An Overview of the ORNL Artificial Intelligence Initiative

CRNCH Summit
November 2, 2018

David E. Womble

ORNL is managed by UT-Battelle, LLC for the US Department of Energy

AI won’t replace the scientist, but scientists who use AI will replace those who don’t*

*Adapted from a Microsoft report, “The Future Computed”

AI will change the way we do science

We are in an “AI arms race”

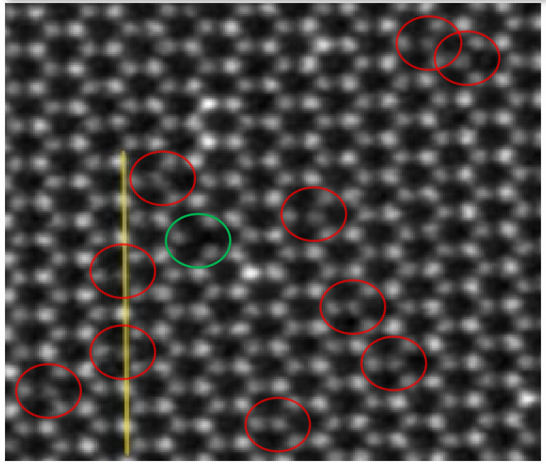
Why now? Why ORNL?

- Data is plentiful
- Computing is ubiquitous
- AI is accessible

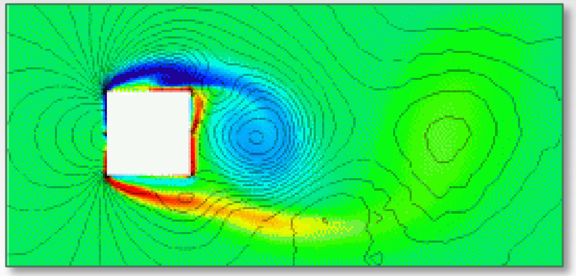
- Mission
- Facilities – we have the ability to produce, manage, store and analyze "extreme" data sets

