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

Response to Request for Information on an Implementation Plan for National Artificial Intelligence Research Resource

John T. Feddema, David J. Stracuzzi, James R. Stewart[1]

Sandia National Laboratories[2]

Sept. 1, 2021

Introduction

This paper is in response to the request for information (RFI) on an Implementation Plan for a National Artificial Intelligence Research Resource (NAIRR) that would provide AI researchers and students across scientific disciplines with access to computational resources, high-quality data, educational tools, and user support. Sandia National Laboratories is supportive of the NAIRR and would like to participate in the direction and use of this resource. Below is a discussion of the questions posed in the RFI. The specific question addressed in the RFI is indicated in brackets.

Goals and Metrics [Question 1.A]

With recent improvements in specialized computing hardware (graphical processor units (GPU's) and Tensor Processing Unit (TPU's)) and Machine Learning (ML) algorithms, Artificial Intelligence (AI) promises to improve U.S. competitiveness, quality of life, and national security. A new NAIRR would accelerate technological advances in AI and enable its use in a broad span of applications. We envision that the NAIRR will develop fundamental capabilities that U.S. industry as well as many U.S. government agencies can build upon to meet customer and public needs.

The goals of the NAIRR should include:

1. Improve U.S. competitiveness
2. Improve U.S. quality of life
3. Improve U.S. Government access to AI

[1] Contact Information: John T. Feddema, [redacted] David J. Stracuzzi, [redacted] James R. Stewart, [redacted]

[2] Sandia National Laboratories is a multimission laboratory managed and operated by National Technology & Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525. SAND2021-10794 O

4. Dramatically increase the number of U.S. universities performing AI research including the underserved.

Metrics that should be measured to evaluate success of NAIRR include:

1. The number of industry, university, and government partners who become members of the consortium
2. Total number of government and consumer products enabled by the consortium
3. The number of technical advances in AI research as measured by publications, open source and copyrighted software, and patents.
4. Volume and diversity of data accessible through NAIRR
5. Capacity and usage of computing resources available through NAIRR

Ownership and Administration [Question 1.B.i]

Advances in AI will have broad impact across many government agencies, including the Department of Energy (DOE), Department of Defense (DoD), Defense Advanced Research Projects Agency (DARPA), Department of Transportation (DOT), and Department of Commerce (DOC). At the DOE Office of Science, AI will enable improved scientific discovery in the areas of light water reactors, carbon capture, combustion research, bioenergy, energy storage, power grid, and advanced batteries. At the National Nuclear Security Agency (NNSA) and the three NNSA National Laboratories, AI will assist in the design, production, and surveillance of weapons systems; aid in the detection of weapons of mass destruction (WMD) proliferation; and protect US assets against cyber threats.

Because of the broad impact of advances in AI, we believe a Public Private Partnership and/or Consortium including multiple government agencies, universities, and industry should be considered. This consortium should be similar to the DOE Public-Private Consortia[3] or FDA's Scientific Public Private Partnerships and Consortia[4]. The consortia would enable joint foundational research while leaving participants free to build on the shared information and software to create proprietary outcomes of value to commercial, public, and national security applications.

Since multiple government agencies could benefit from advances in AI, joint ownership by National Science Foundation (NSF), DOE, DoD, DOT, DARPA, and other government is desirable. If a single lead government agency is necessary, NSF should be considered since it aligns well with their mission to promote the progress of science and provide global leadership in research

[3] https://www.energy.gov/technologytransitions/downloads/doe-public-private-consortia

[4] https://www.fda.gov/drugs/science-and-research-drugs/scientific-public-private-partnerships-and-consortia

and education although DOE also has a proven track-record of leading large public-private consortia and the DOE governance model has attributes worthy of emulation.

Governance [Questions 1.B.ii and 1.C]

The governance structure could leverage the existing public private partnership governance structure from DOE. DOE proactively manages their consortia to maximize their impact and benefits and ensure a successful transition to industry. The DOE consortium management plan defines leadership and decision-making structures, methods for communicating among members, and a sustainability model. Roles, responsibilities, authorities, and accountabilities of key personnel are defined. The management plan enables members to plan R&D activities, share progress and discoveries, and evolve a roadmap to address anticipated needs. [Question 1.B.ii]

The governance board should include a subset of members from NSF, DOE, DoD, universities, and industry. The governance structure would include subcommittees with the necessary expertise to facilitate decision making. This would include decisions around membership, data standards (e.g., formats and privacy), computing ecosystem, prioritization of computing access, etc. It would be expected that the subcommittees be managed through term limits with new members being elected by the NAIRR membership or nominated by the governance board. Further, the NAIRR governance structure would be independently reviewed on a regular basis (perhaps once every three years). [Question 1.C]

Infrastructure [Questions 1.D – 1.G]

A commercial cloud infrastructure would be an excellent choice for computing resources. Grants from government consortium members, e.g., NSF, DOE, DoD, DARPA, would sponsor research performed on the cloud. Industry would join the consortium by either paying a consortium fee or providing commercial cloud infrastructure.

