



TPC 26

Developing best practices for
utilizing AI for scientific discovery
and engineering at scale.

Sunday–Tuesday, May 31–June 3, 2026
Hyatt Regency Baltimore Inner Harbor

WELCOME TO TPC26!

We're delighted that you'll be joining the Trillion Parameter Consortium's annual all-hands meeting in Baltimore. Since its founding in 2023, TPC has united researchers who push the boundaries of AI and science. This four-day conference convenes some 300 participants from over 100 universities, national laboratories, companies, and institutes and more than a dozen countries. Additionally, we have live satellite events and organizational live-streaming set up in Germany (at LRZ), Japan (at AIST), and South Africa (at NICIS), for a combination of live talks, live-streaming, and on-demand content.

All participants involved in TPC, whether in Baltimore or viewing remotely, are convening to pursue that mission on three fronts:

1. Building Community: Sharing Knowledge and Tools

The plenary program brings together more than two dozen international leaders to survey recent breakthroughs in AI-driven science and to frame the challenges — and opportunities — posed by today's frontier models and tomorrow's super-scale systems. Our conference keynote from Dario Gil, Under Secretary for Science at the U.S. Department of Energy, is followed by additional keynote and invited plenary talks from Katie Antypas (U.S. National Science Foundation), Rick Stevens (Argonne National Laboratory), Satoshi Matsuoka (RIKEN), Noah Smith (University of Washington / AI2), and many others.

2. Building Collaborations

Over 100 participants will give "lightning talks" in 20 current or prospective working groups, meeting across 30 parallel breakout sessions. Each of these 90-minute breakout sessions is anchored by a handful of lightning talks that highlight emerging results. Lively discussions will follow, aimed at matching complementary expertise and launching new joint projects.

3. Growing Capacity

The opening day-and-a-half of TPC26 features a hackathon and seven hands-on tutorials focused on agentic systems for science across multiple disciplines. Participants will leave with practical skills and a network of peers on which to draw to accelerate their work.

We encourage you to explore every facet of the TPC26 program — especially the breakouts — where serendipitous conversations often spark the next big idea. Thank you for bringing your expertise and energy to TPC26; we look forward to the discoveries that will follow.



Charlie Catlett
TPC26 Executive Director,
Argonne National Laboratory |
The University of Chicago



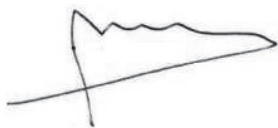
Rick Stevens
TPC Executive Committee,
Argonne National Laboratory |
The University of Chicago



Tom Tabor
TPC26 Executive Producer,
TCI Media



Satoshi Matsuoka
TPC Executive Committee,
RIKEN R-CCS



Mateo Valero
TPC Executive Committee,
Barcelona Supercomputing Center



AGENDA-AT-A-GLANCE

SUNDAY, MAY 31		
9:00	HACKATHON SESSION 1	
10:30	BREAK	
11:00	HACKATHON SESSION 2	
12:30	LUNCH	
14:00	HACKATHON SESSION 3	TUTORIAL SESSIONS 1
15:30	BREAK	
16:00	HACKATHON SESSION 4	TUTORIAL SESSIONS 2
17:30	BREAK	
18:00-19:30	HACKATHON/TUTORIAL NETWORKING GATHERING	

MONDAY, JUNE 1			
9:00	HACKATHON SESSION 5 (BEGINS 8:30)	TUTORIAL SESSIONS 3	EXHIBITION
10:30	BREAK		
11:00	HACKATHON SESSION 6	TUTORIAL SESSIONS 4	
12:30	LUNCH		
14:00	OPENING PLENARY		
16:00	BREAK		
16:30	PLENARY 2		
18:30-20:00	HAMMERSPACE WELCOME RECEPTION		

TUESDAY, JUNE 2		
8:30	PLENARY 3	
10:30	BREAK	
11:00	PLENARY 4	
12:30	LUNCH & PANEL DISCUSSION	
14:00	PARALLEL BREAKOUTS A	
15:30	BREAK	
16:00	PARALLEL BREAKOUTS B	

WEDNESDAY, JUNE 3		
8:30	PARALLEL BREAKOUTS C	
10:30	BREAK	
11:00	PARALLEL BREAKOUTS D	
12:30	LUNCH & PANEL DISCUSSION	
14:00	PARALLEL BREAKOUTS E	
15:30	BREAK	
16:00	PLENARY 5: CLOSING SESSION	

AGENDA

PRE-CONFERENCE VIRTUAL HACKATHON SESSIONS							
FRIDAY, MAY 15				FRIDAY, MAY 22			
DAY 0 INTRODUCTIONS & LOGISTICS (ONLINE ONLY) Team/mentor meeting, introductions, team matching, and cluster access				DAY 1 MENTORING & TEAM INTRODUCTIONS (ONLINE ONLY) Technical deep dive: teams work with mentors to advance their code and complete team introduction presentations			
SUNDAY, MAY 31							
9:00	HACKATHON SESSION 1 Teams work on code with mentors and continue development with mentor support. Room: Baltimore/Annapolis						
10:30	COFFEE BREAK						
11:00	HACKATHON SESSION 2 SCRUM presentations, progress updates, and continued development with mentor support. Room: Baltimore/Annapolis						
14:00	HACKATHON SESSION 3 Teams work on codes with mentors. Room: Baltimore/Annapolis	TUTORIAL SESSIONS 1 AI for Science: Agentic Systems and Scientific Discovery Closed-Loop AI and Self-Driving Laboratories Room: Frederick Enabling Reproducible AI Workflows for the NAIRR Ecosystem Overview of NAIRR Infrastructure and TapisUI Room: Chesapeake A Using AI to Accelerate Innovation & Discovery: Foundations, Models, and Coding Deep Literature Research, Novel Hypothesis Generation, and Innovative Research Planning Room: Chesapeake B Evaluation of AI Model Scientific Reasoning Skills Use Cases and Basic Evaluation Techniques Room: Columbia Building Scalable Agentic Systems for Science: Concepts, Architectures, and Hands-On with Academy Introduction to Agentic Systems and Academy Room: Pratt/Calvert					
15:30	COFFEE BREAK						
16:00	HACKATHON SESSION 4 SCRUM presentations, progress updates, and continued development. Room: Baltimore/Annapolis	TUTORIAL SESSIONS 2 AI for Science: Agentic Systems and Scientific Discovery Agentic AI for Chemistry and Materials Discovery Room: Frederick Enabling Reproducible AI Workflows for the NAIRR Ecosystem Models, Large Models, and Third-Party Registries Room: Chesapeake A Using AI to Accelerate Innovation & Discovery: Foundations, Models, and Coding AI-Supercharged Code Prototyping, Development, Execution, and Optimization Room: Chesapeake B Evaluation of AI Model Scientific Reasoning Skills Advanced Evaluation Techniques Room: Columbia Building Scalable Agentic Systems for Science: Concepts, Architectures, and Hands-On with Academy Hands-On Implementation of Agentic Systems Room: Pratt/Calvert					
17:30	COFFEE BREAK						
18:00 – 19:30	HACKATHON / TUTORIAL NETWORKING GATHERING						
MONDAY, JUNE 1							
8:30	HACKATHON SESSION 5 Teams work on code with mentors, finalize work, and present outcomes, goals, challenges, and results. Room: Baltimore/Annapolis	TUTORIAL SESSIONS 3 (BEGIN AT 9:00) AI for Science: Agentic Systems and Scientific Discovery Agentic Architectures, Frameworks, and Coordination Protocols Room: Frederick Enabling Reproducible AI Workflows for the NAIRR Ecosystem Prompt Engineering: Computer Vision Models with Jupyter Room: Chesapeake A Using AI to Accelerate Innovation & Discovery: Agentic AI and Science Applications Understanding and Using AI Agents Room: Chesapeake B Evaluation of AI Model Scientific Reasoning Skills Hands-On Work Room: Columbia End-to-End and Open-Source AI for Science: Training, Inference, and Agents Using AMD GPUs AMD AI Workflows, Ecosystem, Deployment, and Profiling Stack Overview Room: Pratt/Calvert Heterogeneous Computing and Agentic Workflows for Scientific Discovery on AWS Heterogeneous Quantum and Classical Computing on AWS Room: Camden/Lombard					
10:30	COFFEE BREAK						
11:00	HACKATHON SESSION 6 Final team presentations (5 minutes per team). Room: Baltimore/Annapolis	TUTORIAL SESSIONS 4 AI for Science: Agentic Systems and Scientific Discovery Production Systems: Industry, HPC Runtimes, and Safety Room: Frederick Enabling Reproducible AI Workflows for the NAIRR Ecosystem Fine-Tuning and Analytics Room: Chesapeake A Using AI to Accelerate Innovation & Discovery: Agentic AI and Science Applications Developing and Using AI Agents and Surrogates Room: Chesapeake B End-to-End and Open-Source AI for Science: Training, Inference, and Agents Using AMD GPUs AI4Science Studio: Agent-Driven Workflows for Scientific AI Models Room: Pratt/Calvert Heterogeneous Computing and Agentic Workflows for Scientific Discovery on AWS Agentic Workflows for AI-Driven Scientific Computing on AWS Room: Camden/Lombard					



MONDAY, JUNE 1, continued	
12:30	LUNCH
14:00	Opening Plenary Session TPC VISION Reshaping Discovery: AI, the Genesis Mission, and a New Era for Science Dario Gil, Under Secretary for Science, <i>Department of Energy</i> Accelerating Discovery and Innovation through the National AI Research Resource (NAIRR) Katie Antypas, <i>US National Science Foundation</i> Strategies for International Collaboration Among National and Regional Initiatives Moderator: Deb Goldfarb, <i>Amazon Web Services</i> Dario Gil, <i>Department of Energy</i> Katie Antypas, <i>US National Science Foundation</i> Rick Stevens, <i>Argonne National Laboratory</i> Satoshi Matsuoka, <i>RIKEN R-CCS</i> Per Öster, <i>CSC - IT Center for Science</i>
16:00	COFFEE BREAK
16:30	Plenary Session 2 INDUSTRY / LAB / ACADEMIA Agentic Science and the DOE Genesis Mission Rick Stevens, <i>Argonne National Laboratory</i> Game-Changing Aspects of AI for Science Onto HPC Satoshi Matsuoka, <i>RIKEN R-CCS</i> The Next HPC: Supercompute, Learn, Entangle – Scientific Discovery Without Boundaries Thierry Pellegrino, <i>Amazon Web Services</i> From Exascale to Genesis Mission: Building the Systems That Will Shape Scientific Discovery Thomas Zacharia, <i>AMD</i>
18:00-19:30	HAMMERSPACE WELCOME RECEPTION
TUESDAY, JUNE 2	
8:30	Plenary Session 3 FRONTIER MODELS AND SYSTEMS Olmo, Molmo, and Open Multimodal AI Infrastructure to Accelerate Science Noah A. Smith, <i>University of Washington</i> The Blue Swan AI Gigafactory: Challenges and Potential Dieter Kranzlmüller, <i>Leibniz Supercomputing Centre</i> Scaling from Parameters to Processes: Why Agents Matter for Trillion-Parameter AI Ian Foster, <i>Argonne National Laboratory</i> Building Open Japanese LLMs: From Language-Specific Foundations to Compute-Efficient Reasoning Rio Yokota, <i>Institute of Science Tokyo</i>
10:30	COFFEE BREAK
11:00	Plenary Session 4 WORKFORCE & EMERGING LEADERS Emerging Scientific Drivers for Frontier AI Valerie Taylor, <i>Argonne National Laboratory</i> Educating the Next Generation of Scientists in the Age of AI Karthik Duraisamy, <i>University of Michigan</i> Models for Discovery and Exploration in Chemical Space Anoushka Bhutani, <i>University of Michigan</i> Artificial Hivemind: The Open-Ended Homogeneity of Language Models and Beyond Liwei Jiang, <i>University of Washington</i>

TUESDAY, JUNE 2, continued					
12:30	Lunch and Panel Discussion INDUSTRY AND GOVERNMENT AI COLLABORATION: SUCCESSFUL APPROACHES AND PITFALLS TO AVOID Moderator: Earl Joseph, <i>Hyperion Research</i> Hal Finkel, <i>U.S. Department of Energy</i> Jay Boisseau, <i>Google Cloud</i> Molly Presley, <i>Hammerspace</i> Samantika Sury, <i>HPE</i>				
14:00	Challenge Applications AI for Material Sciences <i>Session 1</i>	Software Stack BOF: AI Frameworks for Multimodal Data Access and Use	Shared Infrastructure BOF: Trustworthy Privacy-Preserved Federated Learning for Science	AI Systems BOF: AI Agents as Scientific Collaborators: Building Human-Agent Research Teams	Evaluation Evaluation of LLMs for Science <i>Session 1</i>
					Workforce BOF: Building the AI-Ready Science and Engineering Workforce: Skills, Systems, and Collaboration Required for Success
15:30	COFFEE BREAK				
16:00	Challenge Applications AI for Material Sciences <i>Session 2</i>	Software Stack DWARF: Data Workflows, Agents Reasoning, and Frameworks <i>Session 1</i>	Shared Infrastructure BOF: 50-State AI Plan: A Grassroots Approach to Building a US AI Continent	Safety & Alignment BOF: Safety and Alignment in Agentic Systems	Evaluation Evaluation of LLMs for Science <i>Session 2</i>
					Software Stack AI Software Infrastructure Frameworks

WEDNESDAY, JUNE 3					
8:30	Challenge Applications GridAI Model Team	Software Stack DWARF: Data Workflows, Agents Reasoning, and Frameworks <i>Session 2</i>	Shared Infrastructure BOF: Self-Driving Labs for Accelerating Scientific Discovery at Scale	AI Systems Toward Scientific AI Platforms: Inference, Agents, and AI Services at HPC Facilities <i>Session 1</i>	Evaluation Agentic Reasoning with Scientific Foundation Models
					Workforce BOF: Who Builds the Future? Workforce Challenges in Trillion-Parameter Scientific Computing
10:30	COFFEE BREAK				
11:00	Challenge Applications BOF: Bio-Foundation Models, Agentic Systems, and Biosecurity	Software Stack BOF: Large Language Models for Scientific Software	Shared Infrastructure BOF: Bold New World of Heterogenous AI Computing <i>Session 1</i>	AI Systems Toward Scientific AI Platforms: Inference, Agents, and Services at HPC Facilities <i>Session 2</i>	Open Models Model Architectures and Performance Evaluation <i>Session 1</i>
					Software Stack BOF: Trillion Parameter Models for the Edge and Computing Continuum
12:30	Lunch and Panel Discussion SCIENTIFIC ADVANCEMENT THROUGH TPC COLLABORATION Moderator: Addison Snell, <i>Intersect360 Research</i> Arvind Ramanathan, <i>Argonne National Laboratory</i> Elliott Jacopen, <i>RIKEN R-CCS</i> Christine Sweeny, <i>Los Alamos National Laboratory</i> Rio Yokota, <i>Institute of Science Tokyo</i>				
14:00	Challenge Applications AI and Strategic Decision Support		Shared Infrastructure BOF: Bold New World of Heterogenous AI Computing <i>Session 2</i>	AI Systems BOF: Human-AI Collaboration	Open Models Model Architectures and Performance Evaluation <i>Session 2</i>
					Shared Infrastructure BOF: Energy-Efficient and Sustainable AI
15:30	COFFEE BREAK				
16:00	Plenary Session 5: Closing Session TPC COLLABORATIVE INITIATIVES AI-Based Scientific Hypothesis Generation Franck Cappello, <i>Argonne National Laboratory</i> Inference Services at the NSF Leadership Class Computing Facility Dan Stanzione, <i>Texas Advanced Computing Center</i> TPC26 Wrap-Up and Conclusions Charlie Catlett, <i>Trillion Parameter Consortium</i>				



Plenary Sessions

Opening Plenary

TPC VISION

Reshaping Discovery: AI, the Genesis Mission, and a New Era for Science

Dario Gil, *Under Secretary for Science, Department of Energy*

This talk marks a pivotal moment in the evolution of scientific discovery, as AI, advanced computing, and human expertise converge to unlock a new era of continuous, real-time innovation. It highlights the Genesis Mission as a bold national effort uniting national laboratories, industry, and academia to build an unprecedented scientific platform capable of accelerating breakthroughs at extraordinary speed. More than a technological shift, it is a call to action — an invitation for researchers and institutions to help shape a future where bold collaboration, shared purpose, and transformative innovation redefine what humanity can achieve.

