Preliminary findings of the SEAB to Secretary of Energy Dan Brouillette regarding the Department of Energy and Artificial Intelligence.

Preliminary Findings

SEAB AIML Working Group
12 March 2020

Secrtary of Energy Advisory Board (SEAB)

Contents

Executive Summary

This document contains the preliminary findings recommended to the Secretary of Energy Advisory Board (SEAB) from a working group dedicated to DOE's capabilities and future in Artificial Intelligence and Machine Learning (AI/ML). The preliminary findings are based on expert opinion and information provided to the SEAB working group regarding DOE AI efforts. The preliminary findings were developed under the leadership of two SEAB members, Dr. Samantha Ravich and Dr. Thomas Rosenbaum who served as co-chairs of the working group.

AI Working Group Co-Chair’s Letter

On August 4, 1977, the Department of Energy Organization Act was passed by Congress**.** The underlying rationale for the creation of the new department was that the United States faced an increasing shortage of nonrenewable energy and that this shortage, coupled with our growing dependence on foreign supplies, presented a serious threat to the national security of the United States and to the health, safety, and welfare of its citizens. To remedy this, the DOE was established to bring a comprehensive, centralized focus to energy policy, regulation, and research, development, and demonstration. The Act underscored the Department’s mission of delivering scientific discoveries, capabilities, and major scientific tools to transform the understanding of nature and to advance the energy, economic, and national security of the United States. Recognizing that adversaries would seek to undermine the country’s scientific research and energy security, Congress ensured that the new department would be properly resourced to implement its security, counterintelligence, and intelligence policies.

Forty-three years later, we are faced with a similar challenge to our national security innovation and industrial base. As important as energy resources and security were – and are—to our national security and economic prosperity, so too is the revolution in Artificial Intelligence (AI) and Machine Learning (ML). Advances in AI and ML are rapidly suffusing all parts of science and technology, as well as promising enhancements in operational efficiencies. Deep learning techniques make discoveries extendable across economic sectors, from harnessing new ways to create renewable energy to optimizing delivery routes to reading medical images to predicting protein folding pathways for drug discovery. The key drivers are large datasets, computational capacity, and mathematical/algorithmic development, all of which are key capabilities of the DOE. Access to cost efficient and reliable energy supplies is critical as well, given the significant energy resources needed for computing and storage.

The United States is again in a race to ensure that we and our allies are preeminent producers of the science and technology that will drive the world’s future advantages. The outcome of this race is not assured. China plans to spend $30 billion per year to achieve the leadership position in AI by 2030, while current estimates anticipate a US governmental level of spending outside of defense at approximately $2 billion per year by 2022.

The American people and the US government, however, have a tremendous resource at their disposal. The DOE national laboratory system has the potential to lead the world in the AI effort, if properly authorized and funded.

With the given existing and planned investment in DOE-aligned and National Lab facilities that generate enormous data sets and exascale computers that can crunch the data, opportunities range from AI-designed workflow, whether in large-scale scientific projects or infrastructure and procurement, to AI-enabled scientific “comprehension,” i.e. grappling with causality and deriving scientific law. The DOE Labs may be the only places that can link high performance computing machines to discovery machines such as the Advanced Photon Source (Argonne) and the Spallation Neutron Source (Oak Ridge) to permit *in situ* machine learning on data and flow. They also have the potential to bring together computer scientists with domain experts on a scale that industry and academia cannot. The National AI Directive calls for efforts to enhance access to high quality cyberinfrastructure and data within a framework that preserves the security and capability of the country's artificial intelligence infrastructure.

The DOE laboratories are a national asset that is a leading hope for generating essential sources of data for AI development outside of the more limited realm of search engines associated with big tech. To be successful, it is essential that bright lines be drawn between controlled data (classified, ITAR, HIPPA) and unclassified data, with minimal gray areas in between per National Security Decision Directive 189.

Prior to the creation of DOE, energy policy and authorities were spread throughout the government, leading to dangerous gaps in our national security infrastructure. The Department was authorized to rectify that peril and to serve as resource to the rest of government.

Over the last four months, the SEAB AI Working Group has assessed the comparative advantages of both the Department and the National Lab complex in helping the country secure its future in an AI-empowered world. We found that to meet this critical national need, the Department of Energy must create an enterprise-wide capability to accelerate the use of AI for scientific discovery at scale while fostering fundamental advancements in AI – all within a secure architecture. This approach is summarized in the phrase: AI for Science and Science for AI. This includes the sciences associated with all aspects of the energy mission, the pure sciences, and the sciences that support DOE's security mission. It also means hosting a national architecture and database that can be accessed by other elements of the USG in the service of their particular missions.

The 2019 Executive Order on Artificial Intelligence stated that, "It is the policy of the United States Government to sustain and enhance the scientific, technological, and economic leadership position of the United States in AI R&D and deployment through a coordinated Federal Government strategy." We believe that the Department of Energy is poised to play a critical role in this strategy as outlined in the attached findings.

Respectfully submitted by SEAB members:

Dr. Samantha Ravich	*Chair of the Center on Cyber and Technology Innovations, Foundation for Defense of Democracies*
Dr. Thomas Rosenbaum	*President, California Institute of Technology*

Introduction

The Secretary of Energy Advisory Board (SEAB) has created a working group dedicated to a charge from Secretary Perry and supported by Secretary Brouillette. The charge states:

> Artificial Intelligence (AI) is a growing force across many sectors of industry, academia and government. The Department of Energy has a long history of leadership in AI and is well positioned to broaden the frontiers of this technology. In light of this, please consider how DOE can best apply artificial intelligence to DOE's energy, science and national security missions and to the effective operation of its enterprise.
>
> Please review mission needs and national capabilities related to next generation artificial intelligence technologies and methods and AI's importance to the DOE. In doing so, please examine the important problems and opportunities that drive the need for next generation AI, as well as the steps by which DOE can create and implement next generation AI, to include NNSA programmatic and budgetary efforts in the field.
>
> The report in response to this charge should include recommendations on whether and to what extent the U.S. Government should lead and accelerate the development of next generation AI technologies and systems.
>
> I request that the SEAB constitute a working group comprised of SEAB members and outside experts to address this charge and to advise me accordingly.
>
> **Purpose of the Working Group:** The SEAB AI Working Group should examine and report on the following:
>
> 1. DOE's opportunities for leading, driving, applying, and promoting AI technologies;
> 2. The growing importance of AI to the breadth of DOE mission, operational; and business functions;
> 3. Workforce development and opportunities that may include, but not be limited to, veterans, apprenticeships, and retraining programs;
> 4. Fundamental and applied research needs across the breadth of technologies including data, sensors, autonomy, proving/deciding, and human/AI interfaces;
> 5. Broader societal benefits from an open AI program and potential market barriers inhibiting private development such as data and technology gaps;
> 6. Related needs in adjacent areas such as data sciences;
> 7. The relationship between current state of technology and plans for the DOE AI effort and other federal agencies; and,
> 8. The need for effective private sector and/or international partnerships that could advance AI based systems beyond current CPU and GPU approaches.
>
> Rick Perry

Artificial Intelligence Working Group Members

The SEAB has designated two members to co-chair the Artificial Intelligence / Machine Learning (AI/ML) Working Group. They are:

- Dr. Samantha Ravich – *Chair of the Center on Cyber and Technology Innovations, Foundation for Defense of Democracies*
- Dr. Thomas Rosenbaum – *President, California Institute of Technology*

Samantha Ravich, Thomas Rosenbaum, Dave Baggett and Kurt Heckman comprise a steering committee to help manage the Artificial Intelligence Working Group. Dave Baggett represents the non-SEAB members of the Working Group as an industry leader, and Kurt Heckman is the DOE's Designated Federal Officer for SEAB.

The AIML Working group members are:

- Dave Baggett
- Alok Choudhary
- Micah Clark
- Erin Hahn
- Jill Hruby
- Ashley Llorens
- Jason Matheny
- Adm. Chuck Munns
- John Piorkowski
- Dave Ward

Preliminary Findings

This section contains the preliminary findings from the Secretary of Energy's Advisory Board (SEAB) Artificial Intelligence and Machine Learning Working Group. The process leading up to this report is as follows:

1. The Secretary of Energy issued a charge to the SEAB regarding the topic of Artificial Intelligence.
2. A working group was formed under the SEAB. The working group includes two members of the SEAB, Samantha Ravich and Thomas Rosenbaum, and a complement of leaders from industry and academia.
3. A steering committee from the working group developed a list of questions that was sent throughout the DOE.
4. The responses to those questions were compiled into a document that is an appendix in this document.
5. The working group used the DOE responses, their knowledge of the DOE mission and capabilities, and the working group members' own experience to develop these initial findings.

Motivation

Leading the world in scientific discovery is vital to the prosperity and security of our nation. Just as computing has changed every aspect of scientific inquiry, advances in artificial intelligence and machine learning are again revolutionizing how we do science. Maintaining world leadership in scientific discovery will require creating, sustaining and continually advancing a national ecosystem for AI-enabled scientific discovery. At the same time, advances must be made in AI approaches themselves in order to facilitate scientific discovery at speed and scale. These two activities form a critical virtuous cycle for scientific leadership on the world stage.

AI: the new SPACE RACE

In a similar vein as the American space-race with Russia, America is now in a new and perhaps more vital race with China regarding data science, datamining, and artificial intelligence. Because of this, our nation requires a comprehensive enterprise capability to accelerate national AI advantages and protect DOE's energy and science equities. The DOE is uniquely positioned to take this on because of the existing, unique and leading lab structure and research capabilities, large managed research teams, and the existing expertise in fundamental and enabling disciplines (algorithms, AI, computational modeling, high-performance computing, etc.). The DOE was originally founded to address a threat to national security based on the energy supply, the Energy Crisis. We are now in AI crisis where data is the new energy.

Some Observations

The AI working group formed some preliminary observations as it interacted with DOE leadership and four of its national laboratories over two multi-day field trips:

- DOE lacks a mechanism for quick-starting AI-related activities and for leveraging lessons learned and technologies developed by prior efforts; instead, many AI-related projects appear to start from the proverbial 'blank sheet of paper' with little maturity and specificity regarding suitable candidate technologies and intended end-state capabilities.
- DOE project managers and principle investigators expressed difficulties in collaborating with other activities and in learning from previous AI-related efforts.
- DOE staff expressed the need for better and more formal AI infrastructure (e.g., common labeling, shared tools and algorithms, listing of SME's, visibility into other efforts).
- DOE does not have a cohesive, enterprise-wide AI workforce; the department's current AI practitioners are smart, well intended, scientific and technically trained individuals whose primary expertise is not AI and whose work in AI is an additional duty on top of their primary responsibilities.
- Cybersecurity appears to be only a secondary or tangential concern for current AI-related DOE projects.
- Over the near-term DOE should shepherd two main categories of activities:
    - Discovery (Basic and Fundamental Science) … those investigations that enhance and expand the knowledge, techniques, and procedures of AI. We call this "*The science of AI*".
    - Application (Applied Science) … those projects that use AI capabilities to provide some meaningful and improved outcome for DOE missions. We call this "*AI for Science*"
    - Both of these activities will encompass:
        - a variety of "scope": local to a DOE unit; enterprise wide across DOE; or whole of government.
        - Various success metrics: Scientific accomplishment; cited papers; DOE site effectiveness or efficiency; commercialization; impact on our other agencies and our Nation's national security, economic security, environmental security or our citizens quality of life.
- AI operational activities, and maybe basic science, can be segment by "use cases". This segmentation may make the administration and management more effective and efficient. Some of these "use cases" could include:
    - Astrophysics exploration of the cosmos
    - Accelerators and imaging techniques of the very small and fast
    - Energy grid management
    - Cyber protections and personal identification

    - Weather risk and mitigation
    - Material engineering
    - Autonomous vehicles and hypersonics
    - Organizational back office efficiencies: finance, HR, budgeting, safety, risk management et cetera
    - Drug discovery, manufacturing, marketing and distribution
    - Communication: transmission/receipt of data, direction, instruction et cetera
    - Speech and vision recognition
    - Target detection
- Some important issues in delivering AI capability seem to be:
    - Develop enabling structures:
        - Data: Training and operational data
        - Models: System, operational landscape, environmental
        - Analytics: algorithms, techniques, procedures
    - Distribute and manage these enabling structures: develop, advertise, test, select, certify, deploy, update, assess
    - Develop trust in the AI capability: language for discussion, measure, test, assure, indemnify
    - A measured description of AI maturity levels: 1) safe secure efficient operations; 2) trusted teaming and faster thought; 3) better decisions; 4) extra human innovations
    - Ethical dimensions… intended outcomes and unintended consequences:
        - As Kissenger, Schmidt, Huttenlucker suggest in Atlantic Monthly August 2019, "AI will change human perception/cognition/interactions; we will need explainable and understandable output; we'll see an evolution of judgment; nation states may be driven to "preemption" to mitigate unacceptable AI effects"
        - The DOE should begin with a review the Defense Innovation Board's recent ethical principles for AI, which the JAIC will incorporate.
        - New discoveries in AI methods will lead to the potential for increased ability to conform to ethical standards.
- DOE is uniquely positioned to enable and guide AI research operational use because of:
    - Basic research facilities and scientists/technicians who appreciate AI
    - Many AI appropriate "use cases"
    - Logical agency for AI research and projects; breadth and scope of current missions
    - Scientific mindset
    - Experience with collaboration and partnerships across: federal government; local governments; academia; business; entrepreneurs
- There seems an urgent need for DOE to establish a coordinating regime to enhance the "Science of AI" and "AI for expanding DOE science":

- Coordinate: Research Projects; AI facilities; AI projects; Enterprise data stores; AI analytics; AI algorithm; AI Models; AI subject matter experts; AI workforce; AI partnerships; Taxonomy/ontology/standards; ethics
- Collaborative: This coordinating regime will require a good balance of "bottom up" collaboration and "top down" visibility and synchronization
- The regime could be modeled after other regimes: like the many DOE "user facilities"; DOE EM's EFCOG (Energy Facility Contractors Group); DOD's Joint AI center (JAIC); GSA center of excellence structure; various consortiums like the nations control of spectrum management; societies and non-profit structures like ASM (International Society of Materials), and EPRI (Electric Power Research Group)

DOE has many strengths to contribute to the advancement of AI – scientific and engineering staff, computing power, computer architecture knowledge, large data sets, and the motivation to work in the public's interest. Many DOE offices and labs have strong and growing activities that will contribute both to the Science of AI and AI for Science. These activities are applauded and should continue and expand as merited.

In order for DOE to accelerate its contribution at a national level, some particularly difficult problems for public good should be organized and championed for DOE leadership. Examples include understanding AI and ML, health monitoring, electric grid resilience, scientific databases for the international science community, and automated cybersecurity.

Preliminary Recommendations

Based on the above observations and conclusions, the SEAB AIML working group has developed five preliminary recommendations that are suggested to the DOE for further study and consideration.

Recommendation 1: DOE-wide AI Capability (DAIC)

To meet the critical national need for AI leadership, the SEAB urges the Department of Energy to create an enterprise-wide capability to accelerate the use of AI for scientific discovery at scale while fostering fundamental advancements in AI. This is summarized in the phrase – *AI for Science and Science of AI*. The scope of *AI for Science* should include the mission-specific development and enterprise-wide use of AI for the advancement of the sciences associated with all aspects of DOE's energy, security, and the science missions. In contrast, the scope of *Science for AI* should include theoretical and technology development efforts that advance the state of AI and the DOE's ability to apply it to problems of national importance.

The SEAB recommends that this be accomplished through the creation of a transformational *DOE-wide AI Capability (DAIC)* that is managed at the highest level in a way that provides insight and influence across the DOE enterprise but also reaches into all laboratories, centers, DOE mission areas and entities where AI-related activities are performed.

To push the frontier, a DOE-wide AI Capability (DAIC) should focus on software, data, infrastructure, and AI workforce. On software, because it is the principal way AI technologies are instantiated. On data, because it is the evidence of the phenomena of interest and, in aggregate, is the grist for AI-enabled scientific insight. On infrastructure, because new hardware architectures, supercomputing resources, software, networking, data sets, and data models that span research domains and geographic locations are essential to pursuit of AI on a national scale. On workforce, because the ready availability of an AI workforce (AI expertise and relevant domain knowledge) is a necessary precursor to meaningful and lasting impact. Building and expanding upon the DOE's long tradition of creating and operating world class, specialized facilities for advanced science, the DAIC will have two essential hub and spoke components: a center of excellence and a network into the DOE complex.

