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

Request for Information (RFI): Hyperion Research's response to RFI on an Implementation Plan for a National Artificial Intelligence Research Resource

Response Date: September 1, 2021

Organization: Hyperion Research

Point of Contact:
Mike Thorp
Sr. Global Account Executive
Hyperion Research

Contributing individuals:
Earl C. Joseph - CEO
Steve Conway - Senior Adviser, HPC Market Dynamics
Mark Nossokoff – Senior Analyst
Alex Norton - Principal Technology Analyst and Data Analysis Manager

Enclosed is Hyperion Research's response to the RFI regarding the development of an implementation roadmap for a shared research infrastructure that would provide Artificial Intelligence (AI) researchers and students across scientific disciplines with access to computational resources, high-quality data, educational tools, and user support.

Hyperion Research is a leading industry analyst, market research firm focused on the High Performance Computing (HPC) market and associated technologies, including deep insights and expertise relative to Artificial Intelligence (AI). We have current engagements and many past successful engagements with federal government organizations, including Department of Energy (DOE), Department of Defense (DOD), National Science Foundation (NSF) and various intelligence agencies. Hyperion Research hosts the HPC User Forum, since 2000, we have held 77 HPC User Forum meetings in the U.S. and around the world. The HPC User Forum was established to promote the health of the global HPC industry and address issues of common concern to users. The organization is directed by a volunteer Steering Committee of users from government, industry and academia, and operated for the users. (www.hpcuserforum.com)

We look forward to addressing any follow-up questions the NAIRR Task Force may have relative to our input. We would also ask to be considered for input for potential future requests and direct engagement with the Task Force.

Hyperion Research proposes the following recommendations to the RFI response questions:

RESPONSE SECTION:

This RFI seeks input from a broad array of stakeholders on the topics set forth below. Comments from the public will be used to inform the Task Force's consideration of options and development of an implementation roadmap.

Responders are invited to provide answers to the following questions (please number your responses accordingly):

1. What options should the Task Force consider for any of roadmap elements A through I above, and why? [Please take care to annotate your responses to this question by indicating the letter(s) of the item (A through I in the list above) for which you are identifying options.]

NAIRR Roadmap Element	Suggestions
A	• *Establish subcommittee to be led by a member of the Task Force to develop metrics for success. Commission a private study of NAIRR constituents to determine their needs and requirements for success.* • *To sustain a successful NAIRR roadmap and success, understanding what foreign entities are doing relative to AI is critical. Developing a formalized mechanism to identify and track non-US based technologies (HW, SW, tools, infrastructure) could provide the necessary global insights.* • *Once goals and metrics are established, developing a methodology for tracking NAIRR progress to attainment of those goals. This could also include comparison and progress against international organizations and countries, including China.*
B i.	
B ii.	The Covid-19 HPC Consortium is an excellent model off which to build a similar governance, oversight, and decision-making authority.
C	
D	• *Commission a private study to identify best practices and pitfalls to avoid for the creation and maintenance of existing shared computing infrastructure to facilitate access to advanced computing resources for researchers across the country.*

	• *Create a consortium or advisory council of public and private sector researchers and organizations to share best practices and develop standardization and benchmarks of AI technologies. By bringing together multiple organizations with different approaches and skillsets, collaboration can facilitate deeper research and capabilities.*
E	*Tracking of progress of NAIRR goals. Tracking NAIRR progress compared to international organizations/countries, including China.*
F	*Commission a private study of NAIRR constituents to identify barriers and propose solutions for the dissemination and use of high-quality government data sets.*
G	
H	*Commission a private study of NAIRR constituents.*
I	*Commission a private study of NAIRR constituents.*

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

A non-exhaustive list of capabilities and services to be prioritized include:

- *Standardization and hosting of public data sets to enable broader access to high quality, large data sets for training.*
- *A concerted focus on developing the necessary AI skillsets in undergraduate and graduate-level courses and degrees. If there are more skilled researchers out there, there is less competition for building AI research teams.*
- *Access to the necessary hardware solutions, especially access to the latest and greatest accelerators from market staples to emergent technologies.*
- *NAIRR should heavily leverage and model the existing US infrastructure for advanced scientific and engineering research, especially the country's unrivaled network of NSF centers and DOE labs.*
    - *These centers/labs have long relied on HPC to support their advanced research. Hyperion Research studies, including for DOE, show that nearly all these sites already perform AI research to a greater or lesser extent. It therefore makes good sense in our opinion for the NAIRR to tie into this work.*
    - *America's next-generation exascale and pre-exascale supercomputers, intended for DOE and NSF sites, have been expressly designed to support both existing scientific-engineering workloads and new and emerging AI*

workloads. Hyperion Research studies funded by federal agencies or private-sector firms also support leveraging these current efforts.

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?

   Many corporations and organizations establish and communicate a set of corporate values to which they aspire and live. Following this model, the NAIRR could:

   - *Identify a set of key "AI" values, provide examples, and actively, repeatedly communicate and promote them.*
   - *Require partner organizations and constituents to also pledge to support and comply with the values.*

   Additionally, the NAIRR could lead the effort to create a global set of standards and benchmarks by:

   - *Developing a set of benchmarks to measure things like data bias for race and gender.*
   - *Establishing a third-party organization to review and regulate outputs of models to ensure civil rights are not violated.*

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

   There already exist some public data sets made available through published work and from government agencies, like the National Cancer Institute (NCI) and National Institute of Health (NIH). Continuing to expand those, as well as working out the funding and management of shared public datasets, is crucial to enabling future developments in AI.

   Private companies like Google, OpenAI, IBM, Facebook, and others have created consortiums to impact standardization in certain areas of the AI space.

5. What role should public-private partnerships play in the NAIRR? What exemplars could be used as a model?
   a) *Public-private partnerships should play a key role in the NAIRR. Collaboration between industry, academia, and government has been a hallmark of the HPC community for decades. One of the most recent and on-going exemplars is the*

COVID-19 High Performance Computing Consortium. Under the leadership of OSTP, IBM, the Department of Energy, and other federal agencies, the consortium organized a broad array of public and private entities in an unbelievably short period of time to bring the appropriate resources to bear to combat COVID-19. In addition, providing access to the physical advanced computing resources, the consortium's efforts included governance, resource access, project approval process, and resource support.

b) *Review and leverage a prior US and global study performed on behalf of NCSA and NSF to identify and characterize public-private partnerships supporting advanced research. Utilize the findings of this study as a manual of best (and worse) practices for partnerships of this kind. A new, similar study could also be commissioned.*

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?

a. *Private organizations can be protective and competitive when it comes to advances in AI capabilities. It will be important to establish the areas and rules of engagement for collaboration. These organizations need to understand the value of sharing best practices.*

b. *Data privacy laws are getting stricter and stricter. Enabling shared data resources within the constraints of privacy laws is critical.*