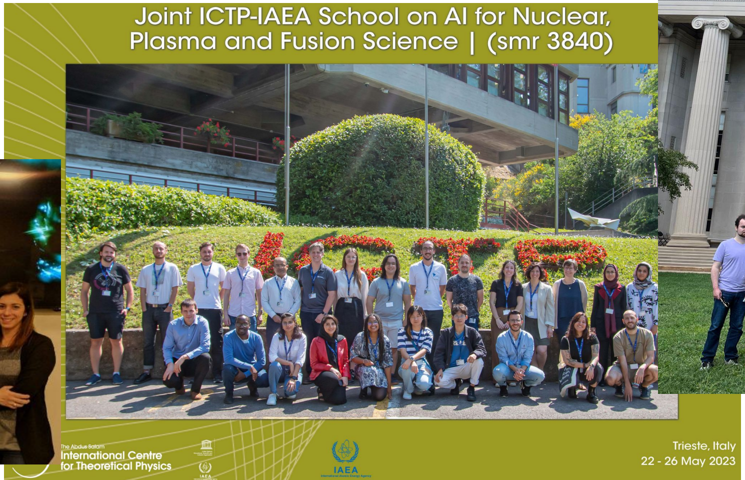
A few words about me: disruptions, Machine Learning, real-time, mentoring, teaching & education

(2014) PhD in Physics, University of Padova, Italy
(2015) Data Scientist, UniCredit, Milan, Italy
(2016) Postdoctoral Associate, MIT PSFC
(2019) Research Scientist, MIT PSFC
(2023) Group Leader Disruption Studies, MIT PSFC



AI & ML in Fusion

Cristina Rea

crea@psfc.mit.edu

MIT Plasma Science and Fusion Center, Cambridge, MA, USA

Introduction to Fusion Energy and Plasma Physics Course, PPPL and remote, June 15th, 2023

PSFC

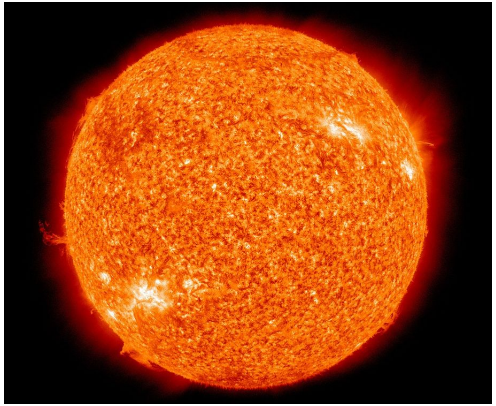
Outline

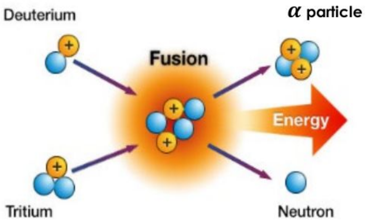
1. Fusion stuff and disruptions!
2. The Universality theorem and brief ML taxonomy
3. Explainable and adaptive ML models – applications in Fusion
4. Current challenges and opportunities for future research
5. Conclusions

Outline

1. Fusion stuff and disruptions!
2. The Universality theorem and brief ML taxonomy
3. Explainable and adaptive ML models – applications in Fusion
4. Current challenges and opportunities for future research
5. Conclusions

Fusion research is tackling transformational technologies to provide alternative, carbon-free electricity generation





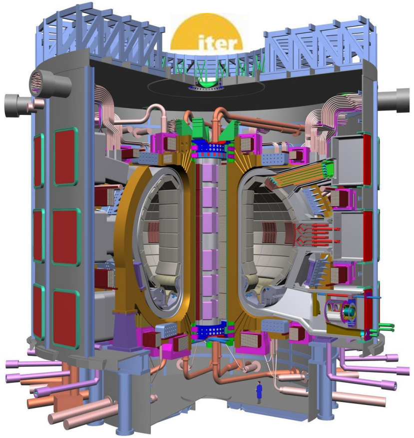
$$D+T \rightarrow \alpha(3.5\text{ MeV}) + n(14.1\text{ MeV})$$

Lab research conducted via:

- Heating and confinement of a plasma of hydrogen isotopes via magnetic fields → **magnetic confinement**

  Moser's and Paul's lecture

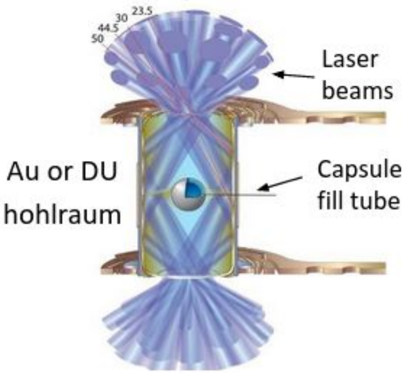
MFE



- Heating and compressing via lasers a fuel target of hydrogen isotopes → **inertial confinement**

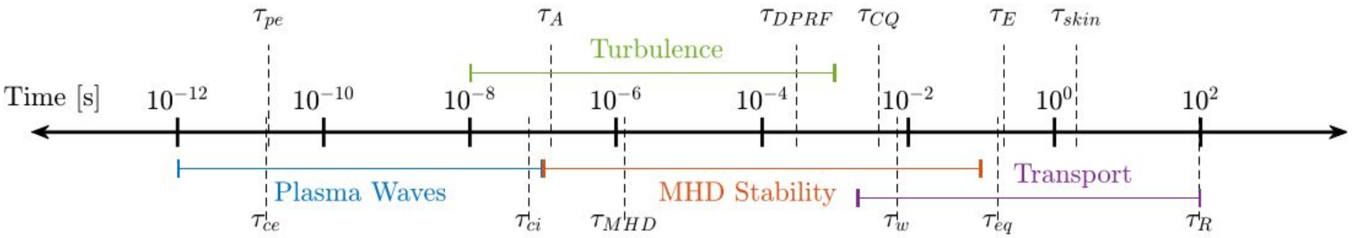
  Malko's and Kritcher's lecture

IFE



Let's take a closer look at **MFE** plasmas.

Fusion Plasmas: nonlinear phenomena really hot or really fast, hard to diagnose, lots theory & exp data, not so easy to bridge



← K. Montes (MIT), PhD Thesis, 2021

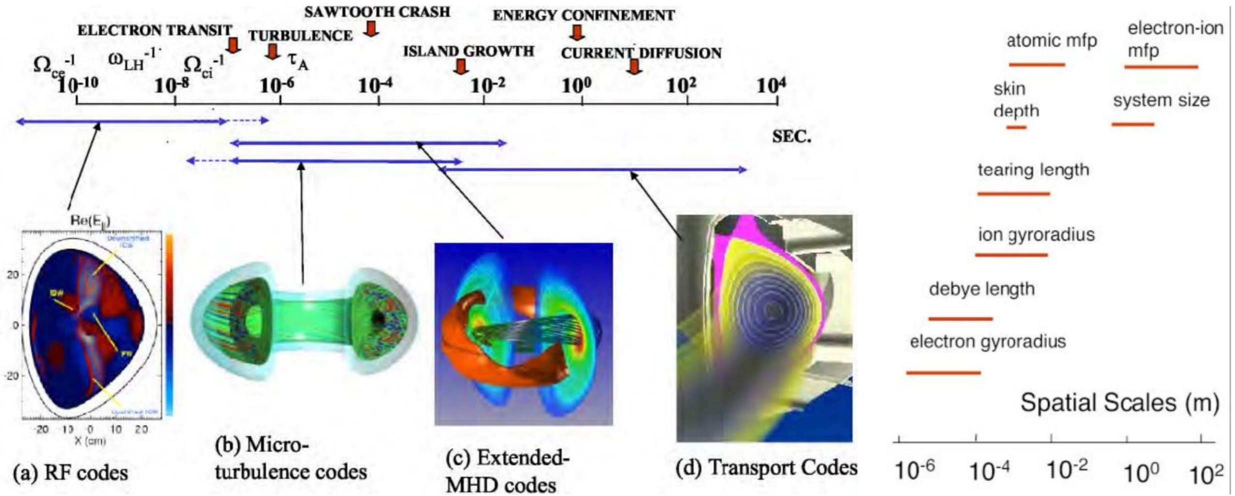
and

↓ L. Chacon 2022 ICTP-IAEA School

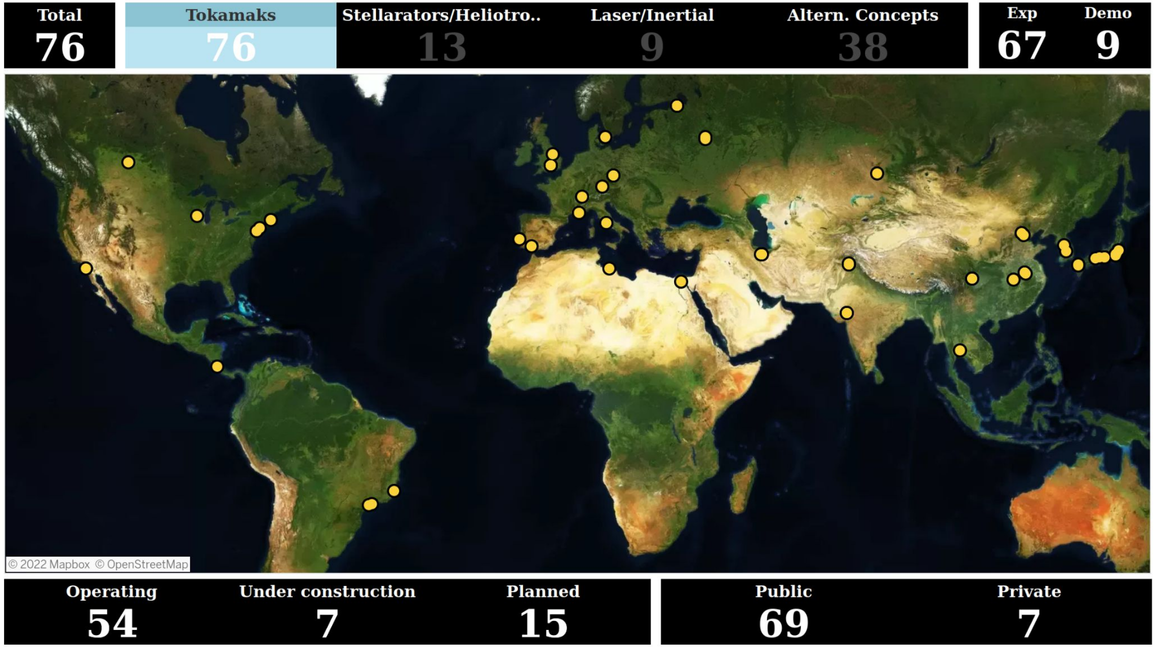
Challenges in thermonuclear fusion simulation: "The tyranny of scales"

Fusion plasma dynamics spanning wide range of spatial and temporal scales

Not so easy to develop first principle solutions!



Many operating experimental devices for magnetically confined fusion research, more planned!

