U.S. DEPARTMENT OF ENERGY | Office of Science

Scientific Artificial Intelligence and Machine Learning

Workshop on Predictive Models and High Performance Computing as Tools to Accelerate the Scale up of New Bio Based Fuels

Workshop Dates: June 9-11, 2020

Steven Lee, PhD
Office of Advanced Scientific Computing Research
Program Manager for Applied Mathematics and AI
Steven.Lee@science.doe.gov

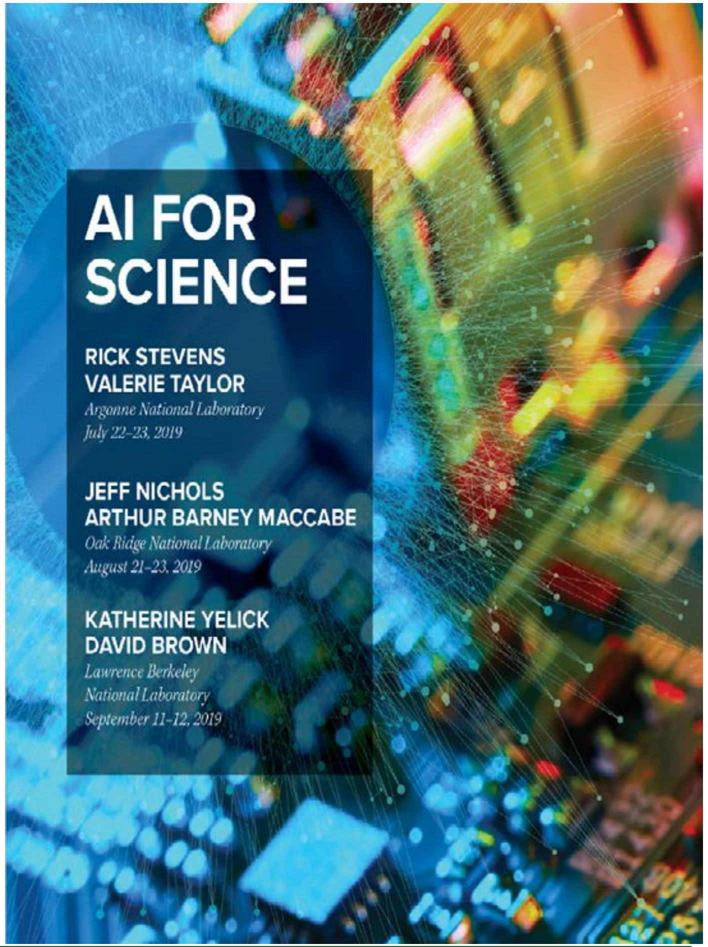
Scientific Artificial Intelligence and Machine Learning - Overview

AI and Machine Learning for Science (AI/ML) broadly refers to the development and use of

AI-enabled facilities, infrastructure, and technologies for transforming science and energy research:

- ❑ Accelerate scientific discovery,
- ❑ Increase scientific competitiveness, and
- ❑ Create innovative scientific and operational capabilities

Examples: AI-assisted science, Autonomous experiments



Trends in AI, High-Performance Computing (HPC), and Scientific Data

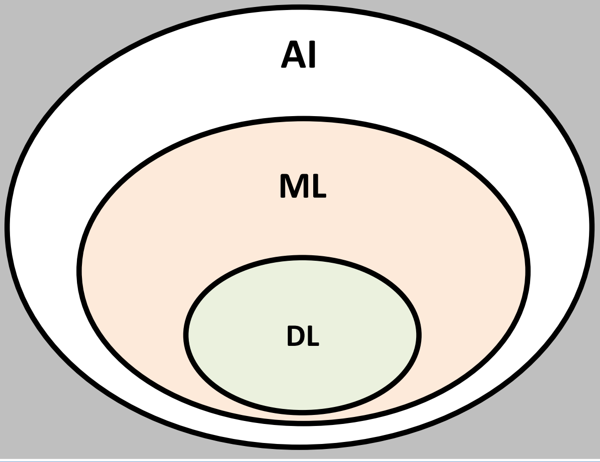
- AI is not a standalone technology
- Near term: Pairwise approaches for HPC-Data, Data-AI, AI-HPC
- Long term: Integrated strategy for the convergence of AI-HPC-Data

Scientific AI/ML will develop & build on DOE's unique strengths & resources.

