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

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.

October 1, 2021

Wendy Wigen
National Coordination Office
Networking and Information Technology Research and Development
2415 Eisenhower Avenue
Alexandria, VA 22314

RE: IBM response, RFI on the National AI Research Resource [86 FR 39081]

Dear Ms. Wigen,
IBM appreciates the opportunity to comment on the Request for Information (RFI) on an *Implementation Plan for a National Artificial Intelligence Research Resource*. IBM strongly supports its development.

IBM is an Artificial Intelligence (AI) and hybrid cloud technology leader and is engaged in research and development across a broad set of scientific and industry domains. IBM has extensive experience developing advanced computing for scientific research, including Summit, a 200-petaflop supercomputer built for Oak Ridge National Laboratory.[1] IBM co-created the COVID-19 High Performance Computing Consortium, where 43 members have carried out more than 100 projects.[2] IBM is also a leading provider of open source and open hybrid cloud architectures and technologies that simplify the integration of heterogeneous multi-cloud environments.[3]

If AI is to deliver on its full promise in advancing health, security and economic prosperity, democratization of its development by the research community and increasing the accessibility of both advanced computing and data will be key. Accordingly, the NAIRR must include the following core components:

1. Federated, hybrid cloud enabled computing resource – an accessible and easy-to-use hybrid- and multi-cloud computing resource built on open architecture that amalgamates various public clouds (such as Amazon, Azure), private clouds, and on-premises resources to create a single, unified, flexible compute infrastructure.[4]
2. Data and models – large scale, high-quality, trusted, AI-ready datasets and pre-trained AI models across the broad AI science and technology landscape.
3. Software and tools – integrated and interoperable software and platform technologies that support AI research and development and enable those with varying degrees of technology and science expertise to be productive.
4. Education – training materials, outreach activities, and user support that ensures easy, efficient, and effective use of the NAIRR.

If designed correctly, the NAIRR would be a pervasive, easily accessible federation of advanced computing resources, combined with a shared data infrastructure, that would bolster American leadership in AI research. Once again, IBM appreciates the opportunity to comment, and we look forward to future engagements. For any questions, please contact Mr. Jeffrey Brown, Science & Technology Policy Executive at jeffrey.brown@ibm.com.

Sincerely,

Dr. Dario Gil
Senior Vice President and Director of IBM Research

[1] "Summit Supercomputer Ranked Fastest Computer in the World." *United States Department of Energy*, June 25, 2018. https://www.energy.gov/articles/summit-supercomputer-ranked-fastest-computer-world.
[2] "The COVID-19 High Performance Computing Consortium." https://covid19-hpc-consortium.org/who-we-are.
[3] "Hybrid Cloud Solutions." *IBM Cloud*. https://www.ibm.com/cloud/hybrid.
[4] "What is Hybrid Cloud?" *IBM*. https://www.ibm.com/cloud/learn/hybrid-cloud.

IBM Response to the Request for Information on an Implementation Plan for a National Artificial Intelligence Research Resource
White House Office of Science and Technology Policy (OSTP)
National Science Foundation (NSF)

1. What options should the Task Force consider for any of roadmap elements and why?

A. Goals for establishment and sustainment of a NAIRR and metrics for success;

To guide the development and sustainment of the NAIRR over an extended time horizon, IBM supports metrics that continuously track progress throughout the lifecycle of the computing resource, including its technical specification, development, testing, rollout, use, maintenance, and eventual upgrade. Similarly, metrics should be established to track the development, promulgation, and use of data resources across the complete lifecycle, including support for data provenance and versioning, and comprehensive management of metadata, classifications and policies. To overcome both compute *and* data constraints to democratize AI research and development, and therefore catalyze innovation, IBM proposes three categories of metrics:

1. Researcher usage and productivity: Initially, metrics should focus on the construction and rollout of the resource, including consumption of open hybrid cloud technologies and the availability of high-quality datasets and their use. As the NAIRR matures, metrics should shift to assess the overall impact of compute and data, including replicas of provided data sets, on researcher productivity. Specific metrics could include the number of institutions using the resource, the increase in scale of research (including amounts of data and compute employed), time taken to conduct and reproduce experiments, the number of research publications facilitated using NAIRR access, and shifts in the citation scores of researchers using the resource.

