Artificial Intelligence for Science Townhall Summary

David Womble

Oak Ridge National Laboratory

Oak Ridge National Laboratory

Setting the Context

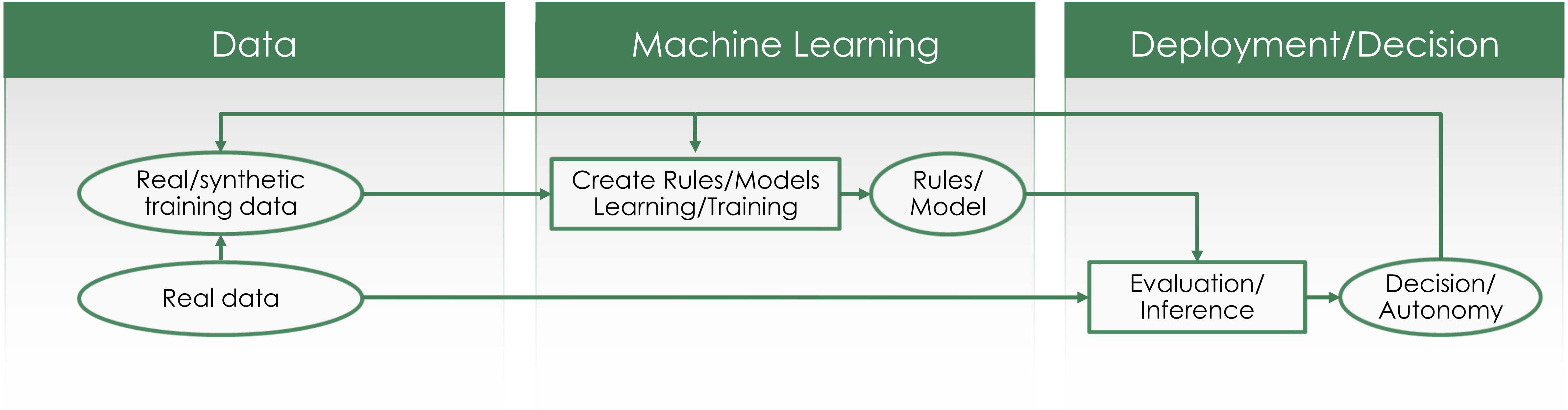
AI won't replace the scientist, but scientists who use AI will replace those who don't.*

*Adapted from a Microsoft report, "The Future Computed"

What is Artificial Intelligence?

Advanced data analytics algorithms in which abstract models are constructed (or learned) from data

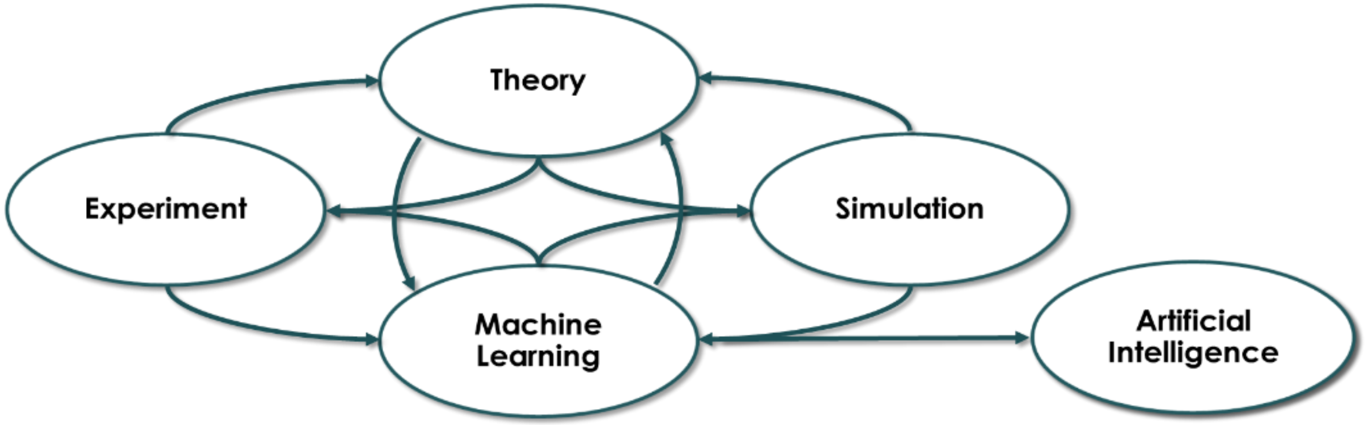
Computers trained to perform tasks that, if performed by humans, would be said to require intelligence



A branch of CS that studies the properties of human intelligence by synthesizing intelligence*

*AAAI

Machine learning requires a strong partnership between the computational scientist and the experimentalist.



OAK RIDGE National Laboratory

We are at a “tipping point” in AI/ML

Data

- Sensors are ubiquitous
- Data is plentiful
- We are “bit-rich”

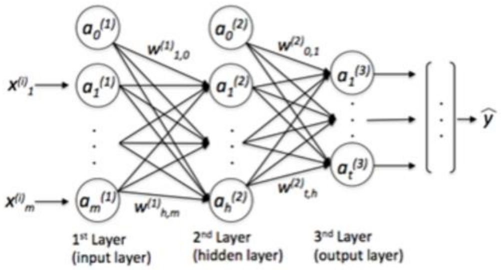
Facilities and data are a distinguishing strength for DOE

Computing

- Computing is “exaflop scale”
- Specialized hardware is being developed for data analytics and “edge” applications

DOE has an HPC mission for science and engineering

Algorithms



- Pre-defined models
- Computationally tractable training for ML

DOE has an HPC mission for science and engineering

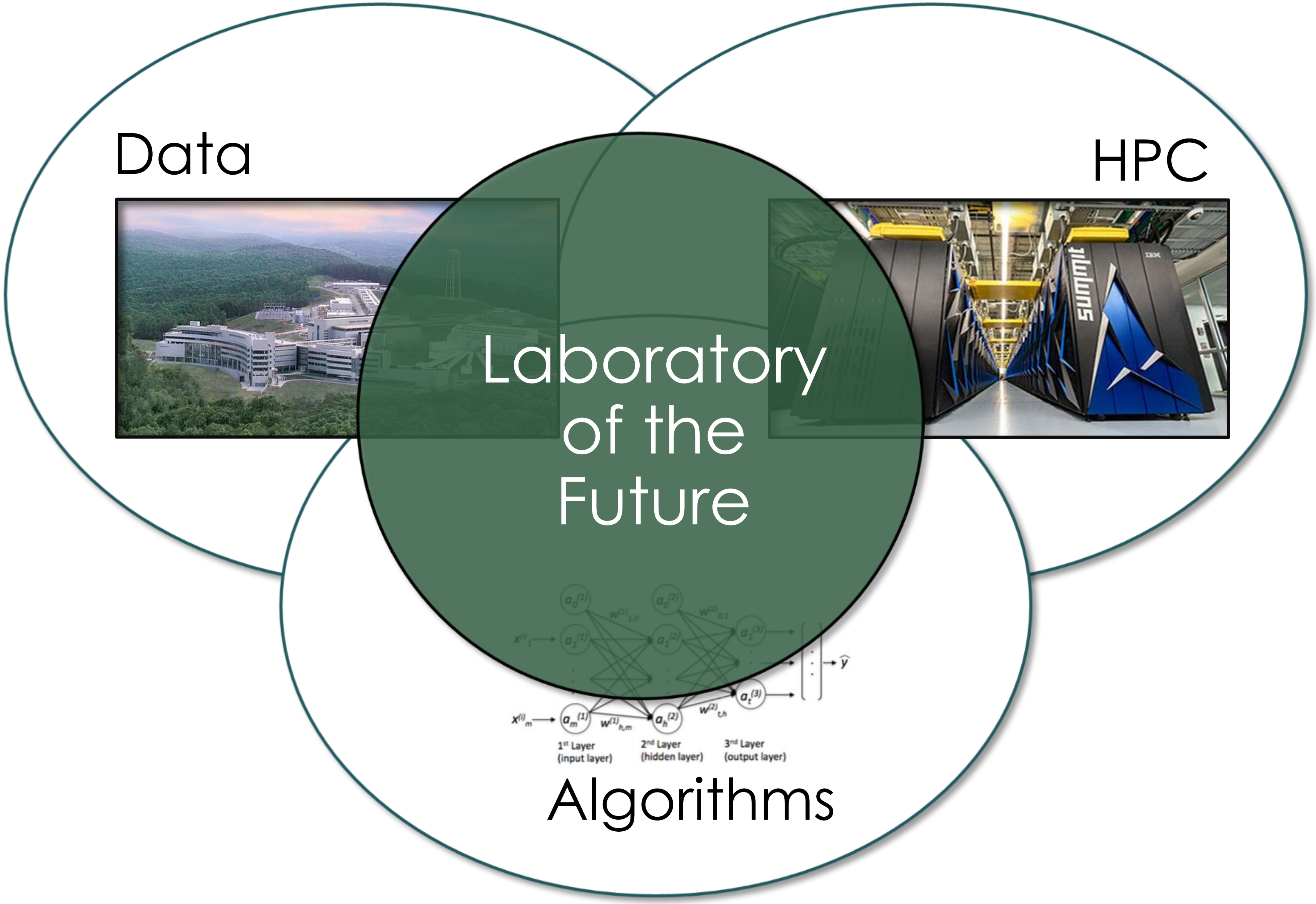
Accessibility

- Everyone has a PC and internet access
- A lot of data is open
- Software is open-source

Assurance is “mission critical”

