

Shared Research Computing Policy Advisory Committee (SRCPAC)

Spring 2026 Meeting

Alexander Urban, *Chair of SRCPAC*

Bi-weekly Meetings with SRCPAC Chair, CUIT, EVPR



Alex Urban

Chair of SRCPAC



Sharon Sputz

AVP for Research Initiatives and
Development, EVPR



Axinia Radeva

Senior Manager, CUIT Research
Computing Services



Halayn Hescock

Sr. Director, CUIT Research Services



R. Yamir Gomez Carrasco

Senior Program Manager, EVPR



Max Shortte

Manager, CUIT High Performance
Computing

GET INVOLVED

Learn More About SRCPAC

Website

research.columbia.edu/content/srcpac

Mailing List

[Join the SRCPAC mailing list](#)

Contact

srcpac_cuit@columbia.edu



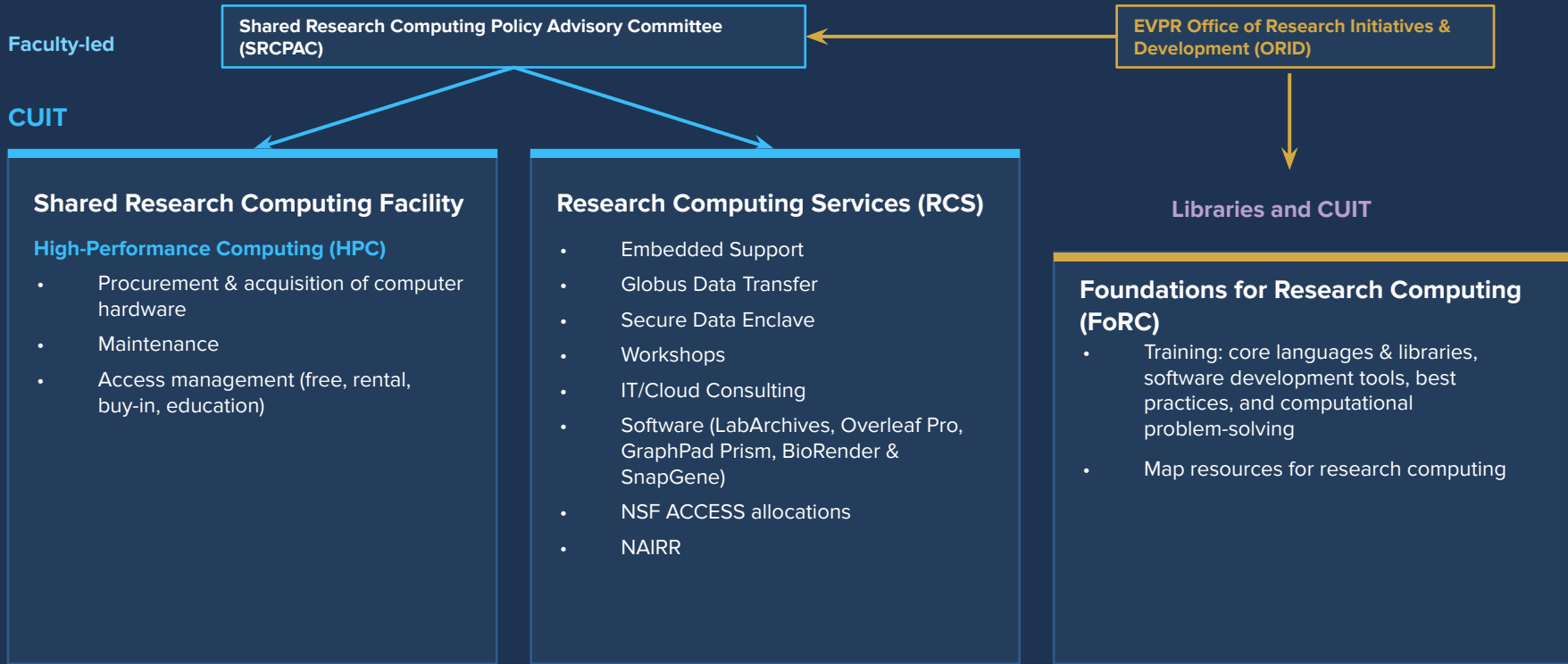
Visit our website, join the conversation, or reach out directly.

Shared Research Computing Policy Advisory Committee — Faculty-led governance of Columbia's HPC resources.

Spring 2026 Meeting

- **Welcome & Introductions**
Alexander Urban, Chair of SRCPAC
- **High-Performance Computing Update**
Max Shortte, Manager, CUIT High Performance Computing
- **Research Computing Services Update**
Axinia Radeva, Sr. Manager, CUIT Research Services
- **Research Computing Training**
Foundations for Research Computing
- **External Research Computing Resources**
Empire AI Update
ACCESS and NAIRR Resources
- **Other Business & Closing Remarks**
Alexander Urban, Chair of SRCPAC

Shared Research Computing at Columbia



High Performance Computing Updates



Alex Urban

Chair, Shared Research Computing Policy Advisory Committee (SRCPAC)



Halayn Hescock

Sr. Director, CUIT Research Services



Max Shortte

Manager, CUIT High Performance Computing

CUIT HPC Resources



CPU | GPU | RAM

Current HPC Footprint

Terremoto (reactivated)

86
NODES

2,064
TOTAL CORES (24/node)

 **68** Standard (192GB)

 **8** High Memory (754GB)

 **10** GPU Nodes

GPU Details

10 2x V100

Ginsburg (Phase 1-3)

284
NODES

9,098
TOTAL CORES (32/node)

 **190** Standard (192GB)

 **56** High Memory (768GB)

 **38** GPU Nodes

GPU Details

18 2x RTX8000
3 2x V100S
9 2x A40
8 2x A100

Insomnia

92
NODES

14,480
TOTAL CORES (80/node)

