

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.



NVIDIA's Response to the Office of Science and Technology Policy Request Regarding An Implementation Plan for a National Artificial Intelligence Research Resource

Federal Register Document: [2021-15654](#); Citation: 86 FR 39081
Submission Date: 1 October 2021

NVIDIA Contacts:

Jack C. Wells, Ph.D., Science Program Manager, NVIDIA
Ned Finkle, Vice-President, External Affairs, NVIDIA

Response to Question 1. What options should the Task Force consider for any of the roadmap elements A through I?

A. Roadmap Element A: Establish and sustain a National Artificial Intelligence Research Resource, and metrics for success.

The goal is to help extend and evolve US leadership in the development, deployment, and usage of information technology (IT) and curated data resources to support artificial intelligence (AI) research and application. In doing so, we will advance both economic and social well-being as society enters a new era in the information age, commonly referred to as the AI age. We believe new thinking should be applied to the existing federally owned or sponsored high-end computing facilities to transition these resources into the AI age. Data is considered the “fuel of AI”, lending to the idea that the curation and availability of public research data and AI models must be transformed, and availability of pre-trained models for reuse is just as important as core data sets in democratizing access to AI capabilities. AI tools and methods require hands-on usage, but are evolving rapidly, increasing the need for infrastructure and expert training. These hardware and software resources will need to be updated frequently until the rate of change stabilizes in the AI ecosystem. Therefore, a National Artificial Intelligence Research Resource (NAIRR) needs to be broadly accessible to individuals with relevant experience and training. AI tools are being applied throughout science, social science, and humanities research and development (R&D). As these tools become ubiquitous in the modern world, AI resources, education, and job-training, should be provided so individuals are equipped to leverage them in an ethical and responsible way. Undergraduate education curricula must be broadened to include introductions to data science and AI so students can achieve literacy in core concepts and the responsible consumption of AI tools.

We recommend establishing a broad, national AI research infrastructure based on a hybrid cloud strategy by integrating “on premises” AI resources, e.g., like those HPC facilities currently provisioned at federally sponsored entities, with the public cloud. NAIRR should leverage existing capabilities on premises at national facilities today, and these capabilities may need to be modernized and/or updated to support complex, data-intensive AI workloads. We recommend that the NAIRR Task Force perform an assessment of data centers to which NAIRR may have access, e.g., at universities, national laboratories, or other facilities to determine the capacity and readiness for NAIRR to potentially leverage. The Task Force should assess the need to construct new, energy-efficient data centers to support the NAIRR mission. To our knowledge, such a national assessment of datacenter capacity for research resources does not exist, and this is needed now.

NAIRR leadership needs to establish clear performance metrics for the success of a National AI Research Resource. For the priority vision of democratizing access to AI with respect to diversity, equity, and inclusion, NAIRR can demonstrate leadership by identifying clear performance targets once an established baseline is determined. Assessing the relevance of these and other performance metrics should be a regular activity for the NAIRR leadership team. Such metrics should include, but are not limited to, the following:

- Measures of data from curated datasets and pre-trained AI models, created and available in a manner to minimize bias in AI.
- The number of unique users per year, and the number of new, first-time users per year.
- User participation (i) by geography on a per population basis, (ii) by institutional type affiliation, (e.g., industry, university, national laboratory, or government agency).
- User participation by groups traditionally underrepresented in AI, or more generally science, technology, engineering, and math (STEM) disciplines.
- The number of students instructed per year.
- The number of training and education programs completed per year.
- The number of scheduled available compute hours, AI model throughput, throughput per unit electrical power expended, and total electricity use.
- The number and severity of data-security events and attacks on NAIRR.
- The number and quality of scientific publications, patents, and achievements highlights published per year.
- The number of small and medium-sized companies participating and using NAIRR.

- B. Roadmap Element B. Create a plan for ownership and administration of the NAIRR. This includes identifying (i) an appropriate agency or organization responsible for the implementation and deployment of the Research Resource; and (ii) a governance structure for the Research Resource, including oversight and decision-making authorities.

