AI for Science

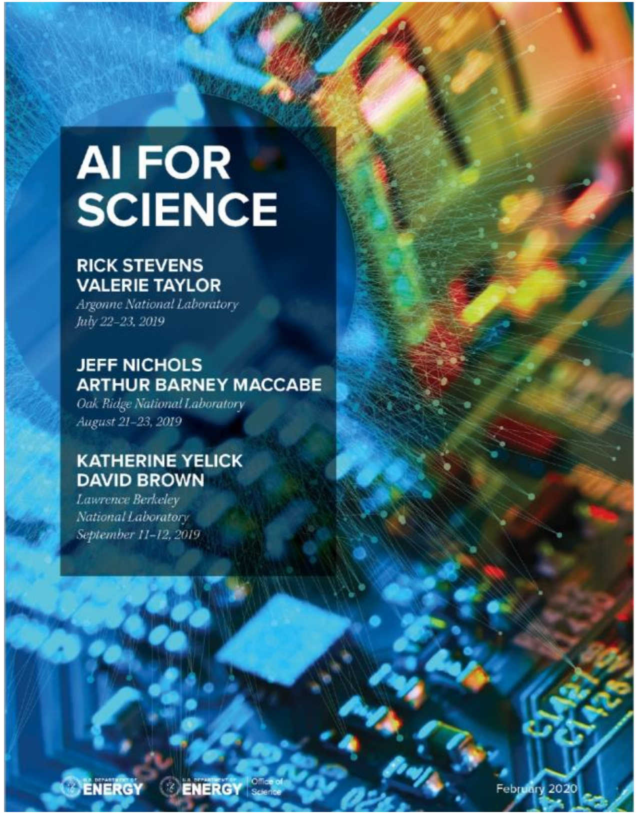
Rick Stevens, Valerie Taylor, ANL

Jeff Nichols, Barney Maccabe, ORNL

Kathy Yelick, David Brown, LBNL

U.S. DEPARTMENT OF ENERGY | Office of Science

AI for Science – What's Next After Exascale



- Over 1,000 scientists participated in four town halls during the summer of 2019
- Research Opportunities in AI
  - Biology, Chemistry, Materials,
  - Climate, Physics, Energy, Cosmology
  - Mathematics and Foundations
  - Data Life Cycle
  - Software Infrastructure
  - Hardware for AI
  - Integration with Scientific Facilities
- Modeled after the Exascale Series in 2007

Use of AI in Science

- **1**: Chemistry, Materials, and Nanoscience
- **2**: Earth and Environmental Sciences
- **3**: Biology and Life Sciences
- **4**: High Energy Physics
- **5**: Nuclear Physics
- **6**: Fusion
- **7**: Energy and Manufacturing
- **8**: Smart Energy Infrastructure
- **9**: AI for Computer Science

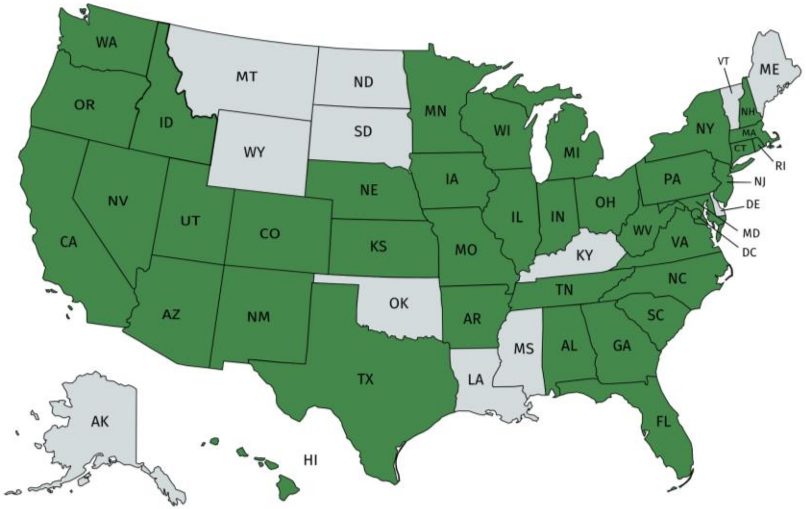
Research to Advance AI

- **10**: AI Foundations and Open Problems
- **11**: Software Environments and Software Research
- **12**: Data Life Cycle and Infrastructure
- **13**: Hardware Architectures
- **14**: AI for Imaging
- **15**: AI at the Edge
- **16**: Facilities Integration and AI Ecosystem