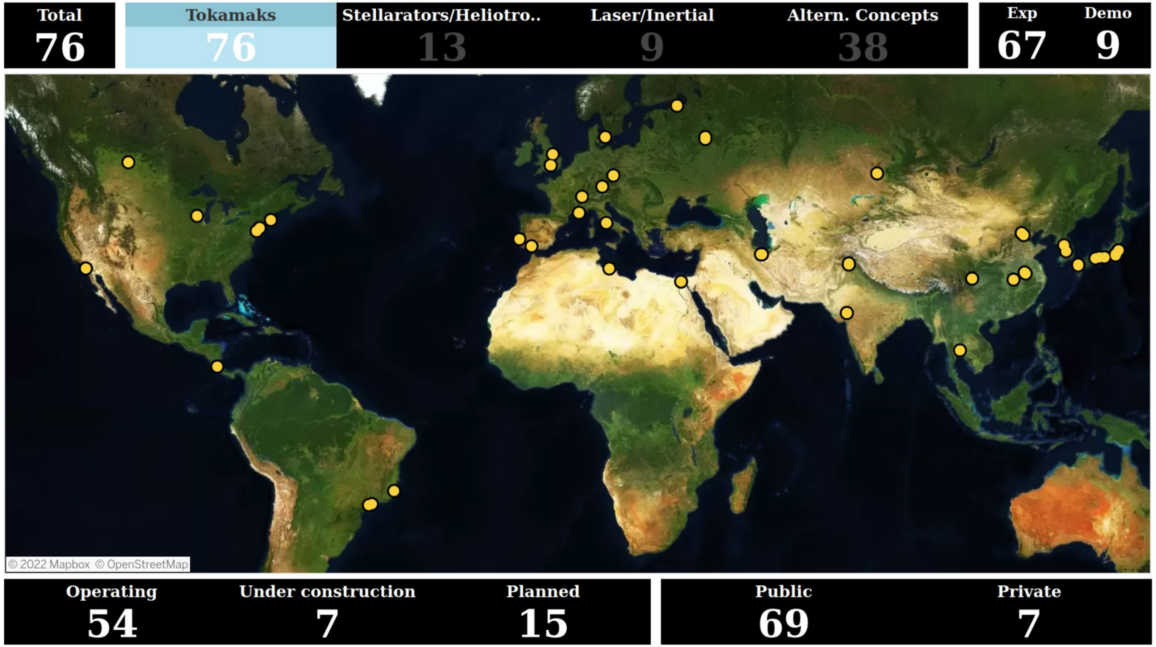


IAEA Fusion Device Information System
https://www.iter.org/of-interest/944

Huge amount of experimental and simulation data available enabling **Machine Learning applications:**

- ❑ optimization of experimental design
- ❑ real-time monitoring of proximity to instability
- ❑ trajectory planning optimization
- ❑ fast surrogates to accelerate simulations
- ❑ …

Many operating experimental devices for magnetically confined fusion research, more planned!

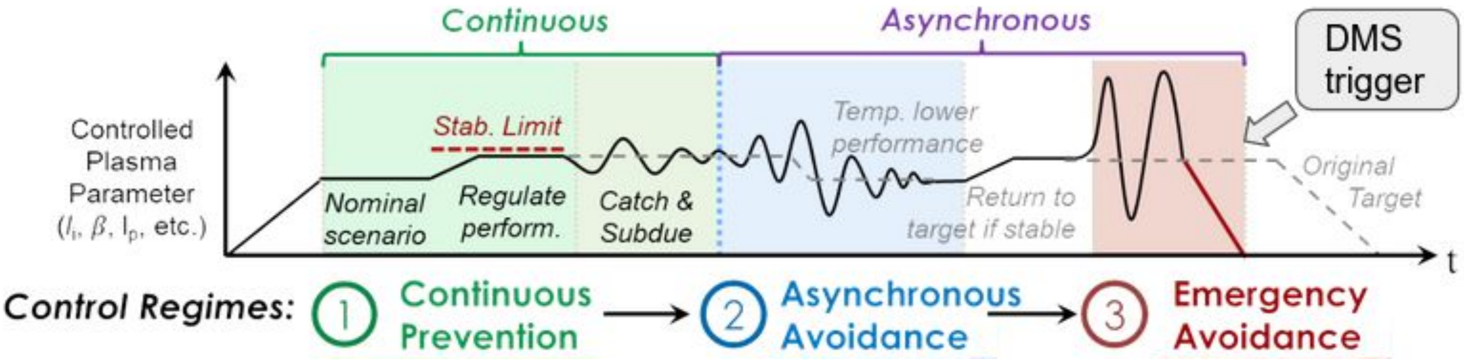


IAEA Fusion Device Information System
https://www.iter.org/of-interest/944

Huge amount of experimental and simulation data available enabling **Machine Learning applications:**

- ❑ optimization of experimental design
- ❑ real-time monitoring of proximity to instability
- ❑ trajectory planning optimization
- ❑ fast surrogates to accelerate simulations
- ❑ …

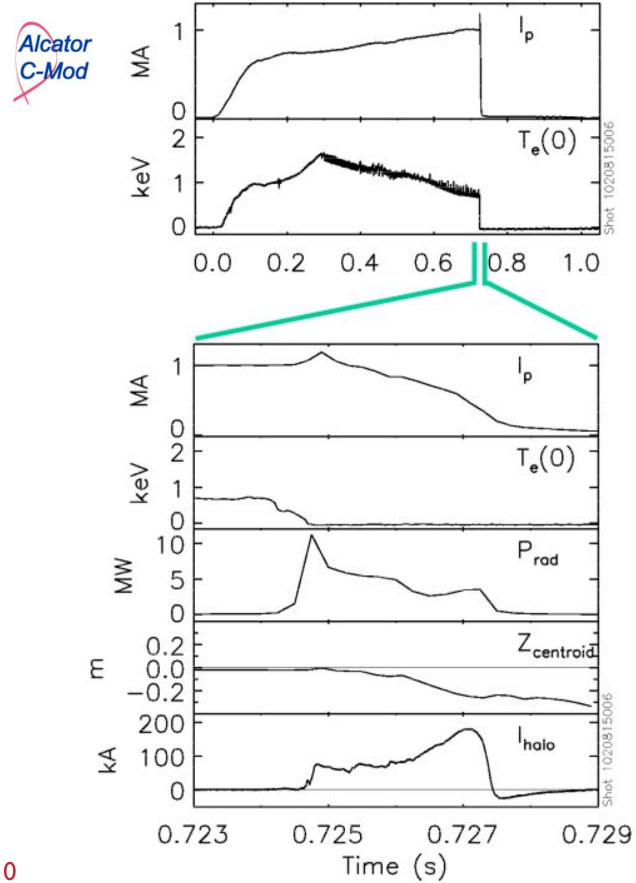
Active monitoring and prediction of soft/hard limits necessary to inform transition across operational boundaries



Adapted from J. Barr IAEA TM PDM 2020

Interpretable/explainable data-driven models provide general proximity to unstable operational space

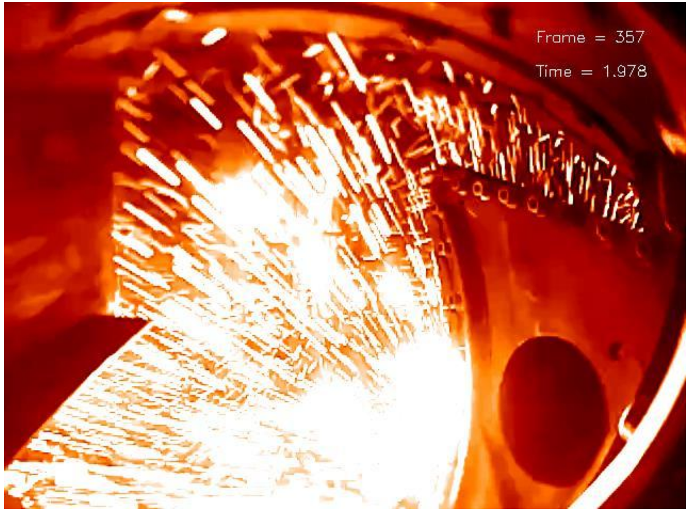
Plasma pushed close to operational limits often leads to instabilities onset or control faults: disruptions!



Major disruption → final **loss of control** evolving on **timescales of milliseconds**:

- Fast drop I_p leads to loss of confining poloidal field.
- Fast I_p transient causes large induced voltages, currents, forces.
- Rapid thermal losses cause surface damage.

Plasma pushed close to operational limits often leads to instabilities onset or control faults: disruptions!



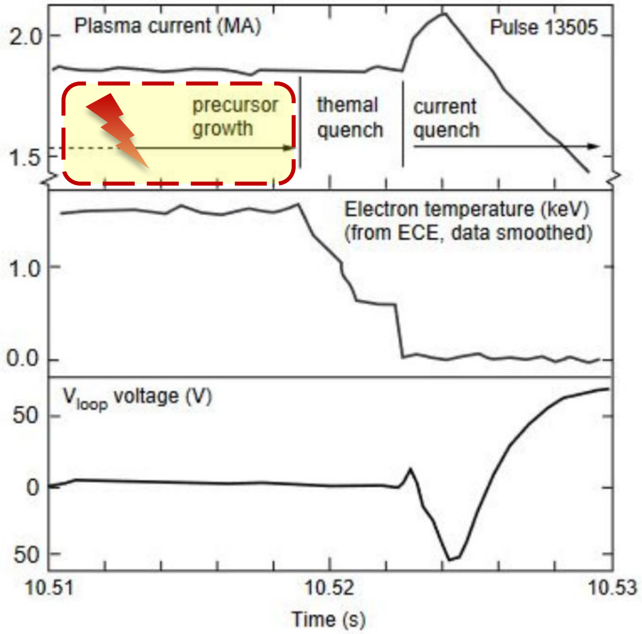
Visible camera view of RE beam hitting Alcator C-Mod first wall. Courtesy R.A. Tinguely

JET runaway electrons damage. https://www.iter.org/newsline/-/2234

How to take care of disruptions:

- **Accept** the damage and live with it.
- **Mitigate** the damage by injecting massive gas or shattered pellets.
- **Avoid** altogether by detecting <u>precursors</u> & steer plasma away from disruptive boundary.

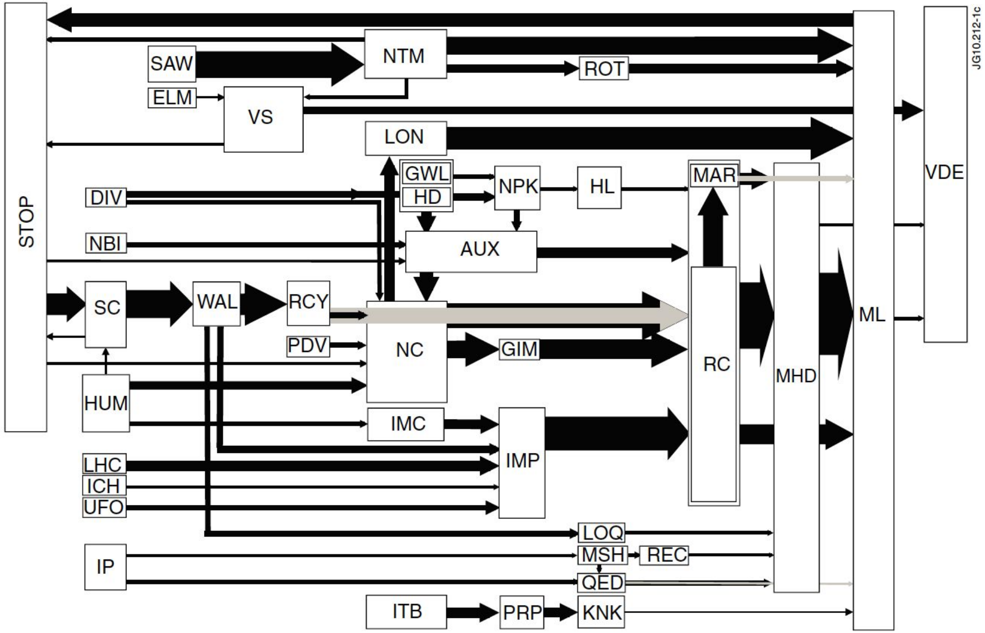
Precursor's growth prediction and detection extremely important…



ITER Physics Expert Group on Disruptions, Plasma Control, and MHD (1999) Nucl. Fusion 39 2251

…but not an easy task!
Statistical studies show complex chains of events:

possible disruptive chains of events



Statistics of the sequence of events for ~10yrs of unintentional disruptions at JET: width of the connecting arrows is the frequency of event occurrence.

