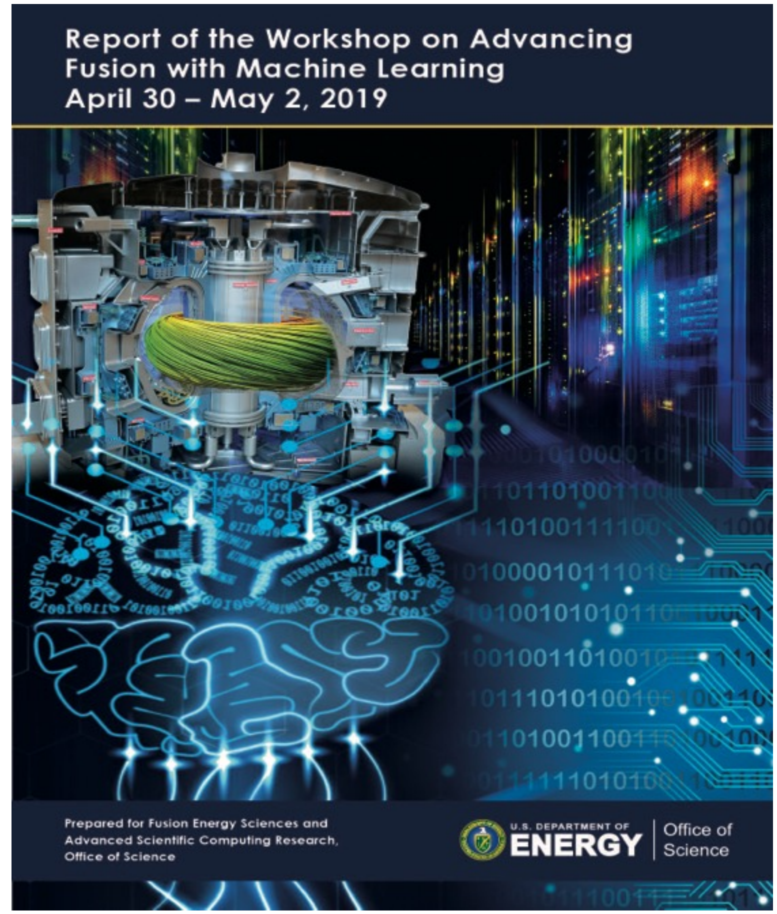
# Research Needs Workshop on Advancing Fusion with Machine Learning

**D. Humphreys (GA), A. Kupresanin (LLNL),**

**Workshop Co-chairs**




**FESAC Meeting**
**Gaithersburg, MD**
**24 August 2020**

# Outline

- Motivation and background for Research Needs Workshop
- Workshop structure, approach, and overview
- Priority Research Opportunities identified
- Foundational resources and activities
- Maximizing effectiveness of fusion machine learning research
- Mathematical solutions and PRO's
- Summary and conclusions

# Growth and Success of Machine Learning Has Motivated Focused DOE Assessments for Potential to Contribute to Areas Including Fusion Energy

- **Machine Learning and Artificial Intelligence:**
  - Rapidly growing fields
  - Extensive/successful commercial and research applications (but also lessons and caveats…)

- **Several DOE assessments and workshops held in the last ~2 years:**
  - FESAC Transformative Enabling Capabilities for Efficient Advance Toward Fusion Energy (2017)
  - Advanced Scientific Computing Research: Scientific Machine Learning (January 2018)
  - Fusion Energy Sciences/ASCR: Advancing Fusion with ML (May 2019)
  - DOE-wide: AI Town Halls (mid-to-late 2019)

