AI for Science

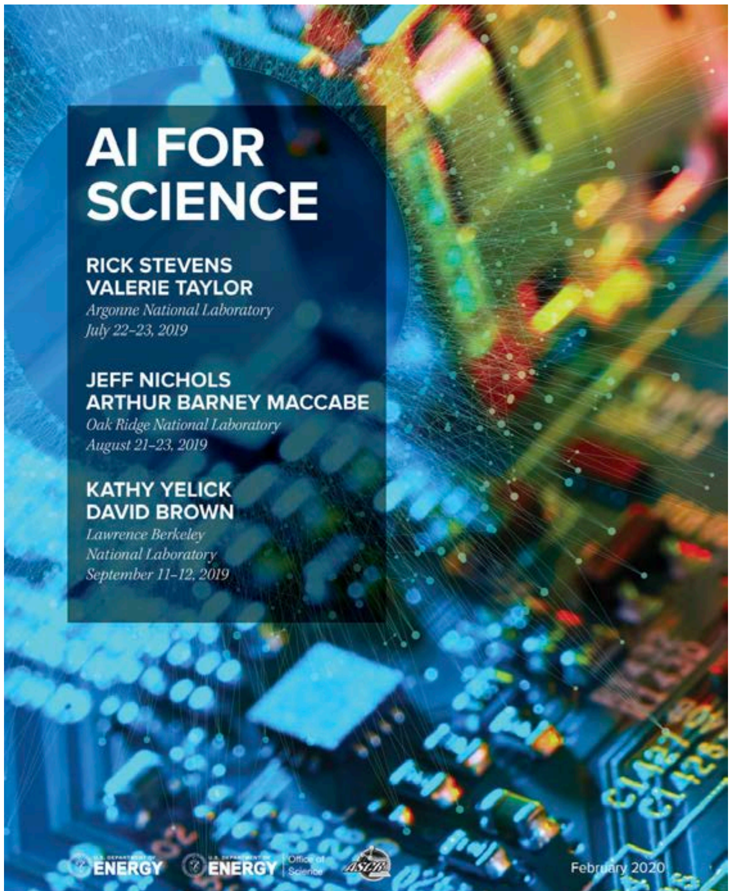
Rick Stevens, Valerie Taylor, ANL

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

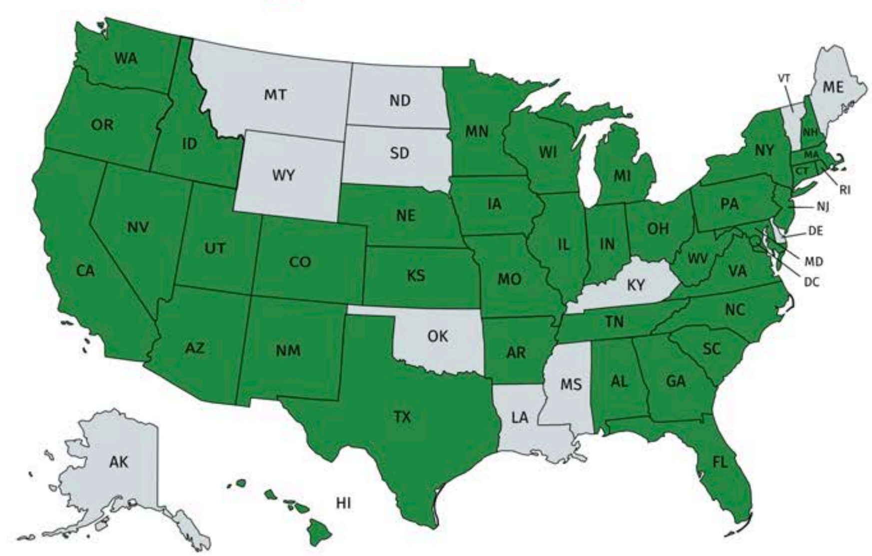


- Over 1,000 scientists participated in four town halls during the summer of 2019
- Research Opportunities in AI
  - Biology, Chemistry, Materials,
  - Climate, Physics, Energy, Cosmology
  - Mathematics and Foundations
  - Data Life Cycle
  - Software Infrastructure
  - Hardware for AI
  - Integration with Scientific Facilities
- Modeled after the Exascale Series in 2007
- ASCAC subcommittee Report due May 2020