 **41** Standard (192GB)

 **19** High Memory (1024GB)

 **32** GPU Nodes

GPU Details

13 8x A6000
2 8x RTX PRO 6000
2 4x A6000
3 2x L40
9 2x L40S
3 2x H100

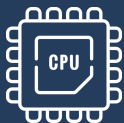
Insomnia Resources Snapshot

Overview of Insomnia Cluster



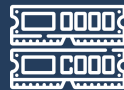
92

Total Nodes



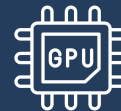
14,480

Total CPUs



60 TB

Total RAM



158

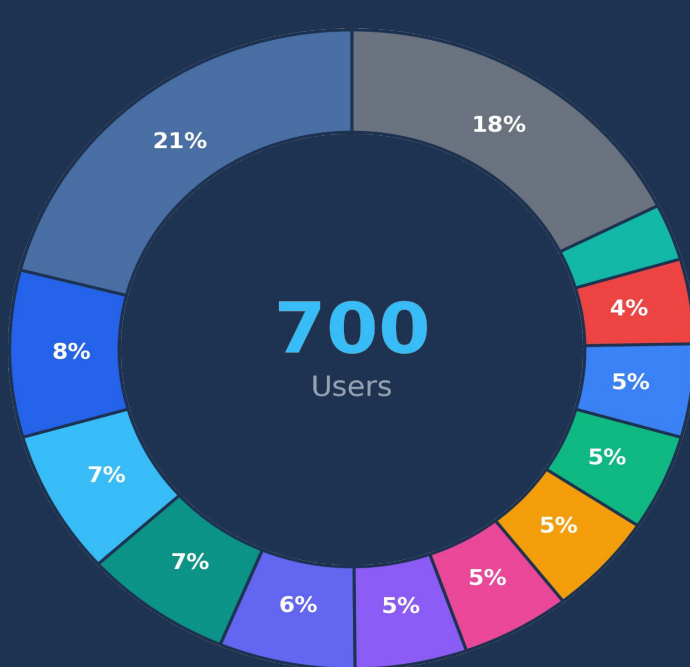
Total GPUs

Insomnia by Department



Compute Usage | Department Breakdown

Insomnia Breakdown | Department By Unique User Count



Evenly distributed, aligning with SRC PAC's shared cluster objectives.

"Other" includes: Biological Sciences, Ecology/Evolution, Pharmacology, Earth & Env. Eng., CISEIS, Applied Physics, LDEO, Psychiatry, Epidemiology, and more.

Top 10 Departments by Completed Jobs

| Rank | Department | Completed Jobs |
|------|---|----------------|
| 1 | Free/Education Tier | 883,799 |
| 2 | Systems Biology | 709,270 |
| 3 | Statistics | 411,582 |
| 4 | Physics | 281,766 |
| 5 | Biological Sciences | 219,079 |
| 6 | Social Science Computing Committee | 211,529 |
| 7 | Mailman School of Public Health | 111,728 |
| 8 | Applied Physics & Applied Mathematics | 38,014 |
| 9 | Chemistry | 16,388 |
| 10 | Ecology Evolution & Environmental Biology | 14,563 |

* Free/EDU workloads consist of many single-threaded jobs that execute and complete rapidly. Does not reflect proportional compute usage.

Top 10 Departments by Total Wall Time

| Rank | Department | Total Wall Time |
|------|---------------------------------------|-----------------|
| 1 | Systems Biology | 287,732 H |
| 2 | Physics | 245,344 H |
| 3 | Mailman School of Public Health | 190,114 H |
| 4 | Biological Sciences | 105,058 H |
| 5 | Chemistry | 93,078 H |
| 6 | Social Science Computing Committee | 91,730 H |
| 7 | Free/Education Tier | 82,084 H |
| 8 | Chemical Engineering | 65,400 H |
| 9 | Applied Physics & Applied Mathematics | 43,888 H |
| 10 | Statistics | 20,947 H |

* More accurate reflection of compute consumption

SRCPAC Autumn 2025



Structured Maintenance Strategy

Structured, predictable maintenance windows to minimise disruption



Longer Slurm Queues

Redesigned job scheduling for fairer, faster workload throughput



Improved Stability

Higher availability through proactive monitoring and rapid incident response



Improved Utilization

Better resource efficiency across compute and storage



All four commitments delivered on schedule

HPC Quarterly Maintenance Strategy



Beginning in March 2026, maintenance will occur four times per year. Each scheduled window is a standardized three-day period (Tuesday – Thursday) during the second full week of March, June, September, and December.

Q1 March

Q2 June

Q3 September

Q4 December

Benefits



Minimize disruptions



Maintain operational discipline



Enables researchers to plan around scheduled maintenance.



Align stakeholders across teams and vendors

HPC Resource Consumption by Queue

2,951,306

Total Jobs

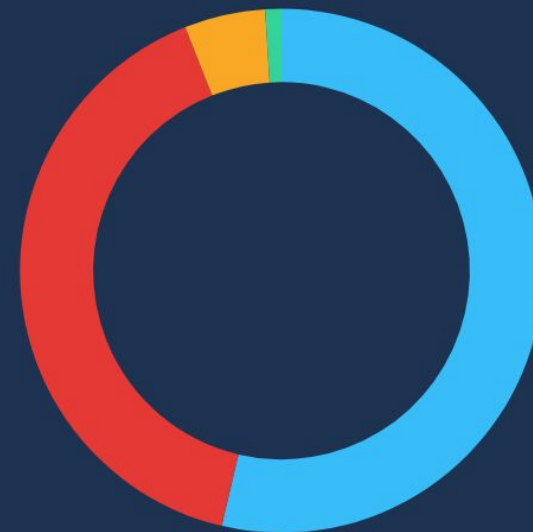
1,310,879

Total CPU-hrs

6-Month Summary

| Queue | Jobs | CPU Hours |
|------------------|-----------|-----------|
| Group Partitions | 1,726,871 | 703,585 |
| Short | 1,211,280 | 528,641 |
| Burst | 10,265 | 65,544 |
| HPC_Test | 2,890 | 13,109 |

Share of CPU Hours



* Fairshare policy to prioritize node ownership is working as intended.

