



MICHAEL RICHARDS

Director

1615 H STREET, NW
WASHINGTON, DC 20062-2000
(202) 463-5518
MRichards@uschamber.com

October 1, 2021

The Office of Science and Technology Policy &
The National Science Foundation
Attn: Wendy Wigon, NCO
Alexandria, VA 22314

Re: Request for Information on an Implementation Plan for a National Artificial Intelligence Research Resource; 86 FR 46278

To Whom It May Concern:

The U.S. Chamber of Commerce's Technology Engagement Center ("C_TEC") appreciates the opportunity to submit comment to the Office of Science and Technology Policy (OSTP) and the National Science Foundation (NSF) Request for Information (RFI) on "an Implementation Plan for a National Artificial Intelligence Research Resource." C_TEC supports OSTP's and NSF's work to develop a National Artificial Intelligence Research Resource (NAIRR), which will produce a roadmap and implementation plan to create a shared computing resource for AI.

One of the guiding AI principles for C_TEC is to "promote Open and Accessible Government Data." C_TEC acknowledges that National Artificial Intelligence Research Resource will look to accomplish just that by leveraging large and robust government data sets, which will help spur further innovation and breakthroughs within the scientific community. We support NAIRR's commitment to making this data available in an accessible manner for the research community.

C_TEC wishes to provide the below feedback on OSTP's and NSF's request for information on the "Implementation Plan for a National Artificial Intelligence Research Resource." These comments include:

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

a. Goals for establishment and sustainment of a National Artificial Intelligence Research Resource and metrics for success;

C_TEC believes that setting goals and metrics are significant for tracking the success of the National Artificial Intelligence Research Resource. We agree that the following metrics and objectives should be considered: research usage, community uptake, and wider impacts.

Research Usage: One of the easiest ways to determine if research resources are being utilized is the overall usage. Therefore C_TEC believes that as NAIRR should use the metric of overall researcher usage.

Community Development: A critical metric to consider is the overall development of the U.S. AI community. Specifically, NAIRR should look at broad participation in workshops related to Artificial Intelligence, investment in R&D, the development of AI tools, the creation of new benchmarks or standards adopted by the community, and the number of experiments shared within the community.

Wider impacts: NAIRR will have the profound ability to assist in the development of resources that can drive innovation and lead the development of significant breakthroughs within scientific fields. This is why C_TEC believes that tracking breakthroughs within the field should be used as a metric, increases in productivity, automation, as well as the number of new products to enter into the market, and the number of startups created.

b. A plan for ownership and administration on the National Artificial Intelligence Research Resource, including:

i. An appropriate agency or organization responsible for the implementation, deployment, and administration of the Research Resource;

C_TEC can foresee multiple models for government ownership for the initial implementation, deployment, and administration phase for NAIRR. Whichever model is chosen, however, we encourage that it undergo review once the research resource has matured to ensure maximized utilization of the resource.

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

C_TEC understands that NAIRR's effort to grow computing resources and high-quality data sets simultaneously is no easy task. For this reason, we encourage NAIRR to review the merits of housing the research resource within a Federally Funded Research and Development Center (FFRDC). FFRDC has a strong track record of spurring research and development and would allow multiple stakeholders to work together on the underlying goal of improving access to data to accelerate scientific discovery. Furthermore, C_TEC believes that multiple agency governance structures could be beneficial, in that it would allow for further buy-in from other agencies. We encourage the FFRDC to be connected with the Department of Energy (DOE), the National Institutes of Health (NIH), and other agencies, with input from OSTP.

c. A model for governance and oversight to establish strategic direction, make programmatic decisions, and manage the allocation of resources;

C_TEC strongly supports allowing the industry to compete for the development and buildout of computing and data resources, AI tools, and platforms. We believe that the procurement of AI software and data management tools should be done in an open, transparent process. Furthermore, we would highlight the need for resources to be interoperable. However, we have concerns regarding total

homogenization of the resources as diversity ensures that researchers access the resources across multiple clouds and interfaces. Diversity is critical in allowing for quick utilization of the resources by researchers.

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

Regarding data sets and secure access control, C_TEC supports the principles of findability, accessibility, interoperability, and reuse of digital assets or better known as the FAIR guiding principle for scientific data management and stewardship. Furthermore, we believe data sets should be easily located and legible for both human and machine consumption.

Regarding compute resources and scalability, C_TEC believes that NAIRR should take advantage of commercially available computing resources. The research community is already familiar with these resources and would allow for quick adoption and utilization. Furthermore, we believe NAIRR should adopt a hybrid, secure, multi-cloud approach to provide cost-effective computing at the necessary scale. NAIRR should also look for opportunities to leverage existing public clouds and security layers and integrate those in the multi-cloud.