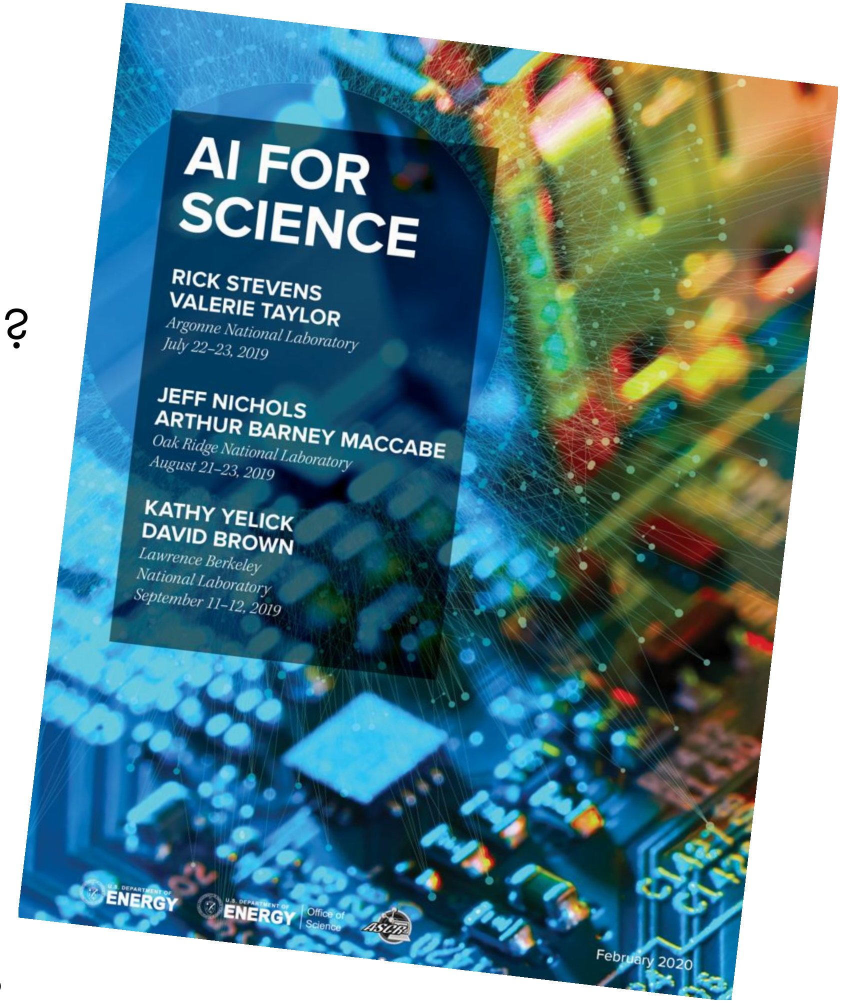
We are bit-rich

and information-poor


Data
HPC
Laboratory of the Future
1st Layer (input layer)
2nd Layer (hidden layer)
3rd Layer (output layer)
Algorithms

AI for Science Townhalls – Focus on Vision

- Meeting objectives
  - Identify transformational uses of AI ∩ HPC
  - Examine scientific opportunities in AI, Big Data, and HPC
  - Lay the groundwork for a program (at the scale of ECP)
- Identify the impact of a sustained push in some domains?
  - Building superhuman capabilities in science
- What scale
  - Big Problems, Big Pushes, Big Data, Big Systems?
  - Fine grain innovation, many thousands of small teams?
- Consider the coupling to experiments, simulations, user and computing facilities?
- What does "scientific production" look like in this space?



AI for Science Town Halls

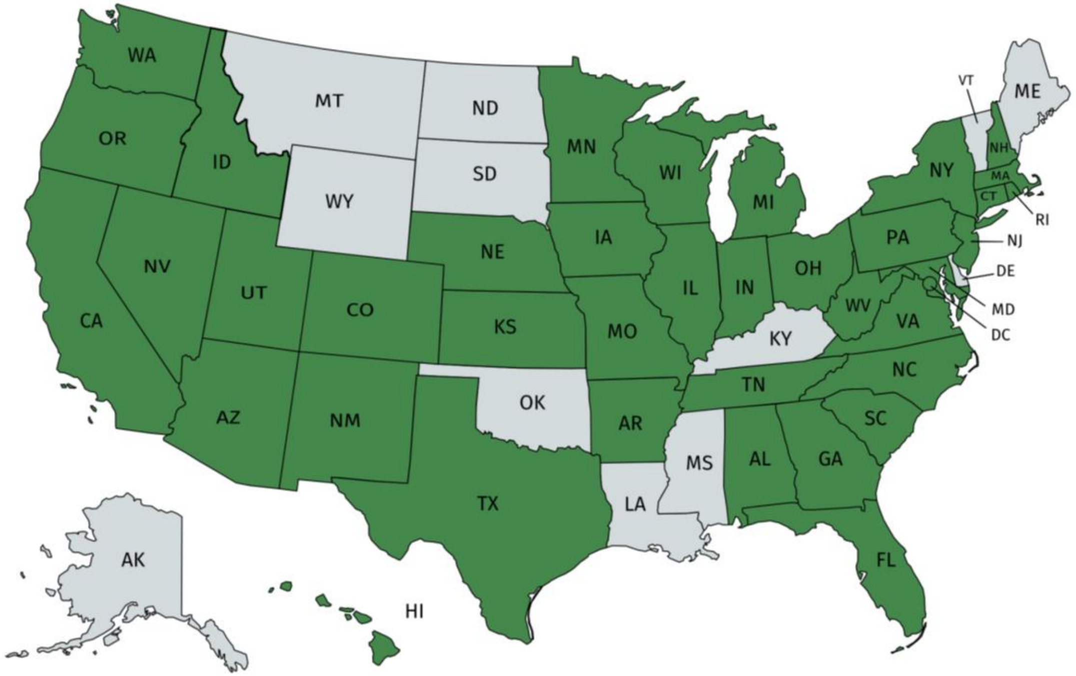
- ~1.5 days each to capture ideas, problems, requirements and challenges for an AI for Science initiative
- Each townhall
  - 1 plenary, 3 keynotes, half-day breakouts on domains, half-day breakouts on crosscuts
  - All breakouts were consistent, with slight tailoring to accommodate what we learned and local influences
- What problems could be attacked?
- What data, simulations, and experiments do we need?
- What kind of methods, software and math do we need?
- What kind of computer architectures and infrastructure do we need?

The AI for Science Town Halls so far

- Over 1000 registrations across 4 Town Halls

ANL	357
ORNL	330
LBNL	349
DC	273
Totals	1309

+100 online
+ ?



- All 17 DOE National Laboratories
- 39 Companies from large and small
- Over 90 different universities
- 6 DOE/SC Offices + EERE and NNSA

An Environmental Scan

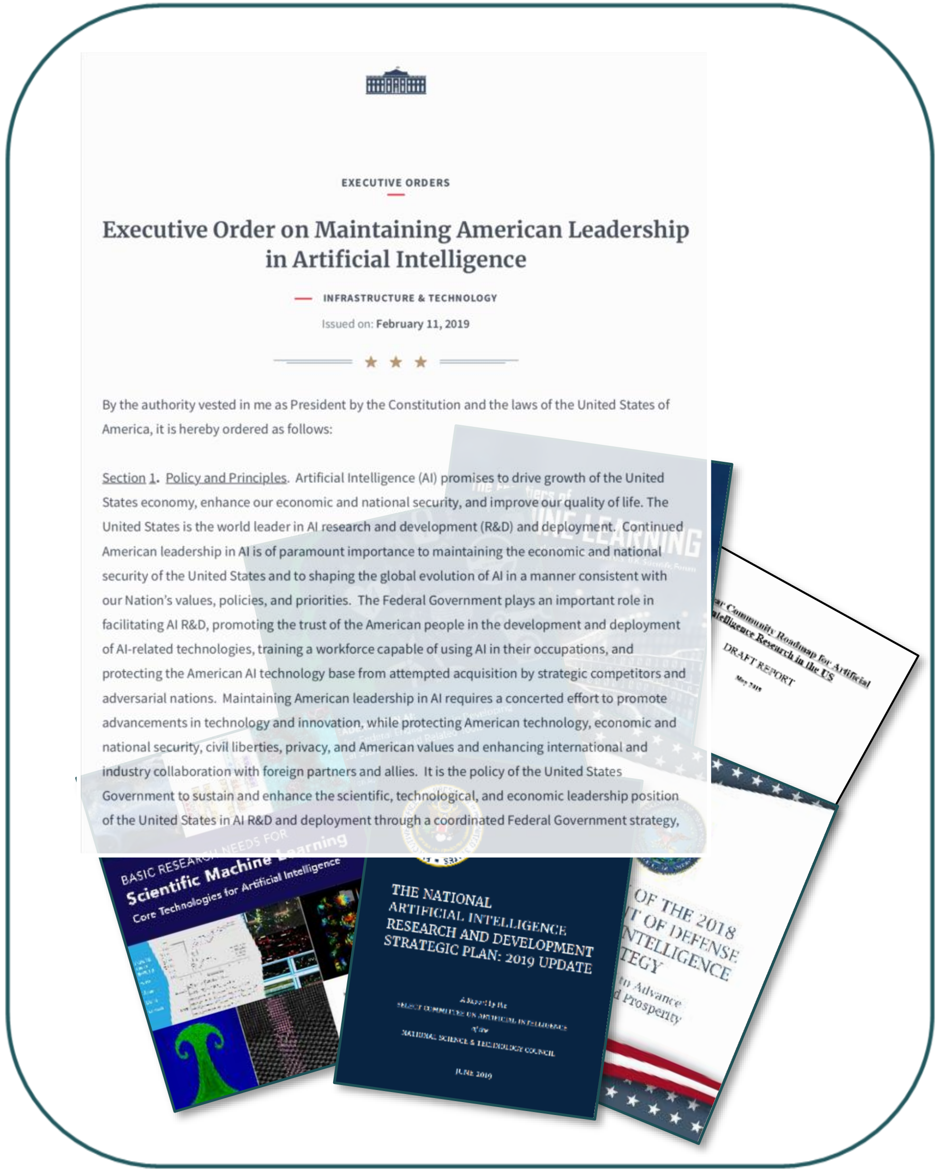
Development and Application of AI Critical For All Government Agencies

- Executive Order on AI

  Policy Statement: Artificial Intelligence (AI) promises to drive growth of the United States economy, enhance our economic and national security, and improve our quality of life.

  … leadership requires a concerted effort to promote advancements in technology and innovation, while protecting American technology, economic and national security, civil liberties, privacy, and American values and enhancing international and industry collaboration with foreign partners and allies.

