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.

CALYPSO AI

CalypsoAI's Response to the Request for Information on an Implementation Plan for a National Artificial Intelligence Research Resource

Calypso AI Corp.
2955 Campus Dr. Ste 110
San Mateo, CA 94403

Under Congresswoman Anna G. Eshoo, CalypsoAI is an official supporter of the bill that created the National Artificial Intelligence Research Resource (NAIRR) Task Force. We are pleased to respond to the National Science Foundation (NSF) and the White House Office of Science and Technology Policy (OSTP)'s request for information (RFI) on an Implementation Plan for a NAIRR. CalypsoAI believes an initial roadmap created by the NAIRR Task Force and informed by academia, industry, and government, is an important step towards achieving the goal of democratizing access to a cyberinfrastructure that fuels Artificial Intelligence (AI) research and development (R&D). CalypsoAI's CEO, Neil Serebryany, highlighted the importance of this initiative in a NAIRR Task Force Act press release stating, "the research infrastructure that will be created by this legislation is critical to our nation's ability to lead the world in building secure and operational AI."

As a software company and thought leader for Secure and Trusted AI, CalypsoAI knows firsthand the importance of access to a shared holistic advanced computing and data ecosystem. Our work on the testing, evaluation, verification and validation (TEVV) of AI and Machine Learning (ML) models resulted in contracts with the U.S. Department of Homeland Security (DHS) Science and Technology (S&T) Office's Screening at Speed Program (SaS) and the Secretary of the Air Force Concepts, Development, and Management Office (SAF/CDM), named Secure Artificial Intelligence. We are also working tangentially with the National Institute of Standards and Technology (NIST) on Software for Trusted Intelligence.

Drawing from our cutting-edge AI R&D team and experience working with industry and government in the AI space, CalypsoAI provides the following comments to inform the Task Force's consideration of options and development of a NAIRR implementation roadmap.

RFI-Specific Question Answers:

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

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

Planning frameworks are frequently adopted out of disciplinary habit and as a feature of cultural production within mature organizations. Just as software engineering will default to agile and iterative methodologies, or military contractors may rely upon the conventions of program management professional trainings (PMPs), national technology policy and planning frameworks typically reflect the beliefs, bias, and organizational structures of the governing body. The challenge, and opportunity, of national AI planning initiatives is to make a clean break, to reflect upon the diversity of planning frameworks, and to construct a planning methodology best attuned to the nature of AI problems – not configured according to the composition of stakeholders.

In this spirit, it is *too early* to settle on key performance or success metrics, or even to articulate the goals, as the problems are still undefined. However, we do know a few things; first, at this point in the nascent 90-year history of artificial intelligence and cybernetics, many core problems include the threatening reproduction, extrapolation, and distribution of misinformation/disinformation, all of which pose a threat to our national security. Second, we know that technology companies are ill-equipped to manage the threats and consequences of their work. Finally, extensive peer review research in the realms of sociology, psychology, and human computer interaction has taught us that many tools, products, and algorithms are deeply divisive and harmful to human relationships, good governance, and social welfare.

Recognizing our limitations, goals for AI resource planning should begin with a review on the emergent socio-technological failures of our AI era. Exploratory multi-stakeholder exercises in scenario planning, speculative design, advocacy planning, and dialogue mapping should precede conventional structured planning and roadmaps. Resource planning, strategy, and investment planning should then align to the composition of the problems that emerge - *not* the whims, expectations, or beliefs of the participants.

Since the research, problems, and key actors shift rapidly within the domain of AI with the emergence of each new technology, user trend, or novel application, metrics should not be established at an overarching policy level. The formulation of metrics should be relevant to the specific bodies that are then tasked to pursue each problem. However, NAIRR should be the body to coordinate oversight or the review of employed metrics, ensuring their implementation, relevancy amidst the rapid shifts within the AI space, adherence to best practices and policy, and consistency with metrics used across the government.

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

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

Governance should be widely decentralized across multiple existing bodies but should coordinate efforts to reflect the same sound principles, such as existing federal memorandums on Responsible AI (RAI). While entities such as OSTP should provide oversight of the Research Resource by leading the conversation on AI regulation, each U.S. government agency should manage its own adoption, implementation, and decision-making. This approach is consistent with other large-scale strategic federal compliance concerns, such as information security and management. Existing bodies such as the Government Accountability Office (GAO) and U.S. General Services Administration (GSA) are already effectively positioned to support the release and support the implementation of standards.