2. Community development: Modern scientific discovery demands the reproducibility of results, the creation of tools, standards or benchmarks, collaboration, and the effective communication of knowledge. Cultivating communities of discovery will be key to the NAIRR's adoption and spread. As the technical infrastructure is developed, parallel efforts should be undertaken to build a community of users whose actions and behaviors contribute to a virtuous cycle of increasing the mainstreaming and usage of the resource. Metrics could include increased participation in workshops related to the use of hybrid cloud for science or R&D, the development of tools for the community, the creation of new benchmarks adopted by the community, and the number of experiments shared within the community.[5]

3. Wider impacts: The NAIRR should result in long-term economic and job growth that bolsters the United States' global competitiveness. After the establishment of its infrastructure and community, metrics should be defined to assess breakthrough technologies and novel services unlocked by NAIRR. These metrics could include increases in productivity and automation, new products and startups created, and the integration of AI by large companies. Such economic benefits will inevitably have workforce impacts, and the resource should measure skills and jobs shifts spurred by its work. Finally, metrics should be developed to determine how the NAIRR has influenced international competitiveness.

[5] A key element of many academic papers is comparison of the new approach to existing approaches, for instance, speed of accessing a data set. Having a standard yardstick to use for this comparison, which is used by many researchers, speeds the advancement of science. Thus, having the NAIRR foster such new benchmarks will contribute to speeding the advancement of science.

B. A plan for ownership and administration of the NAIRR, including:

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

Deep, collaborative partnerships between government, industry, and universities have played a critical role in spurring innovation in the United States for decades. Federal investments in basic and translational research have enabled universities and federal partners to engage in high-risk, high-reward research activities, which have resulted in transformative ideas and commercialization that otherwise would not have happened.

While housing the NAIRR within a single government agency may streamline coordination and procurement negotiations, such a do-it-all approach could also impose short-term constraints regarding data sharing and impede the ability of the resource to grow flexibly over time. For example, a single agency approach could inhibit the pooling of data from a wider set of government agencies, limiting the inclusion of a broad range of diverse data sets as a critical ingredient for AI research and development.

To achieve success, the NAIRR needs to be endowed with flexibility that permits coordination with a diversity of stakeholders both inside and outside of government. And it needs to be built to sustain and evolve over the long run. IBM proposes a federated approach for implementation, deployment, and administration. A federated approach would allow sustained innovation and diversity in the constituent computing resources. For example, a federated approach would allow the NAIRR to take full advantage capabilities provided by the fast-advancing field of AI accelerator hardware and heterogeneous computing systems. A federated, virtualized approach could also better enable seamless access to computing and data infrastructure that allows researchers and scientists to participate in the procurement and deployment of the resource components themselves. Under an open hybrid- and multi-cloud model, the federal government should define how providers make participating computing and data resources available, for example, by documenting open hybrid cloud architectures, application programming interfaces, and standards for data and metadata representation.

ii. A governance structure for NAIRR, including oversight and decision-making authorities.

Given the complex nature of the task assigned to the NAIRR, IBM recommends that it be incubated as a Federally Funded Research and Development Center (FFRDC). The FFRDC model has been effective for similar roles.[6] The NSF, for example, acts as an administrator for five FFRDCs, such as the Science and Technology Policy Institute.[7] FFRDCs further government research, technology development, systems acquisition, and policy guidance. The FFRDC model supports a "special long-term research or development need which cannot be met as effectively by existing in-house or contractor resources." Furthermore, FFRDCs have "access, beyond that which is common to the normal contractual relationship, to government and supplier data, including sensitive and proprietary data, and to employees and installations equipment and real property."[8]

An FFRDC governance structure should be sponsored and empowered by multiple agencies, including the NSF, the Department of Energy (DOE), the National Institutes of Health (NIH), as well as other scientific agencies, with strategic input from OSTP. Since FFRDCs marshal the expertise of government, industry, and academia to solve complex problems, the model is capable of meeting needs that cannot be met solely using government resources and contractors. Most importantly, the FFRDC model would allow multiple stakeholders (and government

[6] "Federally- Funded Research and Development Centers (FFRDC's): Background and Issues for Congress." *Congressional Research Service*, April 3, 2020. https://crsreports.congress.gov/product/pdf/R/R44629/6.
[7] "Master List of Federally Funded Research and Development Centers, by Agency and Type of Administration." *National Science Foundation*. https://www.nsf.gov/statistics/fedfunds/pubs/ffrdc/ffrdc.htm.
[8] Ibid.

agencies) to rally around its shared goal. And an FFRDC would ensure interoperability for communities of discovery to work collaboratively to solve common challenges.

