Diamond: Democratizing Large Foundation Model Training for Science

Zhao Zhang

Department of Electrical and Computer Engineering

Rutgers, the State University of New Jersey

The Progression of the Scientific Method

Increasing speed, automation, and scale

Empirical Science

1st Paradigm

Observation
Experimentation

Theoretical Science

2nd Paradigm

Scientific Laws in Physics, Chem, and others

Computational Science

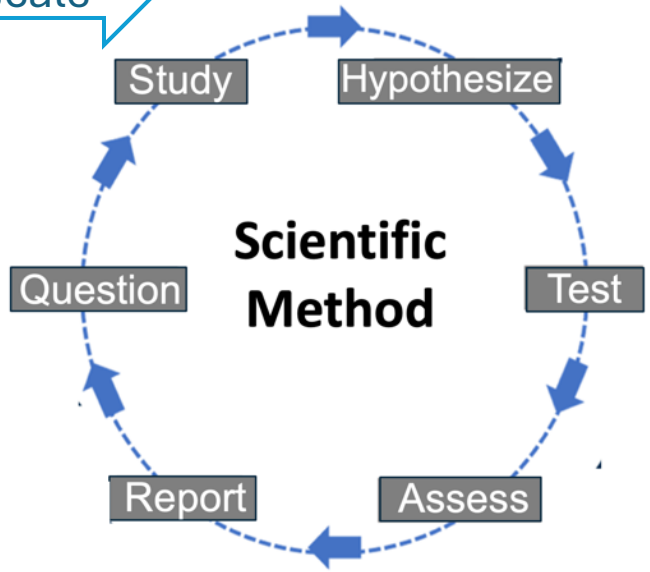
3rd Paradigm

Simulations
Molecular Dynamics
Mechanistic Models

Big Data-driven Science

4th Paradigm

Big data, machine learning
Patterns, anomalies
Visualization



Scientific knowledge at scale
AI-generated hypotheses
Autonomous testing

Credits: Ian T. Foster, UChicago

ML/DL in Science not So Long Ago

Predictions, e.g., transient events