Accelerating Discovery and Innovation Through the National AI Research Resource (NAIRR)

Katie Antypas, *US National Science Foundation*

The NAIRR Pilot, launched more than two years ago to connect the US-based research and education communities to critical AI resources including compute, data, software, models and expertise, is now supporting over 700 projects and enabling more than 7,000 students. The NAIRR Pilot has produced high-impact discoveries and novel AI models for science domains, and spurred start-up companies. The pilot is now transitioning to a sustainable model with a funded operations center that will be announced later this year. In this talk, NSF will discuss this novel public-private partnership and interagency collaboration model, key outcomes, lessons learned from the NAIRR pilot, and future directions.

Strategies for International Collaboration Among National and Regional Initiatives

Moderator: Deb Goldfarb, *Amazon Web Services*

Our panelists collectively steward hundreds of millions of dollars in AI and computing infrastructure, along with world-class scientific talent, vast experimental datasets, major instruments and laboratories, and leadership-class supercomputers across some of the world's leading national and regional initiatives. What could that combined capacity achieve if aimed together at shared scientific challenges? Could coordinated frontier AI systems deliver vaccines in days rather than months the next time a pandemic emerges? Could agentic models working across international infrastructures cut aircraft fuel consumption by 25% or more, or discover energy storage materials and fuels that outperform anything available today? What pre-competitive scientific grand challenges — ones that rise above regional interests — are ripe for a coordinated global push? And how do we organize now to be at-the-ready when the moment demands it?

Panelists: Dario Gil, *Department of Energy*

Katie Antypas, *US National Science Foundation*

Rick Stevens, *Argonne National Laboratory*

Satoshi Matsuoka, *RIKEN R-CCS*

Per Öster, *CSC - IT Center for Science*

Plenary 2

INDUSTRY / LAB / ACADEMIA

Agentic Science and the DOE Genesis Mission

Rick Stevens, *Argonne National Laboratory*

Agentic science reframes scientific practice around human-AI teams that co-generate hypotheses, run experiments, and analyze results. Drawing on examples from our replication project, we develop a resource model spanning input tokens, human review time, HPC cycles for training and testing models, and compute for running experiments and analyzing data. We ask: how do we practically accelerate science, and at what cost? Holding scientist headcount fixed, we explore how per-scientist AI investment shifts productivity — rethinking metrics beyond paper counts toward hypothesis depth, saturation, quality, and “deeper” science with greater per-paper impact. We frame all this within the DOE Genesis Mission.

Game-Changing Aspects of AI for Science Onto HPC

Satoshi Matsuoka, *RIKEN R-CCS*

AI is rewriting science's operating system, replacing human-driven, literature-centric research with autonomous discovery loops. But unlike commercial LLMs, AI for Science demands a vast ecosystem of specialized models and tools — ultra-resolution vision foundation models (ORBIT-2, SC'25 Best Paper), surrogate simulators, gigapixel pathology learners, and agentic co-scientists orchestrating thousands of domain tools. Serving Japan's 500,000 researchers alone would require ~700K Blackwell-equivalent GPUs; agentic workloads remain largely unquantified. No nation or hyperscaler can build this capability solo. Global open collaboration — shared compute, open models, federated data, interoperable stacks — exemplified by the DOE Genesis-RIKEN ARiSE partnership, is absolutely necessary.

The Next HPC: Supercompute, Learn, Entangle – Scientific Discovery Without Boundaries

Thierry Pellegrino, *Amazon Web Services*

Scientific discovery is being transformed by the convergence of high-performance computing, AI for science, and quantum computing, spanning seamlessly from on-premises to cloud. Compressing time to innovation is now the defining challenge, so let's explore real-world examples of how the industry is empowering the scientific community to harness this boundary-free compute model for breakthrough discovery.

From Exascale to Genesis Mission: Building the Systems That Will Shape Scientific Discovery

Thomas Zacharia, *AMD*

The next leap in AI for science will not come from models alone, but from the systems that sustain them. Building on DOE's Exascale Initiative and Project, the Genesis Mission marks a shift from standalone machines to persistent, AI-enabled discovery platforms. AI accelerates hypothesis generation, while high-fidelity simulation ensures validation and trust. As workflows become agentic, computation shifts toward tightly integrated systems spanning orchestration, data, and execution. DOE's LUX and Discovery exemplify this transition, enabling continuous discovery at scale. Leadership will be defined by the ability to build, operate, and sustain trusted scientific ecosystems.

Plenary 3

FRONTIER MODELS AND SYSTEMS

Olmo, Molmo, and Open Multimodal AI Infrastructure to Accelerate Science

Noah A. Smith, *University of Washington*

Modern AI systems are often cast as products: a model, a chatbot, an API. But when we think about using AI for science, that framing is too narrow. What scientific communities need is open AI infrastructure: data, codebases, pre- and post-training recipes, documentation, evaluations, and access to model flows across stages of development, not just a final released set of model weights. This talk will use Ai2's Olmo project portfolio as a case study in what it means to build that infrastructure in the open. Drawing on recent results from our team, including work that exposes and studies multiple stages of model construction rather than only final models, we will show that openness at the level of infrastructure is not only a scientific virtue, but a practical necessity. If researchers are going to build AI systems for their own communities, and if universities, nonprofits, and governments are going to harness AI to serve the public interest, they must be able to invest in, contribute to, and use open infrastructure. Our goal is not to reproduce commercial AI, but to create a healthier open ecosystem to accelerate scientific discovery.

The Blue Swan AI Gigafactory: Challenges and Potential

Dieter Kranzlmüller, *Leibniz Supercomputing Centre (LRZ)*

The Bavarian proposal for the "Blue Swan" European AI Gigafactory addresses the critical need for specialized computing infrastructure to train large-scale foundation models within the European research and industrial landscape. The technical framework integrates high-end GPUs into a coherent, HPC-

oriented cluster architecture, building on Leibniz Supercomputing Center's (LRZ's) pioneering work on holistic energy efficiency. This includes hot water direct liquid cooling, utilization of 100% renewable energy sources, and reuse of waste heat to achieve carbon-neutral operations. On this technological basis, Blue Swan employs a scientific approach to analyze demand and enable and scale industrial applications in dedicated domains. It also integrates national European data spaces to facilitate latency-optimized interoperability between industrial applications and academic research. Thus, Blue Swan intends to serve as a substantial AI resource and a technological validation point for energy-efficient, sovereign AI infrastructures in the exascale range.

Scaling from Parameters to Processes: Why Agents Matter for Trillion-Parameter AI

Ian Foster, *Argonne National Laboratory*

Large models are usually measured by what goes in: parameters, tokens, compute. Agents shift the focus to what happens over time: plans, tool calls, experiments, revisions, failures, recoveries, and discoveries. This talk frames agents as the machinery that turns trillion-parameter models from predictors into participants in scientific and technical work. We'll discuss why agentic systems change how we think about scale, evaluation, reliability, and control, and why the next frontier may be not just bigger models, but larger and more consequential processes built around them.

Building Open Japanese LLMs: From Language-Specific Foundations to Compute-Efficient Reasoning

Rio Yokota, *Institute of Science Tokyo and RIKEN CCS*

Japan's LLM ecosystem is rapidly moving from adaptation to original capability building. This talk will present lessons from



Scaling Intelligence
Breaking Barriers
Building Futures

Your next breakthrough starts with a conversation!
Visit the AWS Advanced Computing team at TPC26.

Learn more



the Swallow and LLM-jp projects: open collaboration, Japanese-centric data curation, multilingual evaluation, and scalable training on domestic infrastructure. Recent work will be presented that shows how pre-training data can be rewritten to improve math and code performance, and how mixture-of-experts sparsity should be optimized for reasoning through active FLOPs and tokens-per-parameter. Together, these efforts point toward transparent, reproducible, and locally-grounded LLM development with global scientific impact.

Plenary 4

WORKFORCE & EMERGING LEADERS

Emerging Scientific Drivers for Frontier AI

Valerie Taylor, *Argonne National Laboratory*

Scientific discovery is entering a new era in which frontier AI is becoming essential for exploring phenomena that are too complex, data-intensive, or time-consuming for traditional approaches alone. From accelerating the understanding of biological systems to enabling new approaches for hardware-software co-design, emerging scientific applications are driving demand for increasingly capable AI systems. This talk will highlight representative efforts from the Genesis Mission Seed Model Teams, the RIKEN Center for Computational Science, and Barcelona Supercomputing Center, illustrating how scientific researchers are shaping the needs for the next generation of AI infrastructure, computing systems, and large-scale computational workflows.

Educating the Next Generation of Scientists in the Age of AI

Karthik Duraisamy, *University of Michigan*

With access to powerful AI, the scarce resource shifts from information and knowledge to judgment and intellectual leaps. The next generation's edge will NOT simply be doing what machines cannot, but holding what must remain non-delegable for a responsible scientist. We argue for redefining "fundamentals" as transferable operations: inference, calibration, causality, verification, taste. Training in the foundational disciplines of physics, chemistry, biology and mathematics will become more important (not less) and will also serve as domains where intellectual thinking and judgment are developed. We educate scientists not only to stay useful in an AI economy, but because understanding is what science is for.

Foundation Models for Discovery and Exploration in Chemical Space

Anoushka Bhutani, *University of Michigan*

This talk presents MIST, a family of molecular foundation models with an order-of-magnitude more parameters and training data than prior works. MIST models predict more than 400 structure-property relationships and demonstrate state-of-the-art performance across diverse benchmarks spanning from physiology to electrochemistry. It will cover MIST's capacity to solve real-world problems across chemical space, from multiobjective electrolyte screening to olfactory perception mapping, along with a systematic application of mechanistic interpretability methods to uncover generalizable scientific concepts learned by the model, which reveals how models encode chemical knowledge. The talk will introduce innovations in training methodology, including hyperparameter-penalized neural scaling laws that reduce model development computational costs by an order of magnitude. Together, these methods and findings represent significant progress toward accelerating materials discovery using foundation models.

Artificial Hivemind: The Open-Ended Homogeneity of Language Models and Beyond

Liwei Jiang, *University of Washington*

Language Models (LMs) often struggle to generate diverse, human-like creative content, raising concerns about the long-term homogenization of human thought through repeated exposure to similar outputs. Yet scalable methods for evaluating LM output diversity remain limited, especially beyond narrow tasks such as random number or name generation, or beyond repeated sampling from a single model. Infinity-Chat is a large-scale dataset of 26K diverse, real-world, open-ended user queries that admit a wide range of plausible answers with no single ground truth. It is the first comprehensive taxonomy for characterizing the full spectrum of open-ended prompts posed to LMs. This talk presents a large-scale study of mode collapse in LMs using Infinity-Chat, revealing a pronounced Artificial Hivemind effect in open-ended generation of LMs. Overall, Infinity-Chat presents the first large-scale resource for systematically studying real-world open-ended queries to LMs, revealing critical insights to guide future research for mitigating long-term AI safety risks posed by the Artificial Hivemind.

TUESDAY LUNCH AND PANEL

Industry and Government AI Collaboration: Successful Approaches and Pitfalls to Avoid

Moderator: Earl Joseph, *Hyperion Research*

This panel explores practical lessons from real-world AI collaborations between industry and government, highlighting what drives successful partnerships — and what causes them to fail. Through case studies and firsthand experience, speakers will discuss key success factors, common pitfalls, and strategies for building effective cross-sector collaboration. The session will conclude with forward-looking ideas for future AI partnerships and actionable recommendations for initiating impactful public-private initiatives in an evolving technological and policy landscape. Attendees will walk away with practical, experience-based guidance for building or strengthening AI partnerships between the public and private sectors.

Panelists: Hal Finkel, *U.S. Department of Energy*
Jay Boisseau, *Google Cloud*
Molly Presley, *Hammerspace*
Samantika Sury, *HPE*

WEDNESDAY LUNCH AND PANEL

Scientific Advancement Through TPC Collaboration

Moderator: Addison Snell, *Intersect360 Research*

The real measure of success for TPC is the extent to which it enables breakthroughs in the global scientific community. One of TPC's goals is to "identify and incubate" collaborations. This panel of experts will describe their work and how it is accelerated by TPC participation. More importantly, this panel opens the door to opportunity. As we near the conclusion of TPC26, the most pertinent insights are the ways in which TPC directly benefits your work. This is where the collaboration begins.

Panelists: Arvind Ramanathan, *Argonne National Laboratory*
Elliott Jacopen, *RIKEN R-CCS*
Christine Sweeny, *Los Alamos National Laboratory*
Rio Yokota, *Institute of Science Tokyo*

Plenary 5

TPC COLLABORATIVE INITIATIVES

AI-Based Scientific Hypothesis Generation

Franck Cappello, *Argonne National Laboratory*

Scientific hypothesis generation is the most consequential step of the research workflow: every downstream action, from experiment design to simulation, observation, and analysis, inherits its quality from it. Researchers are starting to use frontier LLMs and multi-agent co-scientists as Hypothesis Generation Tools (HGTs) to accelerate scientific discovery, with the potential to exploit multi-disciplinary knowledge more effectively than any individual scientist. In practice, HGTs and humans work at different time scales: while a researcher will spend months to come up with a few hypotheses, HGTs can generate thousands of hypotheses for every single scientific question in a matter of hours or a few days. Although many of the HGT-generated hypotheses can be discarded quickly today, rejecting hypotheses will become harder as HGTs progress. This raises a novel question: what should a practical hypothesis generation process for science look like in the AI context? This talk will discuss the current state-of-the-art in hypothesis generation, available HGTs, an analysis of the generated hypotheses, some positive results, and evaluation outcomes. We will also introduce the following questions that the science community will have to answer: How many hypotheses should an HGT generate per problem? What scale of resources (tokens, GPUs) is needed to serve institutions with thousands of researchers? How can the generated hypotheses be filtered/ranked? Do we have the infrastructure to test tens, hundreds, or thousands of hypotheses per research problem?

Inference Services at the NSF Leadership Class Computing Facility

Dan Stanzione, *Texas Advanced Computing Center*

AI inferencing capability is increasingly important inside academia for both inference and training. Although options from industry are available, resources for AI development in academia are substantially smaller than their industrial counterparts. While AI training can be relatively simply supported by the batch-scheduled, GPU-enabled clusters in many HPC centers, inference workflows present new challenges. This talk will cover how inference is being addressed at the Texas Advanced Computing Center (the National Science Foundation's Leadership-Class Computing Facility), examine the unmet needs we see from the user base, and provide a summary of TPC discussions between TACC and other centers around the world on how inference may be addressed in academic research.

TPC26 Wrap-Up and Roadmap

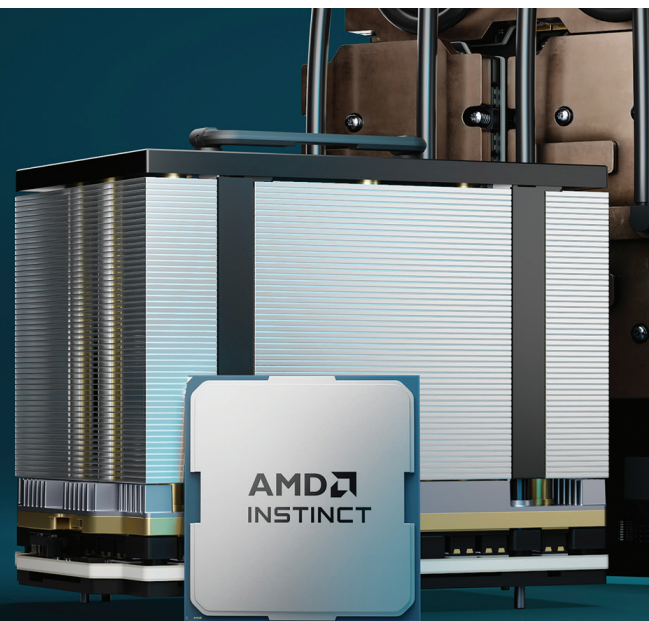
Charlie Catlett, *Trillion Parameter Consortium*

This closing session reflects on the highlights of TPC26 and looks ahead to what's next for the Trillion Parameter Consortium. We'll preview the fall hackathon and the SC26 workshop, which is now accepting 5-page papers through July 10. We'll also showcase collaborations advanced over the past year, including joint work on architecture and data for a large-scale, open frontier-class model partnership with CEPI to accelerate vaccine development, and expanding efforts in evaluation, agentic systems, alignment and safety, and MCP strategies.