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	

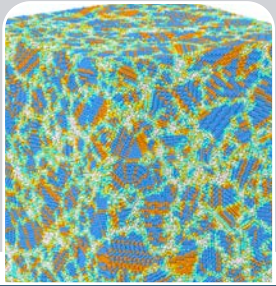


- All 17 DOE National Laboratories
- 39 Companies from large and small
- Over 90 different universities
- 6 DOE/SC Offices + EERE and NNSA

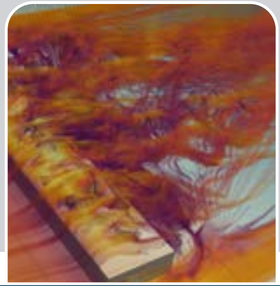
Plenary Presentations at the Town Halls

Argonne

Cosmology
Salman Habib

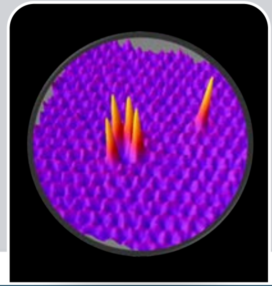


Materials
Ian Foster



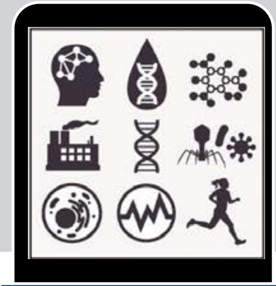
Climate
Rao Kotamarthi

Oak Ridge



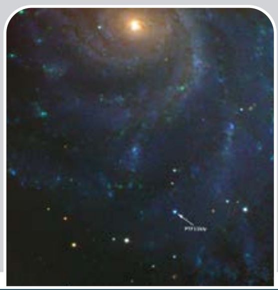
Microscopy
Sergei Kalinin

Manufacturing
Tom Kurfess