Controller

Actions

Telescopes, Edge Devices

Observations

Observations

Storage & Archiving

Raw Data

Data Processing

Processed Data

Model Training

Model

Model Serving

2012-2017 Research Focus

ML Pipeline Diagnosis

Celestial Object Detection with Spark

Parallel Scripting

amplab

Berkeley | EECS
Electrical Engineering and Computer Sciences

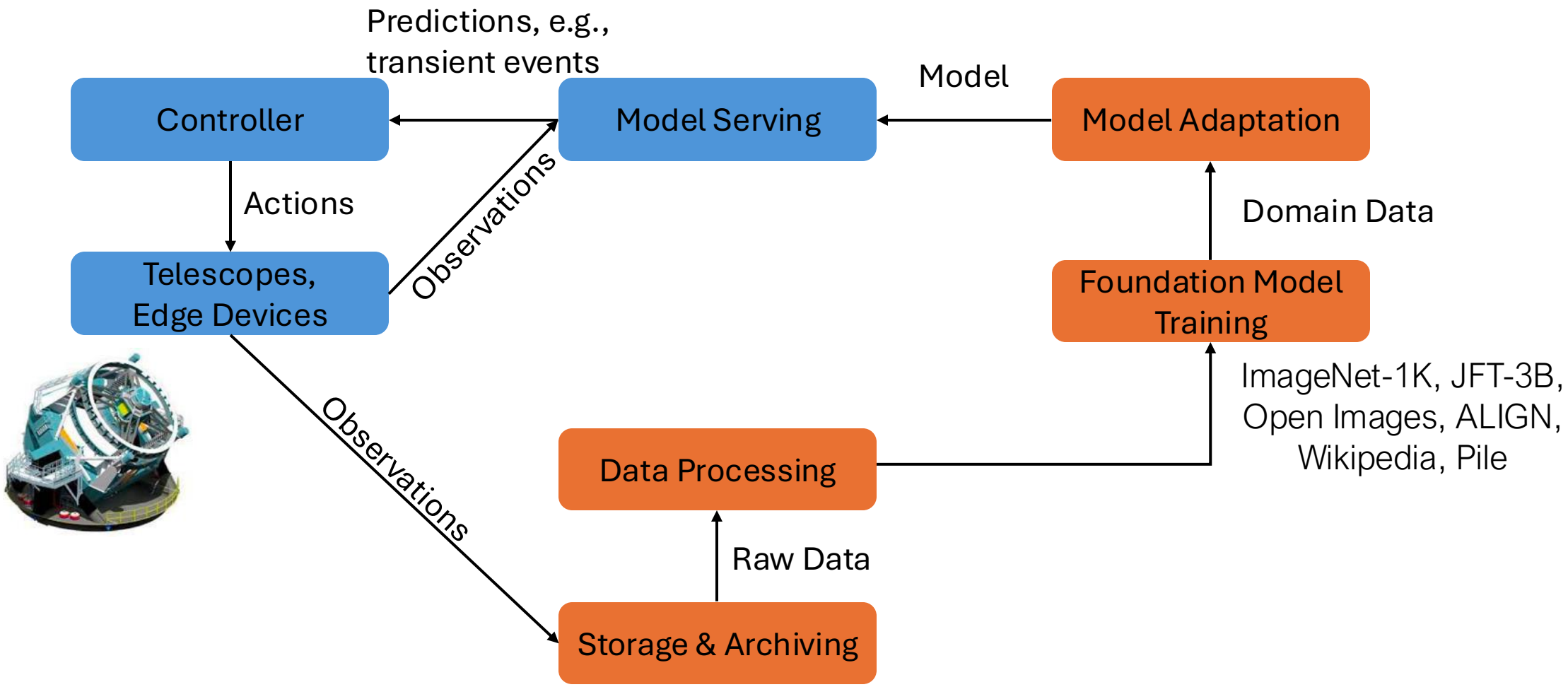
BERKELEY Institute for Data Science

BERKELEY CENTER for COSMOLOGICAL PHYSICS

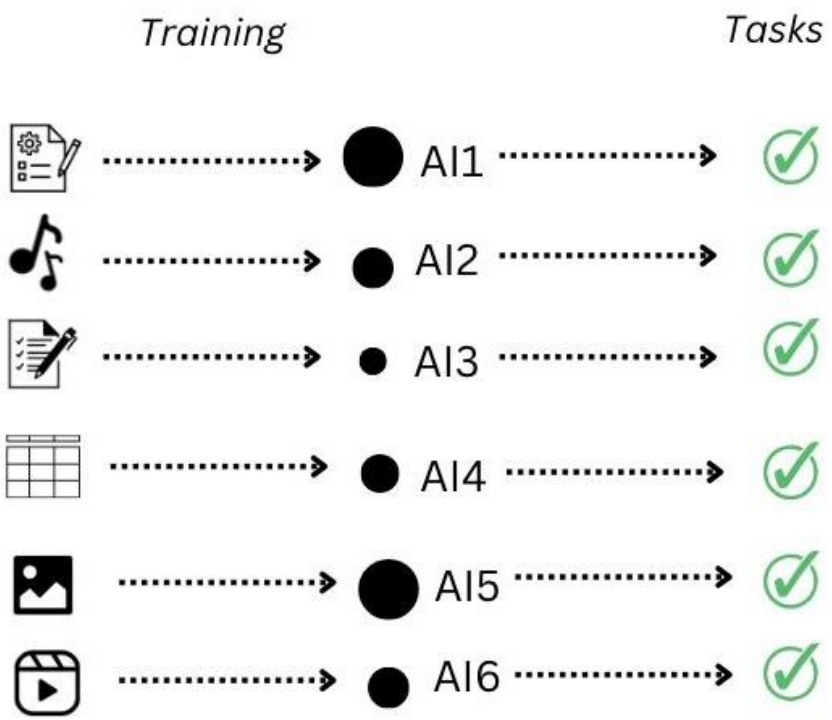
THE UNIVERSITY OF CHICAGO DEPARTMENT OF COMPUTER SCIENCE

Argonne NATIONAL LABORATORY

ML/DL in Science not So Long Ago

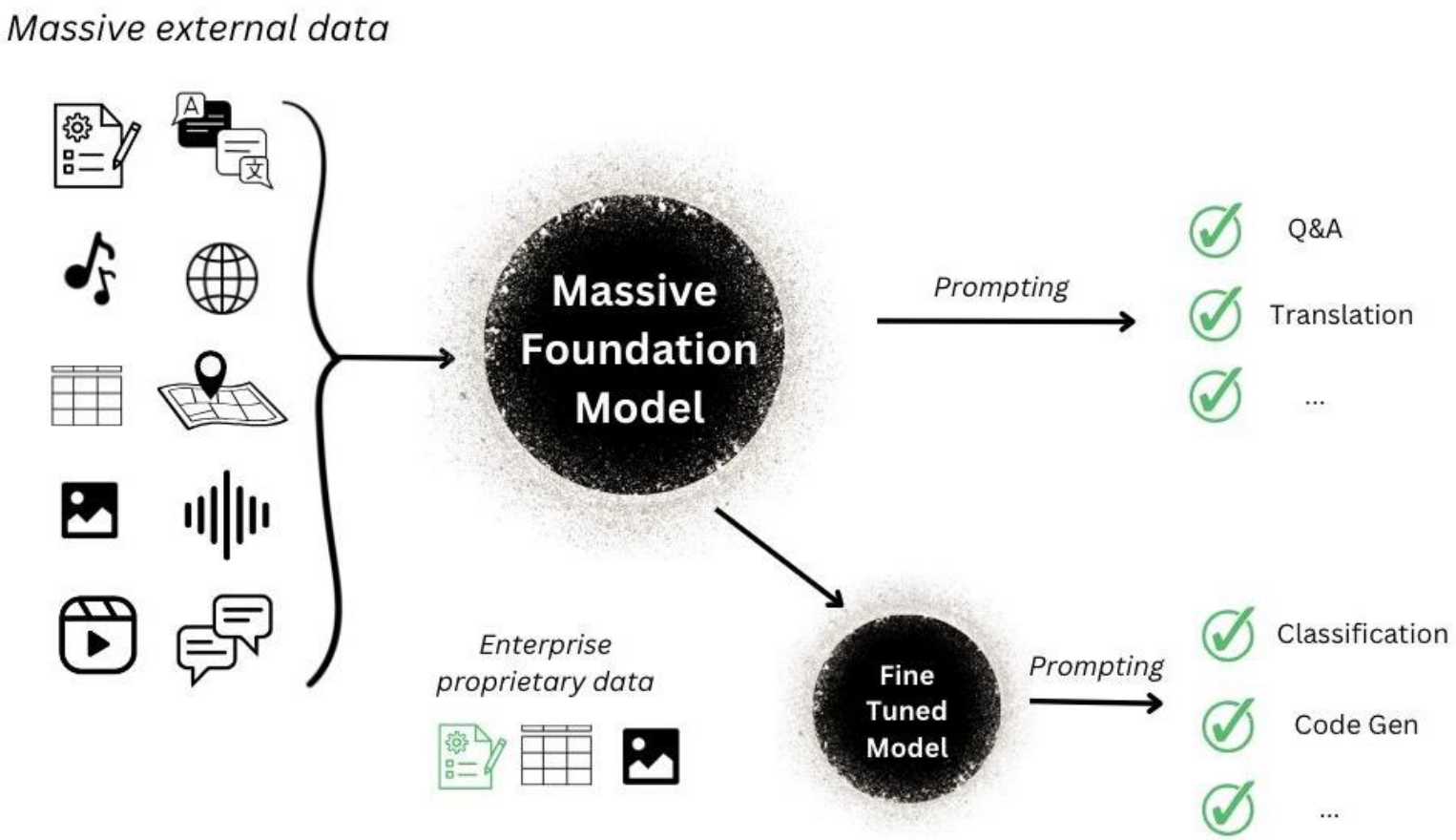


Traditional ML



- Individual siloed models
- Require task-specific training
- Lots of human supervised training

