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.
Introduction

Hewlett Packard Enterprise (HPE) believes artificial intelligence (AI) can amplify human capabilities and that investments into the cyberinfrastructure that fuels AI research and development is in our best national interest. That’s why we appreciate the opportunity to respond to the Request for Information (RFI) to assist the NAIRR Task Force in establishing and sustaining a NAIRR and propose a roadmap detailing how such a resource should be established and sustained. Driven by the promise of AI technologies, scientists across the country are eager to study AI and explore how it may benefit their work. NAIRR should capitalize on these scientists’ enthusiasm and untapped potential by democratizing access to facilities, funds, data sets, knowledge, tools, and any other resources that would allow them to study and leverage AI. For this goal to be actionable, our responses target three key recommendations:

- **Prioritizing, enabling access to AI-ready datasets** – This includes providing clarity on where to find relevant data sets for use in priority areas such as climate change research as well as information about the data sets provenance and metadata to understand details of the data itself.

- **Leverage existing computing facilities and infrastructure wherever possible** – Advancing AI will also require significant computing resources. To maximize government return on investment in these resources, we recommend building on existing computing facilities and infrastructure whenever possible. Additional investments should ideally be targeted to augment these capabilities as needed. This balanced approach would accelerate the availability of NAIRR computing capabilities and also, importantly, enable the Nation to benefit from the decades of experience and expertise that have developed in our National Labs and other national computing centers.

- **Continue support for research into foundational AI methods** – AI is an active field of research, in which everything from processor design and model architectures to overall workload optimization and best practices for mitigating systemic bias remain open areas of innovation. Support for this foundational research should continue to be a priority, and NAIRR should seek to develop processes and pathways for testing and translating new innovations into the resources it makes available.

Response to RFI Questions

**Roadmap Elements**

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) for which you are identifying options.]

**Establishment Goals/Sustainment and Metrics for Success**

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

The NAIRR resource has the potential to add value along the following dimensions for academic, industry and government stakeholders:

- **Scientific** – e.g., make discoveries possible
- **Technological** – e.g., create an ecosystem of innovation that includes for-profit and non-profit uses
- **Data-access** – e.g., democratize access to data resources (e.g. maps, weather, etc.)
- **Scalability** – e.g., ability to grow to more users, more partners, more contributors
• Collaborative – e.g., attract and develop talent

True success on the scientific, technological, access, scalability and collaboration dimensions can be measured using one or more of the following measures of success:

- Increased productivity and reduced time-to-insight
- Number of individual, small-business, non-profit, and corporate users
- Number of papers, press-releases, publications that acknowledge the resource
- Increasing activity among networks of vibrant communities of creative domain scientists and data scientists that can collaborate effectively
- Platforms to orchestrate/instantiate accelerated discovery workflows
- Increase use of AI in addressing high-priority, cross-cutting urgent national challenges such as climate change

Plan for Ownership and Administration

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

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

Envisioning an effective and impactful national artificial intelligence research resource will require federal leadership at the top. Expanding the coordinating and convening role of OSTP and OMB, and to some extent, the NITRD around this resource is likely a good first step. Those agencies and efforts have been successful historically with coordinating federal efforts in areas such as high performance computing, large scale networking and cybersecurity. With Artificial intelligence R&D being a program component area for the NITRD for several years now, it makes sense to leverage ongoing activity and established intergovernmental relationships. Furthermore, the annual OSTP/OMB multi-agency research and development priorities memo continues to urge agencies across government to prioritize, promote and collaborate on AI. Keeping and empowering these mechanisms is likely to lead to faster progress than having to start up a new bureaucracy, from scratch. It also helps provide important top level direction and guidance in support of a national AI research resource.