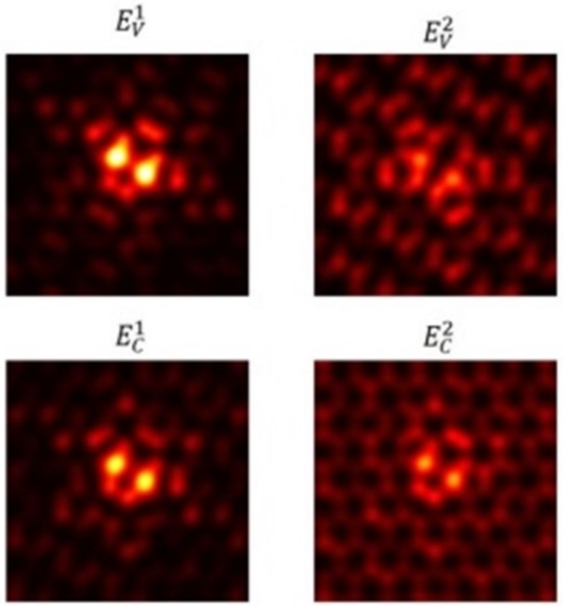
An AI Taxonomy Drives Research Strategy

Classification and regression

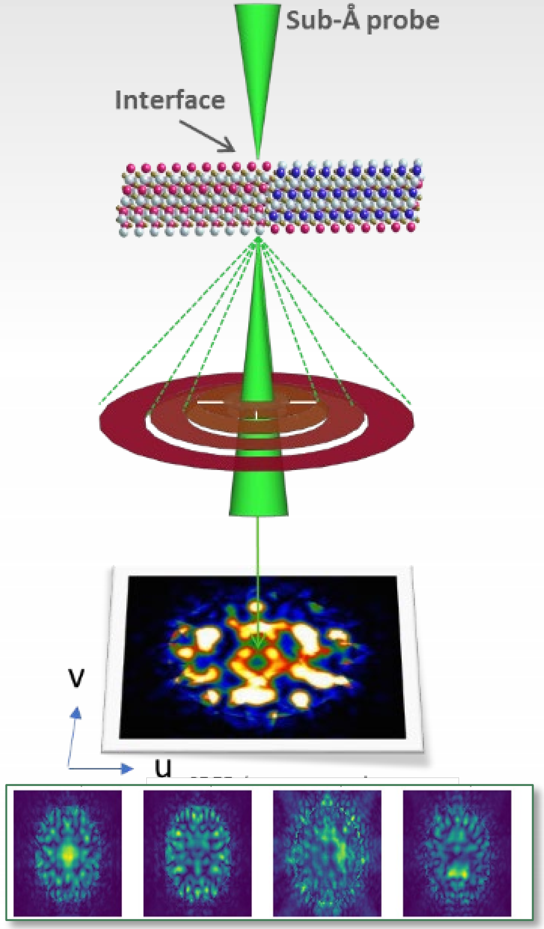


Surrogates

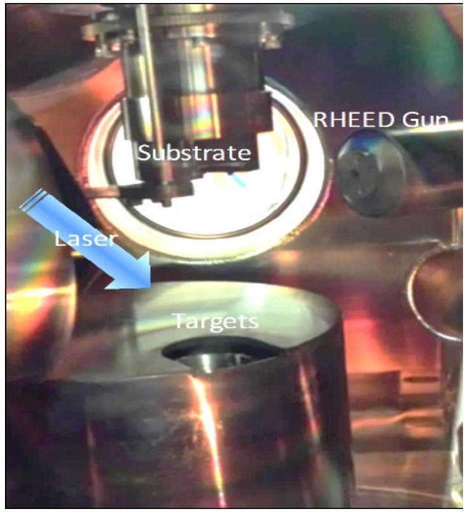




Inverse problems, design and optimization

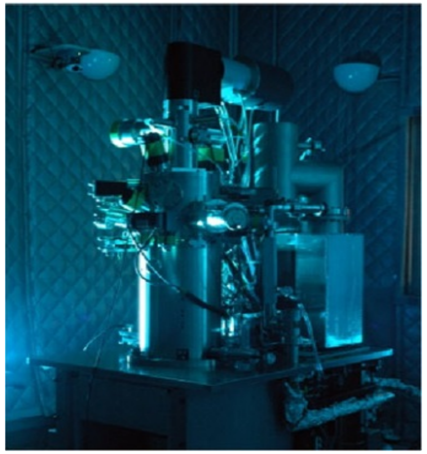


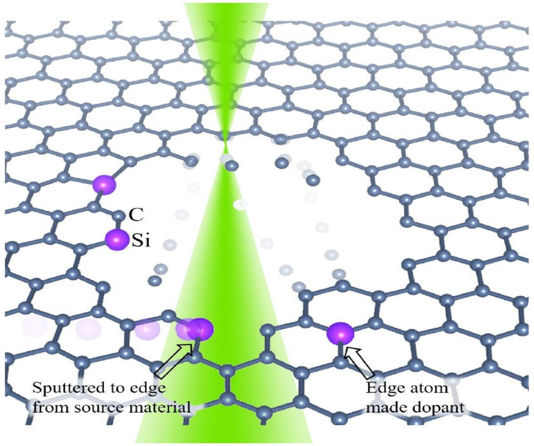
Control systems



Building and exploring libraries of atomic defects in graphene

Imaging and manipulating matter on the atomic level with Scanning Transmission Electron Microscope (STEM)



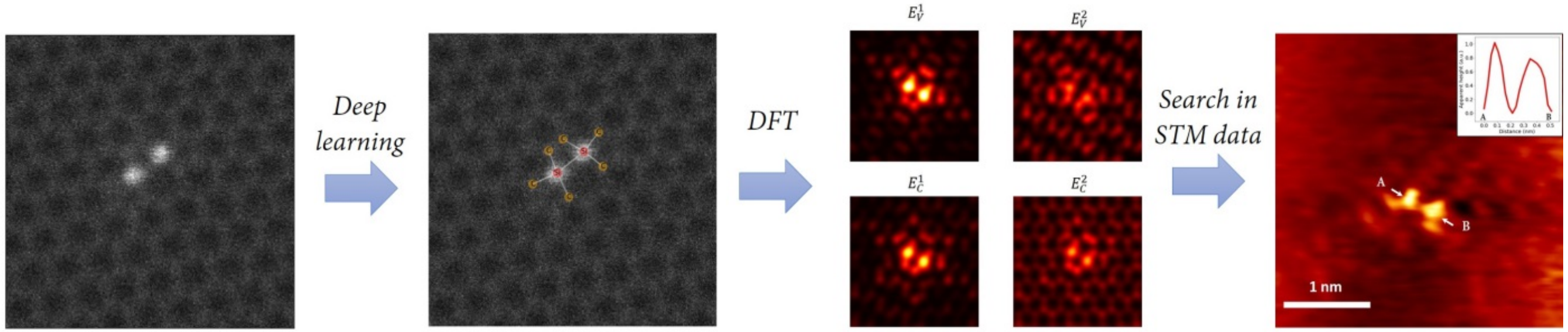


STEM measures local structure

Defect identification

Calculated electronic structure

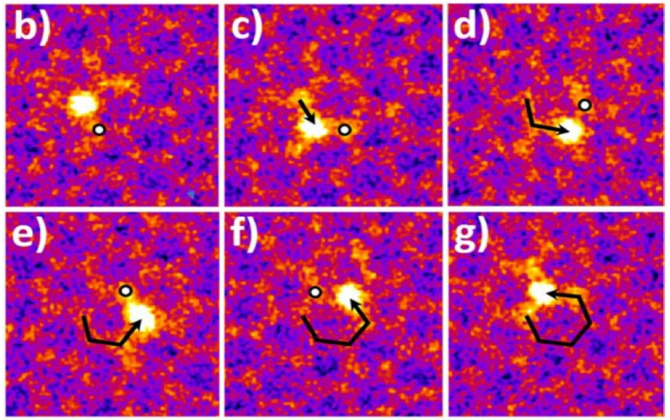
STM (measures local electronic properties)



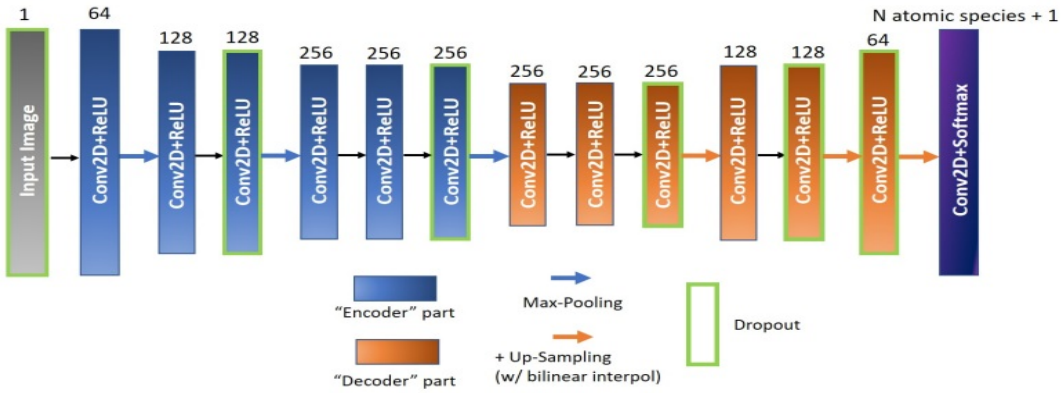
M. Ziatdinov, O. Dyck, B. G. Sumpter, S. Jesse, R. K. Vasudevan, S. V. Kalinin. *Building and exploring libraries of atomic defects in graphene: scanning transmission electron and scanning tunneling microscopy study*. ArXiv:1809.04256 (2018)

Rapid analysis of noisy data with fully convolutional neural network

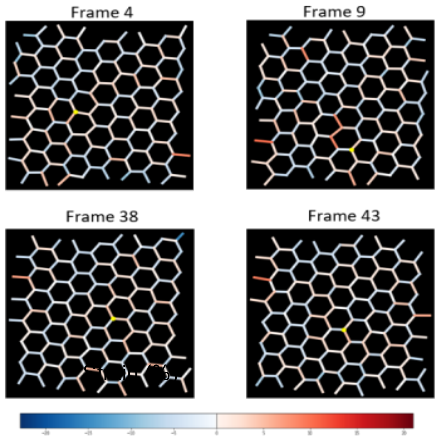
Experimental data (Si in graphene)



