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.

INTRODUCTION

As an independent science, technology, and strategy corporation, Noblis delivers innovation and management expertise from a position of independence and objectivity. For more than 25 years, we have worked across the defense, national security, intelligence, and civilian domains to solve difficult problems that help our government—and nation—operate more effectively and more efficiently. By assembling top talent across many disciplines, we apply the right expertise to support our clients' most critical missions with practical and actionable solutions.

Noblis is an established industry leader in solving the most pressing national security issues of today and tomorrow and routinely delivers algorithms, analytics, tools, and services while sitting side by side with our clients. As a 501(c)(3) company focused on basic and applied research, Noblis is keenly aware of the value of collaboration and works closely with partners from across the public and private sectors, as well as academia. Among its many partnerships, Noblis is a member of eight Other Transaction Authority (OTA) consortia to deliver prototypes, subject matter expertise, and requisite capabilities in the areas of aviation, medical, energy and the environment, counter weapons of mass destruction, information warfare, command, control, communications, and cyber, and systems of systems. As a consortia member, Noblis and its partners share ideas, training, concepts, and capabilities openly. This level of "crowdsourcing" generates enormous benefit to include vastly enhanced subject matter expertise, rapid technology innovation, and best of breed tools, products, and services.

Noblis' Artificial Intelligence (AI) experts leverage our state-of-the-art AI Laboratory to accomplish both basic and applied research. Our experts improve analyst productivity by designing, implementing, and evaluating full stack software solutions using Agile and DevOps practices with a backbone of modern machine learning (ML) and AI algorithms. These solutions allow analysts to better organize, explore, and understand terabytes of data. Our experts focus on developing and deploying analytic tools and capabilities within existing operational environments lowering client risk and enabling them to generate new mission insights, increase speed to delivery, improve confidence, and save financial resources fully and more rapidly. Examples of Noblis expertise in this important field include:

- Improving the recognition performance on unconstrained face imagery.
- Integrating algorithms for cargo imaging and sensor applications.
- Developing algorithms for face morphing and aspect ratio manipulation.
- Leveraging deep learning networks to generate captions for images and videos.
- Identifying combinations of algorithms to counter adversarial machine learning.
- Using natural language processing (NLP) to create and maintain taxonomies.
- Progressing ways to create, track, and present AI rationale and chains of evidence.

The final report from the National Security Commission on Artificial Intelligence (NSCAI) included the following statement, "*the rapidly improving ability of computer systems to solve problems and to perform tasks that would otherwise require human intelligence—and in some instances exceed human performance—is world altering*".[1] Noblis concurs with the NSCAI that AI can have world altering impacts and is highly supportive of the goals of the National Artificial Intelligence Research Resource (NAIRR) Task Force. Our response to question 1 of

[1] NSCAI Final Report, pg. 7

Use or disclosure of data contained on this page is subject to the restrictions on the title page.

the NAIRR Request for Information (RFI) below illustrates the importance of bringing together people, process, data, and technology to achieve NAIRR Task Force objectives.

- *People*: Pairing together multi-disciplinary teams comprised of data analysts, data scientists, technologists, academics, and domain experts is an essential element to ensuring that research will be achievable, impactful, and adoptable.
- *Process*: Effective governance, to include providing direction for priorities and funding, providing training, and ensuring excellent internal communication and external outreach are essential to generating quick-wins and delivering value-added research and capabilities.
- *Data*: Sufficiently structured and validated data made available for training algorithms, testing capabilities, and delivering unambiguous, repeatable, ethical, and trustworthy results is essential for generating new insights and improving confidence in findings.
- *Technology*: Leveraging cloud-based computing power and expanding the use of technology such as NLP and deep learning to enable automated data-curation and automated entity/taxonomy development of large corpuses of data are essential for increasing speed to deliver relevant research and findings.