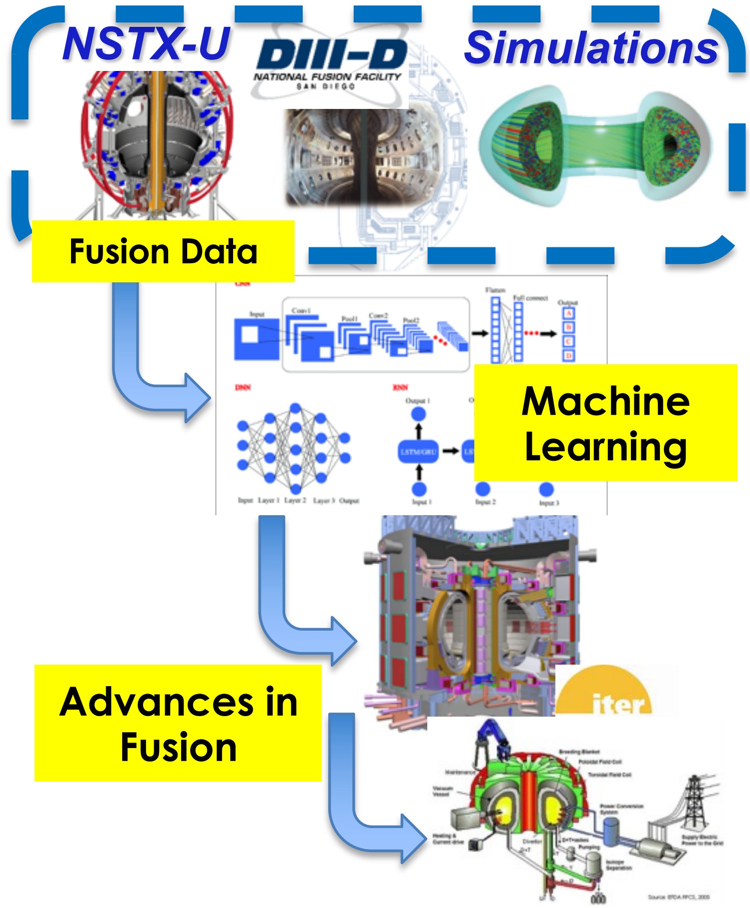

**Workshop & Report: Subject of this talk…**

## Research Needs Workshop on Advancing Fusion with Machine Learning
## Apr. 30 – May 2, 2019; Gaithersburg, MD

- **Goals: Identify Priority Research Opportunities including…**
  - Areas in fusion science where ML/AI can be transformative
  - Gaps in ML/AI for applicability to areas under FES mission
  - Research principles for effective use of ML for fusion

- **Joint FES/ASCR Workshop Team and Participation:**
  - Technical ~ 60;    DOE/Observers ~ 10 → **Total ~ 70**
  - FES/ASCR/Other Community Participants ~ 43%/38%/20%

- **Panels and Participation:**
  - Control 10
  - Models 12
  - Predictors 11
  - Science 12
  - Data 13



# Workshop Structure

- **Approach:**
  - Pre-workshop document defined structure/goals
  - Balance time equally between presentations (plenary) and discussions
  - Start workshop with context, end with identification of draft PRO's in each panel
  - Final output: initial text descriptions of draft PRO's

- **Agenda highlights:**
  - ~50% plenary, 50% breakout sessions: equal time for cross-panels discussion & panel focus
  - Day 1: All plenary…
  - Day 2: ~ All breakout (1 talk AM)
  - Day 3: 50% plenary, 50% breakouts

| Time | Session Title | Chair(s) | Rapporteur |
|---|---|---|---|
| 8:00 – 9:00 | Registration | N/A | N/A |
| 9:00 – 9:05 | General Announcements | N/A | N/A |
| 9:05 – 9:15 | Welcome: James van Dam | N/A | N/A |
| 9:15 – 10:00 | Workshop Intro: Humphreys/Kupresanin | Kupresanin/ Humphreys | N/A |
| 10:00 – 11:00 | Plenary 1: Data Assimilation and Machine Learning as Statistical Physics Problems Henry Abarbanel | Kupresanin | Humphreys |
| 11:00 – 11:45 | Science Panel: Canik/Hittinger | Kupresanin | Sabbagh |
| 11:45 – 13:00 | Lunch | | |
| 13:00 – 13:45 | Control Panel: Kolemen/Patra | Kupresanin | Wild |
| 13:45 – 14:30 | Models Panel: Boyer/Cyr | Humphreys | D'Elia |
| 14:30 – 15:15 | Predictors Panel: Granetz/Lawrence | Humphreys | Rea |
| 15:15 – 15:45 | Coffee Break | | |
| 15:45 – 16:30 | Data Panel: Schissel/Pascucci | Humphreys | Grierson |
| 16:30 – 17:30 | Wrapup/Summary Day 1; Brief for Day 2 | Humphreys | Kupresanin |

# Workshop Results

- **7 Priority Research Opportunities identified:**
  1. Science Discovery with ML
  2. ML-boosted Diagnostics
  3. Model Extraction and Reduction
  4. Control Augmentation with ML
  5. Extreme Data Algorithms
  6. Data-Enhanced Prediction
  7. Fusion ML Data Platform

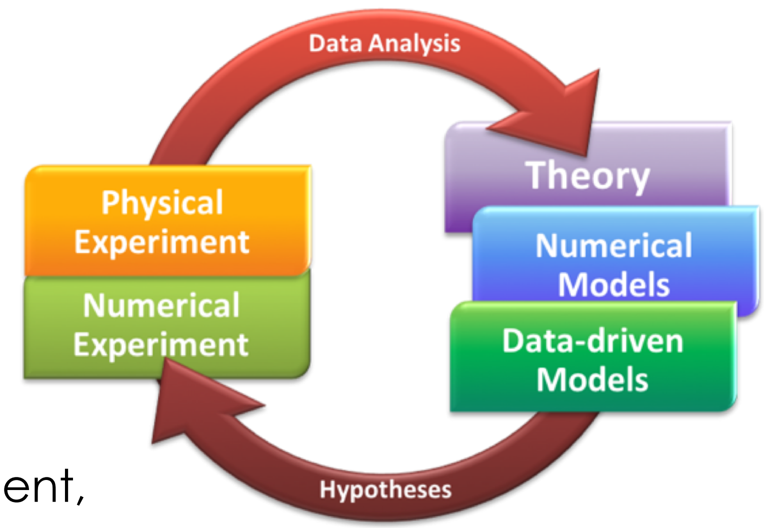
**Workshop on Advancing Fusion with Machine Learning**
**Priority Research Opportunities (PROs)**

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

- **Foundational Activities and Resources:**
  – Experimental fusion facilities/programs
  – Theory & HPC/exascale computing resources
  – Connections among domain experts (fusion science, computer science, statistical inference mathematics)

