



U.S. National
Science Foundation



TWO YEARS IN NATIONAL AI RESEARCH RESOURCE **NAIRR**

*A National Public-Private Infrastructure
Advancing AI Innovation & Leadership*

Led by the U.S. National Science Foundation



TABLE OF CONTENTS

Executive summary..... **3**

Artificial intelligence: A national imperative **4**

The NAIRR Pilot **4**

Accelerating AI discovery, education and innovation **8**

 Highlight: Training the next generation..... **9**

 Highlight: Enabling research across the scientific spectrum **10**

NAIRR Secure **13**

NAIRR Pilot community datasets **13**

Capitalizing on achievements and looking forward..... **14**

Appendix 1. NAIRR Pilot federal and non-governmental partners and contributors..... **16**

Appendix 2. NAIRR Pilot organization and agency investments **21**

Appendix 3. NAIRR Pilot community datasets **29**





EXECUTIVE SUMMARY

To drive global dominance in artificial intelligence, U.S. researchers, educators and innovators must have access to state-of-the-art AI capabilities. The National Artificial Intelligence Research Resource (NAIRR), led by the U.S. National Science Foundation, serves as a national infrastructure designed to provide the research and education communities with critical computing, software, data, models, educational resources and expertise needed to advance AI innovation and discovery in the U.S.

Launched in January 2024 as a public-private partnership, the NAIRR has demonstrated the feasibility and value of a coordinated national AI infrastructure. Over its two-year pilot phase, the NAIRR has:

- Partnered with **14 federal agencies** to provide access to federally funded data, computing, models and educational materials, avoiding duplication of efforts.
- Leveraged approximately **\$100 million worth of in-kind contributions from 28 private sector partners**.
- Supported over **600 research and education projects**, benefiting over 6,000 students across all 50 states, the District of Columbia and Puerto Rico.
- Advanced **high-priority AI and AI-driven scientific research**.
- Funded **demonstration projects** that innovatively integrated or relied upon NAIRR capabilities.
- In partnerships with the National Institutes of Health and the U.S. Department of Energy, supported **NAIRR Secure** demonstration projects that explored the use of sensitive data within secure computing environments while preserving privacy and security.
- Funded **expansion projects** that delivered outreach, training and support to communities new to AI research and education.
- Curated **community-contributed datasets** to strengthen the nation's AI-ready workforce.

The pilot phase of NAIRR validated the importance of a coordinated national AI infrastructure and was recognized in the White House *Science and Technology Highlights: Year One* report. Building on this success, and consistent with the [White House AI Action Plan](#), NSF has released an opportunity to establish the [NAIRR Operations Center](#) to transition NAIRR from a pilot to a sustained national capability.

The next phase of NAIRR will establish a lean, long-term operational structure that, through continued partnerships with leading technology companies, expands access to world-class resources and creates practical pathways for researchers, startups and small businesses to participate in a competitive, open and innovative AI ecosystem, directly advancing the AI Action Plan's directive to foster open, innovative and competitive AI development.



ARTIFICIAL INTELLIGENCE

A NATIONAL IMPERATIVE

Artificial intelligence is rapidly moving from the research lab to everyday life. Its impact can be seen in nearly every discipline, from health care to transportation, finance to education. Industry investments have revolutionized AI at scale by capitalizing on discoveries and innovations produced by thousands of participants across the U.S. research and education community. To ensure global dominance in AI and unlock the enormous opportunity for AI across the economy and society, the United States needs to accelerate foundational research, fundamental breakthroughs and production of new applications that can rapidly translate to economic growth and output — and expand the skilled AI workforce to do so.

To drive AI innovation, discovery and national competitiveness, U.S. researchers, educators and innovators must be empowered with state-of-the-art AI capabilities — computing, data, models, software, tools and expertise. Across the AI research community, demand for these critical resources has exploded and continues to grow. Government and industry investments are each important, but alone will not be sufficient to address the enormous range of critical AI research and innovation questions, challenges and applications. Strong and organized collaboration among the public, private, nonprofit and academic sectors is critical to achieving and maintaining U.S. leadership and dominance in AI.

THE NAIRR PILOT

Goals. Responding to this imperative, the National AI Research Resource (NAIRR) Pilot¹ was a pioneering public-private initiative to jump-start a competitive national AI ecosystem for discovery and innovation. The NAIRR vision (see Figure 1) is to provide U.S. researchers and educators with unified "one-stop" access to critical AI resources to advance the frontiers of AI and train the AI workforce of the future. The NAIRR Pilot aims to build the foundations and pave the way to realizing a long-term national capability that cements U.S. leadership in AI research and development. The principal goals of the NAIRR Pilot are to demonstrate the feasibility of the NAIRR concept and major operational elements; establish strong public-private partnerships that bring together all the critical national AI resources, capabilities and expertise; employ those resources to immediately drive AI innovation and workforce development; and ultimately inform preparations for a long-term NAIRR.



¹ More background on the NAIRR Pilot available at: <https://new.nsf.gov/focus-areas/artificial-intelligence/nairr>.

The NAIRR Vision

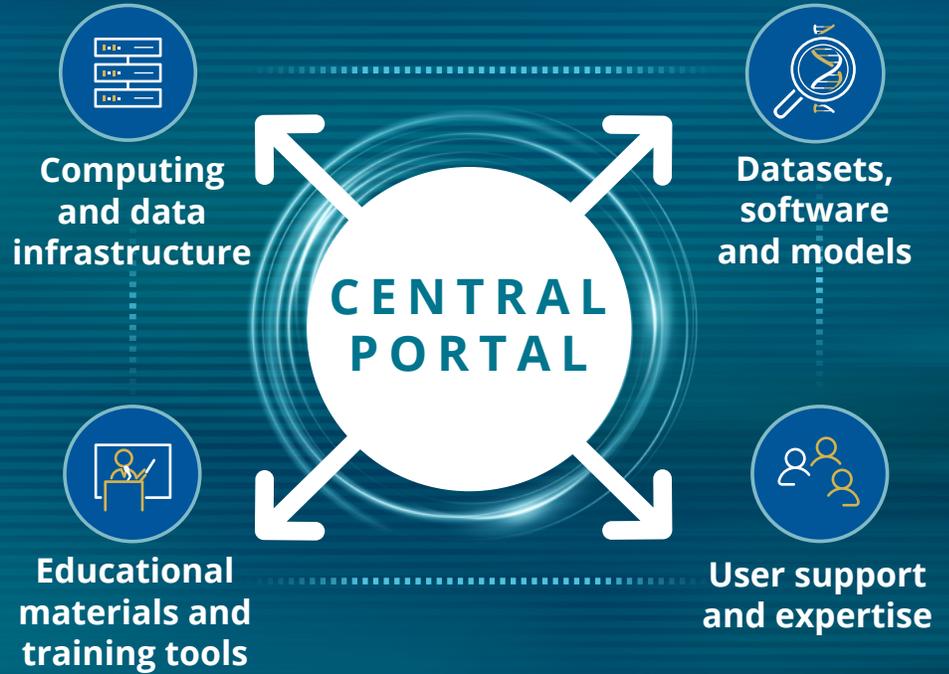


Figure 1: The NAIRR vision.



A public-private consortium. The NAIRR Pilot was formally launched in January 2024 by the U.S. National Science Foundation in partnership with federal agencies, private-sector and nongovernmental organizations, and has since grown into a consortium comprising **14 federal agencies and 28 industry and nonprofit partners.** Together, these partners contribute a unique and heterogeneous array of the most advanced AI resources, including high-performance and cloud computing, models, data resources, software licenses and other tools and capabilities for use by the U.S. research and education communities. **The total value of in-kind pilot contributions from the private sector is estimated at \$100 million, and additional partners continue to join.**

This includes significant computing resources and support from NVIDIA, Microsoft, AWS, IBM and OpenAI, as well as capabilities and expertise from organizations such as the Allen Institute for Artificial Intelligence, Hugging Face, Vocareum, SambaNova, OpenMined, Cerebras and others. See Appendix 1 for a complete list of partners and contributions.

Government-supported resources include NSF-supported advanced computing systems and cloud systems at The University of Texas at Austin, the University of Illinois Urbana-Champaign, UC San Diego, Carnegie Mellon University, the University of Pittsburgh, Texas A&M University, Purdue University and Indiana University. National Institutes of Health-supported capabilities include CloudLab, and Department of Energy-supported capabilities include the Summit system at Oak Ridge National Laboratory and the AI test bed at Argonne National Laboratory (see Figure 2).