It is our judgement that a federation of new and existing resources at federal agencies and other entities is the preferred approach for implementation of NAIRR, as each agency will understand best their mission needs, research community, and data resources. There is no single agency or entity that currently possesses the necessary capabilities for NAIRR mission success. (See subsection D below for discussion of necessary NAIRR capabilities). All agencies and private entities contributing to NAIRR will need to develop internal strategies and action plans to build the requisite capabilities to sustain national leadership in AI.

We recommend a single NAIRR program office be created to provide program ownership, oversight, and administration for each of the federated agencies to ensure that the federated model runs smoothly and so each agency provides the appropriate level of support to achieve the objective of overall national AI leadership. Having one entity with oversight responsibility is imperative for NAIRR to operate successfully and for the needs of the nation to be met in a timely fashion. We believe this NAIRR office should reside within the Commerce Department, possibly within the National Institute for Standards and Technology (NIST), with dotted line connection to each of the participating agencies. (See Roadmap Element I below for brief discussion of possible agency roles.) The Commerce Department plays the key role in the National AI Initiative Act (NAII) of 2020, e.g., in establishing the National AI Advisory Committee (NAIAC) tasked with advising the President and the National AI Initiative Office on topics related to the National AI Initiative.

- C. Roadmap Element C. Develop and implement a model for governance and oversight to establish strategic direction, make programmatic decisions, and manage the allocation of resources.

NAIRR should be a federally governed and sponsored resource available for external use to advance AI research, scientific and technical knowledge, education, and workforce training:

- NAIRR should be open to all interested U.S. based users with appropriate institutional affiliation (government, private industry, higher education, national laboratories, non-profit institutions). Requirement for institutional affiliation facilitates risk management.
- User programs need to have well-defined goals and manage resources in an equitable and transparent manner.
- User fees should not be charged for non-proprietary research and if the user intends to publish the results in the open literature. NAIRR should provide sufficient resources for users to conduct work safely, ethically, and efficiently. Proprietary users are obliged to pay full-cost recovery.

Allocation mechanisms need to be deployed in alignment with National AI Initiative (NAII) goals. For programs focused on large resource consumption, a rigorous merit review of proposed work is required. For research projects funded by federal agencies, NAIRR resources may be requested through the corresponding agency sponsor program manager. For users requiring small resources, an open queuing system could suffice.

We recommend decision-making authority resides with a NAIRR Director, ideally someone who is a member of the research community and recognized for achievements in data curation and sharing, AI, HPC, and leadership of high-end HPC facilities. This director should be appointed by the Secretary of Commerce through consultation with the Director of the National Artificial Intelligence Initiative (NAII) Office. The establishment of a NAIRR Board of Directors can provide oversight on strategy, budget proposals, partnerships, and executive level hiring. We suggest the NAIRR Director appoint a Computing Program Manager and Data Program Manager to integrate and manage computing and data as interdependent but distinct programs. Additionally, two separate advisory boards – one for the social sciences and humanities community and another for the physical, biological, and

computer sciences and engineering communities, shall be established to represent their needs, requirements, and views. To engage the user community and facilitate information sharing, we recommend creating a NAIRR User Group Executive Committee. Members shall be elected to serve staggered three-year terms to preserve and propagate effectively institutional knowledge. Members should elect the User Group Chair and Vice Chair.

- D. Roadmap Element D. Establish capabilities 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.

There are at least five main capabilities required for effective operation of NAIRR: (a) a trusted steward and broker for the curation of complex public and private data sets and pre-trained, trustworthy, privacy-preserving AI models; (b) expert deployment, effective operation, and management of federated computing and data infrastructure; (c) delivery of educational tools and programs in advanced AI methods at the appropriate level (primary, secondary, undergraduate); (d) expertise in the operation of user support programs for research and education; (e) excellence in cybersecurity and AI operations, including a well-balanced security posture and risk-based vulnerability management across diverse program requirements from open to protected.