Note that while there are several well-established codebases for machine learning that cover core capabilities such as classification, regression, and clustering (unsupervised), less common ML algorithms and a majority of AI algorithms are available only as research code and may be unsuitable for non-experts. The infrastructure should allow individual researchers to customize specific algorithms in the shared codebase (locally) to support a broader array of applications than possible with standard implementations.

A separate piece of needed infrastructure relates to training material. If the expectation is for practitioners – people not specifically trained in AI and ML – to apply these algorithms to datasets relevant to their primary research areas, then they will likely require training beyond code documentation to support their efforts. AI and ML algorithms are typically not plug-and-

play. Algorithms must be selected carefully, and data must be appropriately prepared. [Question 1.D]

As consortium members, DOE and its National Laboratories might provide unclassified datasets in the areas of light water reactors, carbon capture, combustion research, bioenergy, energy storage, power grid, and advanced batteries. In many cases, these could be broadly shared among consortium members.

With respect to the training requirement noted above, a key barrier to use of data and compute resources may be wasteful experimentation. Consider for example deep neural networks, which require very large amounts of computation for training. A trial-and-error approach to determining network structure and hyperparameters may require many attempts before success (if it ever comes), each of which uses a large share of available computation. Hundreds of researchers iterating over these large trial-and-error loops can bog down both computation and storage of even a world-class computing resource. Great care and planning will be required to minimize unnecessary experimentation and focus ML training attempts on those that are likely to succeed. [Question 1.E]

Basic research should be performed at an unclassified level. This research would provide foundational software and mathematics upon which industry and government agencies could build products within their own computing infrastructure. The NAIRR network will need to be secure so that only U.S. consortium members have access. [Question 1.F]

Privacy of data must be preserved within the NAIRR cluster and network. Commercial cloud computing resources already have privacy built into their systems, although it is a good idea to check certification periodically. Any experiments including person information must be approved by a human studies board. Anonymized data is preferable as it would be expected that most research benefits would be gained through the sharing of data within the NAIRR amongst researchers without having to worry about privacy. [Question 1.G]

Federal Funding and Partnerships with Private Sector [Question 1.H and 1.I]

An initial allocation for each participating government agency in the consortium should be allocated by Congress. Each government agency would then be responsible for funding industries and universities to perform the AI research that is relevant to their needs. To minimize duplication, representatives from each government agency would share their plans and discuss how to coordinate overlapping research agendas. [Question 1.H]

NSF should be responsible for funding universities including and emphasis on Historically Black Colleges and Universities and other Minority Serving Institutions. DOE, DoD, and DARPA would fund collaborations with industries and universities, also emphasizing inclusion and equity. [Question 1.I]

Capabilities and Services [Question 2]

Highest priority should be the creation of a cloud infrastructure with access to datasets that can be used to test AI algorithms. Second highest priority should be access to educational tools and services. Third highest priority should be standards for measuring and assessing the performance of the AI algorithms on particular applications.

Ethical and Responsible AI Research [Question 3]

NAIRR will need to identify those applications where we must be concerned with racial and gender equity, fairness, bias, civil rights, transparency, and accountability. For some scientific applications such as inertial fusion or combustion research, these may not be a primary concern; however, for many applications that include personal information, this will be a major concern. For those applications, restrictions on the use of AI algorithms will be required.

Building Blocks for NAIRR [Question 4]

There are several building blocks that already exist for the NAIRR in terms of government activities. Within the DOE, NNSA, and Sandia National Laboratories, foundational research in AI is already taking place with an emphasis on DOE and NNSA applications. DOE's Advanced Scientific Computing Research office has initiated projects related to Scientific Machine Learning[5], and NNSA's Advanced Simulation and Computing (ASC) Program has a new initiative in Advanced Machine Learning.

