Report of the Workshop on Advancing Fusion with Machine Learning April 30 – May 2, 2019

Prepared for Fusion Energy Sciences and Advanced Scientific Computing Research, Office of Science

U.S. DEPARTMENT OF ENERGY | Office of Science

The pursuit of fusion energy has led to extensive activities in both experimental and theoretical science. The central goal of all these activities has been to develop the knowledge necessary for the design of successful fusion power plants. Even today, following decades of research in many key areas including plasma physics and material science, much remains to be learned in order to achieve optimization of the tokamak or other paths to fusion energy. Data science methods from the fields of machine learning and artificial intelligence (ML/AI) offer opportunities for enabling or accelerating progress toward the realization of fusion energy by maximizing the amount and usefulness of information extracted from experimental and simulation output data.

From April 30 to May 2 of 2019, a workshop on "Advancing Fusion with Machine Learning Research Needs" was held in Gaithersburg, MD, with joint support from the Fusion Energy Science (FES) and Advanced Scientific Computing Research (ASCR) Programs. The workshop brought together ~ 60 experts in fields spanning fusion science, data science, statistical inference and mathematics, machine learning, and artificial intelligence, along with DOE program managers and technical experts, with the goal of identifying Priority Research Opportunities (PROs) for the application of ML/AI methods to accelerating progress on solving fusion problems. During the workshop, seven key PROs were identified. The description of each PRO includes relevant fusion problems, potentially useful ML/AI approaches, known gaps currently preventing the use of such approaches, and research guidelines to maximize the effective application of ML/AI methods to the PRO. The report containing the full description of the workshop outcomes is available at https://science.osti.gov/-/media/fes/pdf/workshop-reports/FES_ASCR_Machine_Learning_Report.pdf

Workshop on Advancing Fusion with Machine Learning Priority Research Opportunities (PROs)

Accelerating Science	Enabling Fusion Energy
PRO 1: Science Discovery with ML *Hypothesis Generation and Experimental Guidance*	**PRO 4: Control Augmentation with ML** *Diagnostics to Data, Dynamic Models for Control, Fusion Trajectory Design*
PRO 2: ML Boosted Diagnostics *ML Boosted Diagnostics, Physics Enhanced Data*	**PRO 5: Extreme data algorithms** *Extreme-scale Processing, In-situ Data Analysis*
PRO 3: Model Extraction and Reduction *Data-driven Models, Reduction of Complex Code Algorithms*	**PRO 6: Data-enhanced Prediction** *Prediction of Disruption Events and Effects, Plasma Phenomena and State Prediction*
PRO 7: Fusion Data ML Platform	

Priority Research Opportunities

PRO 1: Scientific Discovery with Machine Learning includes approaches to bridging gaps in theoretical understanding through the identification of missing effects using large datasets; accelerating hypothesis generation and testing; and optimizing experimental planning to help speed up progress in gaining new knowledge. These research activities may produce theory-data hybrid models with enhanced explanatory power and enable priority planning of experiments to maximize the effective use of constrained machine time.

PRO 2: Machine Learning-Boosted Diagnostics involves using ML methods to maximize the information extracted from measurements; enhancing the interpretability with data-driven models; systematically fusing multiple data sources; and generating synthetic diagnostics that enable the inference of quantities that are not directly measured.

PRO 3: Model Extraction and Reduction includes the construction of models of fusion systems and plasmas for purposes of both enhancing our understanding of complex processes and the acceleration of computational algorithms. The rationale for this is that data-driven models can help make high-order behaviors intelligible and expose and quantify key sources of uncertainty. In addition, effective model reduction can shorten computation times for multi-scale/multi-physics simulations.

PRO 4: Control Augmentation with Machine Learning is intended to identify those areas of plasma control research that may benefit significantly from augmentation through ML/AI methods. One example is that of control-level models, which can be improved through data-driven methods. Another is real-time data analysis algorithms designed for control, which may enable critical functions such as the evaluation of proximity to magnetohydrodynamics stability boundaries.

PRO 5: Extreme Data Algorithms includes methods for in-situ, in-memory analysis and the reduction of extreme scale simulation data, as well as methods for efficient ingestion and analysis of extreme-scale fusion experimental data into the new Fusion Data ML Platform (see PRO 7). These capabilities are required to manage the amount and speed of data that will be generated by the many fusion codes that use first-principles models when they run on exascale computers.