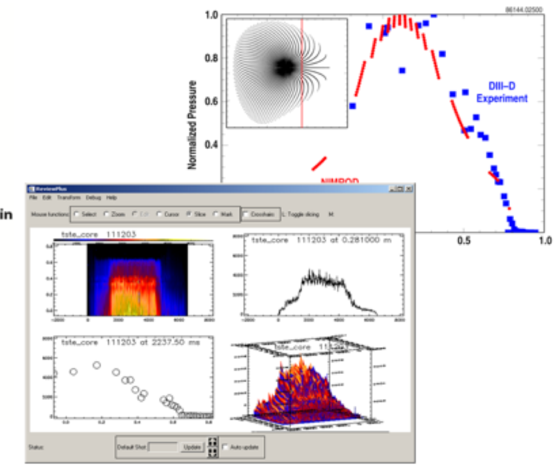
# PRO #1: Science Discovery with Machine Learning

- **Approaches to bridge gaps in theoretical understanding through identification of missing effects using large datasets**
- **Accelerating hypothesis generation and testing**
- **Optimizing experimental planning to help speed up progress in gaining new knowledge**
- **Example activities:**
  - Theory-data hybrid models for tokamak confinement, resistive MHD, PWI
  - Priority planning for tokamak experiments to maximize effective use of limited machine time



# PRO #2: Machine Learning-Boosted Diagnostics

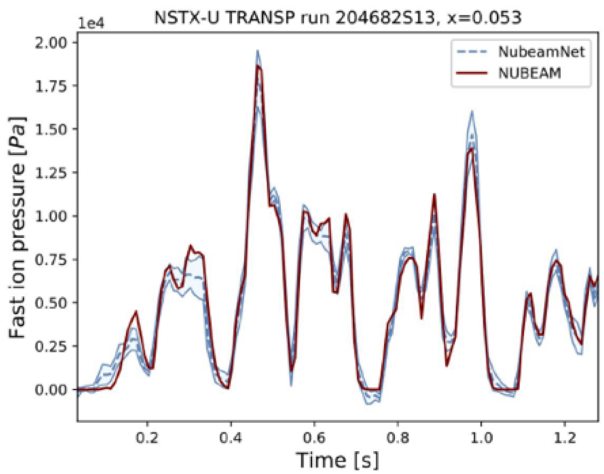
- **Application of ML methods to maximize the information extracted from measurements**
- **Enhanced interpretability with data-driven models**
- **Systematic fusion of multiple data sources**
- **Generation of synthetic diagnostics that enable the inference of quantities that are not directly measured**

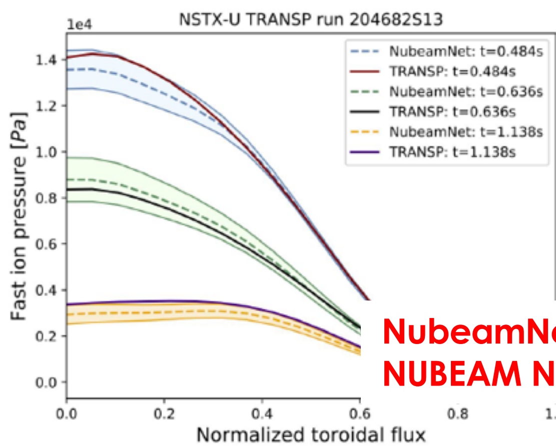




- **Example activities:**
  - Data fusion to infer detailed 3D MHD modal activity from many diverse diagnostics
  - Enhancing 3D equilibrium reconstruction fidelity
  - Extracting meaningful physics from extremely noisy signals

# PRO #3: Model Extraction and Reduction

- **Construct models of fusion systems and plasmas to enhance understanding of complex processes**
- **Accelerate computational algorithms with model reduction for multi-scale/multi-physics simulations**
- **Expose and quantify key sources of uncertainty**