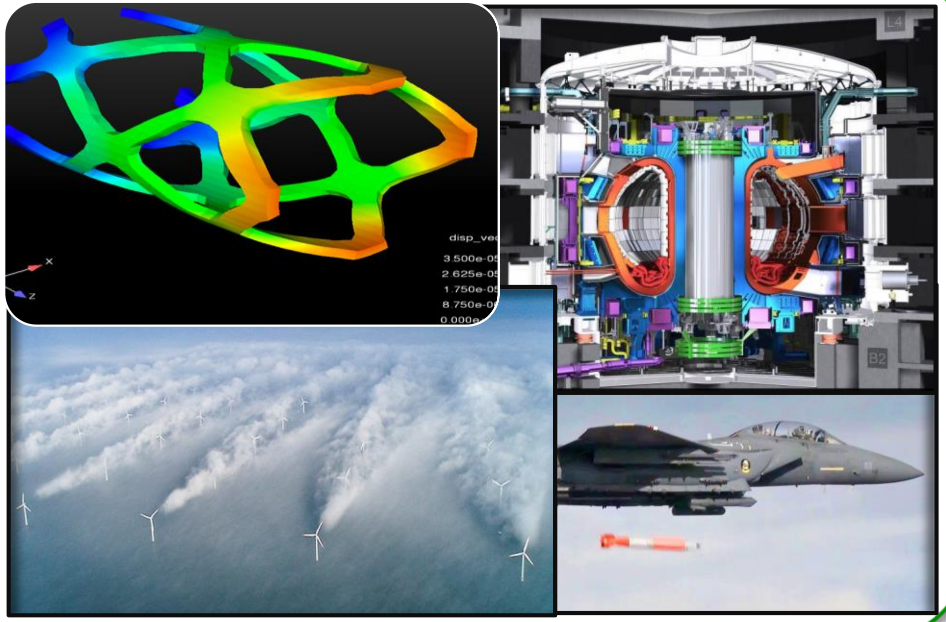
- Supported by multiple agency strategies and programs



OAK RIDGE National Laboratory

DOE builds on historical missions and touches all areas

- The U.S. AI strategy includes
  1. **Long-term investment in research**
  2. Effective methods for human-AI collaboration
  3. Address ethical, legal and social implications
  4. **Ensure the safety and security of AI Systems**
  5. **Develop shared datasets and environments**
  6. Standards and benchmarks
  7. Understand the AI workforce
  8. Expand public-private partnerships
- DOE will play a key role in AI for science and engineering
  - AI Technology office
  - Research and talent development
  - Data to support science and engineering research



DOE's Artificial Intelligence and Technology Office

Secretary Perry Stands Up Office for Artificial Intelligence and Technology

This action has been taken as part of the President's call for a national AI strategy.

SEPTEMBER 6, 2019

VIEW ARTICLE

Vision:

Transform DOE into a world-leading AI enterprise by accelerating the research, development, delivery, and adoption of AI.

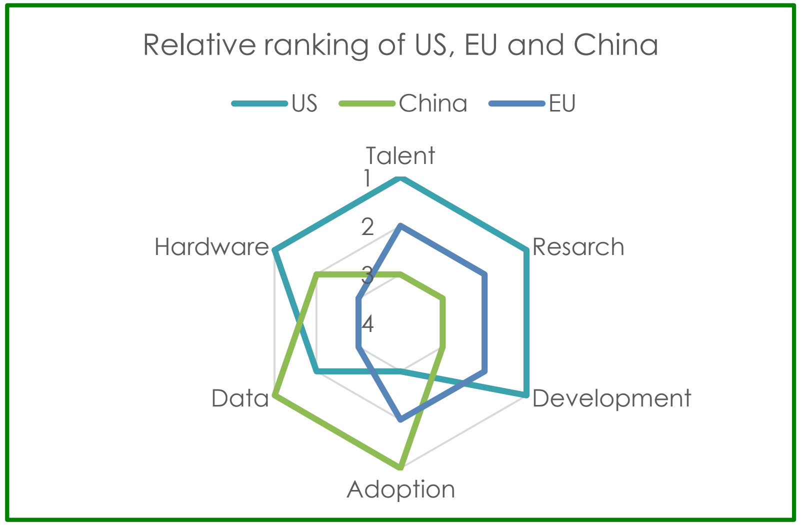
Mission:

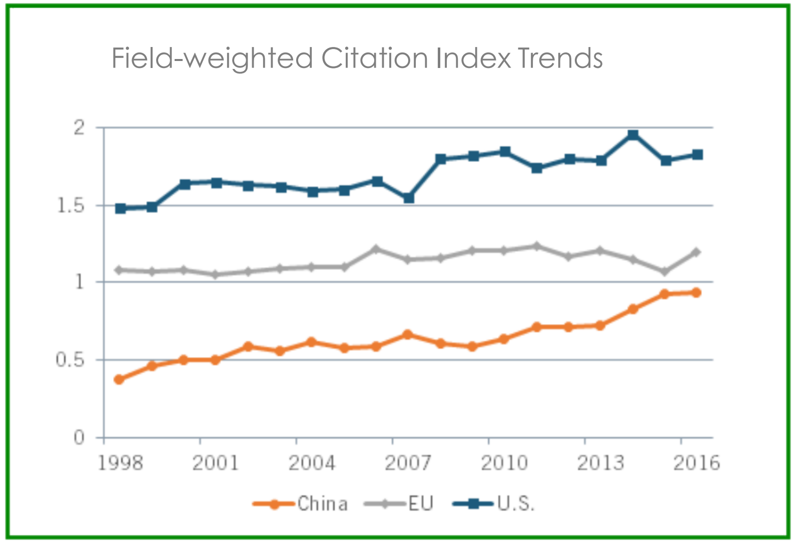
The Artificial Intelligence and Technology Office (AITO), the Department of Energy's center for Artificial Intelligence, will **accelerate the delivery** of AI-enabled capabilities, **scale** the department-wide development and impact of AI, and **synchronize** AI activities to advance the agency's core missions, **expand partnerships**, and support American AI Leadership.

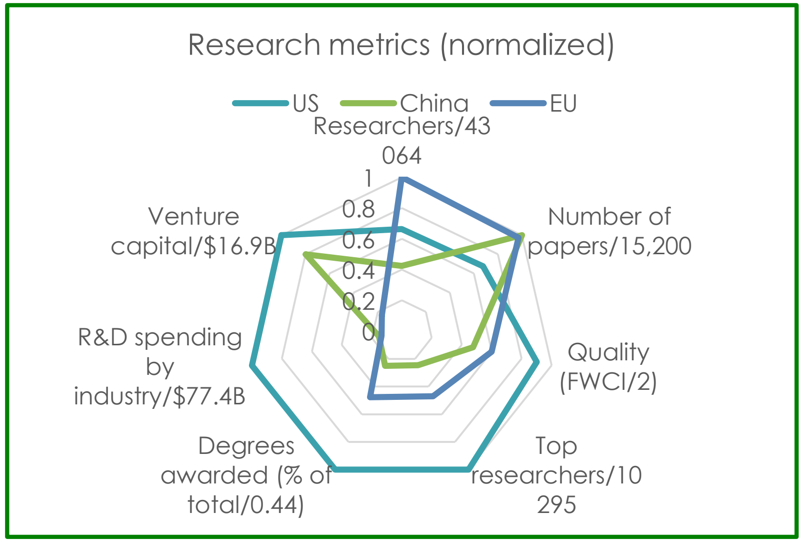
Observations on the international AI landscape

Of the 35 countries that have AI strategies, only three stand out, the U.S., the combined E.U. nations and China.

- The U.S.
  - Leads in research, development and talent (education)
  - Based on historical investments in education, laboratories and the business environment
- China
  - Leads in overall adoption of AI and the collection and use of data
  - Is investing heavily
  - Quality and development is increasing rapidly
- The E.U.
  - Has the most researchers
  - Does not translate this into innovation effectively



