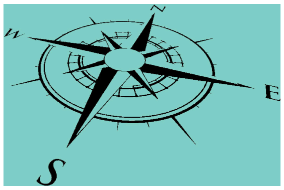
# DATA, ARTIFICIAL INTELLIGENCE, AND MACHINE LEARNING AT DOE SCIENTIFIC USER FACILITIES

The U.S. Department of Energy's (DOE's) scientific user facilities provide access to the world's most advanced research instruments and produce increasingly larger quantities of data. While impressive, reaching the full potential of the rapidly growing data streams will require new innovations to solve a variety of technical challenges in data acquisition, control, modeling, and analysis. Artificial intelligence and machine learning (AI/ML) have opened corresponding new avenues in optimization, efficient surrogate models, data analytics, and inverse problems. These intriguing capabilities suggest that AI/ML can greatly accelerate the quest to probe and understand fundamental phenomena across a vast range of lengths and timescales, potentially leading to transformative advances across scientific disciplines.

We envision a future of AI/ML-enabled facilities that maximize the DOE's scientific impact. As a first step towards this vision, DOE announces its interest in receiving National Laboratory proposals from the twelve Basic Energy Sciences (BES), three High Energy Physics (HEP), and three Nuclear Physics (NP) Scientific User Facilities to support data, artificial intelligence (AI) and machine learning (ML) research and development of tools addressing the priorities identified in the Supplementary Information below. It is expected that the proposed research will impact other Office of Science facilities and programs and envision future partnering with those facilities and programs to further evolve and strengthen the advances.

### *Contents*




# DEPARTMENT OF ENERGY OFFICE OF SCIENCE (DOESC)

## Description:

The Department of Energy (DOE) Office of Science (SC) performs many functions for DOE national laboratory proposals in the Portfolio Analysis and Management System (PAMS), which is available at https://pamspublic.science.energy.gov.

### Stakeholder(s):

**Technical/Scientific Program Contacts**

**Dr. Eliane Lessner** :
*Basic Energy Sciences*

**Dr. John Boger** :
*High Energy Physics*

**Dr. Manouchehr Farkhondeh**

**Industry** :
*Both industry and science already use AI/ML approaches for data analysis.*

**Science** :
*User facilities, however, crucially require AI/ML tools throughout the lifetime of an experiment: not just for data analysis, but also for data creation, acquisition, and storage. In the next 10 years, AI/ML are expected to go beyond traditional data analysis to aid the design and control of complex facilities, enable real-time capabilities to acquire and analyze large data volumes, automatically steer data collection for in-the-loop experiments, and support experimentalists' use of exascale computing. These advances will in turn open new avenues of scientific research in energy sciences and beyond.*

**Scientific User Facilities** :
*In addition, AI/ML approaches can have a significant impact on increasing the operational efficiencies of large, complex scientific user facilities and scientific instrumentation. AI/ML approaches, for example, can be used to predict detector and accelerator component performance which can result in improved performance and higher beam availability for research... The BES, HEP, and NP scientific user facilities (SUFs) annually serve over 21,000 users. Efficient management of the rapidly increasing quantity of data in these complex systems demands increasing human and instrumentation resources. Current and upgraded facilities face a variety of technical challenges in simulations, control, data acquisition, and analysis. AI/ML methods and techniques promise to address these challenges and impart a drastic acceleration of experimental and computational discovery. AI/ML approaches can also facilitate and improve the operations of these complex machines and their instrumentation.*

**Basic Energy Sciences (BES) Scientific User Facilities**

**High Energy Physics (HEP) Scientific User Facilities**

**Nuclear Physics (NP) Scientific User Facilities**

**Eligible Applicants** :
*This is a DOE National Laboratory-only Announcement. FFRDCs from other Federal agencies are not eligible to submit in response to this Program Announcement. Proposals will be accepted only from:*

**BES Scientific User Facilities** :
*Twelve BES Scientific User Facilities (https://science.osti.gov/bes/suf/User-Facilities)*

**Advanced Light Source** :
*at Lawrence Berkeley National Laboratory*

**Advanced Photon Source** :
*at Argonne National Laboratory*

**Center for Functional Nanomaterials** :
*at Brookhaven National Laboratory*

**Center for Integrated Nanotechnologies** :
*jointly managed by Sandia National Laboratory and Los Alamos Laboratory, with locations in Albuquerque and Los Alamos, New Mexico*