* Test queue: Rapid validation, debugging & Sanity check

Group

54%

Short

40%

Burst

5%

HPC_Test

1%

Post Storage Integration Uptime

Outside of scheduled maintenance windows, since the storage integration we've seen:



Insomnia

100%

uptime



Ginsburg

99.5%

uptime



Terremoto

100%

uptime

** The current stability is driven by the storage architecture work conducted in 2025.*

HPC Statistics



HPC Statistics | July 2025 to Present

Insomnia Statistics July 2025 – March 2026

91.0%

Average

80.4%

Minimum

100.0%

Maximum

274

Days Tracked



■ Daily utilization

--- Average (91.0%)

* 91% utilization reflects high demand and adoption across multiple research areas

Ginsburg Statistics July 2025 – March 2026

88.8%

Average

75.8%

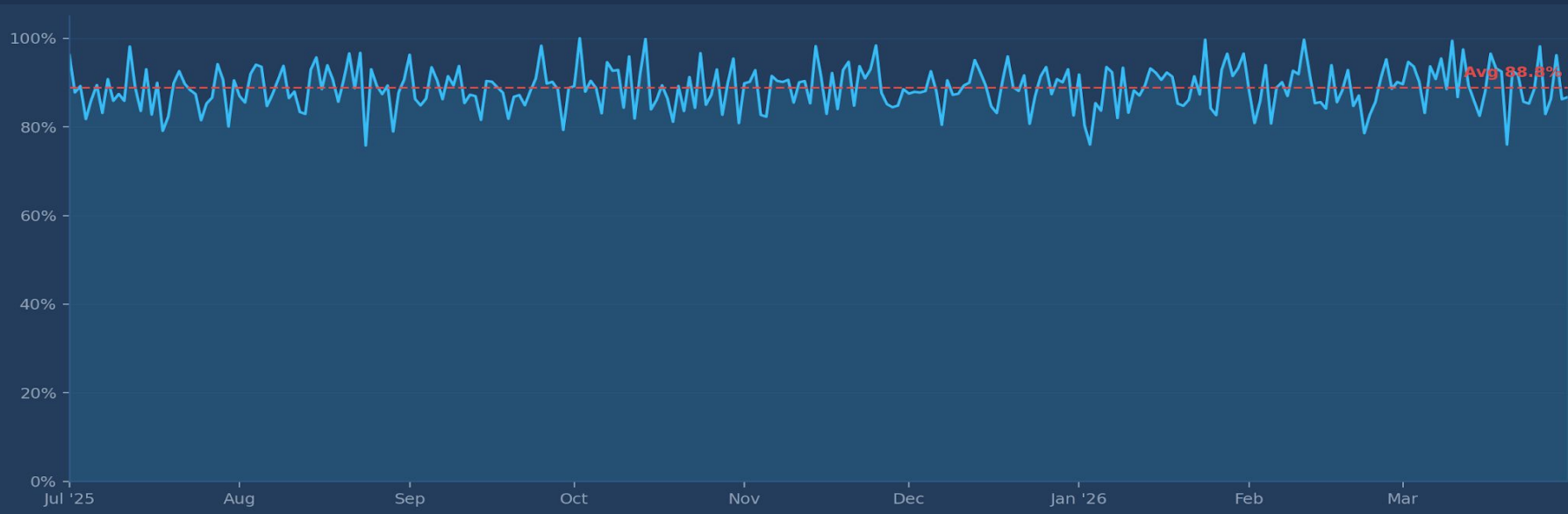
Minimum

100.0%

Maximum

274

Days Tracked



■ Daily utilization

--- Average (88.8%)

Free/Edu Tier Updates

472

Edu Accounts

628

Free Accounts

1,100

Total Accounts

2

Total Nodes

ANNUAL TIER GROWTH Each year **A&S, SEAS, and EVPR** will donate a new node to the free tier.

* *Insomnia is in its third year of operation.*

HPC Support Services



Email

hpc-support@columbia.edu — all HPC questions



Training Workshops every semester (Online)

Hands-on sessions covering Linux, HPC, Python, Singularity and best practices

** Hardware, software, improvements, optimizing workflows, potential purchases, etc*

Research Computing Services (RCS) Updates



Axinia Radeva

Sr. Manager, Research Computing Services, CUIT



Halayn Hescock

Sr. Director, Research Services, CUIT

Portfolio of CUIT RCS Shared Services

| | | |
|---|--|-----------------------------------|
| Embedded Research Computing Support | Dedicated support for research computing | |
| Secure Data Enclave (SDE) | Virtual platform for high security data analysis | |
| DSO-as-a-Service for SDE | DSO securely manages sensitive data onboarding and offboarding | NEW |
| Electronic Research Notebooks (LabArchives) | Organize, store, and share research data with backups & audit trails | |
| Research Data Transfer (Globus) | Secure, high-speed data transfers | NEW: CUIT-Managed Globus Endpoint |
| Research Software License Discounts | Discounted scientific software via CUIT's group subscriptions | |
| Overleaf Professional | Collaborative LaTeX & Rich Text writing tool for scientific publishing | |
| Cloud Research Consulting | Expert guidance on computing resources & onboarding | |
| ACCESS NSF National Resource | Try Columbia's Discover allocation & apply for national HPC resources | |
| NAIRR National Resource | National AI Research Resource | NEW |
| Empire AI | Compute resources and support for AI research workflows | NEW |

CUIT-Managed Globus Endpoint Overview



16

Collections



57

Unique Users



112.56

TB Transferred

Total data transferred since April 2025: 112.56 TB



Who It Serves

Globus helps researchers who lack IT resources to set up an endpoint.



