

Broadening Access to AI resources through the National AI Research Resource (NAIRR) Pilot

Sharon Broude Geva sgeva@nsf.gov

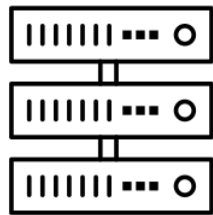
Office of Advanced Cyberinfrastructure, National Science Foundation

Anchorage CISE MSI Workshop

August 5, 2024

Vision for the National AI Research Resource

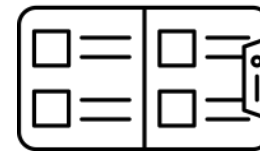
A widely-accessible, national research infrastructure that will advance the U.S. AI R&D environment, discovery, and innovation by empowering a diverse set of users through access to:



Secure, high-performance, privacy-preserving **computing**



High-quality **datasets**



Catalogs of **testbeds** and **educational materials**



Training tools and **user support** mechanisms

Goals:



Spur **innovation**



Increase the **diversity** of talent in AI



Improve U.S. **capacity** for AI R&D



Advance **trustworthy AI**

Why do we need a NAIRR?

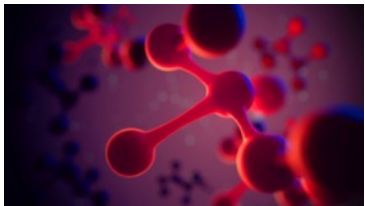
- Many potential contributors lack access to requisite resources which can be costly *as well as* hard to navigate
- Researchers investigating AI to serve the public good require access to resources
- To train the next generation of researchers and AI leaders



Envisioned Users of the NAIRR



AI Researchers



Domain Scientists
Applying AI



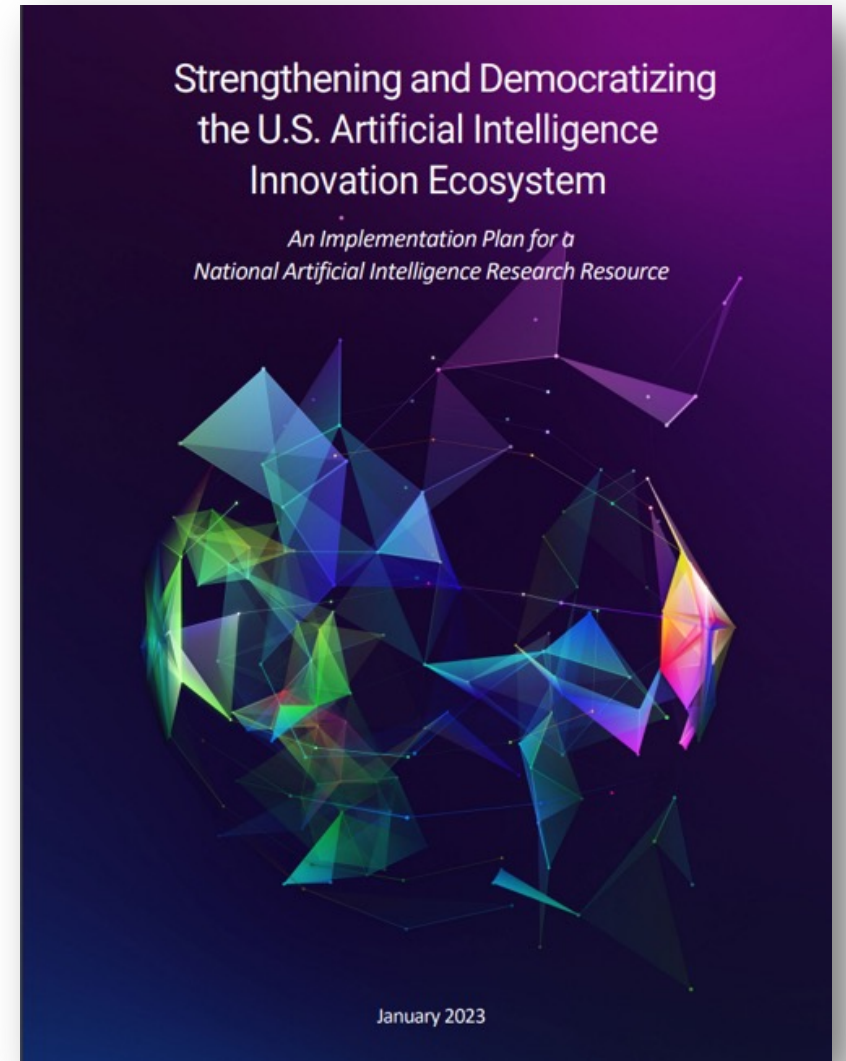
Students and
Educators

Researchers and educators affiliated with US-Based Institutions including:

- Academic institutions
- Non-profits
- Federal agencies or federally-funded R&D centers
- State, local, or tribal agencies
- Startups and small businesses with Federal grants

How did we get here?

