

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

DISCLAIMER: Please note that the RFI public responses received and posted do not represent the views and/or opinions of the U.S. Government nor those of the National AI Research Resource Task Force., and/or any other Federal agencies and/or government entities. We bear no responsibility for the accuracy, legality, or content of all external links included in this document.

Response to the Request for Information on a National Artificial Intelligence Research Resource - RFI 86 FR 39081

From: The Alexandria Archive Institute (Open Context) prepared by Paulina F. Przystupa, with Eric C. Kansa and Sarah Witcher Kansa

Submitted on: September 23, 2021

1-a. Options for the goals for the establishment and sustainment of a NAIRR and metrics for success should include a heavy emphasis in education, demonstrable benefit to communities whose data are used to build the curated data sets used by the resource, an incorporation of inclusive governance for the resource, and improvement to infrastructure that allows easier access to communities currently underserved by AI research.

These should be included because without an infrastructure of education that leads into the use of such a resource it will probably be underused and not serve all communities in an equitable way. This includes education in the resource itself but also in the creation of the resource that allows multiple publics to understand its overall importance so that they can better influence its use and use it for themselves.

A demonstrable benefit to the communities whose data are used is important because scientific research has a history of exploitation that a resource like this has the potential to reinforce. These communities should also help define “metrics for success” as it is their information that is used to create these models.

An incorporation of inclusive governance for the resource is important because understanding what data and models may be appropriate will vary cross-culturally. This is especially important with respect to communities that have experienced colonialism and other forms of inequality. Such communities must be included in the definition and evaluation of any “metrics for success”.

An improvement to infrastructure that allows easier access to communities currently underserved by AI research is important because this will allow for a longer sustainment for the NAIRR. Infrastructure improvements, particularly in underserved AI research communities, is key to the sustainment of NAIRR because it will ensure that a wide variety of communities will continue to use the resource. Additionally, this access will open up as yet unexplored possible avenues for the utilization of such a resource. This includes things like cheaper and faster internet access and computing technology.

1-b-ii. Options for roadmap element 1.B.ii should consider regarding a plan for ownership and administration for NAIR and specifically the governance structure for the NAIR, including oversight and decision-making authorities, are ones that should forefront issues of Indigenous Data Sovereignty and, more generally, a framework of

Inclusive Governance. This is important because it will acknowledge that not all data curated by the resource nor produced by it has the same potential for impact or influence in all communities. It should acknowledge that diverse data should come with diverse structures and cultural requirements for its governance. However, where data pertains to communities not historically minoritized, datasets and curated should strive to follow ethical open data principles and be in the public domain.

Furthermore, the roadmap should also consider a plan for ownership and administration of NAIRR including a governance structure, with oversight and decision-making authorities, that informs subjects who were part of research that created data training sets about when information about them is available or being used. This is important because it will ensure that the research developed out of NAIRR maintains a reciprocal and responsible relationship with the communities who are directly or indirectly incorporated in the work.

Additionally, the roadmap should also consider a plan for ownership and administration of NAIRR including a governance structure, with oversight and decision-making authorities, that includes the possibility for the retirement of data sets. A right to forget should be incorporated into this work and is important because not all information is necessarily collected to be used in perpetuity. In addition, our understanding of privacy may change where information about individuals, places, or other things included in this resource may no longer be considered to have been ethically obtained or relevant to the resource.

Lastly, the roadmap should also consider a plan for ownership and administration of NAIRR including a governance structure, with oversight and decision-making authorities, that acknowledges historical inequalities in the creation and construction of resources and training data sets with regards to this research. Specifically, not all groups have historically been treated with the same fairness when data sets are generated about them. Therefore, the governance structure should actively ensure that equitable application and data use exist. This is important because it will build trust in the resource amongst a wide variety of communities that could benefit from this research by acknowledging histories of differential access to such materials.