NSF SUPPORTED RESOURCES



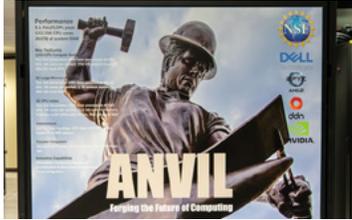
RESOURCE:
ACES

INSTITUTION:
Texas A&M University



RESOURCE:
FABRIC

INSTITUTION:
University of North Carolina at Chapel Hill



RESOURCE:
Anvil

INSTITUTION:
Purdue University



RESOURCE:
NSF Frontera / Lonestar6 / Vista

INSTITUTION:
University of Texas at Austin



RESOURCE:
Bridges-2 / Neocortex

INSTITUTION:
Carnegie Mellon University and University of Pittsburgh



RESOURCE:
Jetstream 2

INSTITUTION:
Indiana University



RESOURCE:
Chameleon

INSTITUTION:
University of Chicago



RESOURCE:
PATH

INSTITUTION:
University of Wisconsin



RESOURCE:
CloudBank

INSTITUTION:
University of California San Diego



RESOURCE:
Prototype National Research Platform

INSTITUTION:
University of California San Diego



RESOURCE:
Delta / Delta AI

INSTITUTION:
University of Illinois Urbana Champaign



RESOURCE:
Sage

INSTITUTION:
Northwestern University



RESOURCE:
Expansive / Voyager

INSTITUTION:
University of California San Diego



RESOURCE:
National Secure Data Service Secure Compute Environment

INSTITUTION:
National Center for Science and Engineering Statistics

Figure 2: Contributed resources available for request during the NAIRR Pilot.





U.S. DEPARTMENT
of **ENERGY** | Office of
Science



RESOURCE:
ALCF AI Testbed

INSTITUTION:
*Argonne National
Laboratory*



RESOURCE:
Summit

INSTITUTION:
*Oak Ridge National
Laboratory*



RESOURCE:
Cloud Lab

**National Institutes
of Health**

PRIVATE SECTOR RESOURCES

- Amazon Web Services
- Anthropic (Model API)
- Cerebras (CS-2 AI Accelerator)
- Databricks (Data Intelligence Platform)
- Eleuther AI
- Google Cloud Platform
- Groq (LPU Inference Engine)
- Hugging Face (Spaces and Compute Grants)
- IBM & the AI Alliance
- Lexset (Seahaven)
- Microsoft Azure
- NVIDIA (DGX Cloud)
- OpenAI API
- OpenMined
- SambaNova (SambaCloud)
- Voltage Park
- Vocareum AI Notebook
- Weights & Biases

Figure 2 (continued): Contributed resources available for request during the NAIRR Pilot.

Partnerships across the public and private sectors were recognized as crucial to the success of the pilot, and major efforts were made to attract, convene and marshal the full breadth of American AI capabilities and expertise to establish the NAIRR Pilot. This effort resulted in broad participation of federal science and technology agencies and the U.S. AI tech industry.

Because the pilot was designed to move quickly and remain flexible, partners contributed resources and expertise voluntarily, without exchanging funds. NSF provided funding for core operational teams, community workshops and events, while participating agencies contributed staffing for joint governance and technical activities. In addition, NSF and several other agencies are making parallel investments through existing programs that contribute directly and indirectly to the pilot's success (see Appendix 2 for details on governance, operations and agency investments).



ACCELERATING AI DISCOVERY, EDUCATION AND INNOVATION

As demonstrated through its pilot phase, NAIRR has rapidly achieved significant scientific and educational impacts, while building the foundation for a sustainable, long-term national capability. These advances were made possible by access to NAIRR's uniquely broad array of public and private-sector AI resources.

Over its first two years, the pilot has accelerated AI innovation and research that would not otherwise have been possible, accelerating progress across scientific domains and supporting AI-ready students, educators and practitioners nationwide. Demand for NAIRR resources remains high and continues to grow.

Today, **NAIRR supports over 600 U.S. research and education projects across all 50 states, Washington, D.C. and Puerto Rico** (see Figure 3), with participation increasing rapidly.



Figure 3: Geographic distribution of NAIRR projects, including research and education resource allocations via NAIRR Research and NAIRR Classroom, respectively, as well as funded demonstration and expansion (Demo/Expansion) projects.

In addition to its geographically distributed national impact, the pilot has supported research and education across a wide variety of institution types and sizes. Users come from a range of institutions, including major research universities that grant many doctoral degrees and conduct extensive research (R1 and R2 universities), smaller master's colleges and universities, associate's colleges, nonprofits, federally funded research and development centers and small businesses/startups. Pilot resources were provided by partnering U.S. federal agencies including NSF, the National Institutes of Health (NIH), National Oceanic and Atmospheric Administration, Department of Energy (DOE), Department of Education, Department of Transportation, Department of War, Defense Advanced Research Projects Agency, Office of Naval Research, Army Research Office, NASA, United States Department of Agriculture (USDA), and Department of Veterans Affairs, as well as from private foundations and nonprofits.

The pilot's extensive set of computing and other AI resources enabled projects of many different scales and types, supporting a broad spectrum of research and educational activities nationwide.

High-impact AI-empowered research projects span a wide range of domains and disciplines. These include fundamental AI research on model robustness and interpretable AI, privacy-preserving methods and cybersecurity, as well as scientific applications of AI methods to advance areas such as farming automation and pest detection, biomedical technology and diagnosis, personalized health care and autonomous navigation.



HIGHLIGHT

TRAINING THE NEXT GENERATION

Through the **NAIRR Classroom**, the pilot is helping train the future AI workforce by enabling educators and students to access pilot resources in their AI class curricula. NAIRR Classroom resources include:

- Vocareum's notebook and Learning Management System.
- GPU, CPU and Jupyter notebook resources through NSF Jetstream 2 and NSF Prototype National Research Platform.
- Commercial cloud via NSF CloudBank access on AWS, Google Cloud Platform and Microsoft Azure.
- CloudLab computing and data resources from NIH for the biomedical community.

The NAIRR Classroom has supported over **60 educators**, and including NAIRR research projects is reaching approximately **6,000 students nationwide** with hands-on experience in advanced AI technologies.

The pilot has also supported outreach efforts via the NAIRR Expansion initiative, including **seven conferences across the U.S.** coordinated by the Computing Research Association to bring together educators working to integrate AI into classrooms (see Table 1).

Educator	Course	Institution	State	# of Students
Chukwu, C.	<i>Computational Method for Applied Mathematics and Data Science</i>	Georgia Southern University	GA	150
Jones, W.	<i>Hands-on Applied Machine Learning</i>	Coastal Carolina University	SC	30
Zhou, J.	<i>Deep Learning</i>	Northern Kentucky University	KY	25
Zeng, D.	<i>Artificial Intelligence</i>	Dakota State University	SD	150
Chang, S.	<i>AI Models for Advanced Quality Engineering Applications</i>	Kansas State University	KS	10
Fernandes, S.	<i>Building Generative AI Applications</i>	Creighton University	NE	25
Gupta, V.	<i>Biomedical Data Science</i>	Meharry Medical College	TN	70
Bosco, B.	<i>Machine Learning</i>	Southern Oregon University	OR	20
Lobaton, E.	<i>Deep Learning</i>	North Carolina State	NC	700
Rahimi, S.	<i>Machine Learning for Cybersecurity</i>	University of Southern Mississippi	MS	30

Table 1: Example of NAIRR Classroom projects.

HIGHLIGHT

ENABLING RESEARCH ACROSS THE SCIENTIFIC SPECTRUM

From a scientific standpoint, NAIRR's extensive public-private partnerships have attracted high user interest and naturally led to unique opportunities for researchers. High-impact AI-empowered research projects made possible by NAIRR span a wide range of domains and disciplines, including fundamental AI research on model robustness and interpretable AI privacy-preserving methods, and cybersecurity. The application of AI methods is also advancing other areas such as farming automation and pest detection, biomedical technology and personalized health care, and autonomous navigation.

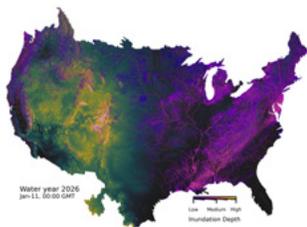
Figure 4: Examples of NAIRR Pilot supported research and education projects.