DAIC - AI Center of excellence – A new DOE user facility that serves as the hub of the DAIC; a center where AI and related skills are developed and supported across professional careers supporting critical workforce development in AI. It also houses the appropriate development, test and integration facilities for AI technologies. The key components of the DAIC include the following:

- *Data & Models* – appropriately architected data repository (Data Lake) containing fully cataloged and AI-ready datasets with multiple levels of security. Also computational models capturing our expanding understanding of physical laws. This includes data curation, dataset coordination, and data policy enforcement.
- *Compute* – national research cloud incorporating hybrid commercial cloud and government-owned multi-scale high performance computing regimes. Also, the home for developing, testing and benchmarking new AI related hardware and computer architectures.
- *Algorithms & Tools* – enabling software for AI development and large-scale computing regimes
- *AI Expertise* – become an employer of choice for AI people who want to reach and build a career in AI. "There is no established path for the researchers to maintain an awareness of the current state-of-the-art AI practices and methods." We must create a supportive environment for leading AI Scientists and Engineers to build a career in AI/ML by helping the DOE further its mission. It is also a place where scientist in other domains can come and learn about the AI technologies and tools that are available to them, and where the AI experts learn about the data and discovery challenges faced by the scientific community.
- AI Resource Optimization – the methods to effectively maximize the utilization of AI resources across many projects.
- *Policy and Data Governance* – the processes, expertise and authority to designate the protection level of data ranging from "full and open" through multiple levels up to and including classified.
- *A center for AI Counter Intelligence* – the ongoing evolution of understanding the cyber threats associated with AI and the appropriate policies and practices.
- *A true DOE User Facility where government sponsored work is performed, but also location where government resources are available to U*.S. industries in a cost recovery or subscription model to accelerate adoption and use of new AI technologies by industries.
- *Physically located where the energy to support AI is not a limiting factor*.

The above list of centralized elements is a preliminary suggestion. A complete set of capabilities to be centralized in a center of excellence versus distributed via User Facility Network is a matter for further study.

> ***DAIC - User Facility Network*** – seamlessly interconnected high-speed network of satellite facilities with unique, but complementary capabilities. Each lab will have a DAIC office that provides AI subject matter expertise and technical support to scientists and serves as a virtual access point to the DAIC systems.
>
> Further, it is recommended that DOE adopt a charge to leverage the DAIC to drive significant fundamental advancements. The proposed DAIC presents the perfect "ground zero" for these advancements. The potential for Turing Awards for fundamental advancements in AI that enable entirely new paradigms for scientific discovery would be substantial in this environment. These advancements will ultimately create artificial intelligence methods that are consistent with the physical laws, robust to gaps or anomalies in data and interpretable in ways that are meaningful and defensible to scientific community. This includes:

- Physics-informed AI
- Next-generation model inversion
- Automated hypothesis generation and causal discovery
- Enabling AI computing regimes and supporting hardware
- Learning algorithms for novel computing regimes (quantum, exascale)
- Robust learning for science (anomaly detection, uncertainty quantification)
- Power-constrained learning and power-efficient learning

Recommendation 2: The Data Problem

The DOE should make progress on the Data problem. Since data is the fuel for statistical AI (e.g., ML) and data science, access to one's data implies ownership of their AI from a security standpoint. The DOE laboratory complex engages with a broad range of partners inside the United States (e.g. universities) and outside of the United States (e.g. international science foundations like CERN). Most of these are excellent partners that provide world class collaboration in many sciences, and data flows out of the U.S. in great volumes. However, there are those that will take advantage of publically available data and use it for purposes not consistent with the values of the American people and in some cases use the data for aggressive military planning. This can no longer be ignored by the broad scientific community. In addition to the efficiencies of centralized processing of data, there needs to be a first level evaluation of the sensitivity of data and the appropriate controls applied. The default disposition should not be "open to all" without a minimal consideration to potential uses of the data.

Furthermore, socialization of data needs to be improved within the appropriate security regiment. It is not a contradiction to want to protect some data from indiscriminant exposure while also wanting universal access to other data. The goal is to capture the data in a way that maximizes efficiency across the mission space, reducing redundancy and waste while protecting vital interests.

Consider the following:

- For selected (relevant) use cases… where is the data, what must be done to uncover it, how is it accessed, how and who to curate it, metrics of success, In this case, the DOE could establish incentives or mandates for data creation, curation, and sharing. NASA's PDS model could be an example.
- Relevant theory, processes, procedures, tools, techniques.
- Develop risk procedures: indemnification, privacy, proprietary, quality, integrity, ownership, boundary conditions

Recommendation 3: Workforce Development

The DOE needs to foster the development of an AI workforce for the DOE and for the U.S. economic base. To this end, the DOE needs a plan to identify, grown, develop, manage, shepherd the US and specifically DOE AI workforce. Today the competition for AI expertise is very strong, and several DOE labs have to compete with industry leaders and their large salaries and benefits for AI talent in their communities. Inspiration in the virtue of the DOE mission should be used as a discriminating factor.

Current DOE STEM workforce development programs should include additional resources to materially impact the pipeline for AI professionals.

Recommendation 4: Regulator Function

Regulatory aspects of AI need to be considered. The AI enterprise will eventually require some assessment, inspection, regulation functions to be done at the right level with the right frequency and for the right outcomes. DOE should start to consider and understand the need and how to proceed.

Recommendation 5: Cybersecurity for AI

- **Assistance outside of DOE:** DOE needs to help drive a secure, efficient, effective AI for the mission of DOE **but beyond that** to be the lead integrating agency for the fundamentals of AI for the broader USG.
- **Non-military Cybersecurity:** The DOE must lead the nation regarding all non-military aspects of cybersecurity for AI since no other USG civilian agency is focused on **creating and deploying** cyber security for AI -- for the data integrity, for the software, for the hardware, for the algorithms.
- **AI and PII Data:** Beyond the actual security of the data, is the issue of PII and anonymization. DOE is seen as a trusted entity so DOE promoting anonymization processes will be critical if we actually want researchers, etc to use the data lake and not have class action suits by the public that their data is used without their permission.

Science of AI

Science of AI refers to the disciplines focused on advancing the foundation of AI. Fundamentally, this is no different from advancing mathematics, or HPC technologies. DOE, and in particular DOE labs are unique and well-positioned to be in advancing *Science of AI.* Why?

- First, DOE has demonstrated fundamental advances at large scale in mathematics, computer science, HPC, various natural sciences which is unique (as compared to any other similar enterprise, government or private.
- Second, DOE has the structure, unparalleled infrastructure of computing, data and networking, people and a well-established model for advancing "science for X".
- Third, DOE enterprise (labs) is the national leader in advancing supercomputing, algorithms and software as scale, energy-efficient computing, and all critical components of advancing *Science of AI*.
- Fourth, DOE has large-instruments and experimental facilities, world-leading HPC systems, thousands of small instruments, massive sensor-based infrastructure, the corresponding domain knowledge and applications of these, and the corresponding datasets; all very critical to applying, evaluating and validating *Science of AI* at a scale and variety that has the potential of advances in *Science of AI* generalizable, trustworthy and validated. This is in comparison to industry based narrow-focused AI research for specific use-cases such as digital-twins, targeted advertising, autonomous vehicles etc.
- Finally, DOE is also well-positioned to leverage the open academic institutions via leveraging their collectively large expertise (individually small within a university) via collaboration and targeted funded projects in *Science for AI* space, similar to those in the current model in HPC, energy, CS, basic sciences and mathematics. Taken together, the afore-mentioned capabilities of the DOE enterprise makes it unique and ready to take on Science *for AI pursuit*.

The Artificial Intelligence and Machine Learning Working Group respectfully submits these preliminary findings and recommendations to the full SEAB.

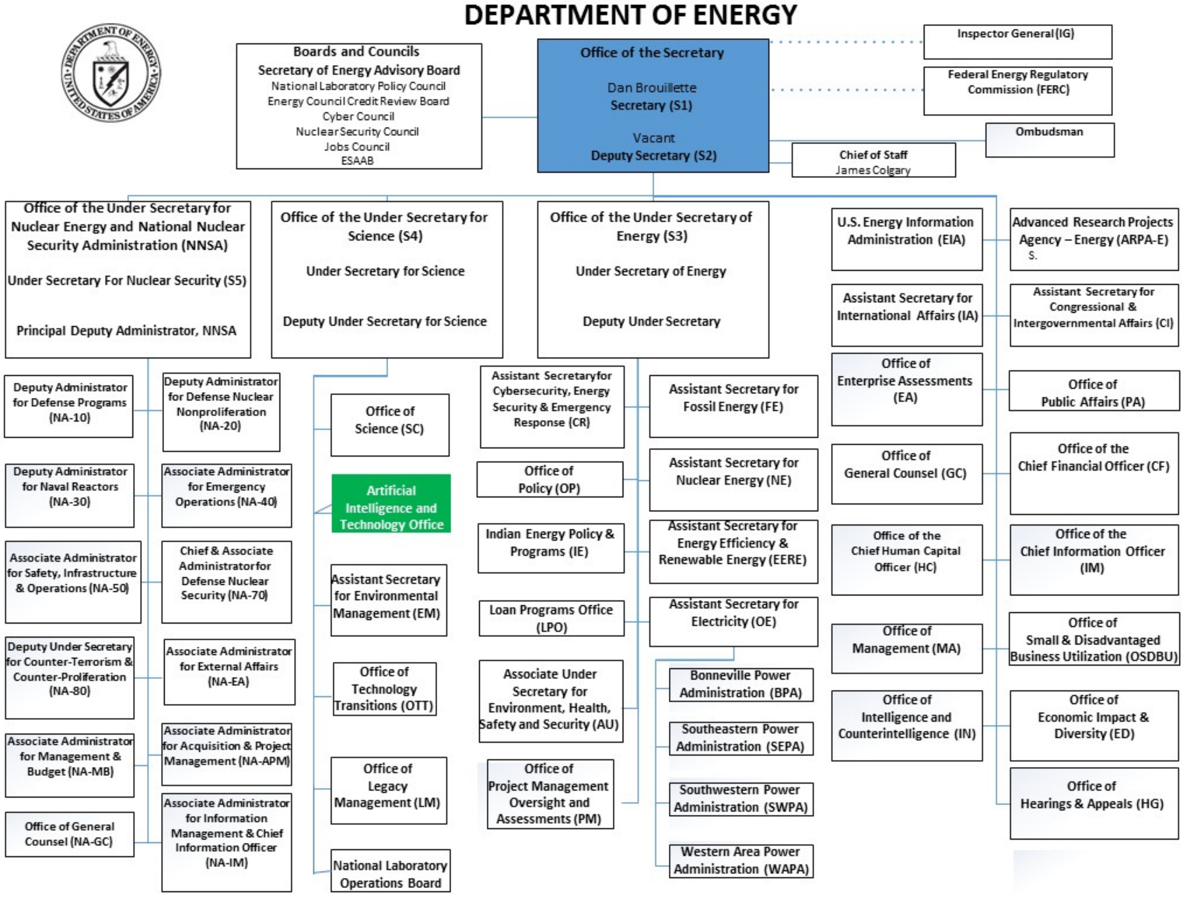
The remaining sections of this document contain information provided to the working group from the DOE.

AITO Data Call Summary

This section contains the response to a call for information from the SEAB Working Group on Artificial Intelligence (AI) on DOE AI efforts. Specifically, it addresses the 9 focus areas of the SEAB AI Working Group, but also provides relevant information on the newly created Artificial Intelligence and Technology Office (AITO), its vision and mission to transform the DOE into a world leading AI enterprise. The far reaching consequences of AI are creating opportunities to accelerate the development of critical AI technologies. The DOE is engaging broadly in AI research and development, as evidenced by the overwhelming response to our data call. The hundreds of projects identified are focusing on addressing immediate core mission needs, and are primarily exploiting existing data seta and available AI methodologies. The initial data call did not include work being done at our 17 National Laboratories. Although many AI technologies from industry have found use in DOE programs, industry will not address the most critical DOE mission needs. In addition, numerous barriers to AI implementation have been identified by the program offices. A DOE AI strategy, currently under development, will identify priority areas, address gaps where greater effort is needed, and mitigate barriers to adoption.

DOE Organization

The Department of Energy has three core missions: Energy, Science and National Nuclear Security. There are three corresponding Under Secretaries: Under Secretary of Energy (S3), Under Secretary of Science (S4) and Under Secretary for Nuclear Security (S5).


DEPARTMENT OF ENERGY
DEPARTMENT OF ENERGY · UNITED STATES OF AMERICA
Boards and Councils
Secretary of Energy Advisory Board
National Laboratory Policy Council
Energy Council Credit Review Board
Cyber Council
Nuclear Security Council
Jobs Council
ESAAB
Office of the Secretary
Dan Brouillette
Secretary (S1)
Vacant
Deputy Secretary (S2)
Inspector General (IG)
Federal Energy Regulatory Commission (FERC)
Ombudsman
Chief of Staff
James Colgary
Office of the Under Secretary for Nuclear Energy and National Nuclear Security Administration (NNSA)
Under Secretary For Nuclear Security (S5)
Principal Deputy Administrator, NNSA
Office of the Under Secretary for Science (S4)
Under Secretary for Science
Deputy Under Secretary for Science
Office of the Under Secretary of Energy (S3)
Under Secretary of Energy
Deputy Under Secretary
U.S. Energy Information Administration (EIA)
Advanced Research Projects Agency – Energy (ARPA-E)
S.
Assistant Secretary for International Affairs (IA)
Assistant Secretary for Congressional & Intergovernmental Affairs (CI)
Deputy Administrator for Defense Programs (NA-10)
Deputy Administrator for Defense Nuclear Nonproliferation (NA-20)
Deputy Administrator for Naval Reactors (NA-30)
Associate Administrator for Emergency Operations (NA-40)
Associate Administrator for Safety, Infrastructure & Operations (NA-50)
Chief & Associate Administrator for Defense Nuclear Security (NA-70)
Deputy Under Secretary for Counter-Terrorism & Counter-Proliferation (NA-80)
Associate Administrator for External Affairs (NA-EA)
Associate Administrator for Management & Budget (NA-MB)
Associate Administrator for Acquisition & Project Management (NA-APM)
Office of General Counsel (NA-GC)
Associate Administrator for Information Management & Chief Information Officer (NA-IM)
Office of Science (SC)
Artificial Intelligence and Technology Office
Assistant Secretary for Environmental Management (EM)
Office of Technology Transitions (OTT)
Office of Legacy Management (LM)
National Laboratory Operations Board
Assistant Secretary for Cybersecurity, Energy Security & Emergency Response (CR)
Office of Policy (OP)
Indian Energy Policy & Programs (IE)
Loan Programs Office (LPO)
Associate Under Secretary for Environment, Health, Safety and Security (AU)
Office of Project Management Oversight and Assessments (PM)
Assistant Secretary for Fossil Energy (FE)
Assistant Secretary for Nuclear Energy (NE)
Assistant Secretary for Energy Efficiency & Renewable Energy (EERE)
Assistant Secretary for Electricity (OE)
Bonneville Power Administration (BPA)
Southeastern Power Administration (SEPA)
Southwestern Power Administration (SWPA)
Western Area Power Administration (WAPA)
Office of Enterprise Assessments (EA)
Office of General Counsel (GC)
Office of the Chief Human Capital Officer (HC)
Office of Management (MA)
Office of Intelligence and Counterintelligence (IN)
Office of Public Affairs (PA)
Office of the Chief Financial Officer (CF)
Office of the Chief Information Officer (IM)
Office of Small & Disadvantaged Business Utilization (OSDBU)
Office of Economic Impact & Diversity (ED)
Office of Hearings & Appeals (HG)

Artificial Intelligence & Technology Office

On September 5, 2019, Secretary Rick Perry announced the establishment of the Department of Energy's (DOE) Artificial Intelligence and Technology Office (AITO). Upon announcing the establishment of AITO, Secretary Perry declared, "The world is in the midst of the Golden Age of AI, and DOE's world class scientific and computing capabilities will be critical to securing America's dominance in this field."[1] The Secretary established the office to serve as the coordinating hub for the work being done across the DOE enterprise in Artificial Intelligence. This action has been taken as part of the President's call (Executive Order 13859) for a national AI strategy to ensure AI technologies are developed to positively impact the lives of Americans.[2] AI is a technology that is capable of preforming tasks that mimic human intelligence, which broadly includes areas such as pattern recognition, decision making, visual perception, speech recognition, information processing, behavior adaptation, autonomous control, optimization, etc. AI has the ability to affect countless technologies used within the DOE and associated national laboratories and has implications for areas such as National Security, Energy Security, Cybersecurity—all which have high consequences for failure. ***This creates a need for trustworthy AI that is accurate with high confidence, proven to be unbiased and reliable.*** These are areas that are not being widely pursued in today's AI industry.