PUT SUPERCOMPUTER POWER BEHIND YOUR HPC WORKLOADS

AMD EPYC™ CPUs and Instinct™ GPUs
advance some of the world's most
powerful supercomputers - making them
perfect to power your most demanding
AI and HPC workloads

[AMD.com/HPC](https://www.amd.com/HPC)



AMD
together we advance_

Hackathon

The general sessions of TPC26 on June 1–3 will be preceded by a day and a half of hackathons, May 31 and June 1. During the TPC26 Hackathon, TPC working groups will be meeting to advance collaborative efforts, building on previous hackathons that have been taking place quarterly since 2024.

TPC26 AI FOR SCIENCE HACKATHON: FROM MODELS TO AGENTS

All day Sunday and Monday morning in the Baltimore/Annapolis Room.

The TPC26 hackathon is part of the Open Hackathon program. An Open Hackathon is a multi-day, intensive hands-on event that helps computational and data scientists, researchers, and developers port, accelerate and optimize their real-world AI and HPC projects on a variety of data center architectures.

The TPC26 hackathon will pair participants with dedicated programming mentors or domain experts experienced in targeted application areas to work collaboratively. This mentoring enables participants to explore various models, libraries, and tools to realize performance gains, improve accuracy, achieve speedups, scale to modern data centers, use the latest features of the accelerated computing stack, or outline a clear development roadmap.

TPC26 Hackathon Schedule

Day 0: 2 hours virtually on May 15	Team/mentor meeting, introductions, team matching, and cluster access
Day 1: 8 hours virtually on May 22	Teams work with mentors to advance their code and complete team introduction presentations
Day 2: 8.5 hours in person at TPC26	Teams continue development with mentor support; SCRUM presentations for progress updates and roadblocks
Day 3: 4 hours in person at TPC26	Teams finalize work and present outcomes, goals, challenges, and results from the hackathon

The TPC26 hackathon will be conducted by Sepideh Khajehei, Senior Solution Architect, NVIDIA, Neil Getty, Computational Scientist, Argonne National Laboratory, Nick Jones, Director of Digital Research Innovation & Artificial Intelligence, University of Auckland, Miguel Vazquez, Principal Bioinformatics Platform Architect, Barcelona Supercomputing Center, Arvind Ramanathan, Computational Biologist, Argonne National Laboratory, and Manjot Singh, Open Hackathons Program Manager, NVIDIA.

Tutorials

These tutorials have been refined over the past 18 months, including the introductory “AI for Science” tutorial that has been presented to hundreds of people at Supercomputing Asia in February 2024, the TPC European Kickoff Workshop in June 2024, the University of Michigan’s annual Conference on Foundation Models and AI Agents for Science, and numerous local training events.

AI FOR SCIENCE: AGENTIC SYSTEMS AND SCIENTIFIC DISCOVERY

Sunday afternoon and Monday morning in the Frederick room

This is a lessons-learned tutorial designed to equip researchers with practical insights and conceptual grounding for applying AI systems to scientific challenges. Twelve talks from TPC participants across national laboratories, universities, and industry present concrete experience building AI systems that reason, plan, execute, and refine across simulation, experiment, and literature. The program covers the full arc from autonomous experimentation to production deployment, organized in four sessions:

Session One	Session Two	Session Three	Session Four
Closed-Loop AI and Self-Driving Laboratories	Agentic AI for Chemistry and Materials Discovery	Agentic Architectures, Frameworks, and Coordination Protocols	Production Systems: Industry, HPC Runtimes, and Safety

AI FOR SCIENCE: AGENTIC SYSTEMS AND SCIENTIFIC DISCOVERY, continued

Talks are exemplars showing AI use in applications ranging from accelerator operations, catalysis, battery science, drug discovery, and computational fluid dynamics, to semiconductor engineering. Each illustrates the use of AI frameworks and tools, including LangGraph, AG2, Claude Code, MCP, ChemGraph, Osprey, SpectraQuery, and Dragon, among others.

Learning Takeaways

Participants will develop a practical understanding of building and deploying AI systems for scientific discovery, including:

- Designing closed-loop, agentic workflows that integrate experiment, simulation, and reasoning
- Selecting and composing agentic frameworks and coordination protocols (LangGraph, AG2, MCP, and beyond)
- Scaling multi-agent systems across HPC platforms and distributed scientific facilities
- Operational realities of moving from prototype to production: runtimes, industrial deployment, safety, and trust

Attendees will come away with concrete patterns, deployment lessons, and emerging directions drawn directly from working AI-for-science systems across the TPC community.

This tutorial is being organized by Dan Stanzione of the Texas Advanced Computing Center, University of Texas at Austin, and will feature 12 lightning talks.

ENABLING REPRODUCIBLE AI WORKFLOWS FOR THE NAIRR ECOSYSTEM

Sunday afternoon and Monday morning in the Chesapeake A room

This tutorial aims to provide researchers with an introduction to the latest reproducible Artificial Intelligence (AI) and Machine Learning (ML) workflows and tools available through the NSF-funded Tapis v3 platform, which provides both an Application Programming Interface (API) and User Interface (UI). Using Tapis, researchers can discover AI/ML models and tools and deploy them directly to compute resources within the NAIRR and ACCESS ecosystem, supporting the goal of providing computation, data, software, models, training, and educational materials to advance research, discovery, and innovation.

Through hands-on exercises, participants will gain experience in developing AI/ML workflows and deploying them on a variety of HPC and cloud resources such as Jetstream2, Chameleon, and Vista, and Stampede3. We will emphasize the utilization of various Tapis core APIs, alongside specialized APIs such as Tapis Workflows, Tapis Pods, ML Hub and FlexServ, all seamlessly integrated within the user-friendly TapisUI. Using these production-grade services, we will demonstrate the creation and facilitation of trustworthy, reproducible scientific machine learning workflows. By the end of this tutorial, researchers will be empowered to efficiently develop, deploy, and maintain their own ML workflows.

Session One

Overview of NAIRR
Infrastructure and TapisUI

Session Two

Models, Large Models, and
Third-Party Registries

Session Three

Prompt Engineering:
Computer Vision Models
with Jupyter

Session Four

Fine-Tuning and Analytics

Learning Takeaways

Participants will learn to securely authenticate with Tapis to access its core and advanced APIs, enabling the creation, execution, and deployment of scientific machine learning AI/ML workflows. By constructing well-defined workflows for real-world use cases, attendees will gain a foundational understanding of how to leverage HPC resources for research and utilize production-quality APIs to build transparent, reproducible ML pipelines.

This tutorial will be conducted by Anagha Jamthe, Wei Zhang, and Christian Garcia of the Texas Advanced Computing Center, University of Texas at Austin.



USING AI TO ACCELERATE INNOVATION & DISCOVERY

Sunday afternoon and Monday morning in the Chesapeake B room

This hands-on tutorial will equip computational scientists, engineers, developers, and students with practical skills for using AI models, tools, and agentic systems for maximum productivity, innovation, and discovery. The tutorial will begin by providing a foundation for understanding both predictive and generative AI methods, including how to minimize errors and to increase accuracy and useful results. The tutorial will then cover the powerful capabilities of LLMs and multi-modal models, with demos and hands-on labs. The majority of the workshop will then show attendees how AI technologies can augment phases from discovery and innovation from start to finish: deep literature research, ideation/hypothesis generation, research/development planning, application prototyping and development, code optimization, surrogate creation, and data analysis. The emphasis for every phase will be how AI technologies can assist and empower (not replace), and which tools are most useful for each task (and why).

Tutorial examples and labs will leverage the latest production and research/prototype Google technologies — Gemini, NotebookLM, Gemini CLI, Code Assist, Antigravity, AlphaEvolve, Co-Scientist, and others — that are available at time of this workshop. However, the core principles and strategies are designed to be portable, enabling scientists to effectively use any comparable AI models and tools in their own endeavors (and even in most of the labs). Multiple AI science applications (WeatherNext, AlphaFold 3, AlphaGenome, etc.) developed by Google DeepMind will be used to show the capabilities of AI-powered scientific discovery.

Session One	Session Two	Session Three	Session Four
Deep Literature Research, Novel Hypothesis Generation, and Innovative Research Planning	AI-Supercharged Code Prototyping, Development, Execution, and Optimization	Understanding and Using AI Agents	Developing and Using AI Agents and Surrogates

Learning Takeaways

Participants will learn how AI-powered tools can help in every phase of the computational research/application development process:

- Prompt engineering: how to get the best results, and why some methods work better than others
- Leveraging context windows to improve accuracy and usefulness of answers
- Providing sources/content to further improve accuracy and usefulness of responses
- Using agentic systems to further improve accuracy and to achieve useful results and make decisions

This tutorial will be conducted by Jay Boisseau, Advanced Computing Strategist at Google.

EVALUATION OF AI MODEL SCIENTIFIC REASONING SKILLS

Sunday afternoon and the first half of Monday morning in the Columbia room

This is a hands-on tutorial designed to equip researchers with practical skills and conceptual grounding in the application of LLMs to scientific challenges. Large Language Models (LLMs) are becoming capable of solving complex problems while presenting the opportunity to leverage them for scientific applications. However, even the most sophisticated models can struggle with simple reasoning tasks and make mistakes.

This tutorial focuses on best practices for evaluating LLMs for science applications. It guides participants through methods and techniques for testing LLMs at basic and intermediate levels. It starts with the fundamentals of LLM design, development, application, and evaluation while focusing on scientific application. Participants will also learn various complementary methods to rigorously evaluate LLM responses in benchmarks and end-to-end scenario settings. The tutorial features a hands-on session where participants use LLMs to solve provided problems.

Session One	Session Two	Session Three
Use Cases and Basic Evaluation Techniques	Advanced Evaluation Techniques	Hands-On Work

EVALUATION OF AI MODEL SCIENTIFIC REASONING SKILLS, continued

Learning Takeaways

Participants will learn the principles and approaches for the use of LLMs as scientific assistants and how these can be evaluated with respect to scientific knowledge and reasoning skills, such as:

- Use cases of LLMs for scientific applications
- Importance of prompting and performance
- Basic of LLM evaluation
- Evaluation of LLMs for science and engineering
- Advanced evaluation techniques of LLMs for Science and Engineering
- Hands-on

This tutorial will be conducted by Franck Cappello, R&D Lead, Senior Computer Scientist, Sandeep Madireddy, Computer Scientist and AI Researcher, Neil Getty, Assistant Computer Scientist and Robert Underwood, Assistant Computer Scientist at Argonne National Laboratory.

BUILDING SCALABLE AGENTIC SYSTEMS FOR SCIENCE: CONCEPTS, ARCHITECTURES, AND HANDS-ON WITH ACADEMY

Sunday afternoon in the Pratt/Calvert room

Agentic systems, in which autonomous agents collaborate to solve complex problems, are emerging as a transformative methodology in AI. However, adapting agentic architectures to scientific cyberinfrastructure — spanning HPC systems, experimental facilities, and federated data repositories — introduces new technical challenges. In this half-day tutorial, we introduce participants to the design, deployment, and management of scalable agentic systems for scientific discovery. We will present Academy, a Python-based middleware platform built to support agentic workflows across heterogeneous research environments.

Session One

Introduction to Agentic Systems and Academy

Session Two

Hands-On Implementation of Agentic Systems

Learning Takeaways

Participants will learn core agentic system concepts, including asynchronous execution models, stateful agent orchestration, and dynamic resource management. A guided hands-on session will help attendees build and launch their own agentic systems. This tutorial is designed for researchers, developers, and cyberinfrastructure professionals interested in advancing AI-driven science with next-generation autonomous systems.

This tutorial will be conducted by Ian Foster, Data Science and Learning Division Director at Argonne National Laboratory, and Yadu Babuji, University of Chicago.

END-TO-END AND OPEN-SOURCE AI FOR SCIENCE: TRAINING, INFERENCE, AND AGENTS USING AMD GPUS

Monday morning in the Pratt/Calvert room

Based on demonstrations and access to AMD Developer Cloud, this hands-on tutorial is designed to equip scientists with the necessary tools to leverage AI in scientific workflows using AMD's open-source ROCm stack (consisting of frameworks, compilers, libraries and tools).

The program will leverage open data (e.g., Wikipedia), open models (e.g., AMD Instella, GPT-OSS, OpenFold), and recipes inspired by real use cases to demonstrate AI model training from first principles, domain-specific fine-tuning, optimized model inference with distillation, interleaving modeling/simulation codes with AI (at full and mixed precision), and orchestrating agentic frameworks on AMD GPUs.

We will also show use of AMD Primus, the flexible training framework enabling large-scale foundation model training, and Enterprise AI Suite, for model hosting and serving, both applied to scientific domains.



END-TO-END AND OPEN-SOURCE AI FOR SCIENCE, continued

Session One

AMD AI Workflows, Ecosystem, Deployment, and Profiling Stack Overview

Session Two

AI4Science Studio: Agent-Driven Workflows for Scientific AI Models

Learning Takeaways

Participants will experience a high developer-velocity guide with practical patterns and performance insights that they can apply to their own scientific workloads and infrastructure. Specifically, attendees will learn how to:

- Get started and learn best practices for porting to AMD systems using open-source tools, models, and code
- Build models, surrogates, and digital twins by training and fine-tuning AI models
- Use tools to profile, troubleshoot, and optimize multi-GPU and multi-node training and inference jobs
- Integrate “Computational Science Assistant” agents within simulation codes for application of AI in autonomous labs scenarios

This tutorial will be conducted by the AMD AI for Science Team, consisting of Ashwin Aji, Ryan Yard, Mike Schulte, and Jon Belof.

HETEROGENEOUS COMPUTING AND AGENTIC WORKFLOWS FOR SCIENTIFIC DISCOVERY ON AWS

Monday morning in the Camden/Lombard room

This hands-on tutorial equips researchers and computational scientists with practical skills for leveraging heterogeneous computing architectures and agentic AI workflows to accelerate scientific discovery on AWS. The tutorial is organized in two sessions.

The first session introduces classical HPC solutions and the quantum computing service, Amazon Braket, covering hybrid quantum-classical resources that support scientific research in academic and private industry settings. Participants will explore AWS HPC services and solutions — AWS Batch, AWS Parallel Computing Service, and AWS ParallelCluster — alongside the integration and deployment of hybrid quantum-classical workloads. The session features a recent implementation of the quantum-classical auxiliary field quantum Monte Carlo workflow and its application to modeling chemical reaction energies.

The second session demonstrates how agentic AI workflows offer a new paradigm for scientific computing: an AI agent receives a research question, reasons about the computational approach, retrieves benchmark datasets from a catalog, configures and launches simulations on cloud HPC, and analyzes the results — synthesizing findings, identifying anomalies, and recommending next steps without manual intervention. Participants will explore how agentic reasoning can accelerate the hypothesis-to-computation cycle, making production-scale scientific computing more accessible and repeatable.

Session One

Heterogeneous Quantum and Classical Computing on AWS

Session Two

Agentic Workflows for AI-Driven Scientific Computing on AWS

Learning Takeaways

Participants will develop practical skills across heterogeneous computing and agentic orchestration for science:

- AWS HPC services and solutions: AWS Batch, AWS Parallel Computing Service, and AWS ParallelCluster
- The AWS fully-managed quantum computing service: Amazon Braket
- Running hybrid quantum-classical applications with Amazon Braket
- Scaling hybrid quantum-classical applications with multi-service solutions
- Designing agentic pipelines that automate the scientific computing workflow from problem formulation through simulation execution and result verification
- Leveraging dataset catalogs and agent tool-use patterns to enable autonomous data retrieval and simulation configuration
- Deploying computational science workloads on elastic cloud HPC infrastructure through agent-driven orchestration

This tutorial will be conducted by Evan Donato, Senior Specialist Solutions Architect, AWS; Evgeny Epifanovsky, Senior Staff Technical Program Manager, IonQ; Tyler Takeshita, Senior Applied Scientist, AWS; and Lowell Wofford, Principal Technical Product Manager, AWS.

Breakout Group Sessions

TPC26 breakout groups are designed to identify, form, and pursue collaborations that will accelerate the development of new AI capabilities and services for scientific discovery. Some sessions are organized by TPC working groups, others are prospective working groups or birds-of-a-feather gatherings. Each session comprises a small set of lightning talks followed by group discussion.