AI-Driven Decision Support in Battlefield Medicine

Kimberly Elenberg, Carnegie Mellon University, trained real-time heart rate estimation algorithms to develop autonomous systems that combine robotics and AI to detect vital signs and injury severity, enabling rapid remote battlefield triage.

Resource: *NSF Bridges-2/GPU, Pittsburgh Supercomputing Center.* [NAIRR240073](#)



National Hydrology AI Digital Twin Pilot

Laura Condon, University of Arizona, and Reed Maxwell, Princeton University, piloted the first AI-powered digital twin of the water system to deliver near-real-time forecasts of flood, drought, surface water and groundwater to help communities and emergency planners make better decisions.

NAIRR Pilot Demonstration Project award NSF-2134892 (supplement).

Resources: *NSF Jetstream2 at the University of Indiana and NSF Anvil at Purdue University.* [NAIRR240400](#)



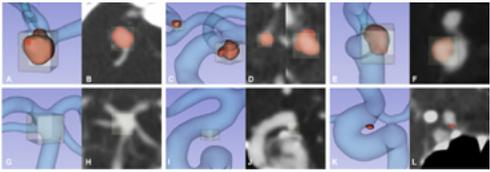
Training Vision-Language Models for Agricultural Resilience

Baskar Ganapathysubramanian, NSF-USDA AIIRA AI Institute, Iowa State U, created a foundation model for agricultural pest detection for farmers.

Resources: *NVIDIA DGX Cloud and NSF Frontera at The University of Texas at Austin.* [NAIRR240277](#)



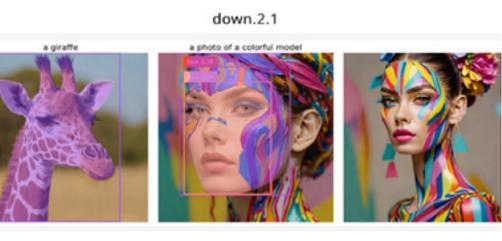
Figure 4 (Continued): Examples of NAIRR Pilot supported research and education projects.



AI-Based Detection of Intracranial Aneurysms in 3D Brain Scans

Huaizu Jiang, Northeastern University, developed a 3D, multi-scale AI model to detect tiny, hard-to-spot aneurysms while reducing false alarms.

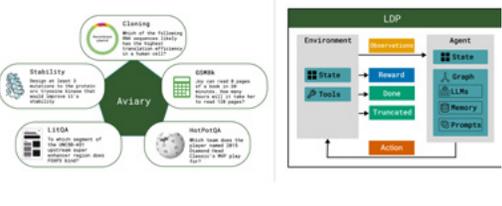
Resources: *NVIDIA DGX Cloud and NSF Frontera at The University of Texas at Austin.* [NAIRR240236](#)



Using the National Deep Inference Fabric to Study AI models

David Bau, NSF National Deep Inference Fabric, is using NAIRR resources to develop the NSF National Deep Inference Fabric, a shared infrastructure that will allow researchers to study very large AI models without each one needing their own expensive setup, reducing costs and making this kind of research possible.

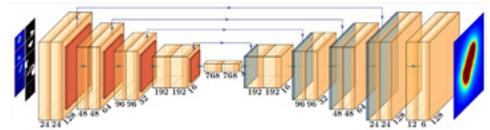
Resource: *NSF DeltaAI at the University of Illinois Urbana-Champaign.* [NAIRR250025](#)



Scaling Laws for Scientific Agents

Andrew White, FutureHouse, is using NAIRR resources to train AI agents on real scientific challenges like protein engineering and DNA construct manipulation.

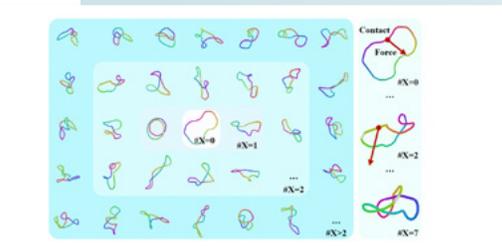
Resources: *Microsoft Azure and the NVIDIA DGX Cloud.* [NAIRR240183](#)



Automated Brachytherapy Planning for Cervical Cancer

Lance Moore, UC San Diego Health, is developing AI models to create precise and safe cervical cancer radiation treatment plans from pre-treatment images by predicting how radiation will affect surrounding tissue in 3D.

Resource: *NSF Voyager at UC San Diego.* [NAIRR250110](#)



KnotGym: A 3D Simulation Environment for Spatial Reasoning

Zizhao Chen, Cornell University, created KnotGym — a 3D virtual environment designed to test AI's ability to understand and manipulate physical objects.

Resources: *NSF Frontera at The University of Texas at Austin, NSF Delta at the University of Illinois Urbana-Champaign.* [NAIRR240194](#)



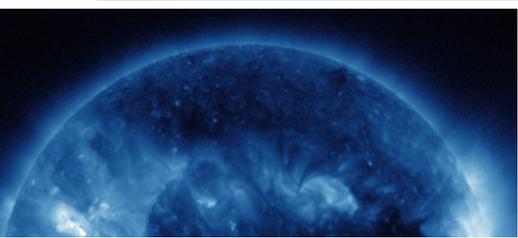
Figure 4 (Continued): Examples of NAIRR Pilot supported research and education projects.



LakeGPT: Building a Foundation Model for Aquatic Sciences

Anuj Karpatne, Virginia Tech, is using NAIRR resources to develop Lake-GPT, an AI foundation model that combines lake sensor data with established scientific knowledge to better understand and predict water quality in lakes and reservoirs.

Resources: Summit at DOE's Oak Ridge National Laboratory and NSF Bridges-2 at PSC [NAIRR240161](#)



AI Foundation Model for Space Weather Applications

Sujit Roy, University of Alabama in Huntsville, trained a foundation model on NASA's Solar Dynamics Observatory for the study of solar dynamics at many spatial and temporal scales.

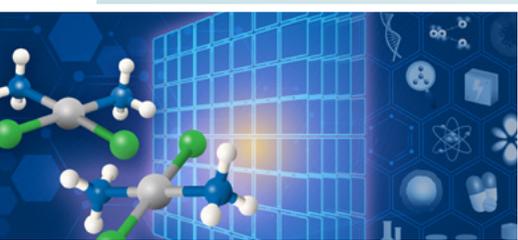
Resource: NVIDIA DGX Cloud. [NAIRR240178](#)



Building Reliable and Secure AI Scientific Applications

Wenqian Dong, Oregon State, developed ways to accelerate the training and inference of LLM.

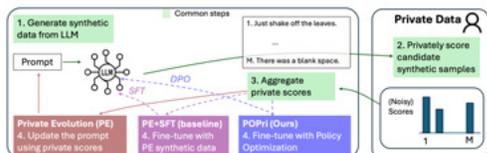
Resource: Cerebras CS-2 system on the AI test bed at DOE's Argonne National Lab. [NAIRR240028](#)



Scaling Foundation Models for Molecular Chemistry

Venkat Viswanathan, University of Michigan, trained foundation models to accelerate the design and discovery of materials for energy storage applications.

Resource: NVIDIA DGX Cloud. [NAIRR240190](#)



Training Foundation Models from Private Federated Client Data

Giulia Fanti, Carnegie Mellon, is using NAIRR resources to develop and test privacy-preserving AI techniques that train large foundation models on decentralized, private data.

Resource: Delta at NCSA. [NAIRR240010](#)



NAIRR SECURE

The NAIRR Secure Pilot was established to enable AI research involving sensitive data requiring protections such as secure computing environments. The NAIRR Secure Pilot interagency federal working group, co-led by the U.S. Department of Energy (DOE) and the National Institutes of Health (NIH), aims to:

- Refine the requirements and infrastructure design patterns for the future NAIRR Secure resources, including integration of diverse and sensitive resources.
- Investigate new opportunities for combining data in ways that preserve privacy and security.
- Identify research and training use cases to inform the development of future NAIRR Secure capabilities.

From the inception of this effort, the federal agency leads and partners for the NAIRR Secure Pilot, in close coordination with the leads for the overall NAIRR Pilot, adopted a goal of sharpening their shared understanding of not only the technical barriers for AI projects with sensitive data, but also the policy and sociological barriers. To this end, the NAIRR Secure Pilot is soliciting perspectives and assembling lessons learned relevant to the assessment and management of risks and opportunities in the rapidly evolving AI landscape.