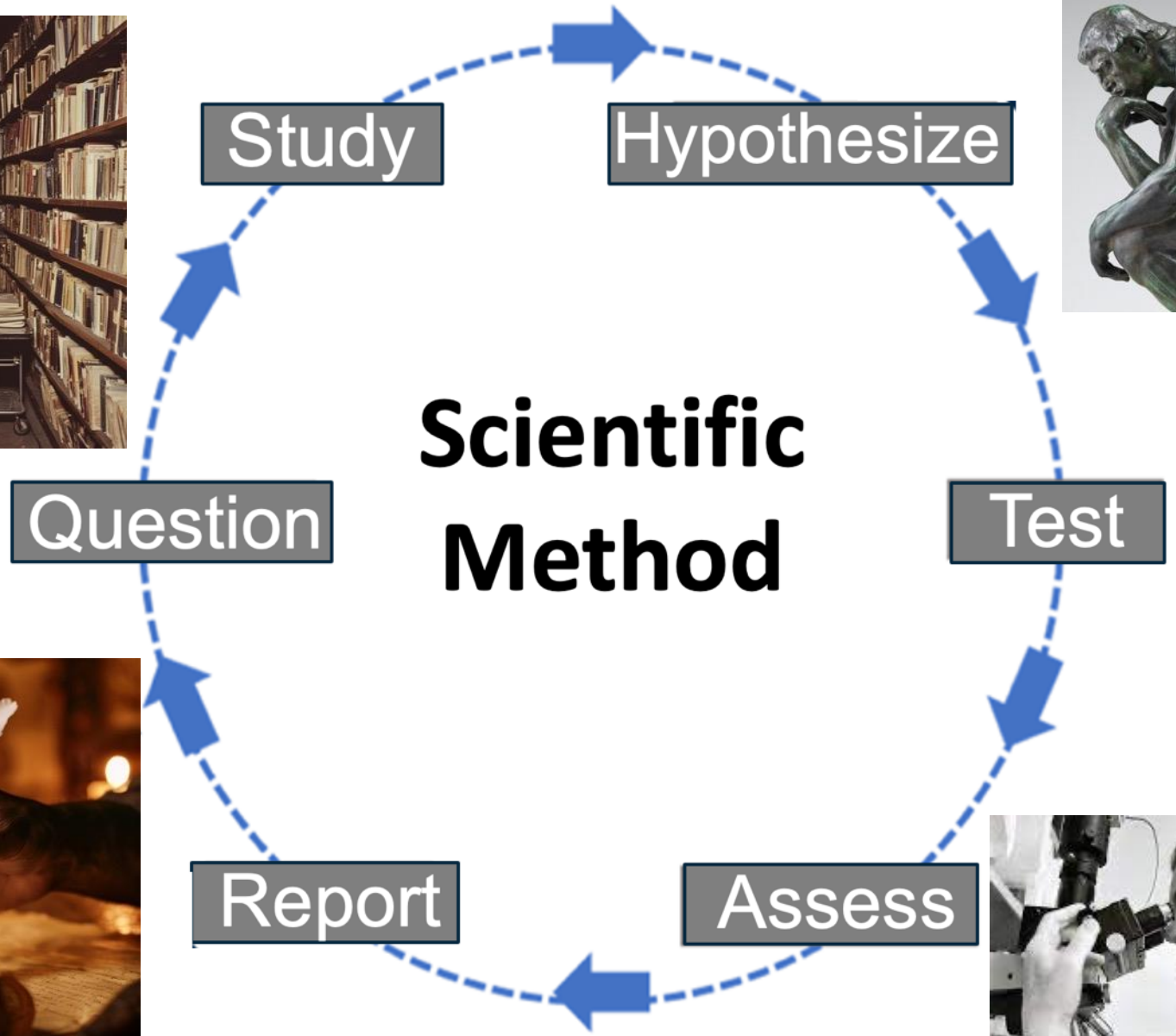
Foundation models

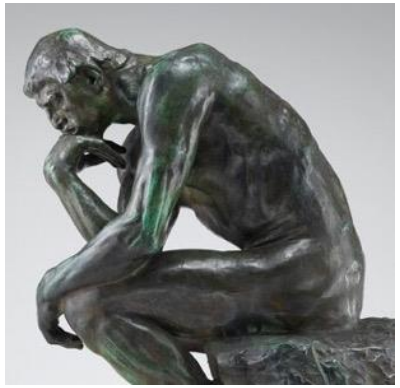


- Massive multi-tasking model
- Adaptable with little or no training
- Pre-trained unsupervised learning

Credits: Ian T. Foster, UChicago

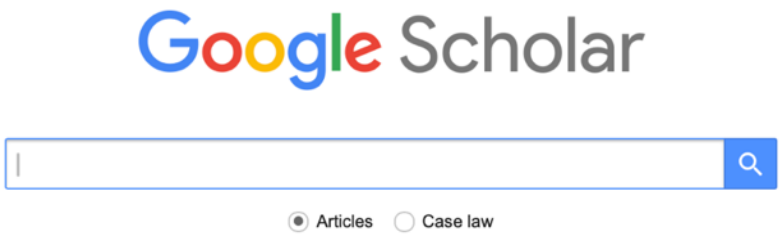
The scientific method remains slow and labor-intensive

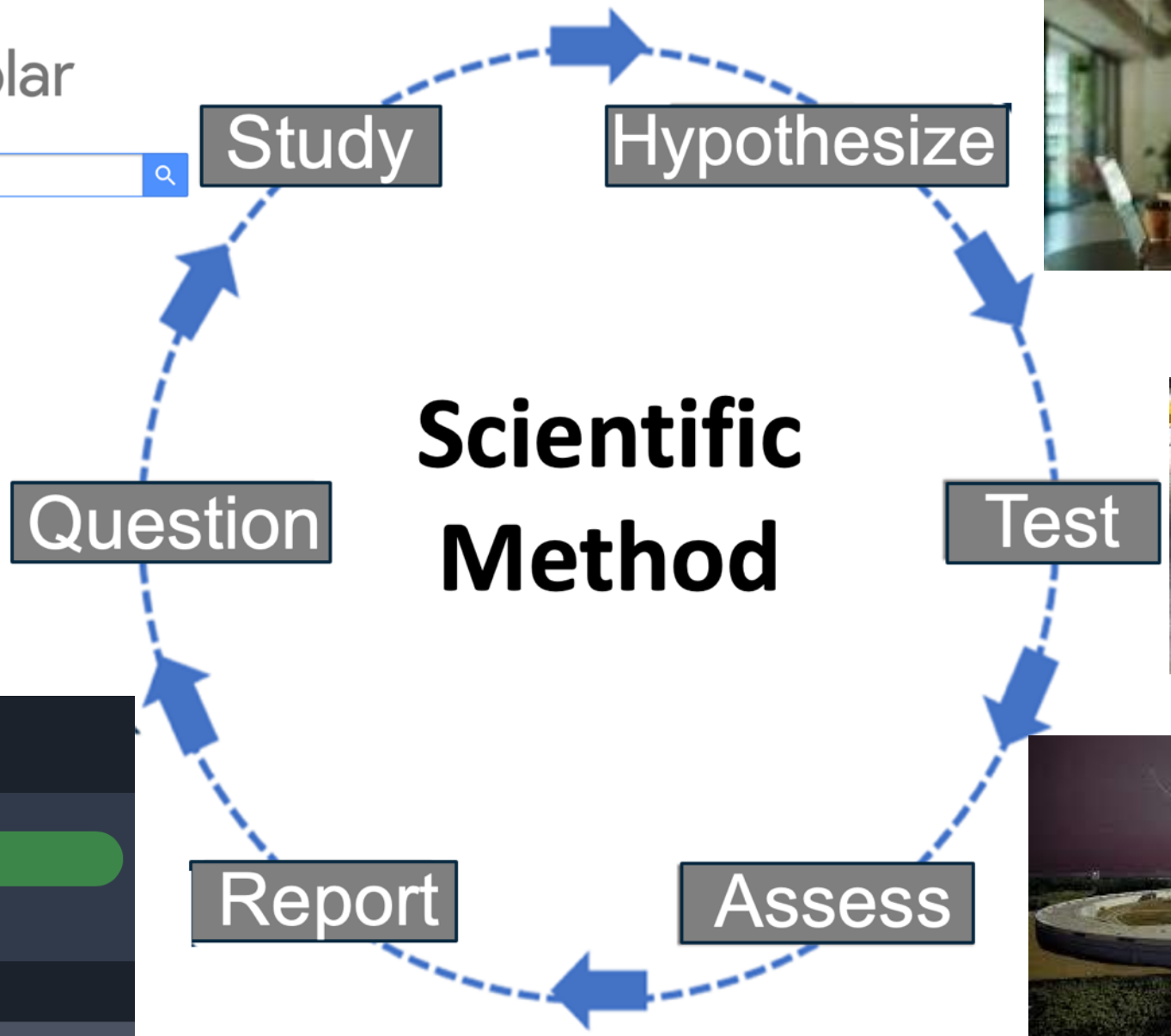


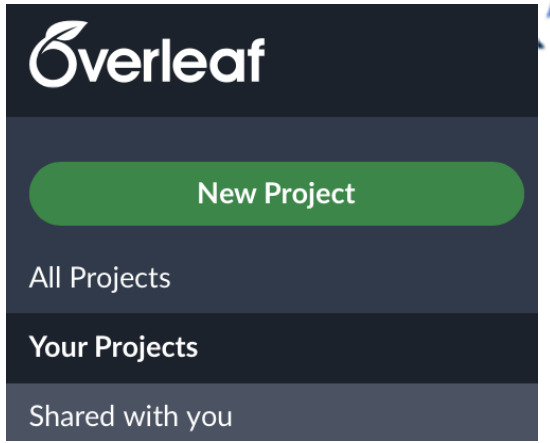


Credits: Ian T. Foster, UChicago

Despite acceleration of some steps via HPC etc.