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

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

The National Science Foundation (NSF) – led National Artificial Intelligence Research Institutes (herein referred to as Institutes), twenty now in total[2] represent an exciting opportunity to research, develop, test, and apply AI to challenging problem sets and to many new fields. The Institutes, which are located across the country and are comprised of academia, government, and industry are excellent AI building blocks. A strong foundation comprised of effective governance, available infrastructure, robust collaboration, and assured funding are necessary to focus ideas, generate value-added research, share best practices and lessons learned, and to ensure the building blocks are solid and sustainable for years to come.

Goal 1: Establish NAIRR Governance. Noblis recommends that a NAIRR Governance Board (herein referred to as the Board) be established and chaired by the Director of the Office of Science and Technology Policy (OSTP). The Board should include the NSF Director and the Director's from each of the Institutes (one per each). The NSF Director, in consultation with Director/OSTP should select the directors of the Institutes keeping in mind that a proper balance between academia, public, and private sector leadership is advisable to ensure optimal Board perspectives and decision-making. Noblis also recommends that the Directorship for each Institute rotate periodically (e.g., annually) between academia, government, and industry to maintain high levels of involvement, ownership, and transparency. The Board should meet monthly within the first year of initiation to establish and implement a strong and sustainable modus operandi. This too will help build networks, break down barriers, identify joint needs and requirements, and allow the members to share, leverage, and apply best practices, both individually and collectively. Significant Board functions include:

- Prioritizing and coordinating Institute research. – See Goal 2

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

Use or disclosure of data contained on this page is subject to the restrictions on the title page.

- Funding and maintaining the underlying infrastructure (i.e., cloud technology and tools for all Institutes). – See Goal 3
- Providing an AI Technology Accelerator Platform (TAP) to track and monitor projects. – See Goal 4
- Implementing an acquisition model that pairs customers (i.e., government) to performers (i.e., Institutes) and enables funding to be put on contract expeditiously. – See Goal 5
- Providing data governance to make relevant data widely sharable and protected accordingly for research through data sharing agreements; Providing technology, data, research, or related policy recommendations; and Communicating NAIRR plans, successes, and requirements to the President, Congress, and the public as appropriate. – See Goal 6

Suggested performance metrics for Goal 1 include:

- Governance Board established / All members identified and in position
- Terms of Reference and Concept of Operations published
- Number of meetings held / Number of decisions made
- Number of performance metrics met (for goals 2-6), and associated results

Goal 2. Prioritize and Coordinate Institute Research. The range of good AI ideas is nearly infinite. Governance is required to prioritize those ideas into research that delivers the greatest impact within a cost and resource-constrained environment. Before launching into a project, Institutes should submit Research Proposals to the Board for their review and approval. Research Proposals should be specific and measurable, and document planned data and technology use, expected performance metrics, likely deliverables (e.g., code, tool, paper) and necessary infrastructure, cost, and resource considerations. Research Proposals should also include a plan for communicating findings to the public (to the greatest extent possible) and methods for sharing deliverables with other Institutes and the public as appropriate.

Despite exciting experimentation and a few small AI programs, the U.S. government is a long way from being "AI-ready[3]*."* While a percentage of basic research must exist to explore and push technology beyond the realm of the possible, the Board's review and approval of Research Proposals should lean towards applied research and focus on criteria such as technical feasibility, positive impact to the problem set, cost, schedule, and risk to accelerate our ability to be "AI-ready" now. A key aspect of the Boards review is to ensure effective coordination of effort between Institutes. For example, if one Institute is developing NLP techniques for entity extraction, another Institute can take advantage of that research and apply it to a different type of problem. The proposed AI-TAP (Goal 4) can be value add in enabling the board to accomplish this function. Suggested performance metrics for Goal 2 include:

- Number of Research Proposals submitted / reviewed / prioritized / funded
- Numbers of proposals that were able to leverage other projects work
- Number of proposal projects completed / shared with public
- Numbers / types of proposal outputs (data, software, publications, patents) delivered / shared (among Institutes, with the public)

Goal 3. Fund and Maintain the NAIRR Infrastructure. AI research requires significant amounts and types of data, which require significant storage and compute power. Noblis, in support of one of our clients for example, required storage capability for up to 10 petabytes of

