Oak Ridge National Laboratory

Artificial Intelligence and Machine Learning: Issues and Opportunities

International Conference on Neuromorphic Systems

July 23-26, 2018

Knoxville, TN

David E. Womble

Director of Artificial Intelligence Programs

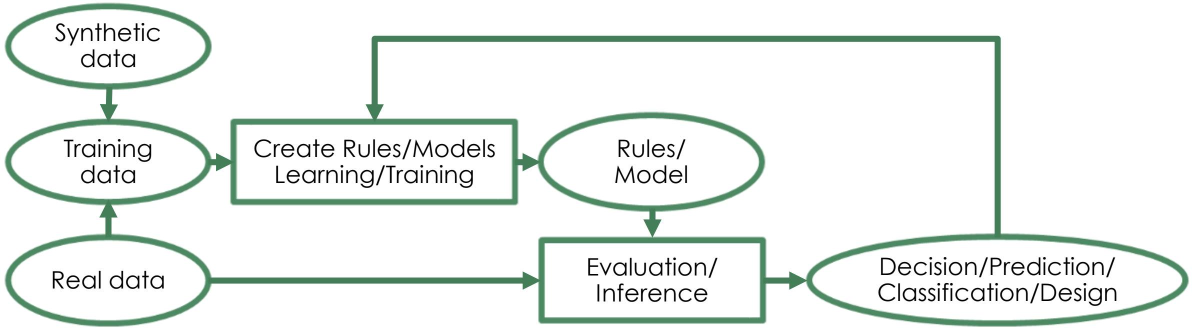
Oak Ridge National Laboratories

What are Artificial Intelligence (AI) and Machine Learning (ML)

- A class of data analytics algorithms in which the rules and/or models are not known a priori and are learned as part of the process
  - Process data to identify correlations
  - Complexity of the model is a potential problem

- Computers trained to perform tasks that if performed by a human would be said to require intelligence
  - Knowledge-based tasks
  - Computers are good at working with data, not “meaning”

Are we trying to reduce “knowledge” to an equation?



OAK RIDGE National Laboratory

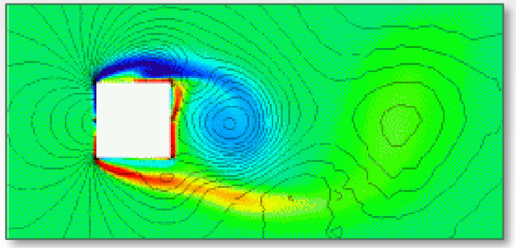
An AI Taxonomy

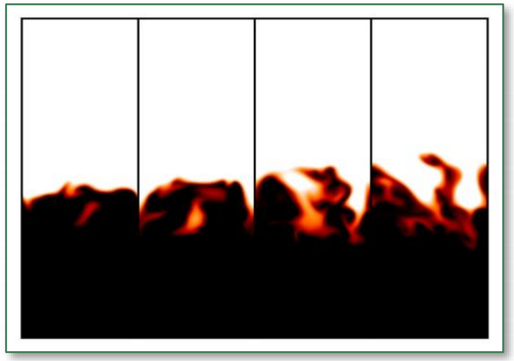
Classification and regression

Near Infrared (single band) WorldView-3 image

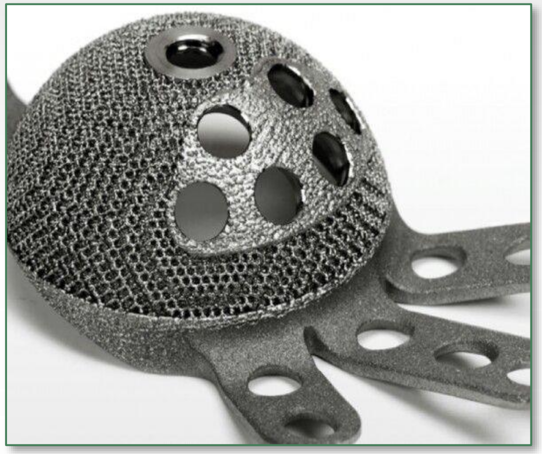
CODA cloud detection saliency map for image above