- 2020 AI Initiative Act created a congressionally mandated NAIRR Task Force
- June 2021 - NAIRR Task Force launched to examine the feasibility of a NAIRR
- Jan 2023 - Task Force final report provides implementation plan for a **\$2.6B** initiative
- March 2023 - NAIRR Interagency Working Group led by OSTP & NSF



NAIRR Pilot

NAIRR Pilot origins and goals



Oct. 2023 Executive Order directed NSF and collaborating agencies to launch a pilot for the NAIRR in 90 days

NAIRR Pilot Goals

1. Demonstrate the value & impact of the NAIRR concept.
2. Support novel & transformative AI research while reaching broad communities.
3. Gain experience to advance and refine NAIRR design.

Budget for 2 year pilot:

- Leverage existing resources and investments from agencies
- FY24 budget \$XXM, FY25 Request \$30M
- In-kind contributions from non-govt partners ~\$80M

Our NAIRR Pilot Plan and Approach



Pilot or investigate all major components of the NAIRR vision but at a limited scale and/or scope



Leverage partnerships



Embrace a pioneer approach while adhering to trustworthy AI principles

Bringing together the strengths of government, private industry and non-profit partners

Contributing Partners

Agencies

- National Science Foundation
- Defense Advanced Research Projects Agency
- Department of Agriculture
- Department of Defense
- Department of Energy
- Department of Education
- Department of Veterans Affairs
- National Aeronautics and Space Administration
- National Institutes of Health
- National Institute of Standards and Technology
- National Oceanic and Atmospheric Administration
- US Geological Survey
- US Patent and Trademark Office (USPTO)

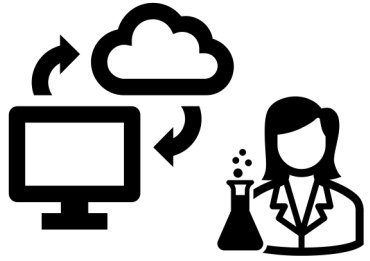
Non-governmental orgs

- AI2: Allen Institute for AI
- AMD
- Amazon Web Services
- Anthropic
- Cerebras
- Databricks
- Datavant
- EleutherAI
- Google
- Groq
- Hewlett Packard Enterprise
- Hugging Face
- IBM
- Intel
- Meta
- Microsoft
- MLCommons
- NVIDIA
- Omidyar Networks
- OpenAI
- OpenMined
- Palantir
- Regenstrief Institute
- SambaNova Systems
- Vocareum
- Weights & Biases

Cross-Cutting AI Research Thrusts

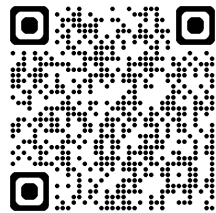
- **Advance AI methods** that enable scientific discovery
- Develop practices for creating **open-source foundation models** for specific applications
- Use large-scale models to **explore complex datasets** interactively
- Advance approaches for **integrating simulations and AI**
- Empower use of scientific data **by other stakeholders and the public.**
- Probe key challenges using **sensitive data** including privacy-preserving methods
- Use **AI at the edge** for scientific research

NAIRR Pilot Opportunities Available to Community



Apply to use NAIRR Pilot resources

Access to computing, data, model, collaboration resources



Contribute to the NAIRR Pilot with a demonstration project

Funding opportunity: Demonstration Project DCL – concept outline required



Train, educate and/or build community around NAIRR pilot resources

Funding opportunity: NAIRR Pilot Expansion DCL – concept outline required

Apply for resources at nairrpilot.org Today!

NAIRR Pilot

National Artificial Intelligence
Research Resource Pilot

NAIRR Pilot
Portal

CLASSROOM AND EDUCATOR RESOURCES

Request access to educational
platforms (such as Jupyter
notebooks).

[Apply](#)

RESEARCHER RESOURCES

Request access to advanced
computing, cloud computing,
models, software, platforms,
and collaborations.

[Apply](#)

DATA, MODELS, AND MORE

Additional government and
government-funded resources
that do not require an
application.