The six-way parallel breakout schedule is loosely organized around eight themes: Driving Challenge Applications; AI Software Infrastructure/Frameworks; Infrastructure to Enable Shared Data & Computing; Open Frontier AI Systems; Open Suite for Evaluating Model Skills, Knowledge, Reasoning, and Safety; Open Frontier Models; Training- and Deployment-Level Safety and Alignment; and Expanding and Deepening the AI Workforce.

DRIVING CHALLENGE APPLICATIONS (CHALLENGE APPLICATIONS)

Main Plenary Ballroom

Identify challenge applications for driving and evaluating the Infrastructure to Enable Shared Data & Computing, Open Frontier Models (Model Architecture & Performance Evaluation), and Open Frontier AI Systems tracks. Not centrally picking winners and losers, but asking the community to volunteer (and drive) scientific challenge applications, aiming for diversity on multiple axes (including industry applications).

AI FOR MATERIAL SCIENCES

SESSION 1: Tuesday, June 2, 14:00

SESSION 2: Tuesday, June 2, 16:00

Organizers: Eliu Huerta, *Argonne National Laboratory*
Xiaoyun Wang, *NVIDIA*

This session will showcase how AI is revolutionizing materials discovery across quantum science, semiconductors, chemistry, energy, and advanced manufacturing. Bringing together world-class leaders from academia and industry, the session features keynote talks by Ted Sargent (Northwestern), Cameron J. Owen (Lila Sciences), Laura McGorman (Meta), and Arvin Kakekhani (PsiQuantum), alongside a special NVIDIA tutorial by Xiaoyun Wang and lightning talks from rising innovators in the field. Expect bold ideas, frontier AI methods, foundation models for science, autonomous experimentation, and next-generation computational workflows that are redefining the pace of scientific discovery.

See insert for associated lightning talks.

GRIDAI MODEL TEAM

Wednesday, June 3, 8:30

Organizers: Kibaek Kim, *Argonne National Laboratory*
Teja Kuruganti, *Oak Ridge National Laboratory*

This working group is organized around GridAI, a Genesis Mission seed project to scope a scalable AI platform for power grid modeling, analysis, and decision support. This session will introduce GridAI to the TPC community, describe the seed project's goals and team, and feature invited contributions from participating institutions on grid-relevant AI directions. An open forum will follow to discuss shared challenges in data, modeling, software, and HPC infrastructure, and to explore connections with scientific ML and energy systems modeling. Researchers in AI/ML, scalable algorithms, optimization, and complex systems are invited to join and help shape the GridAI agenda.

See insert for associated lightning talks.



DRIVING CHALLENGE APPLICATIONS, continued

BOF: BIO-FOUNDATION MODELS, AGENTIC SYSTEMS, AND BIOSECURITY

Wednesday, June 3, 11:00

Organizers: Arvind Ramanathan, *Argonne National Laboratory*
Newton Wahome, *CEPI*
Sarah Carter, *CEPI*

Bio-foundation models and agentic AI systems are reshaping biological discovery — from genome-scale language models and protein structure prediction to autonomous laboratory workflows — while simultaneously surfacing critical biosecurity risks that demand urgent community attention. This BOF convenes leading researchers from industry and national laboratories to examine scalable biological AI architectures, agentic orchestration for autonomous science, and governance frameworks for dual-use risks. Topics span model scaling behavior, biosafety-by-design, and policy-aligned deployment. Attendees will gain actionable insight into responsible bio-AI development, preparing the HPC and AI research community for safe, high-impact biological discovery at scale.

See insert for associated lightning talks.

AI AND STRATEGIC DECISION SUPPORT

Wednesday, June 3, 14:000

Organizers: Frank Alexander, *Argonne National Laboratory*
Manish Parashar, *The University of Utah*

Building on the consortium's collaborative development of foundation models for science and engineering, we'll examine applications in disaster response, supply chain resilience, pandemic management, and other areas. The discussion will connect TPC's work on scalable architectures, scientific data curation, and exascale optimization to breakthrough capabilities in time-critical decision support.

See insert for associated lightning talks.

AI SOFTWARE INFRASTRUCTURE/FRAWORKS (SOFTWARE STACK)

Baltimore and Columbia

Develop software infrastructure and middleware to support the training, deployment, and integration of complex frontier-scale AI models and systems. Provide the technical backbone for the Open Frontier Models (Model Architecture & Performance Evaluation) and Open Frontier AI Systems tracks, while enabling integration with experimentation platforms, laboratories, instruments, and other real-world scientific environments.

BOF: AI FRAMEWORKS FOR MULTIMODAL DATA ACCESS AND USE

Tuesday, June 2, 14:00

Baltimore

Organizers: Ilkay Altintas, *San Diego Supercomputer Center*
Manish Parashar, *The University of Utah*

AI is accelerating discovery, redefining the workforce, and transforming society, yet fragmented data and computing limit innovation. Federated, AI-ready frameworks can bridge distributed data repositories and computing resources through interoperable, production-grade services that enable seamless access, integration, and composability of data, models, and workflows across edge, cloud, and HPC. These frameworks support multimodal analysis, simulation, and end-to-end workflows. This BOF will examine architectures, highlight existing efforts, and explore paths toward a cohesive national ecosystem. Talks and discussions will identify needs, share use cases, and provide practical entry points for leveraging national cyberinfrastructure.

See insert for associated lightning talks.

AI SOFTWARE INFRASTRUCTURE/Frameworks, continued

DWARF: DATA WORKFLOWS, AGENTS REASONING, AND FRAMEWORKS

SESSION 1: Tuesday, June 2, 16:00

SESSION 2: Wednesday, June 3, 8:30

Baltimore

Organizers: Robert Underwood, *Argonne National Laboratory*
Neeraj Kumar, *Pacific Northwest National Laboratory*
Ian Foster, *Argonne National Laboratory*

This multi-session track explores emerging systems and strategies for building intelligent, scalable platforms to accelerate scientific discovery. Talks and discussions will cover the design of agent-based architectures, integration of scientific workflows with large language models, scalable data pipelines, and novel reasoning frameworks. The session encourages both the software infrastructure and the users of that infrastructure. Participants will engage in dialogue on the future of scientific AI infrastructure and the coordination required to realize a distributed, agent-enabled discovery ecosystem.

See insert for associated lightning talks.

AI SOFTWARE INFRASTRUCTURE FRAMEWORKS

Tuesday, June 2, 16:00

Columbia

Organizers: Mohamed Wahib, *RIKEN R-CCS*
Rio Yokota, *Institute of Science Tokyo*

This working group session will discuss the software infrastructure and middleware required to support frontier-scale AI for science, including frameworks for training, deployment, orchestration, and integration with HPC systems, scientific workflows, laboratories, instruments, and experimentation platforms. The session aims to identify common software challenges and collaboration opportunities needed to make large-scale AI systems usable, reliable, and interoperable across scientific environments.

BOF: LARGE LANGUAGE MODELS FOR SCIENTIFIC SOFTWARE

Wednesday, June 3, 11:00

Baltimore

Organizers: Mohammad Alaul Haque Monil, *Oak Ridge National Laboratory*
Keita Teranishi, *Oak Ridge National Laboratory*

Agentic AI is redefining HPC research by introducing intelligent, autonomous capabilities. This BOF explores how large language model (LLM) agents enable code translation, modernization, modeling, and tuning, allowing legacy scientific applications to be efficiently adapted for modern HPC architectures. Beyond code transformation, agentic systems can orchestrate and optimize complex, end-to-end HPC workflows with minimal human intervention. Participants will discuss emerging tools, challenges, and opportunities in deploying LLM-driven agents for scalable, reproducible, and adaptive research pipelines. The session aims to foster collaboration and share insights on advancing autonomous HPC systems powered by agentic AI technologies.

See insert for associated lightning talks.

BOF: TRILLION PARAMETER MODELS FOR THE EDGE AND COMPUTING CONTINUUM

Wednesday, June 3, 11:00

Columbia

Organizers: DK Panda, *The Ohio State University*
Barney Maccabe, *University of Arizona*

Computing continuum at the edge is emerging as a common environment for many applications — transportation, fire prevention, agriculture, medicine, manufacturing, etc. Typical computing environments at the edge (IoT devices, drones, etc.) do not have enough computing or storage capacity, which presents the challenge of how to use TPC Models in this environment. This BOF will focus on the latest state-of-the-art solutions along this direction as well as future opportunities and challenges.

See insert for associated lightning talks.



INFRASTRUCTURE TO ENABLE SHARED DATA & COMPUTING (SHARED INFRASTRUCTURE)

Annapolis and Columbia

Collectively build scientific training data resources and shared computing infrastructure for model training and further fine-tuning for general-purpose and domain-specific settings. Establish scalable and sustainable capabilities that serve as the foundation for the rest of the tracks.

BOF: TRUSTWORTHY PRIVACY-PRESERVED FEDERATED LEARNING FOR SCIENCE

Tuesday, June 2, 14:00

Annapolis

Organizers: Olivera Kolevska, Oak Ridge National Laboratory
Kibaek Kim, Argonne National Laboratory
Ravi Madduri, Argonne National Laboratory

Federated learning offers a promising approach for enabling collaborative scientific discovery while preserving the privacy of sensitive data across institutions. This BOF will bring together researchers and practitioners to discuss trustworthy, privacy-preserving federated learning frameworks tailored for scientific workloads. The discussion will focus on challenges such as secure model aggregation, data confidentiality, system scalability, and integration with distributed research infrastructures. Objectives include identifying common requirements, sharing emerging techniques, and fostering collaborations within the TPC community. The session is particularly relevant to TPC participants interested in distributed computing, secure data sharing, and scalable AI methods that support cross-institutional scientific research.

See insert for associated lightning talks.

BOF: 50-STATE AI PLAN: A GRASSROOTS APPROACH TO BUILDING A US AI CONTINENT

Tuesday, June 2, 16:00

Annapolis

Organizers: Barr von Oehsen, Pittsburgh Supercomputing Center
Jack Wells, NVIDIA

This BOF explores the emerging fabric of state and regional initiatives that will empower American leadership in AI and quantum computing. Using Pennsylvania, Tennessee, Utah, Massachusetts, New York, New Jersey, and California as case studies, we examine how state-led initiatives are aligning regional assets with America's AI Action Plan, the Genesis Mission, and the National AI Research Resource (NAIRR) in building a US AI Continent. Crucially, we discuss how these and other federal initiatives can leverage state-level "factories" to advance research, workforce development, and economic growth by embracing grassroots innovation tailored to local and regional strengths. We highlight essential partnerships between academia, state governments, industry, philanthropy, and federal agencies to drive these efforts. Participants will discuss strategies for democratizing access to high-performance computing resources, streamlining technical deployment, and building the AI workforce pipeline. By fostering inclusive, state-led hubs, we can attract investment, train thousands of workers, and ensure US technological sovereignty in AI.

See insert for associated lightning talks.

BOF: SELF-DRIVING LABS FOR ACCELERATING SCIENTIFIC DISCOVERY AT SCALE

Wednesday, June 3, 8:30

Annapolis

Organizers: Arvind Ramanathan, Argonne National Laboratory
Rio Yokota, Institute of Science Tokyo

Self-driving laboratories (SDLs) are transforming scientific discovery by integrating AI-driven hypothesis generation, robotic experimentation, and closed-loop active learning into fully autonomous workflows. This workshop convenes leading researchers from national laboratories, academia, and industry to examine the computational foundations of SDLs — spanning foundation models for experimental design, agentic orchestration across heterogeneous instruments, high-throughput data pipelines, and HPC integration at scale. Applications span drug discovery, materials design, enzyme engineering, and critical minerals extraction. Attendees will gain actionable insight into deploying autonomous discovery platforms, benchmarking SDL performance, and building the open infrastructure needed to accelerate science at exascale.

INFRASTRUCTURE TO ENABLE SHARED DATA & COMPUTING, continued

BOF: BOLD NEW WORLD OF HETEROGENOUS AI COMPUTING

SESSION 1: Wednesday, June 3, 11:00

SESSION 2: Wednesday, June 3, 14:00

Annapolis

Organizers: Satyam Srivastava, *d-Matrix*
Tom St. John, *Gimlet Labs*

As AI workloads diversity, no single accelerator wins on every axis of cost, power, and performance. Consequently, traditional monolithic architecture is hitting critical bottlenecks in scaling and efficiency. This BoF brings together perspectives from researchers, architects, and hardware vendors on building production systems that combine GPUs, ASICs, and more. Speakers will share insights spanning architecture and integration, software stacks and portability, workload scheduling across heterogeneous fabrics, and real-world performance benchmarks. Attendees will gain a strong grasp of the trade-offs, tooling gaps, and emerging best practices for designing AI infrastructure that treats heterogeneity as a foundational feature.

See insert for associated lightning talks.

BOF: ENERGY-EFFICIENT AND SUSTAINABLE AI

Wednesday, June 3, 14:00

Columbia

Organizers: Siddhartha Jana, *Intel*
Natalie Bates, *Lawrence Berkeley National Laboratory*
Shaohui Liu, *Massachusetts Institute of Technology*

This BOF will focus on the sustainability challenges of large-scale AI, aligned with the mission of advancing responsible and scalable AI for science. As trillion-parameter models demand unprecedented compute, energy, and data resources, critical questions arise around carbon footprint, infrastructure efficiency, and equitable access. The forum will convene researchers, industry practitioners, and infrastructure providers to examine trade-offs between performance and sustainability, share best practices in energy-efficient model design and deployment, and identify collaborative pathways for greener AI. By fostering cross-sector dialogue, this session aims to shape actionable strategies for sustainable AI at extreme scale.

OPEN FRONTIER AI SYSTEMS (AI SYSTEMS)

Frederick

Develop frontier AI systems for science that incorporate reasoning models (start with SOA closed models, eventually include Open frontier model) and develop domain foundation models, knowledge graphs, agentic systems and orchestrations, simulators, and experiments.

BOF: AI AGENTS AS SCIENTIFIC COLLABORATORS: BUILDING HUMAN-AGENT RESEARCH TEAMS

Tuesday, June 2, 14:00

Organizers: Charlie Catlett, *Argonne National Laboratory*
Rick Stevens, *Argonne National Laboratory*

Scientific AI agents are moving from tools to participants — attending conferences, contributing to research, and coordinating across institutions. This BOF features lightning talks delivered by AI agents alongside their human collaborators, showcasing live experiments in agentic research workflows. We then open the floor to explore a proposed international collaboration: a multi-institutional human-agent team working together to accelerate scientific discovery, reduce duplicated effort, and raise the quality of science.

See insert for associated lightning talks.



OPEN FRONTIER AI SYSTEMS, continued

TOWARD SCIENTIFIC AI PLATFORMS: INFERENCE, AGENTS, AND AI SERVICES AT HPC FACILITIES

SESSION 1: Wednesday, June 3, 8:30

SESSION 2: Wednesday, June 3, 11:00

Organizers: Venkatram Vishwanath, *Argonne National Laboratory*
Ilkay Altintas, *San Diego Supercomputer Center*

Building on last year's session on inference-for-science services, this session expands its focus to the broader landscape of scientific AI platforms at HPC facilities. As foundation models, domain-specific AI systems, and agentic workflows gain traction, HPC centers are actively developing infrastructure for scalable inference, AI agents, AI-ready data services, model gateways, and facility-scale AI services. This session will convene members from international HPC centers, application teams, vendors, and the open source community to share emerging best practices for deploying reliable, reproducible, and secure AI services for science. Discussion topics will span simulation and workflow integration, heterogeneous architectures, orchestration, agent skills, observability, sustainability, and workforce development. The session will gather use cases, identify shared technical and operational gaps, and define next steps for continued collaboration across the TPC community.

See insert for associated lightning talks.

BOF: HUMAN-AI COLLABORATION

WEDNESDAY, June 3, 14:00

Organizers: Anurag Acharya, *Pacific Northwest National Laboratory*
Patrick Emami, *National Renewable Energy Laboratory*

This BOF centers on frameworks for Human-AI co-intelligence in scientific discovery, positioning collaboration — not autonomy — as the primary design goal. Building on our concept of Human-AI Virtual Laboratories, we will explore how mixed-initiative interaction, role differentiation, and coordinated workflows could enable scientists and AI systems to function as true teammates. The discussion will hopefully surface key design principles spanning agency, communication, and coordination, alongside open challenges in building such systems in practice. We will also briefly discuss evaluation as a supporting concern, focusing on how to assess collaborative effectiveness in these new paradigms.