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

Ideally, a NAIRR FFRDC would define interoperability requirements including open interfaces for federation of the resource, and allow industry to compete in the development and fielding of leading-edge computing and data resources, as well as AI software, tools and platforms. Flexibility in procurement and use by researchers and "plug and play" by providers would require some degree of interoperability and standardization. However, given the wide diversity user needs and the fast-paced nature of computing and AI developments, care should be taken to avoid homogenization of the resource. A diversity of integrated computing resources would allow researchers to leverage resources across multiple clouds, utilize novel computing devices or hardware accelerators, and take advantage of supercomputers and high-performance clusters to meet their objectives.

To simplify procurement – which has been the Achilles' heel of the FFRDC model – the NAIRR should learn from the NIH STRIDES program, which has greatly simplified cloud resource procurement for its researchers. But the FFRDC for the NAIRR should go further. The FFRDC should specify interoperable architectures and interfaces that allow diverse computing resources to be integrated into the federation. The FFRDC should employ open hybrid cloud mechanisms for managing the federation of multiple clouds (such as Amazon, Azure) and computing resources to allow seamless workflows and workloads to operate across the full diversity of provided resources. The FFRDC should also facilitate the integration of data management and AI software, tools, and platforms into the NAIRR to further aid and increase productivity of AI researchers.

To effectively manage data, users, and the provisioning of compute resources, governance of the compute and data resources should be automated where possible. To this end, it is important to bake in governance mechanisms that enable the automation of the management and the enforcement of policies, FAIR guidelines, regulations, and best practices. IBM suggests a "Software-defined Governance Framework" that can be easily configured and extended while ensuring consistency, transparency, and auditability through the lineage enabled via automation.

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;

Data Sets/Secure Access Control: The NAIRR should adhere to the FAIR guiding principles for scientific data management and stewardship. FAIR provides guidelines to improve the findability, accessibility, interoperability, and reuse of digital assets. In line with FAIR, data made available through the NAIRR should be easy to find and read for both humans and computers. It should be machine readable to enable the automatic discovery of data sets and services.

Once a researcher has identified data sets, the NAIRR should provide direction on how the data can be accessed, so that the resource maintains overall control of security and authentication protocols. To the extent possible, enforcement of the security and governance requirements should be automated using available technologies.

Data acquired through the resource should be interoperable with applications or workflows for analysis, storage, and processing. Crucially, to allow for the reproducibility of experiments, the NAIRR should require interoperability so that workloads can run across diverse data and cloud environments. Further, data should be well-described so that it can be replaced or combined in different research experiments. And the NAIRR should require the replicability and portability of the data brought onto the resource. Replicability and portability need to be enshrined to fend off "data gravity," a phenomenon in which applications, computing, and users gravitate to a sole provider.[9] Requiring interoperability, replicability and portability, in turn, ensures democratization of the resource and prevents overreliance on or lock-in by a single provider. The NAIRR will also need to provide tools and frameworks to enable moving and replicating data to enable computing over geographically distributed data sets.

Compute Resources/Scalability: To stand up and scale the NAIRR quickly, it should leverage commercially available compute resources. Commercial providers offer not only raw computing power, but sophisticated software stacks and user interfaces that have already been widely adopted by the AI research community. This would build on and broaden existing models such as NSF's CloudBank.

Crucially, the NAIRR should adopt an open hybrid- and multi-cloud architecture, which is capable of scaling and delivering compute resources in a cost-effective manner. In short, the goal of hybrid- and multi-cloud is to establish a mix of public (e.g., Amazon and Azure) and private compute resources — and orchestration between them — that would give users flexibility to select the optimal resources for each application or workload and to move workloads freely as circumstances change. This may require making transitions between hybrid- and multi-cloud environments more seamless, for example, by employing a control plane that enables automation for optimal use of hybrid- and multi-cloud resources. Such interoperability would unlock new possibilities for researchers, including computational work that requires integration of data movement and computing across diverse locations.