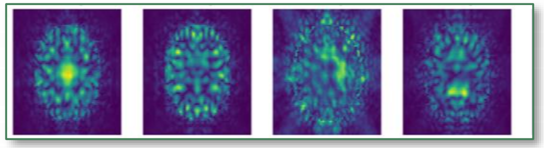
Surrogates





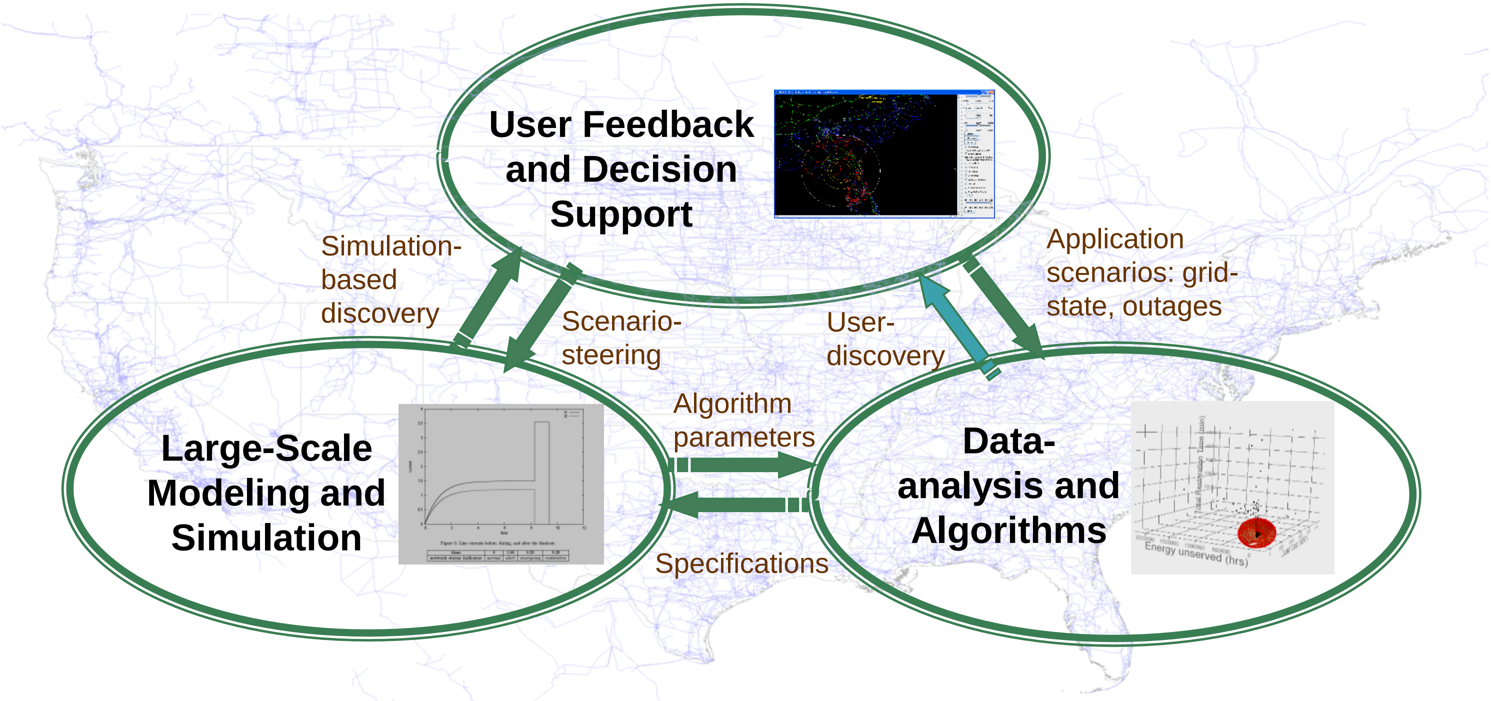
Inverse problems, including design and optimization





Control systems

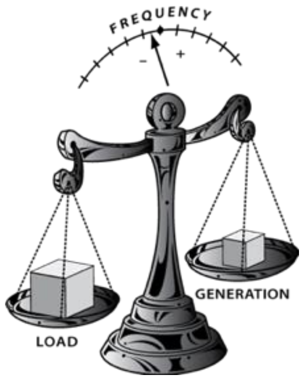
Applications at ORNL: Smart Grid



- Learning algorithms for wide-area, hierarchical information sources
  - Distribution: Intelligent loads, SCADA devices, DERs
  - Transmission: Protection systems, power flow control
  - Generation: Planning and coordination
  - Control: Situational awareness, fine-grained control of DERs, enhanced reliability and resilience

Smart Grid: Challenges

Application challenges	• Integrating variable distributed energy resources (DERs) with intelligent interfaces • Integrating storage at multiple layers • Integrating electric vehicles (EV) • Managing demand – Residential, Commercial, Industrial – Enabling energy coordination and trading between buildings and trading between buildings and grid
Technology challenges	• Connectivity across DERs • Scalable control and diagnostics algorithms that are driven by data • Actionable, real-time situational awareness • Data and physical system security, including privacy

