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

October 1, 2021

COMMENTS IN RESPONSE TO RFI ON AN IMPLEMENTATION PLAN FOR A NATIONAL ARTIFICIAL INTELLIGENCE RESEARCH RESOURCE (DOCUMENT NUMBER 2021-15660)

The non-profit Association for Computing Machinery (ACM), with more than 50,000 U.S. members and approximately 100,000 worldwide, is the world's largest educational and scientific computing society. ACM's US Technology Policy Committee (USTPC), currently comprising more than 140 members, serves as the focal point for ACM's interaction with all branches of the US government, the computing community, and the public on policy matters related to information technology. As such, the Committee strives to serve as an apolitical source of expert information.[1]

Overview

USTPC is pleased to respond to the Request for Information in the above-referenced proceeding[2] issued jointly by the National Science Foundation and White House Office of Science & Technology Policy.[3] Before doing so in detail below, however, the Committee wishes to underscore its fundamental recommendation that the new National Artificial Intelligence Research Resource (NAIRR) should be undertaken *only* if it is independently funded. Creation and operation of the NAIRR thus should not reduce current levels of federal financial support for Artificial Intelligence (AI) research by any other arm of the government.

[1] To arrange for a technical briefing from USTPC and other ACM expert members, please contact Adam Eisgrau, ACM Director of Global Policy & Public Affairs, at acmpo@acm.org or 202-580-6555.

[2] *See* 86 FR 39081 (July 23, 2021).

[3] Principal authors of these Comments for USTPC were: Jim Kurose, Distinguished Professor in the College of Information and Computer Sciences and Associate Chancellor for Partnerships and Innovation at the University of Massachusetts Amherst; and Arnon Rosenthal of Bedford, Massachusetts, a consultant and researcher specializing in data integration, databases, access controls, and their interactions with adjacent technologies. Also contributing were: ACM Technology Policy Council Chair Jim Hendler, USTPC Chair Jeremy Epstein, Vice Chair Alec Yasinsac, and USTPC members Joshua Kroll and Bulent Yener.

Q1. What options should the Task Force consider for any of roadmap elements A - I?

Element C —A model for governance and oversight to establish strategic direction, make programmatic decisions, and manage the allocation of resources

A common infrastructure for all. Currently, computing research is funded at the national level by a relatively few agencies each (understandably) funding access to research computing infrastructure primarily for the researchers that it directly sponsors, reflecting the culture and organization of the funding organization.[4] Thus, for example, while some DOE-funded researchers not supported directly by NSF are able to access NSF-funded resources, and *vice versa,* such resource sharing is not common. As a national resource targeted at research in a specific area, the NAIRR should not be partitioned or "siloed" according to funding agency.

Coupling NAIRR allocation review with agency review. The demand for NAIRR resources is virtually certain to exceed their availability. Accordingly, a process to assess the readiness and appropriateness of proposed resource allocations will be required. Several national models exist for accomplishing such resource allocation[5] that have generally served the community well and profitably might be emulated by the NAIRR. USTPC strongly recommends, however, that NSF's specific practice of de-coupling approval of basic research decisions[6] from others made in the LRAC/XRAC-resource allocation process *not* be adopted. These two sets of decisions, in USTPC's view, ideally should be made in tandem, thus permitting the totality of a project to be fully and accurately evaluated.

Obtaining maximum value from data. Data have fundamental different characteristics, use patterns, and cost structures than the "use it or lose it" characteristics of computational resources. Unlike processors, data can be accessed and used simultaneously by many recipients with minimal marginal costs; "public" datasets can also be downloaded and processed with users' own computational resources. If NAIRR is hosted on a cloud, processing or export might be done on the user's cloud account, sparing NAIRR the burden of billing. It also should be possible, and would certainly be desirable, for the NAIRR to negotiate favorable rates for data export to other clouds. It would be valuable for the NAIRR to support these, and other, approaches towards data access, migration, and use.

---

[4] The national computing research infrastructure is presently funded primarily by the Department of Energy and National Science Foundation with smaller scale investments being made by the National Oceanic and Atmospheric Administration, National Institute of Health, NASA, and other agencies.

