Unlocking AI Research Opportunities

with the

NSF NAIRR Pilot

Penny Atkins, Bill Miller, and Martin Cuma

June 5, 2026

SCIENTIFIC COMPUTING AND IMAGING INSTITUTE
THE UNIVERSITY OF UTAH

Today’s objectives

Start building with local resources,

then **scale** to national resources

- Learn about AI and research computing at the University of Utah including resources at the Center for High Performance Computing
- Introduce the NSF National AI Research Resource (NAIRR) Pilot
- Learn about available resources
- Learn how to apply for an allocation
- Discuss University of Utah faculty experiences with the NAIRR Pilot
- Identify potential use cases for researchers and educators

AI Resources at the University of Utah

Utah AI Infrastructure: Rapidly expanding capabilities tightly coupled to **state-wide** AI research, education, innovation, and workforce goals

Advanced Computing

- $50M HPE/NVIDIA UT-wide Sovereign AI Factory
- Open and secure HPC and virtualized environments

Tools & Resources

- Persistent AI services and AI Sandbox *in development*
- AI Engineers and ambassadors
- Public-private partnerships

Community Uplift

- $100M U of Utah Responsible AI Initiative (RAI)
- Utah Governor's Pro-Human AI workforce initiative
- UT AI Factory enablement for Higher Ed, K-12, and industry

Data at Scale & Speed

- $25M AI expansion of NSF National Data Platform, May 2026
- Advanced storage & archiving – 60 PB and climbing rapidly
- U Science DMZ and state-wide high-speed R&E connectivity

Center for High Performance Computing (CHPC)

Mission *Deliver advanced computing and data systems and services to accelerate discovery and innovation across the University and state, regionally and nationally.*

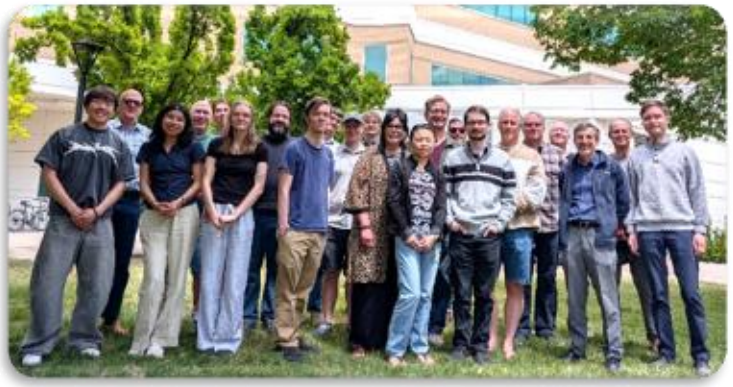
https://chpc.utah.edu/

2000+ Users & Collaborators

Research Computing, Data, & AI Environments

- Open HPC
- Regulated Environments
- Virtual Environs
- Data Systems & Services
- Expertise & Facilitation
- Connectivity, Integration & Interoperability

Innovation Continuum
Research → Piloting & Development → Production



We're hiring!
https://chpc.utah.edu/about/careers.php

- 45 staff, students, and interns
- CPU & GPU, data, and connectivity across scales
- Serves researchers, classes, and innovation

AI resources and tools at the CHPC

Computational and data resources

- **GPUs available by allocation request**
  - RAI Nvidia H200 GPUs plus storage.
  - General GPUs – Many options by allocation.
- **Guest access** to ~400 GPUs on “owner” nodes.
- **Reserved computing** - purchased dedicated service over a multi-year period (replaces owner purchases).
- **Storage** – high-performance storage for purchase.
- **Virtual Machines** – customized for specific needs.

Classroom resources

- **Support for classrooms needing GPUs –** 20–25 classes/semester
- **Classroom-specific environment setup**
  - Web interface to Linux systems; remote desktop
  - Jupyter with custom Python environment
- **General GPU reservations**
  - Modern and older GPUs Modern GPUs
- **More AI resources coming**
  - Inference - LLMs, instructor-run multi-tenant apps

