigational guidance to readers from diverse backgrounds.

Building from the lessons learned from the Diverse Voices team and the work of other responsible AI advocates, PAI launched the Methods for Inclusion research project that aims to enable AI researchers and developers to more effectively and ethically engage with a broad base of constituents and stakeholders in the development of their AI/ML projects. This work seeks to meaningfully include impacted communities in order to enable AI/ML developers to provide an array of products and services that can better meet the needs of diverse populations around the world, without further deepening existing social inequalities or generating harm. A forthcoming publication will identify a broad range of methodologies and practices that can be applied at different stages of the AI development process, drawing on the large body of scholarship that has grappled with the question of how to create inclusive channels of participation in other domains.

Drawing on these experiences at PAI, OSTP and NSF are encouraged to incorporate a focus on inclusion, participatory design, and democratizing access throughout all aspects of the development of the resource. Building a diverse community of stakeholders across sectors to

engage and inform the development and deployment of the resource over time will be central to its success.

Thank you for this opportunity to provide information about the Partnership on AI's work that could help guide the development of the National AI Research Resource. One of the benefits of multi-stakeholder organizations such as PAI is the opportunity to convene and connect diverse perspectives from across sectors, disciplines, geographies, and lived experiences - a critical component to understanding and developing a national resource. The Partnership on AI is happy to provide more details or additional information about the research, workshops, convenings, and other activities we conduct as we continue to develop resources and tools to prevent harms and promote the development of AI that benefits people and society. Please contact Rebecca Finlay, Acting Executive Director rebecca@partnershiponai.org and Mark Latonero, Senior Policy Advisor mark@partnershiponai.org.