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

RFI: National AI Research Resource University of Florida comments

Version 3 9/16/2021 [Note: 10-page limit]

Contact: Erik Deumens

Given the partnership between UF and NVIDIA enabled by the donation of the fastest AI supercomputer in academia, called HiPerGator AI, given the vision of UF to create an AI University that will educate all students in all degree programs in the basics, the applications, the risks, the promises of AI, given that UF makes HiPerGator AI available to faculty and students in the State of Florida under the same policies as it is available to UF faculty and students, given that UF makes HiPerGator AI available to institutions in the SEC and select national universities for the purpose of teaching classes, UF was lead to consider ways to support a large and diverse user community with staff resources sized to support a single university.

These comments will address questions 1. Options, 2. Capability prioritization, and 5. Public-private partnerships.

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

The comment addresses **B. Plan for ownership and administration i. Agency ii. Governance structure.**

Use a distributed model that is based on partnership with each institution willing to use the NAIRR to allow scalable support for all, not just the institutions and faculty who have sufficient expertise to use remote resources with minimal assistance.

Context: UF has 8 years' experience operating a large computing facility for its faculty and their students and national and international collaborators with a model that has proven to be sustainable. With the donation of HiPerGator AI, the Nvidia SuperPOD, to support UF's AI initiative, UF committed to provide access to a much larger user community, namely the faculty and their students and collaborators at all universities in the state. We set up a collaboration agreement with the IT organization at each university so that faculty get local support and the local IT staff works with UF IT Research Computing staff to address issues that go beyond what can be done by the local staff at each institution. This results in a distributed system of ownership of responsibilities and management of the resources and the support and training. UF provides the basic resources, but there are resources committed at each institution. This model is complementary to and distinct from the NSF XSEDE operation as well as the commercial cloud model. The experience during the first few months of operating this way is positive, but the period is too short for a definitive evaluation.

The comment address **C. Security requirements.**

Build the NAIRR to be compliant with security controls for restricted data from the start, such as 800-171 as required by DFARS.

Context: During the first months of operation of the HiPerGator AI, we were immediately faced with the need to handle vast amounts of restricted data for which extra security and compliance controls were required in order for the project to be allowed to move forward. Furtunately, UF has been operating a secure computing environment for some years that is compliant with NIST 800-53 moderate and NIST 800-171, which allowed UF to meet the restricted data requirements.

Question 2. Which capabilities and services (see, for example, item D below) provided through the AIRR should be prioritized?

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.

All these capabilities are required as well as training and consulting services.

Context: In the experience obtained during the first year of the UF AI Initiative, we have learned that researchers and students need all of the above including access to people knowledgeable in AI to build both introductory training as well as advanced training on AI and AI tools and consulting to teams who want to embark on a project (or write a proposal) that uses AI in any subject matter domain. The training is needed in the form of short videos, half-day workshops, and as support to faculty teaching semester-long courses involving AI. The majority of researchers and students need access to single GPUs, in which case the scaling goes with the number of users. In addition, there is the need to support very large AI problems that need the entire 140-node SuperPOD for several days to perform the desired machine learning on massive data sets.

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?

Governance of the NAIRR needs to include oversight from a broad community, unlike other computing resources that are well-managed by governance boards and review panels consisting of mostly technical experts.

Context: The NAIRR will have a governance structure of the NAIRR that sets strategic direction, defines the processes around resource allocation, including setting standards and requirements for proposals requesting access to the NAIRR.

The NAIRR must first determine whether there is a body of principles of ethical and responsible research and development of AI already in existence that is suitable for the needs of the community. We think that does not exist as yet. Therefore, the NAIRR could launch a task force with university researchers, industry leaders, and governmental personnel to draft a set of guiding principles that could be vetted by the community. Upon adoption, the creation and funding of a Center in Equitable AI or AI & Ethics or similar could be tasked with studying and training based on those principles. Further, it could be tasked with vetting AI products to certify that they meet the criteria set forth for ethical and responsible research in AI. In our experience at UF, the Equitable AI initiative is a step in that direction. However, an actual Center in this area could bring together people from a much broader swath of the AI community

To enhance the ability of the NAIRR to reinforce and advance the principles of ethical and responsible research and development of AI, it is imperative that the governance structure includes representatives not only from the technical computing and AI communities, but crucially involves people with background and expertise in all aspects of life and society: that means businesspeople, accountants, auditors, lawyers, philosophers to weigh in on issues of ethics and privacy, social and political scientists, psychologists in addition to the usual people who serve in these positions for computing resources. A characteristic of AI is that it directly impacts all aspects of life, unlike many other computing activities that have indirect impact, for example the computing done as part of building cars or airplanes is not visible to the people driving cars or flying in airplanes. This requirement or representation also extends to the composition of committees and panels that are asked to review proposed use of the NAIRR.

Question 4. What building blocks already exist for the NAIRR, in terms of government, academic, or private-sector activities, resources, and services?

The COVID High Performance Computing Consortium is a perfect example of a resource structure that can address the infrastructure requirements for the NAIRR. The same NSF and DOE supercomputing resources and the expertise to build, operate, and maintain them can be leveraged to build the NAIRR. The commercial cloud providers have infrastructure and processes that are ready to contribute to the capacity and capabilities of the NAIRR. Crucially, the resources in NAIRR are not just computing machinery and data storage systems. In addition, there is a need for education, training, and consulting in AI. Numerous universities are hiring faculty with expertise in AI. The NAIRR needs to formulate a mechanism to leverage them and their expertise as a resource.

Question 5. What role should public-private partnerships play in the NAIRR? What exemplars could be used as a model?

Public-Private partnerships are crucial.

Context: The partnership between UF and NVIDIA in the AI Initiative has been extremely beneficial, beyond the obvious fact that the system was donated to UF. NVIDIA has made training materials available and new materials are being developed together. The partnership

provides access to the engineers in NVIDIA so that when researchers run into difficult problems, caused by hardware or software/applications issues, they can be addressed collaboratively in the partnership to advance the mission of education and research as quickly as possible.

Question 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 limitations are the well-known obstacles of getting started. To make the NAIRR reach underserved communities, NSF needs to provide a complementary program like the CC* program to enable campuses to build the local infrastructure, including training and consulting support staff network bandwidth, to effectively make use of the NAIRR, no matter what its technical implementation and geographical distribution ends up looking like.