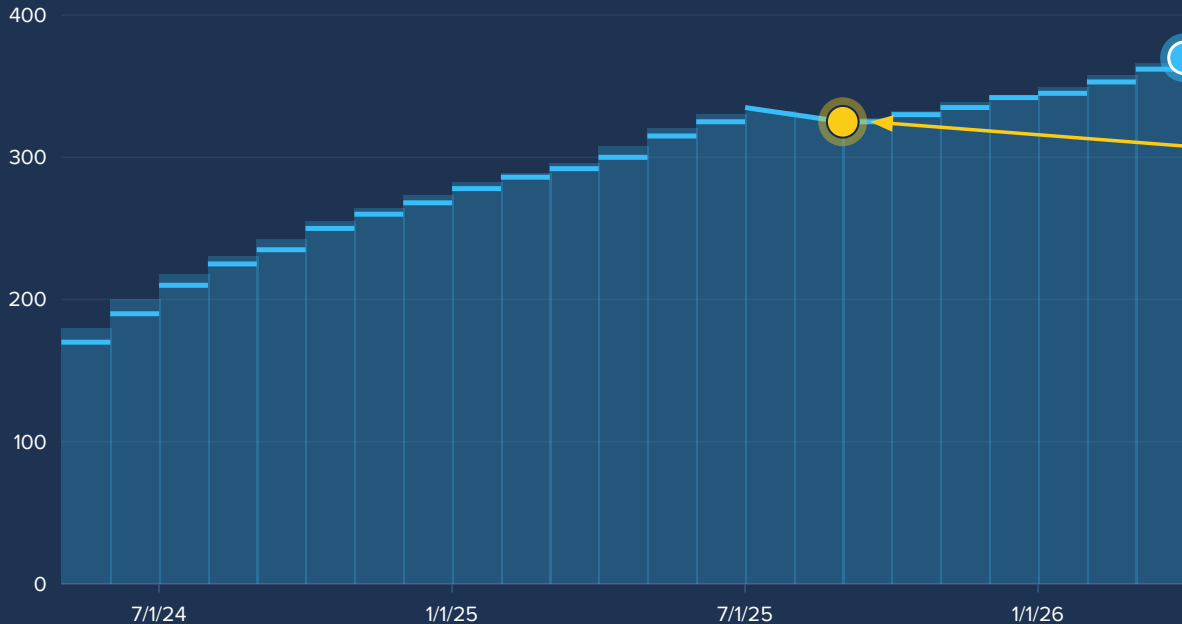
Need a Collection?

Email globus@columbia.edu to get started.

globus@columbia.edu

Globus — Research Data Transfer

Total Active Columbia Globus Subscription Users



370 as of April 2026

335 (August 2025 clean-up of inactive users when launching non-sensitive and Sensitive subscription groups)

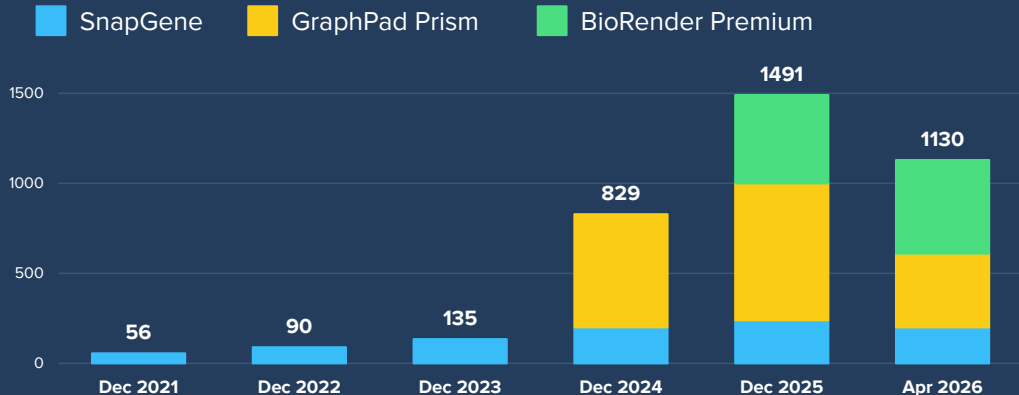
25%
increase since
May 2025

> 1,000 TB data transferred

Discounted Research Software

Discounted Software Licenses

Cumulative license adoption, Dec 2021 → Apr 2026



Price Comparison (YTD)

| Software | CUIT Rate | Academic | Saved |
|-------------------|-----------|----------|-------|
| SnapGene | \$180.00 | \$350.00 | 49% |
| GraphPad Prism | \$200.00 | \$260.00 | 23% |
| BioRender Premium | \$163.40 | \$420.00 | 61% |

Estimated savings over time

| | |
|-------------------|--------|
| SnapGene | \$150K |
| GraphPad Prism | \$189K |
| BioRender Premium | \$139K |

Total Saving to Researchers \$478K



200+ labs/departments served across all campuses

LabArchives Growth & Adoption (2016–2026)

LabArchives celebrates ten years at Columbia as our Electronic Research Notebook service.



4.8M

Recorded Activities



17K

Notebooks



+10%

Growth in 2025

Cumulative usage

- Over 4.8M recorded activities
- Over 17K notebooks
- This platform achieved 10% growth in 2025 with high engagement and activity levels.
- An increasing notebook-to-account ratio indicates improved organization for complex projects with collaboration.

LabArchives Collaboration & Research Impact

3K+ Extensive External Collaboration

3K external collaborators currently working in Columbia-owned notebooks

Partners include Icahn School of Medicine at Mount Sinai, OSU, Mass General Brigham, and Duke University.



Education Edition

Course Notebooks serving 6K students

Consistent pipeline: ~500 new student accounts and ~600 student notebooks created annually.



Strategic Engagement

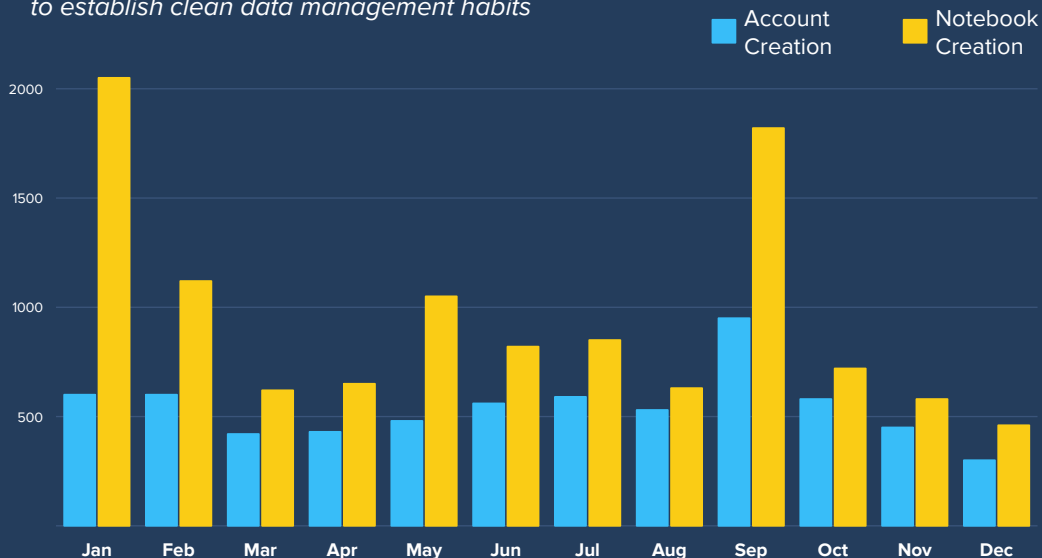
Love Data Week

Annual international event

Partnering with the Libraries and LabArchives for Love Data Week.

