Sandia National Laboratories

Exceptional service in the national interest

Sandia's Trusted Artificial Intelligence Strategic Initiative






John Feddema, Sr. Manager, Enhanced Decision Making Group

Center for Computing Research

U.S. DEPARTMENT OF ENERGY | Office of Science

NRC Data Science and AI Workshop

Tuesday, November 9, 2021

SAND2021-7619 PE

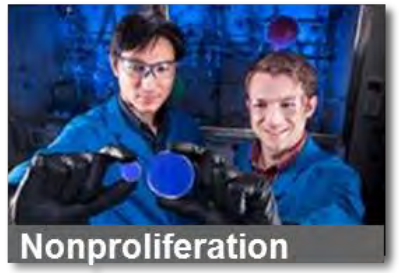
Sandia's ***Trusted AI Strategic Initiative*** is coordinating a series of fundamental R&D projects to lay the foundation necessary for Sandia's scientific and national security applications

Desire to deploy AI/ML technologies are increasing rapidly

- FY20 – 80 LDRD projects (roughly 20%) had a significant AI focus
- FY21 – 126 projects (28%) had a significant focus in AI or were significantly utilizing AI technologies
- FY21 – 9 DOE SC Advanced Scientific Computing Research projects
- FY21 – 7 NNSA Advance Simulation & Computing (ASC) Advanced Machine Learning (AML) projects

Sandia's unique mission needs set us apart from industry

- High-consequence applications require high-confidence decisions
- Solutions require extrapolation beyond the space of available data
- Many national security applications have low volume, incomplete data
- Deployed AI solutions are often in environments under extreme size, weight, and power constraints
- Decisions may need to be made in very short timeframes
- Need to account for potential adversarial issues

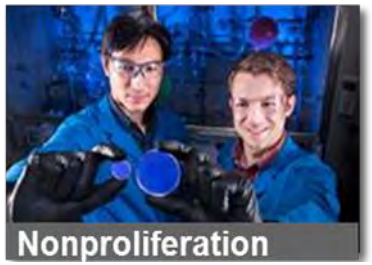


Sandia's history of excellence in core capabilities such as UQ, V&V, optimization, graphs, tensors, and discrete math will enable AI/ML

Sandia Mission Needs for Machine Learning and Artificial Intelligence

Program	Mission Problem	Characteristics
Global Security	Proliferation Detection and Characterization	• Multi-modal sensors, distributed sensors, and real-time behavior • How to extrapolate to cases where we do not have ground truth • Physical models may not exist • Real time monitoring with streaming data
Global Security	Automatic Target Recognition for Military Applications	• Limited data that is likely modified or disguised - extrapolation of models is necessary • Desire to reduce or remove human in the loop • Data available at multiple levels of sensitivity • Adversary withholds differentiating capabilities and tactics exclusively for war
Nuclear Deterrence	Counterfeit and Aging Detection	• Many sources of variation - limits to what can be learned from data are unknown • Lack of a mathematical foundation and physical models • Volume of data is very low
DOE Office of Science	Large Scale Physics Experiments	• Rich but sparse data - can be expensive to obtain • Multi-instrument, multi-experiment, multi-measurement experimental observations • Uncertainty present in experiments and physics models
National Security Programs	Analyst Support for Cyber and Intelligence Operations	• Need to introduce AI into a mature system without disrupting current operations • Very high consequence, very rapid transactions (many per minute) • Streaming data with very dynamic environment
Energy & Homeland Security	Bioscience and Biosecurity	• Multiple types of data requires data fusion • Data collection is often destructive and multiple measurements depend on replication • Theoretical models often don't exist
Advanced Science & Technology	HPC System Management and Operations	• Operations staff don't know much about performance/failure mechanisms • Thousands of instrumentation points but unknown if data provide useful insights • Experiments are typically one-offs due to how resources are allocated and used

Sandia's Trusted AI LDRD Research Campaign Thrusts



AI Usability and Trust

- Generalizability of AI Research to High-Consequence National Security Environments
- Trustworthiness Characteristics of Analytics
- Domain-Informed AI
- Adversarial Impact on User Trust
- Determining Whether to Automate Functionality and to What Level