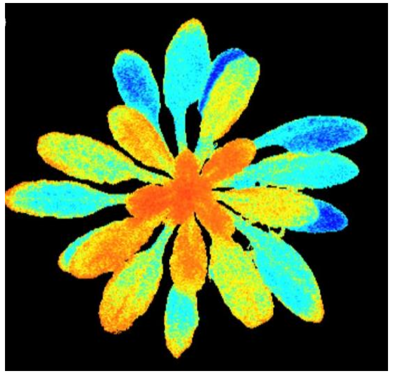


OAK RIDGE
National Laboratory

Applications at ORNL: Biosciences

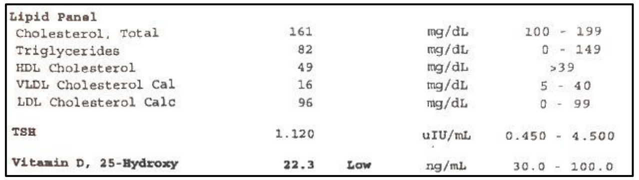
Image analysis

The Cancer Imaging Archive

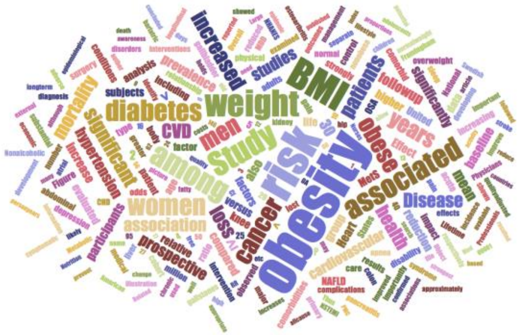


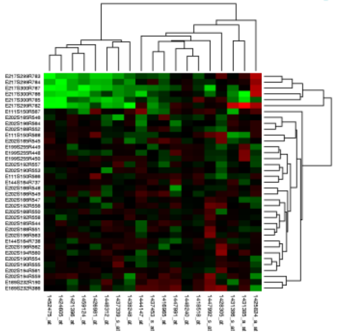
Phil Trans R Soc London B **363** (1491), 639-658

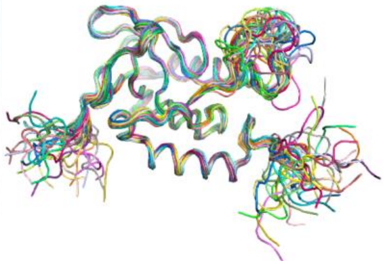
Multimodal data

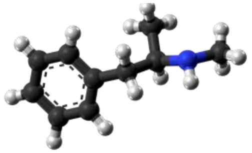


Lipid Panel				
Cholesterol, Total	161		mg/dL	100 - 199
Triglycerides	82		mg/dL	0 - 149
HDL Cholesterol	49		mg/dL	>39
VLDL Cholesterol Cal	16		mg/dL	5 - 40
LDL Cholesterol Calc	96		mg/dL	0 - 99
TSH	1.120		uIU/mL	0.450 - 4.500
Vitamin D, 25-Hydroxy	22.3	Low	ng/mL	30.0 - 100.0









Complex traits

Human health

- Cancer
- Heart disease
- Suicide
- Addiction
- Interventions

Plant science

- Drought tolerance
- Disease resistance
- Biomass yield
- Carbon fixation

Applications at ORNL: Biosciences

Massive multi-modal data sets

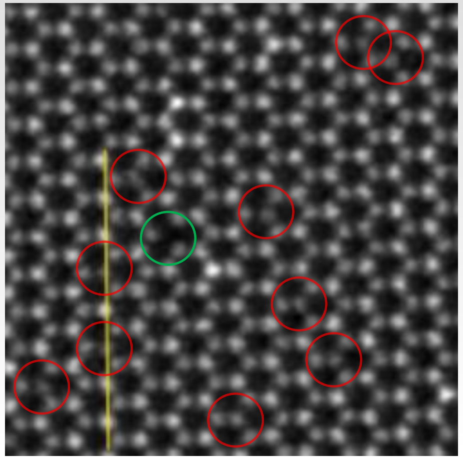
- Million Veterans Program health records
  - Clinical records, genomic data
- Plant phenotyping
  - Imaging data, genomic data
- Genes to cells model
  - ExaLearn, physicochemical data

Data to mechanism

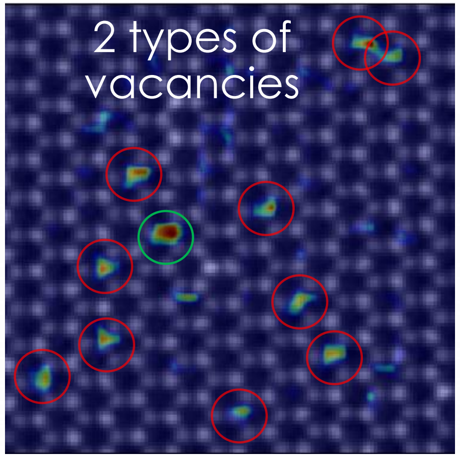
- Experimental/computational feedback loops
  - Hypothesis testing
  - Uncertainty quantification
  - Experimental design
- Explainable AI
  - Mechanistic understanding
  - Hypothesis generation