AITO VISION & MISSION

Vision: Transform DOE into a world-leading AI enterprise by accelerating the research, development, delivery, and adoption of AI.
Mission: The Artificial Intelligence & Technology Office, DOE'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

Most importantly, AI is of great importance to the Nation as it will be a technology that will shape the future of global power (*NSCAI 2019 Interim Report*[3]). AI will be the economic driver of the next decade, projected to add at least 14% ($4 Trillion) to the US economy by 2030 and increasing China's by 26%[4] Much of the world is viewing the development and adoption of AI technology as a race between China and the US; a race that our allies want and need the US to win. Unfortunately, China is outspending the US in AI research development and applying AI across their entire society with an 80% adoption rate in their industry sectors.[5] Failing to lead in this important technology race would directly impact our National and Energy security.

[1] "Secretary Perry Stands Up Office for Artificial Intelligence and Technology," Energy.Gov, September 6, 2019, https://www.energy.gov/articles/secretary-perry-stands-office-artificial-intelligence-and-technology

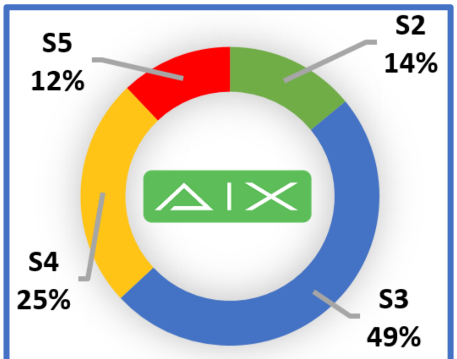
[2] https://www.whitehouse.gov/presidential-actions/executive-order-maintaining-american-leadership-artificial-intelligence/

[3] https://www.nscai.gov/about/reports-to-congress

[4] https://www.pwc.com/gx/en/issues/analytics/assets/pwc-ai-analysis-sizing-the-prize-report.pdf

[5] http://image-src.bcg.com/Images/Mind_the%28AI%29Gap-Focus_tcm108-208965.pdf

There are countless examples across the 17 labs and the entire DOE complex to illustrate DOE's unique qualification to lead this national security mission. Just like our successful leadership of the Manhattan Project led to the nuclear triad, DOE is now forging a technology triad—of exascale computing, quantum information systems and artificial intelligence. Through this technology triad, DOE and America will literally change the world while preserving our most precious freedoms.



The AI Exchange currently holds over 300 DOE active AI-related projects; half of which were reported by Energy programs, a quarter from Science and Environmental Management, and the rest evenly divided between S2 (ARPA-E/OCIO) and NNSA.

In order to evaluate the scope of AI activities within the DOE Programs, AITO created a database known as the AI Exchange (AIX), which contains summaries of over 300 AI-related projects the DOE is currently investing in. The database is designed to ensure projects meet DOE strategic goals, objectives and guarantee mission alignment as Programs move forward in AI initiatives. In addition, AIX will enable the reduction of duplicative efforts across mission spaces, provide points of contacts in AI RD&D, and increase the education and adoption of AI throughout the Department. AIX will be used to capture current AI activities within the DOE programs, sites and the National Laboratories, which will allow the AITO to accelerate the research, development, delivery and adoption of AI across the DOE.

DOE-fueled AI is already being used to strengthen our national security and cybersecurity, improve grid resilience, increase environmental sustainability, enable smarter cities, improve water resource management, as well as speed the discovery of new materials and compounds, and further the understanding, prediction, and treatment of disease. Additionally, AITO has determined that DOE's AI activity is supporting seven of the eight Office of Science, Technology and Policy (OSTP) strategic priorities in AI, which includes long term investments in AI research, developing effective methods for human-AI collaborations, and expanding public-private partnership to accelerate advances in AI. [6]

Global leadership in AI technology is a national security priority and the adoption of trustworthy AI systems is an urgent imperative. AITO has developed a plan of action and is moving forward, working with DOE Programs, associated National Laboratories, industrial and international partners, to accelerate the development of AI capabilities in support of DOE core missions and ensure AI leadership for the Nation.

[6] https://www.nitrd.gov/pubs/National-AI-RD-Strategy-2019.pdf

AITO’S PLAN OF ACTION

1. Create the DOE AI Strategic Plan to define Departmental goals and determine a long-term AI Roadmap to ensure return on investments.

2. Prioritize and develop jointly funded AI partnerships.

3. Institutionalize the adoption of the AI Exchange Database.

4. Develop and implement AI leadership and workforce training.

5. Define future directions in AI through workshops which bring together Programs, National Laboratories, Industry and International Partnerships.

6. Secure authorities, direction and resources needed to successfully execute AITO’s

Data Call to DOE

In response to the Secretary’s charge, the following questions were disseminated throughout the DOE:

1. What AI technologies are being developed in DOE?
2. What AI technologies from industry are being used in DOE?
3. What are DOE’s comparative advantages in AI?
    a. Are there additional AI areas where DOE should have a comparative advantage?
4. What are the comparative advantages in the U.S. industrial base regarding AI?
5. What limitations (hardware, software, data, etc) does the DOE foresee that will slow advancement in AI?
6. How should the DOE address the cyber-security of AI driving elements in data and/or the AI hardware and software supply chain?
7. What systems or processes does the DOE employ to encourage the use of AI internally?
8. How do the labs foster the use of AI?
9. How has or should the DOE built platforms to help other develop or use AI?

The questions are being coordinated through the new Artificial Intelligence and Technology Office (AITO).

DOE Response

AI Technologies Developed in DOE

AI is rooted in the application to specific outcomes. The technologies DOE employs range from application of open source tools, use of open source data, to development of new sensors, AI chips, and learning methods. Technologies and methods are being developed in many areas where available approaches are not adequate or existent. Machine learning, robotics, sensing, expert/intelligence systems and technologies, data science and analytics, and autonomous systems, for example, are important to the broad scope of DOE mission, business and operations, with needs spanning from off-the-shelf to next generation approaches. Much of today's efforts are nascent, project-based approaches that span a growing domain of the AI space. Although the DOE is investing in a broad swath of technologies to address a wide variety of mission issues, we are not yet deeply invested in a number of important, cross-cutting areas that continue to emerge. Advances in such areas as ethics, data and model bias, impacts to privacy, use of (and defense against) adversarial AI, uncertainty quantification tied to confidence in decisions, and sparse data learning will be critical for the nation as it seeks AI supremacy. DOE is also not pushing the envelope to develop next generation technologies at the scale we are capable – current focus is largely on the immediate application of existing technology to amenable problems. AITO was created in recognition of this national need, and one of the defined roles of AITO is to identify gaps in research and development and find the means to drive progress. To sample some of the hundreds of individual efforts:

Data Ingestion:

Multimodal data ingestion from multiple sensors, diagnostic, and control technologies

Models that combine data streams from multiple sensors (e.g. radiation and thermal measurements analyzed together.

Technologies to detect undeclared nuclear facilities using computational models of the nuclear fuel cycle that integrate information from multiple sensors.

Sensor, diagnostic, and control technologies for real-time monitoring of critical plant components to increase coal-fired power plant efficiency, improve unit reliability and availability, and enhance unit capability for flexible operations (e.g., cycling).

Edge device development and deployment

Technologies that merge advanced edge computing, data fusion and machine learning techniques for virtual metering, and a central repository for grid applications such as distributed energy resource control and others on one platform.

Sophisticated compression/rejection data pipelines operating at next to large scientific detectors to save the highest-value data from user experiments.

Learning:

Higher dimensional graph methods (4D, 5D, etc.) for classification and generation problems.

Foundational approaches that leverage math and computer science expertise (e.g. optimization, optimal control, and synthetic data) to improve learning efficiency and accuracy in AI systems and within the context of theoretical limits on computability, expressability, and learnability.

Hybrid molecular dynamics (MD)/machine learning methods that can more readily simulate complex multi-dimensional systems. This include quantum aware MD using active learning.

Application of the DOE Co-design approach to novel AI technologies. We are working with vendors on creating testbeds using early hardware design and rich data environments.

Creating of learning methods which can embed known physical laws so the learning is consistent with the laws of physics, etc.

Scalable learning methods that can operate on DOE petascale and future exascale platforms.

AI techniques for hypothesis generation, design and control of experiments, including the ability to control a large set of experiments or simulations, e.g., to decide what experiments and what parameters.

Methods that work with DOE large facilities to design new devices and control highly sensitive instruments such as accelerators and networks.

Image analysis techniques tailored for the analysis of radiographs, as well as patterns in data and to outputs of simulations.

Methods for estimating nuclear data from heterogeneous data sources ranging from integral to differential data.

Decision Support:

Intelligent approaches to resource management, reinforcement learning in particular, where an agent makes actions given state observations from an environment.

Uncertainty Quantification:

Simplifying learned models to explain their basis, and provide verification and validation of AI systems.

Leverage uncertainty maps to enable generalization of a trained model to shifted domains for High-Consequence Deep Learning Applications.

Uncertainty Quantification on Neural Networks on the reduced order model using active learning for selected problems.

Trust and Adversarial AI

Trust and decisions based on sparse and incomplete/noisy/.. data

Adversarial attacks, describing how a rogue state or actor might use Deep Learning (DL) to devise new attacks, enabling material concealment, delivery, theft, and sabotage

Human/AI Interface:

Visualization and prediction tools are being developed for reservoir management by moving advanced control rooms from visualization of live data to visualization of forecasted behavior for different operational decisions.

Robotics and Autonomy:

Evaluating alternative methods in automation for retrieving and unloading radioactive material shipping containers, including robotics to improve the efficiency of process steps within the surplus plutonium disposition.

Developing automated analysis tools using various techniques, including Bayesian statistical methods and artificial neural networks, to identify anomalous radioisotope measurements in environmental samples.

Autonomous systems and robotics for environmental remediation in harsh and confined environments.

AI Workflow:

Designing workflows that incorporate longer-term solutions for collecting, managing and querying data and learned models, with a focus on sustainability and increased flexibility in ML training.

Drug design has advanced through new AI workflows created in the ATOM partnership with GlaxoSmithKline, NCI and UCSF. Workflow development is fundamental to progress in AI use, and is closely tailored to the data sources, models and available experimental opportunities.

Data Structures, Environments, Labeling:

Rich data environments and architectures that can accommodate the novel AI chips and architectures beyond GPUs and TPUs.

Methods to automatically label data as it is taken so it is suitable for learning.

AITO maintains an active database on current AI activities across the DOE and its labs. A current breakout of the activities across OSTP strategic priorities and by technology type. This does not separate which activities are pushing the technology envelope.

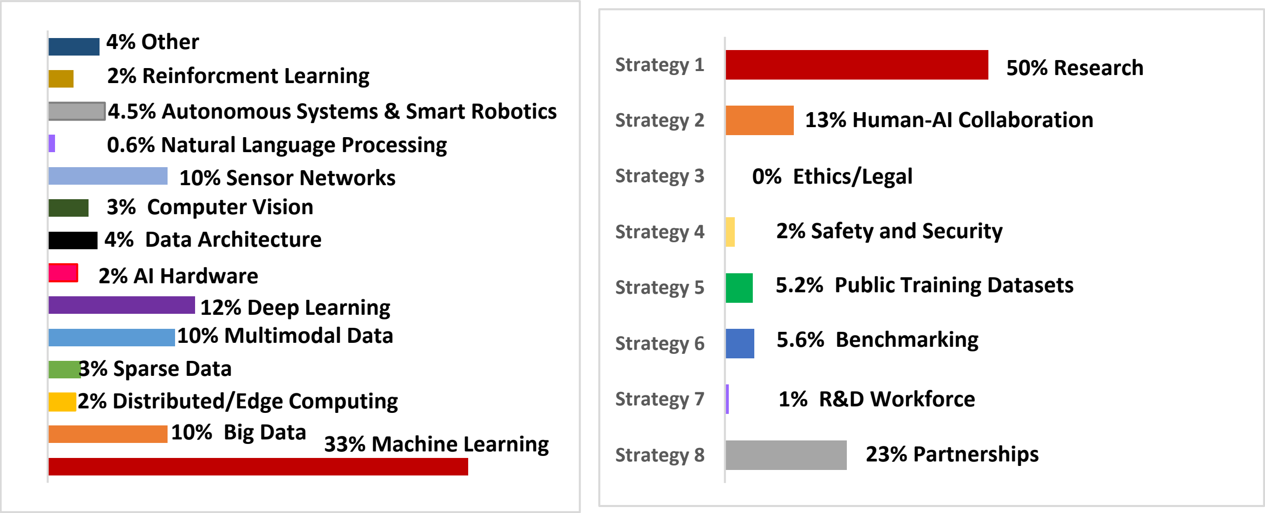


Figure 1: Distribution of AI technologies seeing development in current projects across the agency (left), and the percentage of projects addressing the eight strategies defined by NITRD in their document "The National Artificial Intelligence Research and Development Plan" (right).

There are many sector specific areas as well. For instance, DOE supports R&D efforts that:

- Use control system data and artificial intelligence algorithms to analyze, detect, isolate and automatically respond to threats on the network by identifying anomalous communications and behaviors to ensure the continuous flow of energy;
- Advances artificial intelligence approaches, such as machine learning, that uses to its advantage the data generated by the underlying physical process of energy delivery and

recognizes departures from normal operations resulting from actions of an intelligent adversary;

- Collaborates with energy sector stakeholders on the development of AI/Machine Learning techniques to enable power system equipment to automatically identify a cyber-attack and adapt operations to survive.

Although there are a broad number of technologies represented in current efforts across the DOE (see Figure 1), there are clear gaps emerging (and some technologies not represented at all.) There is scant effort in the important areas of learning from sparse data or edge computing, for example, while a full third of the AI projects reported working on some form of machine learning. Some focus areas (such as computer vision or natural language processing) which seem lightly addressed inside the DOE will be thoroughly addressed in the private sector, while others (such as reinforcement learning and related adversarial approaches) will properly need accelerated government effort.

AI from Industry used in DOE

Off-the-shelf and open source tools are important elements for DOE efforts, as are industrial partners that can focus more specifically on DOE needs. Industrial partnerships are critical to US leadership. In contrast to previous deep shifts in technology, this AI revolution is rooted in private sector innovation and new mechanisms to draw on that innovation engine are imperative. This has to be done keeping in mind that industry is not working many areas important of AI, but understandably developing solutions for their own needs or broad market drivers. They cannot be expected to deliver solutions in our mission spaces

DOE is leveraging technologies such as IBM Watson, General Electric’s (GE) Predix, and other AI tools currently being developed by industry and research organizations.

DOE has partnerships with cloud providers for AI based support including Google Cloud, AWS and Microsoft Azure.

DOE labs are using nearly all the primary deep learning technologies developed by industry including those methods for computer vision, text comprehension, transformers, etc. (the number of AI methods that have been developed is in the hundreds).

DOE labs are using the major AI frameworks (Tensorflow, Pytorch, Mxnet, Horovod, ArcGIS AI family, Jupyter notebooks ...) developed by industry and building systems that leverage them.

DOE labs are also beginning to acquire and test AI hardware accelerators specifically built for 10x-100x acceleration of deep learning applications. We would like these to develop into co-design partnerships that can lead defining the AI technology space.

DOE labs are also experimenting with some of the large-scale industry workflow systems developed to support some of the largest industry projects (MagLev for self-driving cars, Open AI Gym for reinforcement learning, etc.)

DOE is working with General Electric to develop a novel end-to-end trainable AI-based multivariate time series learning system for flexible and scalable coal power plant fault detection and root cause analysis (i.e., diagnosis) and with an eye towards fully remote mining.

AI for Grid Reliability and Resilience

Rapid Damage Assessment & Information Sharing for Power Restoration and Advanced Protective Relaying Technologies and Tools are two efforts drawing on AI. Two companies which have received funding are Brain4Drones LLC and Elintrix. Brain4Drone seeks to reduce outage durations, improve the resilience of the energy sector, and keep linemen and first responders safe by developing an accessory for drones that would enable them to rapidly assess the condition of overhead distribution lines and transmitting that data back to the utility. Elintrix seeks to inform protective relaying operation and reduce incidents of failure-to-trip misoperation by utilizing AI to extract and analyze previously unavailable impulse responses.