See insert for associated lightning talks.

OPEN SUITE FOR EVALUATING MODEL SKILLS, KNOWLEDGE, REASONING, & SAFETY (EVALUATION)

Chesapeake A

Develop an open suite of tools, methods, and benchmarks for evaluating the scientific skills, knowledge, reasoning, agentic capabilities, and safety/security of frontier models and AI systems.

EVALUATION OF LLMS FOR SCIENCE

SESSION 1: Tuesday, June 2, 14:00

SESSION 2: Tuesday, June 2, 16:00

Organizers: Franck Cappello, *Argonne National Laboratory*
Sandeep Madireddy, *Argonne National Laboratory*

One of the main thrusts behind the rapid evolution of LLMs is the availability of benchmarks that assess the skills and trustworthiness of LLMs. Not only do they enable a rigorous evaluation of LLMs skills and trustworthiness from accepted metrics, but they also generate competition between LLM developers. While several frameworks/benchmarks have emerged as de facto standards for the evaluation of general-purpose LLMs (Eleuther AI Harness and HELM for skills, DecodingTrust for trustworthiness), only very few of them specifically are related to science. In this segment, we will discuss the challenges of developing methods to evaluate the skills, trustworthiness, and safety of large Foundation Models for science. This track will include multiple sessions focused on different facets of model evaluation.

See insert for associated lightning talks.

OPEN SUITE FOR EVALUATING MODEL SKILLS, continued

AGENTIC REASONING WITH SCIENTIFIC FOUNDATION MODELS

Wednesday, June 3, 8:30

Organizers: Ayan Biswas, *Los Alamos National Laboratory*
Christine Sweeney, *Los Alamos National Laboratory*

Scientific AI is rapidly shifting from predictive models toward agents that can plan, use tools, run simulations, analyze data, and automate parts of the research workflow. This session asks what “reasoning” means when scientific AI systems act, not just answer. Scientific agents must satisfy constraints that general-purpose LLM agents often do not: physical consistency, numerical validity, uncertainty quantification, provenance, reproducibility, security, and appropriate human oversight. We will discuss how agentic automation changes evaluation for SciML and AI for science. Which tasks should agents automate? How should they decide when to invoke solvers, query data, generate hypotheses, or ask for human input? How do we distinguish scientific reasoning from brittle tool use, brute-force search, or persuasive but invalid explanations? The session will touch upon topics such as community needs for benchmarks, testbeds, safety practices, provenance standards, and evaluation frameworks for reliable, interpretable, and useful scientific agents.

See insert for associated lightning talks.

OPEN FRONTIER MODELS (OPEN MODELS)

Chesapeake A

Build frontier-scale, open AI models using shared data and computing infrastructure (from the Infrastructure to Enable Shared Data and Computing track), harnessing distributed resources across TPC partner institutions. Ensure that all core components are openly available to enable transparency, reuse, and scientific progress.

MODEL ARCHITECTURES AND PERFORMANCE EVALUATION

SESSION 1: Wednesday, June 3, 11:00

SESSION 2: Wednesday, June 3, 14:00

Organizers: Rio Yokota, *Institute of Science Tokyo*
Murali Emani, *Argonne National Laboratory*

Architectures for AI models are evolving rapidly, with frequent innovations in transformer variants, to reduce the cost of attention and kv-cache in long-context/ agentic reasoning, their framework support, parallelism strategies, and system-level optimizations. Identifying the optimal architecture and framework for training foundation models on scientific data is vital to unlocking the next generation of AI for science. Equally crucial is efficient inference, which enables the practical use of pre-trained models in downstream scientific applications. This multi-session track will bring together researchers and practitioners to discuss cutting-edge strategies for large-scale training, inference, and agentic scaling, alongside robust workflows to integrate them.

See insert for associated lightning talks.



Google Cloud Advanced Computing Community

Advance your skills.
Connect with experts.



Scan to join.



TRAINING- AND DEPLOYMENT-LEVEL SAFETY AND ALIGNMENT (SAFETY & ALIGNMENT)

Frederick

Develop methods to embed safety and alignment into the training and deployment of frontier-scale models and AI systems. Focus on system-level mechanisms that maintain alignment with scientific objectives and constraints, and with broader societal values, at extreme scale and in high-impact settings.

BOF: SAFETY AND ALIGNMENT IN AGENTIC SYSTEMS

Tuesday, June 2, 16:00

Organizer: Charlie Catlett, *Argonne National Laboratory*

Agentic systems are being built or proposed for a diverse range of applications, from literature review to laboratory automation. Each domain brings common as well as unique safety requirements and alignment/containment strategies. This BOF seeks to find common approaches that might be applied across different application domains, ideally identifying architectural and design approaches that are broadly applicable to developing operational agentic systems for science and engineering.

See insert for associated lightning talks.

EXPANDING AND DEEPENING THE AI WORKFORCE (WORKFORCE)

Columbia

Identify and report on progress in developing the workforce required to achieve the rest of the tracks, with particular attention to emerging and evolving roles across the frontier AI stack. Examine needs across all career stages and share recent experiences and lessons learned to inform sustainable talent development.

BOF: BUILDING THE AI-READY SCIENCE AND ENGINEERING WORKFORCE: SKILLS, SYSTEMS, AND COLLABORATION REQUIRED FOR SUCCESS

Tuesday, June 2, 14:00

Organizer: Doug Norton, *The HPC-AI Society*

As AI transforms scientific discovery and engineering design, the greatest challenge is not just computational scale, it's human scale. This BOF features a panel that explores how research institutions, industry, and consortia can cultivate and incentivize the next generation of scientists and engineers to harness AI-driven workflows. We will explore the evolving skill sets required for hybrid computing environments that integrate HPC, AI, and emerging quantum systems; strategies for interdisciplinary training that bridges data science, physics, and engineering; and frameworks for workforce development that ensure equitable access to these transformative tools. Panelists will share insights from active programs and collaborations, including CMU/PSC ByteBoost and the Argonne/University of Alaska Wildfire Analysis project, building the talent pipeline for AI-enabled science at trillion-parameter scale.

See insert for associated lightning talks.

BOF: WHO BUILDS THE FUTURE? WORKFORCE CHALLENGES IN TRILLION-PARAMETER SCIENTIFIC COMPUTING

Wednesday, June 3, 8:30

Organizer: Lois Curfman, *Argonne National Laboratory*

The computational science and research software engineering (RSE) communities are at a turning point as trillion-parameter-scale AI systems reshape code generation, simulation, and scientific workflows. Yet software ecosystems, team structures, expected roles, and collaboration practices were designed for a pre-AI era. Addressing this shift requires more than technical integration; it demands rethinking how we organize, incentivize, and sustain research software at scale. While progress has focused on models and infrastructure, critical workforce issues — including cross-disciplinary collaboration, evolving human and AI contributions, team dynamics, and incentive structures — remain underexplored. This BOF invites community input on these emerging challenges and opportunities.

See insert for associated lightning talks.

Plenary Speakers and Panel Moderators

Katie Antypas

Director, Office of Advanced Cyberinfrastructure,
US National Science Foundation

Katie Antypas is the Senior Advisor for Cyberinfrastructure at the US National Science Foundation overseeing the strategic directions for the Office of Advanced Cyberinfrastructure, which supports the development and deployment of advanced computing, data infrastructure, tools, services, and training for the research community. It is also the Office responsible for the launch and deployment of the National AI Research Resource (NAIRR) Pilot. Prior, Katie spent 17 years at Berkeley Lab at the National Energy Research Scientific Computing Center. She has an M.S. in CS from the University of Chicago and a Bachelors degree in Physics from Wellesley College.



Anoushka Bhutan

PhD Student, Mechanical Engineering
and Scientific Computing,
University of Michigan

Anoushka Bhutani is a PhD candidate in Mechanical Engineering and Scientific Computing at the University of Michigan, working at the intersection of chemistry and machine learning. She works on developing molecular foundation models to accelerate materials discovery with an interest in developing model interpretability methods for scientific applications and advancing methods for compute-efficient training of scientific foundation models.



Jay Boisseau

Advanced Computing Strategist,
Google Cloud

Jay Boisseau, an advanced computing strategist at Google, leads the Google Cloud Advanced Computing Community and initiatives to make Google Cloud a premier platform for scientific and technical computing. With 30 years of experience in academia and industry, Jay previously propelled the Texas Advanced Computing Center to global prominence and led strategic HPC projects at Dell Technologies. Jay also directs the Austin Forum on Technology & Society, founded the Austin AI Alliance, and owns Remedy bar in Austin. He holds a doctorate in astronomy from UT Austin, which launched his career in supercomputing and advanced technologies.



Franck Capello

R&D Lead, Senior Computer Scientist,
Argonne National Laboratory

Franck Cappello is an R&D lead and Senior Computer Scientist at Argonne National Laboratory. He leads a research team exploring resilience for HPC and large-scale distributed systems, lossy compression of scientific data, and LLMs for science. Franck is an IEEE Fellow, the recipient of the 2024 IEEE CS Charles Babbage Award, the 2024 Europar Achievement Award, the 2022 HPDC Achievement Award, two R&D100 awards (2019 and 2021), the 2018 IEEE TCPP Outstanding Service Award, and the 2021 IEEE Transactions of Computer Award for Editorial Service and Excellence.



Charlie Catlett

Executive Director, Trillion Parameter Consortium |
Senior Computer Scientist, Argonne National
Laboratory | The University of Chicago

Charlie Catlett is a Senior Computer Scientist at the U.S. Department of Energy's Argonne National Laboratory, and a Visiting Scientist at the University of Chicago's Mansueto Institute for Urban Innovation. His research focuses on building cyberinfrastructure to embed edge-AI in urban, environmental, and emergency sensing and response settings. He was founding chair of Grid Forum / Global Grid Forum from 1999-2005 and director of NSF's TeraGrid initiative from 2004-2007. Charlie was part of the team that established the National Center for Supercomputing Applications (NCSA) in 1985, leading efforts there including the deployment and operation of the NSFNET backbone network, an early component of the Internet, and serving as Chief Technology Officer prior to joining Argonne and UChicago in 2000. He was one of GovTech magazine's "25 Doers, Dreamers & Drivers" of 2016 and in 2019 received the Argonne Board of Governors Distinguished Performer award. Charlie is a Computer Engineering graduate of the University of Illinois at Urbana-Champaign.



Karthik Duraisamy

Professor of Aerospace Engineering, and Director,
Michigan Institute for Computational Discovery and
Engineering, University of Michigan

Karthik Duraisamy is a Professor of Aerospace Engineering at the University of Michigan, where he also directs the Michigan Institute for Computational Discovery and Engineering (MICDE). He holds a Ph.D. in Aerospace Engineering and a Masters in Applied Mathematics from the University of Maryland (U-M). Karthik's research interests span a broad spectrum of computational science and AI, including data-driven and reduced order modeling, statistical inference, numerical methods, and Generative AI for science. Karthik is the PI of the U-M/Los Alamos Center on Advanced Computational Sciences. He is also the founder and chief scientist of the Silicon Valley-based startup Geminus.AI, which is focused on physics-informed AI to accelerate autonomous industrial operations.



Hal Finkel

Director, Computational Science Research and Part-
nerships Division, Advanced Scientific
Computing Research, Office of Science,
U.S. Department of Energy

Hal Finkel serves as the Acting Associate Director for the Advanced Scientific Computing Research (ASCR) program within the Department of Energy's Office of Science, and the Office of Science lead for Genesis Mission, which is the U.S. Government's new national initiative to accelerate and enhance R&D using AI and advanced computing. Prior to joining ASCR, Hal spent nearly a decade at Argonne National Laboratory as a physicist and computational scientist at Argonne's Leadership Computing Facility (ALCF), where he built and led a team focused on advancing the computer science of compiler technology and programming languages. The team contributed significantly to the efforts of the Exascale Computing Project (ECP), in addition to exploring artificial intelligence, quantum computing, spatial architectures, and other forward-looking technologies. Hal holds a Ph.D. in Physics from Yale University, where he focused on early-universe cosmology.



Ian Foster

Data Science and Learning Division Director,
Argonne National Laboratory

Dr. Ian Foster is Senior Scientist and Distinguished Fellow, and also director of the Data Science and Learning Division, at Argonne National Laboratory, as well as the Arthur Holly Compton Distinguished Service Professor of Computer Science at the University of Chicago. Ian received a BSc degree from the University of Canterbury, New Zealand, and a Ph.D. from Imperial College, United Kingdom, both in computer science. His research deals with distributed, parallel, and data-intensive computing technologies, and innovative applications of those technologies to scientific problems in such domains as materials science, climate change, and biomedicine. Ian is a fellow of AAAS, ACM, BCS, and IEEE, and an Office of Science Distinguished Scientists Fellow.



Dario Gil

Under Secretary for Science,
Department of Energy

Dr. Dario Gil is Under Secretary for Science at the U.S. Department of Energy. His office is the nation's largest federal sponsor of basic research in the physical sciences, supporting all 17 National Laboratories of the United States, and responsible for programs including advanced computing, fusion, nuclear and high energy particle physics, basic energy sciences, and biological and environmental research. He is the department's principal advisor on science and technology. Prior to his current position, Dario was IBM Senior Vice President and Director of Research, where he was responsible for one of the world's largest and most influential corporate research labs. He was elected to the National Academy of Engineering "for his contributions to artificial intelligence and quantum computing" and is a globally recognized leader of the quantum industry. Under his leadership, IBM was the first company in the world to build programmable quantum computers and make them universally available through the cloud. Dario is an inventor and an institutional innovator, the force behind the creation of the International Science Reserve, the AI Alliance, the MIT-IBM Watson AI Lab, and the COVID-19 High Performance Computing Consortium. Dr. Gil has served on the President's Council of Science and Technology Advisors (PCAST) and on the National Science Board (NSB), where he was the first member from industry to be elected chairman in 30 years. He has served on numerous boards, including the Center for Strategic and International Studies (CSIS), the New York Academy of Sciences, the Semiconductor Industry Association (SIA), the New York Hall of Science, and Rensselaer Polytechnic Institute (RPI). Dario is the recipient of two honorary doctorates and received his Ph.D. in Electrical Engineering and Computer Science from MIT.



Debra Goldfarb

Director, Products and Strategy, Federal
Supercomputing and AI, Amazon Web Services

Debra Goldfarb recently transitioned from her role as Director of Advanced Computing, Simulation, and Quantum Products and Strategy at AWS to focus on Federal Supercomputing and AI, leading the company's efforts to help transform critical U.S. government and industrial base missions that position America to lead in the next generation of computational discovery. Prior to AWS, Debra was an Intel Fellow and served as Chief Analyst for Intel's Data Platform Group Strategy Organization. In that role, she spearheaded Intel's expansion of HPC into new markets and was instrumental in defining strategy and pathfinding for the Technical Computing Group. Before Intel, Debra led strategy and evangelism for Microsoft's Technical Computing organization,



driving high-profile collaborations including the Gates Foundation's Malaria initiative. Debra launched her career as an industry analyst at IDC, where she founded the HPC practice and the HPC User Forum, established the Life Sciences practice, and built the Bio-IT business portfolio. Her leadership in HPC has helped shape U.S. technology policy and the global HPC ecosystem.

Elliott Jacopin

Postdoctoral Researcher, Center for Biosystems
Dynamics Research, RIKEN R-CSS

Elliott Jacopin is a Postdoctoral Researcher at RIKEN's Center for Biosystems Dynamics Research in Kobe, Japan, in the Laboratory for Biologically Inspired Computing led by Takahashi Koichi. Trained as an engineer at AgroParisTech, his background spans multi-drug resistance modeling at CNRS, multi-agent systems applied to image analysis, and whole-cell simulation infrastructure at RIKEN. Elliott's recent work sits at the intersection of scientific software accessibility and agentic AI. He launched Cracking Shells, an open-source ecosystem for scientific MCP servers, and followed it with research on using Hierarchical Task Networks (HTN) planning as a deterministic coordination layer for multi-agent tool orchestration, bringing formal planning discipline to what is otherwise ad-hoc LLM sequencing.



Liwei Jiang

PhD Student, Paul G. Allen School of Computer
Science & Engineering, University of Washington

Liwei Jiang is a final-year Ph.D. candidate in the Paul G. Allen School of Computer Science & Engineering



MANAGED HPC FOR ENGINEERS

Speed up your simulations by 10x and let TotalCAE handle the IT, so your engineers can focus on engineering.