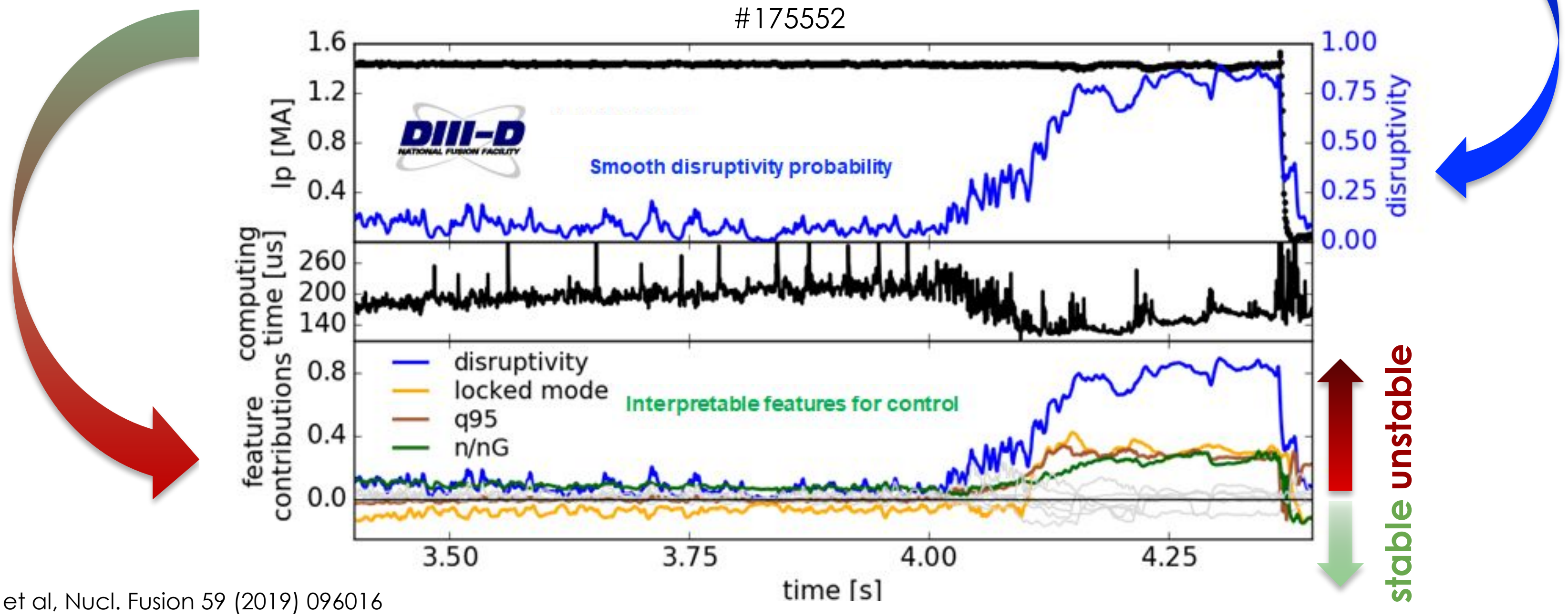
- Similar **statistical studies** not always available across different tokamaks.
- Need **timely identification of precursors** to allow the **plasma control system** (PCS) to take proper action.

Wealth of experimental data from different tokamaks enables Machine Learning applications.

De Vries et al. NF 51 (2011) 053018 "Survey of disruption causes at JET"

Explainable ML models for disruption prediction useful resources to identify stability boundaries in real-time

- DIII-D/EAST: the **D**isruption **P**rediction via **R**andom **F**orest algorithm (**DPRF**) applied to compute the **probability of an impending disruption**, while **interpreting its drivers in real-time**.



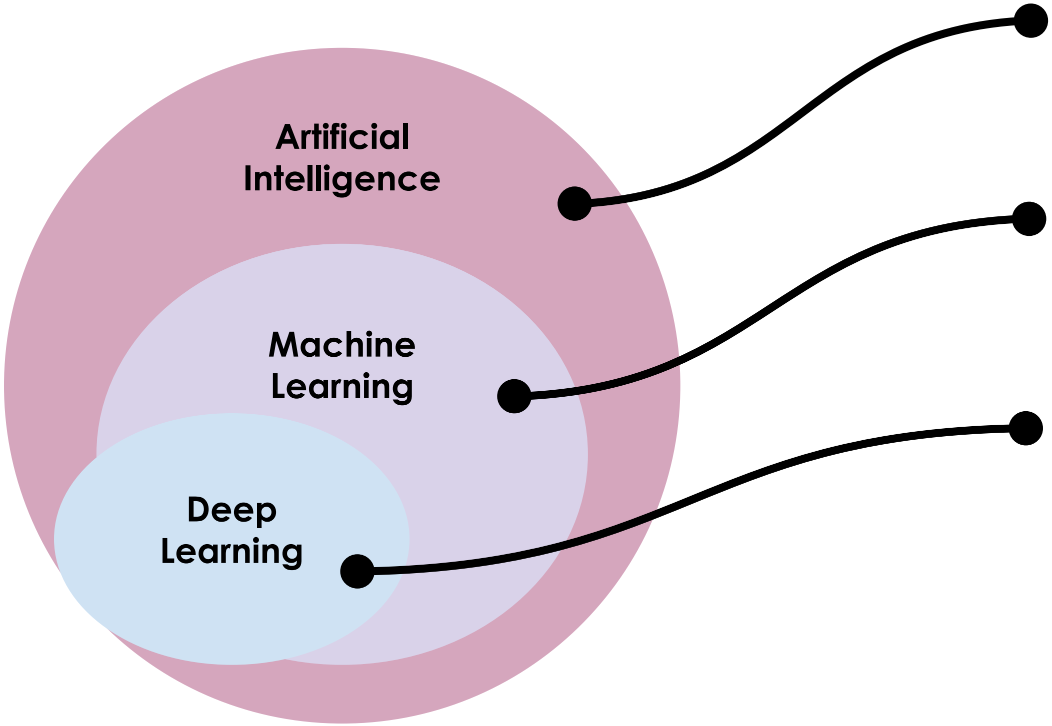
C. Rea et al, Nucl. Fusion 59 (2019) 096016
C. Rea et al, 2021 IAEA EX/P1–25,
J. Barr et al, Nucl. Fusion 61 (2021) 126019

Outline

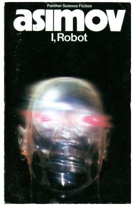
a series of useful concepts for ML practitioners

1. Fusion stuff and disruptions!
2. The Universality theorem and brief ML taxonomy
3. Explainable and adaptive ML models – applications in Fusion
4. Current challenges and opportunities for future research
5. Conclusions

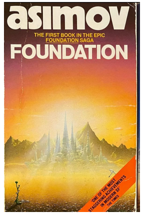
Sometimes there's confusion about terminology, too many buzzwords!



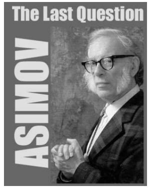
To **mimic human behavior** and functions such as **learning** and **problem solving**.



AI subset using **statistical methods** to enable **learn-from-experience** paradigm.

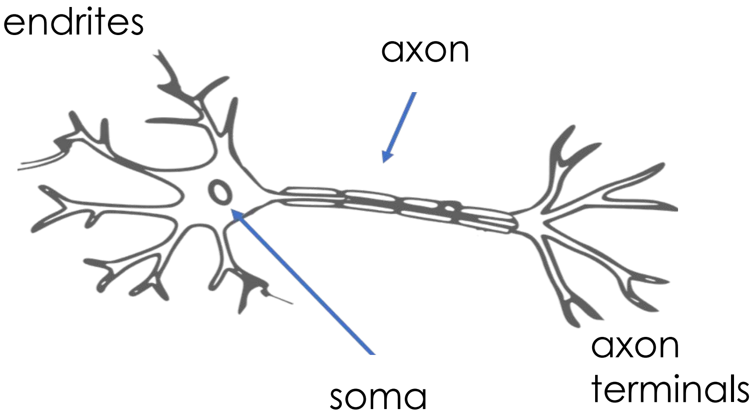


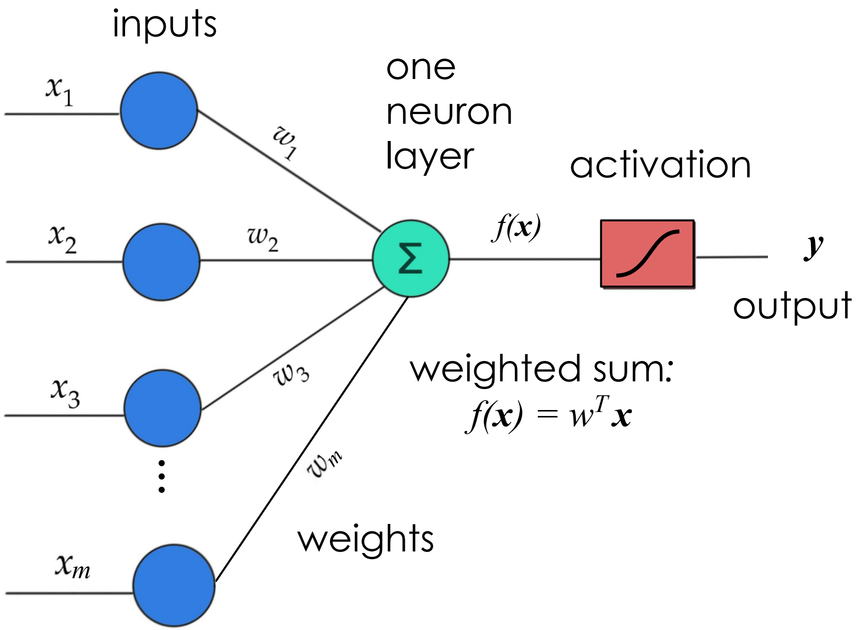
ML subset with broader **generalization** capabilities – **neural networks**.



Adapted from medium.com/@StepUpAnalytics

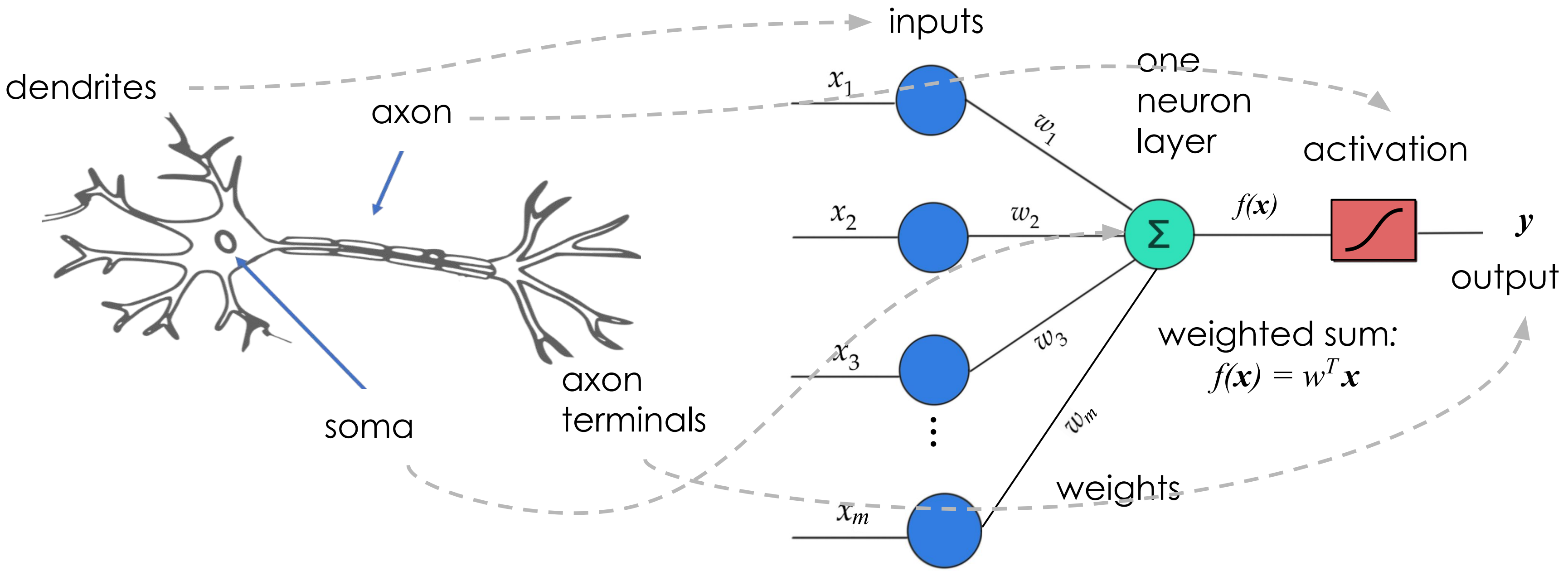
From biological to artificial neurons: the computational graph