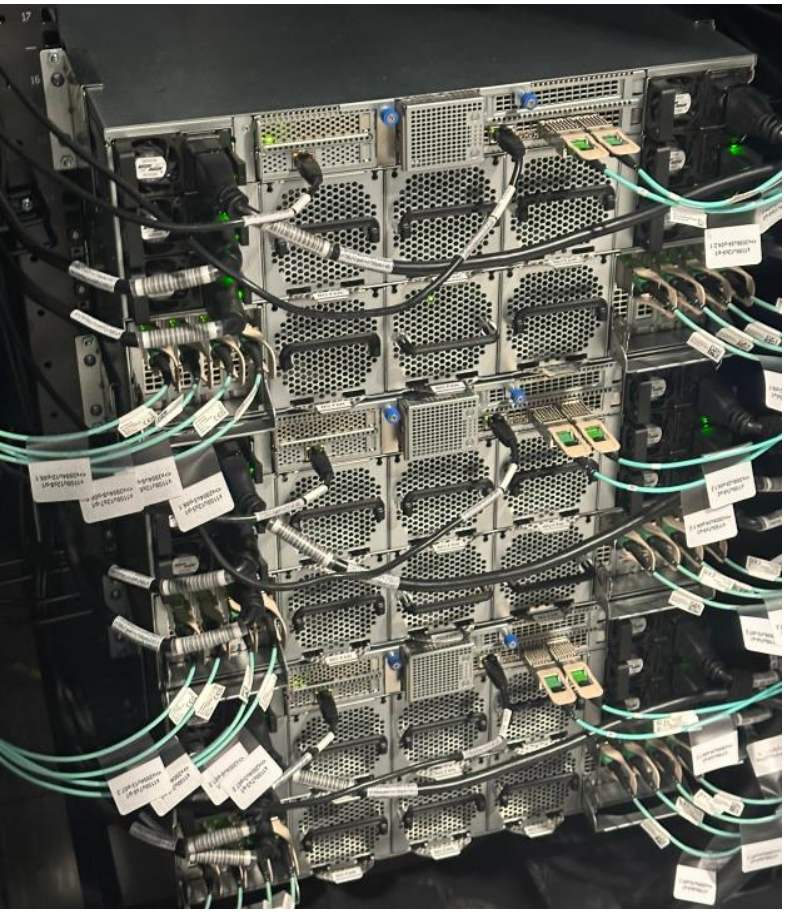
AI tools and applications

- **Tools and models** – Deep learning, User interface tools, Generative AI inference (Ollama, vLLM), Hugging Face generative AI models, NVIDIA Cosmos-Predict, and growing
- **Coding agent tools** – pending approval from AI Office

Free AI learning opportunities

- **CHPC AI trainings** – GPU programming, deep learning, coding agents. And many other live and recorded trainings.
- **Upcoming In-Person**: **Fundamentals of Deep Learning**, Mike Ramshaw, NVIDIA AI Ambassador, August 20, 2026.

Utah AI Factory Supercomputer



***A Statewide AI resource* that advances Utah as an AI Leader**

- Will be available to Utah researchers, educators, and innovators.
- $50 million 5-year public-private partnership by the State of Utah, Univ. of Utah, Huntsman Family Foundation, HPE, and NVIDIA.
- CHPC will conduct operations and coordinate enablement including training, access, and facilitation for all levels of users.
- Early access is anticipated for mid-summer 2026 followed by full operations in the Fall timeframe.
- Stay tuned for opportunities to access the AI Factory.

https://chpc.utah.edu/resources/utah-ai-supercomputer.php

Utah AI Factory Supercomputer

Fact Sheet

Computational resources

In aggregate, the Utah AI Factory Supercomputer has

- **33 HPE Cray XD670 nodes**
- **264 NVIDIA H200 GPUs**
  - Each node has eight H200 SXM5 GPUs, and each GPU has 141 GB of high-speed VRAM to support demanding AI workloads and large models
- **3,696 CPU cores**
  - Each node has two Intel Xeon-P 8570 56-core CPUs
- **66 TB of RAM**
  - Each node has 2 TB of DDR5-5600 RAM

The architecture is optimized for distributed AI training, large language models (LLMs), multi-node GPU computing, and tightly coupled HPC workloads.

High-speed GPU and node connectivity

The system is designed to enable GPUs to work efficiently together, both within and across nodes:

- Within nodes, there is **high-speed GPU-to-GPU communication, providing up to 900 gigabytes per second of bidirectional bandwidth**
- Between nodes, **each GPU is connected at 400 gigabits per second to an InfiniBand fabric** with 24 NDR switches, enabling large AI workloads to scale across nodes
  - The system uses a rail-optimized fat-tree topology to reduce latency and improve scalability

The platform is also **connected to the Utah Education and Telehealth Network (UETN) and the University of Utah's Downtown Data Center at 100 and 400 gigabits per second, respectively,** to facilitate data movement to and from other computational infrastructure in Utah.