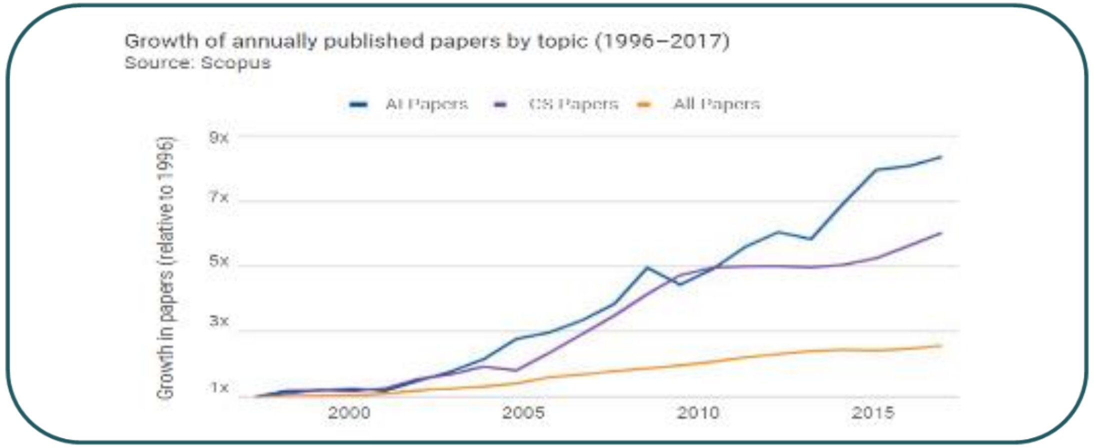


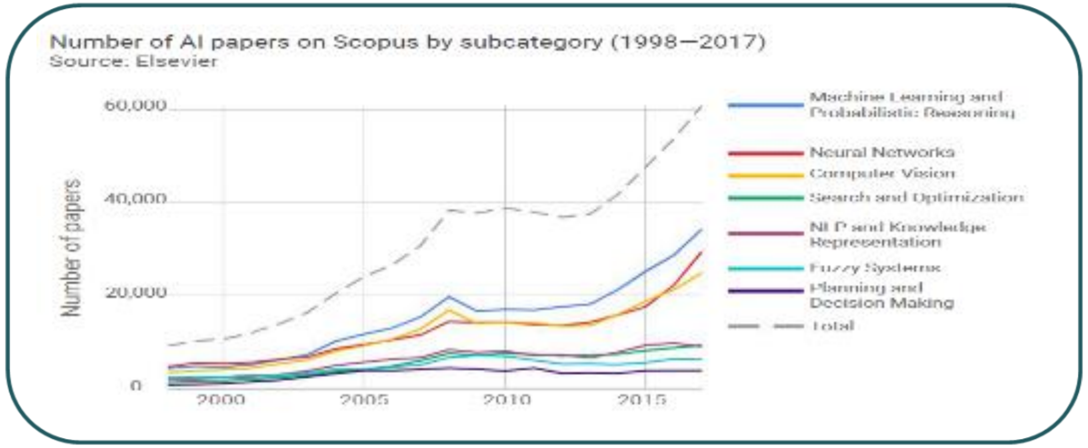


Oak Ridge National Laboratory

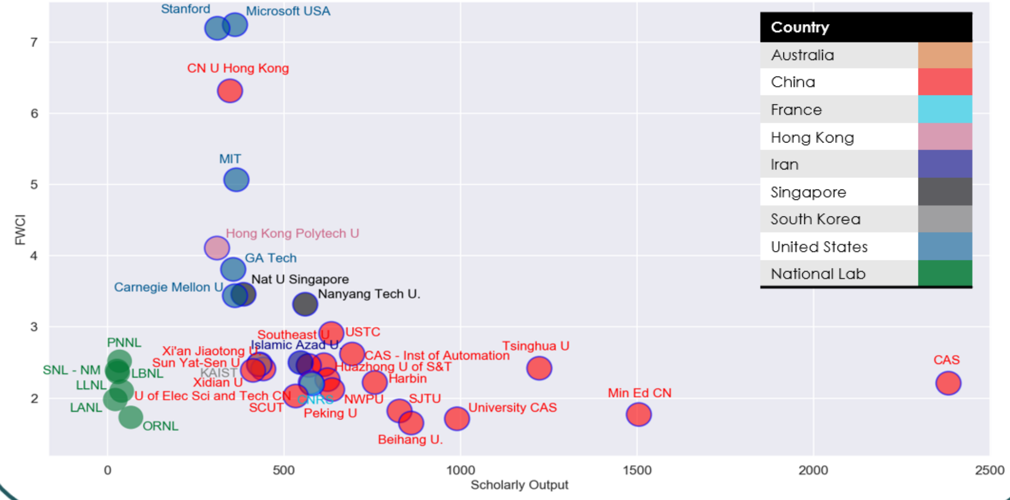
Source: "Who Is Winning the AI Race," Report, Center for Data Innovation

The AI/ML Research Landscape (Measured by Publications)





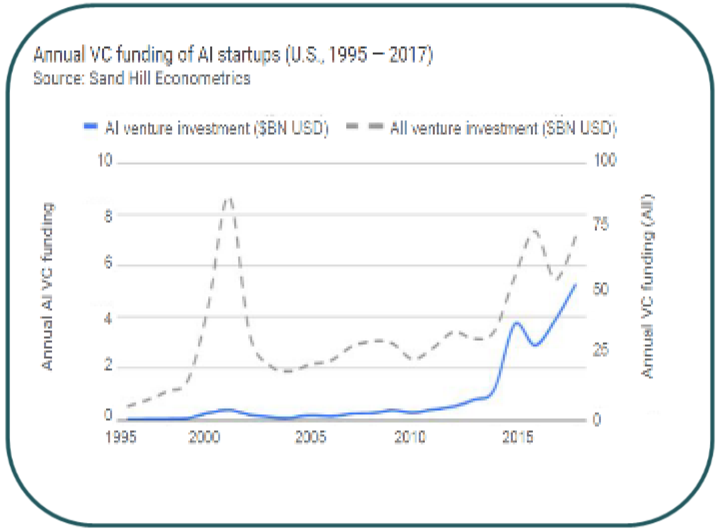
Neural Network Publications
(2016 – Present)



Machine Learning Publications
(2016 – Present)

The Business Landscape

- Business must incorporate AI
  - The “Big 9” dominate, but don’t discount traditional business
  - $7.4B in start-up investments in 488 deals in 2019/Q2 (over $12B over the past 6 months of 2019)
  - $803M in “AI for cybersecurity” VC six months of 2019
- Barriers to insertion
  - Understanding: 37% of executive feel their employees understand the importance of data
  - Trust:
    - 49% of U.S. consumers would trust AI-generated advice for retail,
    - 38% would trust AI-generated advice for hospitality, while only
    - 20% would trust AI-generated advice for healthcare and
    - 19% for financial services
    - Example: 33% of US and 85% of Chinese healthcare professionals have implemented AI into their practice, compared to a 5-country average of 46%.
- Need a consistent approach to regulatory (data and sensitive technologies)
- CB Insights top 100 startups dominated by U.S. (country) and healthcare (sector)



OAK RIDGE National Laboratory

AI for Science Report Summary

Workshop Organization

Science and Engineering

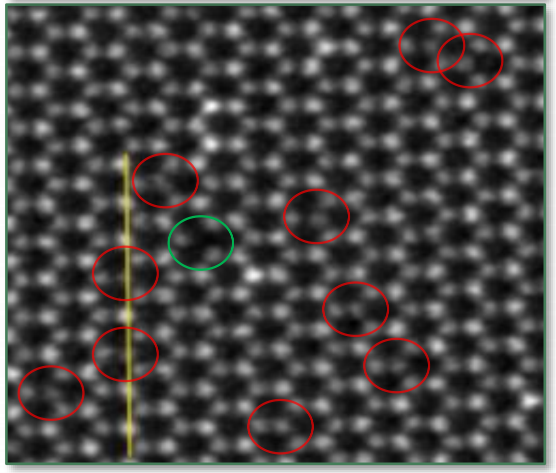
- **High Energy Physics**
  - Cosmology and Astrophysics
  - Particle Physics
  - Accelerator Design and Control
- **Nuclear Physics**
  - Particle Identification
  - Particle Tracking
  - Control
- **Fusion**
  - Feature Tracking
  - Steering
  - Design and Control
- **Energy Sciences**
  - Materials and Chemistry Modeling
  - Photon/Neutron Science
  - Electron Microscopy
- **Earth Sciences**
  - Climate & Carbon
  - Subsurface
  - Water
- **Biological Sciences**
  - Environmental Biology
  - Imaging
  - Syn Biology
  - Health
- **Energy Technologies**
  - Engineering and Manufacturing
  - Smart Grid/Urban
  - Wind and Solar
  - Mobility
- **Computing Sciences**
  - AI for hardware / software
  - AI for networks / computing facilities

Crosscuts

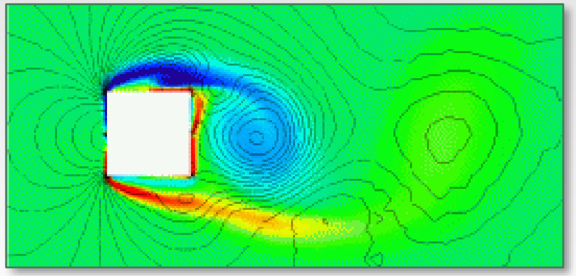
- Foundations
- Data Lifecycle
- Software
- Hardware
- Computing Facilities

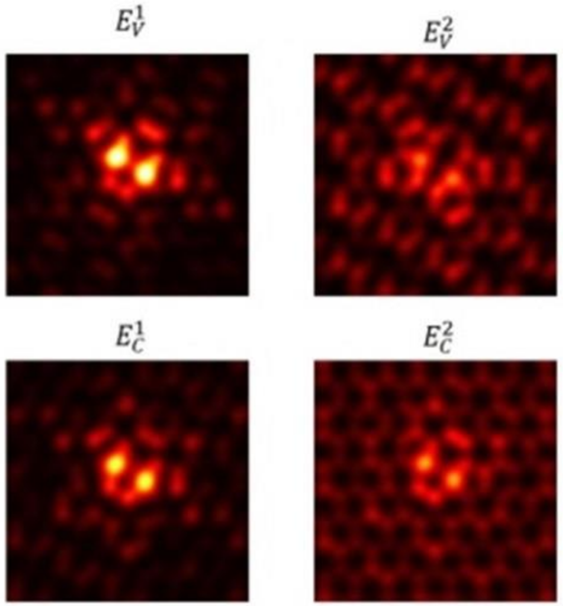
An AI Taxonomy Drives Research Strategy

Classification and regression

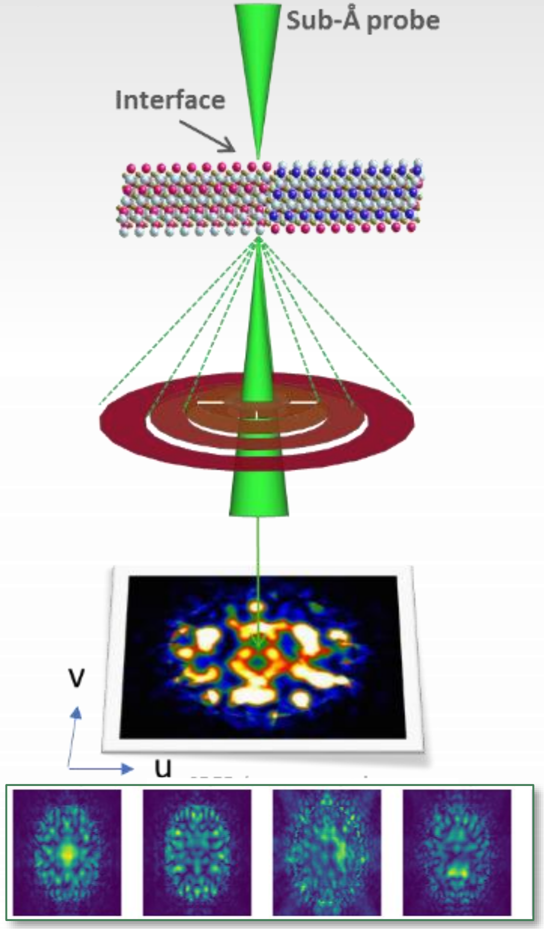


Surrogates

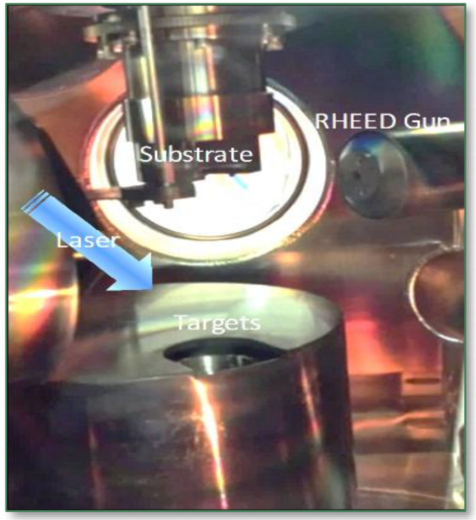




Inverse problems, design and optimization



Control systems



Report Outline

- Materials, Environmental and Life Sciences
  - Chemistry, Materials and Nanoscience
  - Earth and Environmental Sciences
  - Biology and Life Sciences
- High-energy, Nuclear and Plasma Physics
  - High Energy Physics
  - Nuclear Physics
  - Fusion
- Engineering, Instruments and Infrastructure
  - Engineering and Manufacturing
  - Smart Energy Infrastructure
  - AI for Computer Science
  - AI for Imaging
  - AI at the edge
  - Facilities Integration and AI Ecosystem
- Foundations, Software, Data Infrastructure and Hardware
  - AI Foundations and Open problems
  - Software Environments and Software Research
  - Data Life Cycle and Infrastructure
  - Hardware Architectures