Finally, regarding Educational Tools & Services/Resident Expertise, C_TEC believes that effective communication between researchers can spur collaboration and help further drive innovation and determine the reproducibility of results. We encourage collaboration between individual researchers, government entities, and industry and we support engagement necessary to exchange technical expertise and best practices in order to find synergies within their respective work.

e. 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;

C_TEC understands that many issues will need to be resolved before the dissemination of data sets by NAIRR. The following are areas we believe should be further addressed.

First, with the large amount of data coming from multiple agencies being compiled together, there are concerns that such data could have personally identifiable information (PII), Personal Health Information (PHI), or other sensitive data associated with it. For this reason, we encourage NAIRR to deidentify and secure all sensitive data. Also, NAIRR should look at how it can automate the ability to remove and protect sensitive data, as well as anonymize the data.

Second, the mass compiling of datasets could lead to issues with finding the correct data for specific research. C_TEC believes that it would be prudent to develop a hybrid data fabric to assist researchers in finding appropriate data more efficiently. Also, C_TEC would encourage the use of automated tagging and labeling, including looking into opportunities for supervised and unsupervised learning for tagging, to make data search and data understanding more accessible for researchers.

f. An assessment of security requirements associated with National Artificial Intelligence Research Resource and its management of access controls;

C_TEC believes that NAIRR will need to be secured to ensure that the data is not abused, misused, or otherwise compromised. We would encourage the use of security control such as access control, encryption, authentication, logging, and many others that may provide necessary security to the resources.

g. An assessment of privacy and civil rights and civil liberties requirements associated with the National Artificial Intelligence research resource and its research

Data is critical to the development of AI, and the repurposing of personal data may impact consumer privacy and trust in these research efforts. Therefore, clear and consistent privacy protections for individual privacy are necessary. However, should that not be plausible, we encourage NAIRR to implement robust but flexible data protection regimes that enable data collection, retention, and processing for AI development, deployment, and use while ensuring alignment with existing privacy laws.

2. What capabilities and services provided through NAIRR should be prioritized?

C_TEC believes multiple aspects of NAIRR should be prioritized simultaneously. These include the buildout of a hybrid cloud platform, the development of AI and data management software, open standards and open technologies, and consistent methods and tools.

First, NAIRR should prioritize developing a hybrid cloud platform that can provide a seamless user experience across multiple clouds.

Second, NAIRR should further prioritize the development of software that can assist the research resource's productivity, simplicity, portability, and reproducibility.

Third, NAIRR should prioritize the development of standards and open technologies which can assist with the adoption of cloud use and assist in reproducibility.

Finally, NAIRR should prioritize the development of consistent methods and tools to secure the data and resources.

3. How can the NAIRR and its components reinforce principles of ethical and responsible research and development of AI, such as racial and gender equity, fairness, bias, civil rights?

C_TEC supports the focus on fairness and non-discrimination within NAIRR RFI. Fairness and non-discrimination principles are essential for establishing public trust in AI.

4. What building blocks are already for NAIRR regarding government, academic, or private sector activities, resources, and services?

C_TEC believes it is essential to review and incorporate other building blocks from existing collaborations to help NAIRR. Public initiatives such as the European Open Science Cloud, NIH STRIDES, and NSF Cloudbank have all connected researchers to datasets through the cloud. Other efforts that should be considered include the COVID-19 High-Performance Computing Consortium,

which was able to bring together academia, industry, and the government to harness high-performance computing to support COVID-19 research. Finally, we would highlight the Helix Nebula Science Cloud pilot program, creating a shared research space for millions of researchers.

5. What role should the public-private partnership play in the NAIRR? What examples could be used as a model?

C_TEC emphasizes that public trust-building efforts are conducted with government, industry, and other relevant stakeholders. A public-private partnership model will facilitate collaboration between all relevant stakeholders and allow for sharing of best practices.

6. Where do you see limitations in the ability of the NAIRR to democratize access to AI R&D? And could these limitations be overcome?

NAIRR could experience many limitations to democratize the access to AI R & R&D, including a lack of the necessary workforce and skills needed to utilize the resources associated with research funding.

C_TEC believes in the use of commercially available resources. However, some researchers may face the challenges of learning how to use the resources and utilize them to accelerate their research fully. Therefore, an open line of communication between government, industry, and academia to learn best practices and necessary training to reduce the skills gap may be required.

NAIRR will provide further access to the necessary compute and data to allow for breakthroughs in scientific research. Yet, this does not solve the problem of having critical AI scientists and the research and development dollars to assist with their research. As the United States looks to develop this resource, we must also simultaneously look at ways to increase AI research transpiring in the field. This should include looking at federal research grants that may not have previously been used for AI research and developing the AI research workforce.

Conclusion

C_TEC appreciates NSF's and OSTP's ongoing work to develop NAIRR. We encourage further collaboration with stakeholders as we believe the partnership is vital for developing research resources that are fundamental for future scientific discovery. We thank you for your consideration of these comments and would be happy to further discuss any of these issues.

Sincerely,



Michael Richards
Director, Policy
Chamber Technology Engagement Center