Scientific Machine Learning (SciML) has the potential to transform science and energy research. DOE has significant investments in massive data from scientific user facilities, software for predictive models and algorithms, and high-performance computing platforms that will benefit from advances in machine learning and artificial intelligence. Six prioritized research directions are

1. Domain Aware SciML – Integrating human expertise and domain knowledge with scientific machine learning methods
2. Interpretable SciML – new exploration and visualization approaches to interpret complex machine learned models using domain knowledge as well as metrics to quantify model differences
3. Robust SciML – research to show that SciML methods are well-posed, stable, and robust
4. Data-Intensive SciML – developing improved methods for statistical learning in high-dimensional SciML systems with noisy and complex data, for identifying structure in

[5] https://www.osti.gov/servlets/purl/1478744

complex high-dimensional data, and for efficient sampling in high-dimensional parametric and model spaces.

5. Machine Learning-Enhanced Modeling and Simulation – developing new methods to quantify trade-offs and optimally manage the interplay between traditional and machine learned models
6. Intelligent Automation and Decision Support – new mathematically and scientifically justified methods to guide data acquisition and assure data quality, improved SciML methods for multimodal data encountered in scientific applications, and new methods to optimally manage resources using in decision support

Additional building blocks are being developed within the NNSA ASC program. The ASC Advanced Machine Learning (AML) Initiative will improve simulation capabilities in weapons design, production, qualification and certification activities, as well as stockpile assessment through advanced data-driven analyses. This will increase NNSA's agility, enabling greater exploration of design spaces and improved predictive capabilities while potentially lowering the cost of physics simulations and data analytics. Similar investments to the DOE Office of Science prioritized research directions are being made in advancing physics-constrained machine learning, improving our ability to employ machine learning with sparse data, validating and explaining machine learning, exploring learning hardware in a high performance computing environment, creating AML-tailored data environment, and improving simulation workflows.

Finally, the Computing and Information Sciences Research Foundation within Sandia National Laboratories, a DOE multimission NNSA laboratory, has recently started an initiative in Trusted AI. This strategic initiative focuses on the rigorous foundation required to use AI technologies and advancements in high-consequence national security applications. Three thrust areas are being explored:

1. Mathematical Foundations – Improvement of AI methods, relying heavily on abstraction and theory. Emphasis is on understanding the power and limitations of AI methods, creating novel approaches to training and inference especially with limited data, and ensuring robustness especially in the context of extrapolation
2. Efficient and Secure Systems – AI tools, algorithms, methods, architectures, and hardware that mitigate open research challenges in scalability, data management, domain and architecture-awareness, and counter-adversarial security
3. Usability and Trust – Improved decision-making performance and understanding of trust and use in national security applications. Emphasis is on developing a causal, theory-based understanding of trust and use, including when and why they dissociate. This includes creating trust measures necessary for making decisions based on AI results, principled approaches to domain-informed AI that increase user trust and understanding adversarial impacts on overall decision quality.

All three of these efforts plus similar efforts at the other DOE National Laboratories will want to participate in the NAIRR consortium. Many of these efforts are already collaborating with U.S.

universities, and the new NAIRR consortium will allow these current efforts to leverage similar research efforts being sponsored by other U.S. government agencies

Potential Limitations [Question 6]

There are several potential limitations that could prevent the NAIRR from democratizing accessing to AI R&D. First, industry leaders (Google, Amazon, Facebook, Microsoft, IBM, etc.) may not want to participate in this consortium because they are already working with universities and may not find the consortium model to be a competitive advantage to them. Allowing these industry leaders to provide cloud computing resources as an in-kind membership fee will help draw them into the consortium.

Second, rules on the sharing of intellectual property could also be a bottleneck and delay initiation of consortium research activities. As found in Sandia's Combustion Research Facility[6], a variety of consortia models will need to be established to meet the needs of the partnering organizations. Also, the consortium should be focused on advancing fundamental knowledge. Intellectual property and proprietary designs can be created by industry after the consortium's development of the fundamental knowledge.

Third, public funding could also be an issue. A path for sustained funding across government agencies will be a critical part of ensuring continued democratized access to AI R&D.

Fourth, as mentioned previously, AI algorithms are not plug-and-play, regardless of how easy any given software implementation is to use. Most successful AI applications entail extensive use of both domain expertise and AI expertise. The point of developing a national AI computing resource therefore needs to include matching AI expertise with domain expertise in the application spaces to ensure that AI algorithms are being applied appropriately, and that any available domain expertise is incorporated into attempted solutions. If use of a national AI computing resource devolves into an unmanaged set of AI applications that ignore either relevant domain knowledge or mathematical limitations of the algorithms, then the NAIRR will not produce the desired technical advances.

[6] https://www.energy.gov/sites/default/files/2015/09/f26/CRF%20Case%20Study%2008-11-15%20FINAL%20CR.pdf