Federal Register Notice 86 FR 46278, https://www.federalregister.gov/documents/2021/08/18/2021-17737/request-for-information-rfi-on-an-implementation-plan-for-a-national-artificial-intelligence, October 1, 2021.

Request for Information (RFI) on an Implementation Plan for a National Artificial Intelligence Research Resource: Responses

CENTER FOR DEMOCRACY & TECHNOLOGY

October 1, 2021

Office of Science and Technology Policy and the National Science Foundation
Attn: Wendy Wigen, NCO,
2415 Eisenhower Avenue,
Alexandria, VA 22314, USA

Re: Request for Information on an Implementation Plan for a National Artificial Intelligence Research Resource

(submitted by email)

In response to the National Science Foundation and the Science and Technology Policy Office's request for information on an implementation plan for an National Artificial Intelligence Research Resource, the Center for Democracy & Technology would like to offer the following comments:

1. What options should the Task Force consider for any of the roadmap elements A through I above, and why?

B.ii: A governance structure for the Research Resource, including oversight and decision-making authorities.

In designing a governance structure for the Research Resource, the NAIRR Task Force should first consider the types of decisions that decision-making authorities will be called upon to make, and establish a governance structure that will involve individuals with sufficient subject matter expertise in these topics and that will ensure transparency, accountability, and fairness. Among other issues, decision-makers will have to determine:

1. Who should be granted access to the Research Resource, what criteria AI researchers and students must satisfy to gain access to the Research Resource, and how researchers and students will be vetted for compliance with that criteria.
2. What limits will be placed on researcher and students' use of the Research Resource once they are granted access. These limits should be designed to protect individual privacy and prevent unethical uses of the Research Resource.

3. How the limits on use of the Research Resource will be enforced, including how violations will be reported or otherwise detected, investigated, and found to be substantiated or unsubstantiated, and the penalties for substantiated violations of the limits.
4. How the NAIRR can empower researchers and students to make the best use of the Research Resource and ethically use the Research Resource, through technical support, training, and other resources.
5. What data should be included in the Research Resource, what criteria data must satisfy to gain access to the Research Resource—including criteria concerning lack of bias, privacy, and intellectual property rights—and how data will be vetted for compliance with that criteria.
6. What data included in the Research Resource must be kept confidential and accessible only to vetted AI researchers and students, and what data can be made publicly available.

These considerations should guide the development of a diverse and sufficiently empowered governance structure for the NAIRR. That structure should be transparent in terms of who will make decisions, the process for doing so, and how those decisions will be communicated to the public and/or involved parties. The structure should be fair: for example, a researcher that is denied access to the resource should have a means to appeal or contest that denial. And the governance structure should make clear who is accountable for the various decisions that will need to be made.

D. Capabilities required to create and maintain a shared computing infrastructure to facilitate access to advanced computing resources for researchers across the country, including provision of curated data sets, compute resources, educational tools and services, a user-interface portal, secure access control, resident expertise, and scalability of such infrastructure.

On the provision of curated data sets, the NAIRR Task Force should provision and disseminate curated data sets in ways that allow researchers to reproduce and build upon existing research. Not being able to reproduce research results, or what scientists call the "replication crisis", is a significant challenge for science overall, and one of its causes is insufficient access to the underlying data researchers use. One approach to address this is through the use of open data — data that is freely available to use and redistribute with few or no restrictions. Requirements for certain government agencies to open certain data sets already exist at the federal,[1] state, and local levels. However, given that the NAIRR will most likely include collaborations across different research groups from the public and private sectors, an

[1] DCAT-US Schema v1.1 (Project Open Data Metadata Schema), https://www.data.gov/

open data approach is important. In practical terms this means providing data in open formats (i.e., not dependent on particular software), with little or no restrictions (e.g., registration requirements). It also means publishing metadata,[2] creating public APIs where feasible, and being clear about how the data is licensed (ideally licensing similar to that of Creative Commons, or when possible, placing the data in the public domain).