Monthly Notebook and Account Creation

January is the peak month for new notebook creation, aligning with Love Data Week to establish clean data management habits



Research Resources, Training, and Support

Websites



Research Resources Hub

<https://www.cuit.columbia.edu/research>

- Research system login links
- Newly offered tools
- Research Services portfolio overview
- FAQ



Trainings & Events

<https://www.cuit.columbia.edu/rcs/training>

- Calendar of research events across Columbia
- Expanding library of training recordings

Contact / Support

Research Computing Services (RCS)



Email

rcs@columbia.edu



RCS

[cuit.columbia.edu/rcs](https://www.cuit.columbia.edu/rcs)

Foundations for Research Computing Update



Daniel Woulfin

Computational Research Instruction Librarian



Jessica Eaton

CUIT Research Analyst

Next up on the agenda

A continuation of Columbia's investment in research software, training, and collaborative infrastructure.

Mission

Foundations for Research Computing is a collaborative initiative aimed to connect researchers across Columbia University with informal, non-credit opportunities to learn about research computation and how to work within the research computing ecosystem at Columbia University.

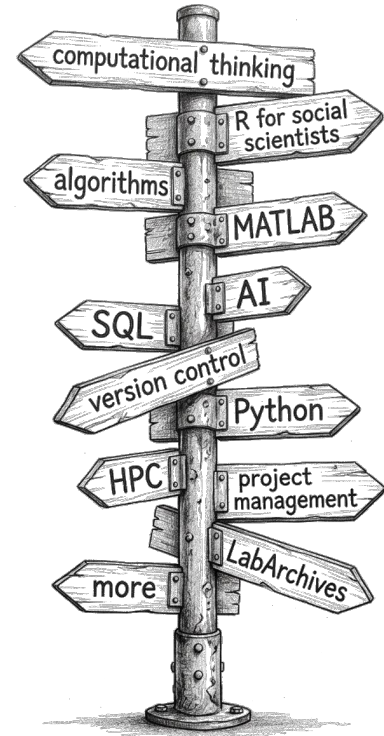


Our goal is to help researchers and the people who support them learn the hard and soft skills needed to do computational research and how to use Columbia's shared computing resources.

A new approach

Prior to this year, Foundations primarily consisted of a two-day Python/R bootcamp based off the Carpentries. The bootcamps were held before each semester.

The new model takes a different approach. We offer sessions throughout the semester using a learning path model where attendees can pick and choose what they want to attend and learn from a longer list of topics offered by RCS, the Libraries, and other partners.



Types of training provided today

Fundamentals

- Python, R, SQL
- Computational thinking
- Algorithms and machine learning
- Data wrangling and visualization
- Version control
- Non-coding research tools (Overleaf, Nvivo, etc.)

Applications

- AI and cloud computing platforms
- Project and data management in the lab (Globus, LabArchives)
- HPC training
 - Linux on the cluster
 - Automating tasks
 - Parallelization with Python
 - Containerization

Program and Trainings Attendance

40 individual trainings

- 27% return rate

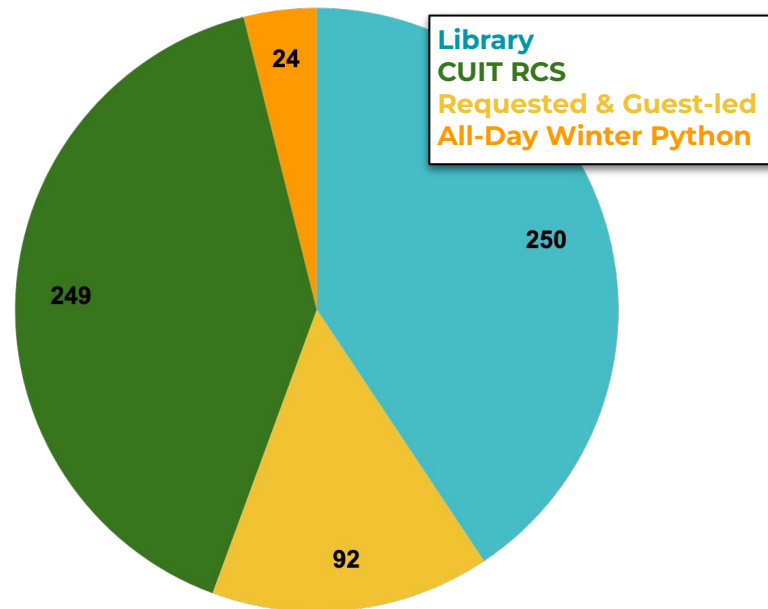
Attendees from all schools

- 26% VP&S and Mailman
- 23% A&S/GSAS

Attendees from all levels

- 40% Grad students
- 18% Postdocs
- 18% Staff
- 9% Undergrads
- 7% Faculty
- 6% Associate Research Scientists