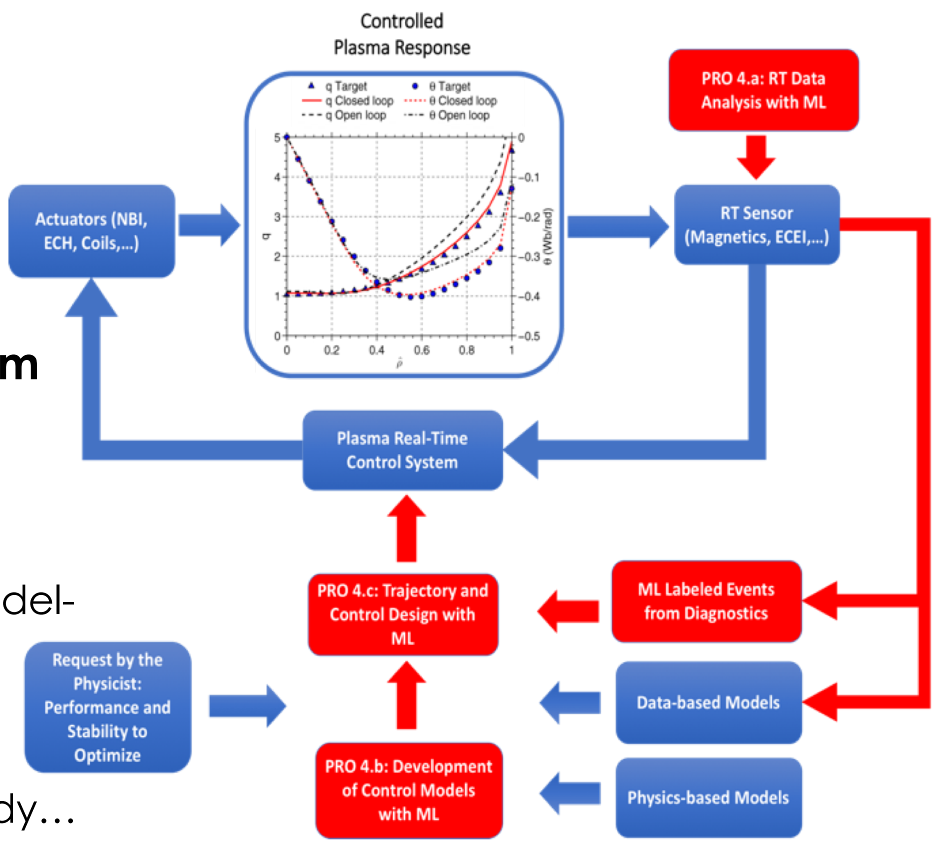




**NubeamNet: NN ~$10^6$ x faster than NUBEAM NBI calculation (Boyer NF 2019)**

- **Support hierarchies of fidelity in computer codes for whole device modeling**
- **Example activities:**
  - Enable faster than real-time execution of tokamak simulations
  - Improve understanding of empirical turbulent transport coefficients

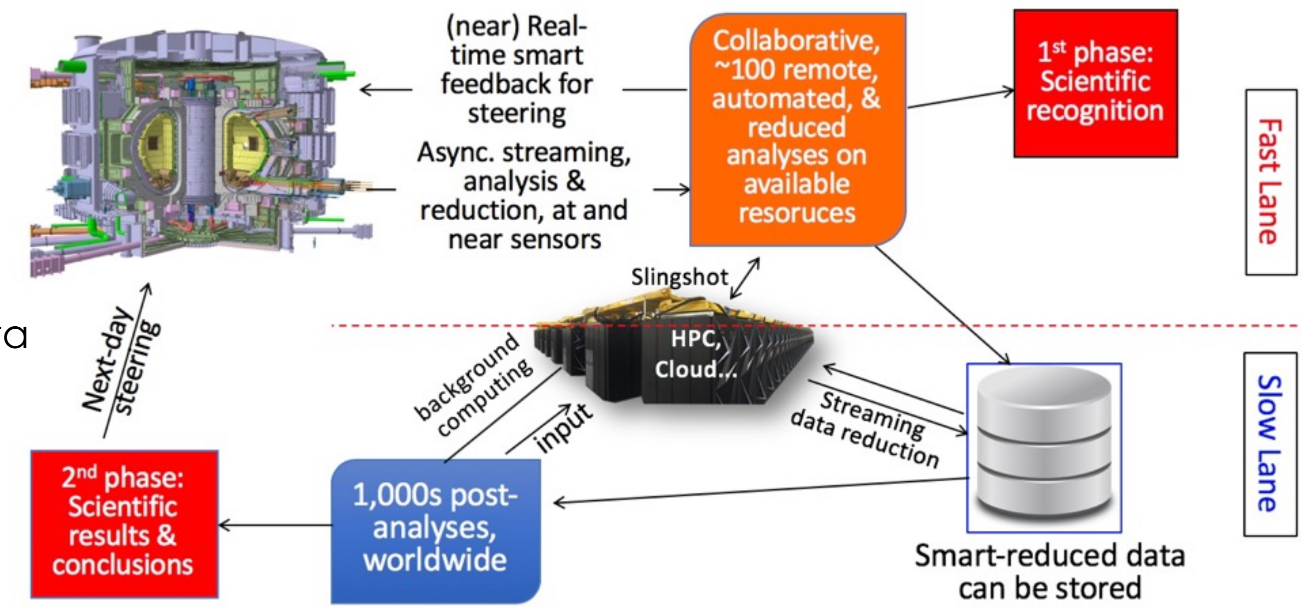
# PRO #4: Control Augmentation with Machine Learning

- **Control-level models for model-based design through data-driven methods**
-
- **Real-time signal analysis algorithms**
- **Optimize plasma discharge trajectories for control scenarios using algorithms derived from large databases**
- **Example activities:**
  - Manage and reduce uncertainty in real-time model-based control algorithms for tokamak operation
  - Enable complex analysis of real-time signals for control decision making and exception handling
  - CAUTION required for reinforcement learning study…