While open data can help address the reproducibility challenge, access alone is not enough.[3] The NAIRR should also develop guidelines on how to create documentation to accompany each dataset that will include how the dataset was constructed (e.g., labelling, calculation of new variables, etc.) and how it was used in the research. These are not common practices for researchers, and so the NAIRR Task Force should also explore ways to incentivize and guide researchers through these steps.

The availability of open data can also promote greater equity in the access and use of the NAIRR, particularly among researchers with limited resources or those outside of research networks of scholars willing to share their data. However, open data practices alone cannot address the equity problem, and the NAIRR will have to be intentional in understanding and addressing the data needs of researchers in non-traditional research organizations (e.g., journalists, civil society) and academic groups that have traditionally collaborated less directly with entities such as the NAIRR itself.

G. An assessment of privacy and civil rights and civil liberties requirements associated with the National Artificial Intelligence Research Resource and its research.

CDT recommends that the NAIRR Task Force conduct a Human Rights Impact Assessment (HRIA) of the Research Resource and build future HRIAs into the governance structure of the Research Resource. HRIAs are a method of reviewing and monitoring particular projects or activities to offer "guidance and practical tools" using a human rights-based approach.[4] While businesses may undertake HRIA to assess consistency of their activities with the United Nations Guiding Principles on Business and Human Rights, governments can and should also engage in HRIAs to analyze the consistency of their activities

[2] See for example DCAT-US Schema v1.1 (Project Open Data Metadata Schema), https://resources.data.gov/resources/dcat-us/

[3] See for example Hardwicke, T. E., Mathur, M. B., MacDonald, K., Nilsonne, G., Banks, G. C., Kidwell, M. C., ... & Frank, M. C. (2018). Data availability, reusability, and analytic reproducibility: Evaluating the impact of a mandatory open data policy at the journal Cognition. *Royal Society open science*, *5*(8), 180448.

[4] The Danish Institute for Human Rights, *Welcome and Introduction, Human Rights Impact Assessment Guidance and Toolbox* (2020) at 4, https://www.humanrights.dk/sites/humanrights.dk/files/media/document/DIHR%20HRIA%20Toolbox_Welcome_and_Introduction_ENG_2020.pdf.

with international human rights principles. The NAIRR Task Force should engage in a HRIA of the proposed Research Resource and incorporate learnings from that assessment into its design of the governance structure for the Research Resource. In addition, CDT recommends building ongoing and periodic HRIAs of the Research Resource into its governance structure, with a particular emphasis on analysis of whether and how the Research Resource is impacting individual privacy or enabling biased, discriminatory, inequitable, or unethical research or application of AI.

In designing and implementing these HRIAs, the NAIRR Task Force may wish to consider the Human Rights Impact Assessment Guidance and Toolbox from the Danish Institute for Human Rights.[5] In addition, the NAIRR Task Force could look to HRIAs conducted of other data-sharing endeavors, such as the HRIA of the Global Internet Taskforce for Combatting Terrorism, through which member companies make use of a hash-sharing database and URL sharing to identify and screen user-generated terrorist content.[6]

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?

We commend the NAIRR Task Force for considering issues of equity, fairness, bias, civil rights, transparency, and accountability during the roadmapping and planning process for the NAIRR. These are challenging and complex issues that should be considered early on in any development process for AI-based frameworks and resources and, as noted above, should be addressed as part of the NAIRR's governance structure.

CDT would like to offer a few suggestions aimed at helping to ensure that issues of equity and accountability are integrated into the NAIRR infrastructure.

1. The NAIRR Task Force should start with its own research into privacy and equity harms from prevalent AI systems. This includes regularly consulting with a range of individuals and

[5] See Danish Institute for Human Rights, *Human Rights Impact Assessment Guidance and Toolbox,* above.