[3] NSCAI Final Report, pg. 2

Use or disclosure of data contained on this page is subject to the restrictions on the title page.

data, processed data in excess of 10 terabytes, extracted, transformed, and loaded 1.5M records monthly, and accomplished processing and analytics of those data on a daily basis using AI to identify unique types of inconsistencies. While not all research projects may require this type of capability, large corpuses of data will need to be structured, stored, and analyzed. Synthetic data may need to be created where data does not exist or cannot be used due to legal or policy purposes. Problem-specific taxonomies will need to created and testing and training data sets developed to account for wide-ranging variables and to address edge-cases. All the above are not trivial tasks and will require significant compute power to apply technologies such as NLP to automate data collection, processing, and synthesis so that research can be accomplished and delivered in a timely manner. These challenges and Noblis' proposed recommendations will be discussed in greater detail in our response to question 1(E).

NAIRR, at the governance level should establish, maintain, and fund a hybrid cloud (e.g., AWS/AZURE) infrastructure that includes data storage and elastic compute and that makes tools available, as proposed in Table 2 (see 1 (D) response below) in support of the Institutes and their associated research. The NAIRR infrastructure should also support the AI TAP discussed in Goal 4. Suggested performance metrics for Goal 3 include:

- NAIRR hybrid cloud environment established / Cost to establish
- Number of Institutes with access to the infrastructure / Time to gain access
- Numbers / types of tools available / Numbers of new tools needed for specialized projects
- Numbers / types of data stored
- Numbers of compute requirements identified / funded
- Numbers / types of cloud and compute training required / funded
- Cost to maintain and use the cloud vs. cost to deliver AI output

Goal 4: Provide an AI Technology Accelerator Platform (TAP). Effective collaboration across the Institutes and more broadly across an AI Ecosystem requires an AI TAP. The AI TAP should provide a one-stop location for tracking and monitoring projects, for matching requirements to contract mechanisms, conducting evaluations, sharing best practices, gathering analytics, and presenting dashboards, all supported by security and access management protocols. Table 1 includes suggested features for an AI TAP.

Table 1. Noblis Proposed AI TAP Features and Benefits.

Features	Benefits
Project Setup	Establish new projects from an Idea Management Module
Evaluation	Capture evaluation findings in a user-friendly dialog window
Contract Management	Store contracts and process modifications at the project level for rapid traceability
Financial Management	Track invoicing data and principal investigator (PI) estimated expenditure data.
Milestone Tracking	Link milestones to dates with automated notifications; Alert PIs of upcoming deadlines
Document Upload	Enable PIs to load their deliverables and link them to milestones
Working Area	Provide secure document sharing and collaboration
Dashboard	Analyze and visualize project and portfolio performance indicators
Realtime Chat	Allow members to communicate in a secure collaboration environment
Security Authentication / Browser Encryption	Provide two-factor authentication with identify verification certificates and security tokens; Access a secure remote access portal using Secure Sockets Layer protocol
Role-based Access	Apply access based on user needs, as approved by the security official

Use or disclosure of data contained on this page is subject to the restrictions on the title page.

Suggested performance metrics for Goal 4 include:

- Platform established; Levels / types of usage
- Numbers / types of projects; Submitted / Evaluated / Funded / Completed
- Numbers / types of users; Numbers / types of collaborations / Resulting in leveraged applications, new ideas, new partnerships, increased speed to delivery