10X
MORE
THROUGHPUT

80%
SIMULATION
TIME SAVED

1hr
SUPPORT
RESPONSE

- ✓ Fully managed HPC cluster, on-premises and in the cloud
- ✓ Simple web portal to submit HPC jobs to cloud and clusters
- ✓ Automatic simulation data management and license analytics
- ✓ Remote visualization for post-processing from anywhere

**CUT SIMULATION TIME FROM DAYS
TO HOURS**

Request a Consultation with TotalCAE

888-268-3930 | info@totalcae.com | www.totalcae.com

at the University of Washington, advised by Professor Yejin Choi. She was previously a graduate student researcher at NVIDIA and the Allen Institute for Artificial Intelligence (AI2). Liwei's research focuses on humanistic, pluralistic, and coevolutionary AI safety and alignment, where she spearheads research on moral and pluralistic value reasoning in language models and develops data-, algorithm-, and system-level solutions to socio-technical challenges in AI safety, security, and large language model alignment. Her work has received Best Paper Awards at NeurIPS 2025, NAACL 2022, and CHI 2024, as well as Outstanding Paper Awards at EMNLP 2023 and the AIA Workshop at COLM 2025, and has been featured in *The New York Times*, *Nature Outlook*, *IEEE Spectrum*, *Wired*, and other major media. Liwei co-organizes workshops, including MP2 (NeurIPS 2023) and SoLaR (NeurIPS 2024; COLM 2025), and co-leads the Guardrails and Security for LLMs tutorial at ACL 2025.

Earl Joseph

CEO, *Hyperion Research*

Earl Joseph, Chief Executive Officer of Hyperion Research, drives research and consulting efforts associated with the United States, Europe, and Asia-Pacific markets for technical computing. He advises Hyperion Research clients on the competitive, managerial, technological, integration, and implementation issues for HPC and AI, and heads up Hyperion Research's high-end HPC user forum activities. Earl's areas of expertise include technical computers (from entry-level servers to high-end capability supercomputers), software, AI technologies, storage, and networking solutions. He has worked for four technical computing companies in multiple marketing and R&D roles. Earl holds a Ph.D. from the University of Minnesota, where his research focus was the strategic management of high technology firms, and an undergraduate degree in business and technology from the University of Minnesota.



Dieter Kranzlmüller

Chairman of the Board, *Leibniz Supercomputing Centre (LRZ)* | Full professor, *Ludwig-Maximilians-University Munich (LMU)*

Prof. Dr. Dieter Kranzlmüller is a full professor of computer science at Ludwig-Maximilians-Universität München (LMU) and Chairman of the Board of Directors of the Leibniz Supercomputing Centre (LRZ) of the Bavarian Academy of Sciences and Humanities. He also serves on the board of the German national Gauss Centre for Supercomputing (GCS). Dieter applies his broad research interests and background in academia as well as industry to emphasize tight integration between IT service operations and progressive R&D. His support for the advancement of innovation and the greater public-private ecosystem has been prevalent throughout his professional career, and manifests in his contributions to, amongst others, the Center for Digital Technology and Management (CDTM), the Münchner Kreis, and in his role as founding member of IT:U Linz or the Scientific Advisory Board at Heidelberg Institute for Theoretical Studies.



Satoshi Matsuoka

Director, *RIKEN R-CCS*

Professor Satoshi Matsuoka has been the director of RIKEN Center for Computational Science (R-CCS), the Tier-1 national HPC center for Japan, since April 2018, developing and hosting Japan's flagship Fugaku supercomputer. Fugaku was the fastest supercomputer in the world in 2020 and 2021, supporting cutting-edge HPC research, including investigating Post-Moore era computing and especially the future FugakuNEXT supercomputer. Satoshi led the TSUBAME series of



supercomputers that received much international acclaim at the Tokyo Institute of Technology, where he holds a professor position pursuing research in HPC, scalable big data, and AI. His longtime contribution was commended with the Medal of Honor with Purple ribbon by his Majesty Emperor Naruhito of Japan in 2022. Satoshi is a Fellow in ACM, ISC, IPSJ, and the JSSST and has won numerous awards, including ACM Gordon Bell Prizes, the IEEE-CS Sidney Fernbach Award, and the IEEE-CS Computer Society Seymour Cray Computer Engineering Award.

Per Öster

Director of Corporate Development, *CSC – IT Center for Science Management Board*

Per Öster is Director of Corporate Development and member of the CSC – IT Center for Science Management Board. He has been with CSC since 2007, leading business activities in the areas of HPC, research infrastructures, and business development. His current focus is on international collaboration and growth opportunities. Over the years, Per has represented CSC and led its participation in several international initiatives and organizations such as European Grid Infrastructure (EGI) and the European Open Science Cloud (EOSC). Presently he is Chair of the European Collaborative Data Infrastructure Council and Chair of the Board of Directors of its legal entity, EUDAT Ltd; board member of ELIXIR, the European bioinformatics research infrastructure; member of the EuroHPC Research & Innovation Advisory Group (RIAG); and board member of NORDUnet, the Nordic NREN. He is also leading CSC's participation in TPC and the associated EuroTPC project. Per has more than 30 years of experience in computational science, spanning both academia and industry. He holds a doctorate in physics from the Department of Physics at the University of Gothenburg /Chalmers University of Technology.



Thierry Pellegrino

Global Head of Advanced Computing, *Amazon Web Services*

Thierry Pellegrino is the Global head of Advanced Computing at Amazon Web Services (AWS), a role in which he oversees HPC, domain-specific ML, IOT, and Quantum for the company. In his last industry role, Thierry was CEO of Penguin Computing, and prior to that he spent 23 years with Dell, where he was the Head of the HPC and AI business. Thierry has held multiple leadership roles over his career, ranging from engineering to strategy, M&A, and business leadership, and has had the privilege to sit on the board of GRC's and Penn State's ICDS.



Molly Presley

SVP Global Marketing, *Hammerspace*

Molly Presley is the SVP of Global Marketing for Hammerspace and the host of the very popular Data Unchained podcast. She brings a wealth of experience from leading product and marketing organizations, user communities, and customer advisory boards for global technology innovators including DDN, Qumulo, and Quantum. Molly is the founder of the Active Archive Alliance, co-author of three books focused on putting data to use in research, analytics, and AI environments, and was a previous board member of the Storage Networking Industry Association (SNIA).



Arvind Ramanathan

Computational Biologist, *Argonne National Laboratory* | Senior Scientist, *University of Chicago Consortium for Advanced Science and Engineering (CASE)*



Arvind Ramanathan is a computational biologist in the Data Science and Learning Division at Argonne National Laboratory and a senior scientist at the University of Chicago Consortium for Advanced Science and Engineering (CASE). His research interests are at the intersection of data science, HPC, and biological/biomedical sciences. Arvind's research spans three areas focusing on scalable statistical inference techniques: (1) for analysis and development of adaptive multi-scale molecular simulations for studying complex biological phenomena (such as how intrinsically disordered proteins self assemble, or how small molecules modulate disordered protein ensembles), (2) to integrate complex data for public health dynamics, and (3) for guiding design of CRISPR-Cas9 probes to modify microbial function(s). Arvind has published over 30 papers, and his work has been highlighted in the popular media, including NPR and NBC News. He obtained his Ph.D. in computational biology from Carnegie Mellon University, and was the team lead for the integrative systems biology team within the Computational Science, Engineering and Division at Oak Ridge National Laboratory.

Noah A. Smith

Vice Provost for AI, Charles and Lisa Simonyi Endowed Chair for Artificial Intelligence and Emerging Technologies, and Professor, Paul G. Allen School of Computer Science & Engineering, University of Washington



Noah A. Smith is the inaugural Vice Provost for Artificial Intelligence and Charles and Lisa Simonyi Endowed Chair for Artificial Intelligence and Emerging Technologies at the University of Washington, where he is also a Professor in the Paul G. Allen School of Computer Science & Engineering. He is Senior Director of NLP Research at the Allen Institute for Artificial Intelligence, directs the OLMo open language modeling effort, and leads the NSF- and NVIDIA-supported project "Open Multi-modal AI Infrastructure to Accelerate Science." Noah's research spans language and music technologies, multimodal AI, and multifaceted evaluation of AI systems.

Addison Snell

Co-Founder & Chief Executive Officer, Intersect360 Research



Addison Snell is a veteran of the HPC industry and the co-founder and CEO of Intersect360 Research, now in its 19th year delivering forecasts and insights for high-performance markets. Intersect360 Research is a premier source of market information, analysis, and consulting for HPC and hyperscale industries worldwide. Addison is a frequent keynote speaker and panel moderator at industry events, has testified before the U.S.-China Economic & Security Review Commission Congressional Subcommittee, and was named one of 2010's "People to Watch" by *HPCwire*. Prior to Intersect360 Research, Addison was an HPC industry analyst for IDC. He originally gained industry recognition as a marketing leader and spokesperson for SGI's supercomputing products and strategy. Addison holds a master's degree from the Kellogg School of Management at Northwestern University and a bachelor's degree in Mathematics from the University of Pennsylvania.

Dan Stanzione

Executive Director, Texas Advanced Computing Center (TACC) | Associate Vice President for Research, UT-Austin



Dr. Dan Stanzione is a nationally recognized leader in the field of high performance computing who has made an impact in the open science community for more than 30 years. He is Executive Director of the Texas Advanced Computing

Center (TACC), one of the leading advanced academic computing centers in the world, and holds the position of Associate Vice President for Research at The University of Texas at Austin. Dan serves as the principal investigator (PI) for the U.S. National Science Foundation (NSF) Leadership Class Computing Facility (LCCF), which will deploy Horizon, the largest academic supercomputer dedicated to open science research in the NSF portfolio. Dan is also the PI for several other NSF-funded supercomputers, including Frontera, the fastest supercomputer at a U.S. university, and Vista, an AI-centric system — all dedicated to open science research. Dan received his bachelor's degree in electrical engineering and his master's degree and doctorate in computer engineering from Clemson University.

Rick Stevens

Associate Laboratory Director – CELS and Argonne Distinguished Fellow, Argonne National Laboratory | Professor of Computer Science, The University of Chicago



Rick Stevens is a Professor of Computer Science at The University of Chicago and the Associate Laboratory Director of the Computing, Environment, and Life Sciences (CELS) Directorate and Argonne Distinguished Fellow at Argonne National Laboratory. His research spans the computational and computer sciences from high-performance computing architecture to the development of advanced tools and methods. Recently, Rick has focused on developing AI methods for a variety of scientific and biomedical problems, and has had significant responsibility in delivering on the U.S. national initiative for Exascale computing and developing the DOE's Frontiers in Artificial Intelligence for Science, Security, and Technology (FASST) initiative.

Samantika Sury

HPE Fellow, Vice President, and Chief Hardware Architect for HPC and AI Infrastructure Solutions, HPE



Samantika Sury serves as an HPE Fellow, Vice President, and Chief Hardware Architect for HPC and AI Infrastructure Solutions. She leads the Future Technologies team, which focuses on advancing hardware and software system innovations. Her expertise lies in architecting, building, and optimizing systems for HPC and AI. Prior to joining HPE, Samantika held prominent roles at Samsung, where she served as Vice President and Chief Hardware Architect for HPC. At Samsung, she spearheaded the development of novel HPC and AI system technologies, including proof-of-concept testbeds and architecture design. Samantika has also previously worked at Intel as a Senior Principal Engineer, driving silicon and system architecture innovations into marketable products. At Intel, she was the Principal Investigator and Lead Architect for the DOE PathForward Program and served as the Lead Architect of HPC-Custom Silicon. In addition to her extensive professional accomplishments, Samantika holds 32 U.S. and international patents, has published 20 peer-reviewed papers, and has delivered numerous invited talks. She is a recipient of ACM and IEEE best paper awards and has actively contributed to several program committees, organizing bodies, and fellowship initiatives. Samantika earned her Ph.D. in Computer Science from the Georgia Institute of Technology.

Christine Sweeney

AI and Scientific Computing Team Scientist, Los Alamos National Laboratory



Christine Sweeney is a computer scientist in the Computing and AI Division at Los Alamos National Laboratory. She currently works on the AI for Mission (ArtIMis) project, producing scientific foundation models for predicting

time evolution of partial differential equations. She also researches AI workflows for the NNSA Advanced Simulation and Computing (ASC) program. Christine has extensive experience in HPC and participated in two Exascale Computing Project (ECP) teams (ExaLearn and ExaFEL) and was part of several ECP-related hackathons that brought together team members from multiple institutions. Christine has a Masters of Science in Computer Science from the University of Washington.

Valerie Taylor

Director, Mathematics and Computer Science
Division, Argonne National Laboratory

Valerie Taylor is the director of the Mathematics and Computer Science Division at Argonne National Laboratory. She received her Ph.D. in electrical engineering and computer science from the University of California, Berkeley, in 1991. She then joined the faculty in the Electrical Engineering and Computer Science Department at Northwestern University, where she was a member of the faculty for 11 years. In 2003, Valerie joined Texas A&M, where she served as head of the computer science and engineering department and senior associate dean of academic affairs in the College of Engineering, as well as a Regents Professor and the Royce E. Wisenbaker Professor in the Department of Computer Science. Valerie's research is in the area of performance analysis and modeling of parallel scientific applications. Currently, she is focused on the areas of performance analysis, power analysis, and resiliency. Valerie Taylor is a fellow of IEEE and ACM.



Rio Yokota

Professor, Institute of Science Tokyo | Team Principal,
RIKEN Center for Computational Science

Rio Yokota is a Professor at the Supercomputing Research Center, Institute of Science Tokyo. He also leads the AI for Science Foundation Model Research Team at RIKEN CCS. His research interests lie at the intersection of HPC and ML. Rio has been optimizing algorithms on GPUs since 2007, and was part of a team that received the Gordon Bell prize



in 2009 using the first GPU supercomputer. He has been leading distributed training efforts on Japanese supercomputers such as ABCI, TSUBAME, and Fugaku.

Dr. Thomas Zacharia

SVP, Strategic Technical Partnership and Public Policy,
AMD

Dr. Thomas Zacharia leads AMD's global expansion of public and private partnerships, engaging with governments, NGOs, and other organizations to accelerate the deployment of AMD-powered AI solutions. These initiatives support a rapidly growing number of global projects and applications focused on leveraging AI for the public good. A distinguished industry leader, Thomas brings extensive experience in scientific research, technology development, public policy, and strategic partnerships. Over his 35-year tenure at Oak Ridge National Laboratory (ORNL), he served as Director and oversaw the creation and implementation of the AMD-powered Frontier supercomputer — the first to break the Exascale barrier. Frontier, powered by AMD EPYC processors and AMD Instinct accelerators, has been recognized as the world's fastest supercomputer by the TOP500 list since May 2022. Under Thomas' leadership, ORNL established itself as the premier high-performance computing center globally, with four successive systems ranked No. 1 on the TOP500 list. From 2012 to 2015, Thomas served as Executive Vice President of the Qatar Foundation for Education, Science, and Community Development, where he set the vision for Qatar's investment of 2.8 percent of its GDP in research and development. He also chaired the Qatar Science and Technology Park, fostering industry and university partnerships to advance initiatives in public health, computing, and energy. Thomas is Chair of the Board of Directors for the Nashville Branch of the Federal Reserve Bank of Atlanta. He holds a B.S. in Mechanical Engineering from the National Institute of Technology Karnataka, Surathkal, India; an M.S. in Materials Science from the University of Mississippi; and a Ph.D. in Engineering Science from Clarkson University.



Sponsors

FOUNDATION SPONSORS



Amazon Web Services

Amazon Web Services (AWS) is guided by customer obsession, pace of innovation, commitment to operational excellence, and long-term thinking. By democratizing technology for nearly two decades and making cloud computing and generative AI accessible to organizations of every size and industry, AWS has built one of the fastest-growing enterprise technology businesses in history. Millions of customers trust AWS to accelerate innovation, transform their businesses, and shape the future.

Learn more at aws.amazon.com and follow @AWSNewsroom.



AMD

For more than 50 years, AMD has driven innovation in high-performance computing, graphics, and visualization technologies. Billions of people, leading Fortune 500 businesses, and cutting-edge scientific research institutions around the world rely on AMD technology daily to improve how they live, work, and play. AMD employees are focused on building leadership high-performance and adaptive products that push the boundaries of what is possible. For more information about how AMD is enabling today and inspiring tomorrow, engage with the AMD website, blog, LinkedIn, and X pages.