Participation at the AI for Science Town Halls

- Over 1000 registrations across 4 Town Halls

ANL	357	
ORNL	330	
LBNL	349	+100 online
DC	273	+ ?
Totals	1309	



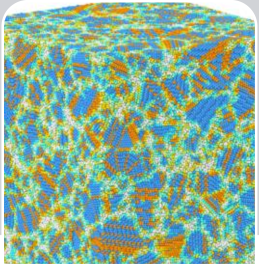
- All 17 DOE National Laboratories
- 39 Companies from large and small
- Over 90 different universities
- 6 DOE/SC Offices + EERE and NNSA

Plenary Presentations at the Town Halls

Argonne

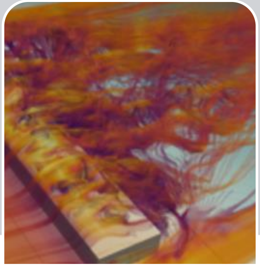
Cosmology

Salman Habib



Materials

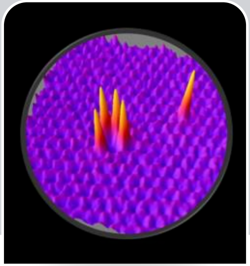
Ian Foster



Climate

Rao Kotamarthi

Oak Ridge

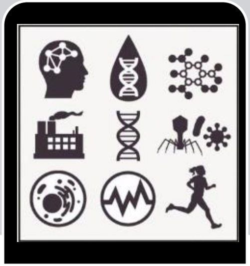


Microscopy

Sergei Kalinin

Manufacturing

Tom Kurfess



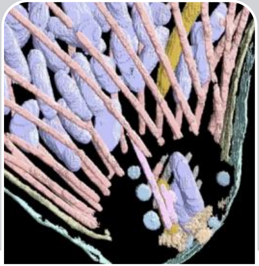
Health

Gina Tourassi

Berkeley

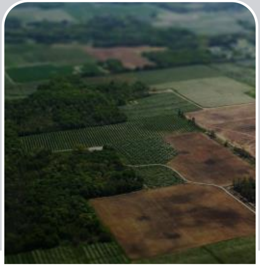
Astrophysics

Josh Bloom



User Facilities

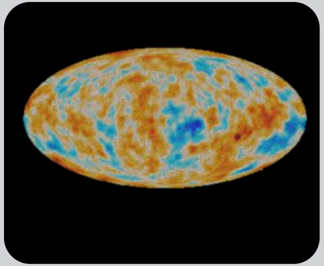
James Sethian



Biology

Ben Brown

DOE’s Role in AI for Science

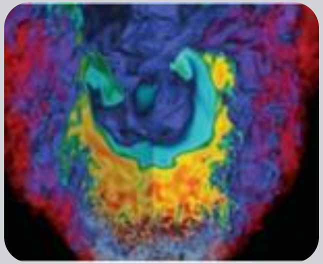


Data Sets

Curated Data for Science

HPC

Most powerful computers for science



Math and CS Research

Science inspired foundations

User Facilities

Observation and Experiment

Team Science

End-to-end science solutions

Expected AI for Science Impacts

- **Acceleration of Discovery Rates**
  - materials, chemistry, biology, physics, engineering, climate, etc.
  - targeted search, optimization, automation, (drug design > 5x rate improvement)
- **From Simple Automation to Goal Directed Systems**
  - semi-autonomous “self driving” laboratories
  - active learning loops
- **Simulation + AI hybrids** $\Longrightarrow$ $''smart\ self - adjusting\ models''$ **Zetta to Yotta**
  - refactoring and restructuring large-scale simulation to leverage ML
  - replacement of functions, optimization of parameters, steering
- **Accessible and Integrated Knowledgebases**
  - new interfaces to the literature and data
  - dramatically lower costs for information extraction and curation
- **Comprehensive transformation of science support and operations**
  - AI everywhere, smart processes
  - dynamic learning organization