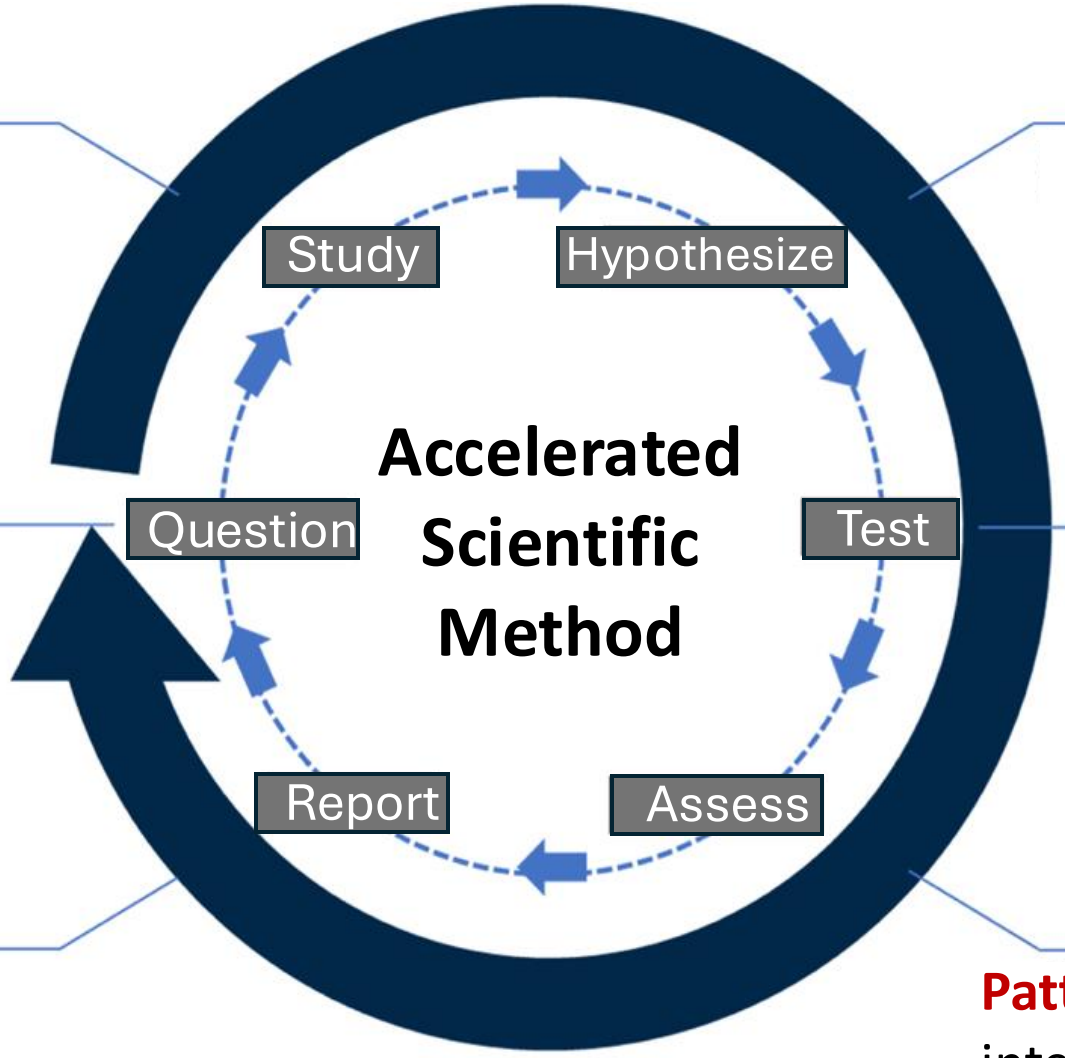


Credits: Ian T. Foster, UChicago

Engage AI assistants to help overcome bottlenecks

Extraction, integration and reasoning with knowledge at scale

Generative models automatically propose new hypotheses that expand discovery space



Tools help **identify new questions** based on needs and gaps in knowledge

Robotic labs automate experimentation and bridge digital models and physical testing

Machine representation of knowledge leads to new hypotheses and questions

Pattern and anomaly detection integrated with simulation and experiment extract new insights

https://doi.org/10.1038/s41524-022-00765-z

Credits: Ian T. Foster, UChicago

Industry Investment in AI Cyberinfrastructure

RESEARCH

Introducing the AI Research SuperCluster — Meta's cutting-edge AI supercomputer for AI research

RSC: Under the hood

AI supercomputers are built by combining multiple GPUs into compute nodes, which are then connected by a high-performance network fabric to allow fast communication between those GPUs. RSC today comprises a total of 760 NVIDIA DGX A100 systems as its compute nodes, for a total of 6,080 GPUs — with each A100

Tesla Unveils Top AV Training Supercomputer Powered by NVIDIA A100 GPUs

'Incredible' GPU cluster powers AI development for Autopilot and full self-driving.

Stability AI, the startup behind Stable Diffusion, raises $101M

Kyle Wiggers @kyle_l_wiggers / 12:01 PM CDT • October 17, 2022 Comment

Stability AI has a cluster of more than 4,000 Nvidia A100 GPUs running in AWS, which it uses to train AI systems, including Stable Diffusion. It's quite costly to maintain — Business Insider reports that Stability AI's operations and cloud expenditures exceeded $50 million. But Mostaque has repeatedly asserted that the company's R&D will enable it to train models more efficiently going forward.

Nvidia and Microsoft team up to build 'massive' AI supercomputer

NVIDIA

Microsoft

/ The companies hope to create 'one of the most powerful AI supercomputers in the world,' capable of handling the growing demand for generative AI.

By JESS WEATHERBED

Meta's Llama 3.1 405B model was trained using **over 16,000 NVIDIA H100 GPUs**. This was the first Llama model to be trained at this scale.

Explanation

- The training process for Llama 3.1 405B required a large amount of computing power.
- Meta optimized their training infrastructure to handle the model's scale.
- The model was trained on over 15 trillion tokens.
- The training process took 54 days.

xAI Colossus is **a supercomputer built by xAI, a company founded by Elon Musk, to train and power the AI chatbot Grok**. It's located in Memphis, Tennessee, in a former Electrolux manufacturing plant.