But, just as the NAIRR leans on commercial providers to build the resource, it should also seek out inclusive frameworks to allow for the integration of existing public clouds. Multiple existing public clouds could be easily integrated into the NAIRR using an open hybrid cloud architecture. Advanced computing, accessible in hybrid- and multi-cloud usage models, will play a pivotal role in accelerating scientific discovery.

IBM has developed a hybrid- and multi-cloud approach for science that addresses key NAIRR requirements:

- *Heterogeneity* supports highly diverse resources including multiple public clouds (e.g., Amazon, Azure), on-premise infrastructure, and edge devices, which may include scientific instruments, sensors, physical devices, and entire labs and research organizations.
- *Consistency* enables portability, interoperability, and ease of management across heterogeneous environments.
- *Reproducibility* enables replication of scientific experiments and results regardless of differences in IT infrastructures or location of data and resources through the adoption of open standards and technologies.
- *Gravity* refers to the strong pull from extremely large data sets – some at a petabyte scale – as well as proximity required due to physical manifestation of experiments and instruments. IBM's open hybrid cloud approach to data aims to optimize workflow deployment consistent with data gravity constraints and to avoid lock-in with specific storage environments.
- *Openness* refers to the prominence of open science practices that may dictate open – or *FAIR* – (findability, accessibility, interoperability, and reusability) data access, including encoding these principles into an open architecture for hybrid cloud through advanced technologies.

Such an approach, underpinned by advanced (and evolving) technologies, would make efficient use of taxpayer dollars, and would boost the usability of the resource. For example, adopting an open hybrid- and multi-cloud framework would allow a university researcher to connect their research experiment to the NAIRR. And government agencies could port their existing infrastructure to the NAIRR using the same hybrid cloud framework. This would make it easier for researchers using the NAIRR to run experiments that span cloud environments and multiple compute resources.

[9] "What is Data Gravity?" *Data Centre Magazine*, November 2, 2020. https://datacentremagazine.com/technology-and-ai/what-data-gravity.

Educational Tools & Services/Resident Expertise: While some researchers are fluent in using tools such as hybrid cloud and AI to conduct research, many are not. Individual researchers, government entities, and cloud providers could benefit from engaging in technical exchanges, sharing best practices, and discussing challenges and opportunities of hybrid cloud for scientific research.

Educational tools and services and resident expertise must be built, including peer-to-peer knowledge sharing across different scientific communities to boost productivity across the board. To further this goal, the NAIRR should establish mechanisms for broad community sharing of best practices through annual conferences or other "birds of a feature" events. The NAIRR should also identify several important areas of shared interest across these communities and foster establishment of accessible testbeds that integrate computing resources, data, experiments, and evaluations for the specific AI application areas.

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;

For data sets to be ready for AI, they need to be high quality and trusted, and they should be representative, transparent, and sufficient for training AI models employed in AI research. Additional challenges include privacy, dissemination, maintenance and automated governance and security policies. Given its ambition, the NAIRR faces numerous – but surmountable – challenges, including:

1. Privacy, Protection, and Anonymization: Amassing data from multiple government agencies to build and train AI models at scale faces overarching constraints imposed by the Privacy Act of 1974,[10] which places limits on data disclosure and sharing. The Act granted certain data privacy waivers in the form of a research exception, which should be explicitly extended to cover the NAIRR and its work. To maintain compliance with the Act, government data gathered as part of the NAIRR should be stripped of personally identifiable information (PII). The NAIRR should consider utilizing mechanisms to automate the anonymization and removal of PII, and it should consider going one step further by ensuring that differential privacy is built into the resource. Furthermore, the NAIRR will need to apply security and compliance standards so that data can only be used for approved use cases. Enforcement of data privacy requirements should be automated to the extent possible using privacy-preserving computing techniques, such as secure multiparty and fully homomorphic encryption.
2. Intellectual Property (IP): Datasets are often viewed as an IP asset, which could constrain data sharing between government, industry, and universities. IBM recommends that the NAIRR adopt protocols outlined in the Patent and Trademark Law Amendments Act of 1980, which would grant FFRDC partners IP rights to innovation produced from the NAIRR.[11]
3. Orchestration & Automation: Given the volume of datasets that will be brought into the NAIRR, users will face the challenge of finding and using the right data for the right task. IBM proposes the adoption of a hybrid data fabric that takes a holistic view of the data lifecycle as it is created and used in federated distributed environments. A hybrid data fabric provides next-generation orchestration of secure and efficient data management. The hybrid data fabric is composed of a data plane and a control plane, which interacts with and manages data centric workloads in a brownfield, multi-vendor, multi-location environment.[12]
4. Dissemination: The NAIRR will need to not only curate data sets at an enormous scale and make them available via appropriate security and governance frameworks, it will also need to employ AI models – such as natural language processing – to solve problems in an open and community-driven way. For example, the HuggingFace Big Science effort uses large language models to solve problems in such a way.[13] HuggingFace shows how a coordinated community effort supported by substantial computing resources (>5 million GPU