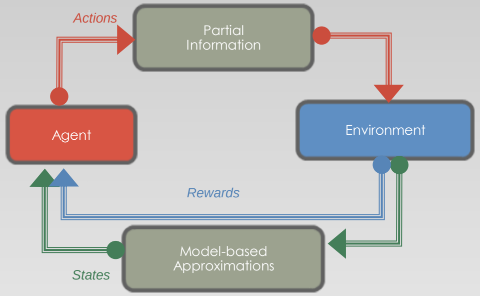
AI Foundations

Foundational Investments

Data

- Experimental design
- Data curation and validation
- Compressed sensing
- Facilities operation and control

Learning

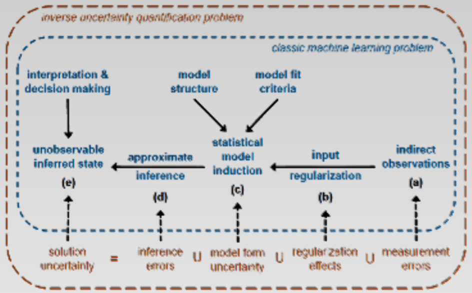


- Physics informed
- Reinforcement learning
- Adversarial networks
- Representation learning and multi-modal data
- “Foundations” of learning, e.g., robustness, stability

Scalability

- Algorithms, complexity and convergence
- Levels of parallelization
- Mixed precision arithmetic
- Communication
- Implementations on accelerated-node hardware

Assurance



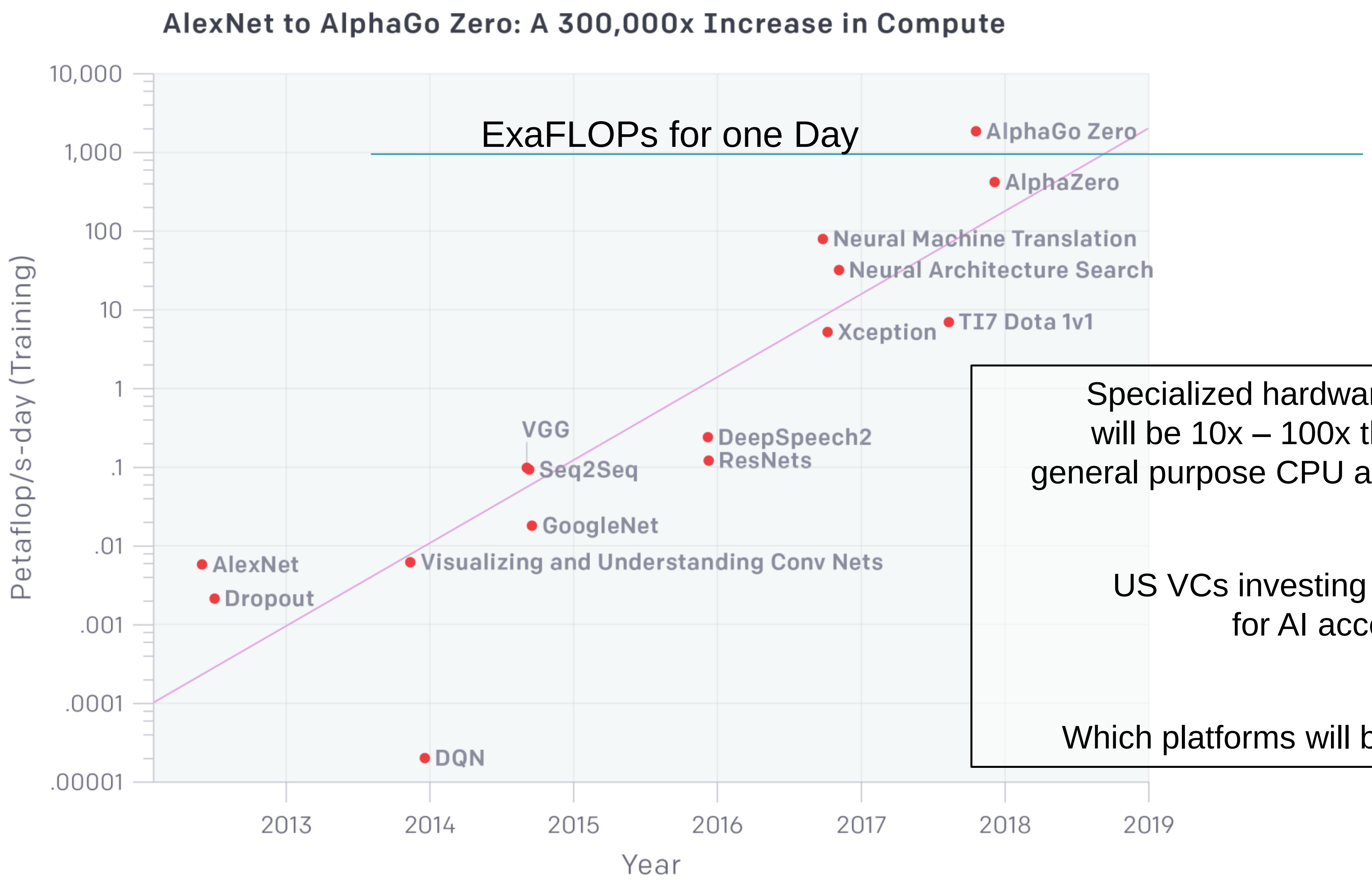
- Uncertainty quantification
- Explainability and interpretability
- Validation and verification
- Causal inference

Workflow

- Computing at the edge
- Compression
- Online learning
- Federated learning
- Infrastructure
- Augmented intelligence
- Human-computer interface

Facilities and AI Ecosystem

Deep Learning Needs High Performance Computing



Specialized hardware is emerging that will be 10x – 100x the performance of general purpose CPU and GPU designs for AI

US VCs investing >$4B in startups for AI acceleration

Which platforms will be good for science?