https://www.osti.gov/biblio/1604756/

AI, Machine Learning, Deep Learning in a Nutshell

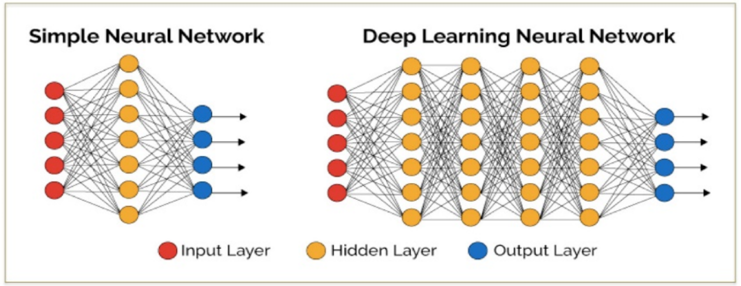
Artificial Intelligence (AI)
How can computational technologies be developed & used to assist, augment or automate human skills?



Machine Learning (ML) - Foundational basis for AI.

ML algorithms make predictions, decisions, & estimates from building a mathematical model or "learning" based on probabilities, samples, or training data.

Deep Learning (DL) - Includes neural network-trained approaches for tasks such as spam filtering, fraud/anomaly detection, image analysis.

Cognitive Skills (AI)	Specific Tasks (ML)	Neural Networks (DL)
Vision & Perception	Classification	
Natural Language Processing	Clustering & regression	
Search & Planning	Simplified or surrogate models	
Problem solving	Feature extraction	
Knowledge reasoning	Pattern recognition	

[Workshop Report on Basic Research Needs for Scientific Machine Learning: Core Technologies for Artificial Intelligence]() (DOE/ASCR, 2019)

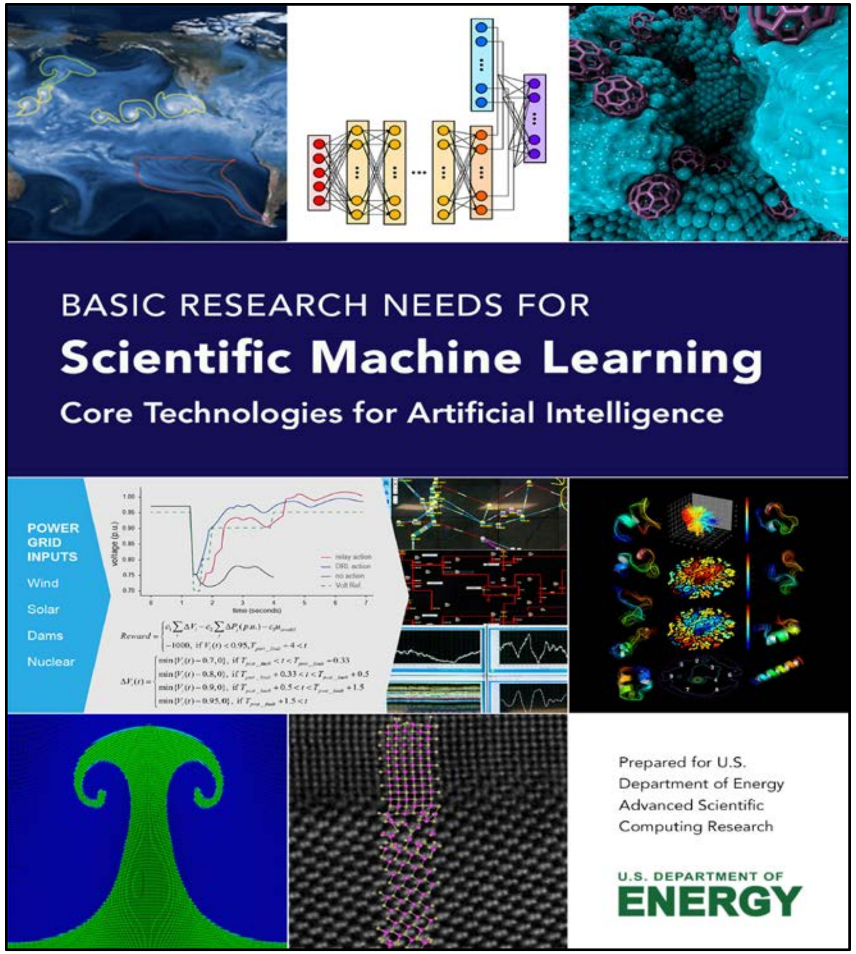
Scientific AI/Machine Learning: Priority Research Needs

Scientific Machine Learning: Foundations

- **Domain-Aware:** Leverages & respects scientific domain knowledge. Physics principles, symmetries, constraints, uncertainties & structure-exploiting models
- **Interpretable**: Explainable and understandable results. Model selection, exploiting structure in high-dimensional data, use of uncertainty quantification with machine learning
- **Robust**: Stable, well-posed & reliable formulations. Probabilistic modeling in ML, quantifying well-posedness, reliable hyperparameter estimation