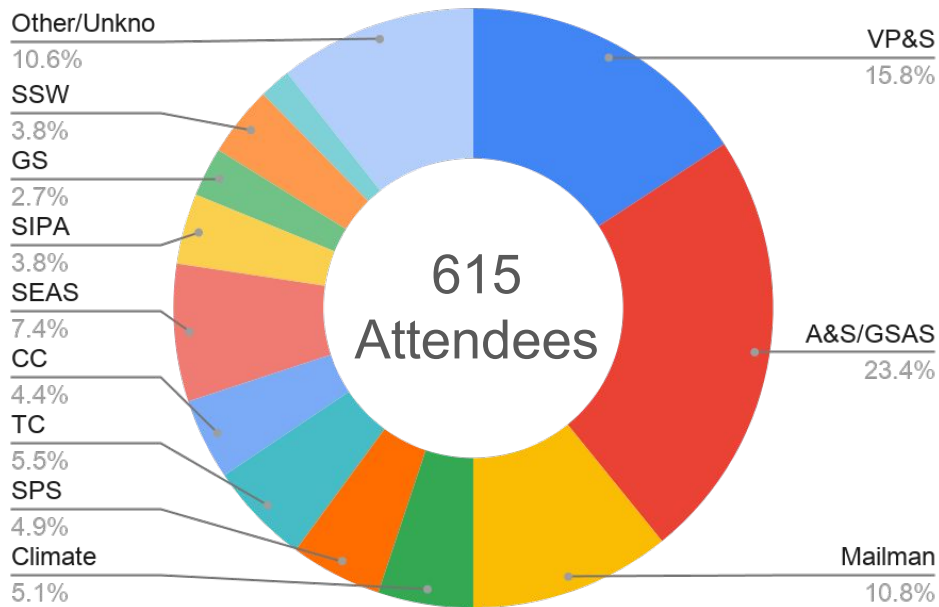
Fall 2025 + Spring 2026*:
615 Attendees



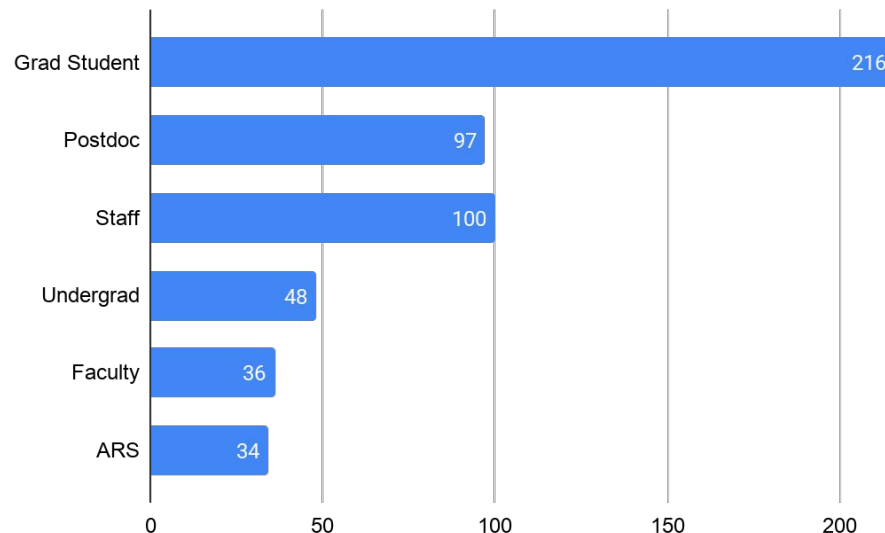
****2 workshops still need to be incorporated***

Attendee Representation

Schools Represented



Attendee Roles



- 40% Grad students
- 18% Postdocs
- 18% Staff

FY26 Feedback (144 respondents)

Positive responses

"Learning how to use the HPC is very useful for working with larger datasets such as sequencing data" [*Intro to Bash*]

'Easy to understand and follow for person who has 0 background like me.' [*Foundations Winter Python Workshop*]

"I LOVED this workshop. It was honestly the most efficient and educational two-hour workshop of all time." [*Learn How to Learn to Code: Python*]

Constructive responses

"It would be ideal to offer a continuation of this session so that one continues learning" [*R 101 for Social Scientists*]

"I think these workshops lack context...a few overview slides could be really helpful." [*Intro to Linux*]

"It might be helpful to use a whiteboard or simple visuals to show how data flows through various clusters and workers..." [*Parallel Computing with MATLAB*]

62 Net Promoter Score:
Recommended to others

5.89 effort score (out of 7)

6.57 teaching satisfaction
(out of 7)

Requested and guest workshops

One area we'd like to strengthen is our requested and guest workshops. These are different from open workshops as they allow us to customize teaching to specific population needs. In the last year we've given guest and requested workshops to:

- Economics faculty, student and staff with MATLAB engineer focusing on AI/ML applications
- Social Intervention Group at the School of Social Work
- Global Mental Health Lab at Teachers College
- Columbia Population Research Center
- SIPA students
- Climate School students

These have been mostly word of mouth requests.

To request, email rcfoundations@columbia.edu

Software
best
practices

AI and cloud
compute
workshops

Project
management
workshops

Coding and
research tool
workshops

Topics by
request

What needs do you see for computational training in FY2027?

Our new model isn't perfect and there's still limitations and capacity issues.

That's why we need to hear from you!

- What do you want to see from Foundations?
- What training needs do you see as lacking?
- Where are we not where we need to be?

Email rcfoundations@columbia.edu



**WE WANT TO HEAR
FROM YOU**

How to keep up with Foundations trainings



[Subscribe to Foundations mailing list](#)

Bookmark for event calendar and resources:
www.rcfoundations.research.columbia.edu

External Research Computing Resources

Axinia Radeva, Sr. Manager, Research Computing Services, CUIT

Empire AI

NSF ACCESS - Advanced Cyberinfrastructure Coordination Ecosystem: Services & Support

NAIRR - National Artificial Intelligence Research Resource

Bridging the Gap to Large-Scale AI Resources

AI is transforming fields from healthcare to public policy, driving demand for large-scale computing, data, and collaboration.

However, much of this infrastructure is concentrated in the hands of a few major technology companies, limiting academic access to the scale required for modern AI workflows.

Access to state and national computing resources has become increasingly critical.

RCS STRATEGIC FOCUS

The focus of the RCS team is to support researchers in obtaining these vital state and national resources.

Empire AI Update

Launched by Governor Kathy Hochul in April 2024 in partnership with New York's leading public and private institutions, Empire AI is a NYS initiative to advance artificial intelligence.

ALPHA System

24 HGX Nodes • 8 H100 80GB GPUs per node • 8 H200 141GB GPUs per node
Fair-share scheduling on a rotating priority access system

CUIT's High Performance Computing (HPC) and Research Computing Services (RCS) teams provide support.

