HPC+ BD + ML
-> Discovery and Innovation

Jim Kurose
Assistant Director, NSF
Directorate for Computer & Information Science & Engineering

Workshop on the Convergence of High Performance Computing, Big Data, and Machine Learning
October 29-30, 2018

WELCOME
THANKS!
Why are we here?

HPC+ BD + ML ->

Science discovery

Health and Welfare

National Defense, National Security

Aligned with Administration Priorities

FY 2020 R&D Budget Priorities Memo

“Agencies should invest in fundamental and applied AI research, including machine learning, autonomous systems, and applications at the human-technology frontier. ... Agencies should prioritize investment in research and infrastructure to maintain U.S. leadership in strategic computing, from edge devices to high-performance computing, ... use of embedded sensors, data analytics, and machine learning.”

National Security Strategy

“prioritize emerging technologies critical to economic growth and security, such as data science, encryption, autonomous technologies, advanced computing technologies, and artificial intelligence.”

National Defense Strategy

“New technologies include advanced computing, “big data” analytics, artificial intelligence, autonomy, robotics, ..”
In addition to scientific modeling and simulation, Summit offers unparalleled opportunities for the integration of AI and scientific discovery, enabling researchers to apply techniques like machine learning and deep learning to problems in human health, high-energy physics, materials discovery ...
HPC + BD + ML: NSF Harnessing The Data Revolution

**Research** across all NSF Directorates

- **Theoretical foundations**
  - Transdisciplinary Research in Principles of Data Science (TRIPODS)

- **Systems foundations**
  - Data-centric algorithms, systems: BIGDATA

- **Data-intensive research**
  - across all science & engineering
  - TRIPODS+X

**Educational pathways**

Innovations grounded in an education-research-based framework
- NASEM: undergraduate data science
- NSF Research Traineeship
- NSF Graduate Research Fellowship Program
- Data Science Corps

**Advanced cyberinfrastructure**

Accelerating data-intensive research.
- Cyberinfrastructure for Sustained Scientific Innovation (CSSI);
- Scalable data-driven Cyberinfrastructure Dear Colleague Letter (DCL);
- Midscale infrastructure (RFI and DCL)

“... bold questions that will drive NSF's long-term research agenda -- questions that will ensure future generations continue to reap the benefits of fundamental S&E research.”

“AI is the universal connector that interweaves all of our Big Ideas; data science is changing the very nature of scientific inquiry, and AI’s use of data has the potential to revolutionize everything we do in science.”

F. Cordova, Director, NSF, Sept. 2017
HPC + BD + ML: NSF Harnessing The Data Revolution

HDR Convergence Accelerator

- Translational, use-inspired convergence research in HDR
- Projects with clear goals, milestones, directed deliverables (e.g., 6-months)
- More intentional, directed management
- Mission-driven evaluation, rather than peer review
- Partnerships: co-funding, co-design, creation
- FY 2019 launch

HPC + BD + ML: Open Knowledge Network

**Goal:** design, develop, prototype, and demonstrate an open knowledge network – an open semantic information infrastructure based on open standards for creating and maintaining a knowledge graph to enable discovery of non-trivial knowledge from multiple disparate knowledge sources, covering thousands of topic areas, especially scientific information

NITRD Workshops on an Open Knowledge Network: https://www.nitrd.gov/nitrdgroups/index.php?title=Open_Knowledge_Network
HPC + BD + ML: Artificial Intelligence & Data

Data (big, and small) often sensed in real-time

IOT  Scientific Instruments  Health

Training, baseline data

HPC + BD + ML: Artificial Intelligence & Data

Distributed Data, Privacy, Security, Data-Sharing

Network, distributed data  Data discovery, access  Security, Privacy

Autonomy  Human-AI interaction

AI Infrastructure

Modeling

Machine Learning Components

Massive Data Management

Sensing / Data Acquisition
### HPC + BD + ML: Artificial Intelligence & ML

- **Deep learning, classification**
- **Reinforcement learning**
- **Transfer learning**
- **Representation learning**
- **Semi-supervised learning**
- **Fairness, Accountability**
- **Transparency, Causality**
- **Hypothesis generation**