Scientific Machine Learning: Capabilities

- **Data-Intensive Scientific ML:** Scientific inference & data analysis. ML methods for multimodal data, in situ data analysis & optimally guide data acquisition
- **Machine Learning-Enhanced Simulations:** ML hybrid algorithms & models for predictive scientific computing. ML-enabled adaptive algorithms, parameter tuning & multiscale surrogate models
- **Intelligent Automation and Decision Support:** Adaptivity, automation, resilience, control. Exploration of decision space with ML, ML-based resource management, optimal decisions for complex systems



January 2019

Advances in 6 Priority Research Directions (PRDs) are needed to develop the next generation of machine learning methods and artificial intelligence capabilities.

U.S. DEPARTMENT OF ENERGY | Office of Science

https://www.osti.gov/biblio/1478744

Science Highlight: AI Joins the Team for Smarter & Faster Experiments

X-ray scattering beamline experiment

- X-rays from each experiment probe a portion of the material specimen
- AI launches best sequence of experiments for revealing the internal structure

Experiment Team & Resources

Autonomous experiments

- DOE Lab researchers: Brookhaven (BNL), Lawrence Berkeley (LBNL)
- BNL facilities & data: Light source (NSLS-II), Nanoscale science (CFN)
- AI & models: LBNL Center for Advanced Mathematics in Energy Applications

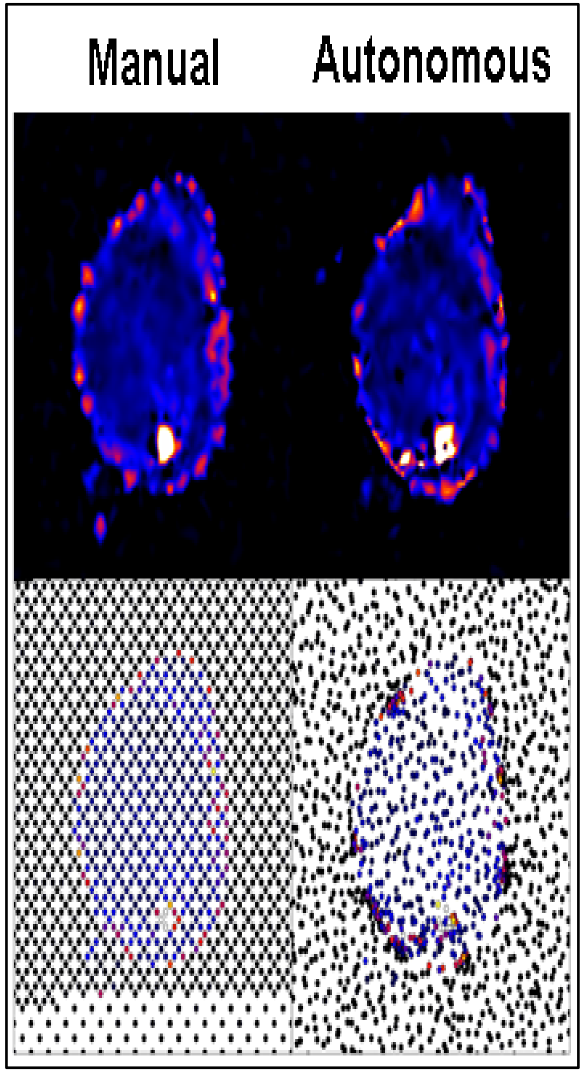
Key Objective: Created an AI ecosystem, a seamlessly integrated suite of resources (AI, Computing, Data, Facilities) for scientific

✓ Discovery – 6x faster

✓ Competitiveness – more accurate

✓ Innovation – smart, self-driving experiments.