Goal 5: Implement a Sustainable Acquisition Model. The NAIRR must create an acquisition model that allows customers (e.g., government) to fund research by performers (Institutes comprised on multi-disciplinary teams) in a highly efficient manner – which largely does not exist today. Amid a crowded federal marketplace, the acquisition model must provide valued services, speed, and ease of use on a single platform, such as what is being proposed via the AI TAP. Noblis proposes that the NAIRR initiate an AI Research Consortia that would operate and be supported by the AI TAP, and implement a decentralized, AI-focused acquisition approach to avoid redundancy, capitalize on best practices, and build a larger and more cohesive team of government and AI partner stakeholders. This type of approach would allow the NAIRR to be contract agnostic and identify, assess, and access a range of acquisition options, including Indefinite Delivery, Indefinite Quantity Task Order contracts, Government-Wide Acquisition Contracts, GSA Multiple Award Schedule contracts and non-FAR-based agreements, such as Other Transactional Agreements (OTA), providing for maximum scalability and accounting for likely fiscal constraints. OTAs in particular represent a useful vehicle as they focus on research, allow for exposure to large numbers of performers, and enable rapid capability transition. Also, the laws and regulations concerning intellectual property (IP) rights and cost accounting/pricing do not apply to OTAs although OTAs generally do contain provisions addressing such topics as IP rights and cost accounting. Suggested performance metrics for Goal 5 include:

- Numbers / types of contracts matched to Institute projects
- Time to establish a contract mechanism per project, Level of funding per contract
- Type of customer (e.g., government, venture capitalist) / Type of desired research or use

Goal 6: Provide Data Governance, Provide Policy Guidance and Recommendations, and Communicate Plans and Success. As noted in Goal 3, substantial numbers and types of data will be required for research. Obtaining those data, unless publicly available, as well as storing, processing, sharing, and noting its use in public are all likely to require data governance. While the identification of needed data and the creation of data sharing agreements should remain the responsibility of the individual Institutes doing the research, the Board should establish and maintain a Data Governance Committee comprised of Chief Data Officers to review the agreements, ensure legal and policy considerations are accounted for, and ensure a coordinated data approach exists across the NAIRR and its associated Institutes.

The use of data and new technology are likely to have policy impacts much like the recent discussions on biometrics. These impacts may result in the need for existing policy to be revised or new policy written. While the Institutes will have a role in unearthing these types of policy requirements, Noblis proposes that the Board establish and maintain a Policy Review Committee that routinely interacts with the Institutes to understand research-based policy implications and that proactively works with the executive and legislative branches of government to affect policy change that advances data and technology use, while ensuring proper protections to citizens.

A recent global consumer study conducted by Pega revealed that *many consumers couldn't even recognize some of AI's most basic tenets [and] that nearly half don't understand that AI*

Use or disclosure of data contained on this page is subject to the restrictions on the title page.

solutions enable machines to learn new things, and even fewer don't know it can solve problems or understand speech[4]. In Noblis' experience, engaging with end-users (i.e., consumers) is critically important to enabling adoption, gaining support, and garnering important funding. For this reason, Noblis proposes that the Board establish a Strategic Engagement and Communications Team. This team would maintain a website that provides transparent information on Institute and NAIRR projects and successes, conduct outreach events to engage with local communities to help explain and highlight the value of AI, assist the Institutes in conducting hackathons, competitions, training, and recruiting events, and aid in the preparation and delivery of quarterly and annual reports and briefings to Congress, the President, the administration, and the public as appropriate. Suggested performance metrics for Goal 6 include:

- Data Governance and Policy Review Committees established / Strategic Engagement and Communications Team in place
- Numbers / types of data requirements / Numbers of Data Sharing Agreements in place / Levels of sharing allowed (e.g., fully open, just among Institutes, etc.)
- Numbers / types of policy requirements / Numbers of proposed vs. accepted refinements to existing policy / Numbers / types of new policy required, drafted, approved
- Communications website established / Numbers and types of listings, comments provided, by segment / Numbers of public meetings held / Briefings provided
- Numbers / types of events held; Numbers / types of attendees; Outcomes produced
- Annual report provided / Other reports (i.e., Institute Research Proposals) posted