That said, it is important to understand the evolution of AI research resources in the US. At HPE, we are intimately familiar with this evolution and see our company as a worldwide leader, particularly at the high end. We are experiencing firsthand how pervasive and rapidly evolving the field of AI is. In our experience, organizations across the country, at various levels of sophistication, are trying to use AI to improve their operations and activities. Most are overwhelmed with what they face. Keeping up with AI’s rapid pace of change is exceedingly difficult for most academic, commercial, government and organizations. Knowing this should be fundamental when trying to offer a nationwide research resource.

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

HPE recommends that an advisory board with membership in a diverse community of stakeholders be established to work with the office responsible for implementation, deployment, and administration.

To ensure the National AI research resource is broadly available throughout the United States, in urban, suburban and rural communities, and especially within communities that are traditionally underrepresented
in the development and use of technology, the federal government will need the participation of nearly every agency. AI will heavily impact all NITRD participating agencies and programs. But there are many more agencies and programs beyond NITRD that need to participate. The fastest way to accomplish this is to keep most of the responsibility for implementation, deployment and administration within existing federal programs and agencies with input from the advisory board to assure the full diversity of the community is recognized.

The big piece that is missing is the coordination piece. AI, including machine learning, deep learning, inferencing and other forms, is already computationally expensive and will get much more expensive in the coming years. Furthermore, data is also growing exponentially, in part because of inexpensive sensors and the ability to collect a lot of data and can be exploited in the use of AI. AI requires storing and moving large amounts of data. When looking at AI users across communities, we see enormous opportunities to share AI artifacts, models, data, computational results and more. This would produce dramatic savings in resources, including power and reduce potentially massive redundancies. A key role of a research resource would be to bring communities together to improve coordination between organizations at the many levels they operate.

Model for Governance and Oversight

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

No one new bureaucracy is likely to be capable of doing all this work in a reasonable amount of time. By leveraging existing agencies and their current oversight and decision-making operations, we can build on existing processes that have been honed over time, and augment with new organizational infrastructure only where needed. Some agencies, like the Departments of Energy, Defense and National Science Foundation can provide additional help with respect to the technical needs of a research resource. As such, HPE recommends taking a “whole of government” approach, similar to what was done under the National Strategic Computing Initiative (NSCI). NSCI had lead agencies (DOD/DOE/NSF) and deployment agencies (NOAA, NASA, etc.) all operating within the NSCI umbrella. NAIRR could build from this model.

Repositories for AI models, metadata, data and other artifacts will be an important component for a national research source. In that respect, we believe that an open-source community-based approach is better way to proceed. NAIRR could consider a “git-for AI data,” modeled after the wildly popular GitHub which provides open source software development and version control. The ubiquitous availability of git server in most IT environments across the United States eases the deployment complexity of AI resources since it relies on already available services and does not create an additional administration burden to shared infrastructure services. For example, git-like simplicity allows AI meta-data to be tracked and managed separately in a decentralized usage model, easily cloned in any environment, following an open-source community-based approach that enables wide adoption and reuse. From there, we can develop technology to build trustworthy AI artifacts that meet desired tradeoffs between accuracy and uncertainty.

Shared Computing Infrastructure

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;
Creating and maintaining a shared computing infrastructure should leverage existing organizations that currently provide shared computing infrastructure to a wide community of students and researchers, including DoE computing user facilities and NSF computing centers. These centers provide many of the capabilities needed to support the NAIRR goals. However, because access to curated data sets is critically important but not all AI research needs large compute allocations, the NAIRR should supplement these with shared computing infrastructure that is designed for and operated with a view toward maximizing the number of projects that can use those datasets, and providing the educational tools and services needed to reach diverse communities of potential users, who may have no experience using existing shared computing infrastructure.