[5] These include: the Leadership Resource Allocation Committee (LRAC) and XSEDE Resource Allocation Committee (XRAC) for computational resources funded via the NSF Office of Advanced CyberInfrastructure (OAC) and INCITE for resources funded via the DOE Office of Science in the U.S. Department of Energy. LRAC and XRAC, which provide merit-based, peer-review allocation processes for resources that would otherwise be oversubscribed.

[6] For example, traditional agency-convened merit-review panels make recommendations to fund research that requires access to computational resources but do not provide funding for access to these resources.

A community-based technical advisory committee. As experience is gained once the NAIRR becomes operational, USTPC recommends that lessons to be learned and efficiencies gained from that experience might best be analyzed by a computing community-based expert advisory committee charged broadly with providing the NAIRR with advice and feedback. By no means incidentally, USTPC also notes that the composition of such a community-based committee should be carefully and deliberately defined to ensure democratized access to the NAIRR's resources by diverse researchers and for diverse socially relevant purposes. Care thus also should be taken to assure that the committee's makeup broadly reflects both the diversity of the research community and the communities intended to benefit from funded research.

Element H — A plan for sustaining the National Artificial Intelligence Research Resource, including through Federal funding and partnerships with the private sector.

Coordinated decision making. Many of the AI researchers utilizing the NAIRR are likely either to be funded by one or more federal agencies or work at a federal agency or laboratory. As noted above, it is important that NAIRR allocation decisions be made in concert with research funding allocation decisions made by these agencies.

Overhead expense relief. USTPC recommends that NAIRR access *not* be subject to overhead recovery costs if such access is explicitly allocated in an agency research award. As noted in the Report of NSF's workshop on *Enabling Computer and Information Science and Engineering Research and Education in the Cloud,*[7] this would "level the playing field" between an organization's option to *(i)* acquire hardware and software computing research assets, which are not subject to overhead, and *(ii)* acquiring access to similar assets in the cloud or elsewhere, which typically is subject to overhead charges at most institutions.

Element I —Parameters for the establishment and sustainment of the National Artificial Intelligence Research Resource, including agency roles and responsibilities.[8]

Multi-agency oversight and budget. Researchers funded by Federal agencies, or those working within them, will likely be among those with access to the NAIRR. Thus, USTPC suggests, multiple agencies should be assigned an oversight and governance role within the NAIRR. Several agencies with significant investment in AI research (e.g., NSF, DOE, NIH, NIST, NOAA) might appropriately be considered "lead" agencies in such a governance structure. Given AI's pervasive and growing application in many spheres, however, it also will be important for other agencies (*e.g.*, addressing Agriculture, Commerce, Transportation, Justice, and more) to play roles in defining and operationalizing the NAIRR.

---

[7] See www.researchgate.net/publication/329911307_Enabling_Computer_and_Information_Science_and_Engineering_Research_and_Education_in_the_Cloud (July 2018).

[8] USTPC notes that much of the following material also potentially is germane to implementation roadmap element B ("A plan for ownership and administration" of the NAIRR) and requests that it also be considered in that context.

Budget flow and agencies. If the budget for the NAIRR is funneled through a single agency, that agency's researchers (whether externally funded or in-house) may be perceived as having priority access to its resources, even if this is not the case in fact. Without budget flow, a resource like the NAIRR may become less visible within the agency. To avoid such an appearance of conflict of interest, and to ensure agency engagement, NAIRR should be structured to route funding through multiple agencies without, USTPC again emphasizes, being required to draw upon their existing budgets. Potential models for such multi-agency funding include major research instrumentation projects, such as telescopes and accelerators.[9] A non-agency, non-governmental entity may serve to more easily and effectively manage the NAIRR, and as an added benefit help facilitate the development of public-private partnerships.

Funding scope. The scope of the NAIRR's support must be carefully defined and tailored to AI *research computing infrastructure* costs rather than to other AI research components. Expenses such as faculty and graduate student compensation, travel, and more all are necessary and complementary to AI research computing expenses. NAIRR funding, however, will be most effective if dedicated to covering purely AI research computing infrastructure costs.