Applications at ORNL: Materials science

Automated defect classification

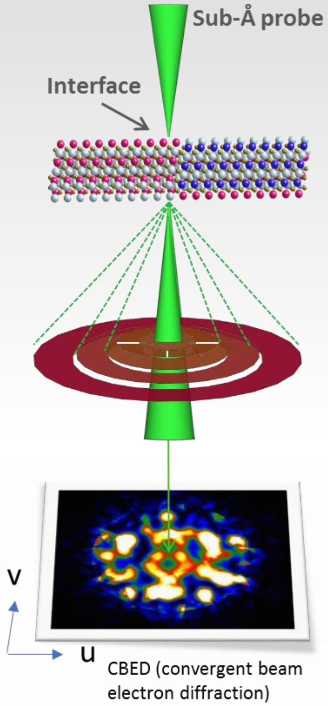


Experimental image

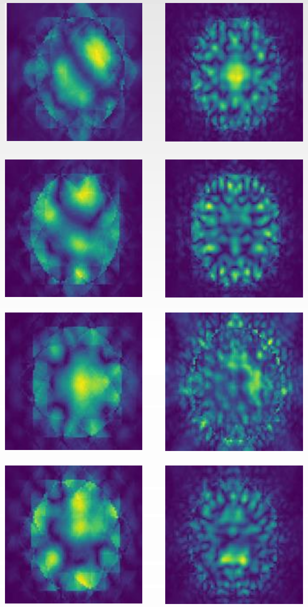


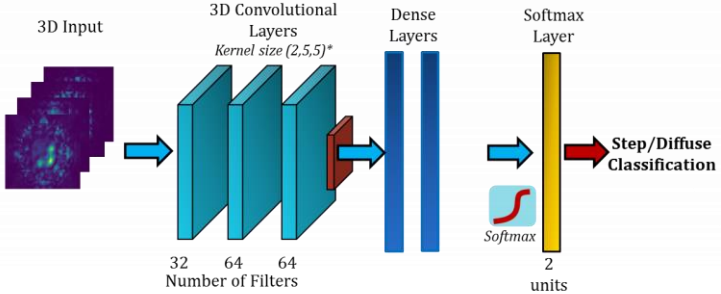
Model output

Inverse problems in imaging

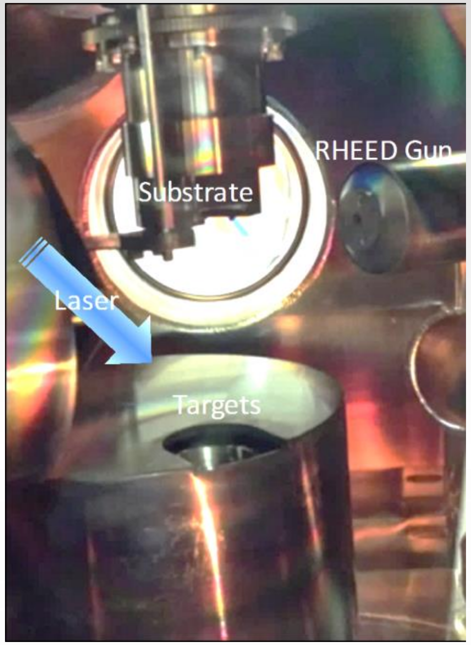


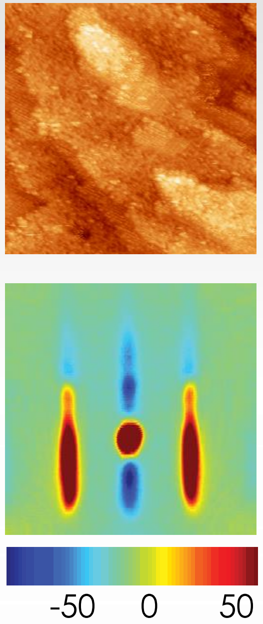
Interface structure classification





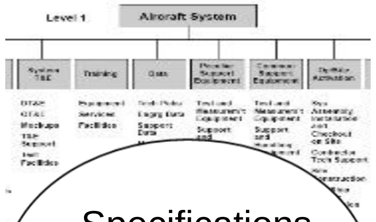
Pulsed laser deposition



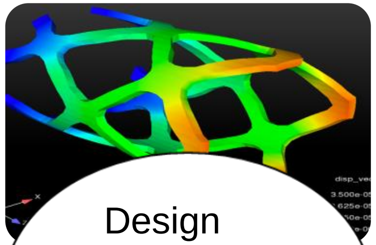


- Design and control of thin-film heterostructures
- Prediction of growth conditions and control of trajectory

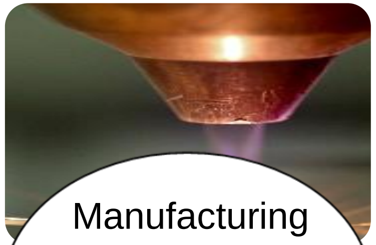
Applications at ORNL: Additive Manufacturing