Credits: M. Kuchera 💖

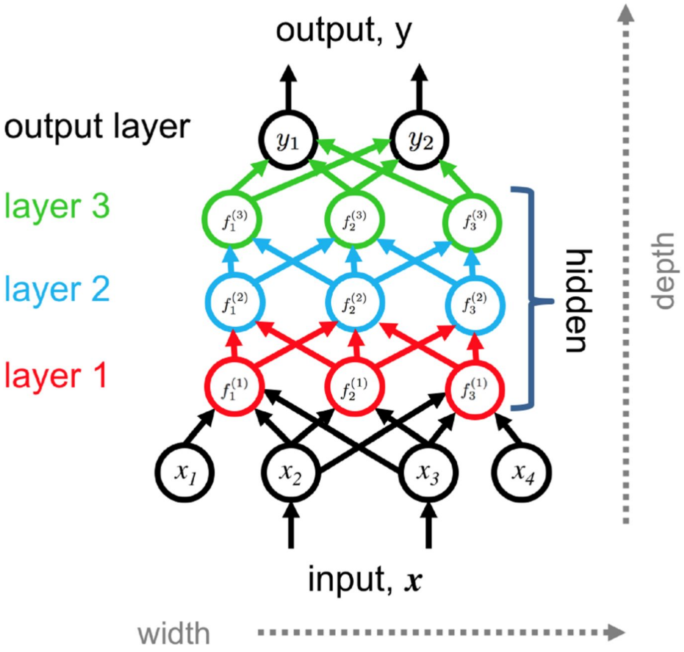
From biological to artificial neurons: the computational graph



Credits: M. Kuchera 💖

The Universality Theorem: for <u>any arbitrary</u> $f(x)$, there is <u>always</u> an artificial neural network that can <u>approximate</u> it

$$y \approx f(x) = f^{(4)}(f^{(3)}(f^{(2)}(f^{(1)}(x))))$$



Deep neural network example, adapted from B. Spears et al PoP 2018

Caveats:

- Increasing the depth can improve the <u>approximation</u>.

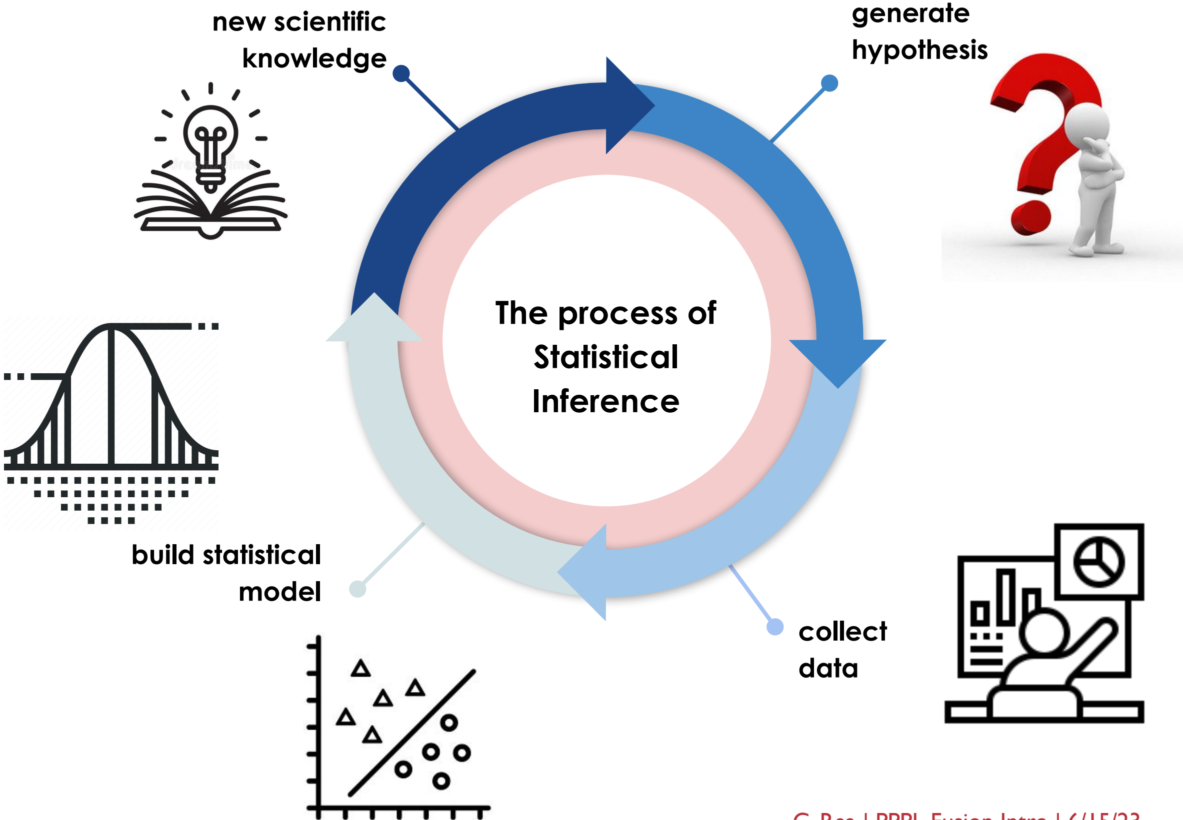
  $|y - f(\mathbf{x})| < \varepsilon$

- Activation must be <u>continuous</u>.

Neural networks provide *nonlinear* **mapping from inputs to outputs**, or a way to **represent your data** through *function approximation and estimation*.

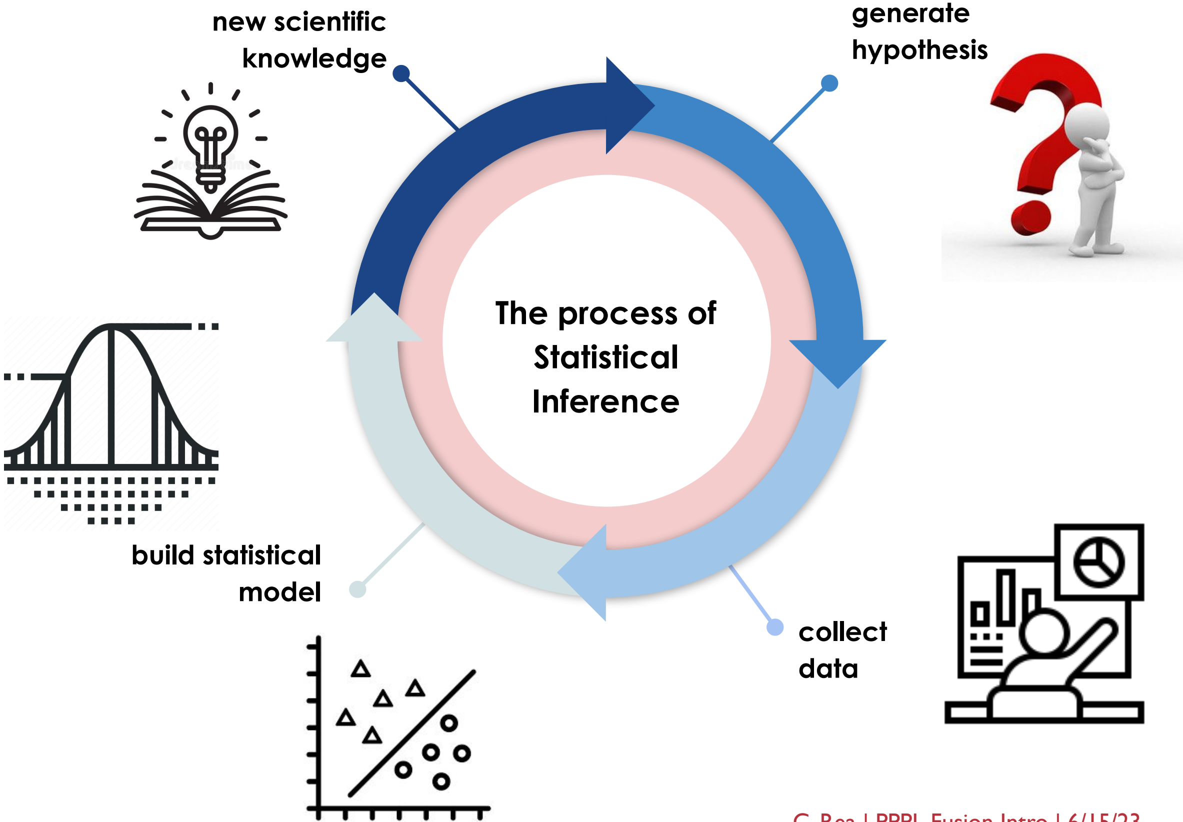
Statistical inference to learn representations from available data

Existing challenges in evaluating the mapping adherence to ground truth for ML models

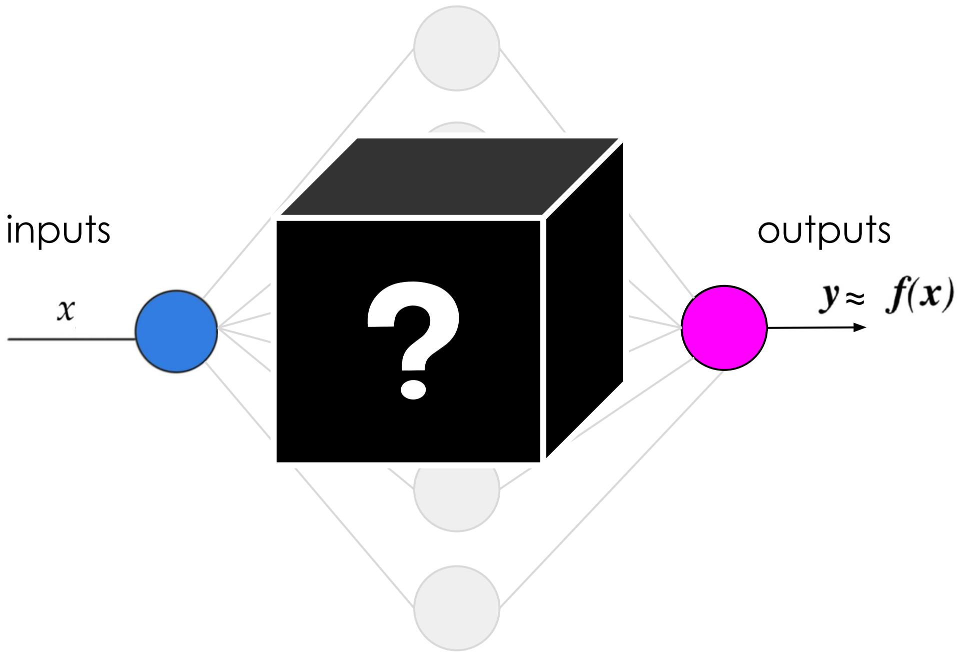


Statistical inference to learn representations from available data

https://xkcd.com/1838/



Introducing the black box: the issue with high-stakes decision making …



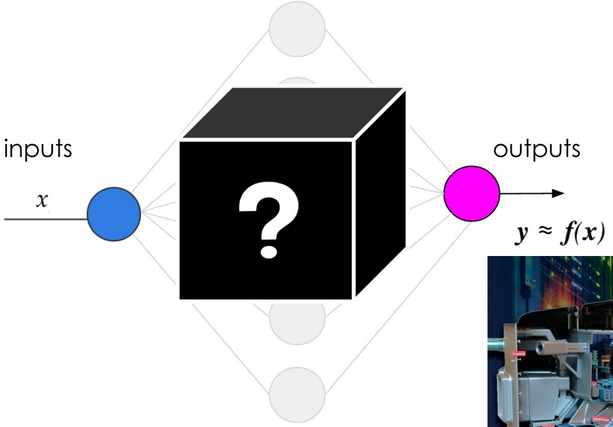
Black box as either
- function **too complicated** for human to comprehend or
- function that is **proprietary**

C. Rudin, Nat Mach Intell 1, 206–215 (2019)

Implications:
- lack of transparency and accountability,
- troubleshooting challenges.

High-stakes decision making:
- healthcare,
- criminal justice,
- child welfare screening,
- self-driving cars,
- …

High-stakes decision making and the parallelism with the fusion context



Fusion energy systems:
Any ML-based decision needs to be **trusted** and **justified**, or *licensed* → high-stake decisions!