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

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

In our experience, the agency or organization responsible for the implementation and administration of a substantial "new start" such as NAIRR, must have significant standing, political backing, and influence to garner broad based support and long-standing commitment among participants, funding backers, and likely AI consumers/beneficiaries. Noblis recommends that the NSF oversee the day-to-day activities of the NAIRR, that they appoint a full-time Director, and that they implement proposed Goals 1-6 discussed in our response to question 1(a). The NAIRR Director should report to the OSTP Director who can provide guidance and convene the Board for higher level oversight and decision-making. Conceptualizing and codifying a plan of action is critical to organizational success. For this purpose, Noblis proposes that the NAIRR Director work with the Board to develop and deliver a NAIRR Program Objectives and Milestones (PO&M) document within the first 60 days of assignment. The NAIRR PO&M should describe agreed-to goals, objectives, activities, milestones, performance metrics, and responsible and accountable officials (by name). While the administrative functions remain critically important for establishing a strong foundation in the early months of "startup", the NAIRR PO&M should include substantive and tangible deliveries of applied research within 180 days that are tied to challenging use cases to generate quick wins and show value. Throughout the first 180 days and for the remainder of the first year, NAIRR leadership should communicate with the President, Congress, the Administration, the Public, and with likely consumers (and

[4] https://www.ciosummits.com/what-consumers-really-think-about-ai.pdf

Use or disclosure of data contained on this page is subject to the restrictions on the title page.

funders) in no less than 45-day increments to report status, success, challenges, actions, and other pertinent information. Noblis proposes that the NAIRR PO&M be revisited on a no less than annual basis to discuss performance and needed refinements.

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

Please see our response to question 1(a), Goal 1: **Establish NAIRR Governance**.

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.

Noblis recommends that an AI TAP be implemented, and that hybrid cloud storage and compute capabilities be made available for effective NAIRR operations. Noblis employs a "lab to mission" methodology that follows three key principles: (1) validate and iterate; (2) simplify and uncomplicate; and (3) develop once, use many. As a company, we are tool and solution agnostic and use DevSecOps approaches that apply numerous capabilities and a wide variety of tools to deliver innovative research and solutions. Table 2 provides a list of frequently leveraged capabilities and is recommended as an initial set of technologies for consideration by the NAIRR and the Institutes. Noblis applies human-centered design and UI/UX processes to transform a stated need into a solution by engaging and collaborating with users/stakeholders throughout the design, development, and transition phases. Our process involves conducting evaluations on the iterative design (e.g., early mockups, prototypes, alpha/beta modules) and follows industry standards and best practices to ensure our tools allow users to accomplish their intended task. Noblis recommends that similar approaches be employed for NAIRR research efforts.

Table 2. Key Functions and Associated Tools Used by Noblis to Deliver Innovative Research.

Capabilities	Tools	Capabilities	Tools
Cloud storage	AWS S3	Multi-platform analytics and visualization	Grafana
High performance object storage	MinIO	Deep neural network training and inference	Tensorflow
Database (DB) management	Apache (Cassandra)	Multi-tasking operating systems	UNIX
Relational DB Management	PostgreSQL	Programming	Python
Relational DB construction	AWS RDS	ML Library	PyTorch
Large DB query	AWS Athena	Data stream processing	Kafka
Agile workflow	Atlassian Jira Confluence	Data integration	AWS Glue
Tester/Developer project management	VersionOne	Combine and query data across varied storage	AWS RedShift
APIs and SDKs for adding cognitive intel to apps	Azure Cognitive Services	Statistical analytics	SAS
Server scripting for dynamic web content	Node Js	Statistical computing	R
Number factoring	GNFS	Data clustering, classification, and regression	SciKit
Store, search, and compute (at scale)	Elastic	Automated code quality inspection	SonarQube

Use or disclosure of data contained on this page is subject to the restrictions on the title page.

Capabilities	Tools	Capabilities	Tools
Networking for large scale data computation	Hadoop	Big data search and access	PowerQuery
Web & micro-service development & integration	Java Spring Boot	Big data (ML and Statistical) analysis	H2O
Automated software development	Jenkins	Data integration and analysis	Splunk
Software development via crowdsourcing	Mechanical Turk	Data visualization and analytics	Tableau
Container building and deployment	Docker	Data visualization	Kibana
Container clustering for app deployment/scaling	Kubernetes	Automated data extract, transform and load	NiFi
Container application management	Portainer	Probabilistic data structure	Bloom Filter

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.