Specifications
- Functional
- Environmental
- Margins



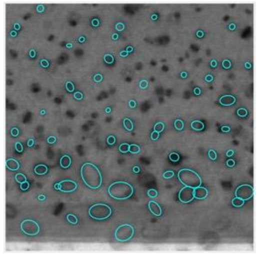
Design
- Shape
- Topology
- Material
- Process

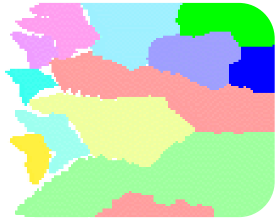


Manufacturing
- RT controls
  - Environment
  - Process
- Diagnostics

Testing
- Validated Process
- Test design

- Impact of machine learning
  - Surrogate models
  - Steering high-fidelity simulation
  - Design, particularly materials and processes
  - Real time diagnostics and control during manufacturing
    - Defect detection and mitigation
    - Control of local structures
  - Predicted performance based on manufacturing data
  - Test design and control



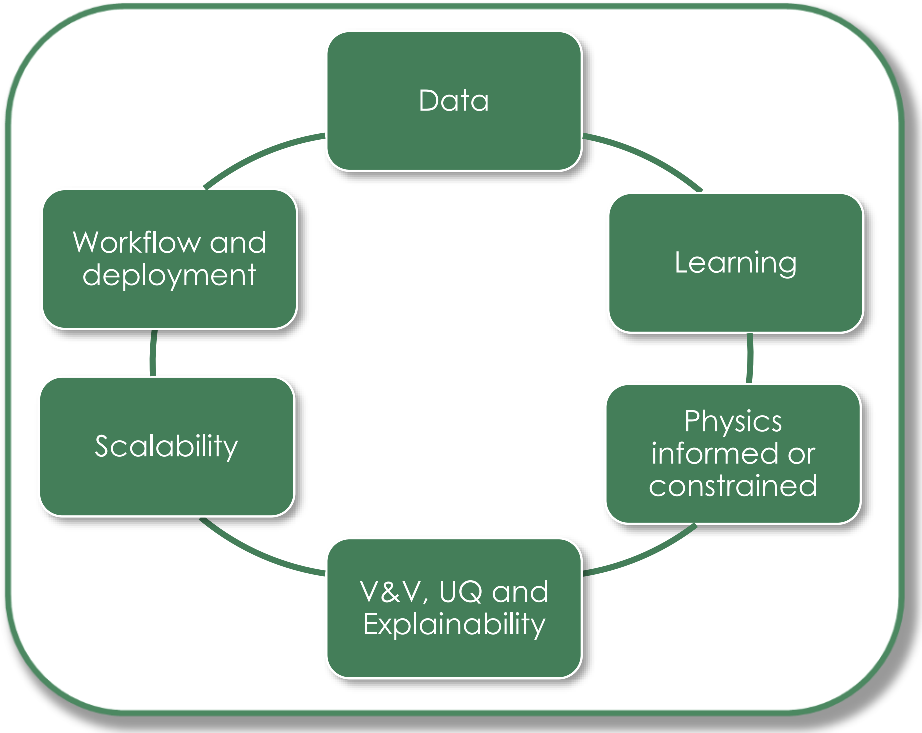


Observations to start a conversation

- Observations and issues
  - AI is effective at for narrowly defined tasks and only identifies correlations in (complex) data
  - Access to and availability of “good” and “labelled” data is one of the biggest challenges for AI
  - We need a sustainable data and compute infrastructure
  - While we have big machines, we don’t have scalable algorithms
  - Vulnerability threats for AI (hacking, intentional manipulation) are a huge concern for deployment
  - We don’t know how to do validation
  - HCI is an important component of the workflow, including explainability and interpretability
  - Deployment of AIs introduces a whole new set of challenges
  - Need to understand the ethics and human impact
  - AI is not synonymous with deep neural networks

Six Research Areas To Be Addressed

- Data quality and statistics
  - Even if we have enough data, it is not necessarily good data
  - Dealing with bias
- Machine learning
  - Needs to accelerate
  - Very model dependent
- Merging physics and AI
  - We can't violate the laws of physics
- Verification, validation and explainability
  - Is the answer right, is the model appropriate, and can we understand it
  - What is the human-computer interface
- Computing
  - How do we use "big" computers
  - How do we use accelerated nodes
- Workflow and deployment
  - Computing at the edge
  - privacy, ethics and regulations



Issue: “Syntactic” Space vs. “Semantic” Space

- Humans tend to think in semantic space, i.e., in terms of the meaning.