1-c Options for the goals for a model for governance and oversight to establish strategic direction, make programmatic decisions, and manage the allocation of resources should include a system similar to those of institutional review boards for human-subjects research in the sciences.

The goals for this model should include a focus on ethical and equitable governance and oversight that allow those in charge to establish ethical strategic directions, programmatic decisions, and allocate resources. This is important because it foregrounds the ways that such a resource can explore inequalities in use, education, and maintenance as well explore who the resource is really serving.

Additionally, the goals for the models for governance and oversight should utilize perspectives from outside direct AI research. This is important because those in the digital humanities and ethics explore the direct and indirect impacts of the resource that will aid in the creation of the decisions.

1-e Options for an 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 should include funding education in ethics and all fields that the resource intends to use data sets from (including education in racial and gender equity, fairness, bias, civil rights, transparency, and accountability), inclusive governance, and lastly, education in the technologies that underpin AI research that begin at the elementary level.

A lack of current funding in the above listed fields is a barrier to the dissemination of NAIRR products because it means a limited community can utilize the resource. A recommended solution to this is to fund education in these fields. Early education in these fields, additionally, will cultivate new and innovative ideas for the use of the resource that will better address problems currently facing American communities.

The lack of inclusive governance of adjacent resources is a barrier to the dissemination of NAIRR products because it means that the products will probably reinforce existing hierarchies in AI knowledge production. A recommended solution to this barrier is to incorporate practices of Inclusive Governance that draw from multiple cultural ways of understanding how knowledge and research should be conducted and the rights of certain communities to withhold information.

Lastly, the lack of education in the technologies that underpin AI research is a barrier to the dissemination of NAIRR products because it means that many people who may be interested in AI do not have the prerequisite skills to undertake this research if they approach it later in their educational careers. A recommended solution to this barrier is to add education in AI as part of elementary education. This should include education in the previously listed fields but also in the technologies that underpin AI so that a holistic approach to the use and importance of the resource is encouraged at the beginning of one's educational career.

1-g Options for an assessment of privacy and civil rights and civil liberties requirements associated with the National Artificial Intelligence Research Resource and its research should include an examination and restriction of the data private companies are able to retain or collect on individuals whose data sets might be incorporated and the right to forget or not preserve all material that goes into this resource.

This is important because our private companies are already able to retain a lot of information about users without accountability and to have a government agency who

will directly benefit from information by or about particular communities do that as well would not support civil rights and civil liberties.

The right to forget or have curated data sets be removed is also important because it acknowledges that not all collected data are necessarily useful forever. Particularly, information that pertains to individuals should have a limited shelf life to preserve their long-term rights to privacy.

2. The capabilities and services provided through the NAIRR that should be prioritized are educational tools and services, including provision of curated data sets, and secure access control.

3. The NAIRR and its components can 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, by being built with the aid of ethical and responsible creators. This particularly means the incorporation from the beginning of ethicists as well as individuals outside of current AI research who do the work on racial and gender equity, fairness, bias, civil rights, transparency, and accountability. Professionals in those fields will be the ones who can speak expertly on the issues most pertinent to AI research that will make the tool actually responsible and ethical.

It can also do this by having the program under an agency that is held more accountable by the public than the current funding and oversight agency. For example, many communities, especially communities with histories of disenfranchisement, would not trust Federal agencies with law enforcement or national security responsibilities to prioritize ethical and responsible research. Federal agencies with missions that directly promote human welfare and scientific progress would have more trust.

Lastly, NAIRR can reinforce those principles by focusing on the serious vetting of any and all training datasets. All training datasets must go through rigorous ethical and responsible scrutiny to ensure that they are not constructing bias that could have severe negative impacts on the previously stated topics. This vetting requires the incorporation of individuals from a variety of backgrounds so that the lifecycle of data sets is explored in terms of creation, management, dissemination, and incorporation to evaluate all possible sources of harm that could come from that data set. To do this requires many voices to be incorporated into the process.