DOE is building on a record of success delivering HPC capabilities

Pre-exascale systems | First exascale systems

2012	2016	2018	2020	2021-2023

ORNL
Cray/NVIDIA

ORNL
IBM/NVIDIA

FRONTIER

ORNL
Cray/AMD

ANL
IBM BG/Q

ANL
Cray/Intel KNL

ANL
Intel/Cray

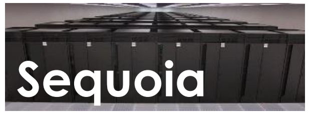
LBNL
Cray/Intel Xeon

LBNL
Cray/Intel Xeon/KNL

LBNL
Cray/AMD/NVIDIA

Dramatically increasing AI/ML capabilities

DOE NNSA ASC Computing



LLNL
IBM BG/Q

LANL/SNL
Cray/Intel Xeon/KNL

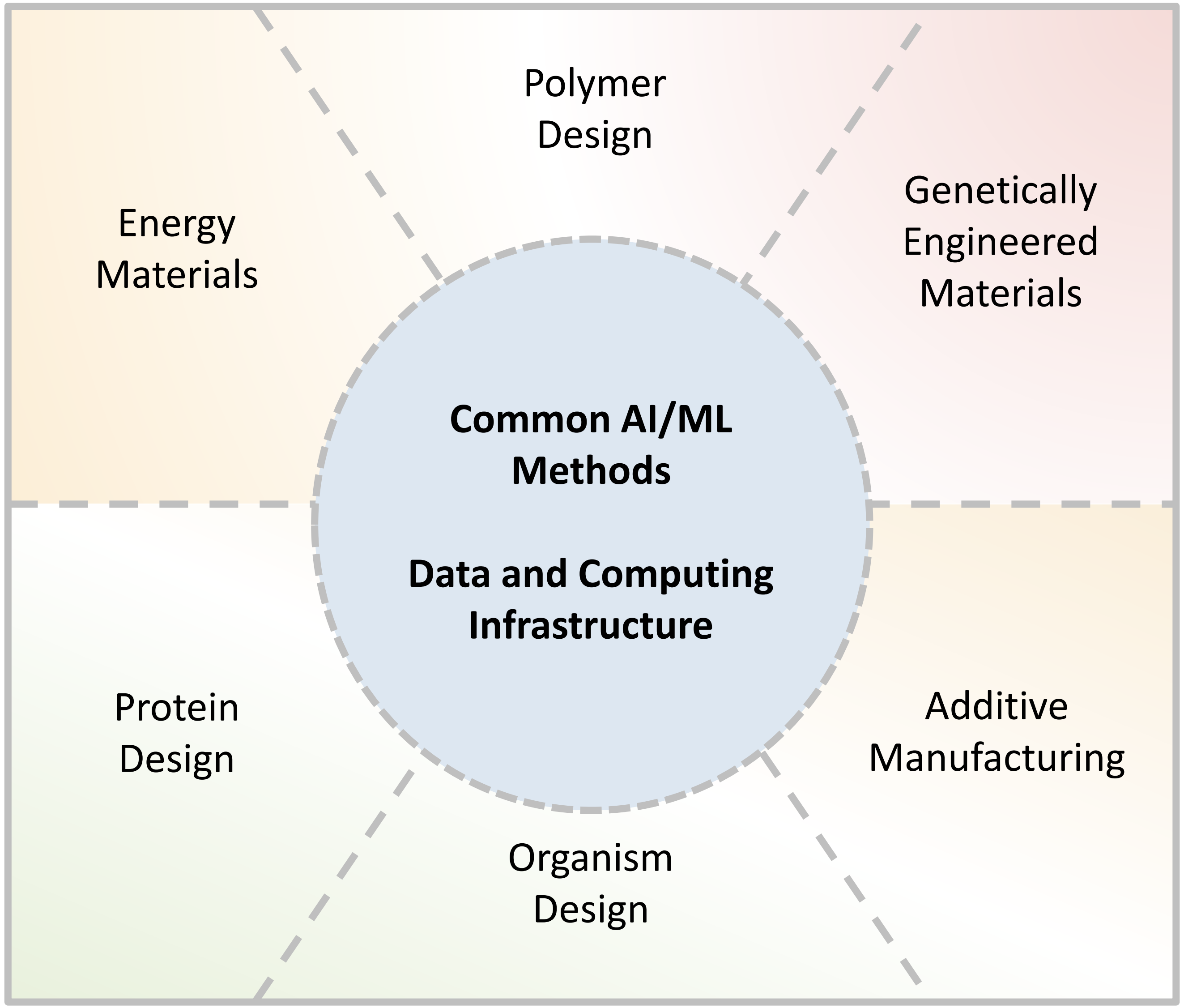
LLNL
IBM/NVIDIA

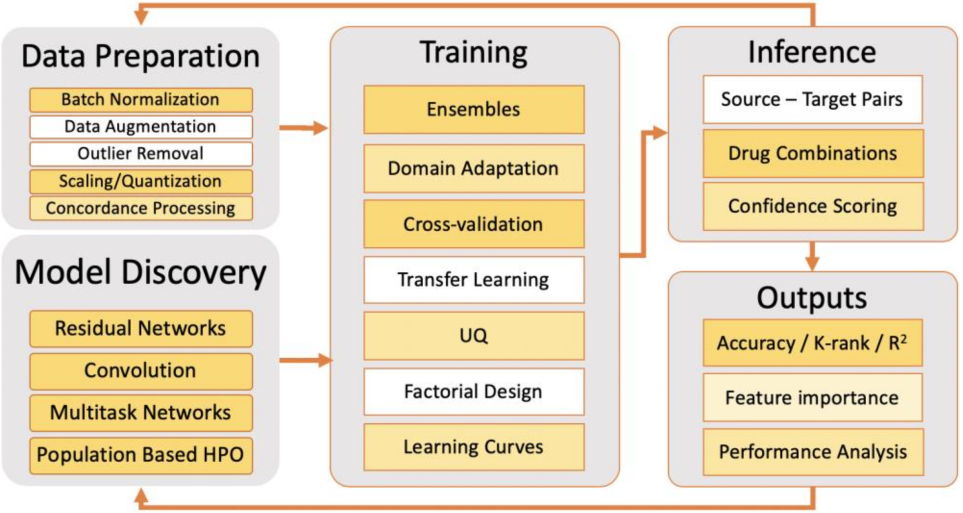
CROSSROADS

LANL/SNL
TBD

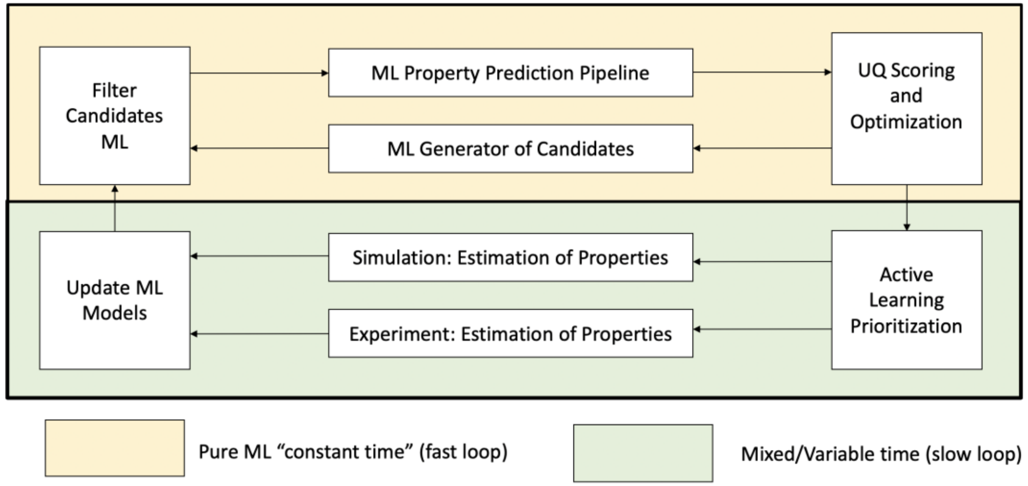
LLNL
Cray

AI Driven Autonomous Laboratory Cluster





Layered workflow combining AI, HPC and HTS



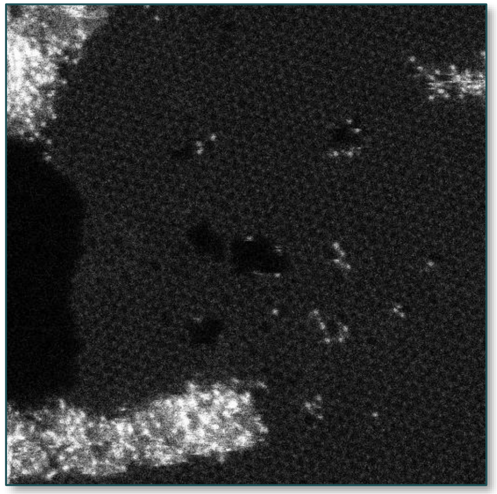
Oak Ridge National Laboratory

Oak Ridge National Laboratory

HPC for Next Generation Materials Science

AI for Defect Characterization and Materials Design

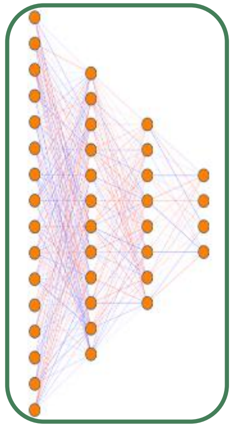
Experiment



Graphene+Si under e-beam
(STEM data from O. Dyck (ORNL))

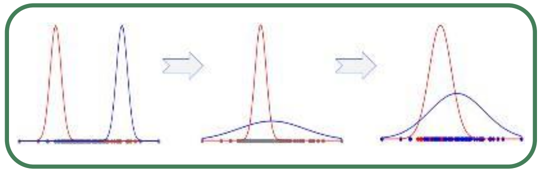
Machine Learning

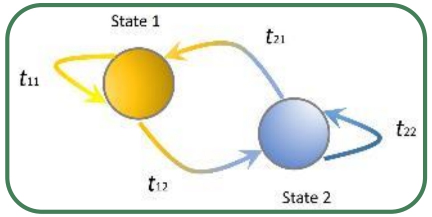
Deep learning



Position and type of
all atomic species

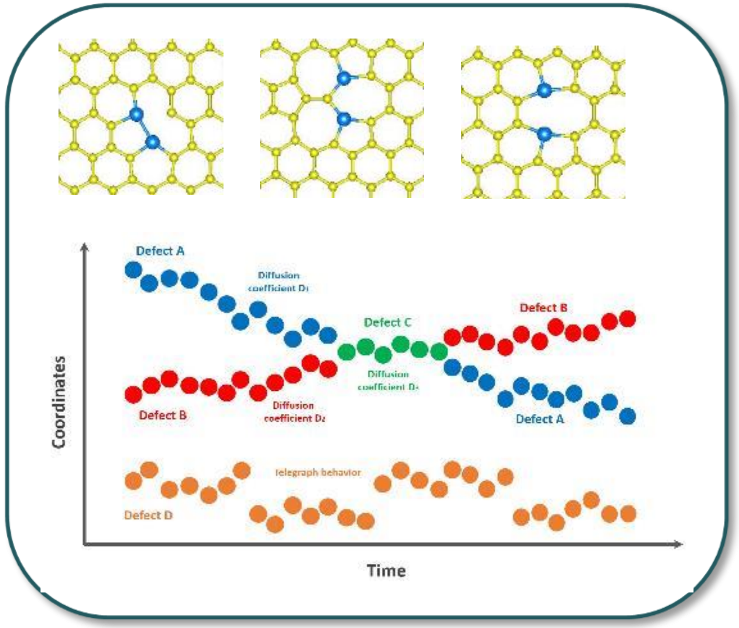
Clustering/unmixing:
Library of structural descriptors





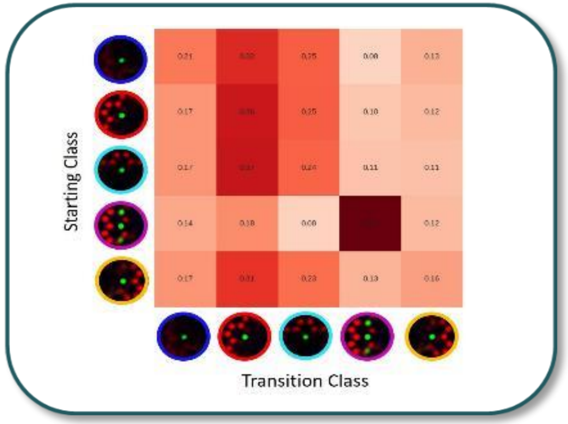
Markov model:
System dynamics

Atomic-scale physics & chemistry



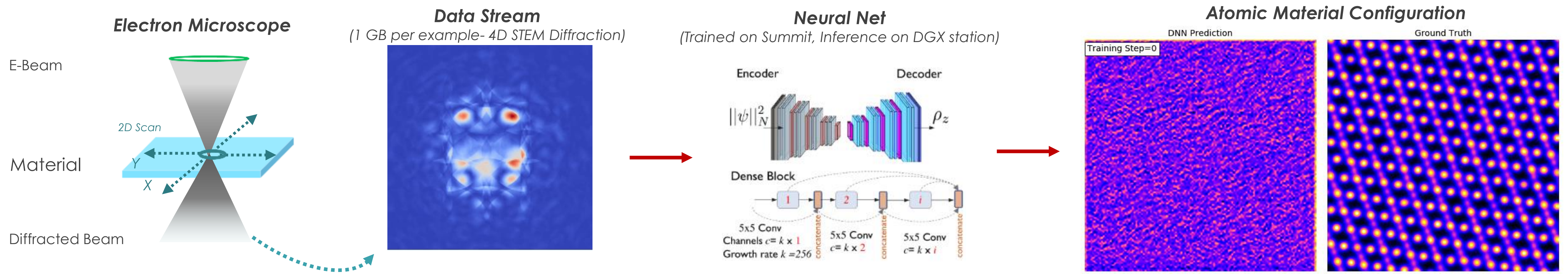
AI/ML fundamentally changes the process

- Convert noisy experimental data into atomic positions/trajectories → CDNN
- Create libraries of structural descriptors → Clustering/unmixing methods applied to the output of neural networks
- Analysis of dynamics and transition probabilities → Markov modelling on the constructed classes