<table>
<thead>
<tr>
<th>Machine Learning Components</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sensing / Data Acquisition</td>
</tr>
<tr>
<td>Massive Data Management</td>
</tr>
<tr>
<td>Machine Learning Components</td>
</tr>
<tr>
<td>AI Infrastructure</td>
</tr>
<tr>
<td>Modeling</td>
</tr>
<tr>
<td>Autonomy</td>
</tr>
<tr>
<td>Human-AI interaction</td>
</tr>
</tbody>
</table>

### HPC + BD + ML: Artificial Intelligence & Modeling

- Machine learning, with domain-specific models
- **Human emotional state**
- **Domain models:** classifying “interesting” astronomical events

<table>
<thead>
<tr>
<th>Knowledge Representation</th>
</tr>
</thead>
<tbody>
<tr>
<td>AI Infrastructure</td>
</tr>
<tr>
<td>Modeling</td>
</tr>
<tr>
<td>Machine Learning Components</td>
</tr>
<tr>
<td>Massive Data Management</td>
</tr>
<tr>
<td>Sensing / Data Acquisition</td>
</tr>
</tbody>
</table>

**Human emotional state**

**Domain models:** classifying “interesting” astronomical events
HPC + BD + ML : Artificial Intelligence & Infrastructure

- People, organizations, & communities
- Software, Algorithms
- Data
- Knowledge Network
- Networking
- Computational Resources
- Scientific Instruments
- Storage
- Sensing / Data Acquisition
- Massive Data Management
- Machine Learning Components
- Modeling
- AI Infrastructure
- Networking
- Scientific Instruments
- Storage
- Sensing / Data Acquisition
- Massive Data Management
- Machine Learning Components
- Modeling
- AI Infrastructure
- Networking
- Scientific Instruments
- Storage
- Sensing / Data Acquisition
- Massive Data Management
- Machine Learning Components
- Modeling
- AI Infrastructure
- Networking
- Scientific Instruments
- Storage
- Sensing / Data Acquisition
- Massive Data Management
- Machine Learning Components
- Modeling
- AI Infrastructure
- Networking
- Scientific Instruments
- Storage
- Sensing / Data Acquisition
- Massive Data Management
- Machine Learning Components
- Modeling
- AI Infrastructure
- Networking
- Scientific Instruments
- Storage
- Sensing / Data Acquisition
- Massive Data Management
- Machine Learning Components
- Modeling
- AI Infrastructure
- Networking
- Scientific Instruments
- Storage
- Sensing / Data Acquisition
- Massive Data Management
- Machine Learning Components
- Modeling
- AI Infrastructure
- Networking
- Scientific Instruments
- Storage
- Sensing / Data Acquisition
- Massive Data Management
- Machine Learning Components
- Modeling
- AI Infrastructure
- Networking
- Scientific Instruments
- Storage
- Sensing / Data Acquisition
- Massive Data Management
- Machine Learning Components
- Modeling
- AI Infrastructure
- Networking
- Scientific Instruments
- Storage
- Sensing / Data Acquisition
- Massive Data Management
- Machine Learning Components
- Modeling
- AI Infrastructure
- Networking
- Scientific Instruments
- Storage
- Sensing / Data Acquisition
- Massive Data Management
- Machine Learning Components
- Modeling
- AI Infrastructure
- Networking
- Scientific Instruments
- Storage
- Sensing / Data Acquisition
- Massive Data Management
- Machine Learning Components
- Modeling
- AI Infrastructure
- Networking
- Scientific Instruments
- Storage
- Sensing / Data Acquisition
- Massive Data Management
- Machine Learning Components
- Modeling
- AI Infrastructure
- Networking
- Scientific Instruments
- Storage
- Sensing / Data Acquisition
- Massive Data Management
- Machine Learning Components
- Modeling
- AI Infrastructure
- Networking
- Scientific Instruments
- Storage
- Sensing / Data Acquisition
- Massive Data Management
- Machine Learning Components
- Modeling
- AI Infrastructure
- Networking
- Scientific Instruments
- Storage
- Sensing / Data Acquisition
- Massive Data Management
- Machine Learning Components
- Modeling
- AI Infrastructure
- Networking
- Scientific Instruments
- Storage
- Sensing / Data Acquisition
- Massive Data Management
- Machine Learning Components
- Modeling
- AI Infrastructure
- Networking
- Scientific Instruments
- Storage
- Sensing / Data Acquisition
- Massive Data Management
- Machine Learning Components
- Modeling
- AI Infrastructure
- Networking
- Scientific Instruments
- Storage
- Sensing / Data Acquisition
- Massive Data Management
- Machine Learning Components
- Modeling
- AI Infrastructure
- Networking
- Scientific Instruments
- Storage
- Sensing / Data Acquisition
- Massive Data Management
- Machine Learning Components
- Modeling
- AI Infrastructure
- Networking
- Scientific Instruments
- Storage
- Sensing / Data Acquisition
- Massive Data Management
- Machine Learning Components
- Modeling
- AI Infrastructure
- Networking
- Scientific Instruments
- Storage
- Sensing / Data Acquisition
- Massive Data Management
- Machine Learning Components
- Modeling
- AI Infrastructure
- Networking
- Scientific Instruments
- Storage
- Sensing / Data Acquisition
- Massive Data Management
- Machine Learning Components
- Modeling
- AI Infrastructure
- Networking
- Scientific Instruments
- Storage
- Sensing / Data Acquisition
- Massive Data Management
- Machine Learning Components
- Modeling
- AI Infrastructure
- Networking
- Scientific Instruments
- Storage
- Sensing / Data Acquisition
- Massive Data Management
- Machine Learning Components
- Modeling
- AI Infrastructure
- Networking
- Scientific Instruments
- Storage
- Sensing / Data Acquisition
- Massive Data Management
- Machine Learning Components