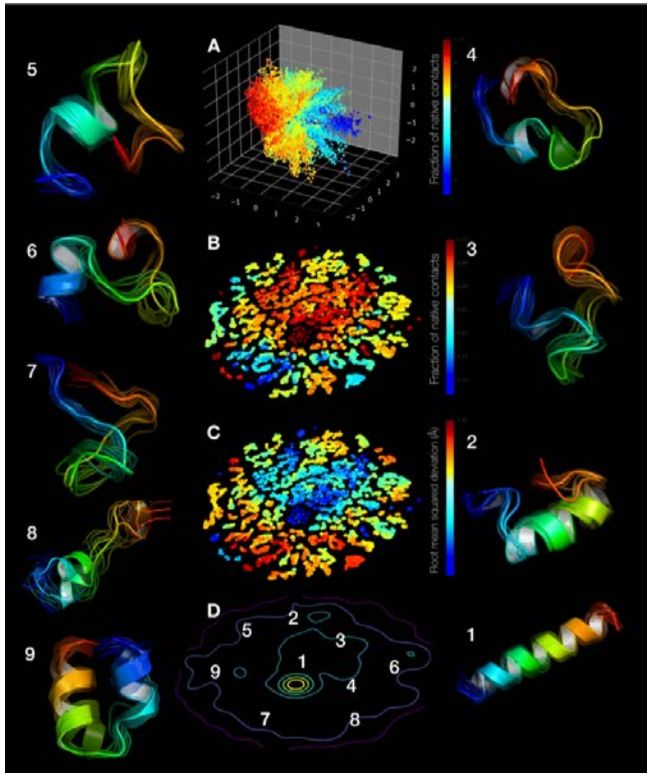
How do we enable other AI ecosystems for science?



PRD #4: **Data-Intensive** Scientific Machine Learning
Automated Scientific Inference and Data Analysis

Key Points: What novel approaches can be developed for reliably finding signals, patterns or structure within high-dimensional, noisy, uncertain input data?

- Scientific ML methods require the development of improved methods for statistical learning in high-dimensional Scientific ML systems with noisy and complex data
- Need approaches to identify structure in complex high-dimensional data
- Scientific ML requires efficient sampling in high-dimensional parametric and model spaces



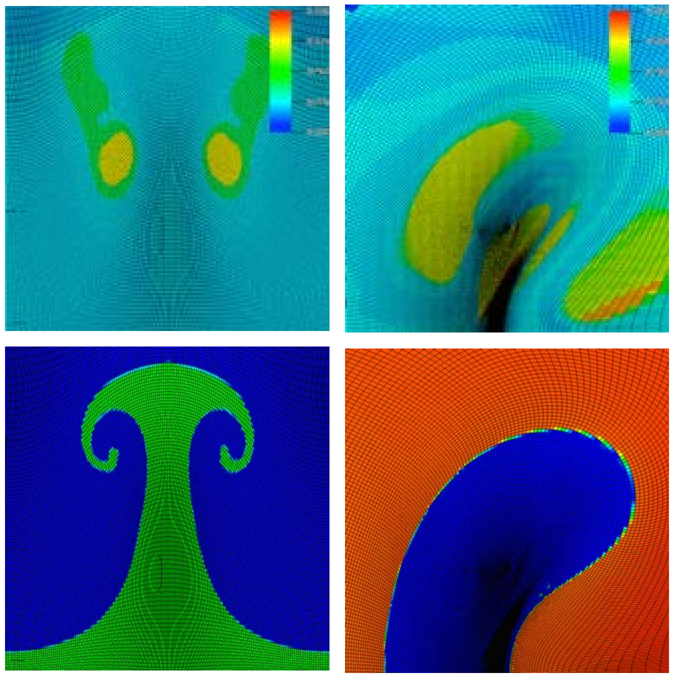
ML techniques reveal Fs-peptide folding events from long time-scale molecular dynamics simulations. A low dimensional embedding of the simulation events reveal transitions from fully unfolded states (blue) to fully folded states (red). A two dimensional embedding using t-test stochastic neighborhood embedding shows the presence of near native states (labeled state 1) versus partially unfolded (2-7) and fully unfolded states (8-9) in the picture.
Image Credit: Arvind Ramanathan, ORNL.

PRD #5: Machine Learning-Enhanced Models and Simulations

Predictive Scientific Computing

Key Points: What is the role and potential advantages of ML-embedded approaches in computational model and algorithm development?

- Combination of scientific computing with learned adaptivity for more efficient simulations
- ML for in-situ parameter tuning
- ML for sub-grid physics models
- Progress will require the development of new methods to quantify tradeoffs and optimally manage the interplay between traditional and ML models and implementations