(a) Steward Data Sets and AI Models: Capabilities to serve curated datasets and pre-trained models owned by both government and non-government entities are required to extend US leadership into the AI era. It is the availability of large, curated datasets, like ImageNet¹, that added an essential element launching the current revival of AI research.² Known "reference" AI models can provide an effective floor of "best in class" to measure against and ensure an increasing level of rigor within specific domains. The primary shift in the new HPC-AI era is the importance of data integration with HPC facilities. These data resources need to be planned, curated, and made available for multiple domains using FAIR principles³ so researchers may develop and validate new AI models and methods. NAIRR should consider the merits of national standards for AI information exchange and the creation of AI tools, (e.g., recommender systems, continuous integration and testing technologies) to facilitate the quality, discovery, and sharing of research data and AI models.

(b) Operate Federated Data and Computing Infrastructure: As described in Roadmap Element A, a hybrid-cloud strategy proves to be the most efficient approach to AI infrastructure deployment. NAIRR should deploy a diversity of production-ready, proven platforms for AI, including necessary software and hardware. NAIRR platform testbeds demonstrate readiness of emerging AI technologies for future deployments. Federal agencies, e.g., National Science Foundation (NSF), Department of Energy (DOE), and Department of Defense (DOD), have expertise in managing HPC platform testbeds to advance their missions. (See Roadmap Element I.)

¹ <https://www.image-net.org/>

² Li, Fei-Fei, [How we're teaching computers to understand pictures](#), retrieved 16 December 2018

³ <https://www.go-fair.org/fair-principles/>

(c) Deliver Educational Tools and Programs: Given the ubiquitous nature of AI, and to reinforce the U.S.'s competitive position in the technology sector, there is an imperative to incorporate AI elements both technical and non-technical into U.S. primary and secondary education (K-12) curriculum, to ensure students can thrive as AI consumers, creators, professionals, and citizens. Youth need to understand how AI algorithms make predictions that impact their experience and decision-making, how this knowledge will ensure their rights are protected, and how working with AI is a viable career path. Unlike computer science (CS) education, AI education includes societal elements that can be taught in earlier years and should be woven across several curricula, including math, science, humanities. The path for NAIRR to incorporate broad AI curriculum has a strong leg up, stemming from the body of work done by several evidence-based organizations as well as after school and summer programs. Examples include AI4K12, AI4All, The AI Education Project, and the Boys and Girls Club of Western Pennsylvania's AI Pathways Toolkit. (Disclosure: NVIDIA has in the past or is currently funding these projects.) (See also Roadmap Element G.) Such programs will contribute to a national pipeline for AI literacy and ensure jobs can be filled as innovation grows.

(d) Operate User Support Programs: User support and experience is critically important and should be offered as tiers for job training, education, and research. For example, today's US federal supercomputing centers provide state-of-the-art HPC user support, such as user portals, expert liaison collaboration, and hackathon-style group learning experiences. NAIRR user support and outreach professionals must represent the diversity of our nation to advance equity and inclusion vision. (See performance metrics in Roadmap Element A.)

(e) Excellence in Cybersecurity and AI Operations: Strong capabilities in cybersecurity are necessary for the safe and secure operation of NAIRR. Given the prominent place NAIRR will play in our nation's innovation ecosystem, NAIRR will be a priority target for malicious actors. First-class capabilities for segregation of infrastructure control layers and user-control layers to enable "zero-trust" architectures⁴ should be deployed. Federal and private-sector investment to make AI models themselves robust, not just the infrastructure running AI⁵. While the AI model may be authenticated and protected from tampering, creating models that are robust to adversarial attacks requires investment in research and practice. (See also Roadmap Element F.)

E. Roadmap Element E. Implement 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.

Open access to vast sources of data available to the scientific and governmental enterprises and curated according to FAIR principles will promote development of new AI capabilities and enable AI researchers to focus on problem-solving. "Curating data in its place"

⁴ S. Rose, O. Borchert, S. Mitchell, S. Connelly. Zero Trust Architecture. NIST Special Pub. 800-207. <https://doi.org/10.6028/NIST.SP.800-207>

⁵ J.M. Wing, "Trustworthy AI," Commun. ACM, October 2021, Vol. 64 No. 10, pp 64-71. 10.1145/3448248

encourages public trust in the data quality and is consistent with the recommended federated approach to NAIRR management.