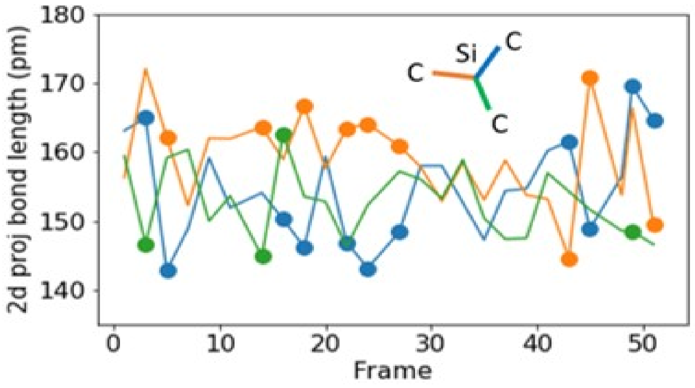
Neural network (schematics)



Decoded results: Exploration of structural symmetry breaking for atomic defects in graphene



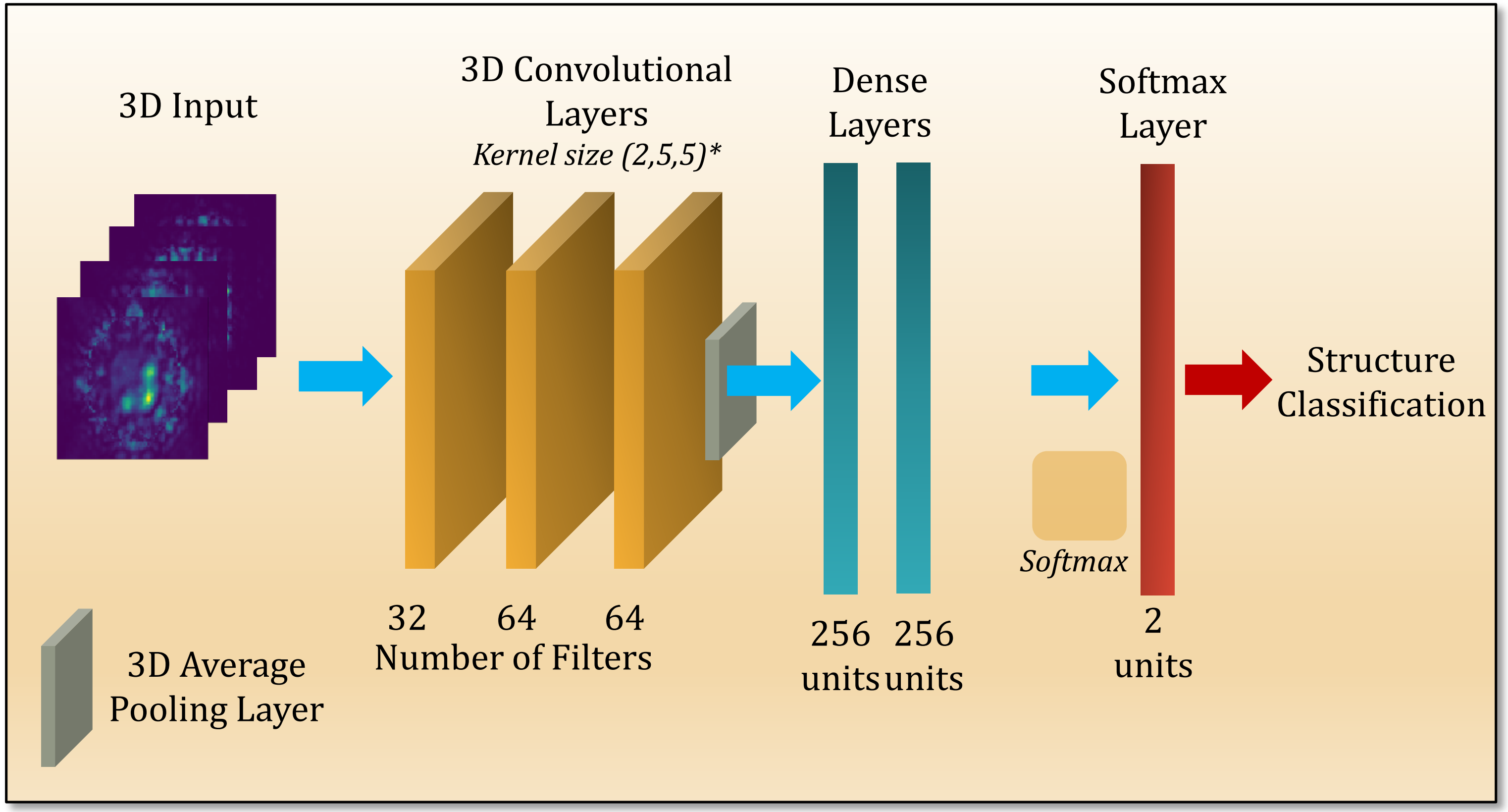
Variations of bond lengths in SiC defect