DOE’s Comparative Advantages in AI

DOE brings a growing expertise in ML/AI and the world’s most capable computing, along with enormous subject matter expertise. In addition, DOE brings challenging problems involving extremely large and complex data sets that can drive creation of new AI methodology and technology.

DOE has advantages in distinct areas: advanced capabilities, access to unique facilities, and a rich diversity of data and use cases that are consequential. The DOE national laboratory complex combines a wide spectrum of skills that are essential to the successful deployment of AI including;

Ownership of important national missions has created knowledge of many supporting applications, data cleaning and preparation methods, data modeling, and deployment challenges. The DOE facilities also provide researchers access to high performance computers; unique energy sector data including plant data, materials data, post irradiation data; and state-of-the-art imaging equipment.

The DOE could provide additional expertise in visualization, a common AI platform, secure cloud service, and use of AI in additive manufacturing.

Many DOE programs operate in the technology spectrum from directed research to market demonstration, including focusing on early stage technology development. In the

Deployment/Market end of the product continuum, DOE uses well-established partnerships and stakeholders in business, industry, markets and policy with whom plans, goals and outputs are closely coordinated.
DOE has convening power across many areas, from within government, to private sector, academia, utilities, international and others which we routinely exercise.
DOE's greatest advantage is in the large number of computational scientists, computer scientists, data scientists, and other subject matter experts (SMEs) that are resident at its national lab system. DOE also has access to large databases generated by its national labs and by universities and industry that execute DOE projects.
For example, DOE labs have:
Over 1,000 computer scientists and mathematicians that understand HPC, AI, machine learning, optimization, data science, data analysis, statistics
World-class domain scientists in DOE science and engineering mission areas that understand science, data and data analytics
Some of the largest and most capable computers for AI training that exist, enabling DOE to work on comparable problems with the main AI players in industry (Google, Microsoft, Amazon, etc.)
Large-scale facilities and science programs that generate huge volumes of data from light-sources, neutron sources, microscopes, detectors, telescopes, radars and other instruments. This data is a rich source of training data for machine learning.
World leading systems software and expertise to build complex workflows needed for data intensive science such as AI training and inferencing.

Innovations Possible with AI in the next decade

Learned Models Begin to Replace Data: queryable, portable, pluggable, chainable, secure
Many Questions Pursued Semi-Autonomously at Scale: from science discovery to supply chain and cyber security.
AI Becomes Common Part of DOE Activities: Infuses into mission, business and operations; into energy, science and national security.
Experimental Discovery Processes Dramatically Refactored: models replace experiments, experiments improve models
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.

DOE's Potential Comparative Advantages

The DOE has additional limitations and constraints on the use of the AI which translate to comparative advantages during development. The desire to develop actionable AI with quantifiable uncertainties on sparse data to inform decisions on rare (or never-seen) events can lead to the development of altogether new methodologies technologies – driving the AI of the future.
DOE has the ability to conduct pre-competitive R&D, drive data generation at remarkable scale and provide management structure and analytic rigor to support AI

efforts. DOE has unmatched expertise in the integration of systems and infrastructures at scale.

AI research at DOE should enable solutions to critical issues in many economic sectors. Leading AI applications include grid optimization for renewable energy sources, transportation network efficiency, quality healthcare access, and advanced manufacturing processes.

DOE can operate at the limits of technology scales through the development of practical AI testbeds and as a data creator and owner for critical national responsibilities. As a member of the intelligence community, it can drive results into important corners of the missions.

DOE approaches problems through team science and use of broad skills to attack problems in enduring ways, through co-design and other means, and knows how to protect and aggregate data.

As a result, DOE would have an advantage in developing and transitioning to practice AI technologies for use by asset owners and operators of critical systems. For example, CESER is uniquely positioned to advance AI tools and techniques addressing energy delivery systems. CESER projects are working towards more reliable security solutions by using AI to protect future energy delivery system devices and networks. These projects will lead to next generation AI tools and technologies that are not available today to become widely adopted throughout the energy sector, reducing the risk a cyber incident could disrupt energy delivery.

DOE has arguably the most experience in the federal government in working with computer vendors and academia to design future technologies. DOE can work with these partners to develop the novel AI algorithms and AI specific architectures needed for the future.

DOE has an excellent track record in organizing and executing large-scale scientific projects and could apply this experience to the development of mission critical large-scale applications of AI. DOE can leverage its track record in leadership facilities to develop a full AI “ecosystem,” including data management, workflow, and “edge” analytics.

Tackling complex technical challenges: Water Security

As with the Human Genome Initiative, DOE has the ability to lead complex technical challenges that are cross-agency and have societal impact. For instance, one could consider how AI could help ensure global water security under a changing environment:

- Water resources are critical for energy production, human health, food security, and economic prosperity
- Water availability and water quality are impacted by environmental change, weather extremes, and disturbances such as wildfire and land use change
- Methods are needed to integrate disparate and diverse multi-scale data with models of watersheds, rivers, and water utility infrastructure
- Predictions of water quality and quantity require data-driven models and smart sensing systems
- Water resource management that accounts for changes in weather extremes, population, and economic growth

The U.S. Industrial Base's Comparative AI Advantages

The US industrial base is among the most innovative and agile in the world, and can be counted on to support elements of important national missions. There are opportunities in the US manufacturing base across the energy use and generation spectrum, through DOE's national labs and partnerships with industry for advanced manufacturing.

The United States is home to large technology companies such as IBM, Microsoft, Google, Amazon, Facebook, and others that are investing heavily in AI. The United States has an innovation ecosystem that is helping to develop a large number of startup companies in AI hardware, software, and services. The United States also has a large energy industry comprising companies of various sizes, from large companies to smaller oil and gas producers. There is an advantage in having energy companies with data and needs and technology companies with solutions to partner and generate competitive advantages for the U.S. energy industry. A recent example is the partnership between ExxonMobil and Microsoft, which will generate billions of dollars in value over the next decade.

External partnerships are key elements of DOE's approach. Examples include:

- General Electric – Cyber Attack Detection and Accommodation for Energy Delivery Systems: The project team is developing and demonstrating an attack detection and accommodation (ADA) system for energy delivery systems initially with high fidelity models running on a threat simulator. The self-learning and resilient ADA system will be designed to continue to provide power to the grid.
- National Rural Electric Cooperative Association (NRECA) – Energy Sector Security Appliances in a System for Intelligent Learning Network Configuration Management and Monitoring (ESSENCE): The ESSENCE project developed tools that facilitate more secure operational network management. ESSENCE enables an electric utility to define and enforce its operational network security policies with fewer IT staff members and less reliance on significant internal security expertise, using a rules engine where the rules are derived from machine learning.

Limitations Foreseen to Slow AI

In theory, DOE has access to data, however, real operational data is not readily accessible for research and development which may be one of the largest impediments to growing AI capabilities and applications. Another limitation within the DOE complex is the high demand for researchers with a strong AI background. Developing the necessary skills to stay up-to-date in the AI field is also challenging because AI is currently a very dynamic science and there is no established path for the researchers to maintain an awareness of the current state-of-the-art practices and methods.

A chief challenge will be in the area of workforce development, similar to that of cybersecurity where not just industry but academia provide higher pay and opportunities for advancement.

While DOE collects large quantities of data, investment levels in data curation, cleaning, validation and data management systems could become limiting factors. AI advancement will also be limited by data quality. Although DOE has access to vast amounts of data and can potentially collaborate with industry to gather additional data, the quality and format of data and protecting the data are issues that need to be overcome. There is a severe shortage of people with expertise in this area. This is felt by industry and more severely by national labs and universities.

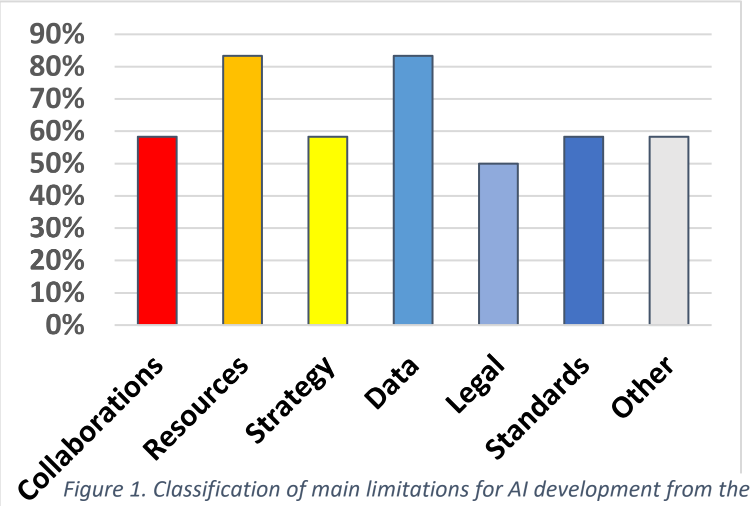
Depending on how fast DOE scientists adopt AI based methodologies the demand for AI computing capability could grow much faster than the AI learning capabilities. There is large uncertainty in the rate of adoption of AI in labs at present.

DOE labs and facilities need to transition to a model where every scientist and engineer has access to AI capabilities to drive productivity and capability. This will require new software and data infrastructure investments at the labs across the program space.

The limitations of current algorithms will become apparent, particularly with respect to the representation and transfer of knowledge, available information content in data, and the assurance of AI systems. This will slow adoption in science and engineering disciplines pending significant algorithmic advances.

Overall, a number of common issues surface across DOE elements (see Figure 2). These include:

- Insufficient collaborations and coordination between programs
- Resources (AI talent and skills)
- Lack of strategy for roadmaps, roles/responsibilities undefined, or top-level decisions
- Data access, trustworthiness, & suitability
- Legal/regulatory (NDA), etc.
- Insufficient benchmarking and standards
- Energy sector reluctance to adopt new AI/ML technologies due to risk of inadvertent disruption of operations.
- Concerns related to AI solutions not being interoperable, scalable, readily manageable and compatible with common methods and best practices.
- Unsuccessful research and development
- Inability to commercialize developed technologies/solutions



Identified Gaps in Approach

Policy Gaps

- Ethics/Privacy
- Safety/Security

Technology gaps

- Adversarial AI
- High-consequence decisions
- AI-data testbeds

Workforce development

- Current training
- Future education
- AI/Data Advisory Group

Figure 1. Classification of main limitations for AI development from the point of view of programs. AI coherence is managed through the Research and Technology Investment Council (RTIC) working group AI subcommittee. This subcommittee, convened by AITO, includes elements from across DOE mission, business and operational functions. This figure summarizes both areas of identified need as well as clear gaps where market and needs are emerging.

Cyber-Security and AI

How should the DOE address the cyber-security of AI driving elements (hardware, software …) in data and/or the AI hardware and software supply chain?

Dynamic Persistent Monitoring

Many of the national security challenges in AI will be built on real-time capabilities.

The Persistent DyNAMICS (Dynamic Nuclear Activity Monitoring through Intelligent, Coordinated Sensing) Venture will designs, builds, and demonstrates an architecture for dynamic persistent monitoring of nuclear processes through intelligently coordinated sensing. It exploits and demonstrates sequences of proliferation activity signatures. In its steady state, collections will occur through autonomous coordination that is dynamically tailored for each event/activity in real-time via AI methods. The collection queuing choices will be driven by inferred nuclear processes and activities.

A significant task of this project is to develop or adapt inference software to address research questions related to the characterization of nuclear facility activity. Persistent DyNAMICS uses data-driven machine learning analysis of sequences, independent of nuclear fuel cycle process models, to provide cross-validation and will develop fast numerical models to train, tune, and test stochastic models and inference engines.

DOE should be investing time into both data-driven and physics-based anomaly detection in its various national security missions that include elements of cyber and supply chain. Adversarial AI and trust, in its growing definitions, as well as the impacts of 5G, will redefine threat envelopes in ways yet to be determined. As an owner of critical systems and responsibilities, DOE cannot just be a customer and must have leadership responsibilities.

As the Sector-Specific Agency (SSA) lead from the 2015 Fixing America's Surface Transportation (FAST) Act, DOE directly collaborates with energy sector utility owners, operators, and vendors to strengthen the cybersecurity of critical energy infrastructure against current and future threats, including cybersecurity of AI driving elements and supply chain issues. DOE aims to deliver game-changing tools and technologies that help utilities secure today's energy infrastructure from advanced cyber threats through competitive funding opportunity announcements (FOA) and research calls to the National Laboratories. CESER emphasizes the selection and investment in emerging technologies from these competitive announcements that will further advance the vision of resilient, cyber-secure energy delivery systems.

Big data can be used to enhance these anomaly detection tools. For instance, process data, maintenance history, corrective action, and industry data can be fused to aid in equipment condition monitoring. Also, real-time data from process systems, atmospheric and oceanic systems, community and first responder systems, and other relevant data can be fused to improve the resiliency and responsiveness of disaster management. Big data could enhance supply chain risk - potentially a fusion of provenance data, supplier evaluations, and known threats.

Cyber security applications for improved intrusion detection systems to identify suspicious activity and potential threats. This includes plant modernization activities to (1) predict plant process anomalies to avoid unexpected plant failure, (2) detect fire in video to automate fire watch activities, and (3) optimize inventory and reduce stocking requirements at nuclear power plants.

In nuclear security, this includes applications such as: equipment condition/health monitoring (detection, diagnostics, and prognostics), reliability assessment, aging management, nuclear safety/transient/accident monitoring, emergency response/disaster management, and cybersecurity intrusion detection.

To ensure the greatest success, cybersecurity controls (technical, physical, administrative) and cybersecurity risk management (including supply chain) need to be considered in the design, implementation, and operation. Aside from this retrospective need, the prospective AI threat considerations will have to become integral to the approach.

While not directly applicable to the narrowly focused question, of note is that DOE launched a Cybersecurity Institute for Energy Efficient Manufacturing in FY19. DOE's EERE will fund the Institute and it will be co-managed by DOE CESER. The main objective of the institute is to address cybersecurity issues most unique and pressing to advanced manufacturers and their supply chain ecosystems. As manufacturing becomes increasingly automated and supply chains become increasingly digitized and interactive, cyber-attacks can negatively affect manufacturer productivity, efficiency, and competitiveness. At the same time, new cybersecurity measures cannot negatively impact manufacturing efficiencies. The institute will be organized around two topic areas: 1) Securing Automation and 2) Securing Supply Chain Networks.

In general DOE should apply existing cyber-security models that apply to data and computation to emerging scenarios in AI including provenance of AI hardware and software.

DOE should support research in AI methods that can compute and train on encrypted (and compressed) data or on data that has been encoded to provide privacy, e.g., differential privacy.

DOE should support research in secure AI training methods that permit the exchange and local training of models rather than exchange of data

Operationally DOE should use a risk-based approach to cyber-security of AI data and models, to balance open scientific exchange in a rapidly moving field with privacy and national security as appropriate to the use case

Processes to Encourage AI Internal Use

DOE encourages enabling data exchange from commercial utilities and others to the research community and access to high performance computing facilities across the national laboratory complex.

DOE's high-performance computing (HPC) system are provided for use of AI internally.

Funding opportunity announcements and annual operational planning in R&D areas that utilize AI and ML are two processes to encourage AI and ML use within DOE.

DOE has a data management policy that encourages the collection and dissemination of data needed for training models

DOE has supported pilot projects to bootstrap AI programs at the laboratories

DOE has created internal teams and AITO to encourage the exploration of AI opportunities across the DOE system

DOE laboratories have made internal investments via LDRD and program development funds to bootstrap AI projects, teams and infrastructure

Community input is crucial for shaping DOE priorities. DOE sponsors workshops and meeting in many venues – internal and external – to provide vision and direction for AI research and development. Some recent ones include:

- Basic Research Needs for Scientific Machine Learning (1/18)
- Fusion Energy Machine Learning / AI Workshop (4/19)
- Data for AI Roundtable (6/19)
- AI for Science Town Halls (7/19-10/19)
- Secretarial AI Roundtable (8/19)
- Basic Energy Science Roundtable (10/19)
- WAPA and Grid AI Meeting (10/19)
- CyberSecurity and Hanford Meetings (10/19)

In addition, AITO is developing workforce development modules that can help both internally at the management level to deeper modules for laboratory scientists and engineers.