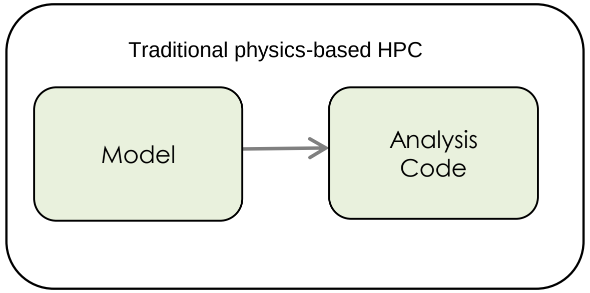

And metrics in semantic space are fundamentally different from those in syntactic space

- Implications
  - Easy to spoof classification systems
  - Transfer learning doesn’t map well.

    (Humans tend to transfer learning in semantic space, e.g., transfer what I learned about human behavior in kindergarten to how I drive. Most AI approaches transfer in syntactic space or transfer parts of the model (a sort of “gene transfer”).

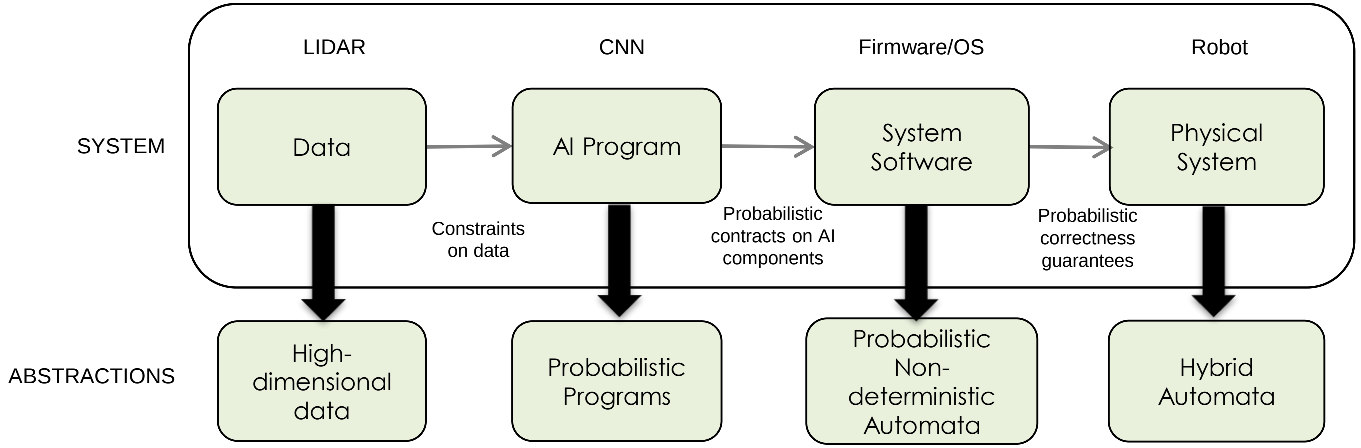
Issue: Verification, Validation, Explainability and Interpretability

- Verification
  - Is the model implemented correctly?
- 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”



Issue: Verification, Validation, Explainability and Interpretability

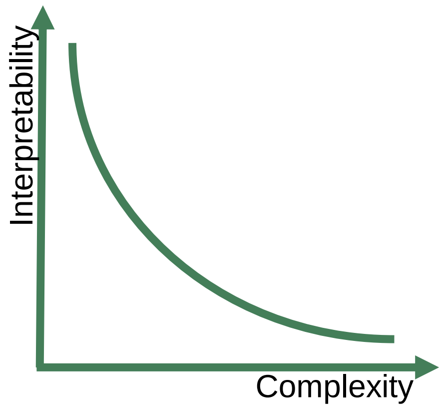
- Verification
  - Is the model implemented correctly?
- Validation
  - Is the model (including training data) appropriate for the decisions being made?
  - Must be evidence based
  - Requires some form of UQ, robustness guarantees and bounds on “distortion”

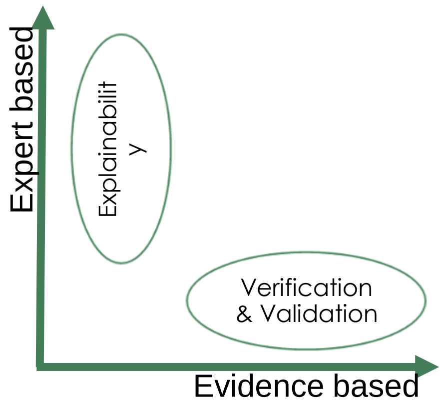


OAK RIDGE National Laboratory

Issue: Verification, Validation, Explainability and Interpretability

- Interpretibility
  - Can a human understand the model? For example, do the basis vectors in a dimension reduction algorithm have a physical meaning?
- Explainability
  - Can the model present a sequence of steps that can justify the answer to an expert?
  - Expert based
- Reproducibility
  - Does the same experiment lead to the same conclusion?
  - Can we run different experiment and not contradict our conclusion?
  - If we create a new model with the same data, do we get the same conclusions?
  - Required for good science
-