[10] "Privacy Act of 1974." *United States Department of Justice*. https://www.justice.gov/opcl/privacy-act-1974.

[11] "H.R.6933 – An act to amend the patent and trademark laws." *Congress.gov*. https://www.congress.gov/bill/96th-congress/house-bill/6933.

[12] "Weaving data fabric into hybrid multicloud." IBM Institute for Business Value. https://www.ibm.com/thought-leadership/institute-business-value/report/data-fabric-multicloud

[13] "A one-year long research workshop on large multilingual models and datasets." *Big Science Hugging Face*. https://bigscience.huggingface.co

hours) can take on difficult challenges like creating trusted open AI-ready data and large-scale training for language models.

5. Data Scalability: As the number of data sets on the platform grows and as data is combined, data becomes harder to discover and use. Therefore, the automated tagging and labeling of datasets with metadata can be important for their discoverability and findability by users. The NAIRR should consider supporting methods for automated metadata generation that analyze data sets and generate metadata tags that provide semantic information about the content of the data such as domain-specific terms and entities.[14]
6. AI Model Trust and Transparency: The NAIRR should incorporate mechanisms such as AI Fact Sheets that capture information throughout the data and AI model lifecycle that is critical for trust and transparency.[15]
7. Self-Service User Access: To reduce the time to value and enable scaling, researchers must be able to find the data they need. Building upon points 3 (orchestration) and 5 (scalability), a logically centralized catalog supporting semantic search and the enforcement of governance and access control will be needed. Such a tool will enable users to find specific data sets and related data sets, see the linage and versions of the data they need, and make data sets visible only to authorized users.

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

The NAIRR must be secure and resilient. Careful steps must be taken to ensure the confidentiality, integrity, and availability of both the computing resources and data provided by the NAIRR, and the computational workloads and derivative artifacts, such as intermediate data, results, models, and so on, from external attackers, administrators and providers of NAIRR, and other researchers. The number of constraints and requirements are too long to list in this RFI response (access control, encryption, authentication, logging, etc.), but we provide an outline below.

The NAIRR should be architected to be compliant with applicable standards commensurate with the classification of the data and workloads to be executed on it. For example, the base platform should ensure baseline compliance with NIST 800-53, provide the means of implementing the necessary controls, the ability to prove compliance, and allow auditors, developers, and users of NAIRR to verify compliance.[16] Additionally, the platform must ensure compliance with additional regulations as necessary, such as HIPAA, FERPA, PCI DSS, and FedRAMP. Due to the open hybrid- and multi-cloud architecture of NAIRR, data and compute resources will be required to attest to their security compliance and integrity prior to admission, and workloads must be restricted to data and compute resources that are allowed by their classifications. For example, healthcare data cannot be processed on a system that is not HIPAA compliant. The resource needs to address unique security and privacy requirements, including:

1. Multi-stakeholders and multiple administrative domains with a shared responsibility for security.
2. Multi-tenant: Researchers represent different organizations and administration domains, and NAIRR must ensure isolation and separation of data and compute workloads.
3. Encryption: Data should be encrypted while at rest and in motion. The NAIRR will maintain full control over security keys and hardware security modules. Data may be encrypted using client-owned keys.
4. Confidentiality of data: Processed through secure enclaves and secure virtual machines, for example.
5. Federated identity, federated authorization, and access management. The platform should support the integration of policy-based data governance.

