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.

FABRIC Testbed Response to the Request for Information (RFI) on an Implementation Plan for a National Artificial Intelligence Research Resource (NAIRR)

Submitted by:
FABRIC Testbed:
Ilya Baldin, RENCI
James Griffioen, University of Kentucky
Tom Lehman, Virnao
Inder Monga, DoE ESnet
Anita Nikolich, University of Illinois-Urbana Champaign
Paul Ruth, RENCI
KC Wang, Clemson University

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

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

- **Goal 1: Infrastructure for AI compute**
  - Emphases: extensible, distributed, federation, clear/single entry/portal, security, open to all (while allowing diversified features, privileges, and governance)
  - Metrics: Capacity, sustainability, number and diversity of stakeholders, clarity and low barrier of use
- **Goal 2: Infrastructure for AI Data**
  - Emphases: extensible distributed, flexible movement, federation, clear/single entry/portal, security, open to all (while allowing diversified features, privileges, and governance)
  - Metrics: Diversity, number and size of accessible data sets, clarity and low barrier of use
- **Goal 3: Infrastructure and tools for sharing and templates for AI Workflow**
  - Emphases: best practice, programmable templates supporting flexible use of any NAIRR resources, community & learning portal.
  - Metrics: Number, types, and quality of templates, use counts of templates
- **Goal 4: Risk Framework and governance processes**
  - Emphasis: balance of open access and advanced, strategic mission uses

- Metrics: coverage of uses, efficiency of approval, supervision, and auditing of proper use of infrastructure

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

A NAIRR can be the enabler to bring together (or federate) disparate, individual system testbeds that specialize in singular technical aspects such as cybersecurity, cloud computing, autonomous vehicles, 5G, network, etc into a larger, overarching AI testbed ecosystem that simulates how AI will work across organizational and technology boundaries. Although discipline-specific testbeds host experiments containing elements of machine learning, combining them under a larger umbrella - a testbed of testbeds - enables a holistic, system-wide view of AI. Forming an overarching “AI Testbed of Testbeds” based on existing new infrastructure such as FABRIC will enable more meaningful analyses of the ways in which results from AI decision making serve as input into and influence upon the decisions made by other systems. This is particularly important as machine learning data from one system feeds other systems; variations in this input are important to test in the structured manner a testbed provides.

Testing AI systems for resilience and safety is currently done in a number of verticals that remain isolated from each other. AI for Medical, Vehicles, IoT devices, Adversarial AI and AI-driven Networks are all tested in varying, specialized testbeds. However, the dependent interconnections of AI systems across these verticals can benefit from a larger testbed ecosystem that enhances repeatability and provides auditability and oversight. An overarching AI Testbed of Testbeds would encourage reproducibility and enable data sharing across AI domains since access to the Testbed ecosystem is expected to be more democratized than current efforts that are restricted to academia or government. An AI testbed will also enable realistic adversarial machine learning and AI environments, the results of which enhance AI resilience and trust.

A National AI Testbed of Testbeds could interconnect Internationally to similar nation-scale systems and AI testbeds around the world. For example, results of ML research based on experimental sensors in one smart city or edge testbed should be easily repeatable using several other smart cities’ testbeds. Any differences in outcomes may indicate opportunities to tune AI/ML algorithms and/or optimize smart cities.

A NAIRR Testbed enables experiments that explore open research questions such as:

Adversarial AI and AI Security. 3000+ papers on attacking (and to a lesser extent defending) ML models have been published in the past ~5 years. While very few of these attacks are detected in the real world, a place for experimentation at scale and across multiple technologies is important in order to build resilient AI systems that are protected from attacks. Additionally, no singular testbed exists to do vulnerability assessments of AI models. A testbed environment in which models exist on multiple types of devices will assist security researchers in assessing new types of security defenses which must be constructed. There is very little knowledge among practitioners on how to secure AI systems.

Developing Machine Learning and AI development best practices. Requirements about how to fairly develop and secure AI models are lacking, as are risk management standards and even widely agreed upon best practices. A testbed is an ideal, sandboxed environment in which to explore emerging work in this area.

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?

Ethical and responsible AI R&D must follow a process which is open and auditable and should be conducted on an independently operated infrastructure open to anyone with the proper credentials. NAIRR can help to establish this national testing resource. AI systems research and development requires testing - ideally on a controlled and well-instrumented testbed under simulated real-world conditions. Current nation-scale systems testbeds focus on individual technologies such as 5G, cloud computing or networking. AI-enabled systems will require testing across multiple areas of computer science. A NAIRR encourages research to be performed on a neutral, nation-wide infrastructure which requires documentation of the training data set provenance, information flow provenance as well as experiment results. This will serve to enhance the transparency of algorithm construction and performance.

Meaningful AI research depends on data. A lot of data. Storing and sharing machine learning training data remains a challenge, especially in medical domains, although they are increasingly using Federated Learning and other privacy preserving methods. However, there is an opportunity during the development process by using a testbed to ensure data contains elements of fairness. Requiring documentation such as data set “nutrition labels” in the course of testing will prompt researchers to consider whether the data set is truly representative of all populations. This will also serve to force transparency in algorithm design and enhance auditability as data set provenance is documented. As part of the approval process for projects using a testbed, testbed owners can encourage researchers to address questionable data set issues if the training data sets and data are made transparent. Part of an onboarding process can include a forcing function to have researchers answer questions about ethical sourcing, vetting, repeatability, etc.

An AI Testbed of Testbeds can uniquely serve as a conduit for an AI Data Exchange or perform a Data Sets as a Service function for those who don't have easy access to large amounts of labelled or unlabeled data. Since resilient and trustworthy AI systems should be tested, the testbed facility is the perfect matchmaker for those who have data and those who need it.