Science discovery → Reconciliation with physical understanding, key ingredient to advance fusion research.

explainable predictions

vs

interpretable models

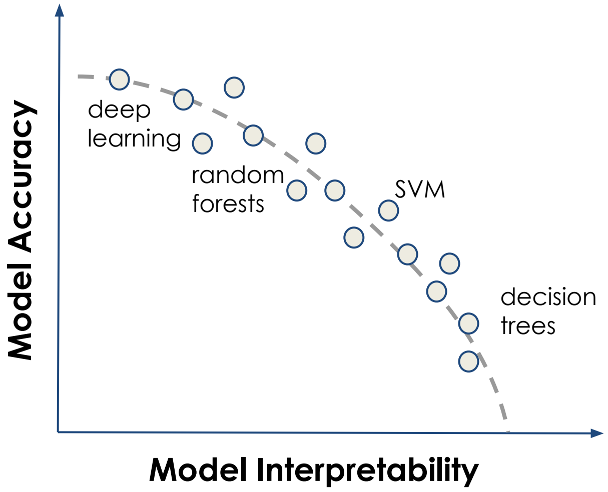
D. Humphreys et al, 2020 Advancing Fusion with Machine Learning Research Needs Workshop Report, J. Fusion Energy 39 123–55

Outline

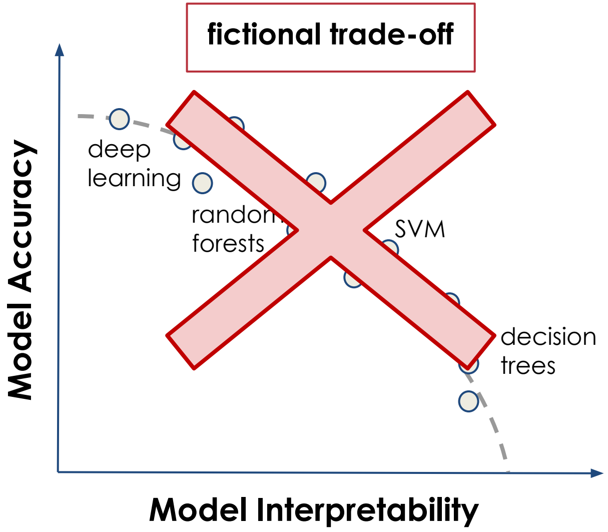
1. Fusion stuff and disruptions!
2. The Universality theorem and brief ML taxonomy
3. Explainable and adaptive ML models – applications in Fusion
4. Current challenges and opportunities for future research
5. Conclusions

https://xkcd.com/2541/

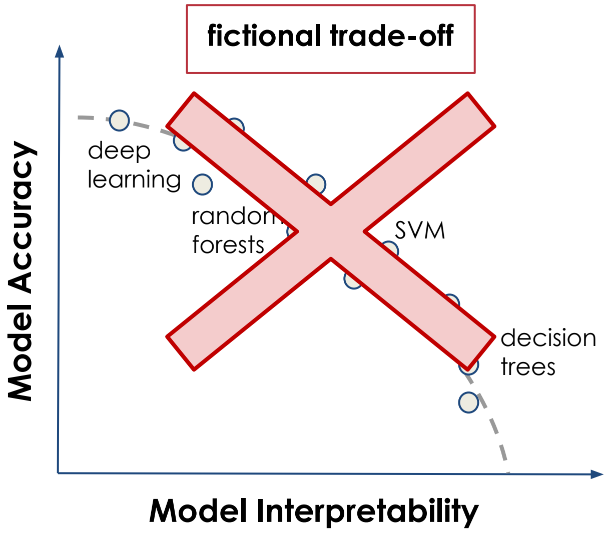
Common perception of accuracy vs interpretability trade-off



More interpretable and simpler models can be as accurate as black boxes



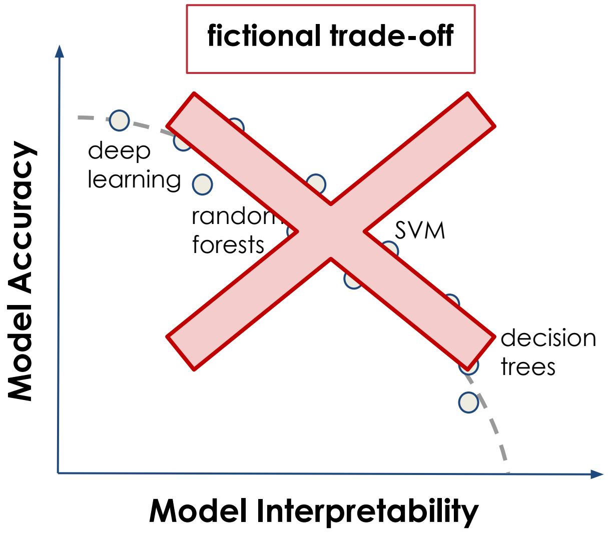
More interpretable and simpler models can be as accurate as black boxes



No unique "interpretability" definition:

- It's algorithm dependent – e.g., possibility to inspect reasons.
- It's domain dependent – e.g. sparsity not good for natural image classification.

More interpretable and simpler models can be as accurate as black boxes

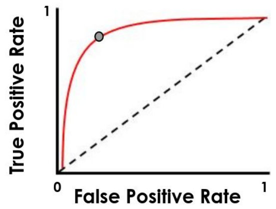


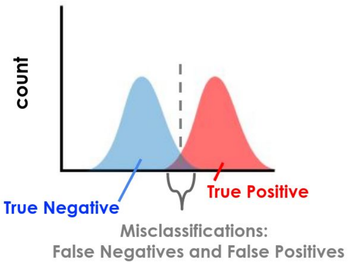
No unique "interpretability" definition:

- It's algorithm dependent – e.g., possibility to inspect reasons.
- It's domain dependent – e.g. sparsity not good for natural image classification.

What about accuracy definition?

- Typically well-defined – e.g., counting statistics of misclassifications, root mean squared error, …

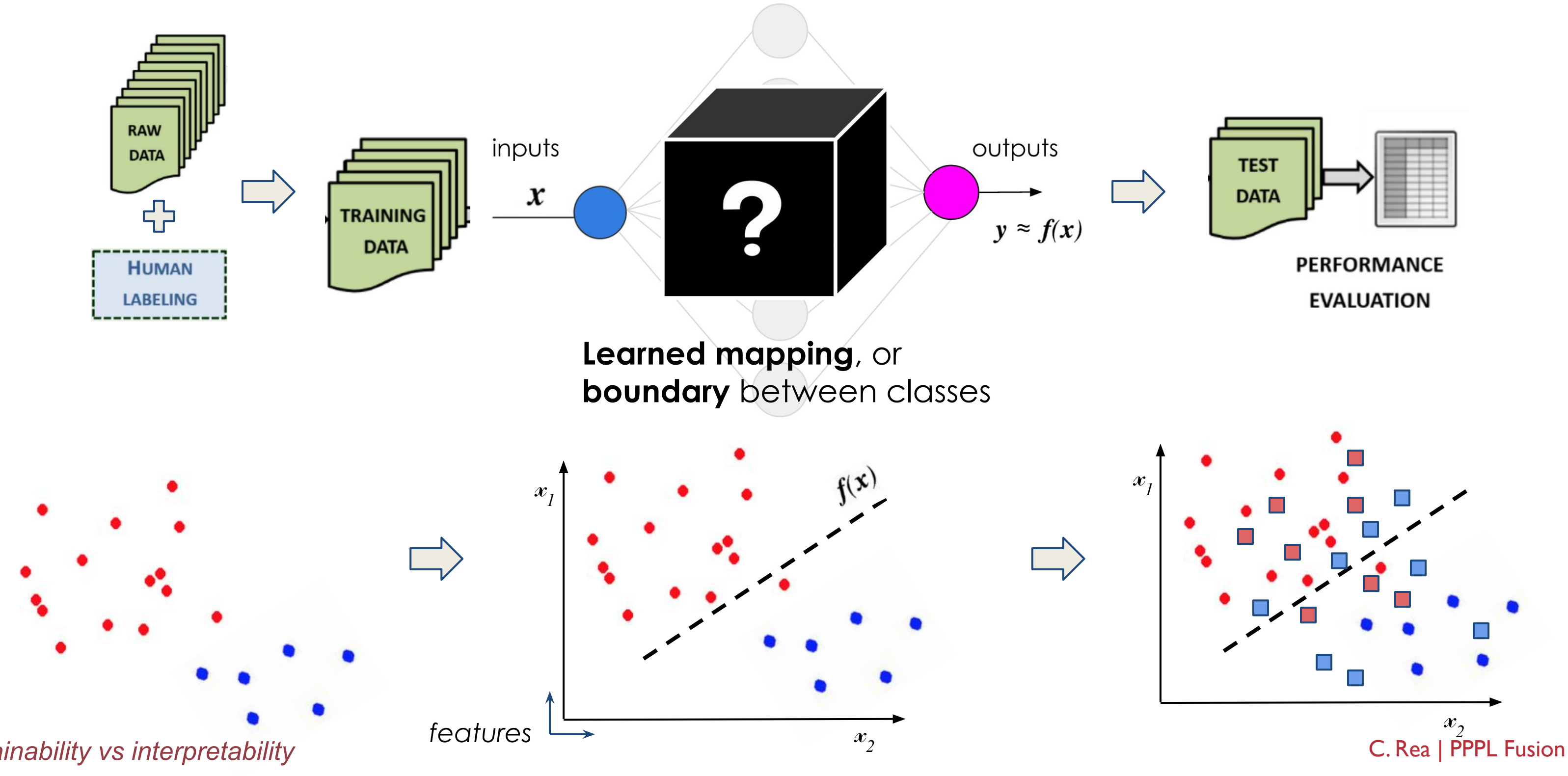




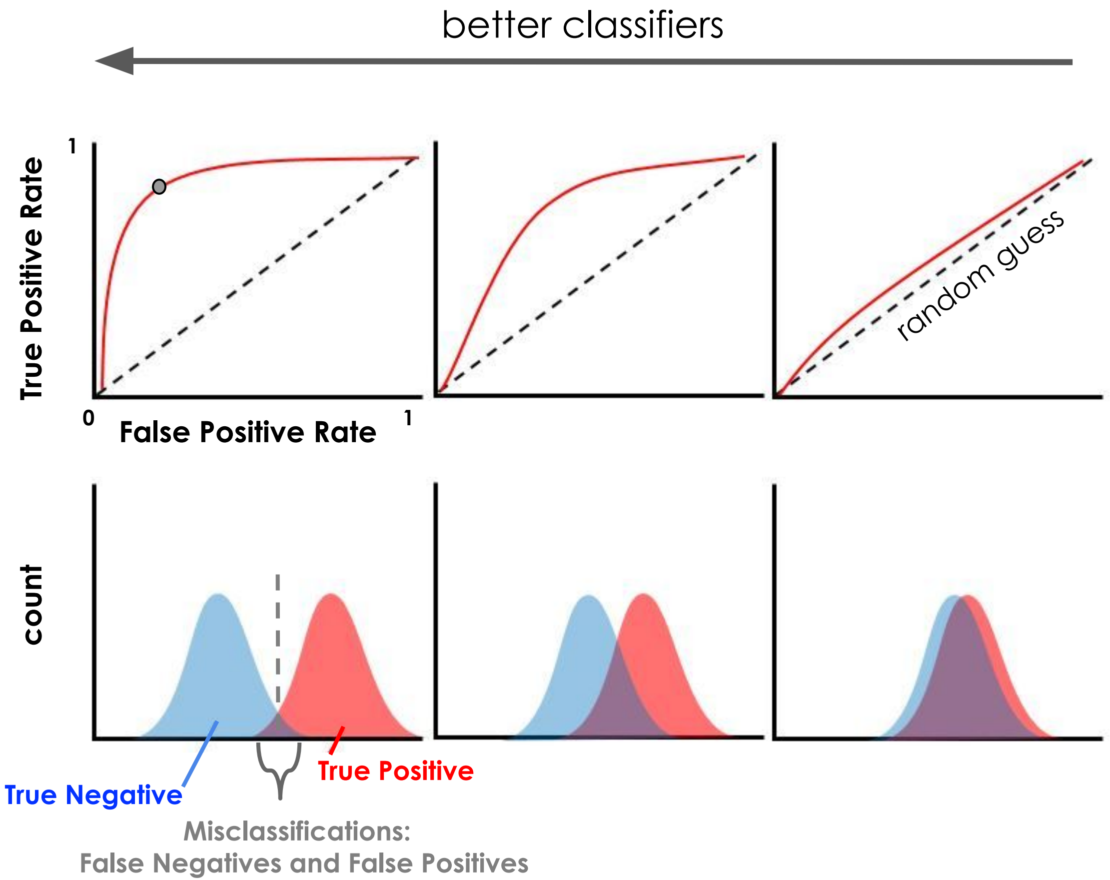
ML systems' prediction accuracy measured on new test data …

Simplified **supervised ML classification workflow**:
2D example (blue vs red)

Adapted from A. Pau et al,
Nuclear Fusion, 59(10):106017, 2019



… by counting how many times the trained classifier is right or wrong!