DOE Labs and AI

The DOE laboratories foster the use of AI through many means. Starting with workforce and AI summer schools and programs, fellowships and research opportunities, to partnerships that drive deployment and implementation.

The national laboratories have been actively pursuing deploying AI methods in various applications that require complex human modeling and analysis. These are funded through Laboratory Directed Research and Development (LDRD) program solicitations as well as various NE programs.

A computational approach developed by the Institute for the Design of Advanced Energy Systems (IDAES), as well as a set of computational modeling tools, algorithms, and frameworks harnessed under the eXtremeMAT consortium, are being utilized for rapid screening of materials and new material discovery efforts across the National Laboratory Enterprise (i.e., to optimize the design of oxygen carriers for chemical looping combustion and to expedite the development and deployment of new alloy materials for fossil energy applications). In addition, a rare earth element (REE) coal assessment methodology integrated with machine learning capabilities is used by the National Energy Technology Laboratory (NETL) to systematically assess REE concentrations in coal and coal-related strata.

DOE laboratories have created internal initiatives (LDRD, program development, etc.) aimed at increasing the strategic development of AI projects and programs

DOE labs have created partnerships with academia and industry to pursue AI driven projects of mutual interest (e.g. ATOM, INSPIRE, etc.)

DOE labs are hiring AI and machine learning researchers to nucleate additional capabilities, including use of various early career awards.

DOE scientists present papers, lead workshops and other activities at major conferences, including Supercomputing, SIAM and IEEE meetings to foster broad collaborations and research with DOE scientists.

DOE has helped promote AI-related research activities through the recent AI XLab Summit, which brought together industry and Lab practitioners to identify challenges and opportunities that would benefit from the application of AI techniques, and to highlight DOE assets that can be leveraged to support these efforts. This large cross sector meeting included themes in healthcare, energy, manufacturing and urban science.

The DOE National Laboratories recognize the AI field is a growing area because it is proven technology with the potential at providing unique insights, skills and perspective to enhance the reliability and resiliency of the Nation's critical energy delivery infrastructure.

CESER is aware the National Labs are developing capabilities in AI and is continuing to leverage these capabilities towards realizing the goals of the CESER mission. CESER sponsors DOE National Laboratories in the research and development of AI technologies, with the goal of transitioning to practice these technologies to the energy sector. DOE National Laboratories engage with industry partners to leverage AI as part of their solutions for improved detection, prevention and mitigation of cyber incidents. Here are a few examples of CESER-supported AI projects:

Driving AI into practice

The goal of **FOA1861**,[1] Big Data Analysis of Synchrophasor Data, is to explore the use of big data, AI, and ML on phasor measurement unit sensor (PMU) data in order to identify and improve existing knowledge and to discover new insights and tools for enhanced grid operation and management, thus enhancing reliability and resilience. Notably, FOA1861 differs from previous FOAs in that it offers pre-packaged datasets to award recipients who possess commercial or near-commercial big data, AI, and ML tools/capabilities. Two such projects funded under the FOA are: Robust Learning of Dynamic Interactions for Enhancing Power System Resilience, performed by Iowa State University of Science and Technology; and Discovery of Signatures, Anomalies, and Precursors in Synchrophasor Data with Matrix Profile and Deep Recurrent Neural Networks, performed by the Regents of the University of California.

Survivable Industrial Control Systems (ICS) is being developed to support the community. This research integrates two approaches. One is the Cyber-Physical Modeling and Simulation for Situational Awareness (CYMSA) project developing a cybersecurity situational awareness technology suite to detect adversarial manipulation of power grid components and communications networks. The CYMSA project involves novel cyber-physical modeling and simulation research on communications networks and substations. The other is Artificial Diversity & Defense Security (ADDSec), a defensive technology that automatically reconfigures network parameters to continuously and dynamically defend energy delivery systems against adversaries who depend on static environments to launch cyber-attacks. Enhancements to both projects include introducing distributed controllers, optimizing Moving Target Defense (MTD) parameters, implementing IPv6-based security, and correlating events observed from Software Defined Networking (SDN) flows combined with the deployed CYMSA field sensors. The team is working on developing machine learning algorithms to perform Software Defined Networking (SDN) behavioral analysis and correlation/distribution of cybersecurity events.

Similarly, the Adaptive Control of Electric Grid Components for Cyber-Resiliency project seeks to develop supervisory control algorithms to counteract cyber-physical attacks that might attempt to compromise multiple independent systems in the electric grid. The project team will analyze the stability of various feedback control systems (e.g., distributed energy resources (DER), and voltage regulation and protection systems) in the electric grid to determine what parameters an attacker would change if DER and utility voltage regulation and protection systems were compromised. The research will develop reinforcement learning algorithms to characterize the complex interaction between the DER and the electric grid to facilitate a better defense against cyber-attacks.

Existing or Potential DOE IA Platforms

The DOE should develop platform agnostic AI capabilities to ensure the greatest applicability across industry. DOE has several active efforts to develop shared public data sets and environments for AI training and testing. For example, in the Computational Consortium for Physics and Chemistry (CCPC), the project uses high-performance computing (HPC) capabilities to model the behaviors of multiple catalyst systems in several pathways. Datasets are made public on the ChemCatBio_Data_Hub.

DOE engages in projects to assist others in utilizing Machine Learning (ML). The Integrated Multi Scale Machine Learning Project creates a foundation for the development of advanced data analytics utilizing existing and new data sources developed through both private and DOE partnerships. ML is the basis for the analysis development, and will allow new levels of visibility and resource integration to be achieved using open and utility datasets.

DOE has developed a rigorous, computational approach for developing new concepts for energy systems by applying models that are both multi-scale and dynamic in nature while incorporating intrusive uncertainty quantification techniques. DOE is investing in several software projects such the CANDLE and ExaLearn ECP project aimed at building AI software tools, AI workflows and libraries for enabling broad collection of AI applications to be developed and run on DOE supercomputers.

DOE has always been a leader in team-based interdisciplinary science with programs like SciDAC, and has additional science-focused projects (CAMERA, ExaSheds, and ECP projects ExaBiome, ExaFEL, and ExaGraph) using AI methods, but these efforts should expand broadly across DOE.

DOE labs are also creating open source AI software and applications that enables third parties to accelerate their AI developments (e.g. LBANN, and many others).

DOE is building platforms to allow bi-directional sharing of threat information to encourage the use of AI approaches to address the energy sector's cybersecurity needs.

- Cyber Analytics Tools and Techniques (CATT™ 2.0) will address both IT and OT infrastructure, and is designed to provide the energy sector with situational awareness and actionable information to support discovery and mitigation of advance cyber threats to U.S. energy infrastructure enriched with classified threat information unique analytical tradecraft owned by the U.S. Government.
- Cybersecurity for the Operational Technology Environment Pilot (CyOTE™) monitors utility data in the complex OT environment to identify malicious actions using an efficient approach to manage data by exception.

In the future, these platforms could be expanded to include AI tools and technologies for faster threat detection and mitigation.

DOE Data Call Program Responses

S3 – Under Secretary of Energy: AI Data Call Response

Office of Electricity

Background: Upon announcing the establishment of DOE's Artificial Intelligence and Technology Office, Secretary Perry declared, "The world is in the midst of the Golden Age of AI, and DOE's world class scientific and computing capabilities will be critical to securing America's dominance in this field."[7] The Office of Electricity, cognizant of the potential advantages and benefits AI affords the U.S. power grid, has sought to develop Machine Learning (ML), Deep Learning (DL), and Artificial Intelligence (AI) capabilities to further this end. In April 2019, OE announced the selection of eight projects to receive almost $7 million in R&D funding.[8] OE's efforts fall under four main categories: 1) FOA1861, 2) grid modernization, 3) SBIR projects, and 4) program-specific projects. The tools/algorithms developed by these initiatives are used in delivering four classes of results pertinent to OE's mission: 1) data analytics in support of fundamental materials research; 2) data analytics on high volume, high velocity grid system data streams; 3) data assisted modeling efforts to auto discover models and relationships in ultra-large grid operational datasets; and 4) machine learning algorithms to deliver faster operations and control mechanisms for digital systems.

The goal of **FOA1861**,[9] Big Data Analysis of Synchrophasor Data, is to explore the use of big data, AI, and ML on phasor measurement unit sensor (PMU) data in order to identify and improve existing knowledge and to discover new insights and tools for enhanced grid operation and management, thus enhancing reliability and resilience. Notably, FOA1861 differs from previous FOAs in that it offers pre-packaged datasets to award recipients who possess commercial or near-commercial big data, AI, and ML tools/capabilities. Two such projects funded under the FOA are: Robust Learning of Dynamic Interactions for Enhancing Power System Resilience, performed by Iowa State University of Science and Technology; and Discovery of Signatures, Anomalies, and Precursors in Synchrophasor Data with Matrix Profile and Deep Recurrent Neural Networks, performed by the Regents of the University of California. The Iowa State project seeks to leverage the team's preexisting big data and machine learning capabilities in order to discover dynamic interactions between electrical grid components. Insights regarding these interactions would result in significant enhancements in system resilience and understanding of anomalous event patterns and cascading outages. The University of California project seeks to apply proven, scalable, multidimensional, and robust big data and machine learning technology on PMU data in order to identify anomalous events, create a catalog of event signatures, predict asset health, and learn precursors to frequency, voltage, and rotor angle instability.

[7] "Secretary Perry Stands Up Office for Artificial Intelligence and Technology," Energy.Gov, September 6, 2019, https://www.energy.gov/articles/secretary-perry-stands-office-artificial-intelligence-and-technology

[8] "Department of Energy Announces $20 Million for Artificial Intelligence Research," Energy.Gov, April 17, 2019, https://www.energy.gov/articles/department-energy-announces-20-million-artificial-intelligence-research

[9] The fact sheet for awarded projects is available at, https://www.energy.gov/sites/prod/files/2019/04/f61/Big%20Data%20Awards%20Fact%20Sheet%20FINAL_1.pdf.

Grid modernization projects seek to further DOE's goal of "[creating] the modern grid of the future."[10] Two such projects are the Grid Resilience and Intelligence Platform (GRIP) and Operational Grid Stability Using Machine Learning (AlphaGrid). GRIP seeks to develop and validate a new software platform, which leverages tools and datasets in order to help operators anticipate, respond to, and recover from extreme weather and distribution system events. Notably, rather than just helping operators recover from extreme events, GRIP will help operators recover faster. AlphaGrid aims to develop a real-time decision support tool in order to help operators assess if they have an opportunity to restore the power system to a safe condition during near-blackout conditions.

SBIR projects currently fall under two funding categories, Rapid Damage Assessment & Information Sharing for Power Restoration, and Advanced Protective Relaying Technologies and Tools. Two companies which have received funding are Brain4Drones LLC and Elintrix. Brain4Drone seeks to reduce outage durations, improve the resilience of the energy sector, and keep linemen and first responders safe by developing an accessory for drones that would enable them to rapidly assess the condition of overhead distribution lines and transmitting that data back to the utility. Elintrix seeks to inform protective relaying operation and reduce incidents of failure-to-trip misoperation by utilizing AI to extract and analyze previously unavailable impulse responses.

Program-specific projects cover an array of different topics, such as grid stability, grid resilience, enhancing situational awareness, and grid reliability. Two such projects are Optimizing Magnetic Devices using Genetic Algorithm, and AlphaClock. Optimizing Magnetic Devices seeks to optimize the design of magnetic devices used in power electronic converters through the use of genetic algorithm. AlphaClock seeks to mitigate the risk of GPS spoofing attacks by developing a measurement-based detection framework for such attacks.

> These eight projects, spanning all four categories, are representative of OE's overall AI R&D effort, which brings together Federal agencies, private business, and academia, in order to both ensure the resilience and reliability of our Nation's grid and to help bring the grid of tomorrow into fruition.

Nuclear Energy

DISCUSSION: On behalf of the SEAB the Designated Federal Officer recently requested the following information related to the development and use of AI within the Office of Nuclear Energy.

a. What AI technologies are being developed in DOE?

   a. There are a number of AI technologies being developed within NE in the areas of cyber security and plant modernization. Within cyber security, the organization is developing improved intrusion detection systems to identify suspicious activity and potential threats. Plant modernization activities include the development of machine learning methods to (1) predict plant process anomalies to avoid unexpected plant failure, (2) detect fire in

[10] "Grid Modernization Initiative," Energy.Gov, accessed October 31, 2019, https://www.energy.gov/grid-modernization-initiative

video to automate fire watch activities, and (3) optimize inventory and reduce stocking requirements at nuclear power plants.

b. What AI technologies from industry are being used in DOE?

  b. NE is investigating the use of technologies such as IBM Watson, General Electric's (GE) Predix, and other AI tools currently being developed by industry and research organizations.

c. What are DOE's comparative advantages in AI? Are there additional AI areas where DOE should have a comparative advantage?

  c. DOE has advantages in two distinct areas; advanced capabilities and access to unique facilities. The DOE national laboratory complex combines a wide spectrum of skills that are essential to the successful deployment of AI including;

knowledge of the application, data cleaning and preparation methods, data modeling, and deployment challenges. The DOE facilities also provide researchers access to high performance computers; unique energy sector data including plant data, materials data, post irradiation data; and state-of-the-art imaging equipment.

The DOE could provide additional expertise in visualization, a common AI platform, secure cloud service, and use of AI in additive manufacturing.

d. What are the comparative advantages in the U.S. industrial base regarding AI?

d. The United States has demonstrated experience in applying AI to specific industry applications.

e. What limitations (hardware, software, data, etc) does the DOE foresee that will slow advancement in AI?

e. In theory, the DOE has access to data, however, real operational data is not readily accessible for research and development which may be one of the largest impediments to growing AI capabilities and applications. Another limitation within the DOE complex is the high demand for researchers with a strong AI background. Developing the necessary skills to stay up-to-date in the AI field is also challenging because AI is currently a very dynamic science and there is no established path for the researchers to maintain an awareness of the current state-of-the-art practices and methods.

f. How should the DOE address the cyber-security of AI driving elements in data and/or the AI hardware and software supply chain?

f. DOE should be investing time into both data-driven and physics-based anomaly detection in nuclear - this includes applications such as; equipment condition/health monitoring (detection, diagnostics, and prognostics), reliability assessment, aging management, nuclear safety/transient/accident monitoring, emergency response/disaster management, and cybersecurity intrusion detection.

Big data can be used to enhance these anomaly detection tools. For instance, process data, maintenance history, corrective action, and industry data can be fused to aid in equipment condition monitoring. Also, real-time data from process systems, atmospheric and oceanic systems, community and first responder systems, and other relevant data can be fused to improve the resiliency and responsiveness of disaster management. Big data could enhance supply chain risk - potentially a fusion of provenance data, supplier evaluations, and known threats.

To ensure the greatest success, cybersecurity controls (technical, physical, administrative) and cybersecurity risk management (including supply chain) need to be considered in the design, implementation, and operation

g. What systems or processes does the DOE employ to encourage the use of AI internally?

    g. The DOE encourages enabling data exchange from commercial utilities and others to the research community and access to high performance computing facilities across the national laboratory complex.

h. How do the labs foster the use of AI?

    h. The national laboratories have been actively pursuing deploying AI methods in various applications that require complex human modeling and analysis. These are funded through Laboratory Directed Research and Development (LDRD) program solicitations as well as various NE programs.

i. How has or should the DOE build platforms to help others develop or use AI?

    i. The DOE should develop platform agnostic AI capabilities to ensure the greatest applicability across industry. Currently, NE is developing a data integration model to enable better data analysis including the use of AI methods.

Office of Electricity Grid Storage Launchpad and Sensor R&D

Energy Efficiency and Renewable Energy (EERE)

a. *What AI technologies are being developed in DOE?*

   EERE is principally a user of AI, developing and adopting data, and its management and analysis, for use in supporting AI functions that enhance technology performance, accelerate development and increase productivity. For the most part, AI technology enables DOE to deliver innovation across the EERE portfolio.

b. *What AI technologies from industry are being used in DOE?*
   Commercial AI and ML programming language software and open-source platforms are widely used across EERE, and includes: Python, TensorFlow, the ArcGIS AI family and Jupyter notebooks.

c. *What are DOE's comparative advantages in AI?* DOE's principal competitive advantages are:
   1) EERE has ready access to high performance super computing systems, equipment and AI SME at our National lab facilities including: NREL, ANL, LBNL, LLNL, ORNL and PNNL for EERE.
   2) EERE's R&D community is in the business of advancing the development of energy technologies with comparative advantages.
   3) R&D drives technology innovation, EERE is principally an applied R&D enterprise accustomed to adopting and applying emerging technology across a broad spectrum of the energy demand and generation.
   4) EERE programs operate in the technology spectrum from directed research to market demonstration, focusing on early stage technology development. In the Deployment/Market end of the product continuum EERE has well-established

partnerships and stakeholders in business, industry, markets and policy with whom plans, goals and outputs are closely coordinated.