- F. Roadmap Element F. Design an assessment of security requirements associated with the National Artificial Intelligence Research Resource and its management of access controls.

In securing AI supply chains for most operating environments, it's helpful to consider datasets as code and models as binaries in evaluating control environments for AI workflows. Strong storage and provenance controls are needed to ensure no erroneous or malicious data is introduced via supply chain or data poisoning attacks^{6,7}. Operational capabilities that reproduce and explain models back to primary data becomes critical in incident response as well as advancing objectives for ethics, bias, safety, and other mission outcomes. (See Roadmap Element G). [Google's "Model Card"](#) proposal is one example how to organize the essential facts of machine learning models in a structured way. As individual AI models are composed into complex systems, model control through attestation and signatures is required throughout the ecosystem to increase confidence and trust in deployed models.

To secure deployed AI training pipelines and inference models, traditional security frameworks for core infrastructure and edge deployments like [AICPA SOC-II](#) and [NIST 8320](#) should be the norm, with additional emphasis on zero trust service design⁴ and strong data-at-rest controls to ensure multi-tenant isolation. Notably, model extraction⁸ and membership inference attacks⁹ may leak information about Intellectual Property (IP) and data used to generate a model, or the model itself, leading to IP loss and potential impact on privacy preservation within deployed models. Likewise, adversarial attacks on model decisions, and AI defenses may be used to evade or influence expected behavior of composed systems. New observations unique to AI derived from efforts such as the [Mitre AI Attack Framework](#) by leading entities like Microsoft, NVIDIA, IBM and others, should be integrated to further refine operational security for AI models.

Vigilance is required to sustain AI security. Vulnerability management and robustness will inevitably change over time as target markets change and novel attack patterns are developed. AI operations must continuously evaluate and update AI models in the field to ensure the overall health of the ecosystem. (See Roadmap Element G.) Active vulnerability management like [FIRST/Mitre](#) should be developed to enable operations teams to effectively manage risk of deployed models. Emerging technologies in confidential compute should be evaluated in testbeds to further strengthen isolation guarantees in shared environments. Standards developed by Trusted Computing Group ([TCG](#)) and Confidential Computing Consortium ([CCC](#)) that unify system level security, trusted enclaves, and operational software layers provide reasonable roadmaps for engagement.

⁶ <https://arxiv.org/pdf/2005.00191.pdf>; <https://arxiv.org/pdf/1905.13409.pdf>.

⁷ <https://www.fireeye.com/blog/threat-research/2020/12/evasive-attacker-leverages-solarwinds-supply-chain-compromises-with-sunburst-backdoor.html>.

⁸ <https://arxiv.org/pdf/1909.01838.pdf>; <https://arxiv.org/pdf/1806.05476.pdf>.

⁹ <https://arxiv.org/pdf/2007.14321.pdf>.

- G. Roadmap Element G. Deploy an assessment of privacy and civil rights and civil liberties requirements associated with the NAIRR and its research. (See also Question 2.)

Societal elements of inclusion and equity are important to address in AI education and research, especially given the potential negative aspects of AI in policing, criminal justice, financial systems, and housing discrimination, which are more likely to impact under-represented communities. The significant lack of representation in the tech industry of people who experience and can work to address the potential harms of AI applications only exacerbates this problem, so it is critically important that AI education is available to communities that have been traditionally underrepresented in technology. The task force should consider whether certain uses of AI systems should be prohibited or out-of-scope for NAIRR because they have the potential to violate privacy rights, civil rights, or civil liberties. Such uses include, but are not limited to, using facial recognition technology (FRT) to infer intent from facial expressions, using FRT to identify a person's sexual orientation¹⁰, and using FRT to identify a person's group affiliation¹¹.

Performance standards are needed for AI systems that adversely affect individual freedom. For example, AI systems that purport to predict a person's likelihood of committing a future violent crime are known to discriminate against certain groups. AI systems that do not meet a minimum accuracy should be prohibited from use by law enforcement and judges. Such models must be continuously revalidated over time to ensure they sustain predictive accuracy in the face of social, generational, evolving regulatory and other changes. Standards are needed for developing synthetic content and supporting the creation of tools for consumers to identify synthetic content.