Open research. Similarly, it should be made clear that the NAIRR's principal aim is enabling open, basic AI research. Funding for computational resources for the commercial development of AI technologies should be provided by other means, such as venture capital investment, economic development fund underwriting of early-stage technologies, or direct industry investment in established technologies by more mature companies.

Q2. Which capabilities and services provided through the NAIRR should be prioritized?

The inquiries made by the RFI in this proceeding clearly are predicated on the authoring agencies' conviction that the NAIRR should support much more than mere access to computational capacity and raw data sets. USTPC broadly concurs with this fundamental premise, and with regard to the following specific capabilities and services:

Computing. Whenever possible NAIRR should leverage commercial cloud offerings, developing its own specialized computing resources only after conducting a comprehensive cost/benefit analysis that justifies doing so. The NSF's 2018 NSF workshop Report on *Enabling Computer and Information Science and Engineering Research and Education in the Cloud* cited at note 7 above provides a thoughtful discussion of the many advantages (and possible disadvantages) of this approach.[10]

---

[9] Examples include AURA, which operates and builds world-class astronomical observatories for NSF and NASA, and US IGNITE, which helps manage NSF's PAWR initiative.

[10] *See also* "Cloud Access for NSF CISE Research" at https://cra.org/cloud-access-for-nsf-cise-research/.

Data. To maximize access to data and its utility to researchers,[11] USTPC suggests that the NAIRR:

- recognize that while money alone can provide access to computing capacity, it cannot always guarantee access to data necessary for a particular research effort.[12] Data thus should be presumed to be a community resource to be managed and made widely accessible accordingly;

- enhance the value of communal data sets by encouraging that user commentary and experiences be associated with all accessible data sets;[13]

- maximize the ways in which all accessible data sets are characterized, categorized, and cross-referenced to enable users to most easily interpret the data, determine its fitness for use, and assess whether and how it may be biased;

- mandate that all accessible data be accompanied by metadata sufficient to understand its provenance, meaning, and trustworthiness;

- assure that the collection of NAIRR-hosted datasets is "FAIR": Findable, Accessible, Interoperable, and Re-usable to the maximum (and likely varying) degree to which each such goal is feasible;

- foster interoperability by devising and encouraging data generators to adopt standards for its form, metadata, and organization but, to avoid discouraging contributions, nonetheless accept data in native form if augmented by sufficient descriptive material to permit others to supplement or complete its description;[14]

- develop the capacity to host anonymized or otherwise "privatized" datasets (created, for example, using differential privacy) derived from confidential origin datasets, and to provide information describing how these data sets were created and might be utilized. The NAIRR also should consider hosting non-public datasets through an interface designed to provide only "sanitized" products (*e.g.*, selected "safe" statistics, not original data);

---

[11] While pursuing and sharing data is a valuable goal, *data alone does not guarantee progress in AI research*. Prioritizing "more data" too often drives policy making when "better data" or "understandable data" should be preferred.

[12] While a great deal of data (often in huge sets) is generated by and accessible within large corporations and other private entities, clearly not all researchers have access to these resources. To the extent possible (e.g., as constrained by competitive commercial concerns of data owners, and privacy or security considerations) the NAIRR can and should make such data broadly available to researchers everywhere.

[13] For example, "Q&A" documentation and shared code (*e.g.*, as currently available through www.stackoverflow.com) would enhance data set value for all researchers.

[14] While this objective might theoretically best be achieved through mandated standards compliance (or by declaring parts of a dataset to be semantically the "same as" part of another set), USTPC is aware that many data owners may be unwilling or unable to conform to such a standard.

- host community-developed code designed to enhance and enable the manipulation of datasets, as well as models built on such information;[15]
- broadly define the term "data resources" to include "higher-level" data products. Such synthesized data structures, like knowledge graphs, are critical for AI research but often are available at scale only within industry settings.[16]

People. Raw computational capabilities and datasets are a necessary, but far from sufficient, set of resources to meet the NAIRR's goals. Perhaps most importantly, people-centric services like education, training, expert consulting, and community outreach[17] all must be key components of an accessible NAIRR that seeks to successfully serve a broad and diverse community of AI researchers.