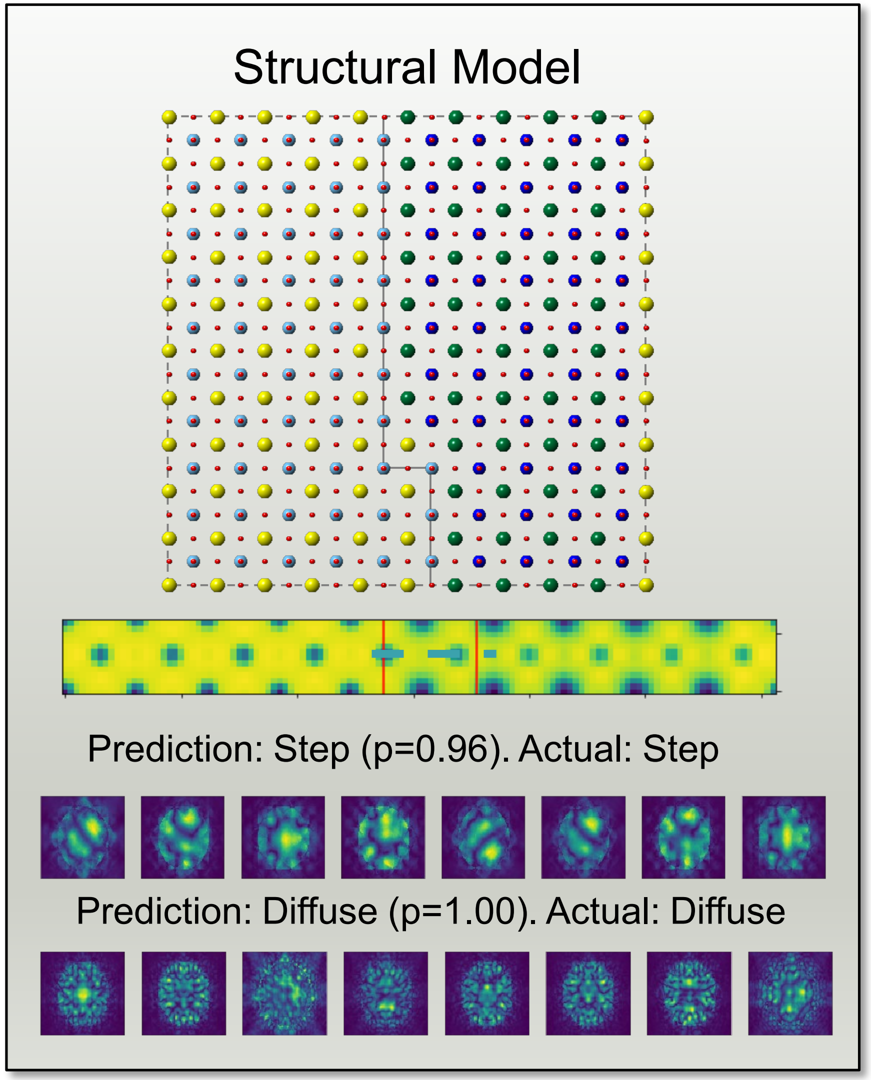
For levels of noise comparable to that in typical experiments, most of atomic position deviations from "ground truth" are below 12 pm (< 10 % of C-C bond)

M. Ziatdinov, S. Jesse, R. K. Vasudevan, B. G. Sumpter, S. V. Kalinin, O. Dyck. *Tracking atomic structure evolution during directed electron beam induced Si-atom motion in graphene via deep machine learning.* ArXiv:1809.04785 (2018).

Deep learning for solving inverse problems in imaging and multi-spectral data



Deep neural networks to distinguish structural features of simulated $LaAlO_3/SrTiO_3$ interfaces



OAK RIDGE National Laboratory

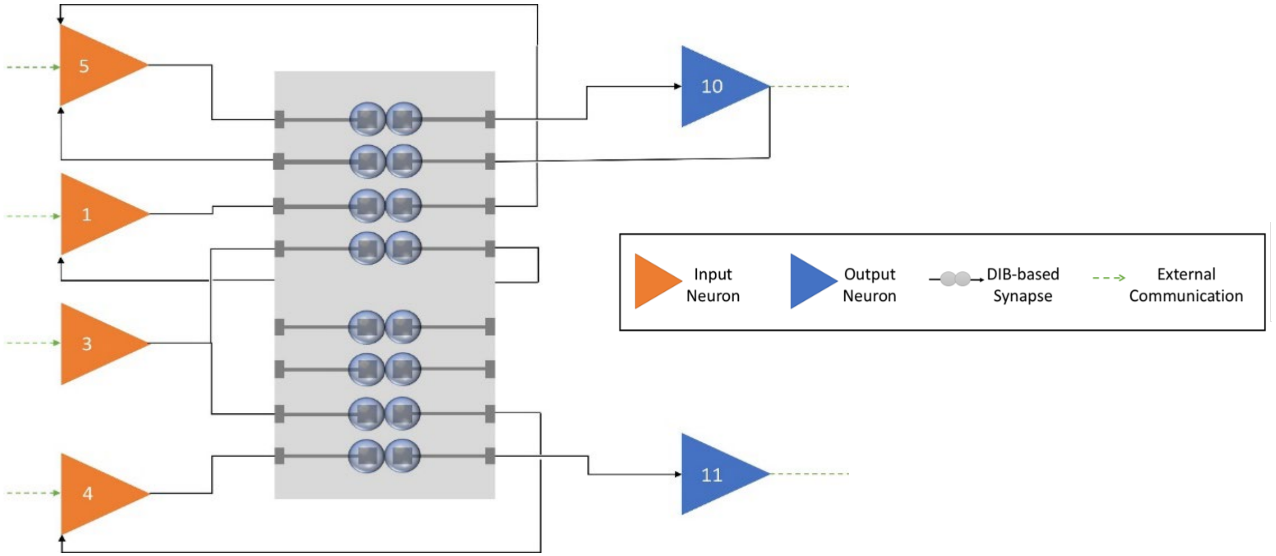
M. Oxley et al., in preparation

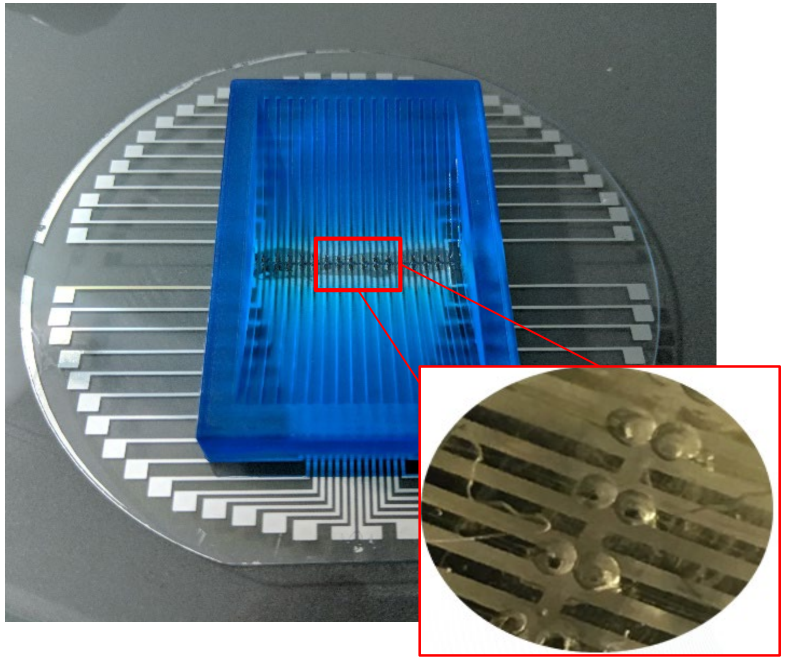
Machine Learning for Biosynaptic Neuromorphic Devices