High-speed storage

The platform includes high-performance temporary scratch storage for active research workloads, with **approximately 1 PB of usable scratch capacity, in addition to high-speed storage local to each compute node**. The WEKA filesystem supports high-throughput AI pipelines and scalable, data-intensive research applications.

The NSF NAIRR Pilot

National Artificial Intelligence Research Resource

- Launched January 2024 by the U.S. National Science Foundation with many federal and non-governmental partners.
- Aims to broaden access to AI tools and resources to accelerate research and innovation.
- Makes computing, data, software, models, and educational resources available to US-based researchers and educators.

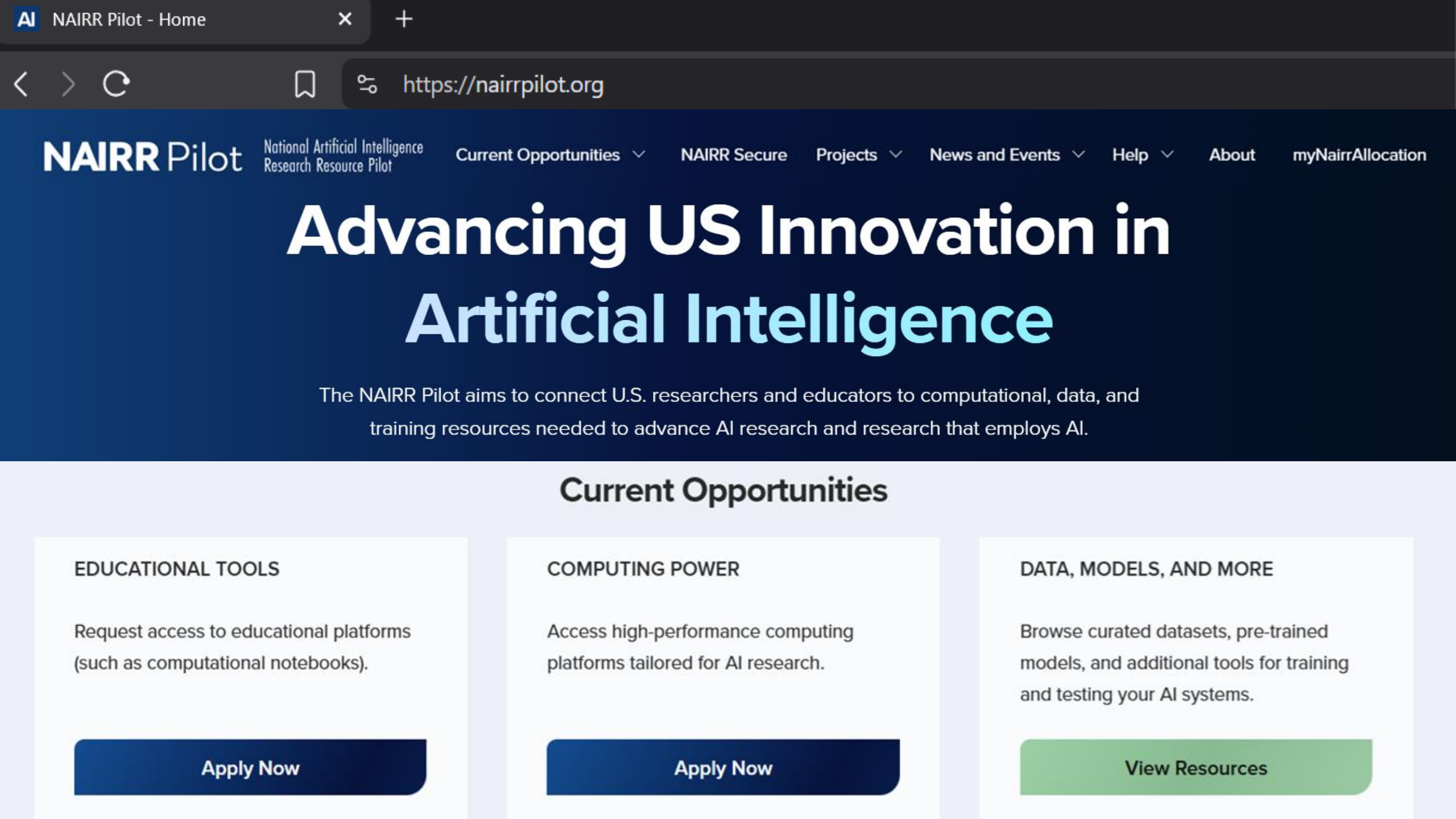
See: https://www.nsf.gov/focus-areas/artificial-intelligence/nairr