NAIRR PILOT COMMUNITY DATASETS

Community datasets. A data and models working group under the NAIRR Pilot Steering Committee developed policies and processes for including nongovernmental community datasets in the pilot. NSF opened an opportunity for the community to suggest public datasets and other resources that would facilitate AI education and researcher skill development. The submissions were evaluated based on their AI use case for newcomers to AI, the user community that would be served or enhanced with inclusion in the NAIRR Pilot, adequate metadata and documentation for newcomers use the dataset, available user support and training and whether appropriate data policy procedures had been followed.

The following first cohort of datasets was announced in August 2025 and is being made available through the pilot. The datasets range from smaller image recognition datasets, such as the [Shipwrecks dataset](#) from the University of Michigan, to larger collections, such as the [microbiome dataset](#) from pregnant women with preterm births from the University of California San Francisco. This initial step piloted the feasibility of how to include and evaluate datasets and how to more fully integrate them into the NAIRR ecosystem in the future. A full listing of the initially included datasets is provided in Appendix 3.



CAPITALIZING ON ACHIEVEMENTS AND LOOKING FORWARD

VALIDATION OF THE NAIRR CONCEPT

The NAIRR Pilot has rapidly demonstrated the potential of a coordinated, public-private national infrastructure to power AI discovery and innovation, setting the foundation for continued U.S. global leadership. From an operational standpoint, the pilot has rapidly demonstrated the feasibility of a coordinated and impactful national AI capability.

- The pilot has demonstrated high and wide-ranging demand across AI research disciplines and across fundamental and applied research for the array of pilot AI resources, as well as services including user training and support. Demand for additional resources that support secure and privacy-preserving workflows is also strong.
- The pilot has pioneered the establishment of a strong public-private consortium on national AI infrastructure, bringing together many federal agencies and nongovernmental partners representing the breadth of the U.S. AI tech industry.
- The public-private collaboration model has been extremely advantageous in delivering a richer array of resources and expertise compared to a public-only model. At the same time, the government plays critical roles – as a convener bringing together a complex array of contributed public and private federated resources using a nimble "all hands in" model, and as a funder, able to leverage existing investments in operational systems and support new pilots, and prototype systems to build out new innovative AI pathways.
- The light-weight governance structure (see Appendix 2) has allowed the pilot to rapidly integrate and make available a unique rich array of critical AI resources for researchers and educators, including cloud and high-performance computing, software and data platforms, datasets, models and educational materials and tools, and AI expertise.
- NSF and partner agencies have successfully leveraged existing investments in cyberinfrastructure capabilities and expertise to rapidly deploy a "one-stop" central portal and user-facing request programs, support, educational resources and community-building opportunities, resulting in the rapid impacts on AI science and innovation discussed above.
- The pilot has nurtured an emerging NAIRR community of researchers, educators, resource providers and operations teams, who are accelerating AI research and innovation, pioneering the use of AI across research domains and disciplines, and creating continent-spanning federated AI infrastructure to enable and support these activities.



PREPARING TO SCALE TO A LONG-TERM NATIONAL CAPABILITY

Transitioning from a pilot to a sustainable national asset that drives U.S. dominance in AI will require more capabilities as well as streamlined and automated operational processes that continue to reduce "time to science." The experience to date from the pilot has highlighted several key opportunity areas that can guide the transition to the full NAIRR:

- **Prioritization.** National goals for AI strongly inform NAIRR. Ensuring alignment of goals with the new administration's AI priorities and the White House AI Action Plan will be a key effort.
- **Coordination.** Intensive coordination has been key to connecting partners, resources and users to achieve success, which, in turn, requires sustained participation from experienced leaders and staff on all sides.
- **Expanding data access, integration and readiness.** Data has emerged as a critical challenge and time-absorber for much AI-driven research and education. There is great urgency to expand access to a broader suite of trusted AI-ready data, tools and platforms for integrating data and preparing data for AI use.
- **Increasing support for the full AI research life cycle.** Researchers and educators require an integrated infrastructure for software and data staging, workflow management, tools to rapidly connect data and computing in ways that preserve privacy and security, and results analysis. Modern workflows that can integrate across myriad users, resource providers and systems are needed — balancing interoperability, security and user experience.
- **Improved user support, training and community building.** More support, tools, and improved workflows are needed for user onboarding, as well as expanding outreach, education and training for researchers and educators to effectively use AI resources. This includes transparently demonstrating the end-to-end utility of NAIRR to build community and reach potential users.
- **Skills training.** The demand for skills training on how to use newly available NAIRR resources is significant and crucial to the effective use of resources, and must be greatly expanded to achieve the full potential and impact of a national AI capability.
- **Measuring impact and success.** Appropriate outcome goals and assessment metrics for a full NAIRR and associated data-gathering mechanisms are critical from the outset and continuing throughout in an integrated manner.

Accordingly, the transition from the pilot builds on successes and lessons to reinforce the public-private partnerships, expand opportunities for AI research and education, and streamline operations to prepare for the opportunity to scale up to a more permanent national AI asset. A lean, sustainable operational capability will be vital for implementing and scaling out these opportunities. NSF is running a competition for such capability through the NAIRR Operations Center funding opportunity (NSF 25-546).

Conclusion

In conclusion, across the AI research community, demand for access to compute, data, models and user training is high and growing. A scaled NAIRR, made possible through public-private partnerships, can unlock the ingenuity of America's research community by providing access to the necessary resources that enable critical work on foundational elements of AI and applications of AI across scientific domains. The NAIRR Pilot has successfully set the foundations and pathways to achieve the vision for a full-scale national asset that will drive regional and national AI innovation, train the next generation AI-ready workforce, and ultimately spark new solutions, products, startups and jobs that will drive economic growth.



APPENDIX 1

NAIRR PILOT FEDERAL AND NONGOVERNMENTAL PARTNERS AND CONTRIBUTORS

(See [NSF NAIRR Pilot website](#) for details of each contribution).

FEDERAL AGENCY PARTNERS

U.S. National Science Foundation

Defense Advanced Research Projects Agency

U.S. Department of War

U.S. Department of Education

U.S. Department of Energy

Department of Veterans Affairs

National Aeronautics and Space Administration (NASA)

National Institutes of Health

National Institute of Standards and Technology

National Oceanic and Atmospheric Administration (NOAA)

U.S. Department of Agriculture

U.S. Food and Drug Administration

U.S. Geological Survey

U.S. Patent and Trademark Office

PRIVATE SECTOR PARTNERS

Allen Institute for Artificial Intelligence

Amazon Web Services

Anthropic

AMD

Cerebras

Databricks

Datavant

EleutherAI

Google

Groq

Hewlett-Packard Enterprise

Hugging Face

IBM

Intel

Meta

Microsoft

MLCommons

NVIDIA

Omidyar Network

OpenAI

OpenMined

Palantir

Regenstrief Institute

SambaNova Systems

Vocareum

Voltage Park

Weights & Biases



FEDERAL AGENCY PARTNERS

AGENCY	CONTRIBUTION
U.S. National Science Foundation	Lead agency, coordinating the operations of the NAIRR Pilot. In addition, NSF will provide allocations on NSF-supported supercomputing systems, AI test beds and access to advanced networking, data systems and datasets. The National Center for Science and Engineering Statistics, a principal statistical agency within NSF, will contribute resources and approaches from the National Secure Data Service Demonstration project, in addition to data and statistical expertise, to inform the NAIRR Pilot.
Defense Advanced Research Projects Agency	Make a set of open-source tools and environments available to the NAIRR Pilot user community, including those that support assurance of autonomous systems, development of machine common sense capabilities, and robustness against adversarial attacks.
U.S. Department of War	Contribute expertise to the management and allocation of NAIRR Pilot computing resources, as well as a responsible AI toolkit to assist NAIRR users in crafting responsible approaches to AI research.
U.S. Department of Education	Furnish education research and statistics datasets, while also providing expertise in the development and project management of foundational models. This support aims to aid the NAIRR Pilot user community in accessing AI model development resources.
U.S. Department of Energy	Provide allocations on the Summit pre-exascale supercomputer as well as access to AI test beds, training resources and associated expertise, and will lead the NAIRR Secure Pilot effort together with the National Institutes of Health.
Department of Veterans Affairs	Support the NAIRR Pilot through its National AI Institute and AI Network, including providing expertise (i.e., training and governance), best practices supporting trustworthy AI, lessons from AI pilots, as well as experience in supporting AI for health use cases. VA will collaborate on NAIRR Secure to explore health-related efforts most important to jointly serving the needs of veterans and Americans.
National Aeronautics and Space Administration (NASA)	Provide datasets to support AI model development, contribute expertise in the development of foundational models, and offer hands-on tutorials to assist the research community in fine-tuning and using these models.
National Institutes of Health	Co-lead the NAIRR Secure Pilot effort with DOE and provide open and controlled access to NIH computing and data platforms and biomedical datasets to support health care-focused AI research.
National Institute of Standards and Technology	Provide expertise on best practices for AI governance, risk management, and testing and evaluation for responsible AI to aid in the management of the NAIRR Pilot.
National Oceanic and Atmospheric Administration (NOAA)	Make datasets of interest available to the NAIRR user community, leveraging the NOAA Open Data Dissemination program.
U.S. Department of Agriculture	Make the forthcoming 2022 Census of Agriculture dataset available and accessible to the research community through the NAIRR Pilot infrastructure.
U.S. Food and Drug Administration	Make AI-specific test methods, datasets and evaluation methodologies available to the NAIRR community, via the Office of Science and Engineering Laboratories (OSEL) in the Center for Devices and Radiological Health (CDRH). These tools, developed by OSEL's AI program, are intended to aid researchers and innovators with safety, effectiveness and performance assessment, in both premarket and real-world settings, of novel AI solutions.
U.S. Geological Survey	Make datasets supporting AI model development available to the NAIRR Pilot user community.
U.S. Patent and Trademark Office	Provide access to rich datasets for AI training and support public challenges to spur the development of novel data products.