[6] Global Internet Forum to Counter Terrorism, *Human Rights Assessment*, https://gifct.org/wp-content/uploads/2021/07/BSR_GIFCT_HRIA.pdf.

organizations with personal experience and technical and policy expertise, from affected consumers and civil society groups to academics and AI researchers. Consumers and civil society groups can speak to the direct impacts of AI systems' discriminatory outcomes, offering valuable insight into the information asymmetry involved in AI systems, the burden that being researched can pose to consumers, and the benefit that direct access to resources like the NAIRR would provide in disputing unfair AI-driven outcomes. Academics and AI researchers can speak to commonly researched types of AI systems, underlying AI issues and goals driving the research, research methodologies, and the ultimate effectiveness of selected data sets and methodologies. This discourse can shape the Task Force's ability to anticipate how the NAIRR's substance and degree of accessibility might contribute to reducing or preventing inequities and privacy violations. The Task Force should build these considerations into its privacy and civil rights and civil liberties requirements, as well as a process to modify the requirements as soon as they are found to be inadequate.

2. The NAIRR infrastructure should include dedicated resources for projects that enable factfinding and research about equity and bias in existing AI systems and methods. There are numerous examples of biased AI-based systems causing harm to people,[7] and the NAIRR should offer priority to researchers and projects that seek to uncover, understand, and, where possible, correct these equity issues.

3. There are many cases where AI systems produce biased outcomes that stem, at least in part, from biased training datasets.[8] The NAIRR should audit any data sets it provides for bias. The NAIRR should seek to build and provide unbiased datasets where possible, while understanding that mitigating bias in datasets is complex and, if done incorrectly, may itself introduce bias concerns. For instance, consider a dataset of test scores, where students of color have disproportionately worse scores due to bias in the test design. It may seem that a solution

[7] Meredith Broussard, When Algorithms Give Real Students Imaginary Grades, N.Y. Times (Sept. 8, 2020), https://www.nytimes.com/2020/09/08/opinion/international-baccalaureate-algorithm-grades.html; A-levels and GCSEs: How Did the Exam Algorithm Work?, BBC News (Aug. 20, 2020), https://www.bbc.com/news/explainers-53807730; Rebecca Koenig, Can Algorithms Select Students "Most Likely to Succeed"?, Slate (July 10, 2020), https://slate.com/technology/2020/07/college-admissions-algorithms-applications.html; Shirin Ghaffary, The Algorithms that Detect Hate Speech Online Are Biased Against Black People, Vox (Aug. 15, 2019), https://www.vox.com/recode/2019/8/15/20806384/social-media-hate-speech-bias-black-african-american-facebook-twitter; Karen Hao, The Coming War on The Hidden Algorithms That Trap People in Poverty, MIT Tech. Review (Dec. 4, 2020), https://www.technologyreview.com/2020/12/04/1013068/algorithms-create-a-poverty-trap-lawyers-fight-back/.

[8] Jeff Larson et al., How We Analyzed the COMPAS Recidivism Algorithm, (May 23, 2016), https://www.propublica.org/article/how-we-analyzed-the-compas-recidivism-algorithm.

would be to remove low scores from students of color to avoid entrenching the original biased test design in the resulting score dataset. However, that dataset would now have more limited information about students of color than about white students, which is itself a form of bias. Consequently, any attempt to correct bias in datasets must be an iterative and rigorous process that avoids introducing new forms of bias.

In many cases, collecting unbiased data may not be possible: bias in datasets typically stems from bias in the environment where the data is collected, meaning that existing and pervasive societal biases will inevitably be reflected in real-world data. In such cases, the NAIRR should identify the purpose for developing each data set and using the types of data involved, the rationale for relying on particular sources of data, and the issues on which each data set provides relevant insights. If these elements are not readily articulable, this could signal privacy and equity risks that need to be more deeply explored, in which case the NAIRR either should not include the dataset in question or should provide clear information about the bias(es) in the dataset so users of the dataset are aware of and can appropriately account for such bias.[9]

The NAIRR Task Force should take advantage of existing efforts within the federal government to understand and limit bias and inequity in AI-based systems. For example, the NAIRR should work in concert with efforts by the National Institute of Standards and Technology (NIST) to improve explainability in AI[10] and understand and mitigate bias in AI.[11]

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?