Features to emphasize:

- Open vs. proprietary technology in the resource
- Sufficient availability, scale, and capability of resources with a commitment to growth as needed
- Strategic approach that encourages market competitiveness and ongoing technological innovation without sacrificing inter-operability
- Seamless interfaces and processes for inter-platform multi-modal data exploration and interrogation
- Tools, queries, methods and algorithms to platforms and data

Assessment of Recommended Solutions

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;

Over the past two decades, the federal government has made significant progress toward increasing the priority placed on data management and curation within federal agencies. These efforts were significantly bolstered by the 2017 Foundations for Evidence-Based Policymaking Act. Despite this progress and these efforts, however, accessing high-quality government data sets remains challenging. One of the key reasons for this is the large number of government portals and initiatives that have been launched, often changing with each administration. This lack of a persistent, global data infrastructure often leads to confusion over where to find data sets, which data sets reflect the most current data and preparation process, and what data are best-suited to various AI applications.

While the NAIRR will need to address questions of ensuring data lineage and provenance when making data sets available, more importantly it must address the challenge of establishing and maintaining a persistent solution for making data sets accessible and available to researchers. This will include establishing capabilities within federal agencies to manage data availability and curation that persist across administrations and through budget cycles. Critically, however, this global solution must also be developed in close partnership with the research community, ideally leveraging non-federally owned or managed infrastructure and capabilities in non-profit entities such as universities or other research institutes. At the same time, the NAIRR should be careful to ensure that data infrastructure does not create a preference for one type of technology or limit competition from future technology and/or data solution providers. To ensure this, the NAIRR will need to reconsider the role of the US Government in managing its data sets, emphasizing the government’s role in convening stakeholders and driving toward shared best-practices while federating data infrastructure management.

Some of the challenges that NAIRR will need through address through broad stakeholder engagement include:
- **Data set discovery** – Making it easy to find data sets and guidance on which data sets are considered the "gold standard" in various disciplines
- **Sharing** – Who can access data and how?
- **Trust** – How can users know they are accessing the data they think they are accessing, understand their lineage, assess their "quality" (e.g. reproducibility etc.), and validate their provenance?
- **Security** – How can data sets be accessed and analyzed securely and in a way that maintains privacy?
- **Incentives** – Many critical data sets are not owned by the federal government. What incentives would help increase safe, trusted, and equitable data exchange?

**Assessment of Security Requirements**

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

The identification, classification and protection of dataset categories, and the categories of the communities that are likely to access them, are the key aspects of NAIRR security requirements.

There are three classifications of assets:

- **Strategic Datasets** – dataset that provide national advantage and security, this includes critical datasets and models and computational techniques (and computational infrastructure and resources)
- **Academic Datasets** - datasets and models that are useful and beneficial to the academic community
- **Community Datasets** - learning datasets and models that are useful for training but not a basis for standard or trusted knowledge datasets

Similarly, there are parallel communities that create and consume these assets:

- Researchers who are authorized to create, consume and derive works from the national advantage and security assets
- Researchers who can leverage and derive works from the academic resources
- Teachers and learners who can consume and leverage training datasets and models.

The boundaries between the three categories are strict and the transmission of data from one to the other must be managed with respect to confidentiality and integrity.

Key security principles to consider are:

- Multifactor authentication to identify and categorize users and manage their appropriate access to the categories of assets
- Protection of the integrity of models and datasets and their inputs
- Encryption of data in motion and data at rest
- Recordation of state changes to assets with accountability to the level of a single individual.

**Assessment of Privacy and Civil Rights and Civil Liberties**

- An assessment of privacy and civil rights and civil liberties requirements associated with the National Artificial Intelligence Research Resource and its research;
As it provides AI researchers and students across scientific fields with access to an advanced computing ecosystem and data infrastructure, the NAIRR has the opportunity and duty to expose researchers and students to holistic tools for assessing the ramifications of applying AI in their disciplines using legal and ethical frameworks which are authentic and inclusive to the populations they serve.

Across the US Federal Government and US-based industries and academic communities, a portfolio of tailored ethical frameworks has emerged that link organizational core concepts to AI ethics principles and then to operational commitments and specifications. The NAIRR should provide access to this expanding portfolio and guidance to researchers and students in selecting and tailoring their AI ethical and legal frameworks. Importantly, the NAIRR should seek to ensure the development of an AI culture in which ethical and legal considerations are foundational to every AI project rather than viewed as a “box-checking” compliance exercise.