PRIVATE SECTOR PARTNERS

Organization	Contribution
Allen Institute for Artificial Intelligence	Access to a fully open ecosystem of data, models, training and evaluation software necessary to support a scientific approach to AI research. Initial access will be provided to AI2 Dolma, the largest open dataset to support language model pre-training.
Amazon Web Services	Support for at least 20 research projects through access to AWS credits for storage, compute, and AI services to build, train and deploy machine learning (ML) models. AWS will also work with the NAIRR Pilot to publicize AWS resources to accelerate AI research, including access to pre-trained and customizable AI/ML models, AI/ML training resources and 500+ datasets through the Registry of Open Data on AWS.
Anthropic	API access to Anthropic's Claude model for 10 researchers working on climate change-related projects. Anthropic will also provide educational resources to help those researchers experiment with prompt engineering.
AMD	Application support for NAIRR Pilot users running applications on AMD hardware and collaboration with cloud vendor sites offering access to AMD hardware.
Cerebras	Access to Cerebras systems and clusters, providing up to four EXAFLOPs of AI compute for NAIRR Pilot projects and users and enabling them to rapidly train AI models. Cerebras will also contribute access to open-source datasets and models, and time from its expert data scientists and AI researchers to facilitate project selection, definition and success.
Databricks	Access to Databricks' Data Intelligence Platform for NAIRR Pilot users. This will facilitate the use of Databricks' data processing tools by the research community to analyze existing datasets and create entirely new ones.
Datavant	\$2.6 million in the form of access to Datavant's privacy-preserving record linkage platform, privacy-preserving data discovery tools, and staff expertise in support of the NAIRR Secure NIH component of the pilot, as well as elements of the future NAIRR software stack.
EleutherAI	100,000 hours of GPU access to support an effort in the research community to train a foundation model for scientific research, in addition to support and expertise for using EleutherAI's large language model training library on high-performance computing systems.
Google	Collaboration across Google Colab, Kaggle and Data Commons programs, including licenses for Colab's virtual notebooks, integration of Kaggle public resources onto NAIRR Pilot infrastructure, and partnership on competitions and red teaming of models. In addition, Data Commons will co-locate an instance of Data Commons with the NAIRR Pilot computing infrastructure to facilitate the ability of the research community to use its diverse, integrated datasets.
Groq	Access for up to 10 research teams to use Groq's language processing unit Inference Engine via GroqCloud.
Hewlett-Packard Enterprise	<i>Proposed contribution:</i> Time on GPU-powered supercomputing platforms and discounts on supercomputing resources for potential expansion of the NAIRR Pilot. In addition, HPE will provide licenses to HPE Machine Learning Development Environment and HPE Machine Learning Data Management Software, along with hands-on training for researchers who will have access to datasets, digital twins and performance and productivity tools.



PRIVATE SECTOR PARTNERS

Organization	Contribution
Hugging Face	100 compute grants for NAIRR Pilot projects and participants to support access to Hugging Face Spaces demos of systems and model evaluation, inference and fine-tuning. Hugging Face will also partner with the NAIRR Pilot to set up sharing and evaluation leaderboards for datasets and models developed through or hosted by the pilot.
IBM	Datasets and benchmarks focused on AI safety and trust evaluation, as well as geospatial, time-series, materials and chemistry foundation models. IBM will also provide expertise and assistance to researchers working with these resources.
Intel	Technical training on Intel server platforms, AI technologies and software optimization for NAIRR Pilot users working with Intel hardware.
Lexset	Access to Lexset's Seahaven synthetic data creation software and tools. Lexset will also provide staff with time and expertise to support research community data creation projects.
Meta	Collaboration with NAIRR Pilot researchers to support research on Meta's Llama suite of models, consistent with applicable model licenses.
Microsoft	\$20 million in compute credits on Microsoft Azure, along with access to leading-edge models, including those available via Azure OpenAI Service. Availability of state-of-the-art resources for developing trustworthy and responsible AI applications, including tools for research and development on AI fairness, accuracy, reliability, transparency, privacy and security, and model orchestration. Offerings include resources to enable HIPAA-compliant computing in support of health care research, access to innovative tools for scientific discovery through Azure Quantum Elements, and opportunities to forge collaborative relationships with Microsoft's scientists and engineers.
MLCommons	Access to the MLCommons technology platform to enable testing of AI systems, as well as access to AI benchmarks and the suite of MLPerf training, inference and storage benchmarks. MLCommons will also provide hosting services for select open datasets developed by the NAIRR Pilot user community.
NVIDIA	\$30 million in overall support for the pilot, including \$24 million worth of computing on NVIDIA's DGX Cloud platform integrated with NVIDIA AI software tools and supported by technical subject-matter experts to assist NAIRR Pilot users. In addition, NVIDIA will provide AI software platform licenses to national supercomputing centers integrated with the NAIRR Pilot and run deep-learning workshops, AI boot camps and AI hackathons for NAIRR Pilot users.
Omidyar Network	\$500,000 in support of the pilot effort, which includes computing infrastructure, dedicated staff time and expertise from portfolio partners. Additionally, Omidyar Network will co-sponsor workshops and future calls for proposals, fostering an inclusive environment for innovation and knowledge-sharing.
OpenAI	Up to \$1 million in credits for model access for research related to AI safety, evaluations and societal impacts, and up to \$250,000 in model access and/or ChatGPT accounts to support applied research and coursework at historically Black colleges and universities and minority-serving institutions. Additionally, OpenAI will provide next-generation AI technology and data processing tools to aid in the digitization and structuring of large-scale datasets that are not currently available for research.
OpenMined	Access to an integrated architecture based on privacy-enhancing technologies and up to \$500,000 in cloud credits to support research partnerships working to develop AI solutions with sensitive or distributed data.



PRIVATE SECTOR PARTNERS

Organization	Contribution
Palantir	Supports the NAIRR Secure Pilot and the expansion of the National Clinical Cohort Collaborative through access to Palantir's Foundry and AIP platforms deployed at the National Institutes of Health, including compute hours and platform support resources.
Regenstrief Institute	Large-scale clinical datasets derived from real-world health care data in support of the NAIRR Secure Pilot to bolster AI research and innovation in health care.
SambaNova Systems	Hosted environment for NAIRR Pilot researchers to access pre-configured generative AI and large language models for domain-specific fine-tuning and experimentation. Researchers will receive training and technical support to ensure project success. Technical assistance will also be provided for researchers utilizing the SambaNova cluster deployed at Argonne National Laboratory.
Vocareum	\$2.4 million in the form of access to Vocareum notebooks and cloud resources for 20,000 students, enabling educators to deliver hands-on AI education in their classrooms.
Voltage Park	One million GPU-hour compute credits, localized storage and associated staff time and expertise to support projects for the research community.
Weights & Biases	Free academic licenses for NAIRR Pilot users to access Weights & Biases' AI Developer Platform, and technical support to maximize impact from use of the platform.