A Bold Vision for AI at DOE: In Ten Years

Learned models begin to replace data

- Queryable, portable, pluggable, chainable, secure

Experimental discovery processes dramatically refactored

- Models replace experiments, experiments improve models

Many questions pursued semi-autonomously at scale

- Searching for materials, molecules and pathways; new physics

Simulation and AI approaches merge

- Deep integration of ML; numerical simulation and UQ

Theory becomes data for next-generation AI

- AI begins to contribute to advancing theory

AI becomes a common part of scientific laboratory activities

- AI is integrated into science, engineering and operations

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

U.S. DEPARTMENT OF ENERGY | Science

AI4Sci

AI for Science Report

- Chapter layout
  - State of the Art
  - Major Challenges
  - Advances in the Next Decode
  - Accelerating Development
  - Expected Outcomes
- Example Chapters
  - Biology and Life Sciences
  - High Energy Physics
  - AI Foundations and Open Problems
  - AI at the Edge

Contents

Executive Summary ... 1

Introduction: AI for Science ... 5
- Materials, Environmental, and Life Sciences ... 5
- High-Energy, Nuclear, and Plasma Physics ... 7
- Engineering, Instruments, and Infrastructure ... 9
- Foundations, Software, Data Infrastructure, and Hardware ... 11
- Conclusions ... 14

01. Chemistry, Materials, and Nanoscience ... 17
1. State of the Art ... 17
2. Major (Grand) Challenges ... 18
3. Advances in the Next Decade ... 21
4. Accelerating Development ... 23
5. Expected Outcomes ... 25
6. References

02. Earth and Environmental Sciences
1. State of the Art
2. Major (Grand) Challenges
3. Advances in the Next Decade
4. Accelerating Development
5. Expected Outcomes
6. References

03. Biology and Life Sciences
1. State of the Art
2. Major (Grand) Challenges
3. Advances in the Next Decade
4. Accelerating Development
5. Expected Outcomes
6. References

04. High Energy Physics
1. State of the Art
2. Major (Grand) Challenges
3. Advances in the Next Decade
4. Accelerating Development
5. Expected Outcomes
6. References

05. Nuclear Physics
1. State of the Art
2. Major (Grand) Challenges
3. Advances in the Next Decade
4. Accelerating Development
5. Expected Outcomes
6. References

Contents (Cont.)

06. Fusion ... 65
1. State of the Art ... 65
2. Major (Grand) Challenges ... 66
3. Advances in the Next Decade ... 69
4. Accelerating Development ... 70
5. Expected Outcomes ... 70
6. References ... 71

07. Engineering and Manufacturing ... 73
1. State of the Art ... 73
2. Major (Grand) Challenges ... 74
3. Advances in the Next Decade ... 77
4. Accelerating Development ... 78
6. References ... 79

08. Smart Energy Infrastructure ... 81
1. State of the Art
2. Major (Grand) Challenges
3. Advances in the Next Decade
4. Accelerating Development
5. Expected Outcomes
6. References

09. AI for Computer Science
1. State of the Art
2. Major (Grand) Challenges
3. Advances in the Next Decade
4. Accelerating Development
5. Expected Outcomes
6. References

10. AI Foundations and Open Problems
1. State of the Art
2. Major (Grand) Challenges
3. Advances in the Next Decade
4. Accelerating Development
5. Expected Outcomes
6. References

11. Software Environments and Software Research
1. State of the Art
2. Major (Grand) Challenges
3. Advances in the Next Decade
4. Accelerating Development
5. Expected Outcomes
6. References

Contents (Cont.)