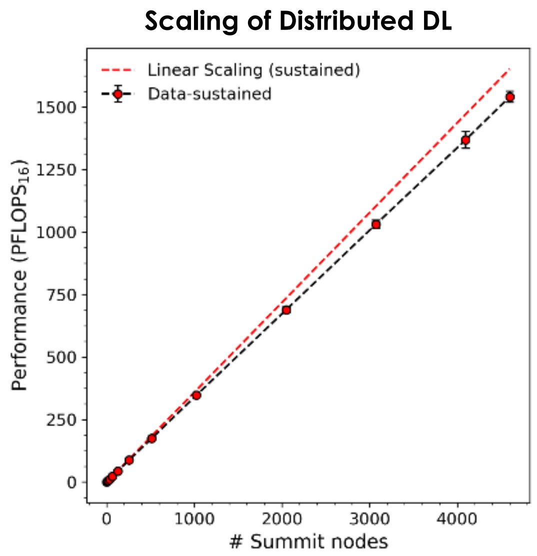
Ziatdinov et al., ACS nano 11, 12742 (2017)
Ziatdinov et al., npj Computational Materials 3, 31 (2017)
Ziatdinov et al., ArXiv:1901.09322 (2019)
Maksov et al., npj Computational Materials 5, 12 (2019)

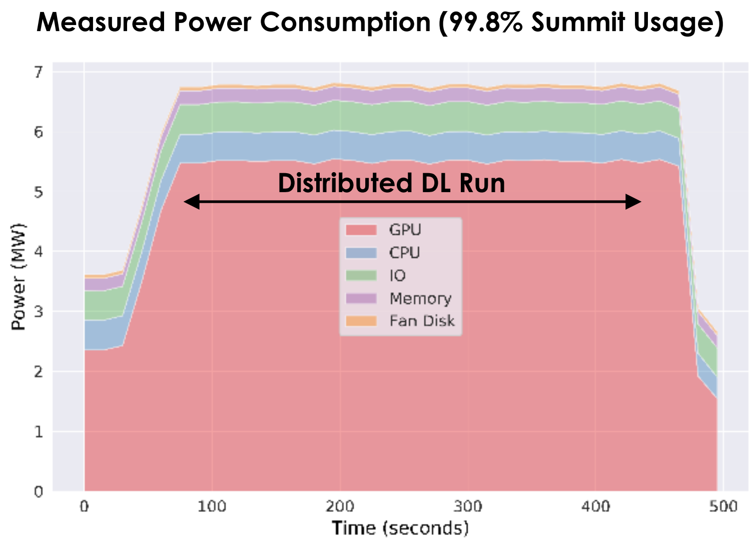
AI coupled with Summit enable structure inversion



One training example is ***4000x larger than ImageNet*** example. Training on ~ 10 GPUs takes days.

- *Distributed Training up to ~ 10,000 GPUs produces near-linear speedup.*
- ***93% scaling efficiency** on full Summit System*
- *Peak Performance of **2.15 ExaFLOPS** (16-bit)*





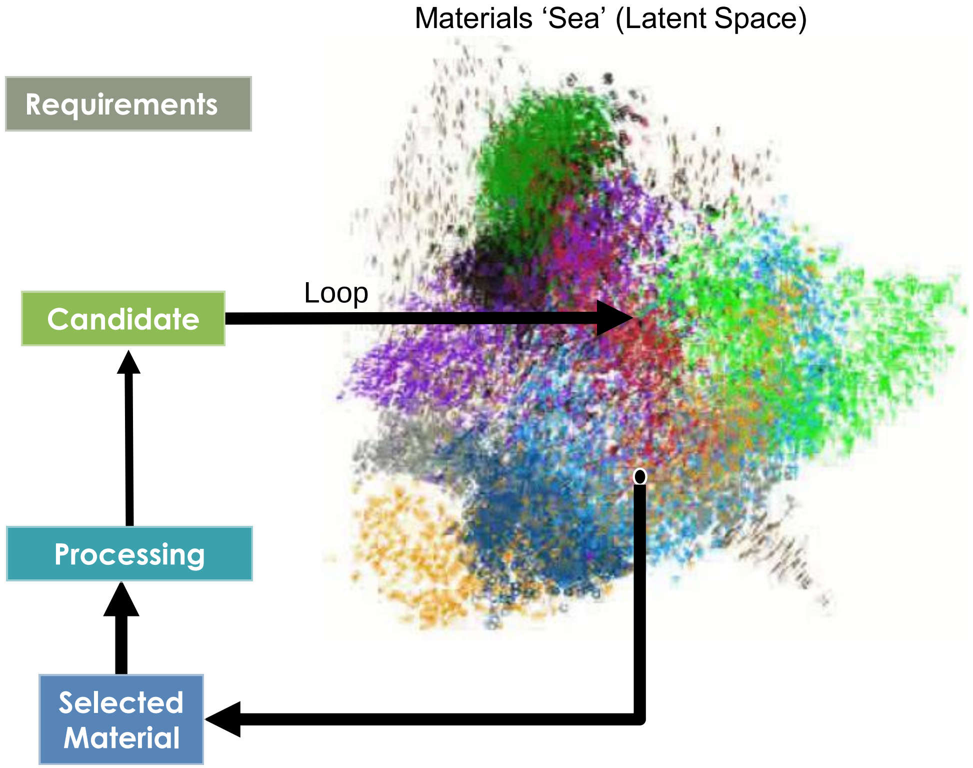
Oak Ridge National Laboratory

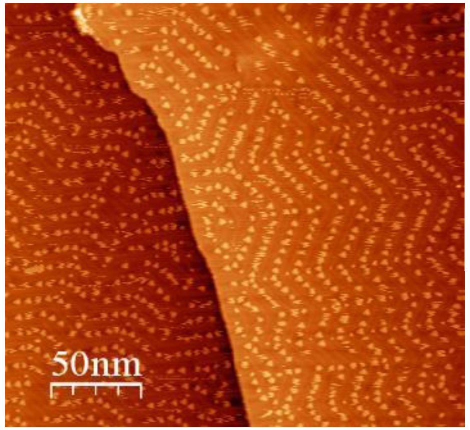
Credit: Numan Laanait

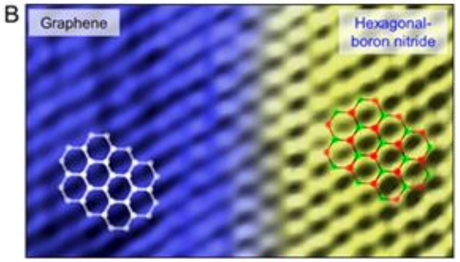
INCITE Leadership Computing

AI-based optimization will enable fundamentally new materials and processing

- Given requirements (high T operation, strength, cost, etc.), can we discover new materials to address these needs?
- After discovery, can we optimize the processing to achieve the aims?
- Scientific challenges
  - Descriptors for materials at all length scales
  - Generation of local structure-property libraries with AI-assisted approaches
  - Impact of processing on subsequent properties
  - Efficient experimental explorations of the design space
  - Cheaper computational modeling for rapid iteration







Science (2014); *Nature Comm.*(2014)

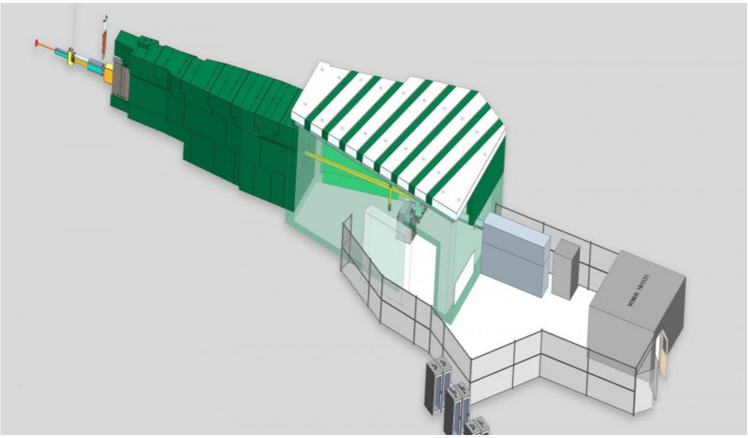
Credit: Rama Vasudevan

Oak Ridge National Laboratory

HPC for Scalable Imaging

Summit-scale U–Net Training

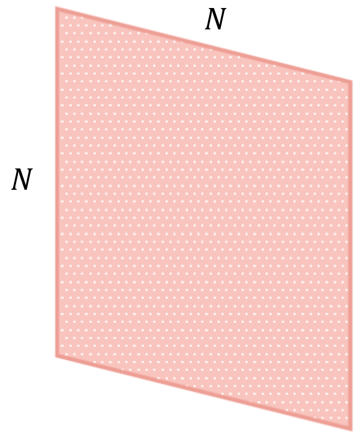
- Goal is not just faster, but also better.
- Satellite images collected at high-resolutions (30-50 cm) yield very large 10,000 x 10,000 images.
- Training ML models on these large high-resolution images is extremely challenging.
- Accurate ML models are needed to resolve multi-scale objects (buildings, solar panels, land cover details).
- U-Net models preferred -- good for training with limited labeled data.
- At present, requires many days to train a single model (even on special-purpose DL platforms like DGX boxes).
- Hyperparameter tuning of these models take much longer.