- Modeling
- AI Infrastructure
- Networking
- Scientific Instruments
- Storage
- Sensing / Data Acquisition
- Massive Data Management
- Machine Learning Components
- Modeling
- AI Infrastructure
- Networking
- Scientific Instruments
- Storage
- Sensing / Data Acquisition
- Massive Data Management
- Machine Learning Components
- Modeling
- AI Infrastructure
- Networking
- Scientific Instruments
- Storage
- Sensing / Data Acquisition
- Massive Data Management
- Machine Learning Components
- Modeling
- AI Infrastructure
- Networking
- Scientific Instruments
- Storage
- Sensing / Data Acquisition
- Massive Data Management
- Machine Learning Components
- Modeling
- AI Infrastructure
- Networking
- Scientific Instruments
- Storage
- Sensing / Data Acquisition
- Massive Data Management
- Machine Learning Components
- Modeling
- AI Infrastructure
- Networking
- Scientific Instruments
- Storage
- Sensing / Data Acquisition
- Massive Data Management
- Machine Learning Components
- Modeling
- AI Infrastructure
- Networking
- Scientific Instruments
- Storage
- Sensing / Data Acquisition
- Massive Data Management
- Machine Learning Components
- Modeling
- AI Infrastructure
- Networking
- Scientific Instruments
- Storage
- Sensing / Data Acquisition
- Massive Data Management
- Machine Learning Components
- Modeling
- AI Infrastructure
- Networking
- Scientific Instruments
- Storage
- Sensing / Data Acquisition
- Massive Data Management
- Machine Learning Components
- Modeling
- AI Infrastructure
- Networking
- Scientific Instruments
- Storage
- Sensing / Data Acquisition
- Massive Data Management
- Machine Learning Components
- Modeling
- AI Infrastructure
- Networking
- Scientific Instruments
- Storage
- Sensing / Data Acquisition
- Massive Data Management
- Machine Learning Components
- Modeling
- AI Infrastructure
- Networking
- Scientific Instruments
- Storage
- Sensing / Data Acquisition
- Massive Data Management
- Machine Learning Components
- Modeling
- AI Infrastructure
- Networking
- Scientific Instruments
- Storage
- Sensing / Data Acquisition
- Massive Data Management
- Machine Learning Components
- Modeling
- AI Infrastructure
- Networking
- Scientific Instruments
- Storage
- Sensing / Data Acquisition
- Massive Data Management
- Machine Learning Components
- Modeling
- AI Infrastructure
- Networking
- Scientific Instruments
- Storage
- Sensing / Data Acquisition
- Massive Data Management
- Machine Learning Components
- Modeling
- AI Infrastructure
- Networking
- Scientific Instruments
- Storage
- Sensing / Data Acquisition
- Massive Data Management
- Machine Learning Components
- Modeling
- AI Infrastructure
- Networking
- Scientific Instruments
- Storage
- Sensing / Data Acquisition
- Massive Data Management
- Machine Learning Components
- Modeling
- AI Infrastructure
- Networking
- Scientific Instruments
- Storage
- Sensing / Data Acquisition
- Massive Data Management
- Machine Learning Components
- Modeling
- AI Infrastructure
- Networking
- Scientific Instruments
- Storage
- Sensing / Data Acquisition
- Massive Data Management
- Machine Learning Components
- Modeling
- AI Infrastructure
- Networking
- Scientific Instruments
- Storage
- Sensing / Data Acquisition
- Massive Data Management
- Machine Learning Components
- Modeling
- AI Infrastructure
- Networking
- Scientific Instruments
- Storage
- Sensing / Data Acquisition
- Massive Data Management
- Machine Learning Components
- Modeling
- AI Infrastructure
- Networking
- Scientific Instruments
- Storage
- Sensing / Data Acquisition
- Massive Data Management
- Machine Learning Components
- Modeling
- AI Infrastructure
- Networking
- Scientific Instruments
- Storage
- Sensing / Data Acquisition
- Massive Data Management
- Machine Learning Components
- Modeling
- AI Infrastructure
- Networking
- Scientific Instruments
- Storage
- Sensing / Data Acquisition
- Massive Data Management
- Machine Learning Components
- Modeling
- AI Infrastructure
- Networking
- Scientific Instruments
- Storage
- Sensing / Data Acquisition
- Massive Data Management
- Machine Learning Components
- Modeling
- AI Infrastructure
- Networking
- Scientific Instruments
- Storage
- Sensing / Data Acquisition
- Massive Data Management
- Machine Learning Components
- Modeling
- AI Infrastructure
- Networking
- Scientific Instruments
- Storage
- Sensing / Data Acquisition
- Massive Data Management
- Machine Learning Components
- Modeling
- AI Infrastructure
- Networking
- Scientific Instruments
- Storage
- Sensing / Data Acquisition
- Massive Data Management
- Machine Learning Components
- Modeling
- AI Infrastructure
- Networking
- Scientific Instruments
- Storage
- Sensing / Data Acquisition
- Massive Data Management
- Machine Learning Components
- Modeling
- AI Infrastructure
- Networking
- Scientific Instruments
- Storage
- Sensing / Data Acquisition
- Massive Data Management
- Machine Learning Components
- Modeling
- AI Infrastructure
- Networking
- Scientific Instruments
- Storage
- Sensing / Data Acquisition
- Massive Data Management
- Machine Learning Components
- Modeling
- AI Infrastructure
- Networking
- Scientific Instruments
- Storage
- Sensing / Data Acquisition
- Massive Data Management
- Machine Learning Components
- Modeling
- AI Infrastructure
- Networking
- Scientific Instruments
- Storage
- Sensing / Data Acquisition
- Massa