12. Data Life Cycle and Infrastructure ... 117
1. State of the Art ... 118
2. Major (Grand) Challenges ... 119
3. Advances in Next Decade ... 122
4. Accelerating Development ... 123
5. Expected Outcomes ... 123
6. References ... 123

13. Hardware Architectures ... 125
1. State of the Art ... 125
2. Major (Grand) Challenges ... 126
3. Advances in the Next Decade ... 129
4. Accelerating Development ... 129
5. Expected Outcomes ... 131
6. References ... 131

14. AI for Imaging ... 133
1. State of the Art ... 133
2. Major (Grand) Challenges ... 135
3. Advances in the Next Decade ... 136
4. Accelerating Development ... 137
5. Expected Outcomes ... 138
6. References ... 138

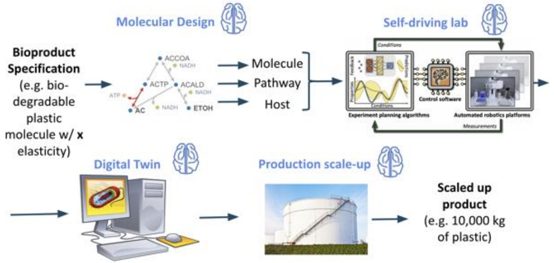
15. AI at the Edge ... 141
1. State of the Art ... 142
2. Major (Grand) Challenges ... 143
3. Advances in the Next Decade ... 145
4. Accelerating Development ... 146
5. Expected Outcomes ... 147
6. References ... 147

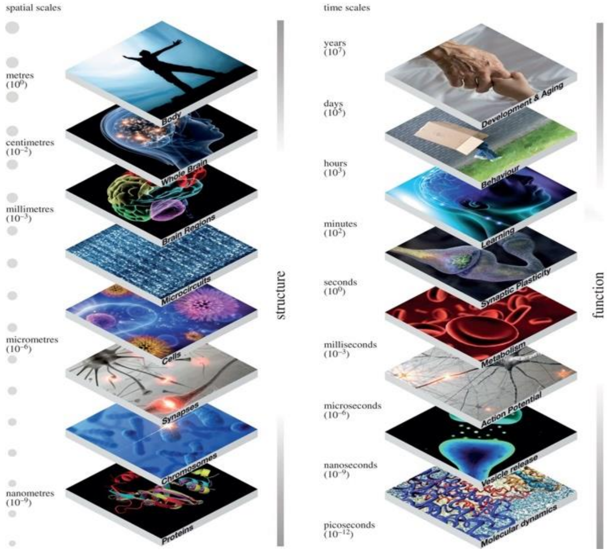
16. Facilities Integration and AI Ecosystem ... 149
1. State of the Art ... 149
2. Major (Grand) Challenges ... 149
3. Advances in the Next Decade ... 152
4. Accelerating Development ... 153
5. Expected Outcomes ... 153

AA. Report Writing Team ... 155

AB. Agendas ... 157
- Argonne National Laboratory ... 157
- Oak Ridge National Laboratory ... 160
- Lawrence Berkeley National Laboratory ... 163
- Washington, DC ... 167

Biology and Life Sciences





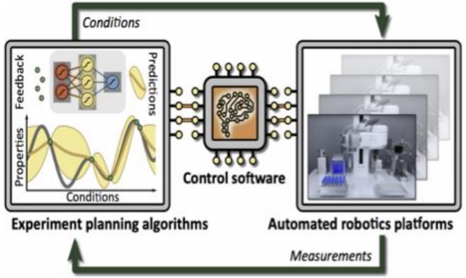
Biological systems, including humans, constitute the integration of many levels of spatiotemporal organization.

Predict, control, and understand biological systems in mechanistic, often molecular detail

Challenges:

- Build the capacity to design custom biological systems capable of addressing major global health and environmental challenges
- Learn to systematically manage and engineer global environmental systems by obtaining a predictive understanding of ecosystems and their services
- Develop AI-enabled self-driving laboratories to enable game-changing advances in the understanding and deployment of biological, chemical, and environmental systems.