Q3. 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?

Democratizing access: beyond supporting agency-funded AI research. USTPC recommends that, as NSF presently does in connection with funding computational resources,[18] the NAIRR set aside a specific and not insignificant percentage of its total available resources for otherwise unfunded developmental, exploratory, and other meritorious projects that demonstrate the need and appropriateness for NAIRR resources. A portion of these reserved resources should in USTPC's view be dedicated to projects that concern issues of racial and gender equity, fairness, bias, civil rights, transparency, and accountability.

USTPC also reiterates here its recommendation responsive to Question 1, Element C above that NAIRR be guided in part by a diverse community-based committee with a broad community perspective (in addition to XRAC/LRAC/NSERC-like bodies) charged with advising on and overseeing resource allocations. The committee should be expressly charged with, among other responsibilities, ensuring democratized access by diverse researchers to the NAIRR's resources. As also noted above, the committee's makeup also should broadly reflect both the diversity of the research community and the communities intended to benefit from funded research.

---

[15] More generally, a model "commons," similar to NIH's notion of a "data commons," might include systems like the Generative Pre-trained Transformer 3 (GPT-3) for use by the research community.

[16] See*, e.g.,* "Open Knowledge Network" NITRD Big Data IWG workshop report (November 2018).

[17] The NAIRR might consider, in particular, outreach by distributed teams of local "champions" modeled, perhaps, XSEDE "campus champions".

[18] In the case of NSF-funded computational resources, 20% are reserved for specific uses not subject to LRAC/XRAC review, or to be awarded at the discretion of the Director of the institution operating that resource.

Finally in this context, USTPC notes that the Association for Computing Machinery's Diversity, Equity, and Inclusion Council also could be an important resource for the NAIRR in populating the recommended committee and developing practices to improve diversity in NAIRR's organization and programs by leveraging ACM's broad and deep connectivity to the nation's AI research community.

Q4. What building blocks already exist for the NAIRR, in terms of government, academic, or private-sector activities, resources, and services?

USTPC recommends that the NAIRR consult existing recent pilot existing programs and resources relevant to the NAIRR's vision and mission, such as NSF Cloudbank, NIH STRIDES, and the 2018 NSF Workshop Report on *Enabling Computer and Information Science and Engineering Research and Education in the Cloud.*

Q5. What role should public-private partnerships play in the NAIRR? What exemplars could be used as a model?

USTPC sees significant potential for NAIRR to maximize the impact of its resources by catalyzing a public-private partnership[19] to collaboratively construct the NAIRR's physical infrastructure in the cloud with access afforded by commercial cloud vendors. Such a structure would allow the NAIRR itself to focus on what may well be its "highest" functions: providing coordination and funding for resource access and use. It also might productively explore offering any of a myriad of services that may be "layered" on top of the physical infrastructure, much as Cloudbank and STRIDES, both referenced above, do so now.[20]

Conclusion

USTPC commends the NSF and OSTP for initiating this proceeding to help launch the NAIRR which, USTPC believes, can bring significant benefits to the AI research community and the nation. As noted in the updated *2019 National AI R&D Strategic Plan,*[21] AI holds tremendous promise for society, and provides important opportunities for national economic competitiveness, health and welfare, defense, and security. The NAIRR can play an important role in achieving these benefits by providing democratized access to AI research computing and data infrastructure, and to services built on and around this infrastructure. USTPC looks forward to working with NSF and OSTP to those vital ends.

---

[19] These Comments also touch upon private-public partnership potential in the context of responding to RFI Q1, Element H above.

[20] This concept is not theoretical. Multiple cloud providers have collaborated to provide resources to researchers funded under the NSF BIGDATA program and commercial cloud services also have hosted large NOAA datasets.

[21] See https://www.nitrd.gov/pubs/National-AI-RD-Strategy-2019.