- H. Roadmap Element H. A plan for sustaining the NAIRR, including Federal funding and partnerships with the private sector.

The predominate funding to establish and operate NAIRR will be federal funds. The private-sector financial investment in establishing NAIRR will be limited and focused, with participation likely to occur through provisioning of open software, procurements to provision AI platform infrastructure, and collaboration on grand-challenge research projects. NAIRR will be sustained through robust budget planning over a 5-year horizon. Long-term operational contracts should be established with universities, national laboratories, and federal agencies to host NAIRR infrastructure, as well as data services providers and public-cloud providers. NAIRR requires long-term access to modern datacenter infrastructure. (See Roadmap Elements A, D(d), and Question 2.) International cooperation with strategic allies, as appropriate, on the testing, evaluation, deployment, and sharing of resources for trustworthy AI systems, is also crucial to sustain NAIRR.

- I. Roadmap Element I. Implement parameters for the establishment and sustainment of the NAIRR, including agency roles and responsibilities and milestones.

¹⁰ Wang, Y., et al. (2018). Deep neural networks are more accurate than humans at detecting sexual orientation from facial images. *J. Per.s Soc. Psychol.*, 114(2), 246–257. <https://doi.org/10.1037/pspa0000098>

¹¹ Kosinski, M. (2021). Facial recognition technology can expose political orientation from naturalistic facial images. *Sci Rep* 11, 100. <https://doi.org/10.1038/s41598-020-79310-1>

NAIRR should be led by a new program office within the Commerce Department, (as indicated in Roadmap Element B), with support from federal agencies and the Networking and Information Technology R&D (NITRD) Program, Interagency Working Group for AI R&D. NAIRR should publish a report on requirements for future resources every two years and an annual self-assessment report.

Other federal agencies have competencies and core capabilities that will contribute within the NAIRR federation. The National Institute of Standards and Technology (NIST) has competency in establishing trustworthy standards for data curation and AI, which is a critical need for establishing broad acceptance of AI solutions. NIST has a strong focus on cultivating trust in the design, development, use, and governance of AI systems by (i) conducting R&D to advance trustworthy AI technologies, (ii) establishing benchmarks, (iii) developing data and metrics to evaluate AI, (iv) developing technical AI standards, and (v) engaging in discussions toward development of AI policies. The National Science Foundation (NSF) possesses leading competency in administering traditional on-premises HPC research resources with availability of public-cloud research resources. The NSF has also demonstrated leadership in a variety of core and crosscutting programs, through facilitating access to cloud computing resources through the CloudBank program, and through a variety of educational programs at all levels – pre-K through colleges and universities. The US Department of Energy (DOE) and the Department of Defense (DOD) who support HPC, data, and high-speed networking facilities. The DOE owns national supercomputing and high-speed networking user facilities and curated scientific databases operated within the DOE complex of multiprogram national laboratories^{12,13}. The DOD's HPC Modernization Program delivers world-class high performance computational capabilities to the DOD's science and technology (S&T) and test and evaluation (T&E) communities. Other agencies, such as the National Institutes of Health (NIH), the National Oceanic and Atmospheric Administration (NOAA), and the National Aeronautics and Space Administration (NASA), also operate federally owned supercomputing data facilities and should be included within the NAIRR federation.

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

NAIRR's highest and most urgent priority must be to establish itself as a trusted steward and broker for curated data sets and pre-trained AI models for use in broad areas of research and education. The complement to this is the exhibition of excellence in cybersecurity and AI operations.

Response to Question 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

¹² <https://science.osti.gov/User-Facilities/User-Facilities-at-a-Glance/ASCR>

¹³ <https://science.osti.gov/Initiatives/PuRe-Data>

gender equity, fairness, bias, civil rights, transparency, and accountability? (See also Roadmap Element G.)