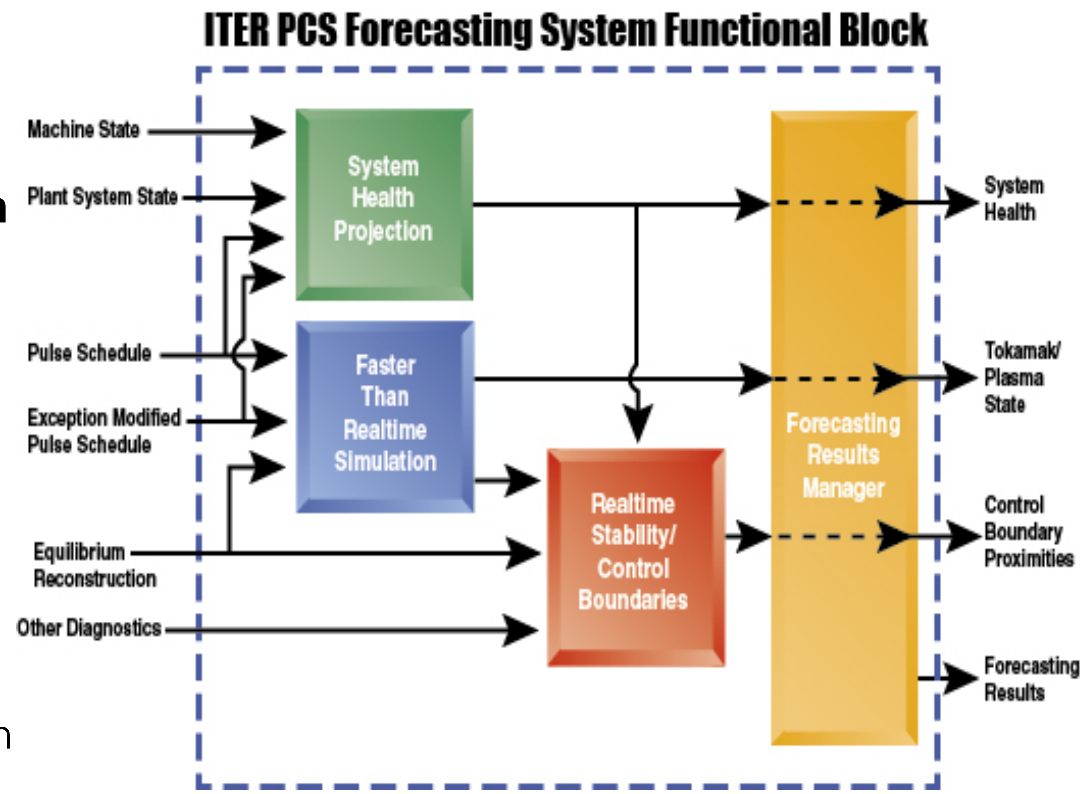
# PRO #5: Extreme Data Algorithms

- **In-situ, in-memory analysis and reduction of extreme scale simulation data**
- **Methods for efficient ingestion and analysis of extreme-scale fusion experimental data into the new Fusion Data ML Platform (see PRO 7)**
- **Manage amounts/speed of data generated by fusion codes on exascale computers**
- **Example activities:**
  - Preprocessing algorithms to increase throughput and efficiency of collaborative analysis and interpretation of long pulse tokamak pulse data as it is produced
  - Enhancing interpretability of data through optimized combination with simulation results



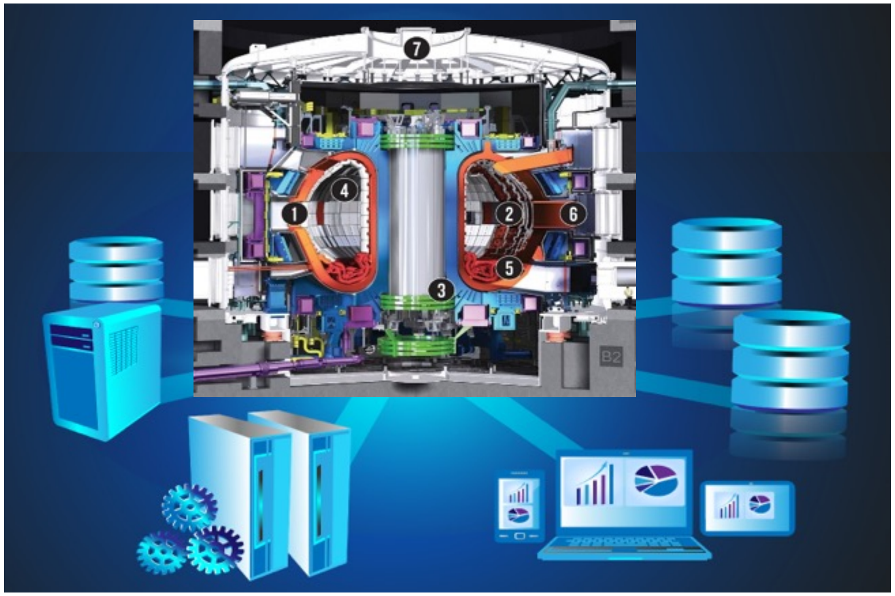
# PRO #6: Data-Enhanced Prediction

- **Algorithms for prediction of key plasma phenomena and plant system states**
- **Projection of complex fault and disruption effects for purposes of design and operational analysis**
- **Reliable disruption prediction with lead time sufficient for effective control action**
- **Example activities:**
  - Quantifiably effective tokamak disruption prevention system enabling control action
  - Tokamak fault prediction and handling satisfying power plant requirements



# PRO #7: Fusion Data Machine Learning Platform

- **Cross-cutting collection of research and implementation activities to develop specialized computational resources to support scalable application of ML/AI methods to fusion problems**