Graphic illustrating NVIDIA DGX™ Cloud, an AI supercomputing service that gives enterprises immediate access to the infrastructure and software needed to train advanced models for generative AI and other groundbreaking applications. NVIDIA DGX Cloud supercomputing is being used by NAIRR Pilot Projects. Credit: NVIDIA



APPENDIX 2

NAIRR PILOT ORGANIZATION AND AGENCY INVESTMENTS

Pilot governance

The pilot employed a lightweight governance and operational model (figure below) involving several entities.

- NSF provided the administrative home for and executive leadership of the NAIRR Pilot through the NSF Office of Advanced Cyberinfrastructure (NSF OAC), working in close coordination with federal partners and the Office of Science and Technology Policy (OSTP) within the Executive Office of the President of the White House.
- The NAIRR Pilot Steering Committee, chaired by NSF and OSTP and with participation of the federal agencies that are contributing *in-kind* resources and expertise to the pilot, provides strategic direction and interagency coordination and input on pilot strategy, structure, activities and forward planning.
- The NAIRR Interagency Working Group includes steering members as well as any other interested federal agencies; it meets periodically to discuss respective agency interests and ideas.
- A self-governed nongovernmental partner forum led by NVIDIA, Microsoft and the Allen Institute for AI, with NSF as *ex officio* observer, facilitates exchange among private sector partners and enables individual partners to provide input and feedback on the use of partner-contributed resources.

The NAIRR Pilot Portal is available at: <https://nairrpilot.org>

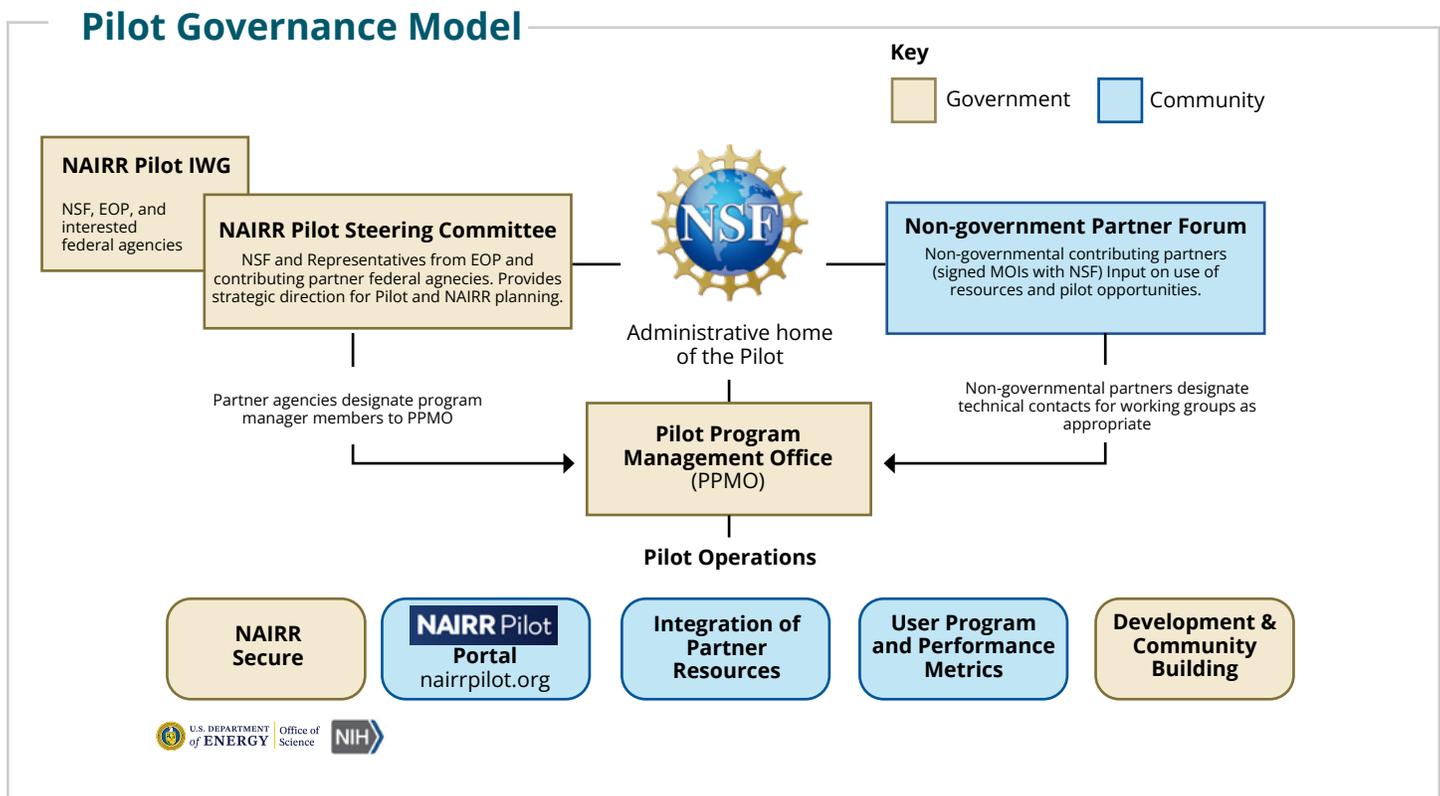


Figure A2.1: NAIRR Pilot organization and operational elements.



Day-to-day oversight of pilot operations is performed by the Pilot Program Management Office (PPMO), led by NSF OAC program officers. PPMO coordinates with partner federal agencies and oversees pilot activities conducted by community-side operations teams and working groups funded by NSF. NSF and the operations teams conduct communications and stakeholder engagement on behalf of the pilot through a dedicated website, press releases, announcements, mailing lists, webinars and other means. In addition, the U.S. Department of Energy Office of Science and the National Institutes of Health (NIH) co-lead the NAIRR Secure Pilot, described in greater detail below.

Pilot operations and user services

The chief operational focus of the pilot is to make contributed AI resources available for use by U.S.-based researchers and educators, including U.S.-based small businesses with federal research grants.

Portal. The public-facing NAIRR Pilot Portal² serves as a one-stop resource for researchers and educators to learn about the pilot and its offerings, request use of available pilot resources, and obtain assistance. The portal was developed and is operated by the Science Gateways Center of Excellence (SGX3) through NSF award 2231406, which coordinates with other pilot teams on the portal's content and functionality.

User program— pilot resource offerings. Three types of user-facing resource opportunities are available to researchers and educators through the portal via a short request form, independent review and matching process:

- **AI Resources for researchers.** NSF- and DOE-supported advanced computing resources and industry partner-contributed cloud computing, model, inference and other types of allocatable resources are available for request.
- **Startup projects.** A rapid entryway into the NAIRR Pilot that enables researchers to access small amounts of computing resources for a limited three-month period to enable prototyping, exploration and facilitate submission of full research resource requests.
- **Deep partnerships.** Collaborative opportunities to engage interactively and collaboratively employ NAIRR contributed resources with industry partners on projects of mutual interest.
- **NAIRR Classroom.** Educators can request access to certain resources to use in their AI-related courses, currently including notebooks and computing offered by the NSF Prototype National Research Platform, NSF Jetstream 2, the NIH CloudLab platform, and Vocareum notebook licenses.
- **Data, models and more.** Users can access a wide variety of federal AI-relevant datasets, pre-trained models, educational resources, toolkits and other resources ranging from U.S. Patent & Trademark Office data, NASA foundation models and datasets, and NIH-sponsored biomedical data-sharing platforms.

²The NAIRR Pilot Portal is available at: <https://nairrpilot.org>

CURRENT OPPORTUNITIES

<p>DEEP PARTNERSHIPS</p> <p>Request access to deeper collaborations directly with select partner resources.</p>	<p>New Opportunity</p>	<p>START-UP PROJECTS RESOURCES</p> <p>Request access to AI resources for start-up projects.</p>	<p>New Opportunity</p>
<p>RESEARCH RESOURCES</p> <p>Access high-performance computing platforms tailored for AI research.</p>	<p>Requires Application</p>	<p>EDUCATIONAL RESOURCES</p> <p>Request access to educational platforms (such as computational notebooks).</p>	<p>Requires Application</p>
<p>DATA, MODELS, AND MORE</p> <p>Browse curated datasets, pre-trained models, and additional tools for training and testing your AI systems.</p>		<p>Freely Accessible</p>	

Figure A2.2: Opportunities available on the NAIRR Pilot Portal.



Request review and matching.