True Positive Rate:

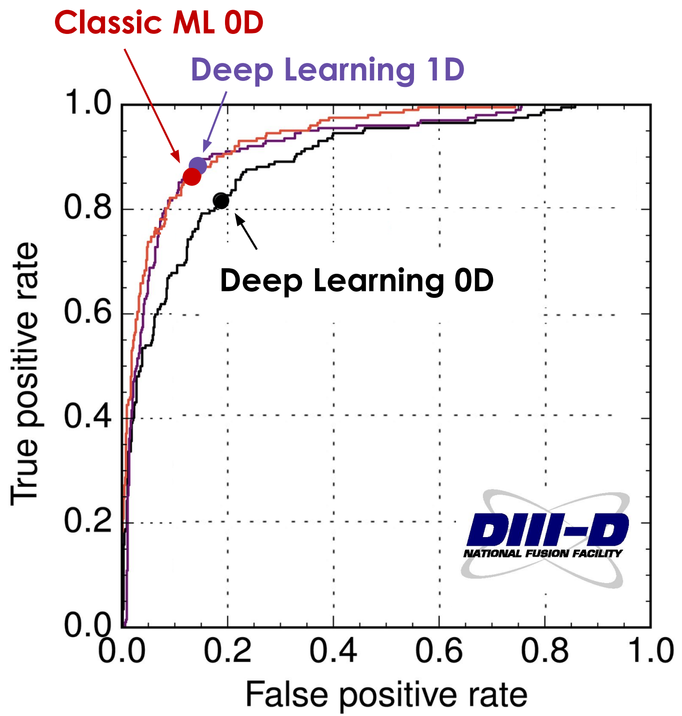
$$\frac{\text{\# correct positive classifications}}{\text{total \# of positive samples}}$$

False Positive Rate:

$$\frac{\text{\# wrong positive classifications}}{\text{total \# negative samples}}$$

ML models of varying complexity can have comparable performances – a disruption prediction example

Fusion



Adapted from J. Kates-Harbeck et al., Nature 568, 526–531 (2019)

- **Rashomon Effect**:
  a multitude of models with approximately the minimum error rate exists, for many problems[1] (also in Fusion!).

  As long as a large Rashomon set exists, it is likely that some are interpretable[2,3], *maybe hard to develop*.

[1]L. Breiman et al, 2001 Statistical Science 16 199–231
[2]C. Rudin et al., 2022 Stat. Surv. 16 1–85
[3]Semenova et al, 2022 ACM Conference on Fairness, Accountability, and Transparency (FAccT'22) arXiv:1908.01755

ML models of varying complexity with comparable performances

Fusion

https://en.wikipedia.org/wiki/Rashomon

<u>Fun fact</u>:
Rashomon term inspired by 1950 Kurosawa's movie!

- **Rashomon Effect**:
  a multitude of models with approximately the minimum error rate exists, for many problems[1] (also in Fusion!).

  As long as a large Rashomon set exists, it is likely that some are interpretable[2,3], *maybe hard to develop*.

[1] L. Breiman et al, 2001 Statistical Science 16 199–231
[2] C. Rudin et al., 2022 Stat. Surv. 16 1–85
[3] Semenova et al, 2022 ACM Conference on Fairness, Accountability, and Transparency (FAccT'22) arXiv:1908.01755

Models interpretable by design vs black boxes that can be “explained”

existing dichotomy

Interpretable by design vs **Black boxes**

- Dimensionality Reduction
- …

- Physics Informed Neural Networks
- …

- Ensemble methods,
- Convolutional Neural Networks,
- Recurrent Neural Networks,
- Autoencoders,
- …

why did we get the output we got? → *post-hoc explainability*

explanation by

simplification

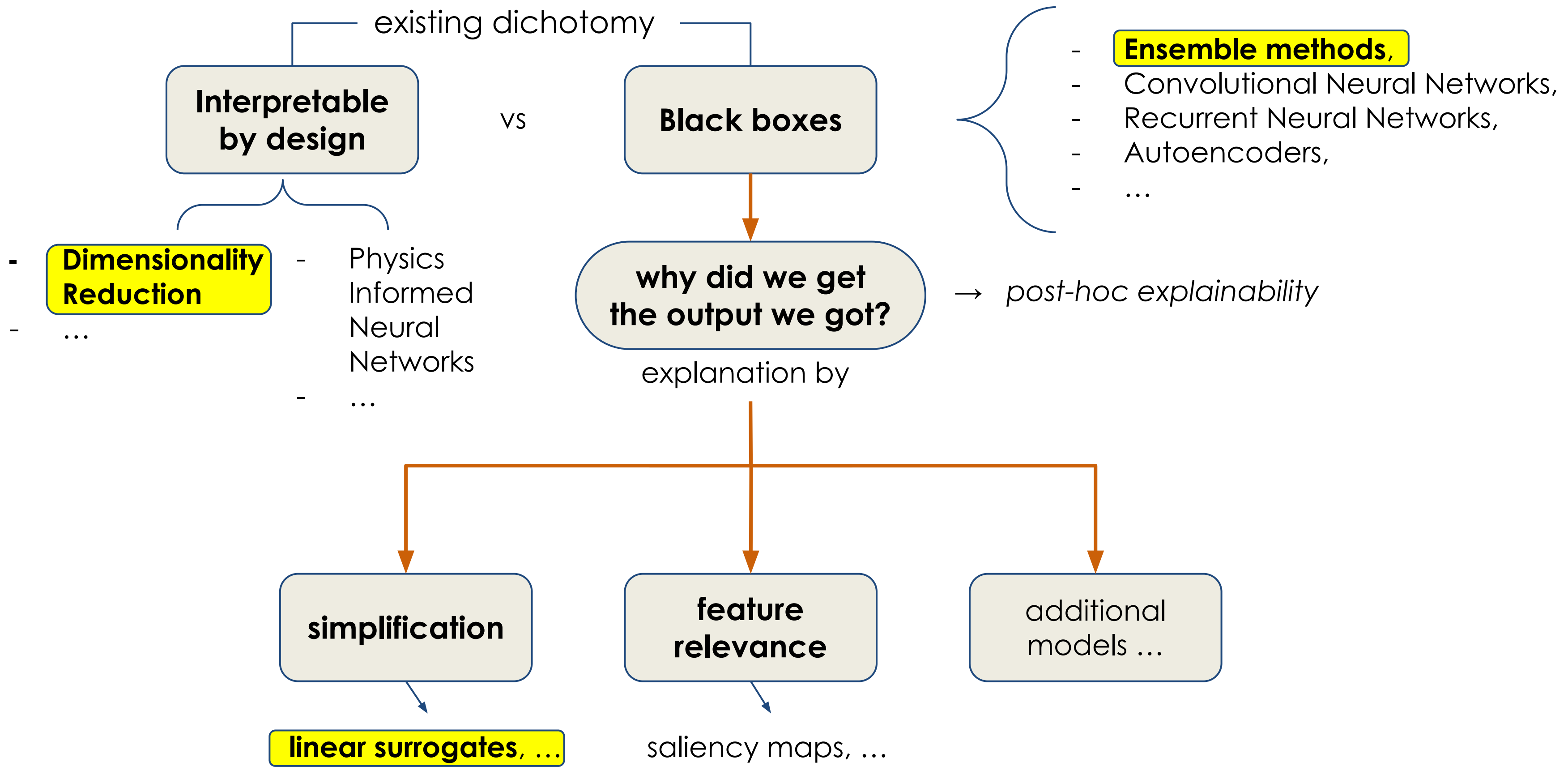
linear surrogates, …

feature relevance

saliency maps, …

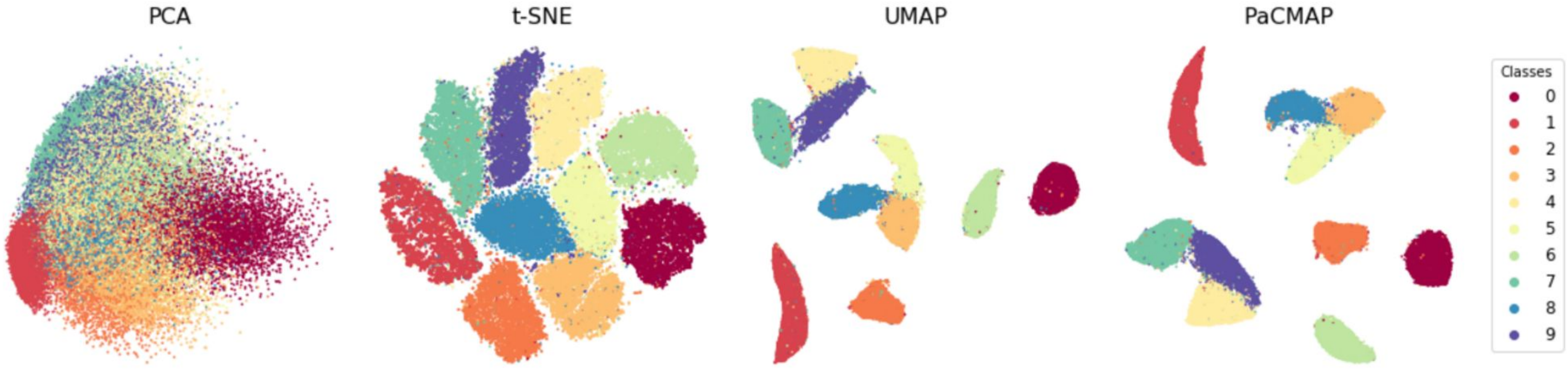
additional models …

Models interpretable by design vs black boxes that can be “explained”



Dimensionality reduction (DR) enables inspection of dataset structure

Dataset of handwritten digits **represented** through different **embeddings in latent space**.



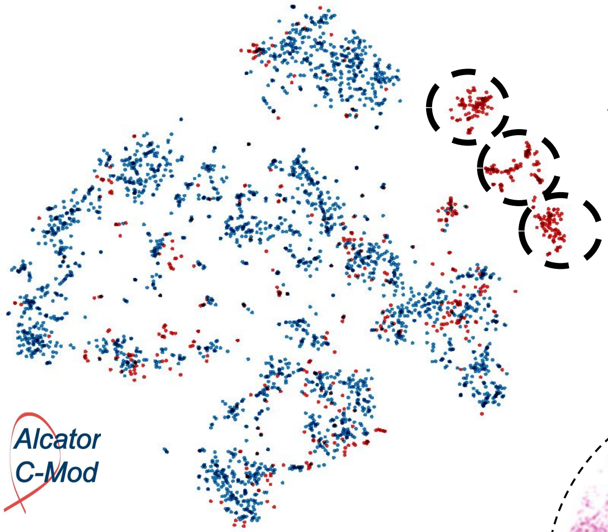
C. Rudin et al., 2022 *Stat. Surv.* **16** 1–85

- Latent space (no physical units) allows 2D visualization of **similar data points** in high-dimensional feature space.
  - All DR methods allow some form of data inspection and understanding.

Coloring done a-posteriori!