Goal: classify time-dependent electroencephalographic data

- Hybrid biomimetic/solid-state devices
  - solid-state neurons
  - biomimetic membranes
  - co-located computing and memory
  - parallel computational operations
  - nanowatts

- Designed and trained using Evolutionary Optimization for Neuromorphic System (EONS) on Summit
- 3D-printed microfluidic-based devices were fabricated on silicon wafers with patterned silver electrodes

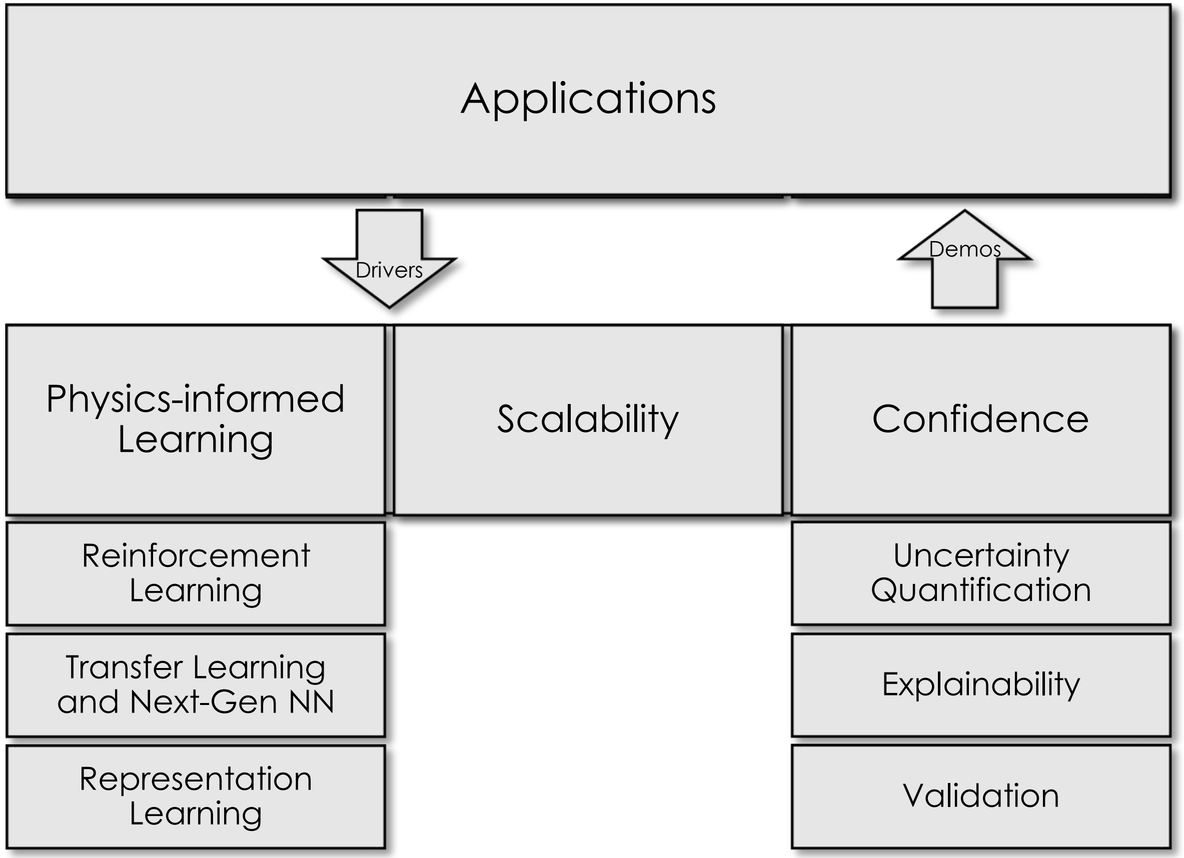




OAK RIDGE National Laboratory

Project Drivers and Research Structure

- Application drivers
  - Bioscience
  - Materials synthesis and additive manufacturing
- Learning
  - Reinforcement learning
  - Transfer learning and next-gen neural nets
  - Representation learning
- Scalability
- Confidence
  - UQ
  - Explainability
  - Validation



Scalability and HPC Architectures (Beyond Moore)

Summit includes

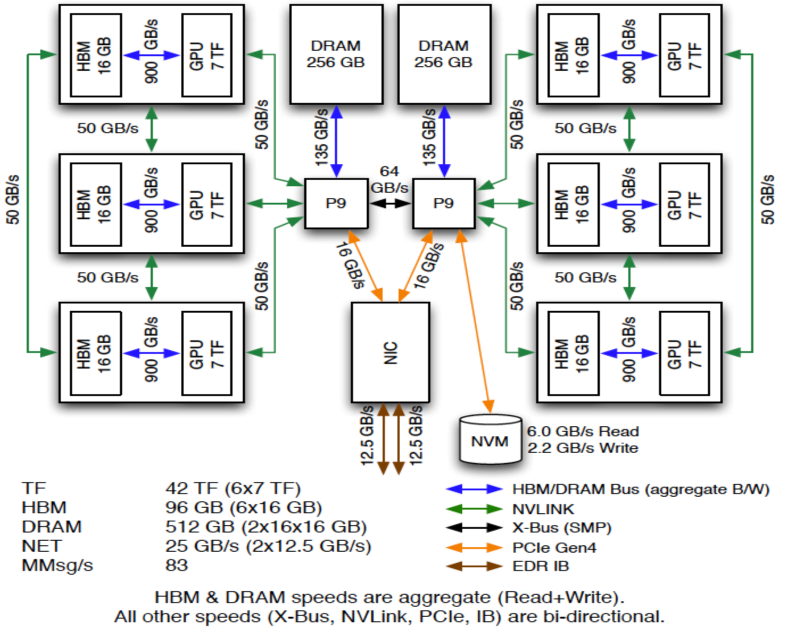
- 4608 nodes
- Dual-rail Mellanox EDR InfiniBand network
- 250 PB IBM Spectrum Scale file system transferring data at 2.5 TB/s