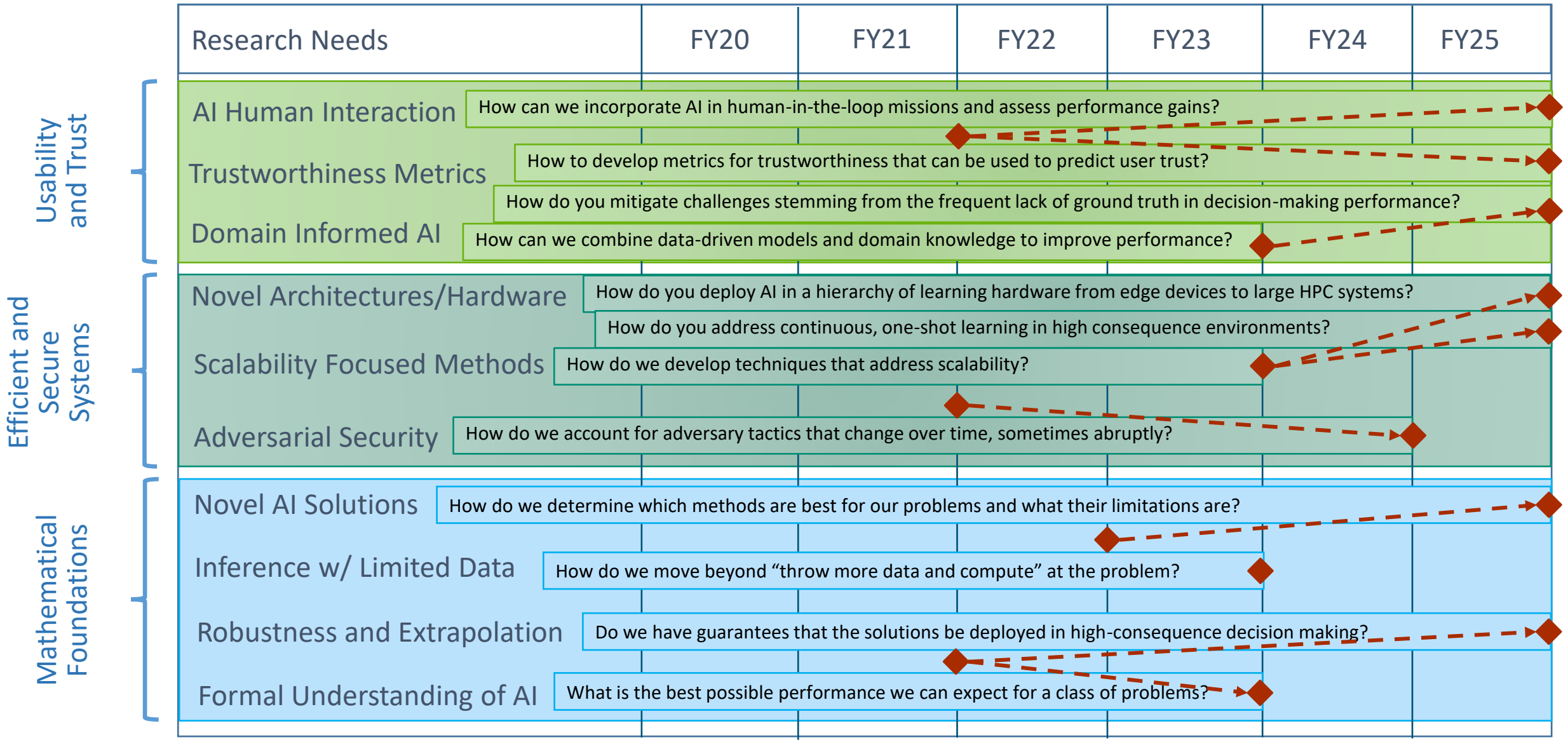
Efficient and Secure AI System

- Novel AI Architectures and Hardware
- Adversarial/Counter-Adversarial Security
- Scalability of AI Methods
- AI Software
- Domain/Architecture Aware AI Methods and Algorithms

Mathematical Foundations of AI

- Mathematical analysis and Directed Improvement of AI Methods
- Acceleration of Training and Hyperparameter Tuning
- Novel AI Solutions
- Robustness & Extrapolation
- Statistical Inference, Especially with Limited Data
- Randomized Methods with Rigorous Probabilistic Guarantees

Successes in Trusted AI will enable Sandia and its mission partners to think differently about current and future mission problems