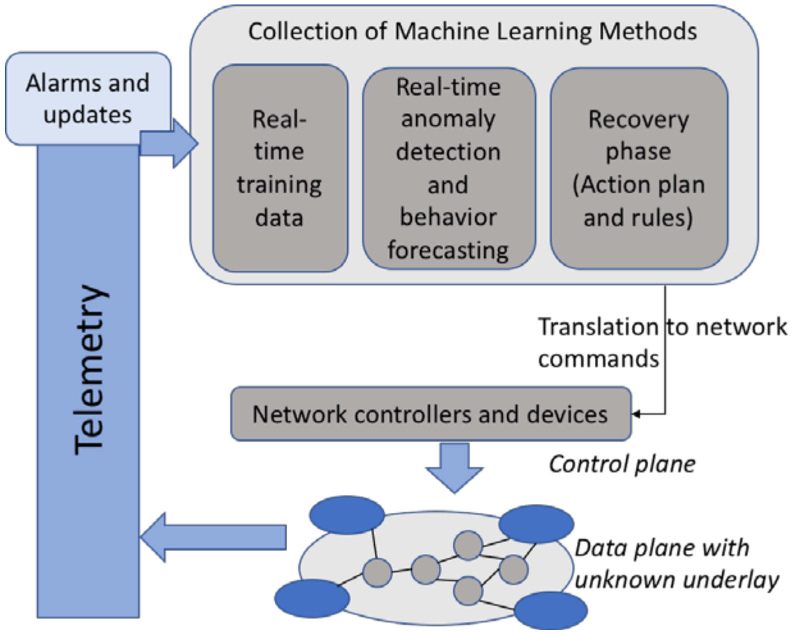
The arbitrary Lagrangian-Eulerian (ALE) method is used in a variety of engineering and scientific applications for enabling multi-physics simulations. Unfortunately, the ALE method can suffer from simulation failures, such as mesh tangling, that require users to adjust parameters throughout a simulation just to reach completion. A supervised ML framework for predicting conditions leading to ALE simulation failures was developed and integrated into a production ALE code for modeling high energy density physics.
Image credit: M. Jiang, LLNL.

PRD #6: Intelligent Automation and Decision Support

Management and Control of Complex Processes and Systems

Key Points: What are the challenges in managing the interplay between automation & human decision-making?

- Outer-Loop applications include optimization, uncertainty quantification, inverse problems, data assimilation, & control.
- New mathematically & scientifically justified methods to guide data acquisition and ensure data quality and adequacy.
- Scientific ML methods for improving system resilience or responsiveness.



Exascale applications are exponentially raising demands from underlying DOE networks such as traffic management, operation scale and reliability constraints. Networks are the backbone to complex science workflows ensuring data is delivered securely and on-time for important compute to happen. In order to intelligently manage multiple network paths, various tasks such as pre-computation and prediction are needed to be done in near-real-time. ML provides a collection of algorithms that can add autonomy and assist in decision making to support key facility goals, without increased device costs and inefficiency. In particular, ML can be used to predict potential anomalies in current traffic patterns and raise alerts before network faults develop. Image credit: Prabhat, LBNL.

DOE Scientific AI and Machine Learning

Foundational Research in Applied Mathematics & Scientific Computing

Capability Themes	Relevant Funding Announcements since 2005
Data-Intensive Scientific Machine Learning	2009 – 2012: Mathematics for Analysis of Petascale Data 2009 – 2012: Joint Mathematics Computer Science Institute 2012 – 2015: Resilient Extreme-Scale Solvers 2013 – 2016: DOE Data-Centric Science at Scale
Machine Learning-Enhanced Scientific Modeling and Simulations	2005 – 2008: Multiscale Mathematics Research and Education 2008 – 2011: Multiscale Mathematics for Complex Systems 2013 – 2016: Uncertainty Quantification (UQ) for Extreme-Scale Science 2019 – 2021: UQ for Scientific Machine Learning & Artificial Intelligence **2020 – 2022: Scientific ML for Modeling and Simulations**
Intelligent Automation and Decision Support for Complex Systems	2009 – 2012: Mathematics for Complex, Interconnected Systems 2010 – 2013: Uncertainty Quantification (UQ) for Complex Systems 2012 – 2017: Mathematical Multifaceted Integrated Capability Centers I 2017 – 2022: Mathematical Multifaceted Integrated Capability Centers II **2020 – 2023: AI and Decision-Support for Complex Systems**

AI/ML for Science and Energy – Snapshot of the Future

- Artificial Intelligence and Machine Learning are rapidly evolving technologies

- Tremendous national and international interest in AI/ML development and use for science:
  - DOE "AI for Science" Town Halls
  - UK Workshop (Feb 2020) – Posing an AI Scientist Grand Challenge: AI Systems capable of Nobel-Quality Discoveries

- Overwhelming response & new ideas from FY2020 Office of Science FOAs and Lab Announcements

- Office of Science planning and AI-ecosystems approach focuses on DOE's strengths and assets:
  - Facilities, Technology laboratories
  - Massive data, HPC, and networking
  - DOE Labs & talented R&D workforce
  - Science at scale