5) Convening power.

a. *Are there additional AI areas where DOE should have a comparative advantage?*

1) Pre-competitive R&D,
2) Data generation, management and analysis,
3) Integration of systems and infrastructures,

d. *What are the comparative advantages in the U.S. industrial base regarding AI?*

There are opportunities in the US manufacturing base across the energy use and generation spectrum, through DOE's national labs and partnerships with industry for advanced manufacturing. .

e. *What limitations (hardware, software, data, etc) does the DOE foresee that will slow advancement in AI?*

A chief challenge will be in the area of workforce development, similar to that of cybersecurity where not just industry but academia provide higher pay and opportunities for advancement.

Additional limitations would be the high-performance computing (HPC) "goalposts", which will continue to move, the need to fund successive computing hardware, and growing demand for access to HPC.

f. *How should the DOE address the cyber-security of AI driving elements in data and/or the AI hardware and software supply chain?*

While not directly applicable to the narrowly focused question, of note is that DOE launched a *Cybersecurity Institute for Energy Efficient Manufacturing* in FY19. DOE's EERE will fund the Institute and it will be co-managed by DOE CESER. The main objective of the institute is to address cybersecurity issues most unique and pressing to advanced manufacturers and their supply chain ecosystems. As manufacturing becomes increasingly automated and supply chains become increasingly digitized and interactive, cyber-attacks can negatively affect manufacturer productivity, efficiency, and competitiveness. At the same time, new cybersecurity measures cannot negatively impact manufacturing efficiencies. The institute will be organized around two topic areas: 1) Securing Automation and 2) Securing Supply Chain Networks

g. *What systems or processes does the DOE employ to encourage the use of AI internally?*

EERE will follow the guide of the AITO as it evolves, but are and have been practically engaged in purpose driven AI for several years.

DOE's high-performance computing (HPC) system encourages the use of AI internally.

Funding opportunity announcements and annual operational planning in R&D areas that utilize AI and ML are two processes to encourage AI and ML use within DOE.

h. *How do the labs foster the use of AI?*

This is a self-driven cycle: research generates the massive data sets that requires analysis, computers process the big data to assist the analysis. Implementing machine learning assists in analysis of expansive data sets. Across the EERE portfolio of technologies, researchers utilize machine learning to aggregate enormous operational datasets in order to benchmark performance and identify areas where current predictive models fall short. The goal is to improve predictive models and reduce uncertainty in performance.

i. *How has or should the DOE built[sic] platforms to help other[sic] develop or use AI?* EERE has several active projects that result in the development of shared public data sets and environments for AI training and testing. For example, in the Computational Consortium for Physics and Chemistry (CCPC), the project uses high-performance computing (HPC) capabilities to model the behaviors of multiple catalyst systems in several pathways. Datasets are made public on the ChemCatBio_Data_Hub.

EERE also engages in projects to assist others in utilizing Machine Learning (ML). The Integrated
Multi Scale Machine Learning Project creates a foundation for the development of advanced data analytics utilizing existing and new data sources developed through both private and DOE partnerships. ML is the basis for the analysis development, and will allow new levels of visibility and resource integration to be achieved using open and utility datasets.

Fossil Energy

FE RESPONSE TO SEAB REQUEST:

Q a. What AI technologies are being developed in DOE?

Sensor, diagnostic, and control technologies are being developed for real-time health monitoring of critical plant components to increase coal-fired power plant efficiency, improve unit reliability and availability, and enhance unit capability for flexible operations (e.g., cycling). Machine learning algorithms are being leveraged through the incorporation of autonomous monitoring and big data management to improve subsurface visualization, detect subsurface carbon dioxide (CO_2) leaks, and improve microseismic detection/analysis during geologic CO_2 storage operations; in addition, they are also being utilized to optimize the recovery of oil and gas in unconventional reservoirs. As part of the largest machine learning initiative (SMART), multiple performers are working to transform "human-in-the-loop" decisions on reservoir management by moving advanced control rooms from visualization of live data to visualization of forecasted behavior for different operational decisions.

Q b. What AI technologies from industry are being used in DOE?

Machine learning, robotics, sensing, expert/intelligence systems and technologies, data science and analytics, and autonomous systems are being used in research within the Office of Fossil Energy (FE) projects. Fully remote mining is a desired outcome that could leverage all aspects of artificial intelligence (AI). As an example, the U.S. Department of Energy (DOE) is working with General Electric to develop a novel end-to-end trainable AI-based multivariate time series learning system for flexible and scalable coal power plant fault detection and root cause analysis (i.e., diagnosis).

Q c. What are DOE's comparative advantages in AI?

DOE's greatest advantage is in the large number of computational scientists, computer scientists, data scientists, and other subject matter experts (SMEs) that are resident at its national lab system. DOE's national labs house high-performance computing systems that are being developed with AI/machine learning applications in mind. DOE also has access to large databases generated by its national labs and by universities and industry that execute DOE projects.

> **a. Are there additional AI areas where DOE should have a comparative advantage?**
>
> AI research at DOE should enable solutions to critical issues in many economic sectors. Leading AI applications include grid optimization for renewable energy sources, transportation network efficiency, quality healthcare access, and advanced manufacturing processes.

Q d. What are the comparative advantages in the U.S. industrial base regarding AI?

The United States is home to large technology companies such as IBM, Microsoft, Google, Amazon, Facebook, and others that are investing heavily in AI. The United States has an innovation ecosystem that is helping to develop a large number of startup companies in AI hardware, software, and services. The United States also has a large energy industry comprising companies of various sizes, from large companies to smaller oil and gas producers. There is an advantage in having energy companies with data and needs and technology companies with solutions to partner and generate competitive advantages for the U.S. energy industry. A recent example is the partnership between ExxonMobil and Microsoft, which will generate billions of dollars in value over the next decade.

Q e. What limitations (hardware, software, data, etc.) does the DOE foresee that will slow advancement in AI?

Two challenges that will slow down the advancement of AI are data quality and workforce availability. Although DOE has access to vast amounts of data and can potentially collaborate with industry to gather additional data, the quality and format of data and protecting the data are issues that need to be overcome. There is a severe shortage of people with expertise in this area. This is felt by industry and more severely by national labs and universities.

Q f. How should the DOE address the cyber-security of AI driving elements in data and/or the AI hardware and software supply chain?

FE defers to the Office of Cybersecurity, Energy Security and Emergency Response (CESER).

Q g. What systems or processes does the DOE employ to encourage the use of AI internally?

The use of AI within FE, is largely within the domain of individual R&D projects performed internally or funded externally, rather than a routine tool for programmatic operations.

Q h. How do the labs foster the use of AI?

A computational approach developed by the Institute for the Design of Advanced Energy Systems (IDAES), as well as a set of computational modeling tools, algorithms, and frameworks harnessed under the eXtremeMAT consortium, are being utilized for rapid screening of materials and new material discovery efforts across the National Laboratory Enterprise (i.e., to optimize the design of oxygen carriers for chemical looping combustion and to expedite the development and deployment of new alloy materials for fossil energy applications). In addition, a rare earth element (REE) coal assessment methodology integrated with machine learning capabilities is used by the National Energy Technology Laboratory (NETL) to systematically assess REE concentrations in coal and coal-related strata.

Q i. How has or should the DOE built platforms to help other develop or use AI?

NETL's IDAES has developed a rigorous, computational approach for developing new concepts for energy systems by applying models that are both multi-scale and dynamic in nature while incorporating intrusive uncertainty quantification techniques. In addition, NETL's Joule 2.0 supercomputer allows researchers to model energy technologies, simulate challenging phenomena, and solve sophisticated problems using computational tools that save time and money to ensure that technology development ultimately proves successful. The High-Performance Computing for Energy Innovation initiative facilitates partnerships between industry and the national labs that use DOE's high-performance computing resources to enable new energy technologies.

Office of Environment, Health, Safety and Security Office of Classification

Advanced Computer Tools to Identify Classified Information: An AU-initiated and NNSA supported joint laboratory (SNL, LLNL, PNNL, ORNL, Y-12) program to develop advanced computer tools to identify classified information embedded in electronic documents and augment human classification reviews. The goals of the program are to develop and deploy AI-based tools that can determine whether a document needs a classification review, if the document is classified, which parts of the document are sensitive, and which classification guides are applicable. Effective tools to augment classification reviews will provide three key information elements: a recommended classification level and category, the rationale supporting the recommendation, and provenance (classification guidance) for the recommendation. A key challenge facing designers of such systems, is that the majority of these documents are written in unstructured natural language. They lack explicit contextual and semantic information that is necessary to correctly assign a classification level and category. Today this is overcome by human reasoning developed over years of experience. . To successfully scale DOE's document review capacity to meet ever increasing demand, AI tools must be able to assign semantic meaning to document content and infer context and semantic relationships therein.

Program challenges include the lack of large, marked data sets that can be shared between the laboratories with regard to need-to-know principles, as well as the lack of a network collaboration infrastructure between laboratories (DOE-wide GitLab, Confluence, Mattermost, etc.).

1) ***Project Title*: Does this document need a review? What is the applicable guidance?**

***Project Summary**:* R&D supervised learning approaches to automatically identify the subject areas of a document, and determine if it needs a classification review and what guidance is applicable. With the massive quantity of classified documents available in electronic format, the U.S. Government is in need of systems able to search and categorize contextual and semantic details of these documents.

Goals: ORNL is testing both a binary classifier to predict whether or not a document should be sent in for a classification review, and a multivariable classifier to which classified subject area the document is in. PNNL is using a deep learning model that was trained on unclassified data and fine-tuning it on a classified data set to determine if a document needs a review. PNNL will utilize syntactic highlighting of relevant words/phrases to provide evidence of applicability of a certain classification guideline.

Point of Contact: Dr. Andrew P. Weston-Dawkes

2) ***Project Title:* Which parts of the document are sensitive?**

Project Summary: R&D to develop both rule-based and example-based AI algorithms to identify which specific parts of the document are sensitive and provide the provenance for assigning a sensitivity level.

Goals: SNL is developing rule-based methods employing ontological models of materials and material properties and semantic queries encoded from the weapons material guide. Current efforts include streamlining the rule encoding process by creating a rule wizard to allow derivative classifiers to assist and accelerate development of ontologies and encoded classification guides. These ontologies will also be used to pre-process documents to consistent semantic standards for the machine learning algorithms. LLNL is developing an example-based algorithm that leverages a multitask question and answering network (MQAN) deep learning model to automatically classify documents, combined with a decision tree structure to provide reasoning for the classification recommendation. LLNL's goal is to replicate the "evidence" DCs use within a document to make classification decisions. It is believed that the decision trees implicitly created by DCs as they reason through this evidence may be explicitly replicated, and applied, electronically..

Point of Contacts: Dr. Andrew P. Weston-Dawkes

3) ***Project Title:* Collaborative Data Set and Multi-lab System Architecture**

Goals: Y-12 is developing the ACTICI Collaboration Data Set, a key resource for all of the ACTICI project teams. High-quality, large-volume data is a necessary prerequisite for modern AI solutions. The collaboration data set consists of DOE documents pertaining to a specific classification guide. All documents have been reviewed and marked for classification level and category. A subset of the documents are also highly annotated with tagged terms and the classifier's decision rationale. This resource is used as training and validation data for example-based methods and as validation data for rule-based methods. Y-12 is also designing and prototyping elements of an integrated, multi-lab document review system. Development efforts are currently focused on user interfaces, document preprocessing, algorithm interface standards, and human-in-the-loop feedback mechanisms.

Point of Contact: Dr. Andrew P. Weston-Dawkes

S4 – Under Secretary of Science: AI Data Call Response

Office of Science

What AI technologies are being developed in DOE?

DOE labs are developing:

- AI methods that have a confidence associated with them that enables predictions – these methods incorporate Uncertainty Quantification, simplify the learned models to explain their basis and provide verification and validation of AI systems
- Foundational approaches that leverage math and computer science expertise (e.g., optimization, optimal control, synthetic data) to improve learning efficiency and accuracy in AI systems and within the context of theoretical limits on computability, expressability, and learnability.
- AI methods that work on problems with known physical laws so the learning is consistent with the laws of physics, etc.
- Methods that go beyond 2D images to work with higher dimensional scientific data (4D, 5D, etc.) for classification and generation problems
- Generative models that are trained on large-scale scientific data (materials, genomes, proteins, etc.) and can be used to generate new instances
- Text processing methods that can work on scientific texts, including mathematics and can be used for concept learning
- AI techniques for hypothesis generation, design and control of experiments, including the ability to control a large set of experiments or simulations, e.g., to decide what experiments and what parameters
- Methods that work with DOE large facilities to design new devices and control highly sensitive instruments such as accelerators, networks, and
- Methods that run at scale on DOEs petascale and future exascale platforms

What AI technologies from industry are being used in DOE?

- DOE labs are using nearly all the primary deep learning technologies developed by industry including those methods for computer vision, text comprehension, transformers, etc. (the number of AI methods that have been developed is in the hundreds)
- DOE labs are using the major AI frameworks (Tensorflow, Pytorch, Mxnet, Horovod, etc.) developed by industry and building systems that leverage them
- DOE labs are also beginning to acquire and test AI hardware accelerators specifically built for 10x-100x acceleration of deep learning applications
- DOE labs are also experimenting with some of the large-scale industry workflow systems developed to support some of the largest industry projects (MagLev for self-driving cars, Open AI Gym for reinforcement learning, etc.)

What are DOE's comparative advantages in AI?

DOE labs have:

- Over 1,000 computer scientists and mathematicians that understand HPC, AI, machine learning, optimization, data science, data analysis, statistics
- World-class domain scientists in DOE science and engineering mission areas that understand science, data and data analytics
- Some of the largest and most capable computers for AI training that exist, enabling DOE to work on comparable problems with the main AI players in industry (Google, Microsoft, Amazon, etc.)

- Large-scale facilities and science programs that generate huge volumes of data from light-sources, neutron sources, microscopes, detectors, telescopes, radars and other instruments. This data is a rich source of training data for machine learning
- World leading systems software and expertise to build complex workflows needed for data intensive science such as AI training and inferencing

Are there additional AI areas where DOE should have a comparative advantage?

- DOE has probably the most experience in the federal government of working with computer vendors and academia to design future generations of computers. DOE could work with these partners to develop AI specific machines for the future

- DOE has an excellent track record in running large-scale scientific projects and could apply this experience to the development of mission critical large-scale applications of AI

- Because of the close proximity of DOE's computer scientists and mathematicians with the physical, biological scientists and engineers, DOE is an an excellent position to drive the development of AI methods for Science and Engineering

- DOE should leverage its distinguishing facilities to develop a full AI "ecosystem," including data management, workflow, and "edge" analytics

What limitations (hardware, software, data) does the DOE foresee that will slow advancement in AI?

- Depending on how fast DOE scientists adopt AI based methodologies the demand for AI computing capability could grow much faster than the computing capabilities planned by DOE. There is large uncertainty in the rate of adoption of AI in labs at present

- While DOE collects large quantities of data, investment levels in data curation, cleaning, validation and data management systems could become limiting factors

- DOE labs and facilities need to transition to a model where every scientist and engineer has access to AI capabilities to drive productivity and capability. This will require new software and data infrastructure investments at the labs across the program space.

- The limitations of current algorithms will become apparent, particularly with respect to the representation and transfer of knowledge, available information content in data, and the assurance of AI systems. This will slow adoption in science and engineering disciplines pending significant algorithmic advances.

How should the DOE address the cyber-security of AI driving elements in data and/or the AI hardware and software supply chain?

- In general DOE should apply existing cyber-security models that apply to data and computation to emerging scenarios in AI including provenance of AI hardware and software

- DOE should support research in AI methods that can compute and train on encrypted (and compressed) data or on data that has been encoded to provide privacy, e.g., differential privacy.

- DOE should support research in secure AI training methods that permit the exchange and local training of models rather than exchange of data

- DOE should support research in using AI to enhance cyber-security methods.