[View](#)

Agency and Non-govt Resources Available to Community



Frontera – UT



Delta – U. Of Illinois



Jetstream 2 – Indiana U



Anvil – Purdue Expance/Voyager UCSD



Bridges 2 – U. Pittsburgh/CMU



ACES - Texas A&M University



AI Testbed – Argonne National Lab



OpenMined



Hugging Face



Google Cloud



ANTHROPIC



Educator Resources

Interactive notebooks: integrated hardware and software environments



Prototype National Research Platform

NRP NATIONAL RESEARCH
PLATFORM



CloudLab

National Institutes
of Health

Vocareum

NAIRR Pilot Classroom Partner showcase was held on 6/13
<https://new.nsf.gov/events/nairr-pilot-classroom-partner-showcase/2024-06-13>



FAQ available on nairrpilot.org

- Project Description (3-pages)
 - A. Scientific/Technical Goal
 - B. Estimate of Compute, Storage, and other Resources **(Justify this!)**
 - C. Support Needs
 - D. Team and Team preparedness
- Supporting Documents
 - References
 - CV / Bio sketch material for team lead

Eligibility for NAIRR Pilot Classroom/Educator Application

This call is open to proposals ...

- by US based educators and researchers
- who are teaching undergraduate or graduate courses or shorter duration training sessions
- to US based students
- that include subject matter in artificial intelligence and require that students use advanced computational resources as part of their coursework.
- Courses from any discipline are eligible for this program. **Courses and training sessions must not allow participants who are not US based.**

Cross-Cutting AI Classroom/Educator Call Focus Areas:

There are **no specific constraints** on the courses eligible for this opportunity. Examples include:

- Computer science (or information science, or data science) courses on machine learning, artificial intelligence (AI), or intelligent systems
- Domain science courses or trainings that are incorporating AI into the curriculum
- Independent study courses of any discipline incorporating AI. This opportunity is also available to Instructors who are developing a course or tools for the classroom and need computational resources to complete testing prior to offering their course.

NAIRR Classroom Resource Application Process

<https://nairrpilot.org/>

- A. Course Description
- B. Estimate of Compute, Storage, and other Resources (**Justify this!**)
- C. Support Needs
- D. Team and Team preparedness
- E. Supporting Documents
 - CV / Bio sketch material for team lead

<https://https://nairrpilot.org/opportunities/education-call>

Some points to keep in mind (for all requests)...

<https://nairrpilot.org/>

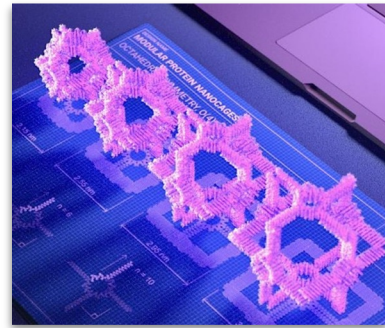
- Not an NSF Proposal!
 - Not a request for funding
- This is a pilot! Resources currently available, may be limited in the future
- Review process:
 - Expectation: Requests reviewed and decisions made in 8-10 weeks
 - Requests are batched to 15th day of each month (i.e. requests submitted by July 15th are currently being reviewed).

~100 NAIRR Pilot Projects announced thus far (as of July)

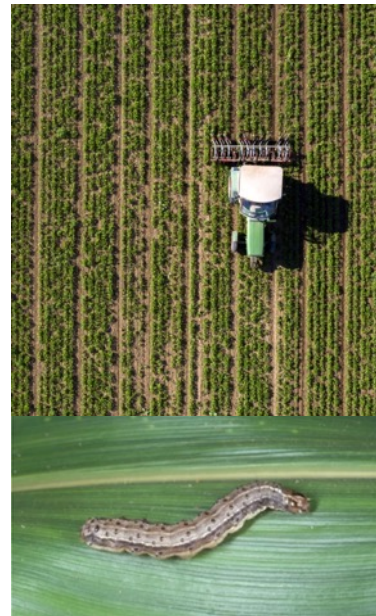
Themes around model safety and security, privacy preserving data generation, and science applications



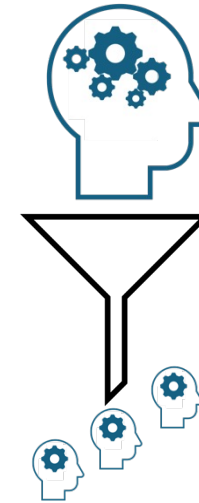
Training Foundation Models from Private, Federated Client Data
PI: Fanti, Carnegie Mellon



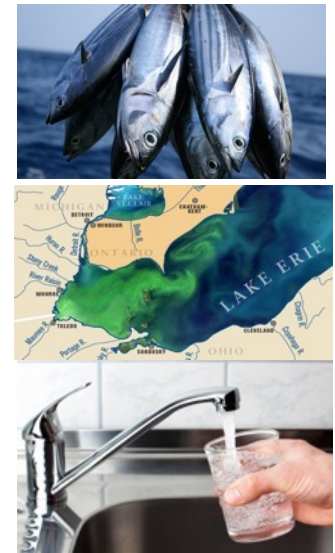
Advanced Training for Protein Diffusion, Binder Prediction, and Antibody Design
PI: Baker, U. Washington



Agricultural Pest Identification with Robustness Guarantees
PI: Ganapathysubramanian, Iowa St.