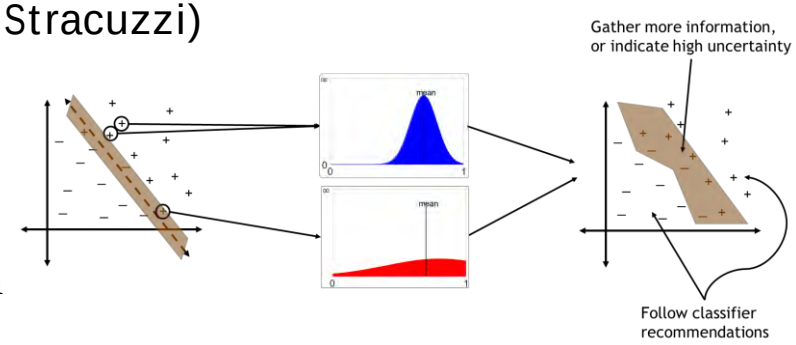


FY21 Trusted AI LDRD Highlights

Usability & Trust

Optimizing Machine Learning Decisions with Prediction Uncertainty (PI: David Stracuzzi)

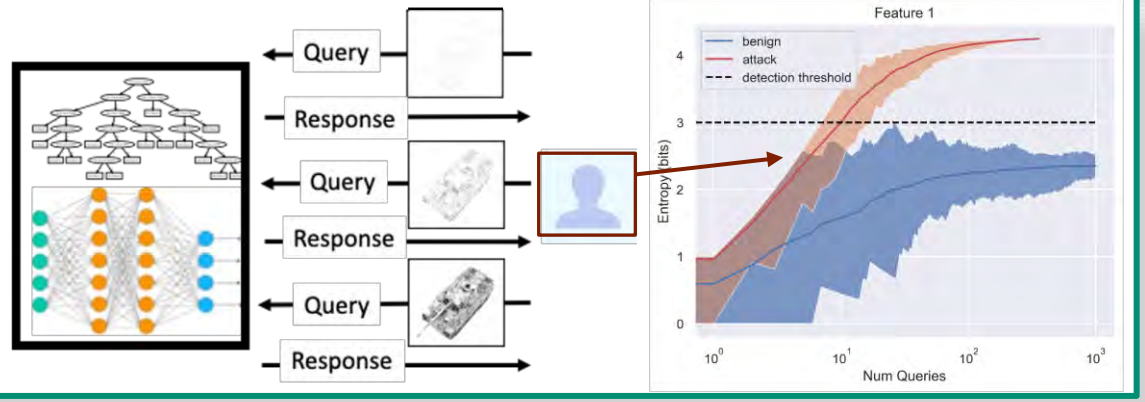
- **Challenge:** The ML prediction problem is often confounded with the decision problem
- **Goal:** Incorporate decision science into ML-based decision making
  - Develop rigorous methods for incorporating prediction uncertainty, error costs, and opportunities to gather additional information to minimize decision errors and costs.
- **Proposed Solution**: Draw on decision science, uncertainty quantification, and information theory to account for possible outcomes and their associated probabilities.



Efficient & Secure Systems

Monitoring Online Adversarial Tampering (PI: Gary Saavedra)

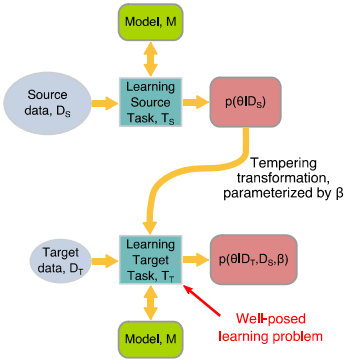
- **Challenge:** Defend against adversarial attacks
  - Little work on detection and less for streaming models
- **Proposed Solution**: Distinguishing factors of our work:
  - Detection rather than model alteration
  - Stream information provides more insight than lone examples
  - Unifies several different attacks into one mathematical framework



Mathematical Foundations

Trust-Enhancing Probabilistic Transfer Learning for Sparse and Noisy Data (PI: Mohammad Khalil)

- **Challenge**: Many Sandia mission domains are defined by a lack of reliable data, effectively precluding the use of many modern deep learning/machine learning techniques for predictive modeling
- **Goal**: Enhance the trust in machine learning (ML) model predictions within sparse & noisy data settings
- **Proposed Solution**: Novel probabilistic transfer learning (TL) framework:
  - Determine when to apply TL, which model to use, and how much (uncertain) knowledge to transfer using new techniques inspired by Bayesian hierarchical modeling, sequential data assimilation, and uncertainty quantification



Thank you for your attention!

For more information, please contact:

John Feddema, 505-844-0827, jtfedde@sandia.gov

Sandia National Laboratories

Exceptional service in the national interest