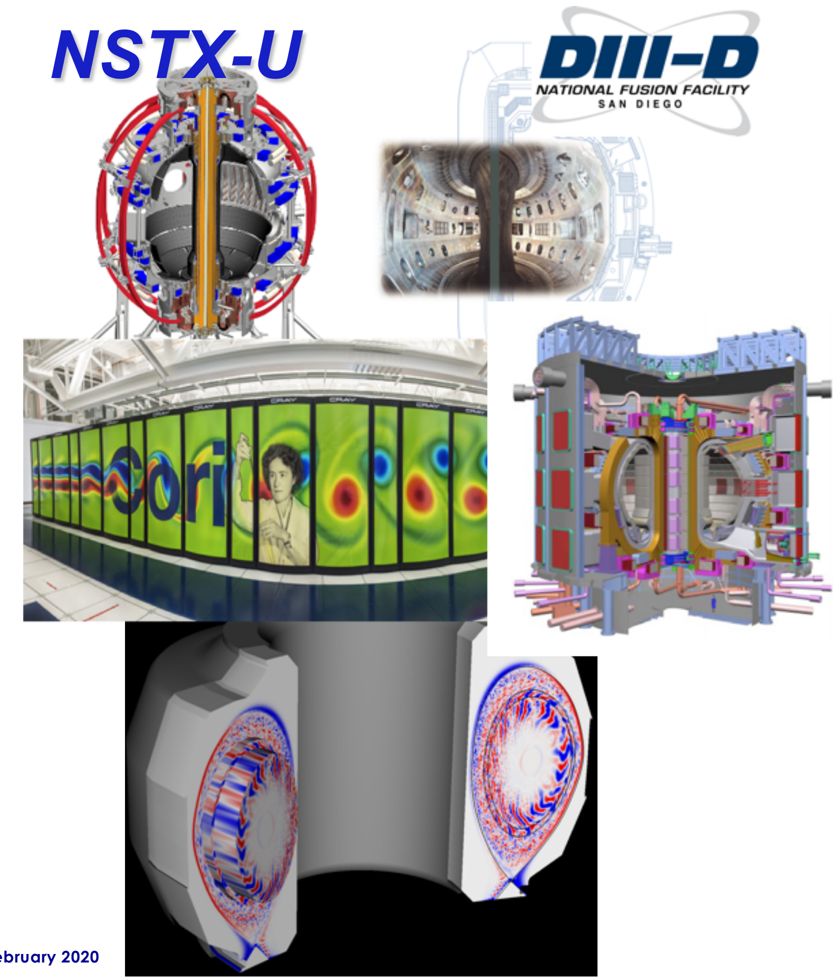


- **Novel system for managing, formatting, curating, and enabling access to fusion experimental and simulation data for optimal usability in applying ML algorithms**

- **Example activities:**
  - Algorithms to automatically populate Fusion Data ML Platform
  - Tools for production and storage of key metadata and labels
  - Methods for rapid selection and retrieval of data to create local training and test sets

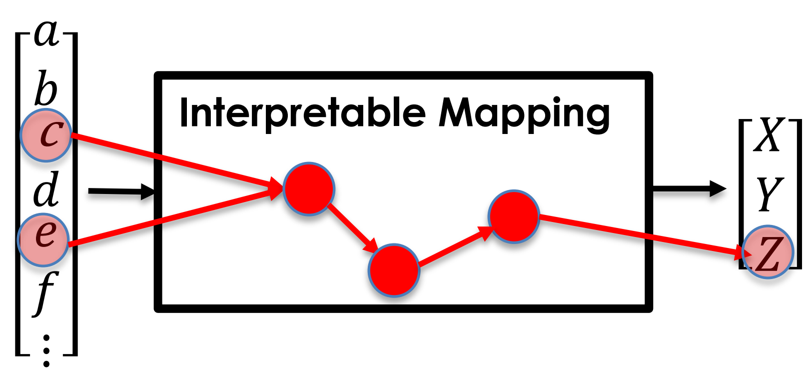
# Foundational Resources and Activities

- **Many resources and activities ongoing in DOE Office of Science were identified, essential to enable effective use of ML/AI to solve fusion problems**
- **Fusion experimental facilities and programs**
  - US & international devices, teams
  - Next-step devices: ITER, DTT, CFETR, STEP, …
- **Theory programs and computational resources:**
  - Theory teams, simulation/analysis codes
  - High Performance Computing/exascale
- **Robust connections among domain experts**



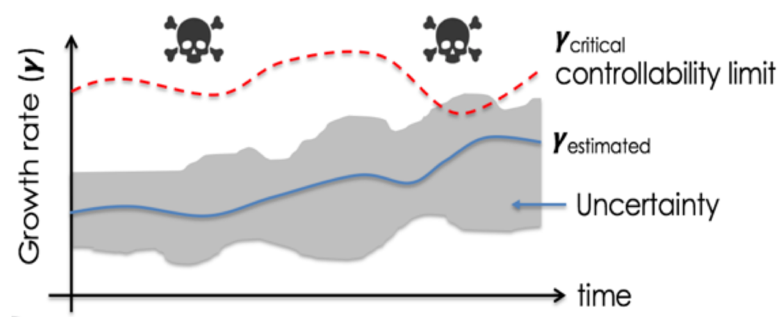
# Research Guidelines: ML/AI for Knowledge Extraction and Models

- **ML goals include:**
  - Extracting maximum knowledge from data
  - Bridging gaps in knowledge using data
  - Creating (compact) models from data




- **Interpretability is important to identify role/relevance of physical parameters**
  - Quantify new physical dependencies