Are compressed language models accurate, robust & fair?
PI: Srikumar, U. Utah

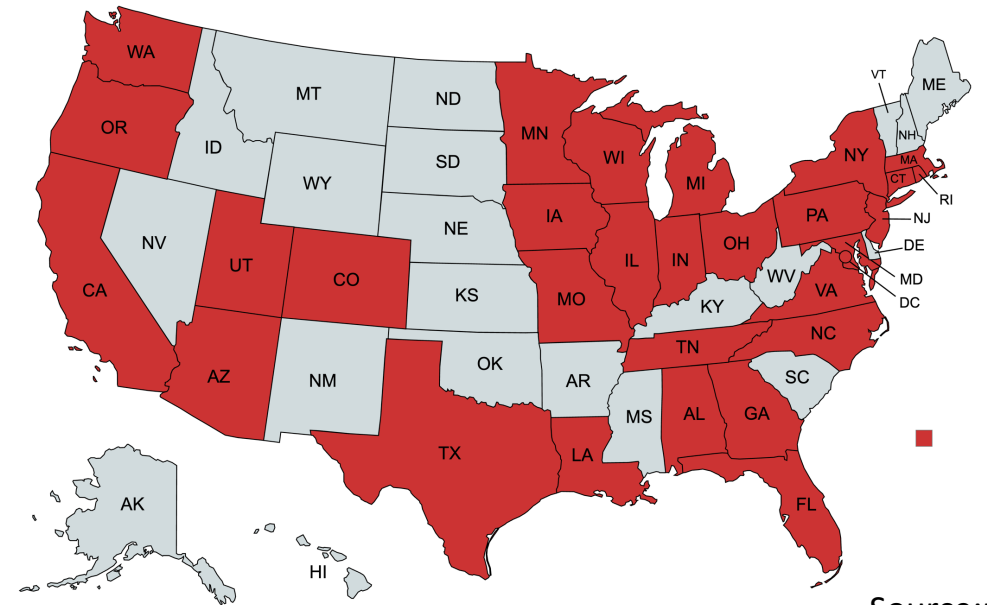


LakeGPT: A Foundation Model for Aquatic Sciences
PI: Karpatne Virginia Tech

NAIRR Pilot user community is growing!

- ~100 projects matched to agency and non-gov partner resources
- New awards announced approximately monthly
- Receiving ~1.5 submissions per day.
- ~80 submissions currently under review
- Demand is high
- Significant outreach needed to communities across the country

Projects supported in 29 states + DC so far (July 25, 2024)



Source:mapchart

Connecting researchers supported by numerous agencies to resources:

- NSF
- NIH
- NIJ
- NOAA
- USDA
- DOE
- DARPA
- DOD
- ARO
- ONR

A number of agency datasets are available through the pilot



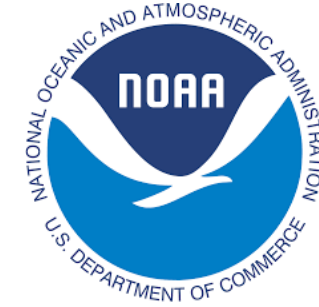
National
COVID
Cohort
Collaborative



IMMPORT
BIOINFORMATICS FOR THE FUTURE OF IMMUNOLOGY



National Institutes
of Health



Tropical Cyclone Dataset



UNITED STATES
PATENT AND TRADEMARK OFFICE



USPTO Research Dataset



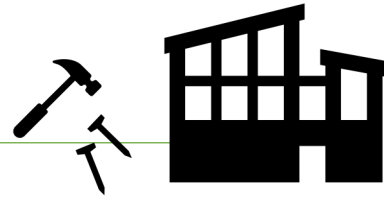
Lake Michigan Substrate
Prediction Dataset