Clustering algorithms enable discovery of data patterns

Fusion

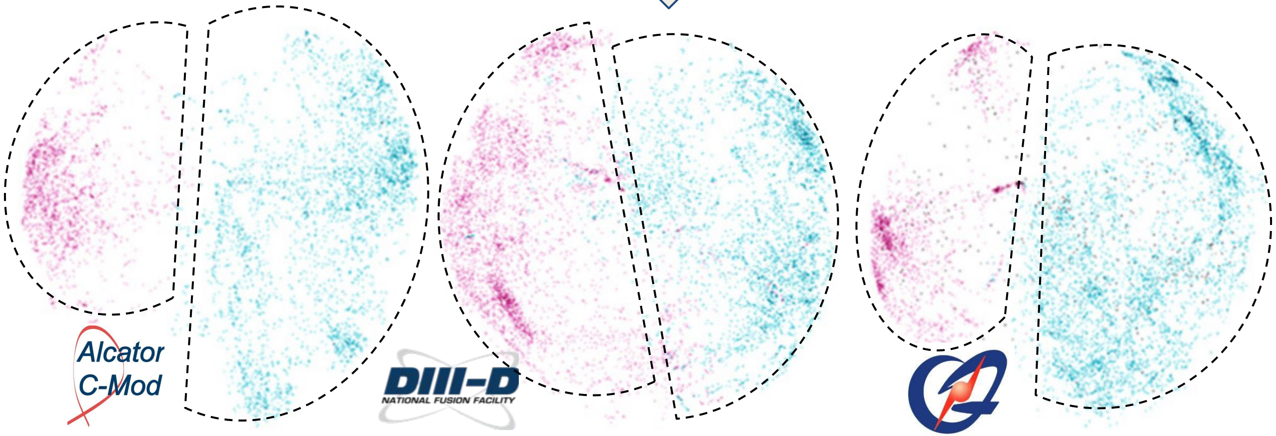


t-SNE clustering of C-Mod disruptive vs non-disruptive time sequences.

PCA clustering of two different performance regimes for three different tokamaks.

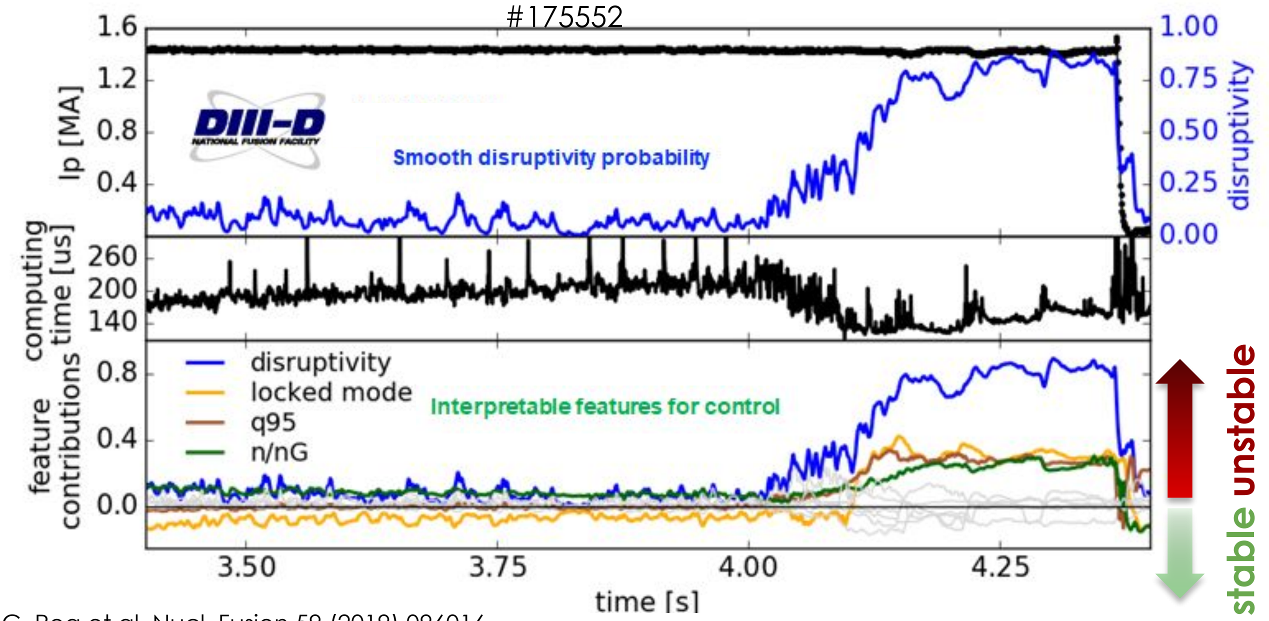
J.X. Zhu, C. Rea et al, 2021 Nucl. Fusion 61 026007

Coloring done a-posteriori!

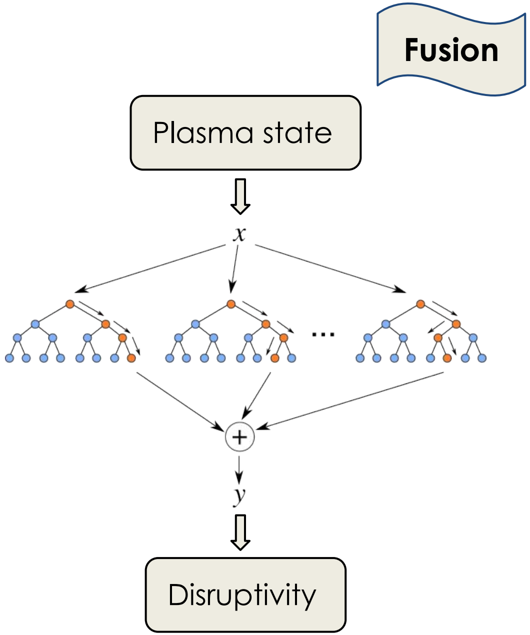


J.X. Zhu, C. Rea et al, 2021 Nucl. Fusion 61 114005

Explainable ML predictions for real-time proximity to instability – a Random Forest example



C. Rea et al, Nucl. Fusion 59 (2019) 096016
C. Rea et al, 2021 IAEA EX/P1–25,
J. Barr et al, Nucl. Fusion 61 (2021) 126019



Adapted from: "Random Forest Regression," *levelup.gitconnected.com*

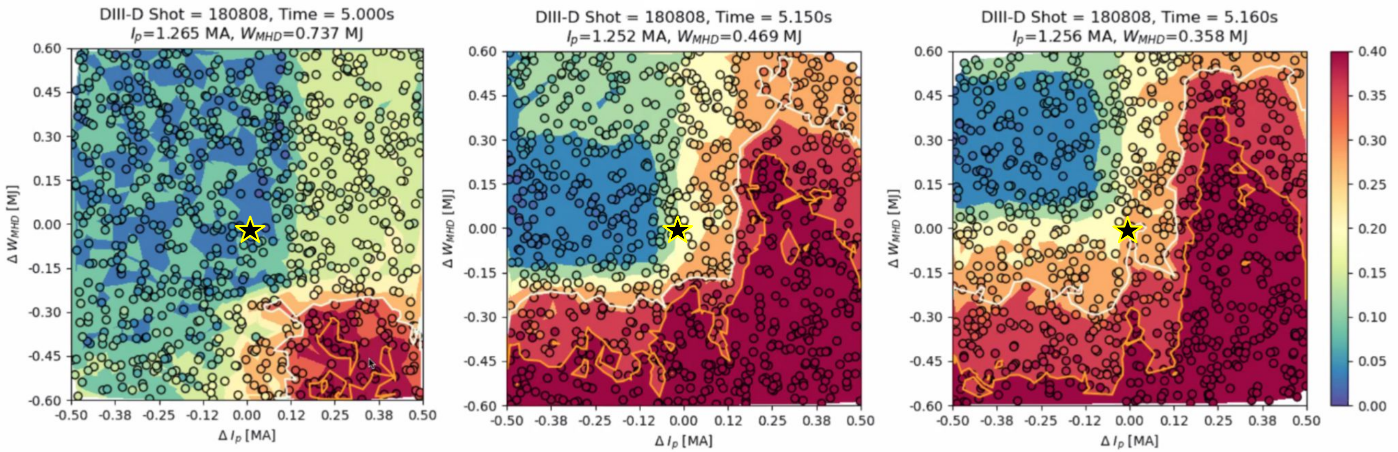
Advantages:

- Identification of stability boundaries in real-time.
- Local explainability metrics leveraged inside controllers to modify plasma trajectory in real-time.

Identification of safe operating region through fast ML enables trajectory planning

time

Fusion

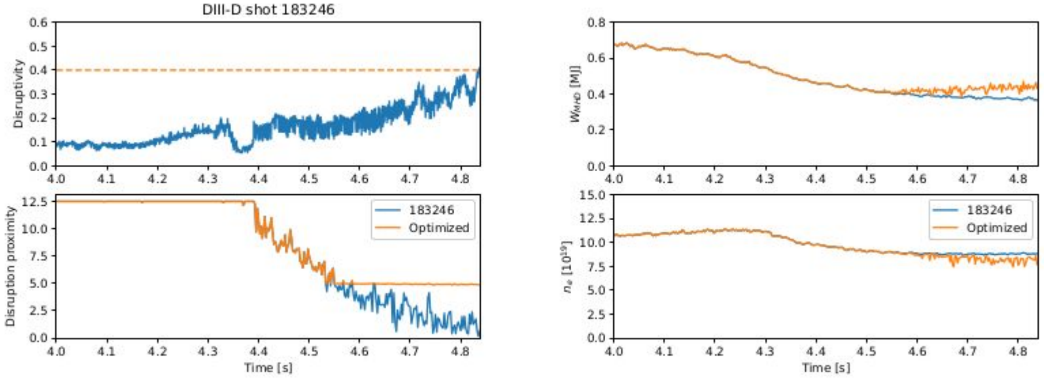


high danger

ML simulations evaluated by sampling from 2D operational regime variations

safe region

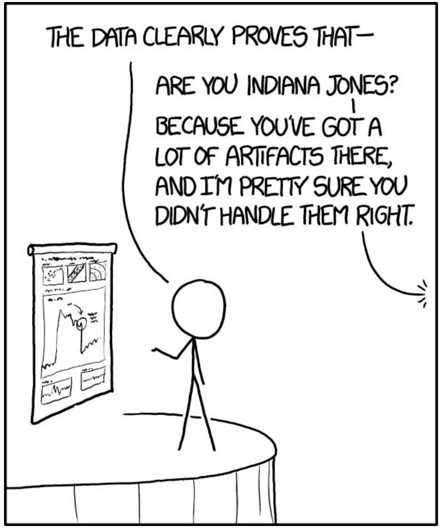
- Goal: leverage ML-driven optimization to identify trajectory across operational space and in real-time control systems.
  - Operating point optimized (Genetic Algorithms) via convex set of linear constraints to calculate disruption proximity.