4. There are a number of building blocks that already exist for NAIRR in terms of government, academic, or private-sector activities, resources, and services. The first suggestion would be for the OTSP and NSF to examine their currently funded projects to see which ones include artificial intelligence research, open data infrastructures, and

studies on related topics in racial and gender equity, fairness, bias, civil rights, transparency, and accountability. These already government-funded projects should provide a significant resource for the creation of curated data sets as well as investigations into secure access control, and the other items listed in 1.D.

Our organization's suggested building blocks primarily focus on organizations involved in the creation of educational tools and services as well as in the provision of curated data sets. Open Context, a project of the Alexandria Archive Institute, provides both educational tools and services as well as curated data sets that could be incorporated or built from for NAIRR. Other possible sources of open data sets include the iSamples project, tDAR, and WholeTale.

5. The role that public-private partnerships play in the NAIRR should be limited to private partnerships that focus on racial and gender equity, fairness, bias, civil rights, transparency, and accountability. Specifically, public-private partnerships should be sought with private entities that focus on the ethics and social justice outcomes that NAIRR may be interested in cultivating. Public-private partnerships that focus on the ethical and just outcomes for the research will be better able to utilize the product as well as ensure the construction of a resource that acts in an ethical fashion.

6. We see the see limitations in the ability of the NAIRR to democratize access to AI R&D in its affiliation, guidance, and funding from the DOD; if it does not incorporate questions of racial and gender equity, fairness, bias, civil rights, transparency, and accountability from the beginning; if scaffolding programs that include improvements to elementary education in AI related fields including racial and gender equity, fairness, bias, civil rights, transparency, and accountability studies are undervalued; and if it doesn't actively seek out methods to ethically democratize access.

To better promote democratization of AI R&D, we recommend that NAIRR should be entirely separate from National security and law enforcement agencies. Administrative control or funding from these agencies would undermine the credibility of NAIRR attempts to develop and promote responsible use of AI methods across the broader civil society. Instead, national security and law enforcement agencies should seek best practice, governance, and ethical guidance from initiatives supported by NAIRR, but NAIRR itself should operate independently from national security or law enforcement.

The limitation to democratizing access to AI R&D due to a lack of the incorporation of questions of racial and gender equity, fairness, bias, civil rights, transparency, and accountability from the beginning occurs because without an investigation of these issues equitable democratization is impossible because it will not be fully defined. Without identifying groups and areas of interest that lack access or have unequal access, it will not be possible to make that access equitable amongst all

groups interested or actively recruited into use of the resource. this can be overcome by incorporating these questions from the beginning of the tool's development process

The limitation to democratizing access to AI R&D due to a lack of scaffolding programs that begin with elementary education means that only certain kinds of researchers will have interest and access. The education required to develop AI technologies and ask ethical questions that AI could help solve, should start at the elementary level as early education often becomes a predictor for later outcomes. The ability to get funding and access to AI research is already an undemocratic system. Therefore, for NAIRR to actually change that, it would need to work to undercut the existing system of tiered access to AI education beginning at the elementary level. This could be solved by expanding elementary education funding in underperforming schools that support STEM, STEAM, and humanities education more generally, that will work together to provide students with the knowledge to succeed and utilize NAIRR.

The limitation to democratizing access to AI R&D due to a lack of inclusive governance could lead to democratization without justice. This could result in the continued marginalization of groups in the Nation. For example, Indigenous peoples make up a small percentage of the population; however, it is possible that data sets built from research that exploited Indigenous knowledge or peoples could be integrated with this resource and continue to do harm by not acknowledging Indigenous Data Sovereignty. In that case, it would be a democratized resource, but its impact could be disproportionate on certain groups. However, if these datasets do not include data about minoritized groups in the United States, because democratization can come into conflict with sovereignty and the rights of minoritized people, this resource will be inequitable, serving or creating research about only some individuals in the Nation. Democratized access that acknowledges the need for tiered access to AI R&D will provide a most just method for the resource.