**Center for Nanophase Materials Sciences** :
*at Oak Ridge National Laboratory*

**Center for Nanoscale Materials** :
*at Argonne National Laboratory*

**High Flux Isotope Reactor** :
*at Oak Ridge National Laboratory*

**Linac Coherent Light Source** :
*at SLAC National Accelerator Laboratory*

**Molecular Foundry** :
*at Lawrence Berkeley National Laboratory*

**National Synchrotron Light Source II** :
*at Brookhaven National Laboratory*

*— continued next page*

*Stakeholders (continued)*

**Spallation Neutron Source** :
*at Oak Ridge National Laboratory*

**Stanford Synchrotron Radiation Light Source** :
*at SLAC National Accelerator Laboratory*

**HEP Scientific User Facilities** :
*Three HEP scientific user facilities (https://science.osti.gov/hep/Facilities/User-Facilities )*

**Facility for Advanced Accelerator Experimental Tests (FACET)** :
*at SLAC National Accelerator Laboratory*

**Fermilab Accelerator Complex** :
*at Fermi National Accelerator Laboratory*

**Accelerator Test Facility** :
*at Brookhaven National Laboratory*

**NP Scientific User Facilities** :
*Three NP scientific user facilities (https://science.osti.gov/np/Facilities/User-Facilities )*

**Argonne Tandem Linac Accelerator System** :
*at Argonne National Laboratory*

**Continuous Electron Beam Accelerator Facility** :
*at Thomas Jefferson National Accelerator Laboratory*

**Relativistic Heavy Ion Collider** :
*at Brookhaven National Laboratory*

## Vision

A future of AI/ML-enabled facilities that maximize the DOE’s scientific impact.

## Mission

To support data, artificial intelligence (AI) and machine learning (ML) research and development of tools

## Values

**Open Science**: SC is dedicated to promoting the values of openness in Federally supported scientific research, including, but not limited to, ensuring that research may be reproduced and that the results of Federally-supported research are made available to other researchers. These objectives may be met through any number of mechanisms including, but not limited to, data access plans, data sharing agreements, the use of archives and repositories, and the use of various licensing schemes. The use of the phrase “open-source” does not refer to any particular licensing arrangement but is to be understood as encompassing any arrangement that furthers the objective of openness.

**Collaboration**: Multi-institutional and multi-facility partnerships to leverage capabilities are encouraged. Collaborative applications, e.g. submissions of identical proposals by different institutions, will not be accepted under this National Laboratory Program Announcement. Group/team efforts must be submitted by the lead institution.

# 1. Strategic Information

*Efficiently extract critical and strategic information from large complex data sets.*

## 1.1. Meaning

*Extract robust and meaningful information from vast and complex data.*

Address how AI/ML can extract robust and meaningful information from the increasingly vast and complex data now being produced at the SUFs. AI/ML techniques have the potential to significantly reduce the effort to process and analyze the data to obtain the desired physical information.

## 1.2. Complexity

*Find connections elusive to human observations.*

In addition, AI/ML can help unmask the complexity hidden in problems in high-dimensional spaces (multi-modal measurements, many experimental variables, etc.) by finding connections elusive to human observations.

## 2. Autonomy

*Address the challenges of autonomous control and experimentation.*

Incorporate use of AI/ML to address the challenges in the real-time operation of large, complex scientific user facilities.

### 2.1. Searching

*Efficiently search large, complex parameter spaces in real time.*

AI/ML based methods are needed to efficiently search large, complex parameter spaces in real time, and ...

### 2.2. Prediction

*Predict the health and failure of instruments.*

to predict the health and failure of instruments that operate at high power sources and experiments that run on these instruments.

### 2.3. Performance & Productivity

*Improve facility performance and maximize productivity.*

Such capabilities could dramatically reduce facility tuning-time and downtime, improve facility performance, and maximize the productivity of the SUFs.

## 3. Facilities & Experiments

*Enable offline design and optimization of facilities and experiments.*

Research to enable offline design and optimization of the facilities and experiments to achieve new scientific goals.

### 3.1. Guidance

*Guide in-silico experiments from conception to synthesis and measurements.*

Physically accurate virtual laboratory environments of experimental facilities (a lab in the computational cloud) will help in guiding in-silico experiments from conception to synthesis and measurements.

### 3.2. Capabilities & Strategies

*Enable the design of new facility capabilities and execution of optimal experimental strategies.*

Digital twins that faithfully mimic facilities, including shared workflows and continuous updates from real experiments, can enable the design of new facility capabilities and execution of optimal experimental strategies to drive physics knowledge acquisition at the SUFs.

**Stakeholder(s):**
**Digital Twins**

# 4. Data

*Use shared scientific data for machine learning driven discovery.*

Address how to catalyze scientific discovery by leveraging the wealth of diverse and complementary data recorded across the SUFs.

## 4.1. Sharing, Curation & Analysis

*Radically improve data sharing, curation, and analysis.*

A radical improvement in data sharing, curation, and analysis is needed to catalyze scientific discovery across all facilities.