Sustainment Plan through Federal Funding

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

Sustained, incremental funding for NAIRR will be critical to its success. Securing this funding will take leadership from the White House, and in particularly from Office of Management and Budget. But it will also be incumbent on individual agencies and programs to re-prioritize their funding proposals in a way that maximizes the use of new and existing research resources toward the broader goal of advancing our national AI capability in a way that continues to serve their core missions. For those agencies and programs already providing a significant amount of AI research resources for the benefit of government, industry and researchers, additional funding will likely still be needed to achieve the goals of the NAIRR. This is because the many programs that we are aware of are already highly oversubscribed and increasing demand will only further stymie progress.

Increasing funding for research resources incrementally will allow these agencies and programs to provide more resources to more users, building on their past experiences and best practices. Furthermore, the NAIRR should consider whether there are mechanisms for facilitating collaboration between grantees at different agencies that already fund AI or AI-enabled research. This kind of cross-disciplinary and cross-mission collaboration has the potential to rapidly advance progress in foundational AI technologies as well as in the way AI is used in a variety of applications by cross-pollinating ideas and enabling the sharing of best practices.

Lastly, HPE is a strong proponent of federally funded partnerships involving the private sector. We see value in having the private sector involved in nearly all aspects of a national AI research resource ecosystem.

Establishment and Sustainment Parameters for NAIRR

**I. Parameters for the establishment and sustainment of the National Artificial Intelligence Research Resource, including agency roles and responsibilities.**

To the extent possible, HPE believes that the NAIRR should leverage processes, infrastructure, and best-practices from existing federal agencies and facilities for managing user facilities. Several science agencies, including DOE, NSF, DoC/NIST, and DoC/NOAA operate shared user facilities with competitively allocated access for use by American scientists. While the closest analog to NAIRR will likely be found in
the computing facilities operated by NSF and DOE, the NAIRR should broadly consider the practices by which federally-funded user facilities (including non-computing facilities such as photon sources and observatories) are made available to develop suitable operating approaches. Most importantly, the NAIRR should prioritize approaches to resource allocation that ensure broad-based access to all Americans, with an emphasis on ensuring that smaller projects and projects proposed by less experienced researchers are not de-prioritized.

To achieve this vision of “democratized” access, the NAIRR will likely need to be a federated resource across several agencies. Within each agency, the NAIRR may prioritize supporting infrastructure and resource access that best relates to that agency’s mission. A similar pattern currently exists with respect to high-performance computing resources, which are made available to the public by DOE and NSF with each agency targeting a different set of researchers and outcomes with the OSTP-coordinated Networking and Information Technology Research and Development (NITRD) providing loose coordination. NAIRR, however, should seek to provide a more robust coordinating function than NITRD currently allows, especially with respect to ensuring sufficient resources are made available in the budget across each agency.

Prioritized Capabilities and Services

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

To broadly advance the use of AI, the education of an AI-savvy workforce, and research in AI, HPE believes that the NAIRR should prioritize making high-quality curated datasets available, with access to expertise to enable use of those datasets. Data ethics and privacy best-practices should be emphasized in the data curation process as well as in the use of that data.

Additionally, HPE believes that the NAIRR should offer enough resource capacity at acceptable capability, that grows over time. Because requirements and technologies are diverse and rapidly evolving, a strategy that allows market competitiveness and the provision of diverse technologies should be adopted.

Other Priorities for consideration:

- Data ethics and privacy best-practices sharing (included in data curation)
- Coordinated access policy across all resources (including across agencies)
- Forum for convening (and also studying the interaction)

Principles of Ethical and Responsible Research and Development

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?