Learn more at www.amd.com/en/products/accelerators/instinct.html and amd.com/hpc.



PLATINUM SPONSORS



Google Cloud

Google Cloud offers a powerful, optimized AI stack — including infrastructure, leading models like Gemini, data management capabilities, multicloud security solutions, developer tools and platform, and agents and applications — that enables organizations to transform their business for the Agentic Era. Customers in more than 200 countries and territories turn to Google Cloud as their trusted technology partner.

Learn more at www.cloud.google.com.



Hewlett Packard

HPE

HPE is a leader in essential enterprise technology, bringing together the power of AI, cloud, and networking to help organizations achieve more. As pioneers of possibility, our innovation and expertise advance the way people live and work. We empower our customers across industries to optimize operational performance, transform data into foresight, and maximize their impact. Unlock your boldest ambitions with HPE.

Learn more at www.hpe.com



Quantinuum

Quantinuum is a leading quantum computing company offering a full-stack platform designed to make quantum computing deployable in real-world environments. The company, which has commercially deployed multiple generations of high-fidelity quantum systems, has active engagements with market leaders across pharmaceuticals, material science, financial services, and government and industrial markets. Quantinuum has a global workforce of approximately 700 employees, including top scientists and researchers. Over 70% of its technology team hold PhDs and Masters. Its headquarters is in Broomfield, Colorado, with additional facilities across the United States, United Kingdom, Germany, Japan, and Singapore.

Learn more at www.quantinuum.com

CORPORATE SPONSORS



Hammerspace

Hammerspace is the data platform to accelerate AI anywhere — on the infrastructure you already own. The Hammerspace hybrid cloud data plane unifies data silos into a single global namespace and orchestrates automated data delivery to GPUs and AI models. With metadata-indexed discovery and a standards-based approach, teams can find and use the right data fast, while maintaining data control and governance. Built to work across sites, clouds, and existing storage, Hammerspace helps organizations make their data AI-ready without replatforming, eliminate lock-in, and accelerate AI initiatives.

Learn more at hammerspace.com.



InspireSemi

InspireSemi provides revolutionary high-performance, energy-efficient accelerated computing solutions for HPC, AI, graph analytics, and other compute-intensive workloads. The Thunderbird “supercomputer-cluster-on-a-chip” is a disruptive, next-generation datacenter accelerator designed to address multiple underserved and diversified industries, including computer-aided engineering, energy, climate modeling, cybersecurity, financial services, and life sciences & drug discovery. Based on the open standard RISC-V instruction set architecture, InspireSemi’s solutions set new standards of performance, energy efficiency, and ease of programming.

Learn more at inspiresemi.com.



Intel

Intel (Nasdaq: INTC) is an industry leader, creating transformative technology that modernizes government operations and improves citizen outcomes. Inspired by Moore’s Law, Intel continuously works to advance the design and manufacturing of semiconductors to help address the public sector’s greatest modernization challenges. By embedding intelligence in government cloud platforms, secure networks, edge computing systems, and citizen-facing devices — powered by advanced analytics and AI — we unleash the potential of data to transform how government serves the American people.

Learn more at newsroom.intel.com and intel.com.



Parallel Works

Parallel Works addresses the complexity of running HPC and AI workloads across fragmented hybrid environments, spanning on-prem systems, multiple clouds, and heterogeneous hardware. These environments are difficult to operate, with inconsistent interfaces, schedulers, and security models that limit scalability and portability. The company focuses on unifying access to distributed compute resources, abstracting infrastructure differences, and enabling consistent workload execution. By improving orchestration,

governance, and cost visibility, Parallel Works increases resource utilization, reduces operational overhead, and accelerates time-to-compute for data-intensive workloads.

Learn more at parallelworks.com.



Rebellions

Rebellions builds what the AI era needs: purpose-built inference accelerators delivering the performance, efficiency, and supply chain resilience that enterprises and nations demand. Its flagship Rebel100™ delivers the best performance per dollar per watt — built for inference from the ground up, not retrofitted from training — with full-stack software built entirely on open source and open standards. With proven commercial deployments already live across enterprises and governments, and a chiplet architecture built for the most demanding AI workloads, Rebellions gives every organization and nation the ability to own and control their AI, not just access it. Learn more at rebellions.ai.



SambaNova

SambaNova is the leading enterprise AI company delivering a full-stack platform — from silicon to software — enabling instant inference, low-power, high-efficiency computing. Our approach allows organizations to confidently, securely, and privately own the trajectory of their data and AI future. Learn more at sambanova.ai.



TotalCAE

TotalCAE specializes in managed HPC solutions for engineering simulation. Driven by a passion to help clients increase their engineering throughput, TotalCAE is at the forefront of HPC solutions. Our solution combines both on-premises and cloud-based HPC cluster management with our TotalCAE Platform that makes submitting to HPC easy as three clicks. Our hybrid approach to HPC lets you generate data faster and up to 60% cheaper for day-to-day workloads than competing solutions. All supported with 1-hour response from a team of computer-aided engineering IT experts. Learn more at totalcae.com

INSTITUTIONAL SPONSORS



Argonne National Lab

Argonne National Laboratory (ANL) is a multidisciplinary science and engineering research center where leading scientists and engineers work together to answer the biggest questions facing humanity — from how to obtain reliable and affordable energy, to how to protect ourselves from emerging threats. The laboratory works in concert with universities, industry, and other national laboratories on questions and experiments too large for any one institution to do by itself. Through collaborations here and around the world, we strive to discover new ways to develop energy innovations through science, create novel materials molecule-by-molecule, and gain a deeper understanding of our planet and the cosmos.

Learn more at anl.gov.



Lawrence Berkeley National Laboratory

Lawrence Berkeley National Laboratory (Berkeley Lab) is committed to groundbreaking research focused on discovery science and solutions for abundant and reliable energy supplies. The lab's expertise spans materials, chemistry, physics, biology, earth and environmental science, mathematics, and computing. Researchers from around the world rely on the lab's world-class scientific facilities for their own pioneering research. Founded in 1931 on the belief that the biggest problems are best addressed by teams, Berkeley Lab and its scientists have been recognized with 16 Nobel Prizes. Berkeley Lab is a multiprogram national laboratory managed by the University of California for the U.S. Department of Energy's Office of Science. DOE's Office of Science is the single largest supporter of basic research in the physical sciences in the United States, and is working to address some of the most pressing challenges of our time.

Learn more at lbl.gov.



Brookhaven National Laboratory

Brookhaven National Laboratory (BNL) is a multidisciplinary laboratory with seven Nobel Prize-winning discoveries, 37 R&D 100 Awards, and countless advances in science and technology since being founded in 1947. BNL leads and supports research teams that address the DOE mission to ensure the nation's security and prosperity by addressing its energy, environmental, and nuclear challenges through transformative science and technology solutions. The lab builds and operates facilities that house unique tools used by interdisciplinary teams comprising researchers from Brookhaven as well as those from other national labs, academia, and industry.

Learn more at bnl.gov.





Los Alamos National Laboratory

LANL is a multi-program, federally funded research and development center for the National Nuclear Security Administration (NNSA) of the U.S. Department of Energy (DOE). LANL's priority roles are serving as a nuclear weapons design agency and a nuclear weapons production agency; addressing nuclear threats; and performing national security science, technology, and engineering. LANL is committed to nuclear deterrence and national security through innovation in research and development.

Learn more at lanl.gov.



Oak Ridge National Lab

Oak Ridge National Laboratory (ORNL) delivers the scientific discoveries and technical breakthroughs required to realize solutions in energy and national security and provide economic benefit to the nation. ORNL addresses national needs through impactful research and world-leading research centers. A wide range of partnerships with other US Department of Energy (DOE) laboratories and programs, universities, and industry pairs ORNL's strengths with others for outstanding contributions to science.

Learn more at ornl.gov.



RIKEN R-CCS

RIKEN, a National Research and Development Agency, is Japan's largest comprehensive research institution renowned for high-quality research in a diverse range of scientific disciplines. Founded in 1917, initially as a private research foundation, RIKEN has grown rapidly in size and scope, today encompassing a network of world-class research centers and institutes across Japan. RIKEN promotes the science of computing, by computing, and for computing — implementing the newest research that integrates simulation, big data analysis, and AI through HPC to solve scientific and social issues and to bring about revolutionary development in our society.

Learn more at riken.jp/en.



Texas Advanced Computing Center

The Texas Advanced Computing Center (TACC) at The University of Texas at Austin is the leading supercomputing center in the country. TACC delivers world-class and innovative systems, tools, software, and expertise to researchers who seek to make an impact in the world, and advance discovery across disciplines. Since 2001, TACC has provided access to a wide range of computational resources and offered staff expertise, thought leadership, support, training, and education and outreach. TACC is also home to the U.S. National Science Foundation Leadership-Class Computing Facility (NSF LCCF), which is poised to revolutionize computational research and discovery.

Learn more at tacc.utexas.edu.



University of Illinois Chicago

The University of Chicago is a world-renowned private research university dedicated to pushing the boundaries of knowledge and fostering innovative solutions to global challenges. Its faculty and students pursue transformative scholarship across the physical, biological, and social sciences, the humanities, and professional schools. Home to pioneering work in economics, physics, medicine, and computer science — including foundational contributions to artificial intelligence — the university partners closely with Argonne National Laboratory and Fermilab to accelerate discovery at scale. With an ethos of rigorous inquiry, interdisciplinary collaboration, and public impact, the University of Chicago is proud to support the Trillion Parameter Consortium's vision and research-driven mission.

Learn more at www.uic.edu.

NON-PROFIT SPONSOR



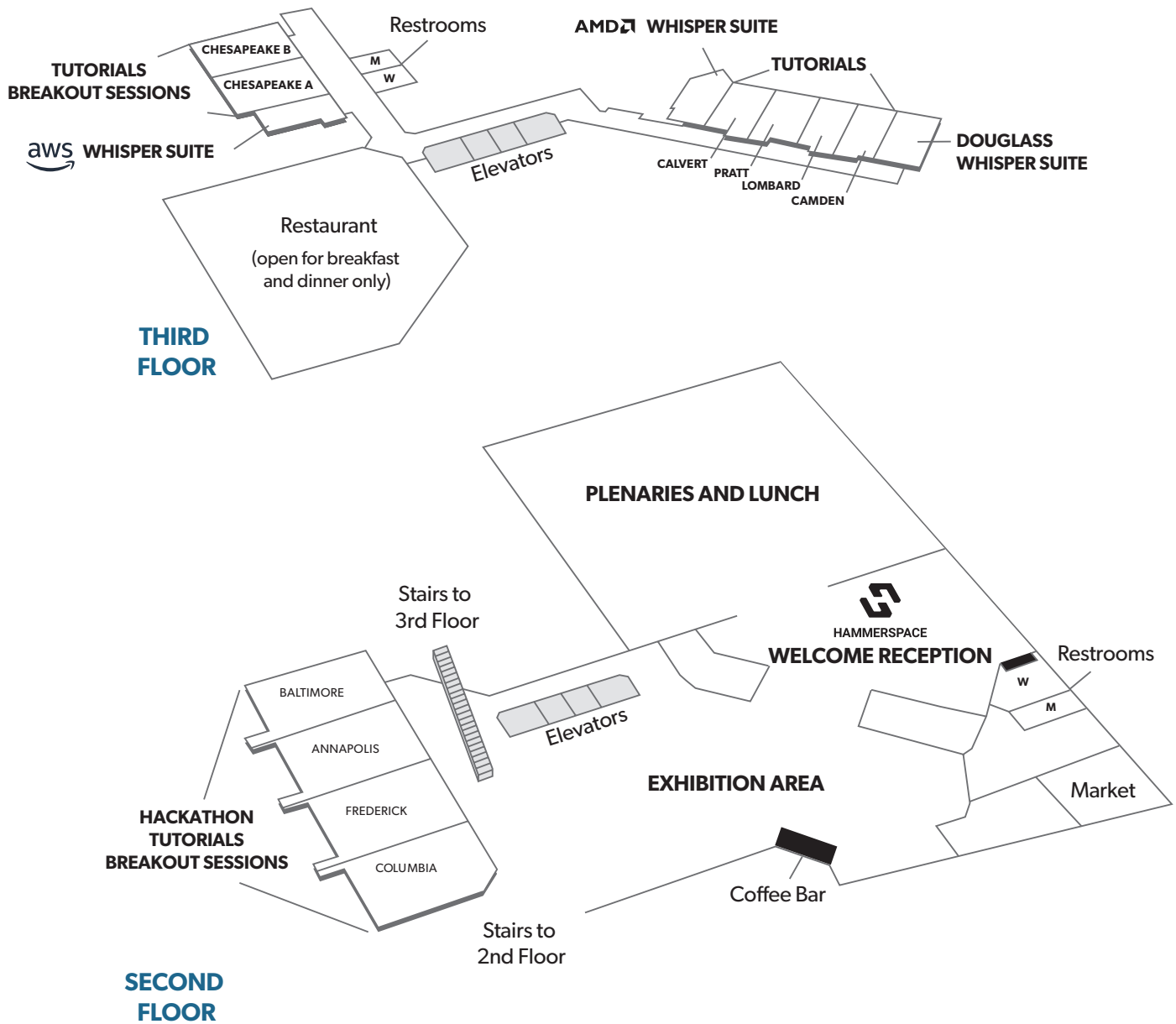
The HPC-AI Society

The HPC-AI Society is a vendor-neutral, non-profit organization whose sole purpose is to educate and promote the common business and technology interests of the HPC-AI community, with the goal of accelerating HPC, AI, data science, and quantum computing through community collaboration. The Society organizes and conducts open forum meetings and manages noncompetitive research activities that address the use, availability, standardization, and evaluation of existing technology, while introducing emerging technology to the HPC-AI community. Membership is open to all HPC and AI professionals serving all industries, government, and academia.

Learn more at www.hpc-ai-society.org.

CONFERENCE MAP

**HACKATHON / TUTORIAL
NETWORKING GATHERING
on the 15th floor**



TPC26 Breakout Group Lightning Talks

DRIVING CHALLENGE APPLICATIONS (CHALLENGE APPLICATIONS): Main Plenary Ballroom

Identify challenge applications for driving and evaluating the Infrastructure to Enable Shared Data & Computing, Open Frontier Models (Model Architecture & Performance Evaluation), and Open Frontier AI Systems tracks. Not centrally picking winners and losers, but asking the community to volunteer (and drive) scientific challenge applications, aiming for diversity on multiple axes (including industry applications).

AI for Material Sciences

Session 1 (Tuesday, June 2, 14:00)

- Human-in-the-Loop and Mixed Acceleration for Next-Generation Catalysis – *Edward (Ted) Sargent (Northwestern University)*
- Discovering Structural Rules in Scattering Amplitudes via Information Lattice Learning – *Hazu Yu (Kocree Inc.)*
- Open Source AI for Innovations in Energy and Materials Sciences – *Laura Gorman (Meta)*
- AIMNet2: A Foundational Machine-Learned Interatomic Potential for General Chemistry, Reactions, and Open-Shell Systems – *Shams Mehdi (Carnegie Mellon University)*
- Toward Scientific Superintelligence: Autonomous Agent Frameworks for Accelerated Materials Discovery – *Cameron John Owen (Lila Sciences)*

Session 2 (Tuesday, June 2, 16:00)

- Nahual: A Sequence Model for Language and Atoms – *Austin Cheng (University of Toronto)*
- Quantum Computing, Embedding, and Machine Learning for Predictive Chemistry – *Arvin Kakekhani (PsiQuantum)*
- Accelerating Scientific Discovery With NeMo Agent Toolkit – *Xiaoyun Wang (NVIDIA)*
- AI and Experimental Automation – *Emma Bouchard (Carnegie Mellon University)*

GridAI Model Team

(Wednesday, June 3, 8:30)

- Scalable Heterogeneous Graph Learning for Grid AI Foundation Models – *Massimiliano Lupo Pasini (Oak Ridge National Laboratory)*
- GridMind: An Agentic Workflow for Weather-Driven Power Grid Risk Assessment – *Hongwei Jin (Argonne National Laboratory)*
- AI Applications in Transmission Planning and Operations – *Anish M. Gaikwad (Electric Power Research Institute)*
- Scaling Laws for Heterogeneous Graph Learning in Optimal Power Flow Applications – *Emon Dey (Argonne National Laboratory)*

BOF: Bio-Foundation Models, Agentic systems, and Biosecurity

(Wednesday, June 3, 11:00)

- Scalable Agentic Reasoning for Designing Biologics Targeting Intrinsically Disordered Proteins – *Archit Vasan (Argonne National Laboratory)*

AI and Strategic Decision Support

(Wednesday, June 3, 14:00)

- From a Napkin To a Workflow — Opportunities and Challenges for Workflow Composition – *Ewa Deelman (University of Southern California)*
- Evaluating Epistemic Non-Triviality of Large Language Reasoning Models in Scientific Hypothesis Generation – *Tirthankar Ghosal (Oak Ridge National Laboratory)*

AI SOFTWARE INFRASTRUCTURE/FRAWORKS (SOFTWARE STACK)

Develop software infrastructure and middleware to support the training, deployment, and integration of complex frontier-scale AI models and systems. Provide the technical backbone for the Open Frontier Models (Model Architecture & Performance Evaluation) and Open Frontier AI Systems tracks, while enabling integration with experimentation platforms, laboratories, instruments, and other real-world scientific environments.