## 4.2. Integration

*Integrate diverse scientific data resources.*

Through the application of new AI/ML platforms to integrate diverse scientific data resources, the SUFs could create extensive new datasets from heterogeneous experimental and simulated data, leading to new opportunities for scientific discovery.

## 4.3. Standards, Formats & Priorities

*Catalyze development of data standards, formats, and priorities.*

Coordinated development of workflows on a shared facility-based data repository could catalyze development of data standards, formats, and priorities across the SUFs.

## 4.4. AI/ML Methods

*Develop new AI/ML methods.*

As a byproduct, these curated datasets can serve as training sets for developing new AI/ML methods.

## 5. Accelerator Complexes

*Improve operations of large accelerator complexes.*

**Stakeholder(s)**

**Fermilab** :
*Work conducted on large HEP facilities, such as the Fermilab accelerator complex, are of particular interest, but work may be conducted on other accelerator facilities that have analogous properties.*

In addition to the above topics, specific interest areas for HEP are implementation of data science techniques such as AI/ML to improve operations of large accelerator complexes, including development of tools and techniques that enable guided optimization, semi-autonomous operations, de-noising and data mining, digital twinning (e.g., virtual laboratories), and failure anticipation... For additional information, go to "Basic Research Needs Workshop on Compact Accelerators for Security and Medicine" https://science.osti.gov/- /media/hep/pdf/Reports/2020/CASM_ WorkshopReport.pdf?la=en&hash=AEB0B318ED0436B1 C5FF4EE0FDD6DEB84C2F15B2, May 6-8, 2019. See sections 2.2 and 5.2.1-5.2.3.

### 5.1. Optimization

*Develop tools and techniques enabling guided optimization.*

### 5.2. Semi-Autonomy

*Develop tools and techniques enabling semi-autonomous operations.*

### 5.3. De-noising & Data Mining

*Develop tools and techniques enabling de-noising and data mining.*

### 5.4. Digital Twins

*Develop tools and techniques enabling digital twinning (e.g., virtual laboratories).*

**Stakeholder(s):**

**Virtual Laboratories**

### 5.5. Failure

*Develop tools and techniques enabling failure anticipation.*

## 6. Accelerator Systems & Detectors

*Support technical development at the intersections between real-time ML and the control and optimization of accelerator systems operation and detector design.*

In addition to the above topics, NP is interested in supporting technical development at the intersections between real-time ML and the control and optimization of accelerator systems operation and detector design using AI models.

### 6.1. Beams & Nuclear Physics

*Address impact beam availability and nuclear physics data collection.*

Specific AI/ML applications are sought in accelerator and detector operations that impact beam availability and nuclear physics data collection. Examples include: predicting and minimizing superconducting radio frequency (SRF) faults, SRF cavity instability detection using trained AI models, and development of data driven models to predict machine and detector behaviors for increased performance.

**Stakeholder(s):**

**NP Scientific User Facilities** :
*Work conducted at NP scientific user facilities are of particular interest. NP held a one-day roundtable on "Machine Learning and Artificial Intelligence (ML/AI) for NP Accelerator Facilities" on January 30, 2020, with a focus on discussing opportunities in AI/ML for improving efficiencies of accelerator operations of NP facilities. Additional information on the workshop and copies of presentations can be found at: https://science.osti.gov/np/Research#ac.*

#### 6.1.1. SRF Faults

*Predict and minimize superconducting radio frequency (SRF) faults.*

#### 6.1.2. SRF Cavities

*Address SRF cavity instability detection using trained AI models.*

#### 6.1.3. Machine & Detector behaviors

*Develop data driven models to predict machine and detector behaviors for increased performance.*

## 7. Capabilities

*Apply and advance state-of-the-art capabilities.*

**Stakeholder(s)**

**Advanced Scientific Computing Research (ASCR) Program**

Addressing the above challenges may require advances in, or applications of state-of-the-art capabilities in, for example, storage systems, I/O, data and metadata management, and advanced computing hardware. Proposed teams should include expertise to address these challenges, as needed. Opportunities for collaboration with the Advanced Scientific Computing Research (ASCR) program may be identified in order to address these challenges.

### 7.1. Storage Systems

*Address challenges, advances, and applications of storage systems.*

### 7.2. I/O

*Address challenges, advances, and applications of I/O.*

### 7.3. Data & Metadata

*Address challenges, advances, and applications of data and metadata management.*

### 7.4. Computing Hardware

*Address challenges, advances, and applications of advanced computing hardware.*



## Administrative Information

**Start Date:** 2020-03-09
**End Date:** 2020-05-01

**Publication Date:** **2020-03-12**
**Source:** https://science.osti.gov/-/media/grants/pdf/lab-announcements/2020/LAB_20-2261.pdf

### Submitter:

**Given Name:** Owen
**Surname:** Ambur
**Email:** Owen.Ambur@verizon.net
**Phone:**