HPC + BD + ML : Artificial Intelligence @ NSF/CISE

- CISE core research programs:
  - Cyber-human Systems
  - Robust Intelligence
  - Information Integration and Informatics
- Cross-directorate programs:
  - BIGDATA
  - NRI-2.0: Ubiquitous Collaborative Robots
  - Smart and Autonomous Systems
  - Smart & Connected Communities
  - Smart and Connected Health
  - Computational Neuroscience
- CISE Expeditions in Computing
- AI+X: ML as a new horizontal
- CISE/IIS budget: $210M
Cross-Government AI Activities

Office of Science & Technology Policy (OSTP)

Lynne Parker
Assistant Director for AI

National Science and Technology Council (NSTC)

Select Committee on AI

Committee on Technology

Committee on S&T Enterprise

Machine Learning and AI (MLAI)

Networking and Info. Tech. R&D (NITRD)

AI R&D Interagency Working Group

Cross-Government HPC+BD+ML: Where?

High End Computing IWG

Big Data IWG

AI IWG

HPC+ BD + ML

-> Discovery and Innovation
Cross-Government HPC+BD+ML: Here!

- Suggestions for paths forward: actionable items
  - technical directions
  - achieving integration (HPC+BD+ML)
  - partnerships (federal, industry, academia)
- Believe it or not: high-quality workshop reports do matter!
  - For impact to last beyond this meeting
  - Community voice into agency priorities, activities

THANKS!

NITRD Staff: Wendy Wigen, Ji Lee, Faisal D’Sousa

NITRD IWG leadership: Jeff Alstott, Chaitan Baru, Rance Cleaveland, Susan Gregurik, Henry Kautz, Sandy Landsberg, Barry Schneider
"Any opinions, findings, conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the Networking and Information Technology Research and Development Program."

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