Each node has

- 2 IBM POWER9 processors
- 6 NVIDIA Tesla V100 GPUs
- 608 GB of fast memory
- 1.6 TB of NVMe memory

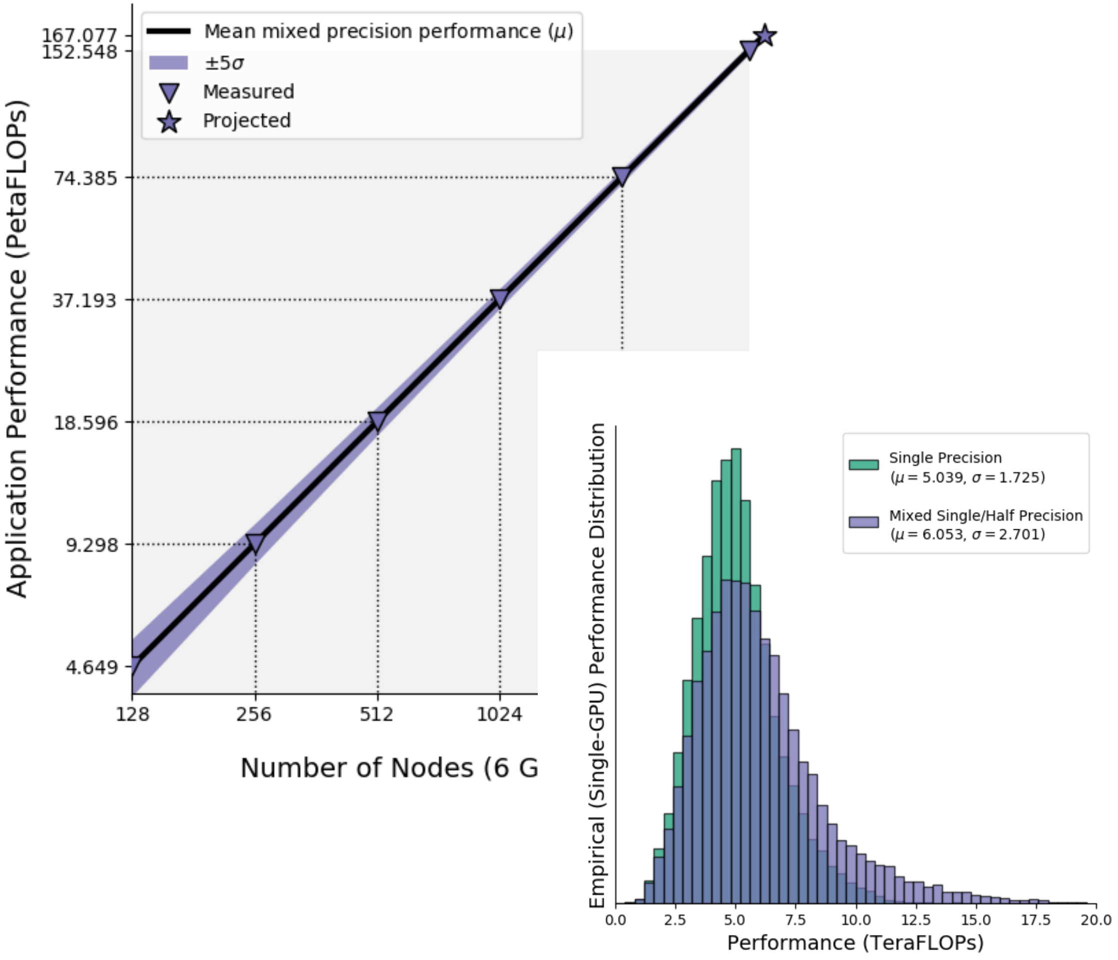
System Performance

- Peak performance of 200 petaflops for modeling & simulation
- Peak of 3.3 ExaOps for data analytics and artificial intelligence