Under the joint leadership of Hewlett Packard Labs and the HPE Chief Compliance and Privacy Offices, a pan-HPE team has recently completed the drafting of the HPE global AI Ethics Principles. That team is now engaged in the even more challenging phase of operationalization of those principles into commitments and specifications to guide our team members, customer and partners in the utilization of AI across our products, our processes and our partnerships worldwide. While there is a large portfolio of AI Ethics principles and frameworks which have emerged from government, industry and academia, we felt that we needed to tailor a framework which was derived from our company purpose, to advance the way that
people live and work, and was authentic to our roles in creating, supplying and consuming AI technologies. As we gain experience applying these principles in practice, we’re uncovering gaps where conventional AI applied to real world situations cannot be applied with confidence in meeting our AI ethics principles, revealing instead issues of Bias, Explainability, Trust and Robustness. It’s not just a matter of being more careful or deliberate with today’s state of the art, these are technology gaps and closing them will take engineering and ingenuity, and this forms the basis of our Hewlett Packard Labs research agenda. That team is developing novel techniques and approaches to Model Synthesis and Analysis, the Data Foundation underpinning ethically robust AI, and Hardware Acceleration that will enable explainable, robust AI to be operated equitably and sustainably.

By embracing a dual mission of providing both providing access to a holistic advanced computing ecosystem and data infrastructure as well as to training and resources to enable researchers and students to establish AI Ethics frameworks tailored to their communities and concerns, the NAIRR could both enable those teams to both demonstrate with transparency how current technology can be applied with confidence and illuminate where current technologies fall short, and innovation is required. As it is intentionally focused on traditionally underserved communities, this dual mission would provide not only access to technology but add the lived experience of those communities directly into the ethical discussions concerning their use.

Without mandating a particular ethical framework, engaging with an NAIRR focused on the dual mission of technology and ethical rigor could be established as a benchmark for due diligence in ethical application of AI technologies.

**Existing Building Blocks**

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

A successful deployment of the NAIRR will require several foundational components which are already in existence. These are important building blocks that the NAIRR can leverage to its advantage. We recognize that as governments and industry actively consider how to approach governing AI, standards will play a key role in establishing the connection between rules and practical implementation.

HPE is currently engaged with a few government agencies that we believe are well positioned to provide the critical building blocks for the NAIRR implementation. These include the DOE National Labs, NSF advanced computing centers, NASA, NOAA and a few other agencies that have high end computing and data infrastructure as well as a surrounding ecosystem of system, applications, and other technical support. This existing infrastructure with some incremental funding could be leveraged to provide AI resources for a broader community. It is significantly cheaper to leverage an existing ecosystem, rather than starting up new centers from scratch.

There will be other considerations for the successful implementation that we’ll identify for your consideration, as follows. Of critical importance are the data repositories. These will need to be federated, and to streamline access for users of the repository, may also need to be imported or securely linked into NAIRR resources. We believe some of the efforts by NIST to cultivate trustworthy AI systems, can also provide the complimentary foundational building block to the efforts the NAIRR Task Force with regards to secure data repository access.

We also recognize that there will be specific use cases where Public cloud providers will provide additional resources to supplement those directly engaged by the NAIRR. Often these can be leveraged for users just
getting started and may not have computing resources beyond the client devices. It is also important to build on existing arrangements with a proven track record like the NSF-funded CloudBank model, which provides individual researchers access to commercial clouds for NSF-funded research. Finally, we strongly recommend engaging with Industry consortia like the ML Commons and other standards bodies such as Open Geospatial Consortium, as well as private sector solutions especially with highly data intensive applications like earth observations.

Public-Private Partnerships Roles

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

- Standards bodies
- Training/tutorials

HPE believes public-private partnership is not just a mechanism to enable capacity or access to experts. Rather, HPE has a long history of developing new technologies, in partnership with our customers, to enable new capabilities. HPE believes that the NAIRR researchers will identify gaps in capabilities that public-private partnership, through NRE funding, can fill. Therefore, there should be provisions allowed for the development of these capabilities through the auspices of NAIRR.

HPE and our government partners aspire to leverage current and future advances in artificial intelligence and high performance computing to dramatically accelerate scientific discovery and expand the frontiers of scientific knowledge to benefit US national security, scientific leadership, and industrial competitiveness.