[14] "IBM expands data and AI excellence with data catalogue technology in Cloud Pak for Data." *IBM*, April 21, 2020. https://www.ibm.com/blogs/journey-to-ai/2020/04/ibm-expands-data-and-ai-excellence-with-data-cataloging-technology-in-cloud-pak-for-data/.

[15] "How IBM is advancing AI governance to help clients build trust and transparency." *IBM*, December 9, 2020. https://www.ibm.com/blogs/watson/2020/12/how-ibm-is-advancing-ai-governance-to-help-clients-build-trust-and-transparency/.

[16] "Security and Privacy Controls for Information Systems and Organizations." *NIST Computer Security Resource Center*. September 2020. https://csrc.nist.gov/publications/detail/sp/800-53/rev-5/final.

6. Mixed access controls: a robust mechanism with a mix of access control models will allow for data sharing while maintaining security and privacy.

G. An assessment of privacy and civil rights and civil liberties requirements associated with the National Artificial Intelligence Research Resource and its research;

The NAIRR should adopt principles for trustworthy AI, including fairness, explainability and transparency, as they are developed by NIST as part of its AI Risk Management Framework.

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

- An open hybrid- and multi-cloud platform which can provide a consistent, seamless user experience across various public clouds (e.g., Amazon and Azure), private clouds, and on-premise resources.
- AI and data management software and workflow tools that simplify the use of cloud for science and R&D, including reducing the barrier to entry for researchers to adopt the cloud.
- Open standards and open technologies which accelerate interoperability, the adoption of cloud, and reproducibility for science and R&D use cases.
- Consistent methods and tools that automate data security and governance as well as identity and access management that can be used consistently across hybrid and multi-cloud platforms.

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?

Building trust in the NAIRR is paramount. The ultimate beneficiaries of the resource – researchers and any member of the public benefiting from a research breakthrough – should be assured that findings have been reached in a manner that is ethical, responsible, and free of bias.

IBM has long adhered to Principles for Trust and Transparency to ensure that new technologies are developed and deployed in a transparent and explainable manner. And IBM has defined a risk-based AI governance policy framework, Precision Regulation that calls for fairness and security validated by testing for bias before AI is deployed and re-tested as appropriate throughout its use, especially in automated determinations and high-risk applications. It also suggests the designation of a lead AI ethics official, a model that the NAIRR should consider as it develops and scales a shared computing infrastructure.

To support bias mitigation strategies, NAIRR should be proactive in creating and implementing AI ethics principles and practices, and ensure appropriate governance is in place to provide ongoing review and oversight of the research resource. Examples of tools to support bias mitigation and the trustworthy use of AI include the AI Fairness 360 toolkit, AI FactSheets, IBM Watson OpenScale, and IBM Watson capabilities designed to help businesses build trustworthy AI. Government, industry, and researchers will have shared responsibility to ensure that AI systems used as part of the research resource are tested and assessed for bias.

Further, the NAIRR must ensure that researchers from minority-serving institutions and those with limited research budgets are guide the science, design, and development and application of AI. While the AI field today may not reflect the demographics of our society, moving toward a more diverse AI research ecosystem could help to avoid and mitigate unwanted AI bias by including and reflecting the interests and values of diverse communities. The NAIRR should also support training that increases understanding, mitigation and recognition of bias and how it could be unintentionally introduced into AI systems during the development pipeline.

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

Academia and government have long pushed the boundaries of advanced computing — and have generated some of the most complex data and computationally intensive workloads along the way. At the same time, the private sector is leading the adoption of hybrid cloud, driven by software application modernization and digital transformation. To maximize the effectiveness of NAIRR, and to ultimately boost the productivity of the research community, it should combine building blocks from existing collaborative efforts to boost research cooperation, as well as leading-edge advances in the private sector.

First, public initiatives — such as the European Open Science Cloud, NIH STRIDES, and NSF Cloudbank — have used the cloud to connect researchers to datasets, software, and processing. Furthermore, public efforts have been made to cultivate communities of discovery that unite to create solutions to large-scale problems. For example, the COVID-19 High Performance Computing Consortium brings together academia, the federal government, and industry to provide access to some of the world's most powerful high-performance computing resources in support of COVID-19 research. And the European Commission's Helix Neubla Science Cloud has piloted a hybrid cloud platform for research, which feeds into the European Open Science Cloud that will create a shared research space for 1.8 million researchers.