Features:

- **GPUs**: The supercomputer has over 100,000 Nvidia H100 GPUs, which are some of the most powerful processing chips available
- **Liquid cooling**: The GPUs are liquid-cooled
- **Networking**: The supercomputer uses Nvidia Spectrum-X Ethernet networking
- **Storage**: The supercomputer has exabytes of storage

National Investment in AI Cyberinfrastructure

- To accommodate the increasing need of HPC for AI, the US government has heavily invested in supercomputers:
  - TACC Horizon, O(1000) GPUs, to deploy in 2026, funded by NSF LCCF
  - NERSC Perlmutter, +7,000 Nvidia A100s, deployed in 2021
  - ALCF Polaris, +2,000 NVIDIA A100s, deployed in 2022
  - OLCF Frontier, 37,888 AMD MI250X GPUs, deployed in 2021
  - ALCF Aurora, 63,744 Intel GPU Max Series, access open a week ago

National Investment in AI Cyberinfrastructure

The National Artificial Intelligence Research Resource (NAIRR) Pilot

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. Federal agencies are collaborating with government-supported and non-governmental partners to implement the Pilot as a preparatory step toward an eventual full NAIRR implementation.

Operational focus areas

NAIRR Open

This focus area, led by NSF, will support open AI research by providing access to diverse AI resources via the NAIRR Pilot Portal and coordinated allocations.

NAIRR Secure

This focus area, co-led by the National Institutes of Health and the Department of Energy, will support AI research requiring privacy and security-preserving resources and assemble exemplar privacy-preserving resources.

NAIRR Software

This focus area, led by NSF, will facilitate and investigate interoperable use of AI software, platforms, tools and services for NAIRR pilot resources.

NAIRR Classroom

This focus area, led by NSF, will reach new communities through education, training, user support and outreach.

Filters

Resource Category

- [x] Federal agency systems
- [ ] Private sector computational resource
- [ ] Private sector model access
- [ ] Other private sector contribution

Resource Type

- [ ] Cloud
- [x] GPU Compute
- [ ] Innovative / Novel Compute
- [ ] CPU Compute
- [ ] Service / Other

Reset Filters

Resources

- Indiana Jetstream2 GPU
- NCSA Delta GPU (Delta GPU)
- NCSA DeltaAI
- PSC Bridges-2 GPU (PSC Bridges-2 GPU)
- Purdue Anvil GPU
- SDSC Expanse GPU
- TACC Frontera GPU
- TACC Lonestar6-GPU
- TACC Vista (NVIDIA GH100 Grace Hopper Superchip)
- TAMU ACES

2017-2025 Research Focus

100-1000 Processors

- Non-convex Optimization
  - Large-batch Training
  - 2nd-order Optimization
  - Gradient Sparsification

Berkeley | EECS
Electrical Engineering and Computer Sciences
UCLA Samueli Computer Science
THE UNIVERSITY OF CHICAGO DEPARTMENT OF COMPUTER SCIENCE
Argonne NATIONAL LABORATORY
Computer Sciences
RUTGERS

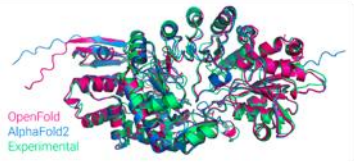
- I/O System
  - Efficient I/O for Neural Network Training with Compressed Data
  - Fair-sharing Remote-shared Burst Buffer

The Grainger College of Engineering Computer Science
NCSA

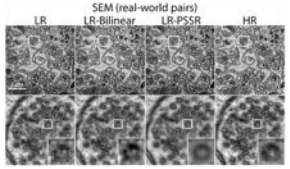
Supercomputer

- Job Scheduling
  - Tradeoff between turnaround and node hour consumption
  - Carbon emission

2017-2025 Research Focus



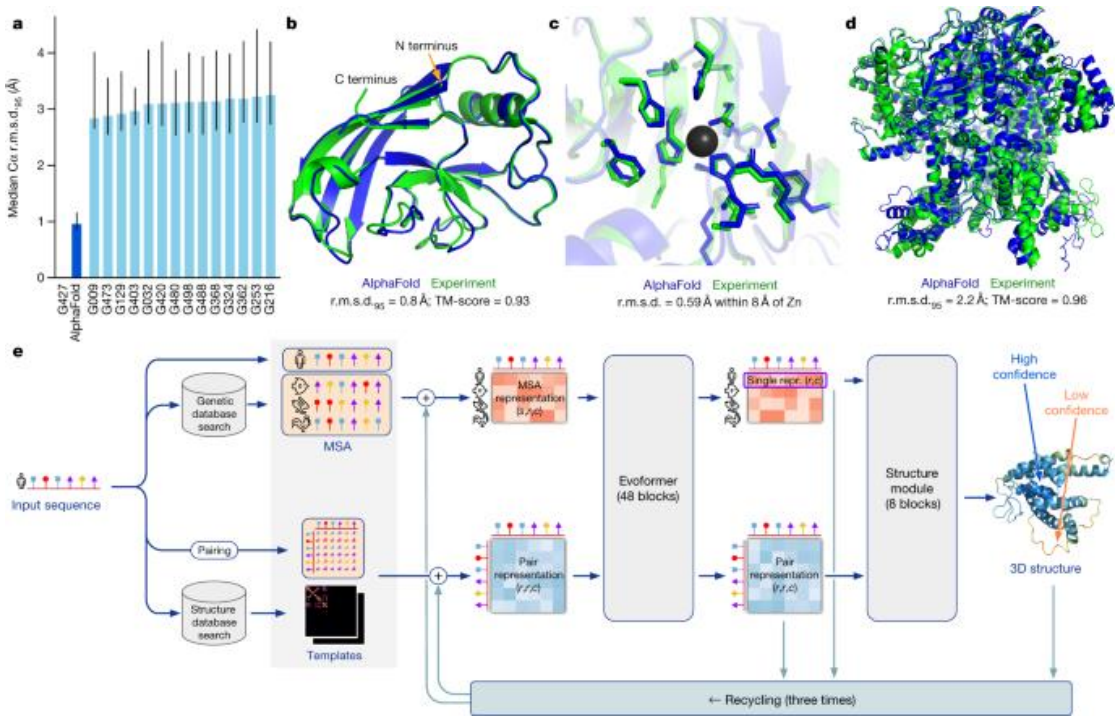
- OpenFold, an open source implementation of AlphaFold



- SRGAN, super-resolution of low-dose electromagnetic brain images

- [In progress] Animal Ecology

- [In progress] Digital Agriculture



AlphaFold2

Sequence origin	Count (approx.)	MSA	Template hits	Structure
PDB (all unique chains)	140,000	✓	✓	Experimentally determined
Uniclust30 (filtered)	270,000	✓	✓	Predicted by AlphaFold2
Uniclust30 (unfiltered)	16 million	✓	×	×

OpenProteinSet

Software Installation

Run on Many GPUs

Job Management

Data Management

Provenance Management

RUTGERS

THE UNIVERSITY OF CHICAGO DEPARTMENT OF COMPUTER SCIENCE

Computer Sciences School of Computer, Data & Information Sciences

NCSA TACC

NSF

Diamond: Train Any Model with Any Dataset on Any Machine!