PRO 6: Data-Enhanced Prediction will develop algorithms for the prediction of key plasma phenomena and reactor system states, thus enabling critical real-time and offline health monitoring and fault prediction. ML methods can significantly augment physics models with data-driven prediction algorithms to provide these essential functions. Disruption prediction represents an important requirement for a fusion reactor amenable to ML/AI.

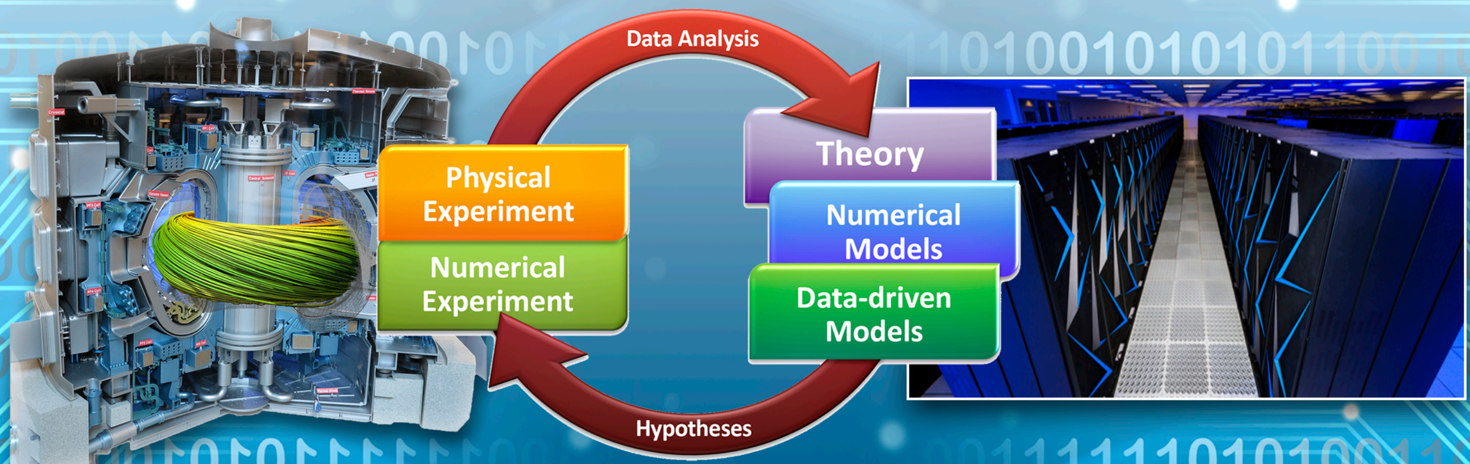
PRO 7: Fusion Data Machine Learning Platform constitutes a unique cross-cutting collection of research and implementation activities aimed at developing specialized computational resources that will support scalable application of ML/AI methods to fusion problems. The Fusion Data Machine Learning Platform is envisioned as a novel system for managing, formatting, curating, and enabling access to fusion experimental and simulation data for optimal usability in applying ML algorithms.

Summary and Foundational Activities

The set of Priority Research Opportunities identified in the Advancing Fusion with Machine Learning Research Needs Workshop fall into two broad areas: Accelerating Science, and Enabling Fusion Energy. The six PRO's in these two areas all rely on the cross-cutting PRO #7 (Fusion Data Machine Learning Platform), which includes both research and deployment of computational infrastructure supporting application of ML/AI methods to fusion problems.

In addition to these seven PROs, a set of foundational activities and resources were identified as essential to the execution of effective ML/AI research that would address fusion problems. These foundational activities and resources include experimental fusion facilities and ongoing research; advances in theoretical fusion science and computational modeling; high performance and exascale computing resources; established and supported connections among university, industry, and government expert groups in the relevant fields; and the establishment of connections to ITER and other international fusion programs.

The set of high-impact PROs identified in the workshop, together with the foundational activities highlighted (including the extensive experimental and theoretical science research presently funded by the DOE Office of Science), will significantly accelerate and enhance research towards solving outstanding fusion problems, helping to maximize the rate of knowledge gain and progress toward a fusion power reactor.



U.S. DEPARTMENT OF ENERGY | Office of Science