- Require transparency of AI that can cause harm to humans, by prohibiting trade secret protections of AI used in law enforcement and criminal justice applications, particularly when individual freedom or equal access to resources are at stake. Adopt standards against which AI outcomes are measured. Robust transparency will include robust review, including accessibility to Freedom of Information Act (FOIA) requests.
- Bias in AI system outcomes often can be traced back to dataset inputs, in training or in production. Facilitate the creation of and access to large, clean datasets for R&D.
- Provide consistent, practical definitions of key principles, e.g., fairness, safety, accuracy, performance, etc. Create tools to assess and rate AI systems on such key principles. Require transparency through institutional review board (IRB) review and/or community stakeholder board (CSB) review/approval, as appropriate, of AI systems effect on humans.

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

- Leading edge AI platforms from US-based accelerated-computing vendors, in partnership with universities, laboratories, and facilities. An example is the University of Florida-led partnership around the HiperGator AI Supercomputer, (<https://ai.ufl.edu/>).
- Widely adopted software frameworks¹⁴ supporting AI development and applications.
- World-class datacenter capabilities at NIST, NSF, DOE, NIH, and DOD HPC Centers.
- World-class, high-speed, wide-area networking (WAN) through, DOE's Energy Sciences Network (ESNet), and other national scale WAN resources.
- World-leading, US-based cloud service providers (CSP) of compute and data infrastructure.
- Public data from federally funded research promoting FAIR principles, e.g., at Data.gov, NIH's Open Domain-Specific Data Sharing Repositories¹⁵, DOE's PURE Data Initiative¹⁶.

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

Public-private partnerships will enable NAIRR to start faster, deploying today's robust AI infrastructure platforms quickly. Federal agencies have long partnered with the private sectors in procuring and/or operating HPC research resources through management and operating (M&O) contracts. This approach has been effective in delivering world-leading capabilities by transferring selected risk management – and related risk premiums – to universities or private contractors. These represent positive exemplars available for NAIRR to model, and long-term M&O contracts should be established within the NAIRR federation as communicated in response to Roadmap Element H. Newer programs designed to make public-cloud research resources

¹⁴ <https://www.nvidia.com/en-us/ai-data-science/>

¹⁵ https://www.nlm.nih.gov/NIHbmic/domain_specific_repositories.html

¹⁶ <https://science.osti.gov/Initiatives/PuRe-Data>

available (e.g., CloudBank or Strides) indicate that long-term contracts are appropriate to establish with multiple cloud-service providers.

Significant international exemplars exist of AI research resources promoting public-private partnerships. The AI Bridging Cloud Infrastructure¹⁷ (ABCI) center in Japan is a designated AI supercomputer resource open to public and private research offering cloud access to compute and storage capacity for AI and data analytics workloads. ABCI's software environment is a container-based ecosystem with core programming and other tools as part of its standard offering. NVIDIA launched Cambridge-1¹⁸, the UK's most powerful supercomputer, enabling top scientists and healthcare experts to use the combination of AI and simulation to speed the digital biology revolution and bolster the UK's world-leading life sciences industry.

Response to Question 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?

- In the absence of a well-understood definition of success, NAIRR may be challenged over time to maintain focus on its mission. The vision to “democratize access to AI R&D” is both compelling and vague. Well defined success metrics are needed to sustain NAIRR, and, in the process broaden access of resources for the US research community.
- Growing the participation of traditionally underrepresented groups will be a challenge requiring full participation and coordination with many local, regional, and national programs. Some potential limitations include (i) lack of visibility into entire AI R&D landscape; (ii) marginalized and underrepresented communities not at the table to create the plan, and (iii) absence of an actionable pipeline framework to build AI R&D readiness. A comprehensive strategy should be created in which NAIRR has specific, coherent actions to achieve the program targets in these, but with underrepresented and historically marginalized communities included up front and contributing to that plan.
- The user programs and communities for NAIRR are not well defined or understood. NAIRR should create well-defined user programs with public outreach for participation to bring focus and clarity to the community.
- The US is limited in access to science data repositories provisioned to support AI. A high priority is to leverage existing science-data-generating programs to produce world-leading data repositories ready for use in building AI models. These data repositories should be created in collaboration with the subject-matter experts and in adequate proximity to HPC resources.

¹⁷<https://abci.ai/>

¹⁸ <https://blogs.nvidia.com/blog/2021/07/07/ceo-unveils-cambridge-1/>