344

Completed Jobs

4,428h

Job Run Time

3,731

GPUs Allocated

43

Accounts

14

Projects

12

PIs

Future Phases of Empire AI

May 2026

Beta System Operational

Early 2027

Gamma Facility Construction Starts

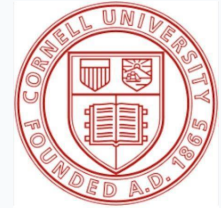
Mid-late 2027

Gamma Facility Opening

Early 2028

Gamma System Operational

Consortium Members



CUIT-Led Implementation for Empire AI

SU-Based Billing Engine

Engineered a service-unit-based billing model for CPU, GPU, and storage usage

ColdFront Integration

Intake workflows, SU allocation, and top-up mechanisms built into ColdFront

iLAB Validation Layer

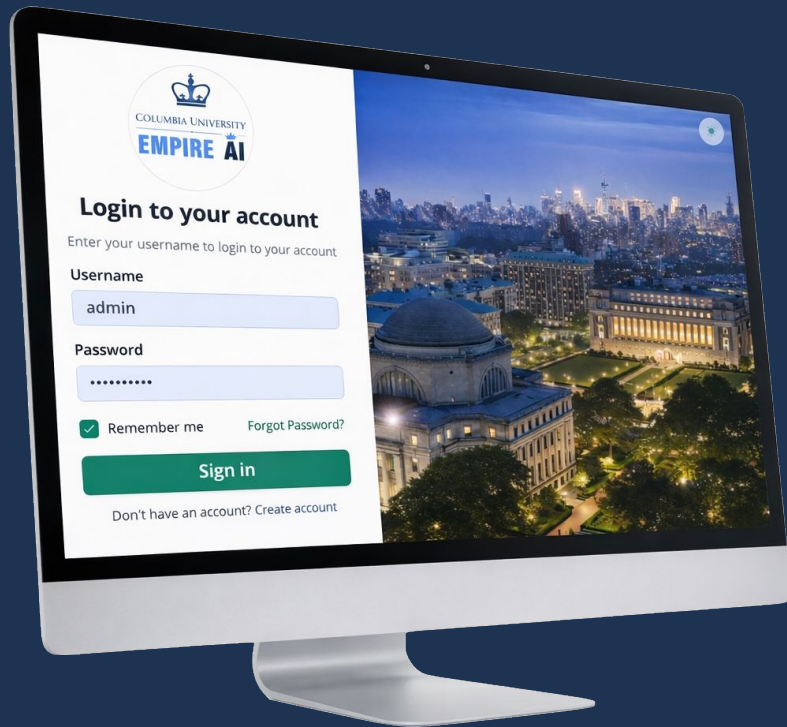
iLAB integration for financial validation and reconciliation with ColdFront

Unified Consortium Portal

Single portal for all consortium members to manage resources and accounts

Responsive UI

Develop a mobile and desktop friendly interface for member access and management



Empire AI | Queue Adoption

Empire AI has adopted CUIT's queue architecture, reducing the learning curve and making cross-pollination between clusters easier.



Familiar Workflow

Researchers moving between CUIT HPC and Empire AI encounter the same SLURM queue patterns they already know.

Reduced Learning Curve

Consistent queue naming and scheduling policies mean less onboarding friction for new Empire AI users.

Cross-Pollination

Shared architecture enables workloads to move fluidly between on-prem clusters and Empire AI resources.

ACCESS Resources



ACCESS (Advanced Cyberinfrastructure Coordination Ecosystem: Services & Support) is an NSF program offering supercomputing resources to researchers.



The Research Computing Services (RCS) team provides crucial support for finding, using, and navigating these resources.

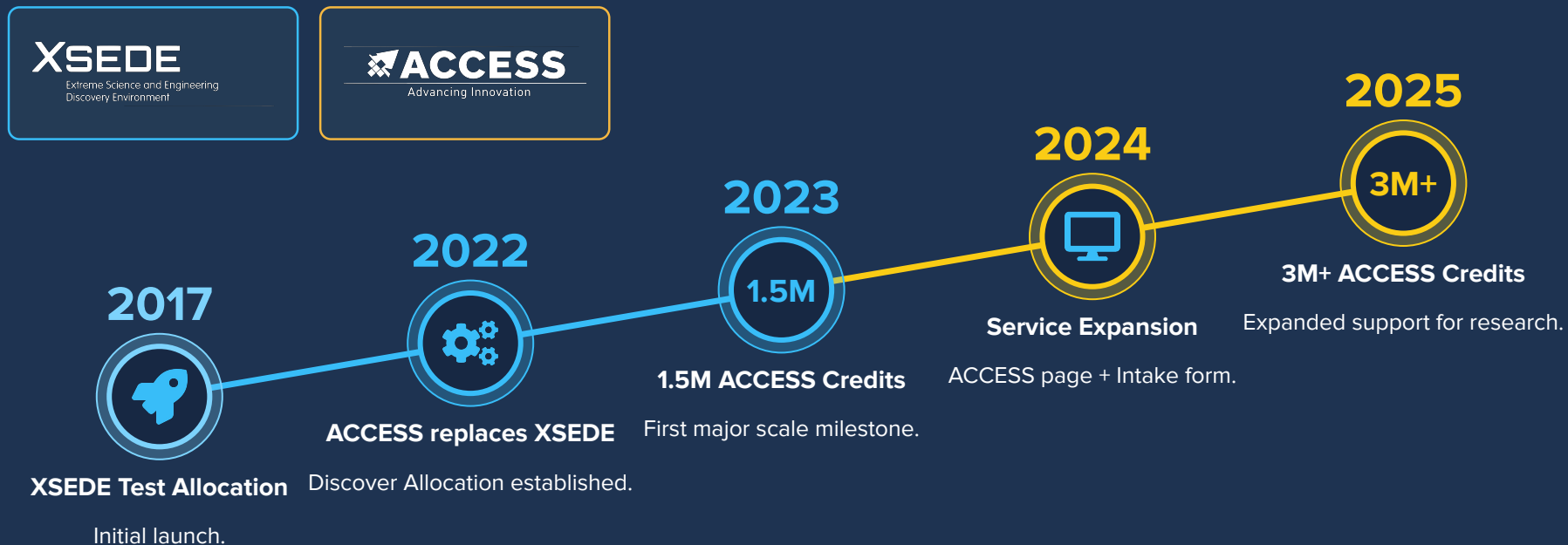


Researchers can onboard quickly with a low-barrier-to-entry resource line the **Columbia Discover Allocation**.



RCS assists researchers transition from the **Discover Allocation** to their own sustained, scaled-up workflows.