Using Tensor Cores is key to high performance

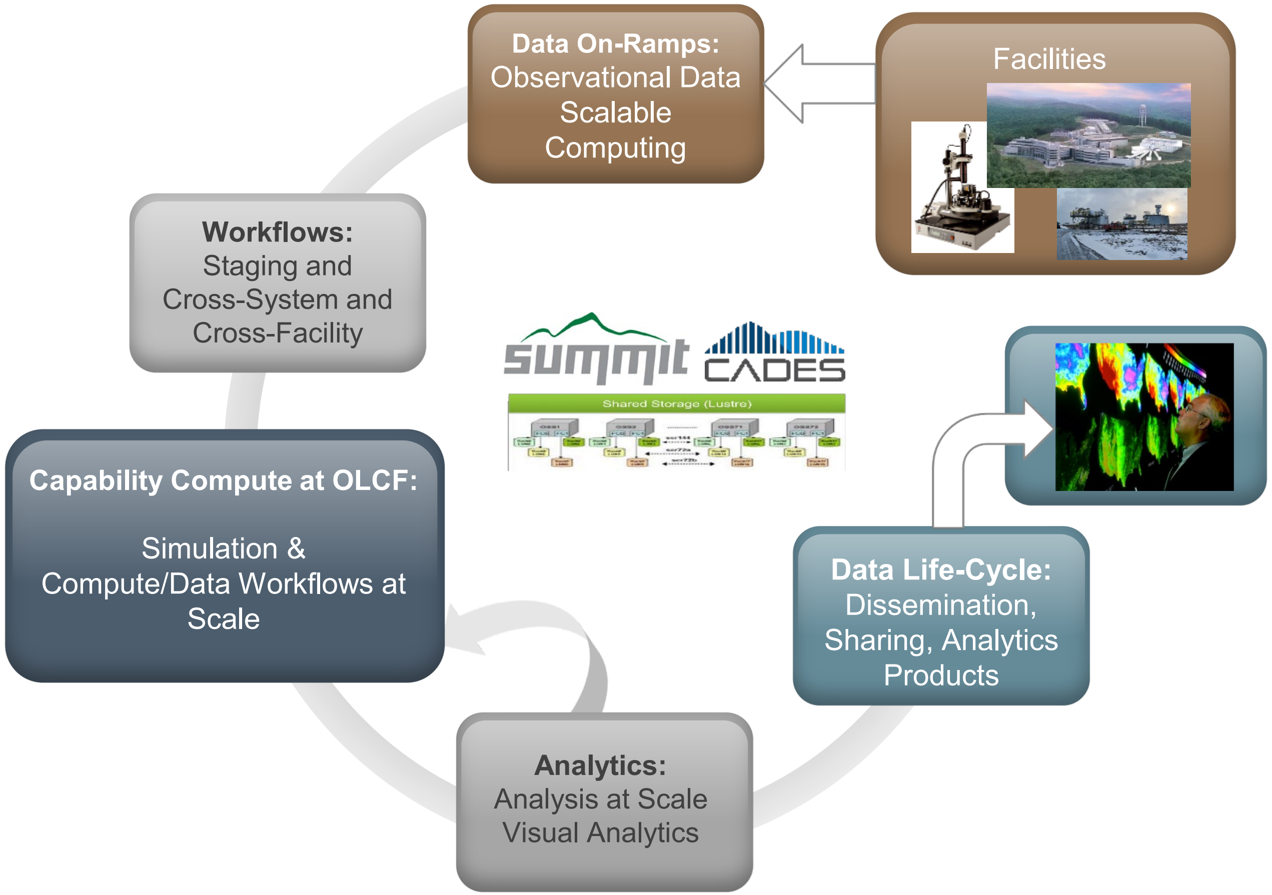
- MENNDL is the kernel of one of our Gordon Bell finalists
  - Determines optimal hyperparameters for a DNN
  - Relatively easy to parallelize
  - Effectively demonstrates the power of the Tensor core units (as well as the challenge of using them)
- Mixed precision presents algorithmic challenges
  - What accuracy is actually needed for simulations
  - Performance must now be correlated to accuracy



Travis Johnston, ORNL

OAK RIDGE National Laboratory

A Good Infrastructure Is Required To Manage Data

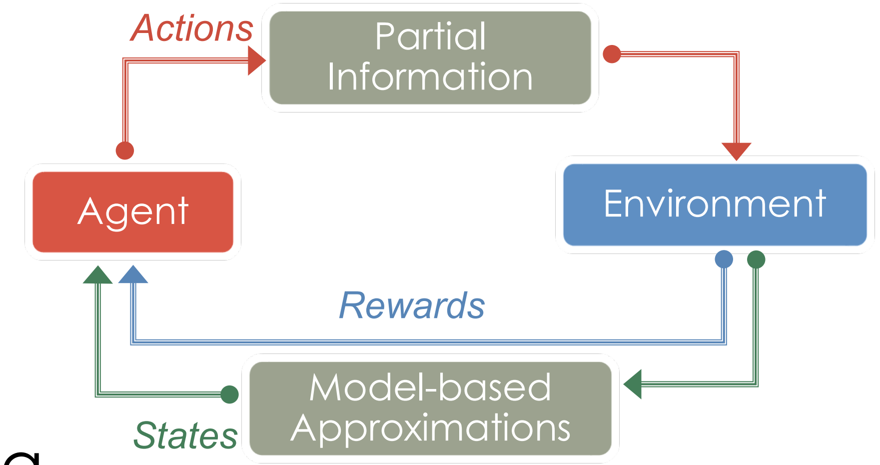


Fast, Efficient Learning with Limited Data

- Data is
  - Limited
  - Poor quality with large uncertainties
  - Unlabeled
  - Physically constrained

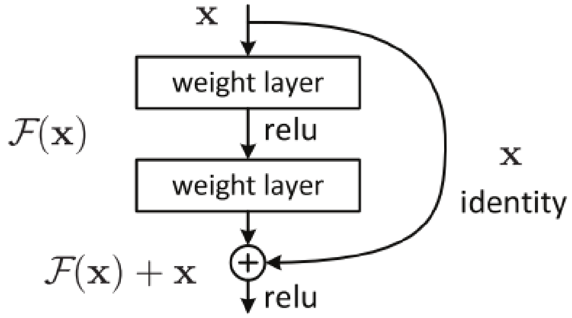
Address through reinforcement learning

- Agents in an environment learn actions that maximize some cumulative reward



- Learning algorithms
  - slow
  - networks are non-optimal

Address through stochastic optimal control algorithms



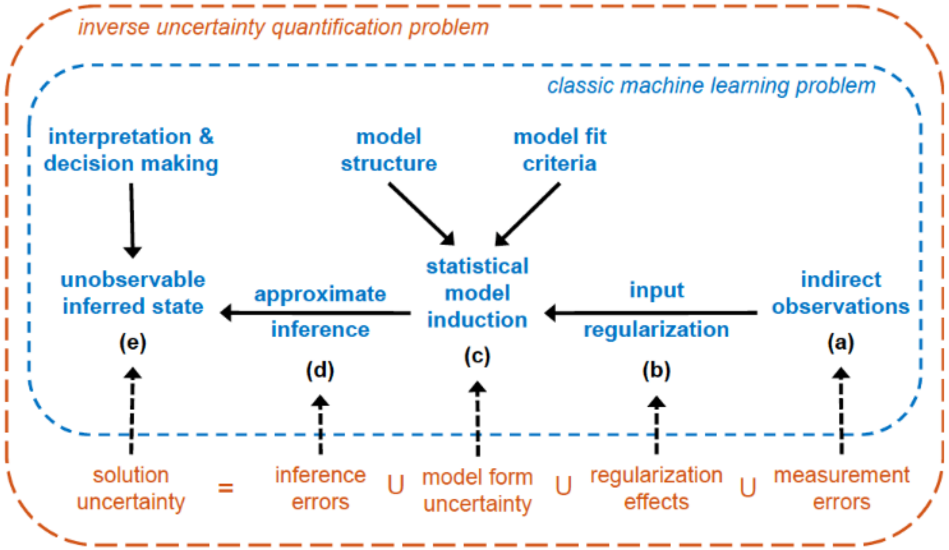
$$X_{N+1} = X_N + \Delta x F(X;\theta)$$