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

To points B and C of the initial implementation roadmap, NAIRR could refer to the Federal Lab Consortium (FLC) as a governing structure example of ownership, administration, governance, and oversight. It is governed by an Executive Board with four nationally elected positions, six regional coordinators, six Members-at-Large, and chairs of standing committees. This board determines policy, direction, and budget. Furthermore, this Executive Board is advised by the National Advisory Council (NAC) informed by user communities (industry, academia, government, federal laboratories).[1]

For the NAIRR, these executive board members, regional coordinators, Members-at-Large, and chairs of standing committees should be nominated via an open call for nominations, similar to the U.S. Department of Commerce's current call for nominations for its National AI Advisory Committee. This provides members of government, industry, and academia opportunities to promote the best and brightest in their fields and to provide representation of various stakeholder entities. After nomination, a selection committee from NSF and OSTP should review nominations and select the NAIRR governing body members based on established records of distinguished service and eminence in the field. These members should serve for two-year terms in order to provide adequate time to execute initiatives while still accounting for the need to bring in new talent within such a fast-paced field.

This hierarchical structure of governance will be important for NAIRR stakeholders because it establishes authority and accountability. Creating a single entity like the FLC also helps accomplish the NAIRR's goal of democratizing access to the cyberinfrastructure that fuels AI research and development by establishing buy-in and visibility of stakeholders across government, industry, and academia.

However, if this body is to go beyond strategic planning and participate in programmatic decisions, it should do this in direct coordination and co-design with the federal agency, or agencies, with equities.

[1] Federal Lab Consortium, https://federallabs.org/about-us/organization

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;

Initiatives such as Data.Gov or USA.Gov, initially developed by GSA Technology Transformation Service (TTS), demonstrate the successful instantiation of shared public infrastructures, and should be both replicated and scaled to create a shared computing infrastructure. With both examples, multiple government agencies participate in managing computational resources with high quality user experience, upholding relevant standards, establishing modern development practices, and responsibly implementing open-source licensing. While it does not matter which agency leads the creation of a similar initiative, it is imperative to establish rules and standards for how information is stored and shared. This consistent curation, coupled with a simple user-interface portal, will enable researchers across the country and various industries to better use the information we have. It makes the information more usable, both across data sets and as a way of tracking over time, which promotes good governance and auditability. Without one particular agency leading the initiative for setting standards on the storing and sharing of information, every agency would set their own standards, losing the power of the NAIRR's shared infrastructure.

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;

Strong precedent exists for the federal management of high-quality data. Over the last ten years, agencies such as the U.S. Department of the Interior, the National Aeronautics and Space Agency (NASA), the U.S. Department of Health and Human Services (HHS), and the Environmental Protection Agency (EPA) have all taken active steps to make computable quality data available to the public. Yet solutions such as Data.Gov have been more effective because the intentional design – such as the use of Restful APIs, GeoJSON, robust metadata and similar normative data standards – have enabled rapid adoption and consistent updates. Rather than create new data resources, the NAIRR should invest in and scale the existing proven solutions, such as Data.Gov and USA.gov, to best align with the needs of the AI research and commercial communities.

A large barrier to the dissemination of high-quality data for the NAIRR is the quick and constant evolvement of metrics within the AI space. Without an agile approach to a data repository, data could already be outdated by the time it gets into the hands of the NAIRR stakeholders. To mitigate this problem, agencies should be required to input data within a certain timeframe after acquiring it. Additionally, the NAIRR should advocate for individual data efforts --including those of NASA, HHS, and EPA --to be consolidated into Data.Gov. This will provide the NAIRR stakeholders with one go-to database for high-quality government data sets that have breadth and depth of the most cutting-edge information. Finally, the storage and sharing of the data within such a data repository should follow the standards set by the particular agency chosen to create and implement the standards, as discussed above.

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

Typically, privacy and civil rights and civil liberties requirements are met through large reporting instruments. Although this approach is comprehensive, it risks infringement because it is not monitored consistently. Consequently, rather than implementing large reporting instruments, privacy and civil rights assessments should be more consistent and more modular in composition. This includes implementing prioritized checklists in place of sprawling compliance documents, constant discussion across stakeholders, and quickly collaborating when an issue arises to ensure it does not spread or recur. By aligning these privacy and civil rights and civil liberties checks with the pace of AI development, we may better protect citizens and build their confidence in AI systems.

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

AI governance suffers from extensive thought leadership and minimal real-world experimentation. There are few examples of companies or universities leading exploration in everyday human environments – such as the construction of an experimental intersection at the University of Michigan for autonomous vehicle testing – on the implementation of concepts in AI ethics or design principles. Rather than continuing to allocate money into the publication of white papers, reports, and large documents that explore ideas, the NAIRR should be sustained through action. This means that federal funding and partnership investment should support creative physical exploration and experimentation of AI governance concepts. For example, it is myopic to invest funding or pursue partnerships only with machine learning engineers, given that their work sits within human environments constructed and informed by a range of other professions. Partnerships should be broad in spectrum and interdisciplinary though targeting finite goals.