Scaling National HPC ACCESS for Columbia Research

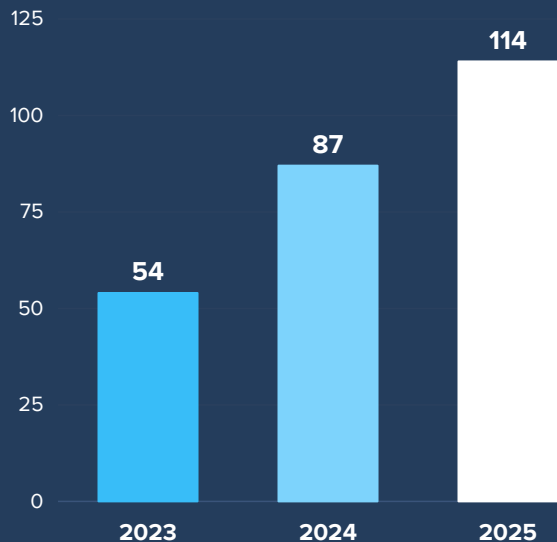


From initial access to 3M+ credits — eight years of expanding national HPC for Columbia researchers

ACCESS Growth

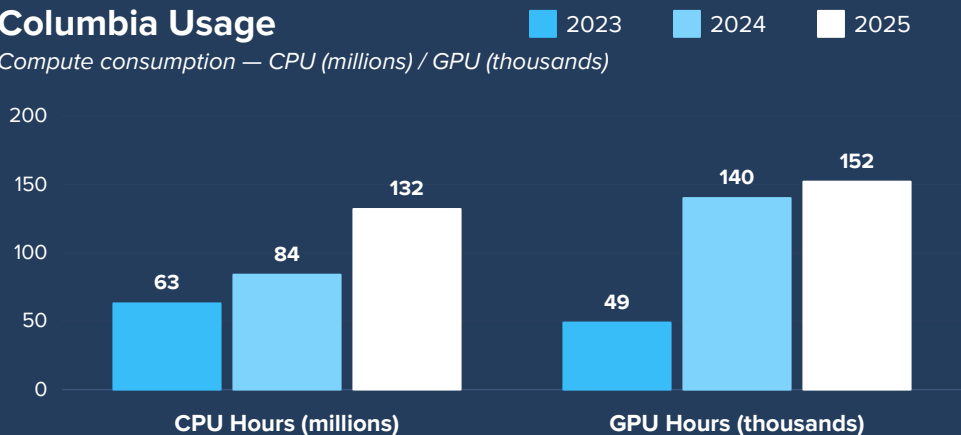
Columbia Users

Active users — annual snapshot



Columbia Usage

Compute consumption — CPU (millions) / GPU (thousands)

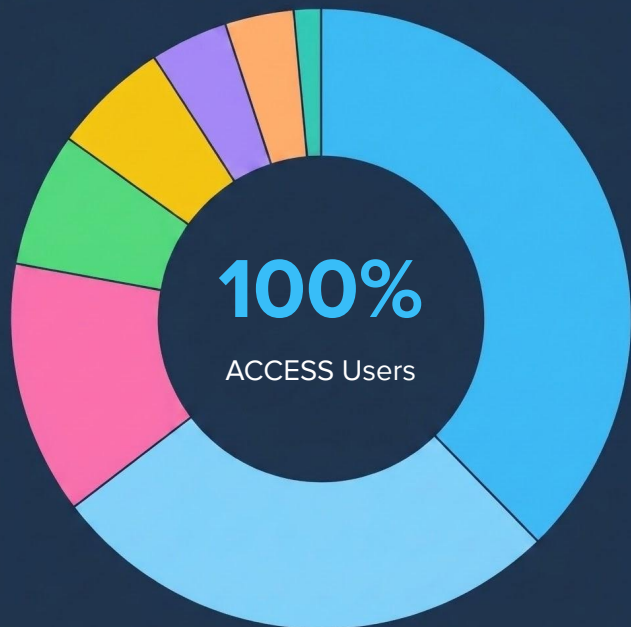


Most Popular Resources



ACCESS Metrics

Users by School



| | |
|---------------------------------|-------|
| SEAS | 37.7% |
| Arts and Sciences | 36.0% |
| CUIMC | 10.1% |
| Zuckerman | 7.0% |
| Barnard | 3.5% |
| Lamont-Doherty - Climate School | 3.0% |
| SIPA | 1.8% |
| Business School | 0.9% |

NAIRR Resources



NAIRR (The National Artificial Intelligence Research Resource) is a U.S. government initiative for national AI computing resources. Its purpose is to democratize access to high-end AI technology for research.



Research Computing Services (RCS) supports researcher in discovering, accessing, and navigating NAIRR.



This support includes guidance on proposal submissions and resources utilization.



Columbia has already secured **three research allocations** through the 2024 Pilot.

Research Computing Ecosystem at-a-glance

| | SRCPAC On-Prem HPC | Empire AI | ACCESS | NAIRR |
|--------------|--|--|---|--|
| Scope | Columbia shared HPC clusters | NYS consortium | NSF cyberinfrastructure | National AI research resource |
| Eligibility | Columbia researchers, faculty, students (via PI or allocation) | Researchers affiliated with the consortium | U.S. researcher/educator, graduate-student level or higher | U.S. researchers/educators from U.S. institutions, non-profits, or federal agencies |
| Resources | CPU + GPU (A6000/RTX/V100) | H100/H200 GPUs; AI stacks | CPU/GPU/HTC cloud/storage/software resources | GPU + datasets/models; NAIRR Secure |
| Best fit | General HPC + moderate AI/ML | Large-scale AI training/foundation models | Scale-out HPC + moderate AI/ML | AI model dev/training/benchmarking |
| Type of Data | Non-Sensitive at this time | Non-Sensitive at this time | Non-Sensitive | Non-Sensitive data; Secure option |
| Cost | Condo model | Usage-based; university subsidized | Free via allocation (NSF-funded) | Free via allocation |
| Support | Dedicated CUIT HPC Team Support, training workshops, documentation | Dedicated institutional support (HPC teams, onboarding, consultations) | Centralized support (help desk, documentation, training, community support) | Coordinated support across agencies + providers (training, onboarding, AI-specific guidance) |



Questions?