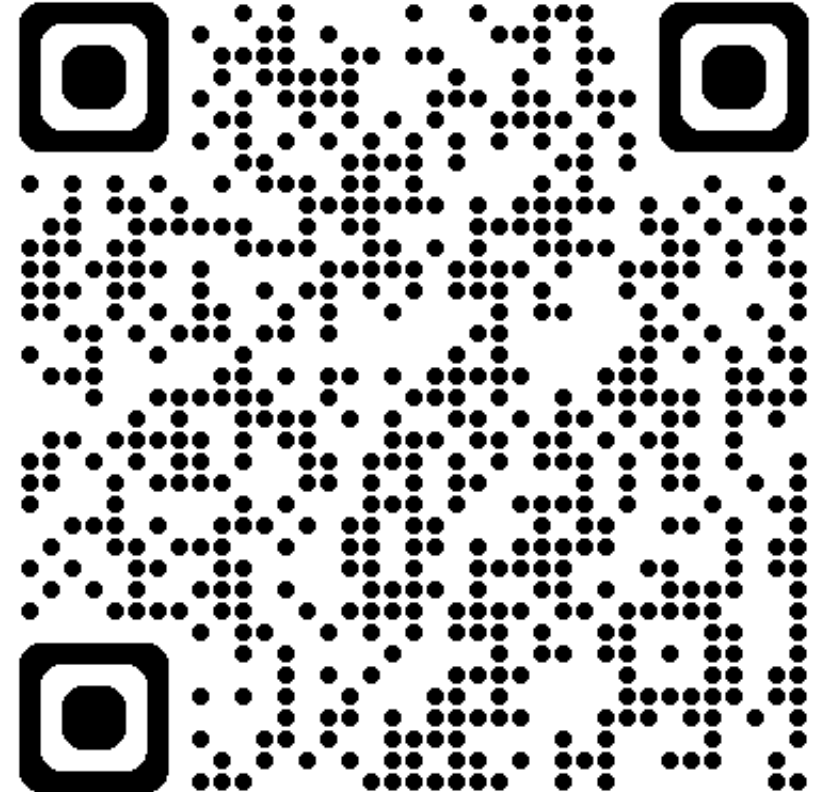
US Census of Agriculture

Note: NAIRR is envisioned to support a search and data discovery system – not direct hosting of datasets

Contribute with a Demonstration Project



- **NSF 24-093: NAIRR Pilot Demonstration Projects**
- Separate from allocation requests!
- Funding types
 - EAGER proposals
 - Supplemental funding requests
- NAIRR Pilot Demonstration Projects
- Led by teams of AI researchers, domain scientists, and cyberinfrastructure specialists
 - **Innovatively use, integrate with, or rely on one or more NAIRR Pilot requestable resources.**
- Interested teams should submit a Concept Outline to NAIRR_Pilot@nsf.gov



<https://www.nsf.gov/pubs/2024/nsf24093/nsf24093.jsp>

NSF 24-109 NAIRR Pilot Expansion



- NSF released a 'Dear Colleagues Letter' calling for proposals that expand the NAIRR pilot community
- Two focus areas:
 - New and emerging researchers
 - Supplements
 - EAGER proposals
 - Conference and workshop proposals
 - Educators
 - Conference and workshop proposals
 - Research (or education) Coordination Networks
- Interested teams should submit a Concept Outline to NAIRR_Pilot@nsf.gov



<https://www.nsf.gov/pubs/2024/nsf24109/nsf24109.jsp>

Plan to evolve the NAIRR Pilot in next 18 months

- Support additional workshops and community activities in different focus areas: portal, data, software, education, training and outreach
- Complete integration of partner resources
- Support budding NAIRR Pilot user community
- Incorporate trustworthy AI policies into pilot as possible
- Mature NAIRR Pilot operations and assess gaps
- Integrate additional resources or demonstrations projects

Additional (Free) Resources



What: Compute, Storage, SW Services & Support (not just for HPC)

Who: US researchers & educators (NSF funding not required)

Cost: Free (provided by NSF)

Welcome to ACCESS.
Advanced Cyberinfrastructure Coordination
Ecosystem: Services & Support

GET STARTED

How: <https://access-ci.org/about/get-started/#start>

How to get access to the NSF CI ecosystem?

- ACCESS – <https://access-ci.org/about/get-started/#start>
- Frontera allocations - <https://frontera-portal.tacc.utexas.edu/allocations/>
- PATH High Throughput computing credits - <https://path-cc.io/>
- Cloudbank - <https://www.cloudbank.org/options-requesting-cloud-resources>
- Cloudlab - <https://cloudlab.us/>
- Chameleon experimental testbed - <https://chameleoncloud.org/>
- National Artificial Intelligence Research Resource (NAIRR) - nairrpilot.org

Thank You

Sharon Broude Geva, sgeva@nsf.gov

To subscribe to the OAC Announce Mailing List

Send an email to: OAC-ANNOUNCE-subscribe-request@listserv.nsf.gov