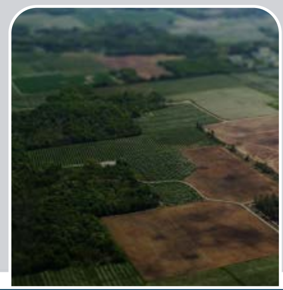
Health
Gina Tourassi

Berkeley



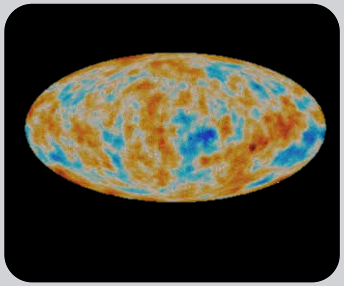
Astrophysics
Josh Bloom

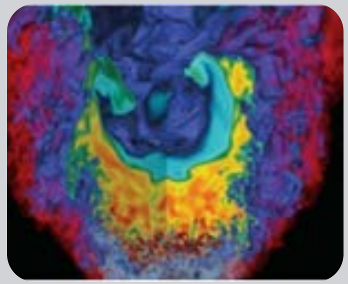
User Facilities
James Sethian



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

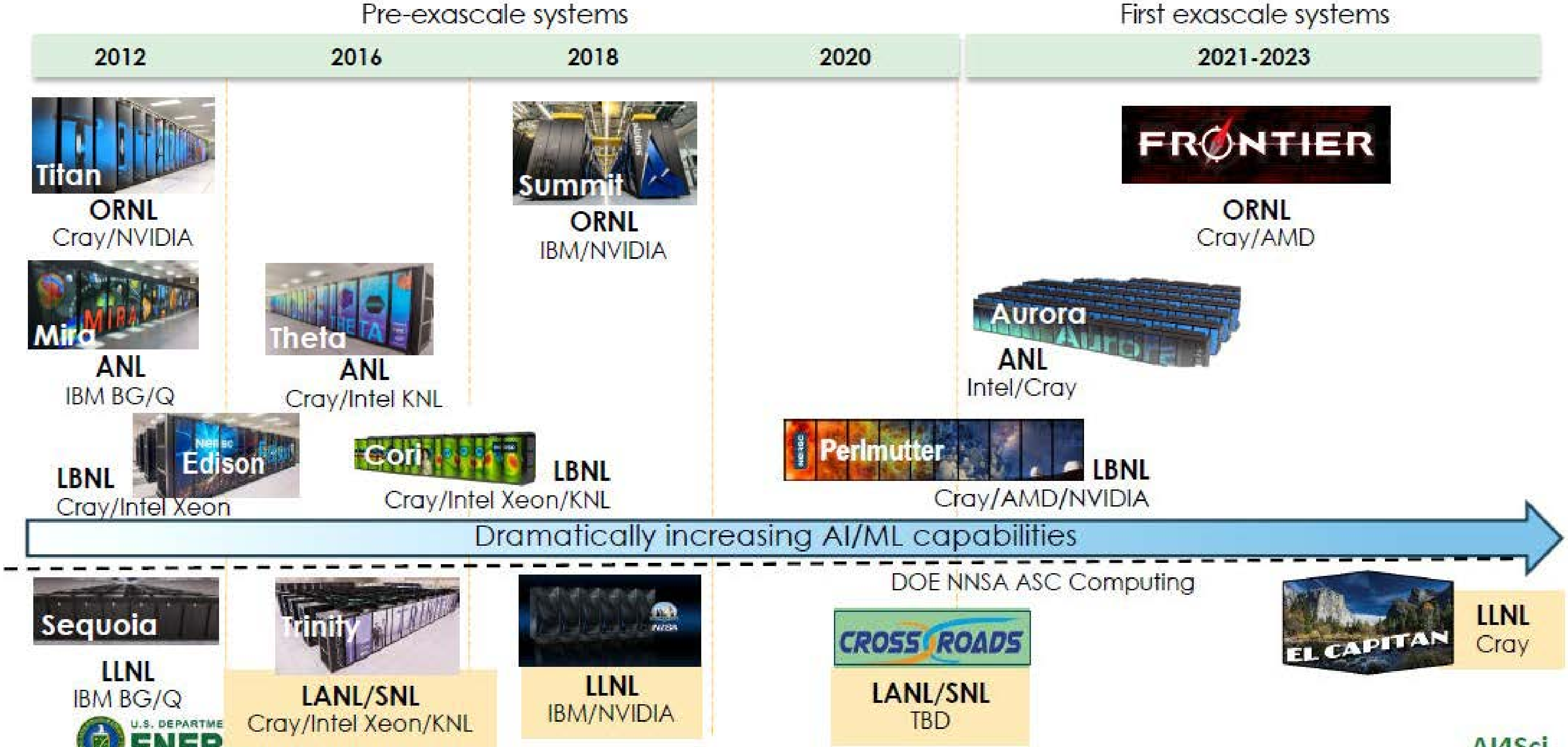
Theory becomes data for next-generation AI

- AI begins to contribute to advancing theory

AI becomes a common part of scientific laboratory activities

- AI is integrated into science, engineering and operations

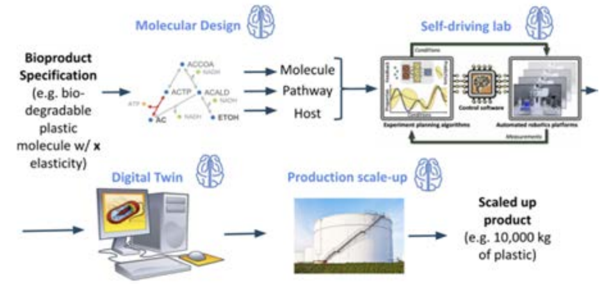
DOE is building on a record of success delivering HPC capabilities