Federal agency partners

NSF	DOE	NIST	USPTO
DARPA	FDA	NOAA	VA
DoD	NASA	USDA	
DoEd	NIH	USGS	

Non-governmental partners

Allen AI Inst	EleutherAI	Meta	Palantir
AWS	Google	Microsoft	Regenstrief Inst
Anthropic	Groq	MLCommons	SambaNova Sys
AMD	HPE	NVIDIA	Vocareum
Cerebras	Hugging Face	Omidyar Net	Weights & Biases
Databricks	IBM	OpenAI	
Datavant	Intel	OpenMined	


NAIRR Pilot - Home
https://nairrpilot.org
NAIRR Pilot
National Artificial Intelligence Research Resource Pilot
Current Opportunities
NAIRR Secure
Projects
News and Events
Help
About
myNairrAllocation
Advancing US Innovation in Artificial Intelligence
The NAIRR Pilot aims to connect U.S. researchers and educators to computational, data, and training resources needed to advance AI research and research that employs AI.
Current Opportunities
EDUCATIONAL TOOLS
Request access to educational platforms (such as computational notebooks).
Apply Now
COMPUTING POWER
Access high-performance computing platforms tailored for AI research.
Apply Now
DATA, MODELS, AND MORE
Browse curated datasets, pre-trained models, and additional tools for training and testing your AI systems.
View Resources

NAIRR Pilot National Artificial Intelligence Research Resource Pilot

Research focus areas:

- Advancing AI methods that enable scientific discovery and improve AI interpretability, security and trust.
- Accelerating time to science and innovation through AI enabled automation, autonomy and novel design and control processes.
- Applying AI to use, share, or integrate sensitive data from multiple sources to enable new experimental methods and discovery.
- Advancing approaches for integrating simulations and AI.
- Creating or developing open-source AI tools, models, datasets, and methods.
- Training and educating the next generation AI-savvy workforce.

Other projects that align with the broader objectives of the NAIRR Pilot, as well as projects in other areas of AI research and applications, may secondarily be considered for allocation. See website: https://nairrpilot.org/opportunities/allocations

Slide courtesy of David Hart, Lauren Michael, Timothy Middelkoop (https://nairrpilot.org/help/presentations)

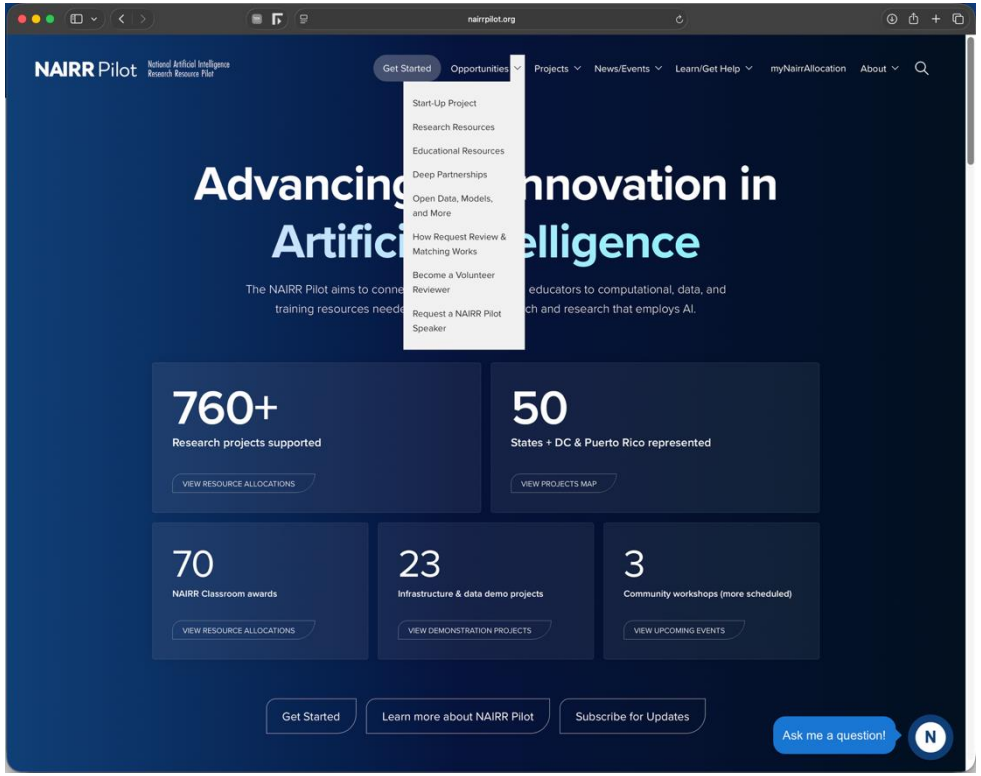
Example Use Cases

- Hands-on AI course requiring GPU resources for undergraduate or graduate students to develop and train models.
- Training AI models on large medical imaging datasets for disease detection.
- Evaluating and improving the use of LLMs for code synthesis.
- Benchmarking existing models for capabilities, security, and trust.
- Scaling up AI model training, including using simulation and synthetic data
- Use of novel AI hardware for training and inference.