Noblis has extensive lifecycle data experience that spans collection, cleaning, transformation, and analysis involving a vast array of data types. Examples of government and commercial data that Noblis has worked with include web traffic, cell records, digital video, hard drive and removable media, social media, healthcare and geospatial data, basic safety messages and vehicle trajectory data, taxpayer, facility, financial, and air traffic information, and biometrics. Based upon our experience, too much of a data scientists time is spent discovering and preparing data for analysis vs. time spent on analysis to deliver valuable insights. The volume, velocity, and variety of data available today contains incredible value and can be applied to address myriad issues. Those same traits however make data search, access, retrieval, use, and analysis overwhelming. As part of the NAIRR governance process, and in accordance with relevant data sharing agreements, Noblis recommends that a thorough review be conducted prior to collecting, procuring, or distributing any datasets to the research community. Data acquired or generated should be specifically tied to use cases defined as high priority by the Board to avoid the risk of bad[5] data entering the system. Table 3 lists challenges and recommended approaches to the use of high-quality open source and government data for NAIRR goals and objectives.

Table 3. Challenges and Recommended Approaches to the Use of High-Quality Open Source and Government Data for NAIRR Goals and Objectives.

Challenge	Recommendation
Data Structure: Estimates say that just 20% of data is structured, while unstructured data accounts for 80-90%[6]. AI/ML or any type of computing is best accomplished when data is structured and can be compared/analyzed within a relational database.	**Leverage NLP** to auto-curate and structure data that is not yet structured. **Create mission or topic specific taxonomies** that are continuously updated via automated entity mapping. **Use training datasets** for NLP of tagging pipelines such as Parts of Speech, topic modeling, and Named Entity Recognition.
Data Sensitivity (i.e., Personally Identifiable Information (PII) and / or classified government data): The escalation of security breaches involving PII has contributed to the loss of millions of records over the	**Create and use synthetic data** coupled with real data that can be validated to not bias the data samples. **Incorporate API toggles** that leverage multiple registration tags and that use AI to teach the system the range of available enrichments while also structuring the

[5] Bad data may include data that is poorly structured, sensitive, bias, etc.
[6] https://monkeylearn.com/blog/structured-data-vs-unstructured-data

Use or disclosure of data contained on this page is subject to the restrictions on the title page.

Challenge	Recommendation
past few years'. While large amounts of information are made available via social media, unauthorized use of those data for studies can have legal ramifications. AI/ML research often requires access to sensitive information that may contain PII or in other ways be deemed sensitive. If the sensitive fields are removed, the remaining data often has limited utility.	APIs so that they limit or restrict queries based on role-based access. **Conduct pre-correlations between data sets to enable research**. Removing specific addresses but leveraging zip codes from medical records for example can still provide information about economic and medical demographics of the community.
Data Stagnation, Corrupt Data, and Incomplete Large Data Sets: Data sets will need to continuously evolve in both scope (e.g., number of records) and breath (e.g., number of fields) to avoid research being overcome by new data before it's even released. Data are often corrupted and simply removing observations that are not perfect may bias results.	**Leverage NLP** to auto-curate and structure data. **Create mission or topic specific taxonomies** that are continuously updated via automated entity mapping to reduce the need for researchers to formulate their own taxonomies prior to working on new algorithms.
Data Value: Enormous amounts of data are being produced daily. Most of it however is not mission or topic relevant and many sources (e.g., algorithms, bots, influencers) are not trustworthy, explainable, or credible.	**Create a data rating schema** to assess mission and topic value. **Incorporate crowd-sourcing software and sentiment analysis** to deliver credible and relevant data. **Leverage deep and recurrent neural networks** to create detection models that extract features and index and produce objects and relationships. **Implement clustering models** that generate terms of interest, import custom entities, and create customizable graphs.
Data Bias in Training Data Sets: Biases may be deliberate or inadvertent and may be reflected in techniques used to collect, generate, normalize, update, or analyze.	**Leverage statistical techniques** to validate that bias are not present. **Incorporate and consistently apply a Data Quality Assessment** to ensure that approaches to data collection ensure appropriate representation.
Data Robustness: Often when faced with AI development challenges, researchers will use a portion of the available data as a training data set, and the remaining data as a testing or validation data set.	Clearly **distinguish between training data sets and testing data sets** which may be specifically sampled to include deliberate attempts to stress algorithmic approaches to ensure they are robust to edge cases.