To ensure fairness in the apportionment of the pilot's requestable resources, a NAIRR Pilot Allocation Working Group, comprising experts in the allocation of advanced computing resources, is supported by NSF to manage the request, review, and allocation processes in close coordination with PPMO. This includes creating public calls for requesting pilot resources and associated policies; organizing and conducting independent peer review of the requests; creating and managing a process for matching meritorious requests to available contributed resources; and communicating the outcomes of the requests/matching to the requestors, resource providers and other stakeholder operations groups including user experience, pilot portal managers, as well as the PPMO. The eXtensible Resource Allocation Service platform was adapted via NSF funding to manage the pilot's allocation processes.

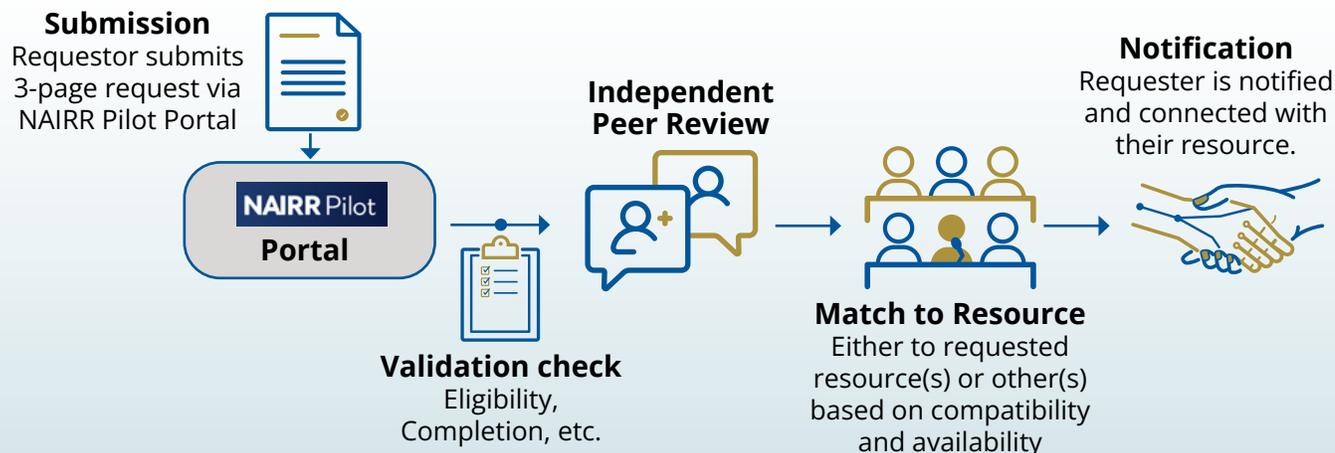


Figure A2.3: Pilot resource request and allocation process.

Policies.

The NAIRR Pilot leadership and steering committee set the basic policies for resource allocations. In keeping with the goals of the pilot, eligibility is restricted to U.S.-based researchers and educators, defined as individuals at U.S.-based institutions, including academic institutions, nonprofits, federal agencies or federally funded research and development centers, state, local or tribal agencies, and startups and small businesses with federal grants. The NAIRR Pilot makes AI-related resources available to users on a merit basis at no cost; however, it does not support the costs associated with the scientific or educational activities that are being pursued. Resource requesters in academia are encouraged but not required to indicate federal research funding associated with the pilot projects being requested. Researchers from small businesses must indicate relevant funding support from a federal small business grant.

User support, onboarding and training.

User support for each allocated project is provided by the assigned resource provider(s). A User Experience Working Group is supported by NSF to assist with onboarding pilot users, ensure good end-to-end user experience, coordinate user training opportunities and resources, and coordinate across resource providers and pilot operations groups as needed. User training opportunities include live sessions, videos, webinars, hackathons and office hours provided by resource providers and user support personnel, all advertised on the pilot portal events and training page.³

Usage and performance measurement.

Through NSF funding, the Open XDMoD tracking system has been deployed to monitor pilot allocations and resource usage across NSF-supported computational systems and nongovernmental partner cloud resources. Overall tracking of pilot user experience and progress is also gathered by the User Experience Working Group.

³ List of pilot events and training available at: <https://nairrpilot.org/pilotevents>



Resource integration and capacity enhancements.

To achieve a rapid launch of the NAIRR Pilot's user program, NSF and DOE leveraged existing advanced computing and cyberinfrastructure investments, including the coordinated services provided by the NSF ACCESS⁴ program and time on NSF-supported academic supercomputers, and DOE's contribution of time on the Summit supercomputer at Oak Ridge National Laboratory and the AI test bed at Argonne National Laboratory. Within the first year of the pilot, all the government-supported resources and approximately 80% of the nongovernmental contributed resources were integrated into the pilot to be requested by researchers and educators, with the remaining integrations currently underway. NSF also invested in expanding the computing capacity available on systems funded under the NSF Advanced Computing Systems and Services program⁵.

Agency investments in novel AI pathways and building NAIRR community.

In parallel with supporting the above core operational activities, to maximize the pilot's scientific and technical impacts and outcomes, NSF and other agencies are investing to demonstrate novel AI research infrastructure pathways and nurture the emerging NAIRR community of researchers, educators, resource providers and operational teams.

NSF AWARD	INSTITUTION	ACTIVITY
2501232	University of Illinois at Urbana-Champaign	Pilot operational coordination
2138259 <i>supplemental funding</i>	Carnegie Mellon University	Resource allocations
2138286 <i>supplemental funding</i>	University of Colorado Boulder	User engagement
2138307 <i>supplemental funding</i>	University of Illinois at Urbana-Champaign	Ticketing System
2137603 <i>supplemental funding</i>	SUNY Buffalo	Metrics
2436057	Internet2	Outreach and communication
2231406 <i>supplemental funding</i>	UC San Diego	nairrpilot.org web portal

Table A2.1: Acknowledgement of various NAIRR Pilot operations teams.

⁴ NSF ACCESS Program available at: <https://access-ci.org/>

⁵ NSF Advanced Computing Systems and Services program available at: <https://www.nsf.gov/funding/opportunities/advanced-computing-systems-services-adapting-rapid-evolution>.



Demonstration projects

Pilot demonstration projects⁶ explore how pilot resources may be used and combined in innovative ways, how their capabilities may be enhanced with additional cyberinfrastructure software, and how higher-level and discipline- and community-specific capabilities can be built on top of NAIRR Pilot infrastructure.

NSF demonstration projects.

NSF supported 20 new demonstration projects through an open call for proposals⁷ and invested in expanding ongoing national-scale distributed data infrastructure projects to support AI-driven research, including

- The National Data Platform (NSF award 2333609) provides a catalog of AI-ready open data from community providers and repositories, which can be used to explore new AI techniques.
- The Pelican software platform (NSF award 2331480), together with the Open Science Data Federation, provides a national data delivery system that allows NAIRR Pilot users to rapidly connect data to computing.
- The SAGE distributed infrastructure (NSF award 2331263) supports NAIRR Pilot integration for AI-at-the-edge.
- The Partnership to Advance Throughput Computing (PATH; NSF award 2030508) expands support for NAIRR Pilot researchers.

NAIRR Secure Pilot led by DOE and NIH.

DOE and NIH co-lead the NAIRR Secure Pilot with a working group of members from interested federal agencies. NAIRR Secure explores issues and challenges for AI projects that involve sensitive or protected data and secure computing environments. Specific objectives are to refine the requirements and infrastructure design patterns for the future NAIRR Secure resources, including integration of resources with differing sensitivities or architectures; investigate new opportunities for combining data in ways that preserve privacy and security; and identify research and training use cases to inform the development of NAIRR Secure in the future.

The NAIRR Secure Pilot has focused its work on several interagency demonstration projects involving NIH, DOE, NSF and the National Center for Science and Engineering Statistics (NCSES), the latter through its National Secure Data Service Demonstration Project. The demonstration projects currently include:

- Synthetic Data Generation with Large Real-World Data (Agency partners: NSF/NCSES, NIH/National Center for Advancing Translational Sciences, DOE).
- Democratizing AI for cancer with privacy-preserving synthetic data generation for cancer case identification (Agency partners: NIH/NCI, DOE).
- Althena: Institutional LLM-Assisted Biomedical Question Answering (Agency partners: NIH/NCATS, NSF, DOE).

In addition, the NAIRR Secure Pilot has surfaced and documented lessons learned from a variety of AI research projects involving sensitive data and secure environments that are germane to the assessment and management of risks and opportunities in the rapidly evolving AI landscape.