Biology and Life Sciences



Accelerating Development:

- Improve scalability of datasets with respect to quantity, quality, and provenance
- Establish the infrastructure required to make communal use of data that cannot be moved or revealed due to privacy concerns
- Develop foundational technologies to promote a rigorous statistical framework to integrate knowledge across disciplines, including data-efficient learning
- Understand how data biases or inaccuracies threaten model performance on subgroups in heterogeneous settings

Outcomes: Capacity to understand, engineer & control biological systems

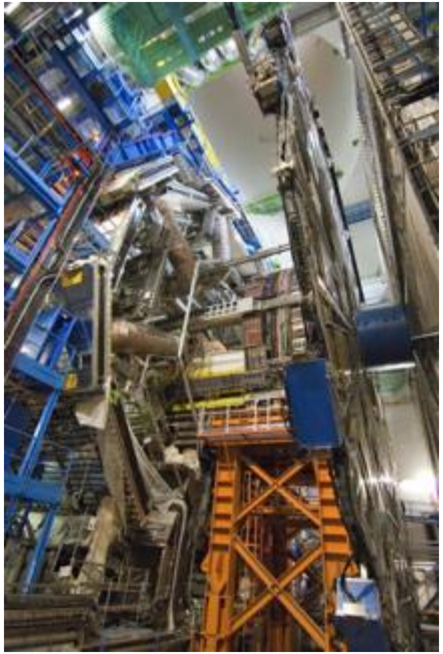
- Deliver accuracy to "precision medicine" for healthcare
- Discover the controls of massively multi-scale, dynamic biosystems
- Build life to spec
- Engineer our troposphere

High Energy Physics

Discovering the ultimate constituents of matter and uncovering the nature of space and time

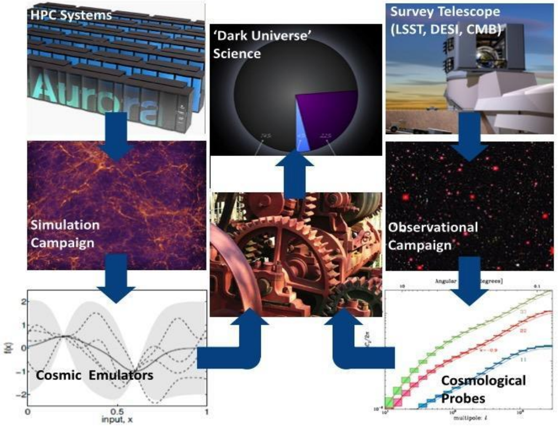
Challenges:

- Reconstruct the history of the universe using AI techniques
- Advance knowledge of cosmic structure formation with the AI-driven Automated Cosmology Experiment (ACE)
- Zettascale AI to uncover new fundamental physics



The ATLAS detector at the LHC under construction in 2007.

High Energy Physics



Accelerating Development:

- Usable tools for large-scale distributed training and optimization of ML models
- Training methodologies that are able to detect rare features in high-dimensional spaces while being robust against systematic effects
- Tools to quantify the impact of systematic effects of the accuracy and stability of complex ML models

Outcomes:

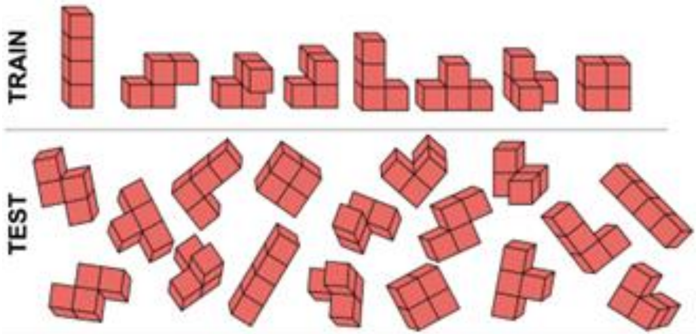
- Enable the exploration of the data from the next-generation surveys
- Make a movie of the universe from its earliest moments until today
- A new era of precision physics at the Energy and Intensity frontiers

AI Foundations and Open Problems

Advancing the mathematical, statistical, and information-theoretic foundations of AI

Challenges:

- Incorporate domain knowledge in ML and AI
- Establish assurance for AI
- Achieve efficient learning for AI systems



Specially designed neural networks can satisfy domain properties such as 3D rotation-equivariance, allowing one to train on shapes and molecules in one orientation while still identifying shapes and molecules in any orientation. Adapted from N. Thomas, NeurIPS18 [1]