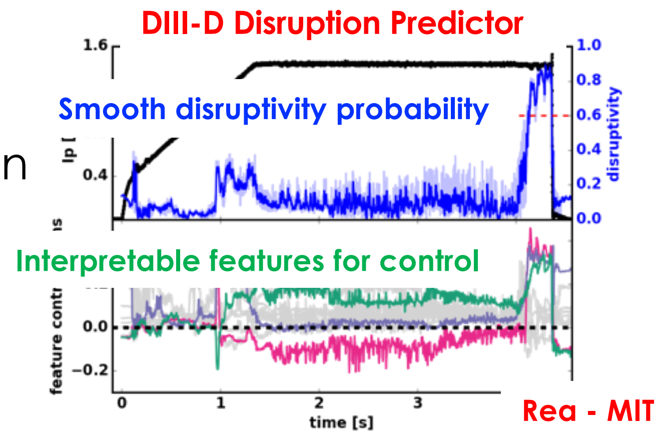
- **Uncertainty quantification is essential for ALL results**
  - Quantify accuracy/reliability of new knowledge
  - Quantify performance of data-driven theory




- **Validation measures must be identified in general**
  - Quantify space of validity and extrapolability

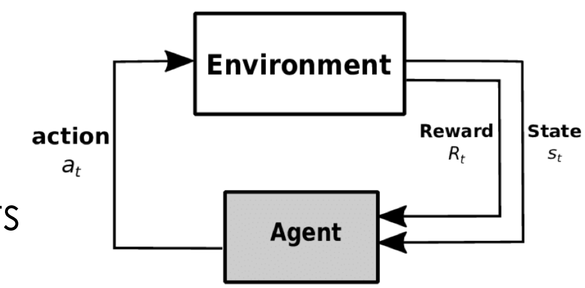
# Research Guidelines: Unique Requirements of ML/AI for Realtime Control

- **Requirements on Realtime Predictors for Control:**
  - Predict quantities for control of specific phenomena
  - Output parameters that enable realtime control action
  - Well-behaved, defined validity and extrapolability

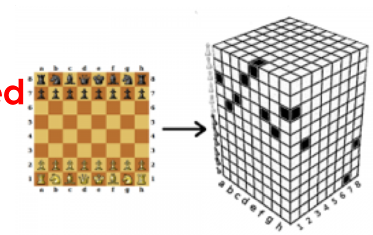



- **Models for Control Design:**
  - Quantifiable region of validity

- **Caveats on Reinforcement Learning for Control:**
  - Successful in domains that can be completely characterized (board games, video games, elements of autonomous vehicles…)
  - Valid and important area of ML research…
  - BUT not appropriate without detailed performance guarantees, UQ, validation… for controllers generated

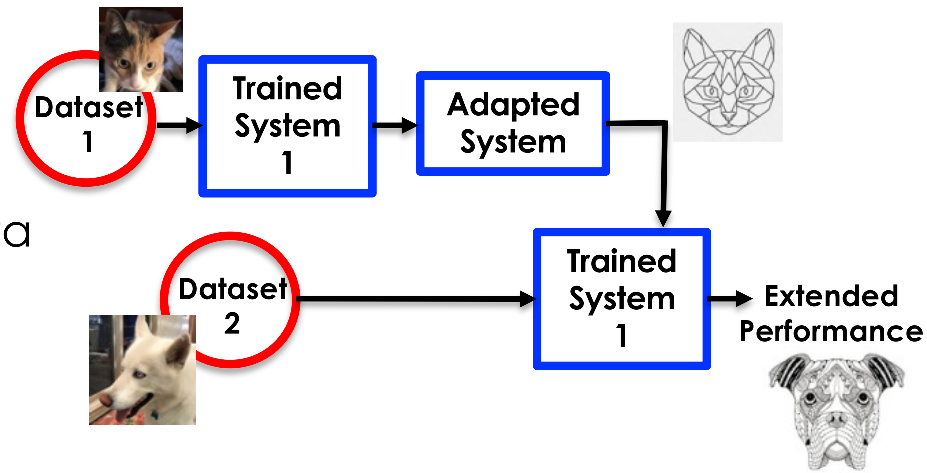



Controllers for Completely-Specified Domains



# Mathematical Approaches for PRO's 1-2: Scientific Discovery & Models

- **Transfer learning:**
  - Bridge gaps in existing theories with data
  - Hybrid models: theory + data-driven
  - Extend initial models with incremental data




- **Unsupervised learning:**
  - Discover trends and dependencies
  - Classification studies
- **Structured equation parameter identification:**
  - Analytic physics descriptions
- **Surrogate models:**
  - Encapsulate physical behavior with (compact) models