OpenFold

Sequence origin	Count (approx.)	MSA	Template hits	Structure
PDB (all unique chains)	140,000	✓	✓	Experimentally determined
Uniclust30 (filtered)	270,000	✓	✓	Predicted by AlphaFold2
Uniclust30 (unfiltered)	16 million	✓	×	×

OpenProteinSet

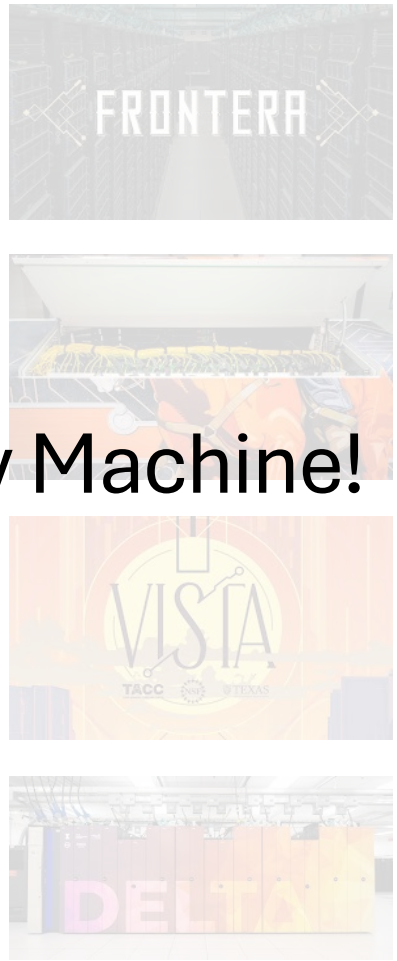
Software Installation

Hardware Variety

Job Management

Data Management

Provenance Management



RUTGERS

THE UNIVERSITY OF CHICAGO DEPARTMENT OF COMPUTER SCIENCE

Computer Sciences School of Computer, Data & Information Sciences

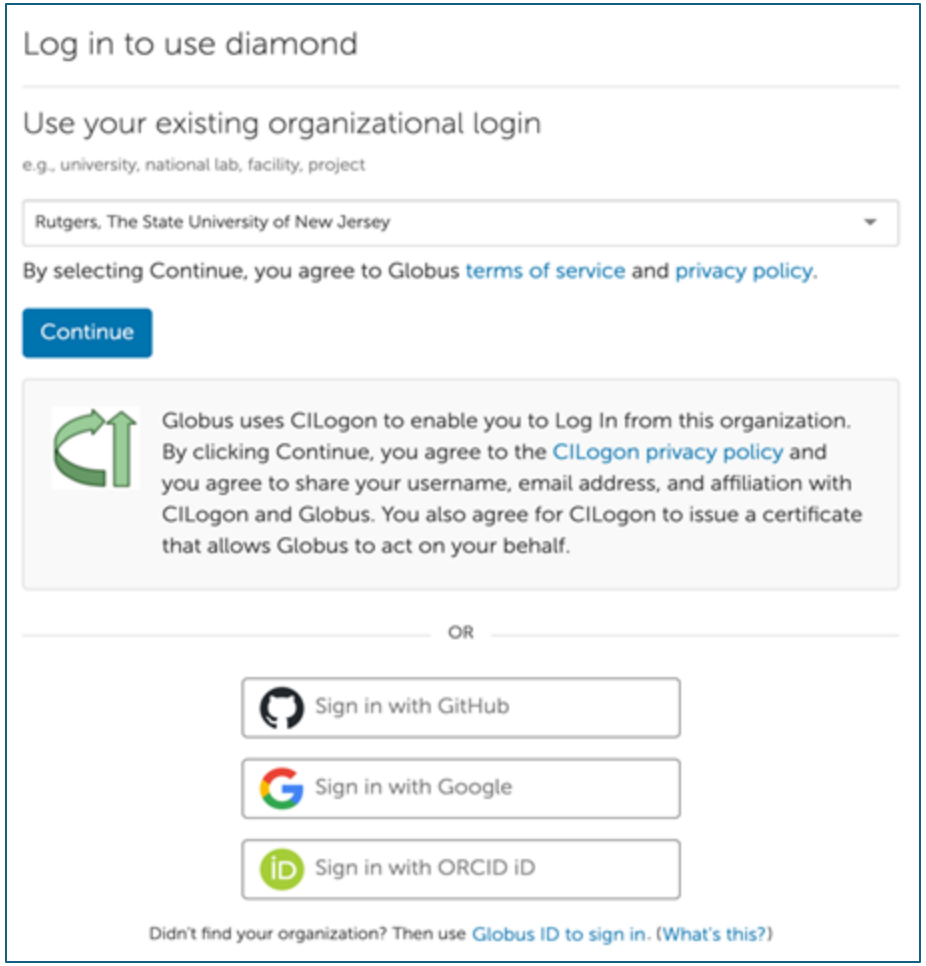
NCSA

TACC

NSF

Diamond: Democratizing large foundation model training for science

- https://diamondhpc.ai/
- Web UI training management on NAIRR GPU Resources
  - Container Image Builder
  - Job Composer
  - Task Management
- Future functionalities
  - Data management across clusters
  - Provenance tracking



Diamond: Democratizing large foundation model training for science

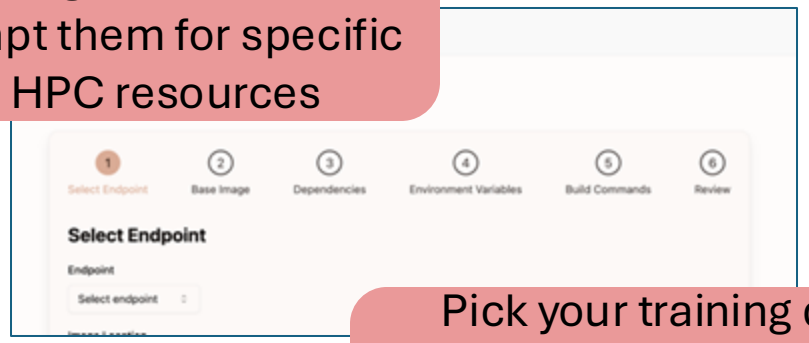
Train and fine-tune large neural networks *anywhere*

Leverage cutting-edge training libraries and state-of-the-art training practices

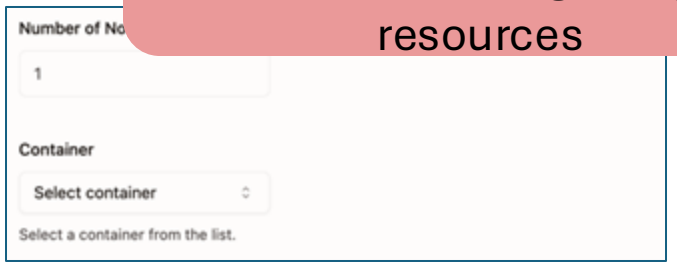
Scale training pipelines to the largest HPC resources

Manage training datasets across all your computing resources

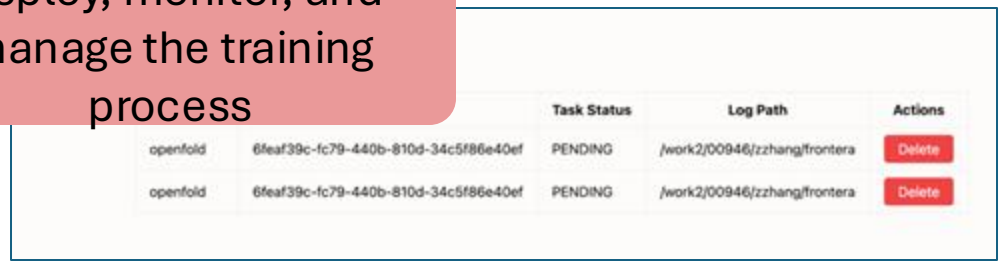
Discover or build training containers and adapt them for specific HPC resources