- Operationally DOE should use a risk-based approach to cyber-security of AI data and models, to balance open scientific exchange in a rapidly moving field with privacy and national security as appropriate to the use case

What systems or processes does the DOE employ to encourage the use of AI internally?

- DOE has a data management policy that encourages the collection and dissemination of data needed for training models

- DOE has supported pilot projects to bootstrap AI programs at the laboratories

- DOE has created internal teams and AITO to encourage the exploration of AI opportunities across the DOE system

- DOE laboratories have made internal investments via LDRD and program development funds to bootstrap AI projects, teams and infrastructure

Community Input is Crucial :DOE has sponsored workshops in many programs to provide vision and direction for AI research and development

- Basic Research Needs for Scientific Machine Learning, January, 2018

- Fusion Energy Machine Learning / AI Workshop, April 30 – May 2, 2019

- Data for AI Roundtable, June 5, 2019

- AI for Science Town Halls, July, August, September and October, 2019

- Basic Energy Science Roundtable to identify opportunities for artificial intelligence (AI) methods and machine learning (ML) tools to address the production, mining, analysis, and control of large data sets generated at the existing and future BES scientific user facilities, October 22-23, 2019,

How do the labs foster the use of AI?

- DOE laboratories have created internal initiatives (LDRD, program development, etc.) aimed at increasing the strategic development of AI projects and programs

- DOE labs have created partnerships with academia and industry to pursue AI driven projects of mutual interest (e.g. ATOM, INSPIRE, etc.)

- DOE and the labs held a AI innovation summit XLab to bring industry, labs and academia together around a set of a AI topics in healthcare, energy, manufacturing and urban science

- DOE labs are hiring AI and machine learning researchers to nucleate additional capabilities, including use of various early career awards

- DOE scientists present papers, lead workshops and other activities at major conferences, including Supercomputing, SIAM and IEEE meetings to foster broad collaborations and research with DOE scientists.

How has or should the DOE built platforms to help others develop or use AI?

- DOE is investing in several software projects such at the CANDLE and ExaLearn ECP project aimed at building AI software tools, AI workflows and libraries for enabling broad collection of AI applications to be developed and run on DOE supercomputers.

- DOE has always been a leader in team-based interdisciplinary science with programs like SciDAC, and has additional science-focused projects (CAMERA, ExaSheds, and ECP projects ExaBiome, ExaFEL, and ExaGraph) using AI methods, but these efforts should expand broadly across DOE

- DOE labs are also creating open source AI software and applications that enables third parties to accelerate their AI developments (e.g. LBANN, and many others).

Office of Science AI/ML research

The Office of Science (SC) is uniquely positioned to not only benefit from but also advance current AI activities. Each program identified their unique areas of interest in AI for FY 2020:

- *Advanced Scientific Computing Research* -- Activities will focus three basic areas: foundational research in computer science and applied math to improve the reliability, robustness and interpretability of big data and AI technologies through the development of new algorithms, methods and software tools; co-design of a distributed computing ecosystem to ensure seamless integration of ML and computing resources and partnerships to broaden the applicability of AI and big data solutions across a broad range of DOE applications.

- *Biological and Environmental Research* -- The Data Management effort will initiate new research on using applying and demonstrating artificial intelligence (AI) and machine learning (ML) tools to enhance predictability of the atmospheric boundary layer and watersheds, based on data derived from modeling, simulation, and field observations.

- *Basic Energy Sciences* — A core research priority is data analytics and machine learning for data-driven chemical and materials sciences; FY 2020 core program funds will support continuations of awards that are expected to be initiated in FY 2019. The FY 2020 facilities budget includes $10M for applications of artificial intelligence methods and machine learning techniques to accelerator optimization, control, prognostics, and data analysis, which will bring new software and hardware advances to help address the BES user facilities' data and information challenges.

- Fusion Energy Science— FES has been supporting pilot efforts in ML/AI and data science in FY 2018/FY 2019. FES, jointly with ASCR, plans to conduct a workshop on "Advancing Fusion with Machine Learning" on April 30 – May 2, 2019, to identify priority research opportunities in ML/AI and data science for fusion and plasma science. In FY 2020, informed by the findings of this workshop, FES plans to issue a Funding Opportunity Announcement and a companion Announcement to DOE national laboratories to competitively select research projects in this critical area.

- *High Energy Physics* Research priorities will leverage DOE high performance computing resources to scale up the performance, validation, and analysis of simulated and physics data; improve production-quality tracking with pattern recognition for online triggering systems of HEP experiments; and design specialized hardware for computation and curation of data sets. HEP will seek crosscutting AI/ML solutions across the HEP experimental frontiers, theory, and technology thrusts and will partner or coordinate with other DOE funded research, facilities, and projects as relevant.

Innovations Possible with AI

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 Common Part of Scientific Laboratory Activities:** Infuses scientific, engineering and operations

Earth and Environmental Sciences: State of the Art

A *predictive understanding* of the Earth system is crucial for utilizing its energy and water resources while mitigating costly environmental hazards. AI approaches are playing useful roles in:

- Geophysical characterization and change detection
- Data assimilation and model–data integration

- Data-driven and physics-informed machine learning
- Surrogate modeling

Water Security

Ensure global water security under a changing environment

- Water resources are critical for energy production, human health, food security, and economic prosperity
- Water availability and water quality are impacted by environmental change, weather extremes, and disturbances such as wildfire and land use change
- Methods are needed to integrate disparate and diverse multi-scale data with models of watersheds, rivers, and water utility infrastructure
- Predictions of water quality and quantity require data-driven models and smart sensing systems
- Water resource management must account for changes in weather extremes, population, and economic growth

Biodesign

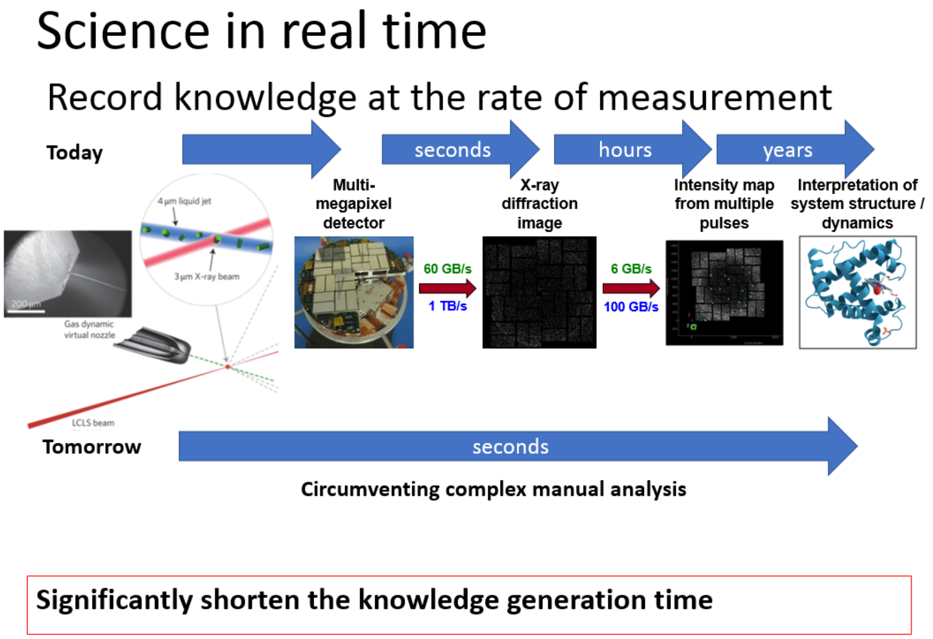
With a robust biodesign capability we could…

- Replace chemical factories with small safe portable biomanufacturing
- Democratize and accelerate drug development
- Produce novel food grade protein and fiber sources
- Produce biological carbon capture systems
- Produce designer polymers that are environmentally benign
- Harness bespoke biological systems for water purification
- Integrate 3d printable bioinks with biological computing and control to produce new types of smart matter

Science in real time

Record knowledge at the rate of measurement

Maximizing the Value of DOEs largest scientific facilities with AI Accelerators rely on operation and high precision tuning of hundreds of thousands of parameters simultaneously. We can use AI to create self-driving, high-performance accelerators.



Science in real time

Record knowledge at the rate of measurement

Circumventing complex manual analysis

Significantly shorten the knowledge generation time

- Customizable shot by shot configurations for xFEL experiments
- Reducing time to configure experiments from one day to one minute
- Orders of magnitude improvement of source to detector stability

On-the-fly materials & chemical synthesis

Transformative advances are possible. How? Use autonomous-smart experiments and simulations enabled by AI

- requires a new class of machine learning algorithms that continually learn and update their predictions based on new data sources, encode physical constraints/laws and models, and learn to estimate fidelity

Delivering materials/chemicals potentially up to ~**1000x faster** and with desired performance/properties

Universe: The Movie

Reconstruct the past from the Big Bang until today and predict the future of our visible Universe, from the largest scales down to our own galaxy.
Use all existing data (galaxy positions, stellar mass, velocities, dark matter maps, gas distribution, tSZ, kSZ, X-ray).
What is dark energy? What is its density evolution in time? What is the nature of dark matter? Did inflation happen?
Provide tightest possible constraints on fundamental physics questions by solving optimal inference problem.

- GANs for image emulation
- GP and DL-based emulators for summary statistics
- CNN-based image classification
- AI-based photometric redshift estimation

- AI methods for inference and reconstruction

Office of Environmental Management

Office of Technology Transfer

What AI technologies from industry are being used in DOE? OTT has used AI to improve LPS accuracy. This information has been shared with the RTIC subcommittee. Over the course of FY19, the Lab Partnering Service (LPS) began improving the quality of LPS content by introducing machine learning techniques to the classification of patents. Prior to the introduction of a machine learning models, and by using an Automated Form Submission process, LPS working with the National Renewable Energy Laboratory (NREL,) was achieving an overall accuracy rate of 48.6% when matching patents and applications to Assignee National Laboratories. After the application of a machine learning model based on the OneVsRest model where an independent classification model is trained for each possible label, NREL was able to achieve an overall accuracy rate of 99.8% with a mean accuracy score of 98.8%. In FY20, LPS will continue to improve the performance by developing neural networks to assign patents to national labs, to improve classification of patents by domain of endeavor, and import other Government Agencies' innovations into the classification system which we have designed for LPS.

Also, DOE/LPS has now added a NASA curated list, of NASA patents to the Visual Patent Search, and the same AI capabilities were used to ensure that each patent was correctly matched. Additionally LPS will begin using neural networks and ML to improve the classification of those innovations (patents) on LPS. By classification we mean, which area of Science or Engineering any single patent might fall into. That effort is underway and will complete at the end of this year, if not earlier.

What limitations (hardware, software, data, etc) does the DOE foresee that will slow advancement in AI? Computing capacity. We are heavily focused on the science of cutting-edge computing at the DOE, and you don't need the latest and greatest computer to run AI simulations. A 200 PF computer certainly doesn't hurt, but you can do a lot with a 2 PF computer as well. The investment approach for maximizing AI outcomes from computing capacity would not align with DOE's current approach of constantly breaking new ground in computing power/speed; rather, an AI-optimizing investment strategy would like to drive down hardware costs through replicability/scalability. A great AI portfolio wouldn't have machines showing up on the Top500 supercomputers ranking; they would be work horses. Marcos has shared this perspective with the RTIC AI Subcommittee as well and it's part of the discussion around developing an AI Strategy for the DOE.

How do the labs foster the use of AI? OTT has helped promote AI-related research activities through the recent AI XLab Summit, which brought together industry and Lab practitioners to identify challenges and opportunities that would benefit from the application of AI techniques, and to highlight DOE assets that can be leveraged to support these efforts.

S5 – Under Secretary of Nuclear Security: AI Data Call Response

NNSA Advanced Simulation & Computing (ASC) Program (NA-114):

Project Title: Advanced Machine Learning (AML)

Project Summary: Starting in FY20, ASC will stand up an Advanced Machine Learning (AML) portfolio with the overall goal of improving efficiencies in weapons design and stockpile assessment through advanced data analysis. This program will increase NNSA's agility, enable greater exploration of design space, and improve predictive capabilities, while potentially lowering overall costs.

Lawrence Livermore National Laboratory (LLNL)

1. *Design Workflow* - Define and implement longer-term solutions for collecting, managing and querying data and learned models, with a focus on sustainability and increased flexibility in ML training.

2. *Hardware Investigation* **-** Work closely with hardware vendors to understand and influence next generation hardware ML and its integration with CPUs and GPUs. Examples include coordination of datasets with vendors, exploration of early hardware design, and setting up testbeds.

3. *System Software for Machine Learning -* Create and integrate interfaces between applications and advanced hardware, with a goal of meeting scaling and performance needs while reducing the burden on the application developer. Examples include working with Sandia's FUGU (neuromorphic) interface and exploring porting of frameworks to advanced hardware.

4. *Data Collection & Management* - Define and implement longer-term solutions for collecting, managing and querying data and learned models, with a focus on sustainability and increased flexibility in ML training.

5. *Surrogate Models* - Use advanced ML techniques to create surrogate models that can be used as low-computational-cost approximations of more expensive calculations. The focus of this work is to enable fuller exploration of design space and improve current solutions for low cost approximations. Examples include improved constitutive models and optimizing simulation parameters for high cost calculations.

6. *Augmented Physics Algorithms* - Use advanced ML techniques to augment our current physics algorithms to create more accurate solutions, with a focus on reducing computational cost and adhering to physical constraints. Examples include better capturing interatomic potentials, improved opacity lookups at runtime, and improving turbulence modeling.

7. *Uncertainty Quantification (UQ)* - Evaluate emerging UQ techniques for ML, focusing on Deep Learning, for ASC applications. This includes research that is ongoing under LDRD and through interagency collaborations.

Los Alamos National Laboratory (LANL)

1. *Quantum-aware Molecular dynamics potentials using active learning* - Use first principles Density Functional Theory (DFT) calculations to learn potentials and predict electronic state.
2. *Upscaling crack interactions for improved damage models* - Predict crack interaction, coalescence and growth statistics to calculate improved effective moduli to be used in continuum.
3. *Uncertainty Quantification on Neural Networks* - quantify uncertainties on the reduced order model using active learning for the above 2 multi-scale problems.
4. *Radiograph Analysis* - Use image analysis techniques to analyze synthetic radiographs, as well as analyze patterns in data and relate to outputs of simulations.
5. *Nuclear Data Evaluation* - Extend model calibration methods for estimating nuclear data from heterogeneous sources including integral and differential data, using ML methods.

6. *Materials Model Prediction using Advanced Learning*- Optimize parameters in various strength models using multi-source experimental data.

7. *Artificial Viscosity* - Use intelligent adaptive sampling techniques to reduce the number of simulations needed to tune artificial viscosity parameters, compared to space-filling ensemble designs for shock tube problems.

8. *Nuclear Emergency Response* - Accelerate the emergency response workflow using ML to estimate key metrics

9. *High Performance Computing (HPC) Analysis* - Support of ML infrastructure and data collection for early identification of HPC system issues

Sandia National Laboratories (SNL)

1. *Machine Learning for Reduced Order Models* - develop a faster, more efficient means of simulating complex physics problems and perform high-fidelity simulations.

2. *Accelerating Calculations of Fluid Flow via Physics-Informed Machine Learning Models* – Train a ML model using Sandia's CTH code, with end goal of computing pressure waves from 10,000 fragments with a much shorter time-to-solution.

3. *Deep Learning Enables Discovery of Anomalies in High-Reliability Components* - enable discovery of anomalies in high-reliability electronic components.

4. *Parallel Training of Deep Residual Neural Networks* - develop a faster method of training deep neural networks on a HPC system and to provide a rigorous theoretical understanding of the training process.

5. *Uncertainty Quantification for High-Consequence Deep Learning Applications* - leverage uncertainty maps to enable generalization of a trained model to shifted domains.

6. *Machine Learning Enabling Automatic Mesh Generation* - use ML to enable automatic meshing of parts to reduce set-up time.

7. *Machine Learning for System Software* - use ML to infer intelligent approaches to resource management, reinforcement learning in particular, where an agent makes actions given state observations from an environment.