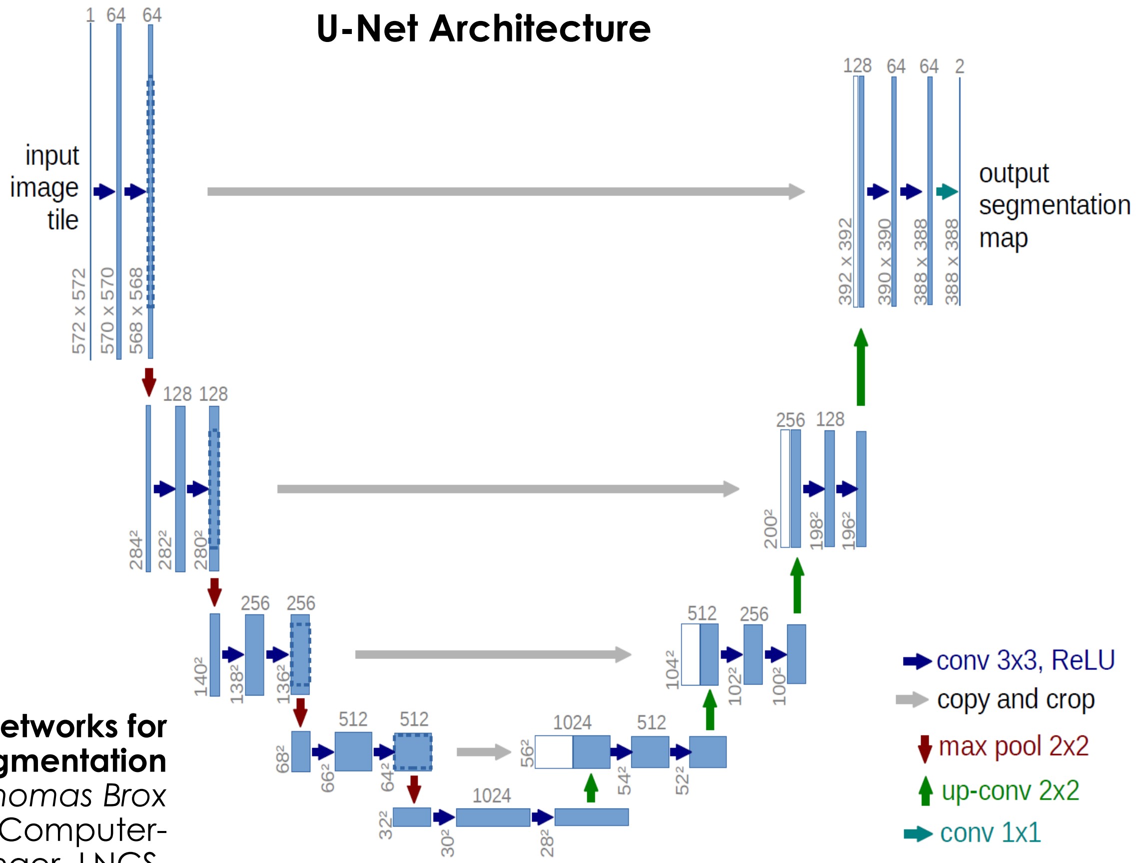
ML technologies developed will be here applicable to other large-scale image analytics domains (e.g., as anticipated from the VENUS neutron imaging instrument).

Semantic Segmentation with U-Net

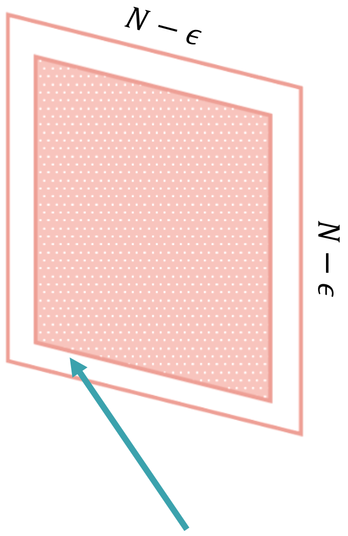
Input Image



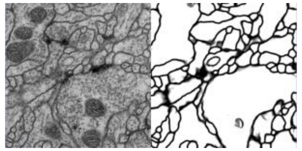
U-Net Architecture



Output Image



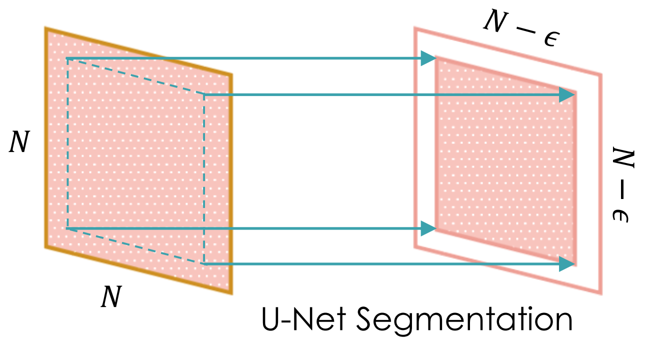
- Refer to this as the halo region.
- Halo width is a function of U-Net architecture (depth, channel width, filter sizes, etc.).
- Halo width determines the receptive field of the model.
- Larger the receptive field, wider the length-scales of identifiable objects.

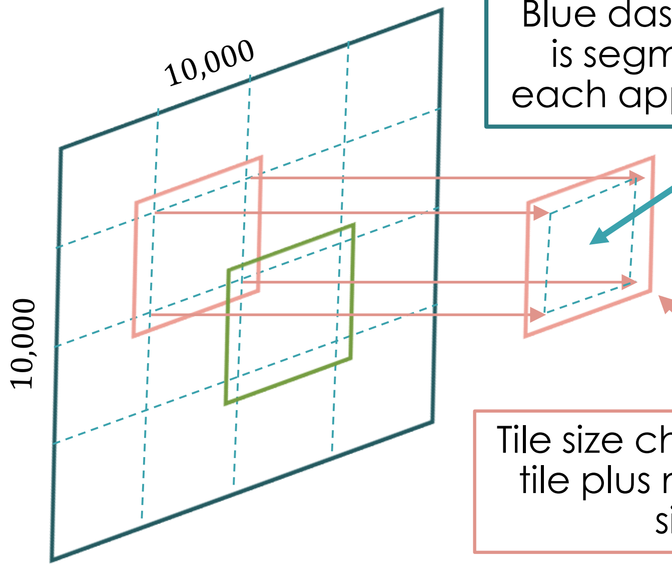


U-Net: Convolutional Networks for Biomedical Image Segmentation
Olaf Ronneberger, Philipp Fischer, Thomas Brox
Medical Image Computing and Computer-Assisted Intervention (MICCAI), Springer, LNCS, Vol.9351: 234--241, 2015.

Scalable Sample Parallel Segmentation
Leveraging Summit's Vast GPU Farm

- Given a $N \times N$ image, U-Net segments a $(N - \epsilon) \times (N - \epsilon)$ inset square.

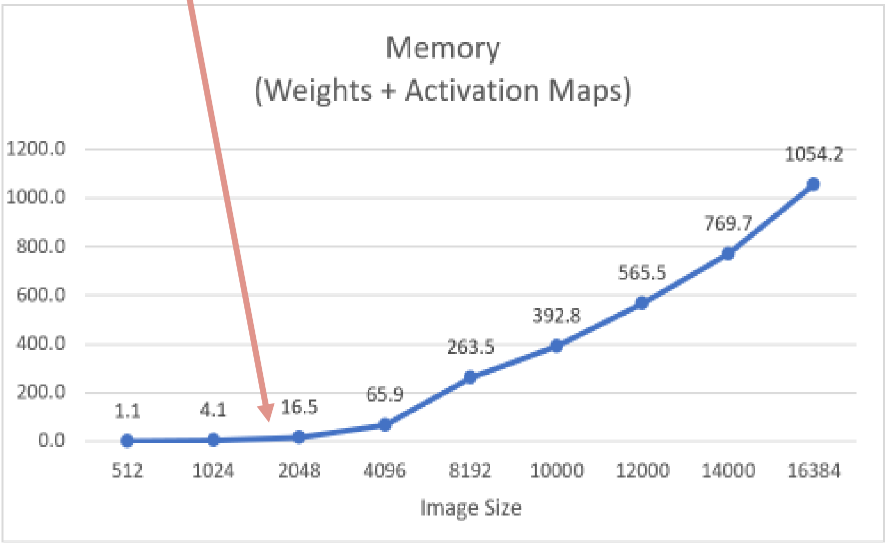




Blue dashed square is segmented for each appended tile.

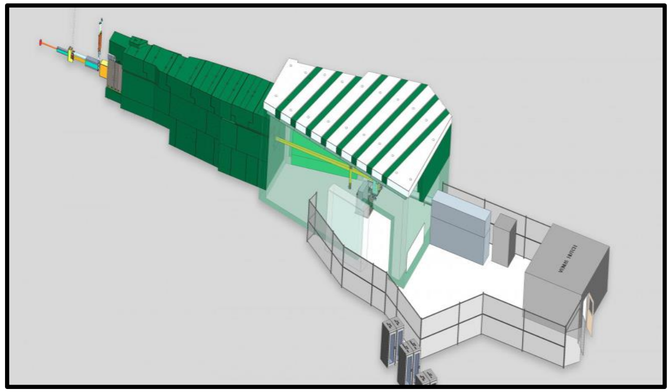
Tile size chosen such that appended tile plus model parameters fit on a single Summit GPU.

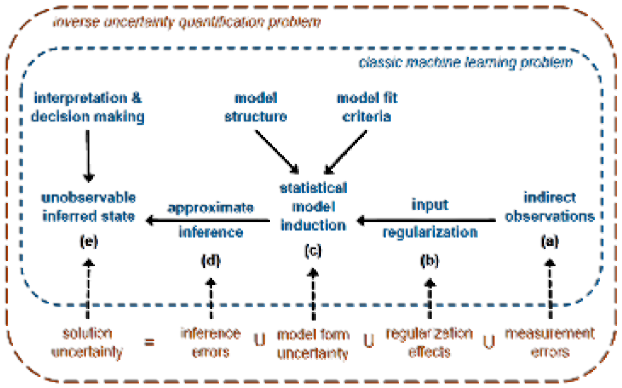
- Partition each 10,000 x 10,000 image sample into non-overlapping tiles.
- Append an extra halo region of width ϵ along each side of each tile.
- Assign each appended tile to a Summit GPU. Use U-Net to segment appended tile.
- Each GPU segments an area equal to that of the original non-overlapping tile.

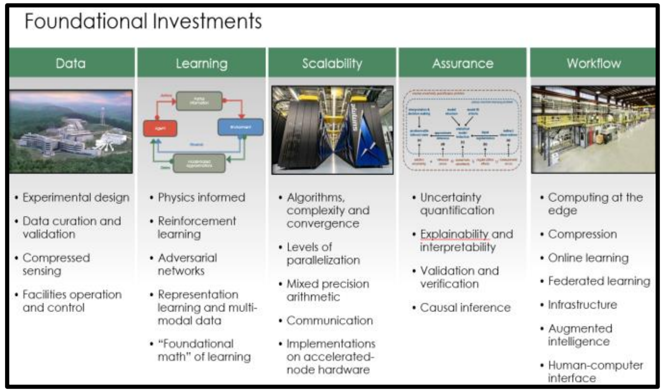


Summary: My Identified Themes From The Townhalls

- Laboratory of the Future. AI changes
  - How we do science
  - How we manage facilities
  - How we manage data
- Assurance. We are going to have to think differently and more intentionally.
  - Quality of science
  - Verification and Validation
  - Societal impacts
- Fundamental technical challenges
  - Need foundational investments in math and computer science
  - Need to deliver a software and hardware infrastructure
  - Need a co-design mentality







Thank you