BOF: AI Frameworks for Multimodal Data Access and Use (Baltimore)

(Tuesday, June 2, 14:00)

- Operationalizing Sovereign Data for AI Inference and Physical AI – *Molly Presley (Hammerspace)*
- From Access to Action: The NAIRR Pilot Portal and Its Interactive Sandboxes – *Sandra Gesing (San Diego Supercomputer Center)*

DWARF: Data Workflows, Agents Reasoning, and Frameworks (Baltimore)

Session 1 (Tuesday, June 2, 16:00)

Session 2 (Wednesday, June 3, 8:30)

- Best Practices for Scientific Workflows – *Robert Underwood (Argonne National Laboratory)*
- When More Cores Hurts: The Vector Database Scaling Paradox in HPC – *Seth Ockerman (University of Wisconsin)*
- Agents with Agency – *Yadu Babuji (University of Chicago)*

- Deploying Agentic AI Across the DOE Accelerator Complex: The MOAT Experience – *Thorsten Hellert (Lawrence Berkeley National Laboratory)*
- Unified User Experience Across Heterogeneous GPU Clusters with Diamond – *Zhao Zhang (Rutgers University)*
- Enabling Autonomous Scientific Discovery Through Agentic AI and High-Performance Computing – *Murat Keceli (Argonne National Laboratory)*
- Building Reusable and Trustworthy AI Co-Scientists: Lessons From Multi-Domain Scientific Deployments – *Chandrachur Bhattachar (Argonne National Laboratory)*
- The Semiconductor Engineering Inflection – *Srikanth Gubbala (Applied Materials)*

AI Software Infrastructure Frameworks (Columbia)

(Tuesday, June 2, 16:00)

- AI and CI Services From NSF-AI Institute ICICLE – *D.K. Panda (The Ohio State University)*
- Planned Reproducible Orchestration of Tools for LLM-Based Agents – *Elliott Jacopin (Center for Biosystems Dynamics Research, RIKEN R-CSS)*
- MoE-Inference-Bench Performance Evaluation of Mixture of Expert Large Language and Vision Models – *Murali Emani (Argonne National Laboratory)*
- NAPA Project: AI Testbed Deployment in Japan – *Jason Haga (National Institute of Advanced Industrial Science and Technology (AIST))*
- Scaling MoE Inference with vLLM on Aurora – *Padma Apparao (Intel)*
- EIMerFold: Exascale Distillation Workflows for Protein Structure Prediction on El Capitan – *Nikoli Dryden (Lawrence Livermore National Laboratory)*

BOF: Large Language Models for Scientific Software (Baltimore)

(Wednesday, June 3, 11:00)

- Applications of Deep Knowledge Graph – *Geoffrey Fox (University of Virginia)*
- Curating Agentic Workflows with Knowledge Graphs and Operational Experience – *Ana Gainaru (Oak Ridge National Laboratory)*
- A Tale of Two Agentic Frameworks: Empirical Studies of LangChain/LangGraph and AG2 in Autonomous Scientific Workflows – *Meifeng Lin (Brookhaven National Laboratory)*
- Enabling Autonomous Scientific Discovery Through Agentic AI and HPC – *Murat Keceli (Argonne National Laboratory)*

BOF: Trillion Parameter Models for the Edge and Computing Continuum (Columbia)

(Wednesday, June 3, 11:00)

- Catalyzing AI-Enabled Science Across the Digital Continuum for Science – *Manish Parashar (University of Utah)*
- Supporting the Continuum AI Ecosystem in the National Data Platform – *Ilkay Altintas (San Diego Supercomputer Center)*
- Catalyzing AI-Enabled Science Across the Digital Continuum for Science – *Manish Parashar (University of Utah)*
- Bridging the Computing Continuum – *Sean Shahkarami (Northwestern University)*

INFRASTRUCTURE TO ENABLE SHARED DATA & COMPUTING (SHARED INFRASTRUCTURE):

Annapolis

Collectively build scientific training data resources and shared computing infrastructure for model training and further fine-tuning for general-purpose and domain-specific settings. Establish scalable and sustainable capabilities that serve as the foundation for the rest of the tracks.

BOF: Trustworthy Privacy-Preserved Federated Learning for Science

(Tuesday, June 2, 14:00)

- NeuroFL: OBI's Intelligence Network for Brain Health – *Bryce Pickard (Ontario Brain Institute)*
- OmniFed: Towards Configurable Cross-Silo Federated Learning – *Sahil Tyagi (Oak Ridge National Laboratory)*
- Differentially Private Federated Averaging with James-Stein Estimator – *Minseck Ryu (Arizona State University)*
- Socio-Technical Infrastructure: Operationalizing FL Systems – *Mohammed Manzari (Deloitte)*
- Are You Ready for Production Federated Learning? – *Holger Roth (NVIDIA)*
- Federated LLM Training Across NNSA Lab – *Max Carlson (Sandia National Laboratories)*
- Scalable Cross-Facility Federated Learning for Scientific Foundation Models on Multiple Supercomputers – *Yijiang Li (Argonne National Laboratory)*
- Towards Trustworthy Federated AI: Privacy, Ownership Protection, and Model Editing – *Olivera Kolveksa (Oak Ridge National Laboratory)*
- The Next Frontier: Federated AI with Flower – *William Lindskog-Munzing (Flower Labs)*

BOF: 50-State AI Plan: A Grassroots Approach to Building a US AI Continent

(Tuesday, June 2, 16:00)

- Catalyzing the Utah Responsible AI Innovation Ecosystem – *Manish Parashar (The University of Utah)*
- Experimental Design for Foundation Models: From Uncertainty to Risk – *Yu Wen (Stony Brook University)*
- The Pittsburgh Supercomputing Center, An Integrated National Resource for AI for Science – *Paola Buitrago (Pittsburgh Supercomputing Center)*
- From Vision to Momentum: The AI Tennessee Blueprint – *Tabitha Samuel (University of Tennessee)*
- The Keystone AI + Quantum Factory – *Barr von Oehsen (Pittsburgh Supercomputing Center)*

BOF: Bold New World of Heterogenous AI Computing

Session 1 (Tuesday, June 3, 11:00)

- Macroheterogeneity: Enabling Hybrid HPC and AI Workflows – *Samantika Sury (HPE)*
- Inference-First AI Design – *Tim Clarke (SambaNova Systems)*
- AI, HPC, Precision, and the Uncertain Future – *Robert (Bob) Robey (AMD)*
- Off the Leash: Device-Native Heterogeneous Execution – *Saliya Ekanayake (d-Matrix)*
- MoE-Inference-Bench Performance Evaluation of Mixture of Expert Large Language and Vision Models – *Murali Emani (Argonne National Laboratory)*
- MoE at Scale: From GPUs to Wafer-Scale AGI – *Daria Soboleva (Cerebras)*

Session 2 (Tuesday, June 3, 14:00)

- Optimizing Inference at Wafer Scale? *Daria Soboleva (Cerebras)*
- Efficient and Scalable Agentic AI with Heterogeneous Systems – *Zain Asgar (Gimlet Labs)*
- Prefill Here, Decode There: Disaggregated LLM Serving Across GPUs and LPUs – *Vineeth Gutta (NVIDIA)*
- One Abstraction, Many Transports: Streaming Communication for Disaggregated Inference – *Sai Rahul Chalamalasetti (d-Matrix)*
- The NAPA Project: Inference Systems – *Jason Haga (National Institute of Advanced Industrial Science and Technology (AIST))*

OPEN FRONTIER AI SYSTEMS (AI SYSTEMS): Frederick

Develop frontier AI systems for science that incorporate reasoning models (start with SOA closed models, eventually include Open frontier model) and develop domain foundation models, knowledge graphs, agentic systems and orchestrations, simulators, and experiments.

BOF: AI Agents as Scientific Collaborators: Building Human-Agent Research Teams

(Tuesday, June 2, 14:00)

- Agent-Enabled Paper to Code Generation for ML Reproducibility – *Zhao Zhang (Rutgers University)*
- Towards Intelligent CFD Workflow in the Era of Large Language Models – *Shaowu Pan (Rensselaer Polytechnic Institute)*
- An AI Research Assistant for Automating the Computational Catalysis Pipeline – *Ruchika Mahajan (Stanford University)*
- Model Capabilities Driving New Paradigms of Agentic Patterns – *Matt Baughman (Princeton Plasma Physics Laboratory)*
- Toward Agentic Closed-Loop AI for Battery Science: From SpectraQuery to Multimodal Experimental Agents – *Sreya Vangara (Stanford University)*

Toward Scientific AI Platforms: Inference, Agents, and AI Services at HPC Facilities

(Wednesday, June 3, 8:30)

- The NAPA Project: Inference Systems – *Jason Haga (National Institute of Advanced Industrial Science and Technology (AIST))*
- Agentic AI-Driven Inference Service for Scientific Applications – *Shaojia Fan (National Center for High-Performance Computing)*
- DragonHPC: A High-Performance Distributed Runtime for AI-Coupled HPC Workflows – *Samantika Sury (HPE)*
- ALCF Inference Service: Facility-Scale AI Services for Science – *Venkatram Vishwanath (Argonne National Laboratory)*
- Optimizing Inference at Wafer Scale? – *Daria Soboleva (Cerebras)*

BOF: Human-AI Collaboration

(Wednesday, June 3, 14:00)

- Building for Human-AI Scientific Co-Intelligence – *Patrick Emami (National Laboratory of the Rockies)*
- Evaluating Memory Condensation Strategies for Coding Agents in Data-Driven Scientific Discovery – *Sameera Horawalavithana (Pacific Northwest National Laboratory)*

OPEN SUITE FOR EVALUATING MODEL SKILLS, KNOWLEDGE, REASONING, & SAFETY (EVALUATION): Chesapeake A

Develop an open suite of tools, methods, and benchmarks for evaluating the scientific skills, knowledge, reasoning, agentic capabilities, and safety/security of frontier models and AI systems.

Evaluation of LLMs for Science

Session 1: Evaluation of Code Generation, Translation, and Parallelization (Tuesday, June 2, 14:00)

- Reliable Evaluation of LLMs for Scientific Parallel Code Translation – *Le Chen (Argonne National Laboratory)*
- Benchmarking LLM-Generated Parallel Code for Task-Based Workflow Programming Models – *Eduardo Iraola de Aceve (Barcelona Supercomputing Center)*
- CelloAI Benchmarks: Evaluating LLMs for HEP/HPC Software – *Kriti Chopra (Brookhaven National Laboratory)*

Session 2: Principled Approaches to AI Scaling, Reasoning, and Benchmarking (Tuesday, June 2, 16:00)

- Why Neural Scaling Laws Require an Information-Theoretic Foundation – *Sujata Sinha (Virginia Polytechnic Institute and State University)*
- Building Reusable and Trustworthy AI Co-Scientists: Lessons from Multi-Domain Scientific Deployments – *Chandrachur Bhattacharya (Argonne National Laboratory)*

- Difficulty-Oriented Reasoning Effort Modeling and Systematic Evaluation of Science Problems for Large Language Models – *Martin Foltin (HPE)*
- Pier: Efficient Large Language Model Pretraining with Relaxed Global Communication – *Zhao Zhang (Rutgers University)*

Agentic Reasoning with Scientific Foundation Models

(Wednesday, June 3, 8:30)

- Accelerating Scientific Discovery with NeMo Agent Toolkit – *Xiaoyun Wang (NVIDIA)*
- Simulation Surrogates Using Gaussian Splatting and Diffusion – *Han-Wei Shen (The Ohio State University)*
- Evaluating Agents Without Ground Truth – *Kirsten Hofmockel (Pacific Northwest National Laboratory)*

OPEN FRONTIER MODELS (OPEN MODELS): Chesapeake A

Build frontier-scale, open AI models using shared data and computing infrastructure (from the Infrastructure to Enable Shared Data and Computing track), harnessing distributed resources across TPC partner institutions. Ensure that all core components are openly available to enable transparency, reuse, and scientific progress.

Model Architectures and Performance Evaluation

Session 1 (Wednesday, June 3, 11:00)

Session 2 (Wednesday, June 3, 14:00)

- Scaling MoE Models to Exascale With Open Software on Aurora – *Padma Apparao (Intel)*
- Olmo Hybrid: From Theory to Practice and Back – *William Merrill (Tokyo Technological Institute)*
- RingX: Scalable and Efficient Long-Context Learning for Scientific Foundation Models on HPC – *Junqi Yin (Oak Ridge National Laboratory)*
- AIMNet2: A Foundational Machine-Learned Interatomic Potential for General Chemistry, Reactions, and Open-Shell Systems – *Shams Mehdi (Carnegie Mellon University)*
- Closing the Loop: Federated Learning for Distributed Scientific Systems – *Chandreyee Bhowmick (Oak Ridge National Laboratory)*

TRAINING- AND DEPLOYMENT-LEVEL SAFETY AND ALIGNMENT (SAFETY & ALIGNMENT): Frederick

Develop methods to embed safety and alignment into the training and deployment of frontier-scale models and AI systems. Focus on system-level mechanisms that maintain alignment with scientific objectives and constraints, and with broader societal values, at extreme scale and in high-impact settings.

BOF: Safety and Alignment in Agentic Systems

(Tuesday, June 2, 16:00)

- A Comprehensive Multilingual Jailbreak Evaluation of Open-Source Large Language Models – *Kashyap Manjusha (University of Illinois Urbana-Champaign)*
- A Full-Stack Approach to Frontier Model Safety: Red-Teaming, Interpretability, Unlearning, and Formal Verification – *Sumit Kumar Jha (University of Florida)*
- The Need for New Safety Measures for LLMs in Scientific Applications – *Saket Chaturvec (Argonne National Laboratory)*

EXPANDING AND DEEPENING THE AI WORKFORCE (WORKFORCE): Columbia

Identify and report on progress in developing the workforce required to achieve the rest of the tracks, with particular attention to emerging and evolving roles across the frontier AI stack. Examine needs across all career stages and share recent experiences and lessons learned to inform sustainable talent development.

BOF: Building the AI-Ready Science and Engineering Workforce: Skills, Systems, and Collaboration Required For Success

(Tuesday, June 2, 14:00)

- Igniting Alaska's AI Future: Developing Alaska's AI Talent Through Wildfire Analysis – *Arghya Kusum Das (University of Alaska Fairbanks)*
- ByteBoost: Building the AI Workforce for Advanced Computing Systems – *Paola A. Buitrago (Carnegie Mellon University | Pittsburgh Supercomputing Center)*

BOF: Who Builds the Future? Workforce Challenges in Trillion-Parameter Scientific Computing

(Wednesday, June 3, 8:30)

- How Fast Can We Evolve the Workforce? – *Roscoe Giles (Boston University)*
- From Writing Software to Ensuring It's Written Well: The Evolving RSE Role – *Arfon Smith (Schmidt Sciences)*
- Enabling Research Software Engineers to Leverage AI at APL: Policies, Tools, Training, and Use Cases – *Jon Vandegriff (Johns Hopkins University)*
- Designing Scientific Computing Ecosystems for the AI Era – *Lois Curfman McInnes (Argonne National Laboratory)*
- Lesson Learned from Hosting a Frontier AI and LLM Tutorial Series for a Mixed Audience – *Meifeng Lin (Brookhaven National Laboratory)*
- Iterative Co-Design, Co-Development, and Co-Delivery: Accelerating S&T Productivity – *Mary Ann Leung (Sustainable Horizons Institute)*