AI Foundations and Open Problems

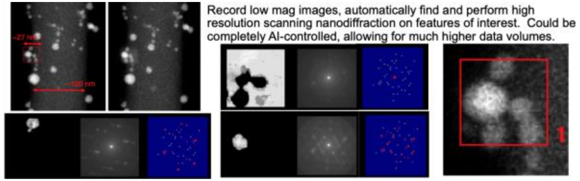
Accelerating Development:

- The use of scientific principles, modeling and simulation, and domain-specific knowledge to inform and advance AI
- Addressing robustness, uncertainty quantification, and interpretability of AI systems
- Learning for inverse problems and design of experiments
- Reinforcement and active learning to develop AI for control and data acquisition system

Outcomes:

- Increase trust in ML and AI as scientific techniques
- Provide efficient computational algorithms for ML and AI
- Maximize the understanding realized from science-informed AI

AI at the Edge



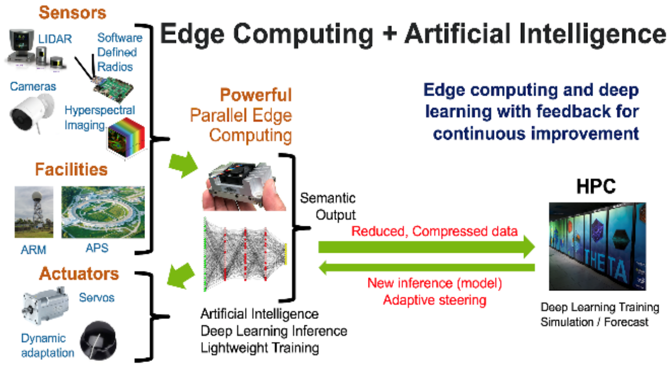
Instrument steering using AI

Local computing resources to analyze and compress experimental and observational data while additionally enabling remote steering and intelligent responses to changing conditions

Challenges:

- Improve scientific productivity with high-speed data through AI at the edge
- Enhance scientific discovery through integration of multiple data sources
- Enable smart scientific infrastructures through AI at the edge
- Integrate systems of systems using AI at the edge

AI at the Edge



Accelerating Development:

- Learning under limited resources
- Understanding errors, failures, and correctness
- Dealing with all aspects of the computing continuum
- Modeling interactions
- Managing dynamic resources and interacting systems

Outcomes:

- Enable data collection and analyses at scales not previously possible
- Enable large-scale experiments in harsh environments
- Enable real-time control/steering of experiments
- Change the way DOE scientists work

AI Science Applications

AI Enabled

Design Workflows

(what to make)

→ …materials, polymers, organisms…

AI Enabled

Experimental Workflows

(how to make it)

→ …self-driving labs, synthesis search…

AI Enabled

Scientific Comprehension

(what it means)

→

- data Sets
- literature
- science "news"
- strategy

→ *Cleaned*
Updated
Annotated
Aggregated
Interpreted

→ Insight?

AI4Sci

AI for Science: AI Building Blocks (examples)

Design
- Biology
- Materials
- Chemistry
- Batteries
- Devices
- Drugs

Control
- Accelerators
- Experiments
- Reactors
- Simulation
- Mobility
- Energy

Augmented Simulations
- Landscape Search
- Surrogates
- Optimize

Science and Math Comprehension
- Physics
- Mathematics
- Biochemistry

Generative Models
- Cosmology
- Detector Simulations
- Biodesign

Inverse Problems
- Spectra 2 Structures
- Image2Phase
- Waveform 2 Source

Multimodal Learning
- Text
- Waveforms
- Images
- Graphs
- Structured
- Time-series

Decision-Making
- Research Priorities
- Risk Assessment
- The Next Problem

AI: a fundamental shift in the economic and military landscape

- Executive Order brings focus to national strategy
- Industry focuses on AI-based products for business, especially social, financial, health and security
- Universities focus on basic research and education
- DOE has a unique role
  - Mission-driven development and application of AI/ML, i.e., innovation in, for example
    - *Science*
    - *Energy*
    - *National security*
  - Build on its HPC mission
  - Large-scale scientific data for research
  - Talent development

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

Thanks!

Questions?

DOI: 10.2172/1604756

U.S. DEPARTMENT OF ENERGY | Office of Science