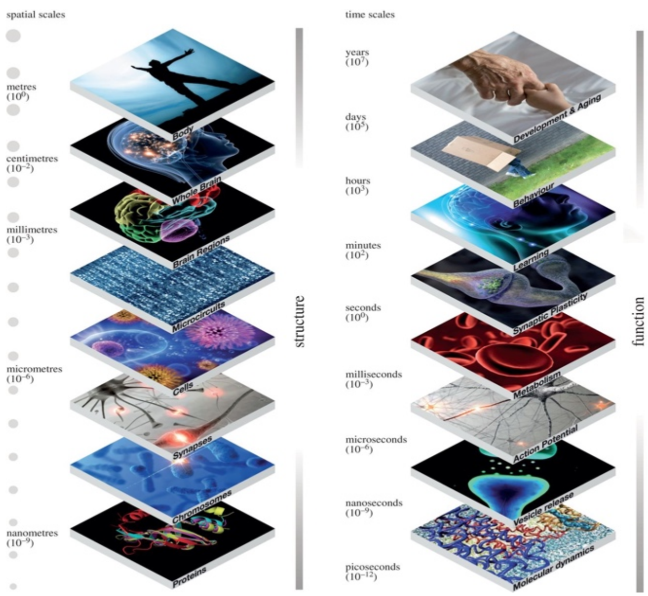


U.S. DEPARTMENT OF ENERGY | Science

AI4Sci

Biology and Life Sciences





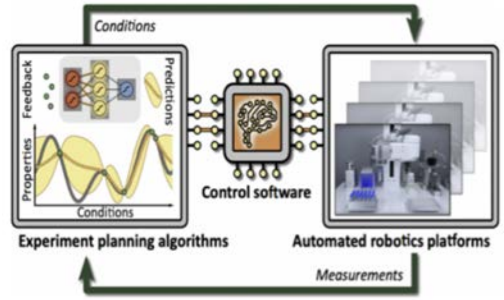
Biological systems, including humans, constitute the integration of many levels of spatiotemporal organization.

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

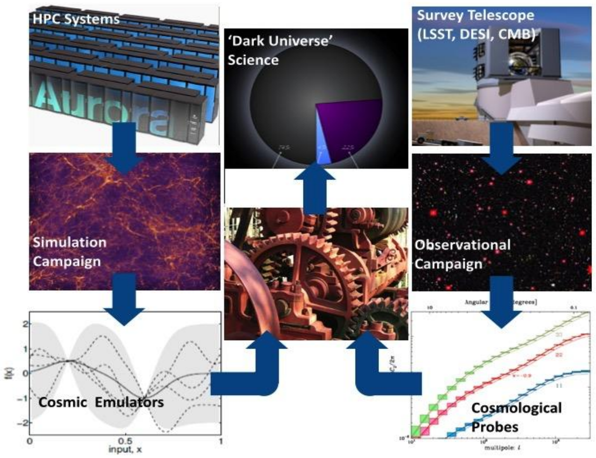
Discovering the ultimate constituents of matter and uncovering the nature of space and time

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:

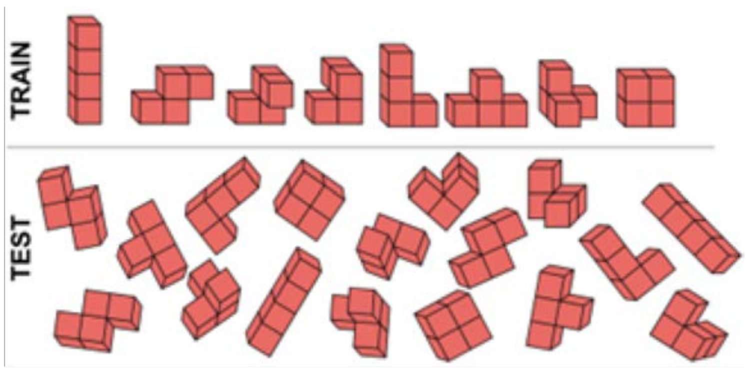
- Enable the exploration of the data from the next-generation surveys
- Make a movie of the universe from its earliest moments until today
- A new era of precision physics at the Energy and Intensity frontiers

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, NeurlPS18 [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 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…