Pick your training dataset, configure training parameters, and select training compute resources



Deploy, monitor, and manage the training process



Diamond: Democratizing large foundation model training for science

NAIRR Pilot

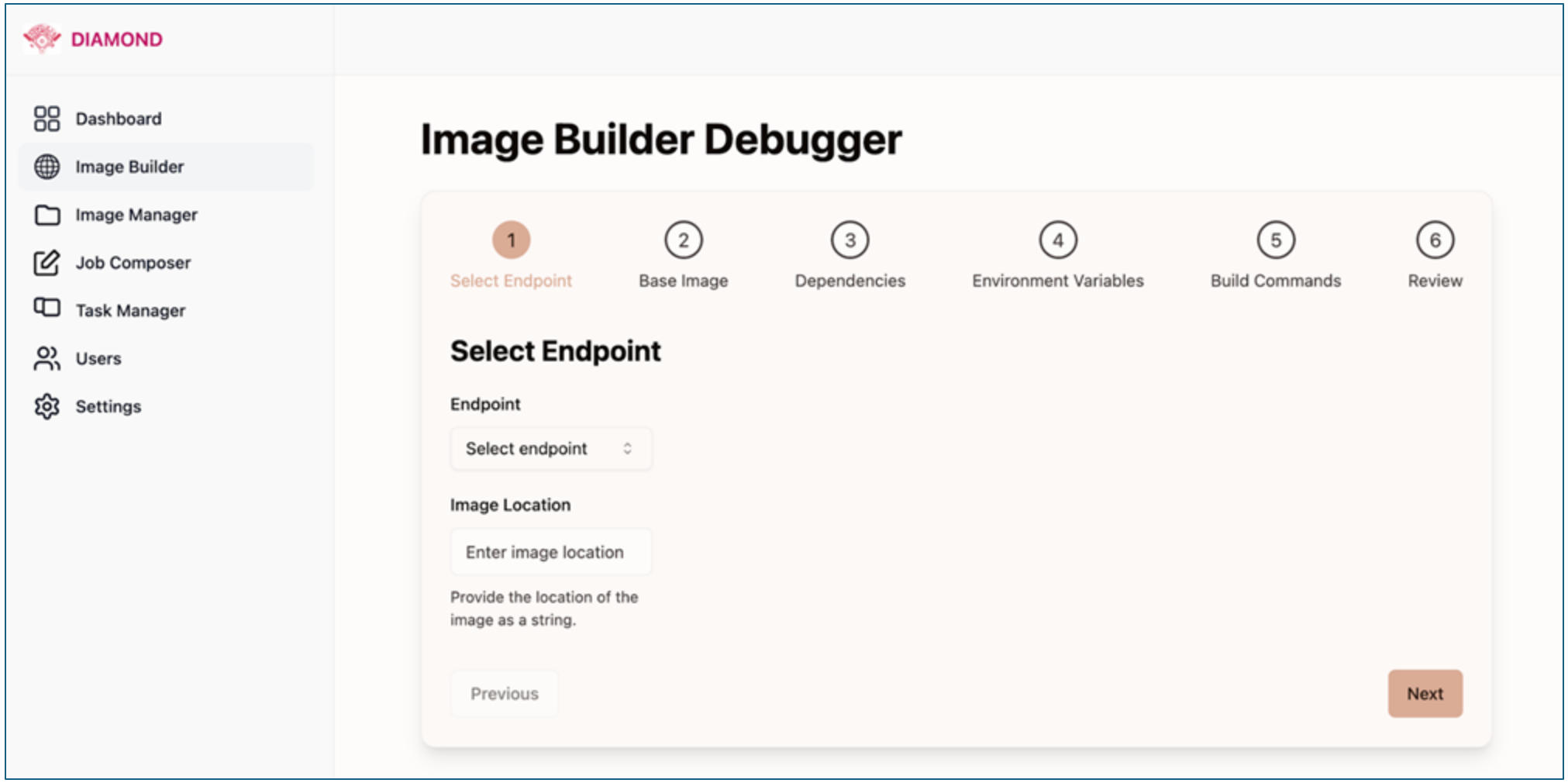
Resources

- Indiana Jetstream2 GPU
- NCSA Delta GPU
- NCSA DeltaAI
- PSC Bridges-2 GPU (PSC Bridges-2 GPU)
- Purdue Anvil GPU
- SDSC Expanse GPU
- TACC Frontera GPU
- TACC Lonestar6-GPU
- TACC Vista (NVIDIA GH100 Grace Hopper Superchip)
- TAMU ACES

Diamond: Train Any Model with Any Dataset on Any Machine

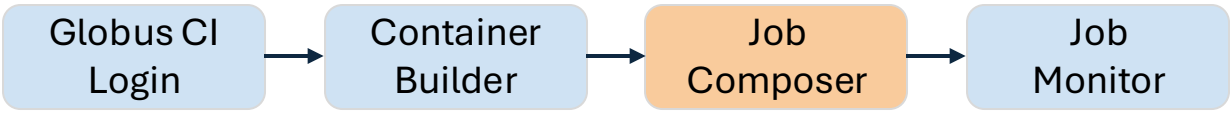
- Custom Container Image

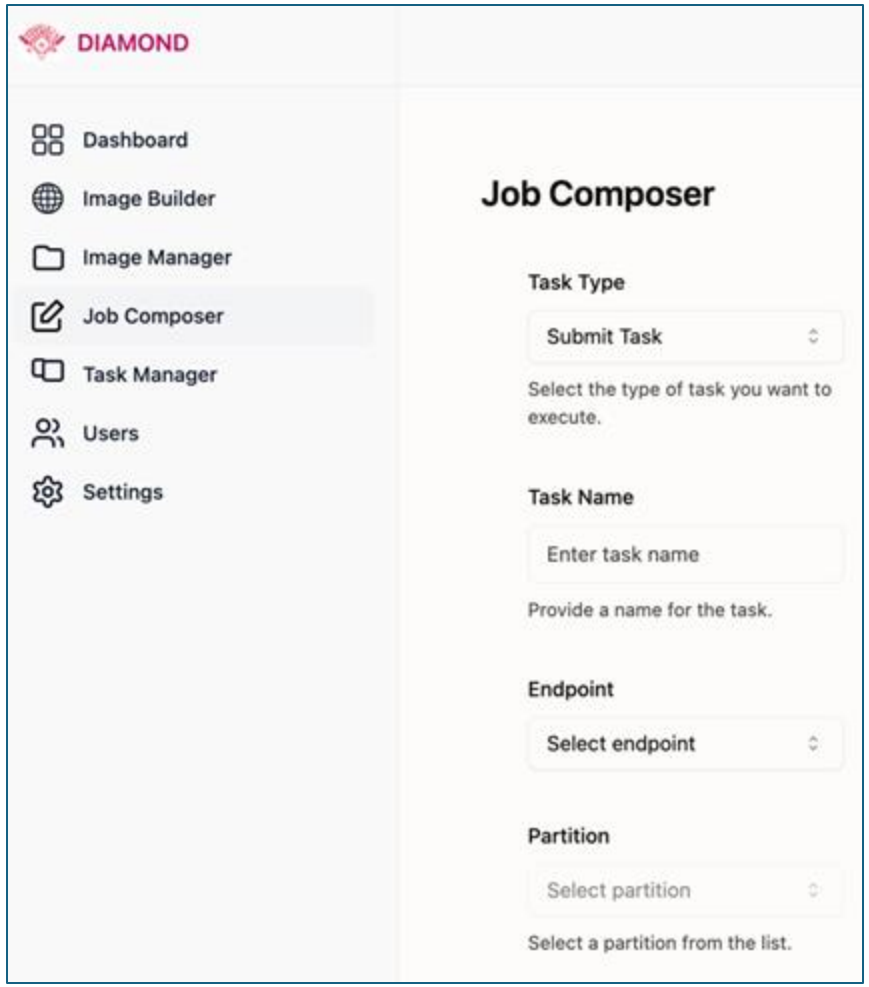
Globus CI Login → Container Builder → Job Composer → Job Monitor