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

Noblis proposes that the NAIRR consider employing a model like the highly successful Fraunhofer-Gesellschaft (Fraunhofer Society)[8] network of institutes for applied research. The NAIRR funding sustainment model, much like that of the Fraunhofer should include funding derived from diverse sources, including federal, state, and public funding fees. In Noblis' experience, obtaining government provided funding, allocated through multi-year Congressional appropriations are the most consistent form of funding. Other potential sources of funding to consider include those gained from Foundations such as the McCain Foundation, Bill and Melinda Gates Foundation, and the Warren Buffett Foundation, and through engagement with Venture Capitalist (VC). Critically important to ensuring industry and venture capitalist involvement is providing maximum flexibility for maintaining intellectual property (IP) rights, providing licensing fees for continued development and deployment of unique technologies and solutions, and encouraging spin-off companies – which in turn will add more expertise, development, and competition in the market. Allowances for maintaining and / or continued funding of IP will encourage private industry and VC investment. NAIRR funding should be provided via contract mechanisms, such as those discussed in Goal 5, pg. 5. Finding the proper

[7] https://nvlpubs.nist.gov/nistpubs/Legacy/SP/nistspecialpublication800-122.pdf
[8] https://www.nap.edu/read/18448/chapter/13#225

Use or disclosure of data contained on this page is subject to the restrictions on the title page.

contract vehicles and / or establishing effective contract mechanisms can be burdensome which is why we recommend the AI-TAP (Goal 4) as a means to rapidly match performers to customers via contract mechanisms and deliver applied research.

Customer involvement in research is critical to ensuring understandable, usable, and adoptable AI. For this reason, Noblis recommends that government (local, state, federal) be intimately involved in the creation of technology innovation pilots [based on well-defined and approved use cases – see 1(e), pg. 7-8] and that they and the Research teams routinely interface throughout the project, akin to the process described in 1(d), pg. 7. This type of involvement will generate quick wins, provide greater interest and value-add, and result in additional funding for projects.

We should embrace the AI competition. Competition already infuses the quests for data, computing power, and the holy grail: the rare talent to make AI breakthroughs[9]. Engagement with academia, specifically with students interested in AI technologies, in a way that implements a "dual system" of education/apprenticeship, should be a particular focus for NAIRR. As part of their undergraduate or graduate curriculum, students often must, or seek internships. This "free" or relatively inexpensive labor force infuses perspectives, grows talent, and results in opportunities for future employment. Noblis employs this type of "dual system" and is a proud recipient of WayUp's Top 100 Internship Program award in 2020 and 2021. Noblis' Internship Program gives undergraduate, Masters, and PhD students an opportunity to support direct client work and contribute to our internal research and development programs. This year, our internship program received nearly 7,500 applications for our 75 projects. Our interns came from over 50 colleges and universities and 18 states. In 2020, Noblis had an 80% acceptance rate on our intern-to-full-time offers. The opportunity for private industry to gain access to this type of talent and potentially hire home-grown experts will drive industry engagement and funding for NAIRR efforts. Noblis' experience in establishing state of the art labs and provisioning them with a vast array of capabilities, tools, data, and expertise generates quality research and appeals to prospective interns and applicants alike. Noblis proposes that the Institutes have and maintain similar capabilities and tools as suggested in Goals 3 and 4 and in question 1(D).

We look forward to the opportunity to discuss our knowledge, skills, and capabilities to help the Task Force make progress in furthering the value of AI!

[9] NSCAI, pg. 2

Use or disclosure of data contained on this page is subject to the restrictions on the title page.