Data Constraints

$$\frac{dx_1}{dt} = A_{11}x_1 + A_{12}x_2 + C_{112}x_1x_2$$

$$\frac{dx_2}{dt} = A_{21}x_1 + A_{22}x_2 + C_{212}x_1x_2$$

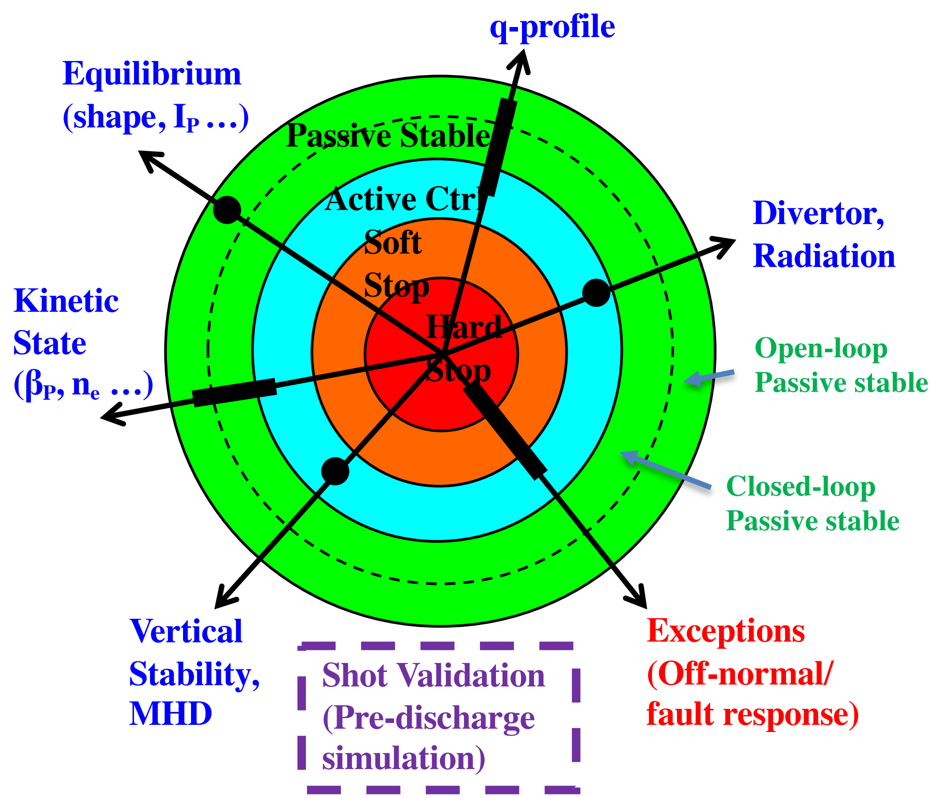
# Mathematical Approaches for PRO's 3-6: Control/Data/Predictors

- **Models for Control Design, Realtime use:**
  - Data-driven models
  - Surrogate models from simulations
  - Model reduction: speed, reduced accuracy

- **Performance quantification:**
  - Uncertainty quantification
  - Dynamic accuracy/precision
  - Regions of validity

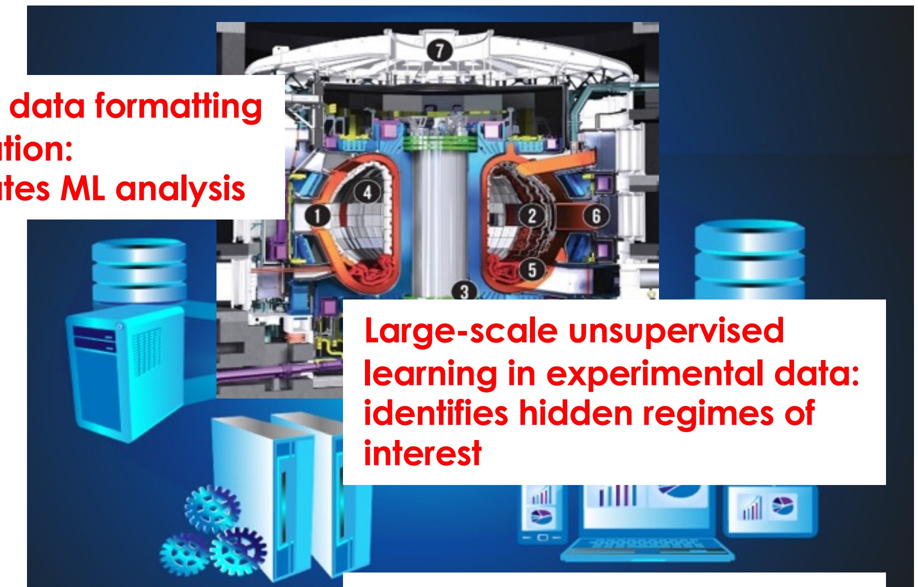
- **Math implications of predictor principles:**
  - Interpretability to identify actions
  - Classification for specificity of response
  - Realtime-capable computation

**Tokamak Control Operating Regime Map**



# Mathematical Approaches for PRO 7: Fusion Data Platform

- **Automated meta-data extraction:**
  - Create additional data labels to enable large-scale ML analysis
  - Federated availability of all data
- **Unsupervised/supervised learning:**
  - Identify data and regimes of interest in fusion experiments
  - Use meta-data to create better understanding of operational impacts
- **Tools/Workflows for Fusion Data Platform:**
  - Rapid data staging for ML applications
  - Curation for efficient access and use
  - ML-specific queries and visualization



# Summary and Conclusions

- **Growth and success of machine learning/AI have motived several DOE assessments of potential for advancing many areas including fusion energy science**

- **Joint FES/ASCR Research Needs Workshop, held on "Advancing Fusion Energy with Machine Learning", April 30 – May 2, 2019, identified:**
  - Areas in fusion science where ML/AI can be transformative
  - Gaps in ML/AI for applicability to areas under FES mission
  - Research principles for effective use of ML for fusion


- **Seven Priority Research Opportunities were identified:**
  1. Science Discovery with ML
  2. ML-boosted Diagnostics
  3. Model Extraction and Reduction
  4. Control Augmentation with ML
  5. Extreme Data Algorithms
  6. Data-Enhanced Prediction
  7. Fusion ML Data Platform