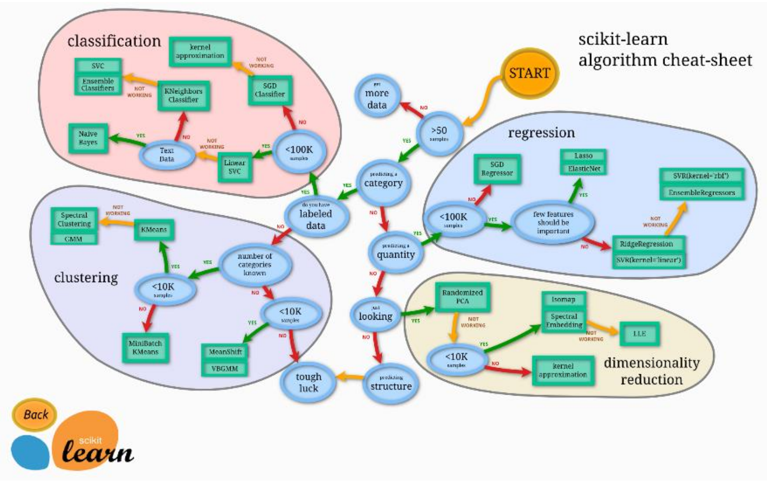


Issue: Big Data Is A Big Problem

- Need more data than was imagined just a few years ago
  - We are looking for complex correlations
  - Using primarily statistical methods
- Labelled data is a problem
  - Generating labels is expensive and labor intensive (e.g., Mechanical Turk)
  - Need to move toward reinforcement learning
- Noisy and missing data is a problem
- Data needs to be curated
- Synthetic data and simulated environments are partial solutions
  - But an AI can learn the flaws in these systems