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

The NAIRR, in promoting and democratizing access to the infrastructure needed to fuel AI research and development, must address the global gap between AI development investment and AI deployment. While billions of dollars were spent on AI in 2021, AI deployment is remarkably low. Therefore, the NAIRR should specifically prioritize research and investment in capabilities and services that address the security and risk management of algorithms. This will promote trust in AI/ML systems and speed up their development and deployment. This prioritization by the NAIRR will strengthen the United States' national security and maintain its global lead in AI by putting AI/ML security at the forefront of research, development, and deployment of AI and AI-enabled solutions across U.S. industry, government, and academia. As an industry partner in the AI/ML

TEVV and risk management space, CalypsoAI has seen firsthand the importance of this prioritization of AI security and holds a suite of tools to assist NAIRR and its stakeholders in this space.

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?

AI/ML products developed with minimal standard tooling leads to algorithms of uncertain quality, subjective trustworthiness, and potential vulnerability to attack. Across all sectors, organizations lack the tools to adequately assess and monitor AI/ML products, often leaving such solutions undeployed.

Built-in oversight tools and mechanisms to support progress and accountability such as periodic evaluation, design choices that enable metrics collection, and transparent reporting will be necessary to reinforce principles of ethical and responsible AI research and development. Specifically, with CalypsoAI's expertise in TEVV of AI/ML, the Task Force's roadmap must provide solutions that enable the following for the development and deployment of AI/ML models:

I. *Responsible Application:* Provide rapid algorithm testing to encourage transparency and accountability in algorithm development and deployment, while respecting the intellectual property rights of participating vendors. Testing should also place the least possible amount of burden upon the vendor.

II. *Accessible Results:* Can be easily accessed and used by sophisticated experts who are not data scientists specialized in model testing. The testing process must be swift to expedite the role of model review within the acquisition process. Test results should be concise, easy to understand, and contextualized so that model review teams can rapidly ascertain, compare, and act upon model insights.

III. *Adaptability*: Solutions will need to accommodate an extensive range of models, designed for many use cases, and often with little insight into the training of those models. Every scenario cannot be planned for or built into the model. Tests are representative of a range of feasible scenarios to demonstrate how an individual is resilient and tolerant considering emerging unknown threats.
   - *a.* *Real-world Acuity*: Models must be assessed according to the trade-offs of mathematical performance and real-world use. Other metrics may need to be created.

IV. CalypsoAI recommends that the NAIRR use TEVV to assess robustness, security, reliability, and bias of AI/ML algorithms during development and deployment. Examples of ways to assess AI/ML performance are as follows:
   - *a.* Model evaluation

   i. Demonstration of best practices used in the machine learning lifecycle
   ii. Recommendation on thresholds for acceptable/tolerable performance
   iii. Accuracy, Recall, Precision, Specificity
   iv. ROC Curve
   v. F1 Score

b. Verification and Validation measures
   i. Model simplicity: model based – Measure to determine how simple or complex a model is
   ii. Model simplicity: feature based – Evaluation of the model performance relative to the percent of input features
   iii. Noise injection – Measure to determine how much of the data used in the model is corrupt or not understood by the system (i.e., outliers, randomness)
   iv. Bias Detection / Fairness Evaluation– Evaluation of the correlation between biased/unethical features, such as race and gender, and the other features used to train the model
   v. Parameter modification – List of what model parameters were modified independent of the training data to control the learning process and justification for doing so

c. Way to validate AI/ML performance test results
   i. Proof of data spreadsheets used
   ii. Demonstration of the model using new data to showcase performance matches the results of the previous testing

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

Government, academia, and the private sector have all recognized the need to harness the volume, velocity, variety, veracity, and value of data. As a result, they have invested significantly into various activities, resources, and services for artificial intelligence to compute data. For example, in 2018, the Defense Advanced Research Projects Agency (DARPA) announced a $2 billion multi-year investment in AI projects that has included R&D efforts into High Performance AI.[2] Over the summer, the National Science Foundation helped establish strong building blocks with universities through creating 11 new National AI Research Institutes, which includes the NSF AI Institute for Intelligent Cyberinfrastructure with Computational Learning in the Environment at the Ohio State University.[3] Moreover, Georgetown University stood up the Center for Security and Emerging Technology (CSET), which is addressing a host of relevant issues related to

[2] AI Next Campaign (darpa.mil)

[3] https://www.nsf.gov/news/news_summ.jsp?cntn_id=303176

AI, such as data and computational power and cybersecurity.[4] Finally, in the private sector, there are countless specialized startups that can leverage their expertise to solve different aspects of the computing and data infrastructure problem.