The NAIRR's commitment to offering training to bring more people into the AI R&D community will be an important element of democratizing the field, but in order to be effective, that training will need to take into account the existing inequities in the AI community, and the tech sector more broadly. For instance, trainings or resources targeted at university computer science students to encourage them to

[9] Timnit Gebru et al., Datasheets for Datasets, (Mar. 19, 2020), https://arxiv.org/abs/1803.09010.

[10] P J. Phillips et al., Four Principles of Explainable Artificial Intelligence (Draft), (Aug. 18, 2020), https://www.nist.gov/publications/four-principles-explainable-artificial-intelligence-draft.

[11] NIST Proposes Approach for Reducing Risk of Bias in Artificial Intelligence, (June 22, 2021), https://www.nist.gov/news-events/news/2021/06/nist-proposes-approach-reducing-risk-bias-artificial-intelligence.

specialize in AI will not address the lack of diversity amongst computer science students at the university level, and thus will be limited as far as democratizing and diversifying the field. The NAIRR can help to overcome these limitations by acknowledging and working to combat existing challenges. For example, this could mean offering trainings and resources designed to pull students and the general public into AI from a broader variety of fields than just computer science, including those with a higher concentration of populations that are currently underrepresented in the AI field, such as nursing or teaching.

The Task Force should also consider the need for a clear process allowing equitable access to the NAIRR and to resources within the NAIRR that help contextualize and explain data sets for consumers' understanding. Providing avenues for engagement and participation from civil society and advocacy groups that work on behalf of marginalized communities (rather than just academic or industry AI researchers) is critical to avoiding discriminatory outcomes and enabling those communities to self advocate.[12] This is particularly important given that marginalized people experience greater barriers to entry and advancement within the AI research sector.[13] Without their own access to resources like the NAIRR, marginalized communities would have to depend on qualified researchers who may not prioritize the AI-related issues that concern and harm consumers most.[14] In certain situations, researchers' analysis may also not be easily available to the public due to other interests involved in the research endeavor.[15] Overall, the Task Force must actively examine its processes for building, curating, and providing access to AI resources to determine whether their use remains appropriate and continues to serve the public interest.

Overall, CDT commends the Office of Science and Technology Policy and National Science Foundation for undertaking the development of a resource as complex and potentially valuable as the NAIRR, and for considering important questions of equity, privacy, and democratic access from the beginning.

---

[12] Kathryn L.S. Pettit et al., Urban Institute, Putting Open Data to Work for Communities (2014), available at https://www.urban.org/sites/default/files/publication/22666/413153-Putting-Open-Data-to-Work-for-Communities.PDF.

[13] Ebony O. McGee, *Let's Remake Racially Unsafe STEM Educational Spaces*, Higher Education Today (Feb. 11, 2021), https://www.higheredtoday.org/2021/02/11/lets-remake-racially-unsafe-stem-educational-spaces/.

[14] Florence Ashley, Accounting for Research Fatigue in Research Ethics, 35 Bioethics 270, 272 (2021), available at https://www.florenceashley.com/uploads/1/2/4/4/124439164/ashley_accounting_for_research_fatigue_in_research_ethics.pdf.

[15] Brian Resnick and Julia Belluz, *The War to Free Science*, Vox (July 10, 2019), https://www.vox.com/the-highlight/2019/6/3/18271538/open-access-elsevier-california-sci-hub-academic-paywalls.

Sincerely,

Hannah Quay-de la Vallee, *Senior Technologist, CDT*

Gabriel Nicholas
Research Fellow, CDT
Ridhi Shetty
Policy Counsel, Privacy & Data Project, CDT
Dhanaraj Thakur
Research Director, CDT
Caitlin Vogus
Deputy Director, Free Expression Project, CDT