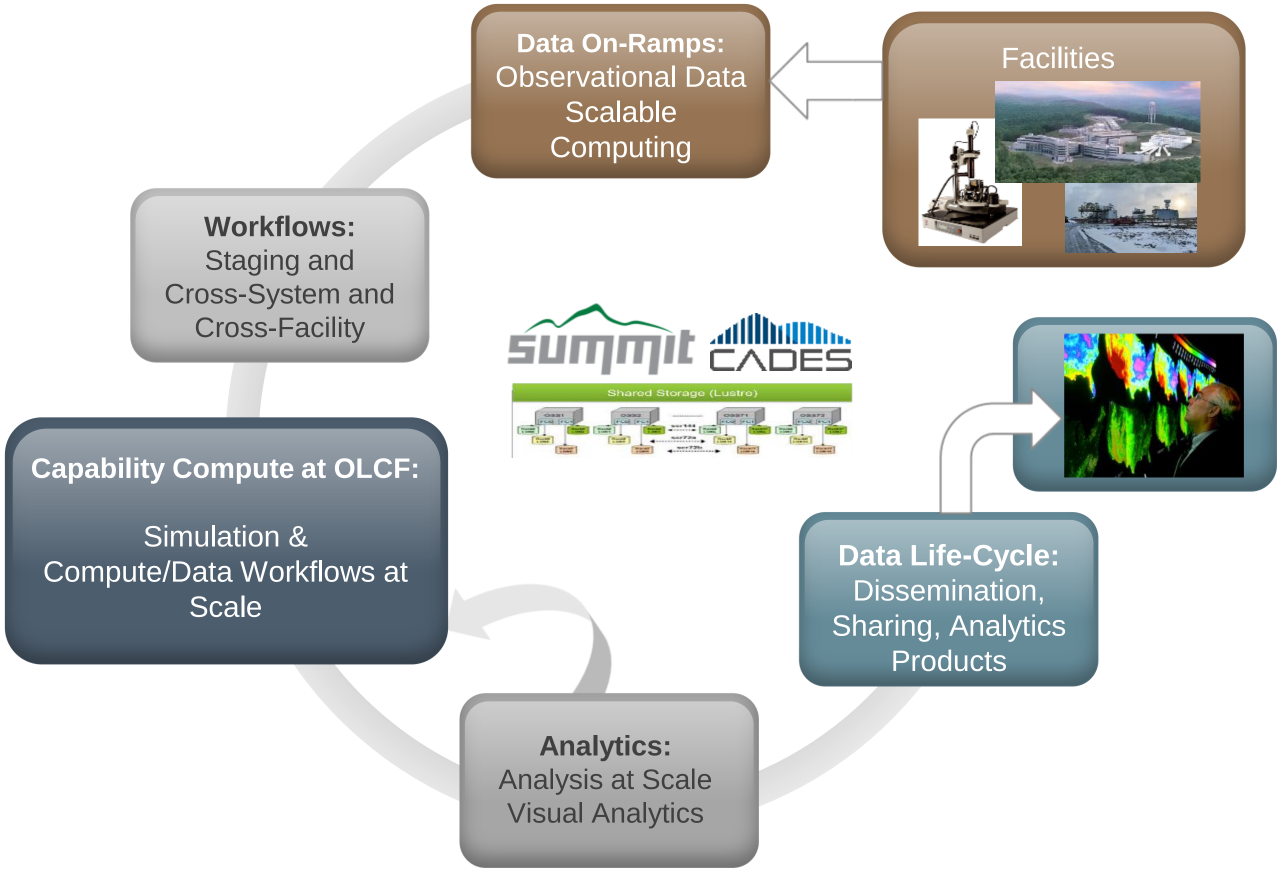
Issue: AI Is An Art

- Choosing the model form and hyper parameters is often ad-hoc and requires experience and insight
- AI models must be tuned
- Neural networks design is difficult and often requires tuning
- Interpreting the results requires expertise



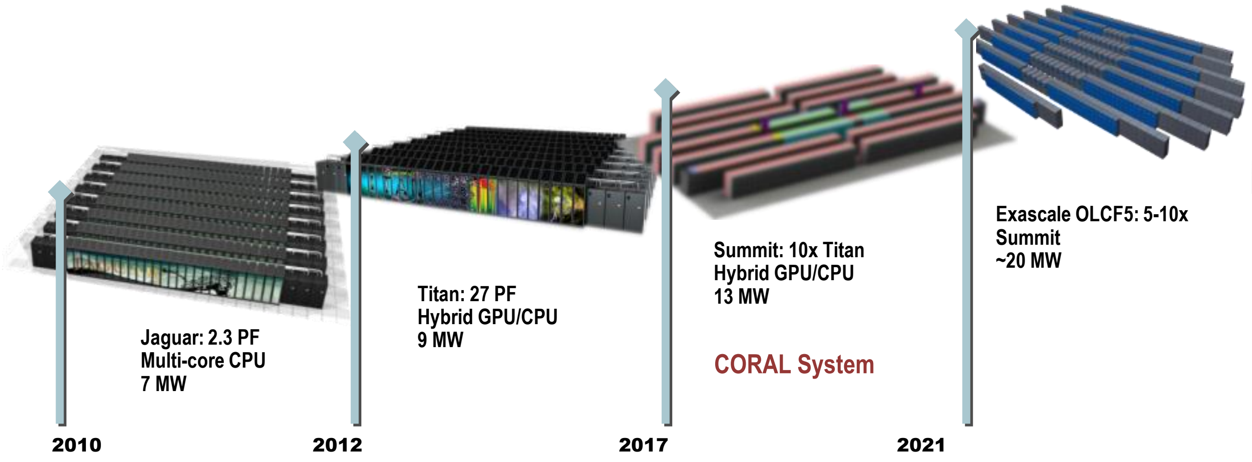
> "Machine learning methods are often described in papers at an abstract level, for maximum generality. However, a good choice of hyperparameters is usually necessary to make them work well on real-world problems, and tricks are often used to make most efficient use of these methods and extend their capabilities."
>
> G. Montrevan, et.al., "Methods for Interpreting and Understanding Deep Neural Networks."

Infrastructure is critical to a good workflow



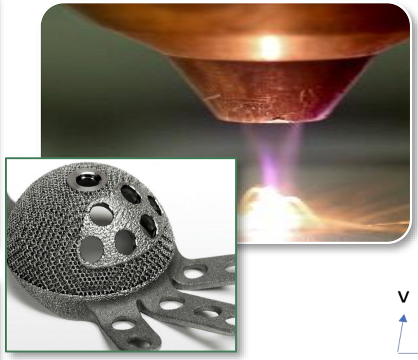
New HPC capabilities enable AI

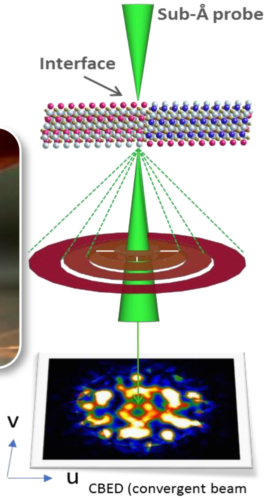
- We have the ability to collect and store large amounts of data
- Computational power continued to increase, with architectural improvements that are amenable to neural networks and AI more generally
  - For example, GPU became practical for accelerated computations.
  - Reduced-precision tensor core units are included



AI is changing the way we do science and engineering

- Science and national security drive the applications at ORNL
  - New materials and processing
  - Autonomy and control
  - Predictive and personalized health systems
  - Energy and the environment
  - Disaster response and security
  - Unique data sets and user facilities
- Research in AI and machine learning establishes the capability
  - Scalability for world-class facilities
  - Innovative learning algorithms
  - Bringing together mathematics, statistics, computer science and data science





CBED (convergent beam electron diffraction)