Project Goal: The ASC program is already beginning to see successes in limited evaluation of machine learning (ML) techniques, such as the use of surrogate models for fast evaluation, improved accuracy in reduced-order turbulence modeling, and simplifications of workflows, but many possible technical advances remain. Combining our efforts to tackle fundamental research challenges while developing a multi-disciplinary data science workforce is essential. ASC is well positioned to leverage investments in experimental facilities, next-generation computer architectures, algorithm development and simulation data collection. Other internal ASC initiatives, such as the Large-Scale Calculations Initiative (LSCI), allow ASC to develop ML workflows that utilize multi-source, multi-fidelity data for answering questions about the stockpile.
Points of Contact: Thuc Hoang, David Etim

Artificial Intelligence Threats and Protections for Global Material Security (NA-21)

AI enables exciting new technologies for protecting nuclear assets and detecting illicit materials and activities, but it also provides adversaries with a new and powerful tool to design attacks against nuclear facilities, systems, and supply chains.

Project Title: Impacts of AI for Global Material Security

Project Summary: The Emerging Threats and Technologies Working Group in DNN's Office of International Nuclear Security (INS) is currently studying the following aspects of AI.

1. Adversarial attacks, describing how a rogue state or actor might use Deep Learning (DL) to devise new attacks, enabling material concealment, delivery, theft, and sabotage,
2. Nuclear material protections, describing how the US and its partners might employ DL for enhanced security or to how to respond to DL-driven attacks, enhancing defenses and protections and detection of material outside regulatory control, and
3. Vulnerabilities from Deep Learning, describing new risks and new attack approaches that come from adopting AI-based systems, potentially enabling theft, concealment, and sabotage.

With respect to adversarial attacks, the focus is on applications that are enabled by Deep Learning, meaning new attack pathways that didn't exist prior to the current prevalence of DL capabilities. The working groups is also developing a threat likelihood and severity assessment for each use case, in order to provide a framework to prioritize initiatives and investments.

Project Goal: Recommend actions for Global Material Security (GMS) to mitigate risks associated with each aspect of AI considered, and identify areas for further detailed studies.

Point of Contact: Allison Johnston, Director, DNN Office of International Nuclear Security

DNN R&D Data Science (NA-22)

The Data Science portfolio focuses on the development of novel methods and technologies for real-time and early detection of proliferation activities leveraging emerging technologies, artificial intelligence, and applied mathematical and statistical methods. Current projects in the Data Science portfolio include artificial intelligence approaches applying machine learning, optimization, probability and statistics, and decision theory as well as advanced applied mathematics such as tensor factorization and graph analysis.

Project Title: Advanced Analytics for Proliferation Detection (ADAPD)

***Project Summary*:** Current state of the art in proliferation detection uses models, simulations, and experiments of single physics-based and chemical phenomena to develop and validate expected signatures of illicit and nefarious activities related to the proliferation of nuclear weapons and threats. The ADAPD Venture will improve USG capability to detect proliferation by developing new signatures that incorporate multiple phenomena, incorporate temporal- and spatial-based patterns, and leverage pattern-of-life and human-generated data streams. R&D in ADAPD leverages AI and to overcome both big data and sparse data problems in this domain. Additionally, ADAPD leads DNN in developing statistically-based approaches to quantify uncertainty and attribute features in AI, particularly in methods combining multiple mathematical and machine learning methods.

Accomplishments: ADAPD has improved DNN R&D's capability to model and detect EM signals of interest by training a deep learning classification model using simulated data combined with real-world data collected from NNSA experiments as well as a generative neural network approach to improve emulated waveforms of background EM. ADAPD has also developed analysis pipelines using machine learning, natural language processing, and graph analysis to identify communities of interest in global-scale data streams of technical publications, Finally, ADAPD has developed models of complex activities combining Bayes-net based generative models and constraint-based reasoning to predict the timing of activities of interest.

Point of Contact: Ms. Angela Waterworth, Technical Advisor,
DNN Office of Proliferation Detection

Project Title: Multi-Informatics of Nuclear Operations Scenarios (MINOS)

Project Summary: The MINOS Venture seeks to employ and develop advanced analytic techniques to use disparate data sets from measurable phenomena associated with nuclear reactor and nuclear target processing facilities to characterize operations similar to those at a proliferator nuclear fuel reprocessing facility. To this end, MINOS is establishing a test bed at a DOE co-located nuclear reactor and target processing facility from which signals data will be gathered along with all necessary ground truth; creating an innovative, data-sharing infrastructure, with necessary metadata, that is readily accessible, useable, and scalable; exploring existing and developing new data fusion techniques to establish a capability to better characterize operations at the test bed; investigating the transferability of the methods to other nuclear operation facilities of interest.

***Accomplishments**:* MINOS has developed a suite of machine learning-based algorithms to characterize reactor power levels using single sensor data streams (e.g. seismic waveforms or thermal spectra). Data scientists on MINOS have begun to explore the performance of models that combine data streams from multiple sensors (e.g. radiation and thermal measurements analyzed together). Toward supporting the future use of AI, MINOS has also generated a huge set of discoverable, machine-readable, and temporally/spatially synchronized measurements by 10 different sensors types of operational activities at the research reactor at Oak Ridge National Laboratory. This dataset is labelled and includes ground truth of the operations at Oak Ridge. Additionally, MINOS has developed a cloud-based data management infrastructure accessible to researchers across the Lab Complex and academia, and visualization and data analysis software products.

***Point of Contact**:* Ms. Angela Waterworth, Technical Advisor,
DNN Office of Proliferation Detection

Project Title: Shadow

Project Summary: The Shadow project is developing AI-based methods to improve the US capability to detect and characterize ground-based events of interest using seismic waveform data. In particularly, to overcome the prohibitive cost of generating labeled data of such events to train ML models, Shadow investigates the use of semi-supervised learning approaches that allow unlabeled data to be integrated into a supervised classification methodology to improve performance.

Accomplishments: Shadow has developed machine-learning and deep learning models using a semi-supervised training approach that outperforms existing, manual approaches to classify ground events for earthquakes and other events in Utah. Shadow has also developed software tools for automated pre-processing of seismic data to support data analysis that have been made available to the community of seismologists and proliferation detection researchers within the Lab Complex. Finally, the Shadow project team has worked with a targeted operational end user of the developed technology to adapt their methods to characterize events in country of interest.

***Point of Contact**:* Ms. Angela Waterworth, Technical Advisor,
DNN Office of Proliferation Detection

Project Title: Ecosystem for Open Science (eOS)

Project Summary: The development of DNN R&D capabilities requires world class infrastructure and expertise in traditional fields as well as emerging computational and analytics disciplines. Leveraging state-of-the-art technologies and capabilities developed in industry, academia, and the Lab complex, eOS will pilot a cloud-based workspace where program researchers can explore program data and develop and apply advanced data analytics, tap into deep repository of domain knowledge generated by experts at the Labs, and easily and efficiently collaborate with multidisciplinary teams across the Lab complex.

***Goals**:* Build COTS-sourced, government-certified cloud platform that delivers one-click access to industry-class data management infrastructure for NA-22 research community; knowledge management including data repository, data catalog and extraction of program documents and data; digital collaboration through customizable workspaces throughout project lifecycle; analytics workspace with compatible with state-of-the-art computing and analytics software and techniques. eOS will achieve initial operating capability of an unclassified platform in 18 months; a classified platform with cross-domain pipelines will be rolled out in later phases of development.

***Point of Contact**:* Ms. Angela Waterworth, Technical Advisor,
DNN Office of Proliferation Detection

DNN R&D Remote Detection (NA-22)

The Remote Detection portfolio seeks to provide tools which locate, detect, and characterize nuclear facilities and activities from the fence line to 100s of kilometers away. There are two active AI/ML projects looking at advanced modeling and predictive sensing of proliferation activities.

Project Title: Modeling and Integration of Remote Sensing

Project Summary: The Modeling and Inference for Remote Sensing (MIRS) project objective is to evaluate methods for nuclear non-proliferation assessments - in particular, detect undeclared nuclear facilities using computational models of the nuclear fuel cycle that integrate information from multiple sensors. The MIRS project will test the hypothesis that computational modeling and inference methods can be used to integrate sparse remote sensing information across the fuel cycle to draw stronger conclusions about the activities at unknown facilities than can be done with sensors alone. A key contribution of the computational method is the integration of observables from facility activities and from the trail of goods, including vehicles, equipment, and materials, moving from one facility to the next.

Accomplishments**:** Machine Learning algorithms have been evaluated to distinguish between material diversion scenarios and determine which features (sensors) are critical for distinguishing between scenarios. The team also designed a convolutional neural network (CNN) model aimed at classifying whether or not a particular simulated facility scenario was misreporting its association with the nuclear fuel cycle.

Point of Contact**:** Dr. Christopher Ramos, Physical Scientist, DNN Office of Proliferation Detection

Project Title: Persistent Dynamic Nuclear Activity Monitoring through Intelligent, Coordinated Sensing (DyNAMICS)

Project Summary: The Persistent DyNAMICS (Dynamic Nuclear Activity Monitoring through Intelligent, Coordinated Sensing) Venture will design, build, and demonstrate an architecture for dynamic persistent monitoring of nuclear processes through intelligently coordinated sensing. Persistent DyNAMICS will exploit and demonstrate the mission relevance of sequences of proliferation activity signatures. This project will establish analytical techniques that couple sequenced observations of diverse physical phenomena with predictions based on subject matter expertise to inform collections and infer nuclear activities. In the final product, collections will occur through autonomous coordination that is dynamically tailored for each event/activity in real-time via Artificial Intelligence. The collection queuing choices will be driven by inferred nuclear processes and activities.

Goals**:** A significant task of this project is to develop or adapt inference software to address research questions related to the characterization of nuclear facility activity. Persistent DyNAMICS will use data-driven machine learning analysis of sequences, independent of nuclear fuel cycle process models, to provide cross-validation and will develop Fast Numerical Models to train, tune, and test stochastic models and inference engines.

Point of Contact**:** Dr. Christopher Ramos, Physical Scientist, DNN Office of Proliferation Detection

DNN R&D Safeguards (NA-22)

The Safeguards portfolio develops a set of technical measures designed to verify the correctness and completeness of declarations made by States about the use of their nuclear materials, their fuel cycle activities, as determined through the legal framework outlined in the Nuclear Nonproliferation Treaty and each State's safeguards agreement, as well as to detect undeclared nuclear material and activities.

Project Title: Data Science Methods to Improve the Efficiency and Effectiveness of Safeguards Verifications

Project Summary**:** Nuclear safeguards is a data-rich field that spends significant amounts of labor obtaining complicated sets of data from disparate sources to ensure safeguards obligations are met. NA-22 Safeguards R&D is investing in the development of modern data analytics techniques (machine learning and artificial intelligence) to determine: 1) the value of traditional and non-traditional datasets, and 2) evaluate the most effective and efficient use of these datasets for determining material diversion or facility misuse. Three national laboratories (LANL, ORNL, SNL) are working together to develop and test methods that investigate machine learning and other data science approaches for data assurance, signature discovery, and patterns of activity.

***Accomplishments*:** Preliminary results have shown the use of machine-learning methods for integrating disparate data streams from patterns-of-life sensors and various datasets can provide indications of activities in safeguards training facilities, the ability to identify and characterize predictive signatures in material processes, and rapidly find anomalies in large datasets that can be related to facility and plant operations.

Point of Contact: Dr. Christopher Ramos, Physical Scientist, DNN Office of Proliferation Detection

DNN R&D Nuclear Test Detection (NA-22)

The Nuclear Test Detection Portfolio seeks to develop new methods and technologies to detect, locate, and characterize underground nuclear tests and provide technologies to enable timely notification of probable nuclear events with accurate location estimates.

Project Title: Low-Yield Nuclear Monitoring Dynamic Networks (LYNM-DN)

***Project Summary**:* The Dynamic Networks Venture seeks to exploit the power of open network data and new data types to significantly lower detection thresholds of low-yield, evasive underground nuclear explosions without increasing time-to-detection or number of human-analyst reviews. Project goals are to identify key technological, algorithmic, and scientific challenges to processing traditional and open-network, multi-phenomena data and develop and evaluate new technologies, algorithms, and architectures to meet these challenges. The focus is on exploiting signatures of low-yield, evasive tests from dynamic sensor networks at local distances (<200 km).

Accomplishments: Data scientists within the Dynamic Networks have developed a suite of machine learning algorithms to classify earthquake and explosions at catalogued by the University of Utah Seismograph Stations (UUSS network). Additionally, Dynamic Networks has created a benchmark standard dataset of these events (10,644 events from 2012-2017 at 132 stations) including ground-truth labels generated by a subject matter expert to support the development of AI models to classify ground events.

Point of Contact: Mr. Brian Paeth, Physical Scientist, DNN Office of Proliferation Detection

Surplus Plutonium Disposition Research (NA-23)

There is one current project underway and one potential project being evaluated. Both projects will evaluate the use of robotics to improve the efficiency of process steps within the surplus plutonium disposition. This includes packaging and glovebox operations.

Project Title: Automated Processing of 9975 Shipping Containers

Project Summary: This project explores alternative methods in automation for retrieving and unloading radioactive material shipping containers. Automating this process will utilize mature commercially available technologies.

Project Goals: The Automated handling of the shipping containers. The basic steps are:

a) Retrieve the pallet with the desired drum with an Automated Guided Vehicle (AGV) forklift and transport the palletized drums to an unloading station.

b) Unload the 9975 and remove the 3013 container.

Point of Contact: Paloma Richard, Program Analyst, DNN Office of Material Disposition.

Safeguards Technology Development (NA-24)

1) Project Title: Using Deep Learning Algorithms to Enhance Image-Review Software for Surveillance Cameras

Project Summary: Surveillance cameras are a core monitoring technology used by the International Atomic Energy Agency (IAEA) Department of Safeguards. Current image-review software has limited automated functions requiring time-consuming manual inspection performed by inspectors. Advanced machine learning algorithms can flag objects and anomalies for more efficient inspector review.

Goals: Adapt AI technologies to automatically review surveillance camera images. Current technologies being adapted are neural networks for object detection and deep convolutional recurrent neural networks for change prediction. These techniques are being incorporated into a new review software module that is suitable for evaluation by IAEA inspectors.

Point of Contacts: Dr. Arden Dougan, Program Director, DNN Office of Safeguards Technology Development

Project Title: Nondestructive Assay Fingerprinting of UF_6 Cylinders

Project Summary: Monitoring UF_6 enrichment and mass throughout its life within the nuclear fuel cycle relies on periodic inspections and through-the-wall measurements on cylinders that are completed with varying accuracy. Radiation emissions from UF_6 cylinders are complex. At any given time and cylinder location, the emissions profile is a function of the: geometry and filling/washing history of the cylinder; filling, sampling, handling, and environmental details of the cylinder; and fuel cycle history of the uranium. This project is working to determine the feasibility of detecting UF_6 cylinder declaration inconsistencies using measurements of radiation emissions.

Goals: Apply a variety of data analysis techniques, including support vector regression (a popular machine learning tool), to determine a set of requirements for measuring UF_6 fingerprint signatures that lead to high confidence in safeguard measurements.

Point of Contacts: Dr. Arden Dougan, Program Director, DNN Office of Safeguards Technology Development

Project Title: Statistical Methods for Evaluation of Environmental Sample Data

Project Summary*:* Analytical results from environmental samples are used to evaluate and verify safeguards declarations. Results are compared to historical data and evaluated in the context of declared activities. Simplicity of model assumptions may limit interpretation of the data. A robust, mathematically based, statistical model is being developed to evaluate the consistency of new environmental sample results with historical and declared activities.

Goals: Develop an automated analysis tool using various techniques, including Bayesian statistical methods and artificial neural networks, to identify anomalous radioisotope measurements in environmental samples.

Point of Contacts: Dr. Arden Dougan, Program Director, DNN Office Safeguards Technology Development

Safeguards Concepts & Approaches (NA-24)

Project Title: Artificial Intelligence, Machine Learning, and Safeguards

Project Summary: This study will analyze two machine language (ML) methods: the one-class support vector machine (OCSVM) and a long short-term memory (LSTM) neural network for application to safeguards at reprocessing or enrichment plants. The two methods were chosen because of their ability to learn in an unsupervised environment with unlabeled data, an important factor for safeguards because most data from reprocessing and enrichment facilities will be heterogeneous and unlabeled. This study will include an evaluation of direct impacts such as possible time savings for inspectors, resource maximization, and effectiveness of pattern or anomaly detection.

Goals: Improve awareness and understanding of AI and ML and their potential benefits, challenges, and how their application could improve the effectiveness and efficiency of international safeguards implementation.

Point of Contacts: Wayne Mei, Program Director, DNN Office of Safeguards Concepts & Approaches