Knowledge Representation and Representation Learning

- Knowledge is stored in “syntactic” space. Humans tend to think in semantic space, i.e., in terms of the meaning.
- Representation learning and encodings are necessary for
  - Multimodal AI
  - Transfer learning
- Understanding and learning how to represent knowledge is the key to human-computer interactions (HCI)
  - Explainability
  - “Document” summarization
  - Causal inference
  - Robustness
  - Validation
- Focus on graph-based representations

Confidence: Uncertainty Quantification

- An AI is simply a model and has uncertainty. UQ can also enhance explainability and support decision making
- Types of Uncertainty
  - Data
  - Model form
  - Learning
  - "Semantic"

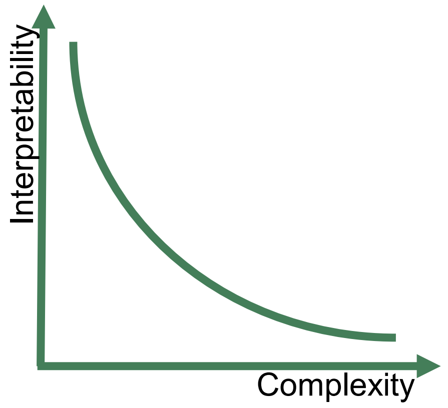


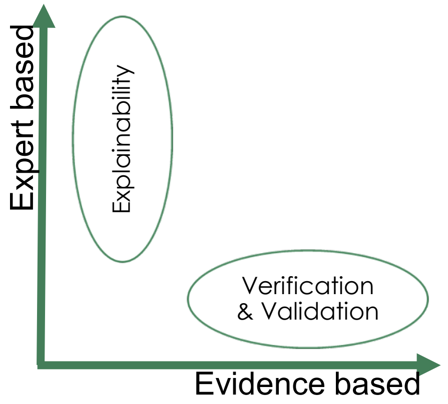
- Bayesian methods/variational inference common, but are computationally expensive

Oak Ridge National Laboratory

Confidence: Validation, Explainability and Reproducibility

- Validation
  - Is the model (including training data) appropriate for the decisions being made?
  - Must be evidence based
  - Requires some form of UQ, robustness guarantees and bounds on "distortion"
- Explainability
  - Can the model present a sequence of steps that can justify the answer to an expert?
  - Expert based
  - Depends on representations
- Reproducibility
  - Does the same experiment lead to the same conclusion?
  - Can we run different experiments and not contradict our conclusion?
  - If we create a new model with the same data, do we get the same conclusions?





OAK RIDGE
National Laboratory

Observations and Issues That Drive Our Research Investments

- Data quantity is easy, data quality is hard
  - Access to and availability of "good" and "labeled" data is one of the biggest challenges for AI
  - We need a sustainable data infrastructure and data policies
- Validation, validation, validation and explainability
- Resilience and robustness are critical. Highly optimized systems are often the most sensitive to disruption
- Vulnerability threats for AI (hacking, intentional manipulation) are a huge concern for deployment
- AI may change how we do science, but it can also cover up bad science
- We need to understand and address the ethics and human impact of AI
- Scalability is important, but it needs to be put into the context of the entire workflow
- We need to manage expectations
- HCI is critical
- AI is an art

Join the ORNL Artificial Intelligence Team.

Oak Ridge National Laboratory is currently seeking researchers and leaders in Artificial Intelligence (AI) and Machine Learning (ML) and hiring entry-level and postdoctoral positions.

ORNL's AI & ML Initiative harnesses the Laboratory's unique suite of expertise, compute capabilities, and facilities to solve some of the nation's most daunting challenges in R&D and national security.

Specifically, ORNL offers . . .

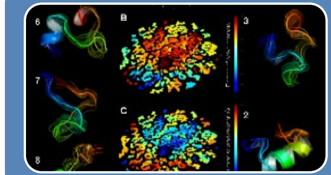
Leadership in high-performance computing: The Laboratory's IBM system, dubbed Summit, is currently the most powerful and "smartest" computer in the world with a tensor core GPU architecture that makes it ideal for rapidly developing and training scientific artificial intelligence and machine learning networks.

Unparalleled scientific facilities and datasets: ORNL hosts some of the world's most advanced scientific facilities and instruments, including the National Transportation Research Center, the Spallation Neutron Source, and the Manufacturing Demonstration Facility, to name a few. These assets generate the large and unique scientific datasets perfectly suited for AI analysis.

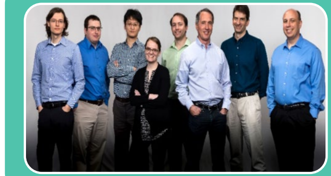
Leadership in scientific disciplines: As a Department of Energy laboratory, ORNL staff drive innovation in additive manufacturing, neutron science, national security, materials science, and artificial intelligence applications for biology and medicine.

Harnessing these assets in concert, however, requires interdisciplinary research teams and a collaborative environment such as the one enabled by ORNL's diverse scientific research programs.

Join ORNL's AI Summer Institute!

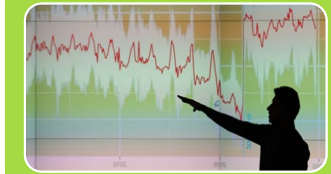


Diverse teams of students who will collaborate to solve real scientific challenges



Collaboration with leading scientists and mentors who have expertise in AI and ML or domain sciences such as materials or physics

Educational seminars, professional development, and career opportunities



Opportunities to present work to and interact with ORNL's scientists in an open forum

Opportunities to work on the world's smartest supercomputers

Requirements:

- US Citizen or LPR
- Rising junior, senior, Master's or PhD
- Prior experience in machine learning and artificial intelligence, data science and analytics, statistics and mathematics, high performance computing, and/or visualization and visual analytics

Apply now at
ai.ornl.gov/summerinstitute