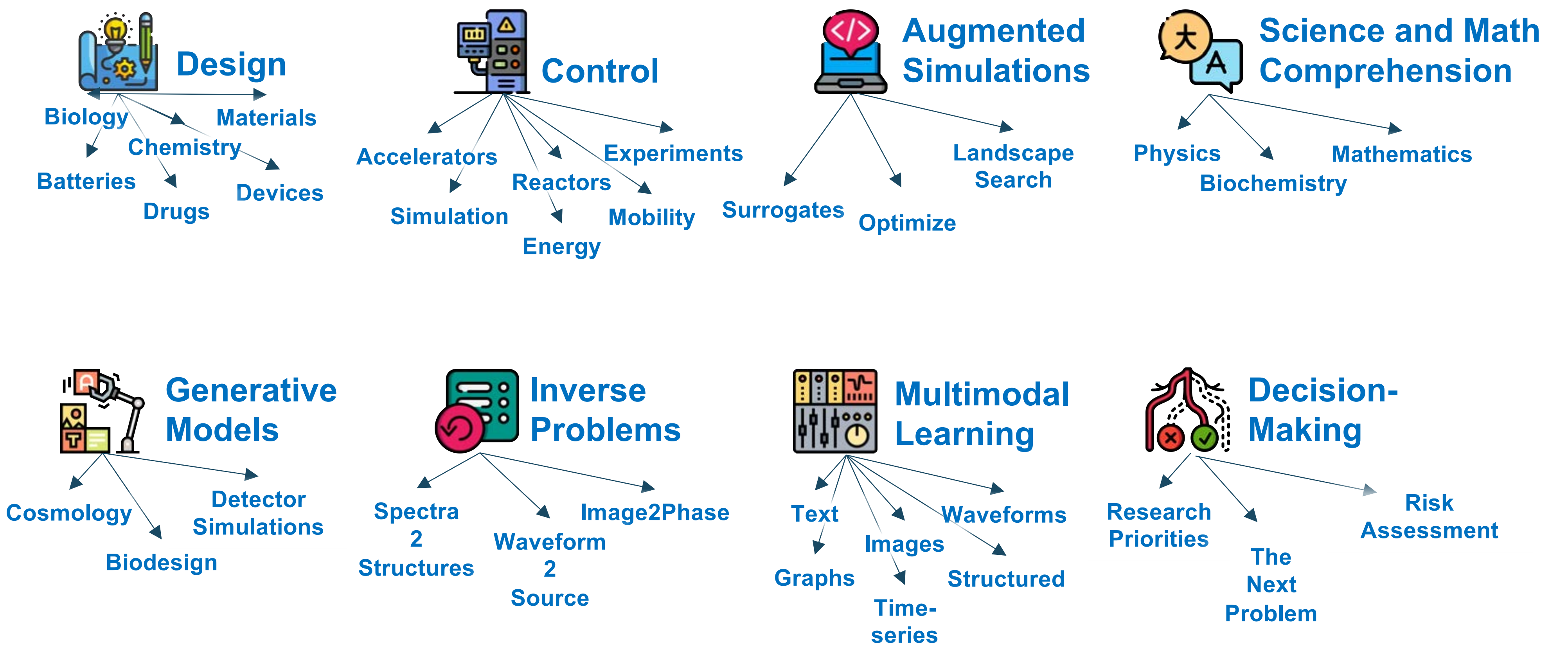
AI Enabled
Scientific
Comprehension
(what it means)

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

→ *Cleaned*
Updated
Annotated
Aggregated
Interpreted
→ Insight?

AI4Sci

AI for Science: AI Building Blocks (examples)



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!

U.S. DEPARTMENT OF ENERGY | Office of Science

The Networking and Information Technology Research and Development (NITRD) Program

Mailing Address: NCO/NITRD, 2415 Eisenhower Avenue, Alexandria, VA 22314

Physical Address: 490 L'Enfant Plaza SW, Suite 8001, Washington, DC 20024, USA Tel: 202-459-9674, Fax: 202-459-9673, Email: nco@nitrd.gov, Website: https://www.nitrd.gov