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

The NSF community has accumulated a substantial amount of practical experience in building and operating large distributed testbeds for their research communities (GENI, DETER, NSF Clouds, PAWRs, SAGE, FABRIC). Lessons learned include architecting, deploying large distributed infrastructures, proper federated identity management, security procedures, control, measurement software, dealing with operational and experimenter data and, importantly, methods of operating these infrastructure by inter-institutional teams - lessons that can be transferred onto NAIRR to make it a sustainable, multi-institutional effort.

These testbeds contain important resources that can be tied together to support a variety of applied AI research, which when coupled to NAIRR infrastructure would create unique opportunities for cross-disciplinary experimentation in applying AI techniques to multiple technologies and research domains with significant impact on our society - 5G, autonomous infrastructure management, autonomous vehicles, cyber-security. Such resources, which a standalone NAIRR infrastructure would otherwise have to replicate at great cost.

Specifically, the NSF-funded FABRIC Testbed represents an important element of this ecosystem intended to tie together the many pieces, constructing a continuum of computing from large centralized resources to small edge resources and dedicated instruments, with an intelligent, programmable and stateful network in-between. As many data intensive disciplines seeking novel AI applications transition from the idea of processing data in situ to operating on streaming data as it is produced by instruments and sensors, FABRIC provides a unique platform on which such ideas can be tested at scale.

FABRIC with its broad reach to a wide variety of research communities also has significant convening powers through workshops and other outreach activities, providing a platform to discuss different types of research relevant to NAIRR goals.

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

Technology companies and content providers conduct much of the same machine learning research in parallel with academia but with a larger amount of both hardware resources and data. An increasingly large percentage of accepted papers at the top machine learning and AI conferences are from industry. But there has been little incentive to work together. Industry researchers have ample resources, while academic researchers struggle for funds. However, both industry and academia lack a nationwide multi-disciplinary testbed with resources dedicated to conducting machine learning and AI research. The benefit to industry is increased access to highly skilled researchers, while the benefit to academia is access to hardware resources and data. Both sides benefit from access to large scale, multi-disciplinary resources on which to conduct research. Several successful NSF programs such as PAWR and the recent Fairness in AI have been formed by cooperation between agencies and industry.

A joint, multi-agency effort, including NSF, DoE, DoD, DHS, CISA, FDA, FAA, and multiple others, would ensure that efforts at creating an AI Testbed aren't duplicated across Federal agencies.

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?

Insider-only Data Access: Many of the advances in AI/ML/DL will be based on access to rich, extensive, data sets that, today, are often held and managed by companies or government agencies where access is restricted to approved members of the organization. In order to gain access to these data sets, researchers must often become a member of the organization (e.g., summer employment) or work with someone within the company or agency. These types of selective engagements do not scale and do not facilitate democratized access to the data. While industry is known to have some of the most extensive data sets, there will also be data sets needed for research in niche areas. These will often be collected by smaller sets of researchers who, historically, have been hesitant to share their research data, or will only share their data after they have studied it and some amount of time has passed. Such practices could also significantly limit data democratization efforts.

Screening Research Intentions: Good work is being done by organizations with research goals and intentions that might not align with those of the data holders. For example, non-profits, human rights organizations, and citizen scientists are all exploring questions related to fairness, particularly as it relates to AI/ML models, and may experience challenges accessing the data they need to carry out their research.

Resource Limitations: AI/ML research often requires massive data sets. The resources needed to store, transport, and process these data sets can be so burdensome that some portions of the research community are not able to participate. While some of these issues can be addressed with increased funding, others, such as local network infrastructure may be difficult to address for any number of reasons.

Legacy Access Methods: Existing data access techniques (e.g., remote file services and repositories, web services, network transport protocols, etc.) are not designed to provide efficient and precise access to the massive data sets needed by AI/M/DL. Many AI/ML/DL systems are designed to pull in data sets from centralized or cloud storage and perform local computations, oftentimes pulling in the entire data set when only a small portion of the data is needed. This leads to slow data retrieval times and many redundant (local) copies of the data. Intelligent access mechanisms and protocols are needed to strategically pull in precisely the data that is needed. Testbeds such as FABRIC could be used to help develop these types of intelligent access mechanisms.

Viewing Data as Infrastructure: Data is problematic in terms of access to training data, data labeling, data sharing, etc. One cannot assume that the data can be moved to the processing. Current approaches to working with data, particularly data that is intentionally distributed and must remain where it is, need to be rethought. For example, Federated Learning and other new data and model pipelines as an architecture need a place to experiment at scale and testbeds such as FABRIC offer such a place. Storage of the data, the AI models, the results should be funded by a Federal entity.

Converged AI Research: As AI research rapidly evolves in the scope and scale of data, several critical considerations have emerged. Data are often ingested from multiple sources, and ensuring effective access, transfer, staging, and quality control of such data requires high quality tools, infrastructure, and best practices. A nationally shared testbed like FABRIC is not just about the state-of-the-art infrastructure it provides. Even more important is its provision of a shared infrastructure where **templates of curated toolsets, data, and computing infrastructure** can be created based on best practices and shared with researchers with diverse backgrounds and interests, so that all can begin developing their cutting edge AI research on **a common, high-quality, validated foundation**. The template approach is also an opportunity to bootstrap and broaden **transparency and ethical considerations** in AI research by publishing, sharing, and teaching such tools and methods amongst AI researchers to inspect their own AI data and solutions. Sharing of data and software tools has been a longstanding practice in many research communities. NSF XSEDE, for example, has facilitated sharing of High Performance Computing (HPC) tools in a national community. What FABRIC is offering is one step beyond the sharing of tools, instead, providing a complete environment with tools, data, instructions, and programming interfaces for developing new AI solutions.