These communities will capture the next dominant workflows and workloads for accelerated discovery and will drive a robust supply chain for innovation and value creation and achieve a scale of impact that is critical for society. Existing collaborative efforts should be boosted by private sector efforts such as IBM's OpenShift, which allows for building and moving of workloads to boost responsiveness, scalability, and price.

In addition, there are many enabling technologies, most of which are based on open source, that help address some of the pain points that researchers have experienced under the above programs in their early stages of adoption of cloud. For example, hybrid- and multi-cloud platforms like OpenShift based on Kubernetes, and emerging technologies like Ray[17] and IBM's CodeFlare[18] that further simplify the user experience for AI and data science on cloud can not only reduce the barrier to entry and enhance productivity for users but can also simplify the management and operations for providers. Emerging open technologies that help automate data security and governance, for example Fybrik[19] or modular encryption, can be used to ensure that data is used and accessed according to policies defined by NAIRR.

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

IBM recommends that the research resource be incubated using an FFRDC governance model. However, as the NAIRR evolves, it should consider leveraging public-private partnerships to accelerate its work. For example, in March 2020, OSTP initiated the public-private COVID-19 High-Performance Computing Consortium, a public-private partnership to marshal computing power to develop responses to the COVID-19 pandemic. In March 2021, IBM and the Cleveland Clinic announced a ten-year partnership to advance and apply open hybrid cloud architectures, AI, high-performance and quantum computing to accelerate health care and life sciences research.

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?

[17] "Fast and Simple Distributed Computing." RAY.io. https://ray.io/.

[18] "CodeFlare drastically reduces time to set up, run, scale machine-learning tests." IBM Research, July 7, 2021. https://www.research.ibm.com/blog/codeflare-ml-experiments.

[19] "A native platform to control data usage." Fybrik.io. https://fybrik.io/.

The NAIRR could face several roadblocks in its ability to democratize access to AI R&D:

Vendor lock-in: One of the greatest hurdles that researchers face as they begin to adopt the cloud is the difficulty of using more than one cloud platform for R&D. Researchers from NIH and other government agencies in particular point to the tremendous overhead of trying to operate across more than one cloud when it comes to secure data access and computing. We believe this problem can be solved through requiring interoperability through open standards and technologies that allow users to work more seamlessly across heterogeneous cloud environments to reduce this burden on the part of the user. For example, if data security, governance, and access management challenges could be made interoperable, regardless of cloud provider, it would allow researchers to become considerably more productive when moving from one platform to another, e.g. because of the availability of data in a particular location. In addition, over time we expect to see more and more development of algorithms that enable federated computation over geographically disjointed data sets. The NAIRR thus needs to support extensibility of the algorithms and frameworks it supports to break down barriers and fend off data gravity.

Inequity of funding: The high cost of compute resources and well-curated datasets means that researchers at large, well-funded universities benefit from access to these compute resources and well-curated datasets. However, smaller universities with finite research budgets often struggle to access – and make use of – these same resources. Lacking access to reasonably priced compute or data at scale will crimp the ability of the research resource to accelerate research and scientific discovery. While the NAIRR will theoretically increase access to compute and data, it must also be augmented by an increase in research and development spending that targets under-served communities. By moving to an open hybrid- and multi-cloud model that enables the integration of resources across multiple providers and on-premise IT, resources may become available at different price points but with the same user experience.

Human capital and user support: Boosting access to affordable compute and data is a major building block in democratizing access to AI R&D. But once the infrastructure of the NAIRR is in place, researchers will face the challenge of developing the skills and competencies needed to use the resource to accelerate their research. Closing the research gap with large, well-endowed institutions necessitates the creation of user resources and support communities that build human capital to operate and apply the capabilities of the research resource. To overcome these barriers to adoption and application, IBM proposes the creation of a research resource skills academy, along the lines of the IBM Quantum Education and Research Initiative, which partners will 12 historically black colleges and universities to develop education, community resources, and technical communities to power the field of quantum computing.