Shared Principles:

- Keep AI-related software open source as much as possible and prevent vendor lock-in
- Establish an inviting, standards based open infrastructure environment to foster AI hardware and software innovations from the greater community. Standards proposed will take into account the need to ensure system-wide functionality with competitive performance relative to proprietary solutions
- Hide complexity from the user in order to make high end AI accessible to a wider number and range of domain scientists
- Greatly improve the collection, curation, storage, movement and analysis of AI relevant data at scale, taking into account the expected greater volume, velocity and variety.
- Enable real-time and near real-time control/steering of experiments and scientific based operations through the use of AI connecting the system of systems
- Emphasize the development of high productivity software and tools that perform will at scale

Lab Collaboration:

HPE believes that lab and other key customer collaborations will be critical for an AI for Science at scale initiative to succeed. HPE’s efforts must be driven by advanced future use cases from customers in order to be relevant and attack the problems that matter most to our user community. Our large customers have extensive experience in advancing science and have access to ground breaking use cases, and connections to the researchers and the greater scientific community developing them.
Limitations to Democratize Access to AI R&D

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?

Efficiently executing AI workloads is a very computing and data intensive endeavor. In the last decade, the AI R&D community has aggressively moved to a specialized hardware and software infrastructure, primarily using GPU acceleration, to achieve the desired cost/performance target. Democratizing access to AI R&D requires an open approach to accessing this specialized infrastructure, and the tools (such as runtime software, programming environment, and libraries) that are required to operate it.

There are a few obstacles to making this happen that the NAIRR will have to consider. The AI accelerator market is experiencing a high degree of variability regarding openness in how accelerators are integrated into larger computing systems. In traditional accelerated systems, accelerators cards (such as GPUs, FPGAs, or AI-dedicated ASICs) are attached to standard servers through open interfaces (like PCIe) and multiple integrators can offer comparable systems. In newer, tightly integrated "AI appliances", entire systems are offered as proprietary black boxes and only accessible through network interfaces, thus limiting the end user choice and the integration points. A last, but not less important category of AI systems are those that are kept proprietary, not available in the open market, and only accessible through a specific interface. Among these are for example what is offered by cloud service providers with dedicated proprietary accelerators. In this case, the lack of openness raises to the software level, and while users are not impacted at the hardware integration level, they are constrained in the way in which they can access and orchestrate the accelerators, and where the data they process can live.

Overcoming these limitations requires a "community call to action" that the NAIRR could sponsor, to preserve a healthy and open ecosystem that offers AI R&D users choice at all levels. While proprietary innovation is fundamental for the AI ecosystem to thrive as a business, there are important dimensions that can benefit from an open approach, such as storage, networking, virtualization, runtime, workflow, scaling and security. For example, the NAIRR could sponsor a set of workshops or other activities with the objective of producing open recommendations (possibly standards) and a reference architecture, for future accelerated systems. By provide an open reference architecture to everyone as part of the requirement of future systems, the NAIRR could help create a blueprint for an open AI R&D infrastructure that can motivate all players to participate.

One last important challenge to the democratization of AI R&D is the lack of established approaches the combine AI and scientific computation. The main development of the AI frameworks (such as PyTorch or TensorFlow) is happening outside of the scientific community, driven by different players and incentives. The NAIRR could play a vital role to act as a catalyst for the scientific community to help define, design, and publicize desirable AI and HPC integration models, such as AI by the side, outside-the-loop, and inside-the-loop of HPC codes. Publishing open benchmarking practices and measurements for AI in science is another important set of objectives that can help AI R&D to establish common ground and be able to compare alternatives. Finally, working with the open-source community so that lower level of accelerator software can be penetrated by open-source compilers, optimizers, and programming tool chains.

Democratize access by allocating a portion of the resources to be made available through a light-weight request process, to enable researchers and students without a history of using large centralized HPC resources to start using NAIRR resources without substantial barriers.