Slide courtesy of David Hart, Lauren Michael, Timothy Middelkoop (https://nairrpilot.org/help/presentations)

Opportunities to use NAIRR resources

- Visit https://nairrpilot.org
- Look under “Current Opportunities”
- There are four project options
  - **Research Resources** — The original “NAIRR Pilot” opportunity, appropriate for most research activities
  - **Educational Resources** — a.k.a. “NAIRR Classroom,” for courses and training activities on a targeted set of resources
  - **Start-Up Projects** — Best for getting started, especially if you have limited access to resources to prepare a larger request
  - **Deep Partnerships** — Unique opportunity for close collaborations with selected resource providers and specialized offerings



nairrpilot.org > Current Opportunities

Slide courtesy of Dave Hart, NAIRR Pilot Allocations

Request and allocation process

1. **Proposals accepted at any time** (<3 pages)
2. **Proposals peer-reviewed approximately monthly**
   - soft deadline: 15th of the month
   - 2+ reviewers per proposal, per request size/complexity
3. **Selected proposals matched to resources.**
   - Matching committee includes reps from every resource-provider.
   - Proposers requesting only gov-funded resources can ask to be invisible to non-gov partners.
   - Matching may adjust the requested quantity or resource(s), per reviewer suggestions and/or resource availability.
4. **Selected awards notified.** (within 3 months)
5. **Resource onboarding between NAIRR Pilot Resource and Proposer.** (1-3 weeks)
6. **PIs provide required reports on progress.**

See https://nairrpilot.org/opportunities/allocations and https://nairrpilot.org/reporting-guidance

Slide courtesy of David Hart, Lauren Michael, Timothy Middelkoop (https://nairrpilot.org/help/presentations)

Getting Help

- NAIRR Pilot Website https://nairrpilot.org/
  - The Resource Opportunities provide lists of resources, documentation, and proposal instructions.
  - The Previous Awards page provides information on what projects and resources have been awarded.
- Read the Frequently Asked Questions https://nairrpilot.org/help/faq
- Check the Events and Training page for training events and office hours
  - https://nairrpilot.org/pilotevents
- Attend the NAIRR Pilot Office Hours
  - Currently 2nd Tuesday of every month, 12 PM PT/3 PM ET – check the NAIRR website for registration
- Submit a NAIRR Pilot Ticket
  - On the NAIRR Pilot website, navigate to “Help” and “Submit a Ticket”, or visit https://nairrpilot.org/open-support-request

Scan for NAIRR Pilot slides

Slide courtesy of David Hart, Lauren Michael, Timothy Middelkoop (https://nairrpilot.org/help/presentations)

Local Allocations

- Research: An Empirical Evaluation of Accuracy, Robustness and Bias of Compressed Language Models (Vivek Srikumar, KSoC)
- Research: Detecting Misalignment and Ensuring Robustness in LLMs (Daniel Brown, KSoC)
- Research: Multi-Directional Conditional Generative Models as Efficient Surrogates for Large-Scale Spatiotemporal Simulations (Shandian Zhe, KSoC)
- Research: A Foundation Model for Hematopathology: Advancing Diagnostics for Blood Disorders (Brendan O'Fallon, ARUP Labs)
- Research: Bringing Edge Computing to the National Data Platform using Sage and the Science Data Exchange (Jess Tate, SCI Institute)
- Research Start-Up (Olivia Fisher, Medicinal Chemistry)
- Demo Project: National Data Platform (Ilkay Altintas, UCSD/SDSC; Manish Parashar and Jess Tate, SCI Institute)
- Expansion Project: Creating a Community of Practice for Building Resources and Access to AI Education (Mary Lou Maher, CRA; Manish Parashar and Penny Atkins, SCI Institute)

Scaling from local to national resources:

We can help you!

Research computing and data facilitators at the Center for High Performance Computing can help you scale your work to national resources.

Please contact the CHPC at helpdesk@chpc.utah.edu to relay your needs and ask about our services and support.

- ACCESS Pipeline
- AI Ambassador

Questions and Discussion