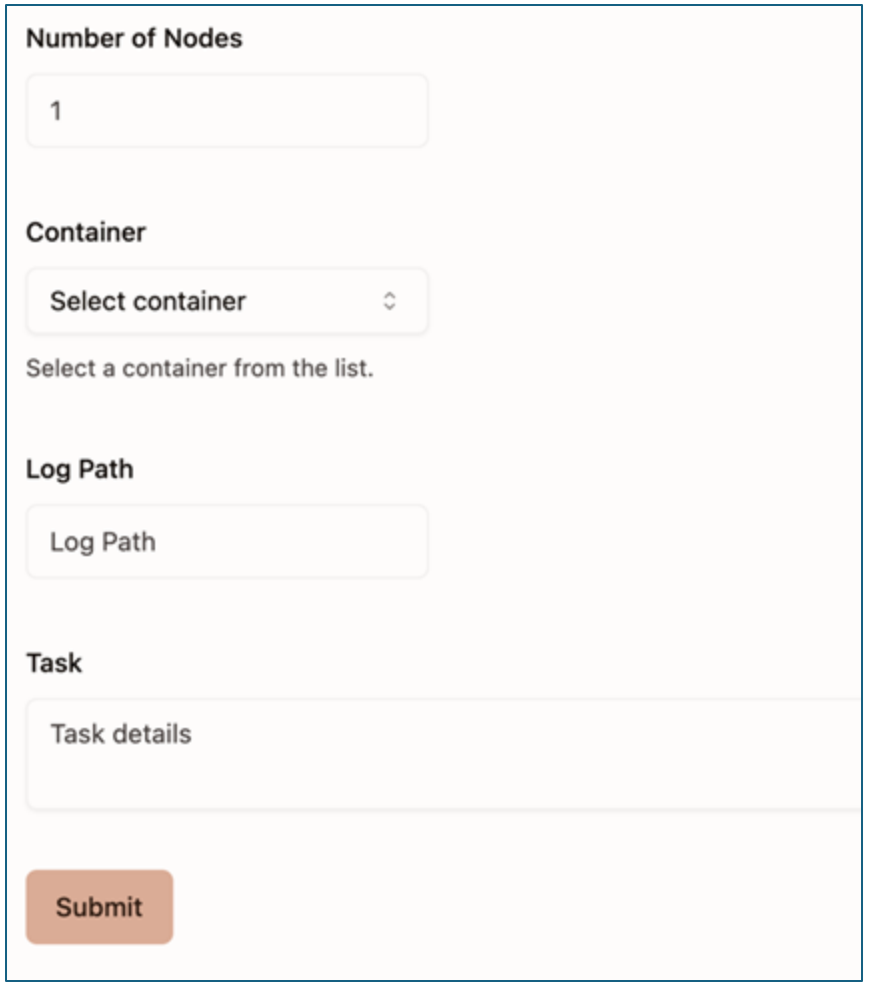


Diamond: Train Any Model with Any Dataset on Any Machine

- Compose and Run a Job

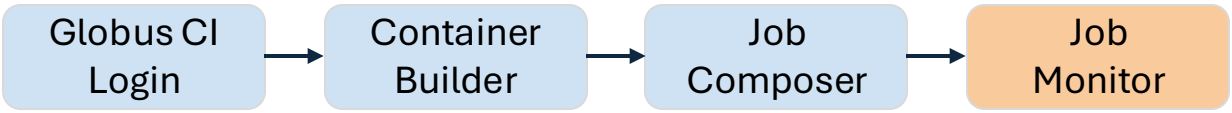


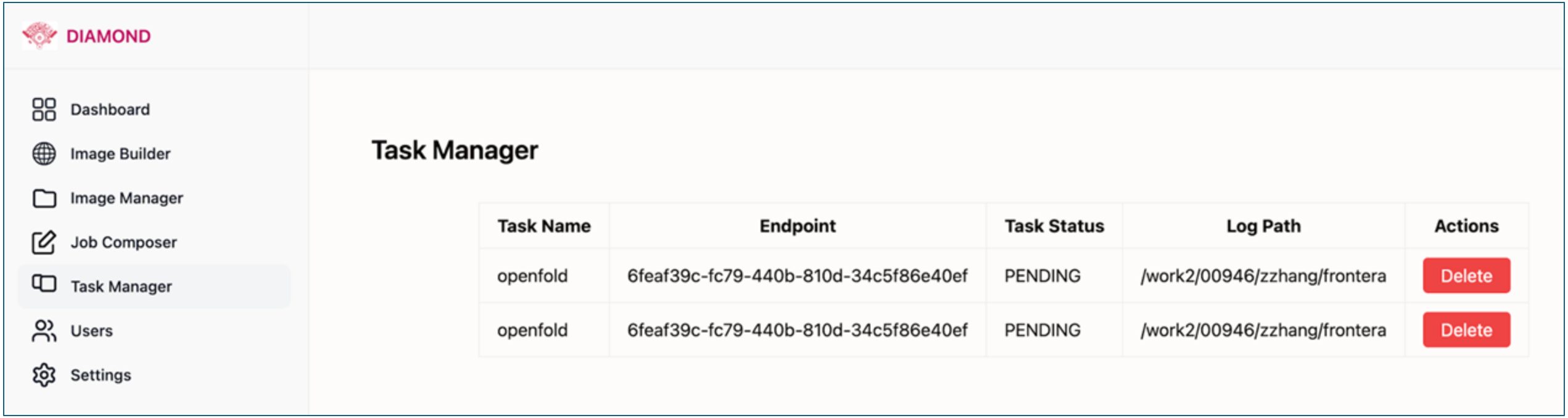




Diamond: Train Any Model with Any Dataset on Any Machine

- Monitor Jobs





Diamond: Democratizing large foundation model training for science

- https://diamondhpc.ai/
- zhao.zhang@rutgers.edu