⁶ Full list of NAIRR Pilot Demonstration Projects available at: <https://nairrpilot.org/projects/demo>

⁷ NAIRR Dear Colleague Letter (DCL) 24-093 available at: <https://www.nsf.gov/funding/opportunities/dcl-national-artificial-intelligence-research-resource-nairr/nsf24-093>



Community building

Outreach.

Significant outreach has been conducted by NSF on behalf of the pilot, including through announcements, public webinars and numerous presentations to university consortia, state and regional cohorts of researchers and institutions, the small business community, and at major scientific and AI conferences. Additional outreach has been conducted by the supported community-side pilot operations teams through surveys, workshops, showcases to highlight research and education projects, training opportunities, and organizing community outreach sessions at major conferences. Participating agencies also provided outreach to their respective research communities.

Gathering user needs and use cases.

An important goal of the pilot is to gather information on the needs and use cases of researchers and educators for AI resources to inform not only about the pilot activities but also the design of the eventual full-scale NAIRR.

- **NAIRR request for information.** At the launch of the NAIRR Pilot, NSF coordinated with the NAIRR federal partners, with support from the Science and Technology Policy Institute, to develop a *Request for Information (RFI) on Researcher and Educator Use Cases for the National Artificial Intelligence Research Resource (NAIRR)*, which was released through an NSF Dear Colleague Letter⁸ and cross-advertised by the partners. Several hundred responses were received from across the science and engineering research and education domains. The results informed various aspects of the Pilot activities and emphases, particularly concerning the need for user training and data support.
- **Community survey of AI resource needs.** Subsequent to the RFI, the SGX3 Science Gateways Center of Excellence (which manages the NAIRR Pilot Portal) and the NAIRR Pilot User Experience Working Group undertook a survey of both current and prospective users of NAIRR Pilot resources. This produced over 400 responses, including insights into the pain points and needs of the research community.

Other information about user requirements, use cases and challenges was gathered informally through meetings of the NAIRR Pilot steering committee, partner forum, community meetings and workshops (see below) and operations working groups.

⁸ NAIRR Dear Colleague Letter (DCL) 24-051 available at: <https://www.nsf.gov/funding/information/dcl-request-information-researcher-educator-use-cases-national/nsf24-051>



NAIRR Pilot research partners provide hands-on demonstrations of groundbreaking AI innovations from around the nation during a Capitol Hill event hosted by the Congressional Artificial Intelligence Caucus.

Credit: Charlotte Geary/NSF



Community building (continued)

Topical workshops.

Investments were made to catalyze and nurture an emergent NAIRR community of users, resource and service providers, developers, and thought leaders. Several workshops were supported on topics associated with NAIRR goals, culminating with a major community gathering: the inaugural NAIRR Pilot Annual Meeting.

- A 2024 workshop on **Community-Informed Policies and Best-Practices for the National Artificial Intelligence Research Resource (NAIRR)** was held at New York University, supported by NSF (Award 2432040) and Omidyar Network, to explore policies and practices for trustworthy and interpretable AI as applied to the pilot and beyond. Participants included AI policy experts, pilot users and representatives of NAIRR Pilot resource providers. The workshop report summarizes the discussions, key takeaways and suggestions which informed pilot activities and planning for the full NAIRR.
- NSF and DOE sponsored a 2024 **NAIRR Software Workshop**⁹ at Argonne National Laboratory (NSF award 2440948) with 110 participants from academia, national laboratories, nonprofit and industry organizations, as well as federal agencies, developer and user communities. The goal was to begin the process of defining the needs, scope and requirements for an AI software stack (or stacks) that could be implemented as part of the eventual full-scale NAIRR, and to identify a few areas and activities that could be explored and piloted as part of the NAIRR Pilot toward the development of such a capability.
- A **NAIRR portal workshop** was held Feb. 4-5, 2025, at the San Diego Supercomputer Center. The 84 participants discussed and developed recommendations for the design and functionality of a central portal for a future full-scale NAIRR. The workshop was supported as part of a supplementary award¹⁰ to the Science Gateways Center of Excellence (SGX3), "Community Discovery for Near-Term NAIRR Pilot Portal Functionality" (NSF supplementary award 2438297), which also involved gathering community feedback through an online survey that received 1,234 responses, six two-hour online focus groups, and two in-person focus groups at the annual SGX3 Gateways conference.
- NSF supported an **inaugural NAIRR Pilot Annual Meeting** in February 2025 in Arlington, Virginia, organized by Internet2 (NSF award 2452148), which brought together 378 attendees from 47 states, including researchers, educators, resource providers, operations working group representatives and federal agencies. The goals were to foster connections and share experiences and insights from the first year of the NAIRR Pilot, gather feedback to inform pilot next steps, and plan for an eventual full-scale NAIRR. Participants provided valuable feedback¹¹ on streamlining proposal processes, enhancing resource accessibility, and supporting a wide range of users.
- Additional NSF supported community events and training activities are supported by the **NAIRR Expansion**¹² initiative. This initiative included a call for proposals¹³ with two focus areas: one for community outreach and training to new and emerging researchers, and one for educators bringing inclusive AI-educational experiences to classrooms nationwide.

⁹ Final Report of the 2024 NSF/DOE Workshop on NAIRR Software available at: https://events.cels.anl.gov/event/529/attachments/584/2351/NSF-DOE-NAIRR_SoftwareReport-Final.pdf

¹⁰ NAIRR Portal supplementary award report, including report out of the NAIRR Portal workshop available at: <https://zenodo.org/records/15792892>

¹¹ NAIRR Pilot Inaugural Annual Meeting Report available at: <https://zenodo.org/records/15263283>

¹² Full list of NAIRR Expansion Projects available at: <https://nairrpilot.org/projects/expansion>

¹³ NAIRR Dear Colleague Letter (DCL) 24-109 available at: <https://www.nsf.gov/funding/opportunities/dcl-nairr-pilot-expansion-researcher-training-community-outreach/nsf24-109>



Policies and Practices workshop

- *Include tools for assessing transparency of datasets, models, and research outputs.*
- *Incentivize good practices by resource providers and users.*
- *Foster interactions of responsible AI and infrastructure communities.*
- *Establish metrics and measurement protocols to assess progress.*



Software workshop

- *Support diverse hardware platforms and accelerators.*
- *Address both HPC and AI software concepts and applications (e.g., layered stack, integrated ecosystem for data, user support, training, etc.).*
- *Offer easy-to-use interfaces to lower barriers for novices and expert (e.g., notebooks, chatbots, educational resources, support, etc.).*



Portal workshop

- *Organize the portal to ensure easy navigation, awareness of, and access to NAIRR resources.*
- *Simplify the initial user experience while supporting advanced functionalities.*
- *Provide services for data analysis, data sharing, and data storage.*
- *Continue gathering user feedback and monitor platform usage.*

Figure A2.4: Participant suggestions from three topical NAIRR Pilot workshops.



APPENDIX 3

NAIRR PILOT COMMUNITY DATASETS

The datasets selected in the NAIRR Pilot community datasets section of the main text are listed in figure A3.1. These datasets were selected via an NSF Dear Colleague Letter¹⁴, and further updates to the NAIRR Pilot community data sets will be found at the Open Data, Models, and More section of nairrpilot.org¹⁵.

NAIRR Pilot Datasets

John Hopkins <u>Turbulence Datasets.</u>	<u>AI4Shipwrecks</u> from the University of Michigan.
<u>FathomNet</u> from Monterey Bay Aquarium Research Institute.	University of California San Francisco <u>Industry Documents Library.</u>
<u>Cell Painting Gallery</u> from Broad Institute.	<u>SecureChain</u> from Purdue University.
Lidar-derived high-res raster <u>topography dataset</u> from the Open Topography Facility.	<u>PatchDB</u> from George Mason University.
<u>Microbiome data from pregnant women</u> from the March of Dimes Prematurity Research Center at the University of California San Francisco.	<u>Phase-Field Fracture Simulation Dataset</u> from John Hopkins.

Table A3.1: List of NAIRR Pilot Community Datasets.

¹⁴ NAIRR Dear Colleague Letter (DCL) NSF 25-018 available at: <https://www.nsf.gov/funding/information/dcl-national-ai-research-resource-nairr-pilot-seeks-datasets>

¹⁵ Open Data, Models, and More section of nairrpilot.org, filtered for community datasets, available at: <https://nairrpilot.org/pilotresources#community>



U.S. National Science Foundation

www.nsf.gov