While assessing data quality is important, it is equally vital to take active steps in tandem with these efforts to secure the data we have. Given the skill gap that we see across emerging technology sectors, risk mitigation features must be as easy-to-use and accessible as possible. Recognizing this, CalypsoAI provides users of various skill levels the ability to assess the quality of their models and make adjustments based on the factors that are most important to them. This can include adjusting for datatypes, environments, and natural or adversarial corruptions to determine model performance in real-world deployment scenarios. In effect, CalypsoAI can help the NAIRR protect privacy and, when applied to cybersecurity systems, democratize cybersecurity infrastructure.

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

Public-private partnerships should be central to the NAIRR because the private sector brings innovative perspectives and technologies that would enable the NAIRR's goal of creating a holistic computing and data infrastructure. Some exemplars of productive partnership include leveraging the knowledge of Federally Funded Research and Development Centers (FFRDCs) and other think tanks with topical research expertise; drawing upon Small Business Innovation Research (SBIR) and Small Business Technology Transfer (STTR) opportunities that flow through innovation hubs; and implementing rotational programs to embed private sector knowledge into government.

Additionally, public-private partnerships should be used for advancing AI educational opportunities, particularly if it can work with universities to create a course or module geared towards improving the data and computing ecosystem. In the course, the NAIRR can work with private sector companies to expose students to emerging and innovative capabilities. For example, since CalypsoAI's AI Model Risk Management platform, VESPR™, can be used by non-technical individuals, the NAIRR can leverage it as a teaching tool to broaden access to AI research and education opportunities.

Finally, the Task Force could again refer to the FLC as a model for public-private partnerships. For example, FLC offers a comprehensive database highlighting funding resources for federal agencies, academia, and industry. FLC also implemented the SBIR/STTR programs mentioned earlier, which have been widely successful in funding emerging technologies and R&D from small businesses and universities. If the NAIRR does not directly leverage SBIR/STTR programs, it could follow suit with its data sharing database and create similar programs to encourage public-private flow of ideas, engagement, and technologies.

[4] Publications - Center for Security and Emerging Technology (georgetown.edu)

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?

There are a few limitations that could potentially affect the NAIRR's ability to democratize access to AI R&D.

One limitation is the Task Force's ability to identify the specific data types and specific tools needed for the NAIRR toolbox. For guidance, the Task Force should look to industry for the most common and sought-after data types and tools.

Second, balancing funding and demand within the NAIRR is challenging because the cost of purchasing the correct data, computing power and processing, and software can be expensive. To address this issue, there should be task force liaisons responsible for engaging industry, government, and academia for specific topic areas such as Test and Evaluation (T&E), Verification and Validation (V&V), and data quality to assess demand. Additionally, to balance funding and demand insight can be drawn from partnerships with industry, annual budget analysis and adjustments, and utilization of existing and complementary resources to maximize funding.

One of the biggest limitations of implementation is the lack of common understanding and universal language between AI policy makers, users, stakeholders, and decision-makers. Without a common understanding between actors or universal language, AI-related risk cannot be properly managed. AI jargon is often not understood by key decision-makers, who assess the real impact of the risks associated with AI and are responsible for determining the level of acceptable risk. Furthermore, AI teams and vendors typically work on a discreet part of an overall operational problem. They target AI-specific metrics, such as high performance or accuracy, without necessarily understanding the complete context in which models will run. Therefore, the NAIRR should work in conjunction with other government agencies, industry, and academia to define universal language as it relates to AI.

Finally, AI/ML initiatives continue to be siloed, with isolated actors attempting to build and deploy models with limited access to standard industry tooling. This often results in inefficiencies and redundancy in AI efforts, as well as a lack of communication across organizations on best practices. The NAIRR can be one way of addressing these siloes through a collaborative effort, bridging the gap between government, industry, and academia. However, it will also require a great investment of resources from all stakeholders. As resource investment increases, greater regulation and oversight is needed. Sound oversight includes established evaluations, measures, and considerations for areas of interest such as accuracy, explainability and interpretability, reliability, privacy, robustness, safety, security, and mitigation of unintended and/or harmful bias. These all contribute to ensuring accountability and AI model robustness which, although more than a checklist, creates parameters for established good practices.

Conclusion:

CalypsoAI firmly supports NSF and OSTP's effort in establishing the National Artificial Intelligence Research Resource and appreciates the opportunity to provide our thoughts and feedback on the path forward. We welcome any opportunity to work with the Task Force, industry, and broader government agencies to assist in developing an accessible, inclusive, responsible, trustworthy, and secure NAIRR for the benefit of all sectors.

For further questions or for more information please do not hesitate to reach out to Hannah Mezei at [redacted]