M. D. Boyer, C. Rea, M. Clement, Nucl. Fusion 62 (2022) 026005
Continued work by Z. Keith, C. Rea (MIT)

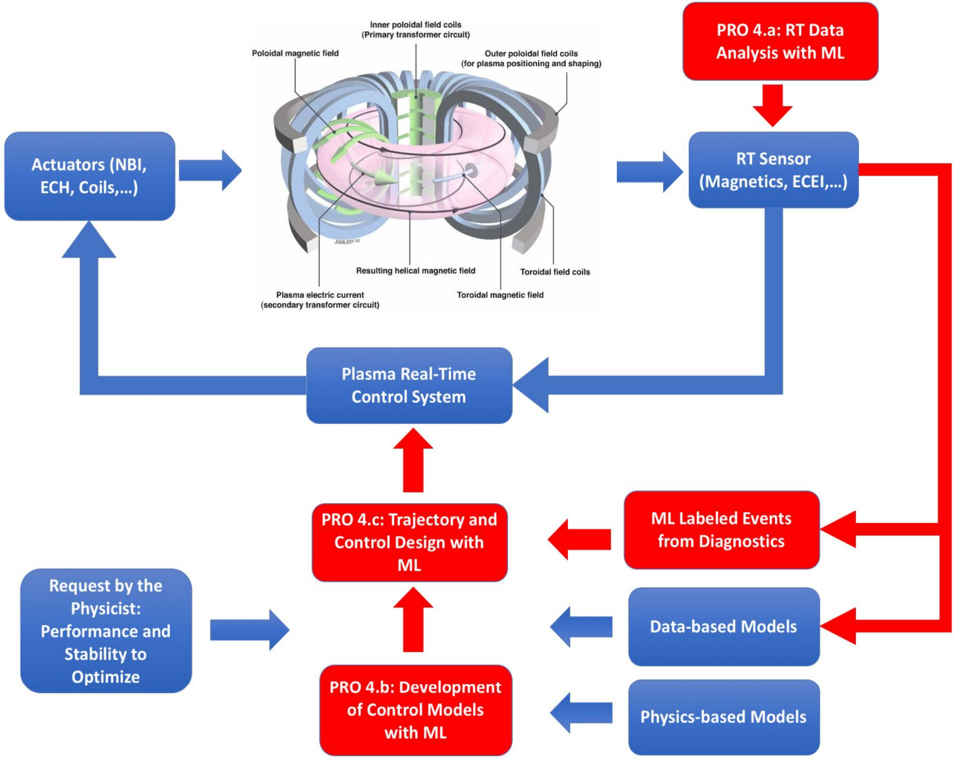
Outline

1. Fusion stuff and disruptions!
2. The Universality theorem and brief ML taxonomy
3. Explainable and adaptive ML models – applications in Fusion
4. Current challenges and opportunities for future research
5. Conclusions

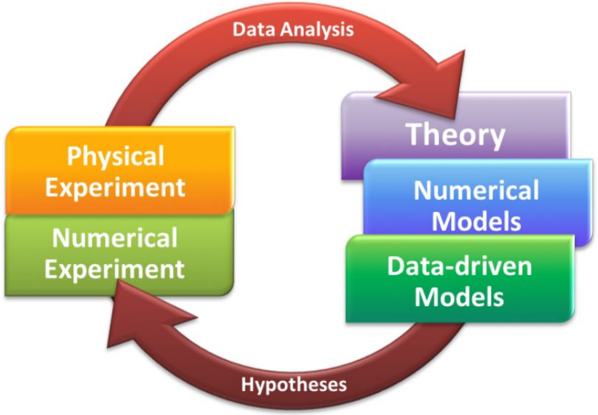


https://xkcd.com/1781/

Data drives fusion experiments' design, simulation, analysis, control and optimization – enabling science discovery



Adapted from D. Humphreys et al. "Advancing Fusion With Machine Learning" DOE Workshop (2020)



D. Humphreys et al. "Advancing Fusion With Machine Learning" DOE Workshop (2020)

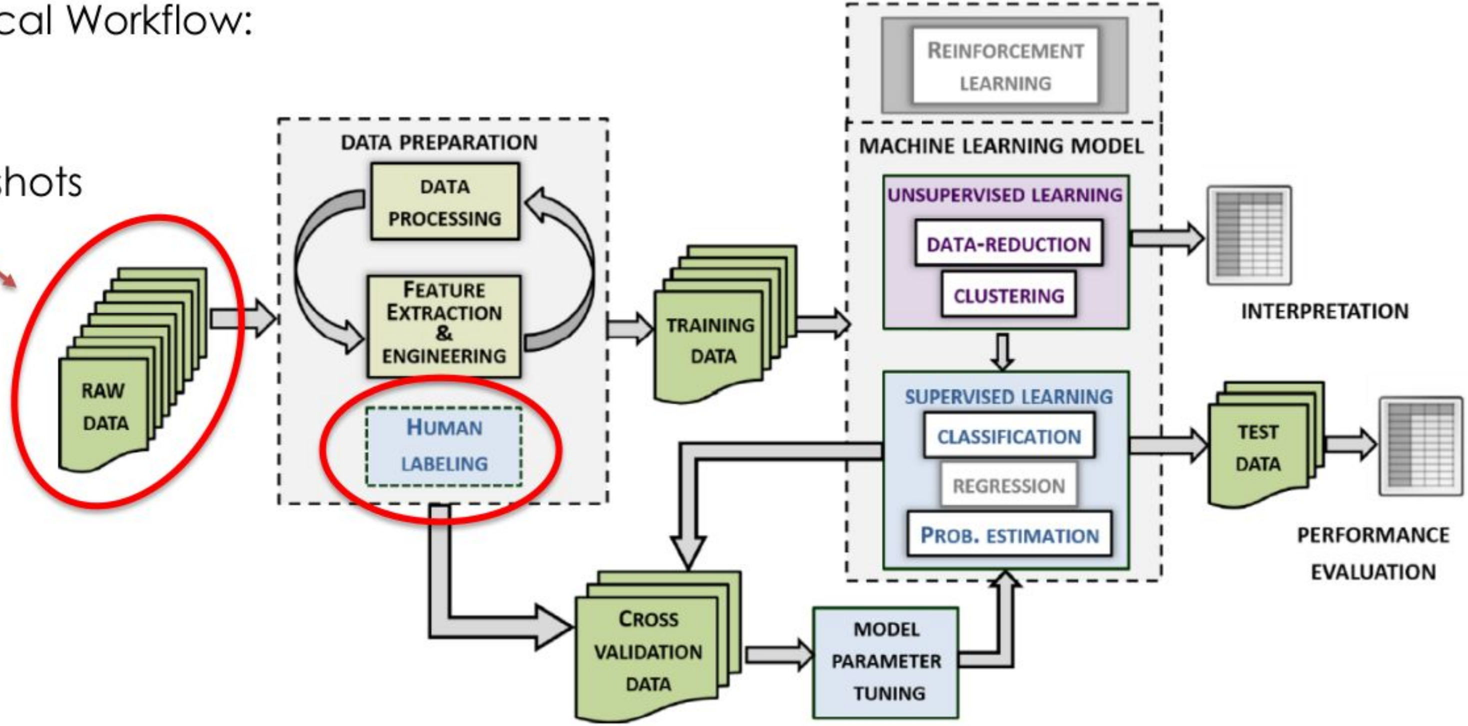
Pilot plants and next gen devices need robust models featuring

- **interpretability/explainability**
- well-defined validity and extrapolability boundaries
- …

ML requires <u>large</u> datasets with target events <u>labeled</u>

Typical Workflow:

$\sim 10^3 - 10^4$ shots



A. Pau et al, Nuclear Fusion, 59(10):106017, 2019

Labeling events often requires manual inspection of multiple signals: semi-supervised ML accelerates event labeling

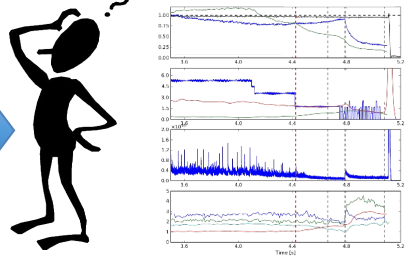
Most shots go unexamined

Diagnostic Data

- From MDSplus, etc.
- Many ($\sim 10^3$) shots

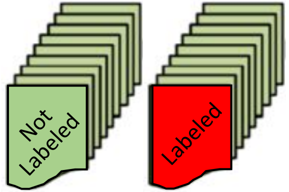
Some shots analyzed

Human Expert



- Use diagnostic signals to identify events
- Record times at which event occurs
- Validate 'learned' event detections

Partially Labeled Event Data



- Common signals from many shots
- A fraction (at least 1 shot) has labeled event occurrence

Semi-Supervised Learner

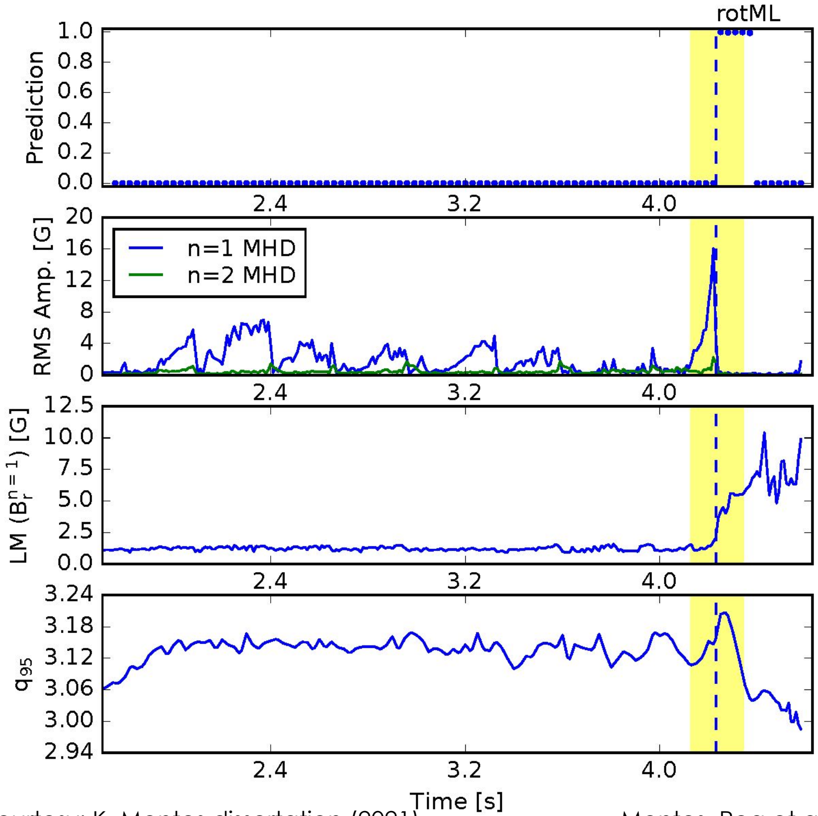
- Use unlabeled data distribution to find other events
- Output times when event likely occurred

Predictions & performance results

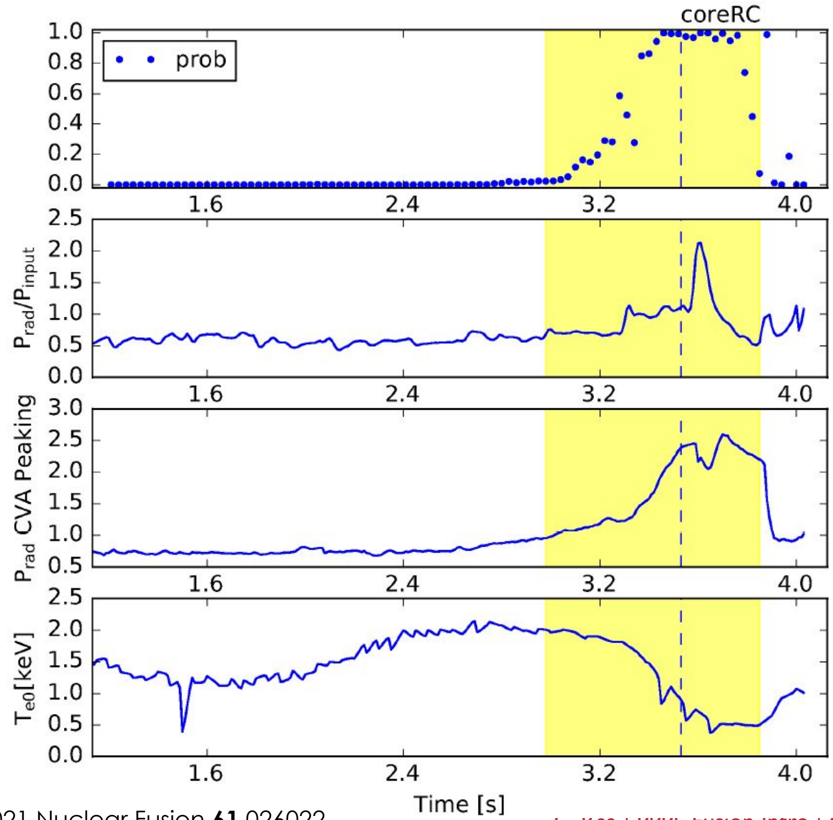
Courtesy: K. Montes (MIT) dissertation (2021)

Semi-supervised label spreading algorithm to automate detection of physics events

Rotating mode locks (**rotML**)



Core radiative collapse (**coreRC**)

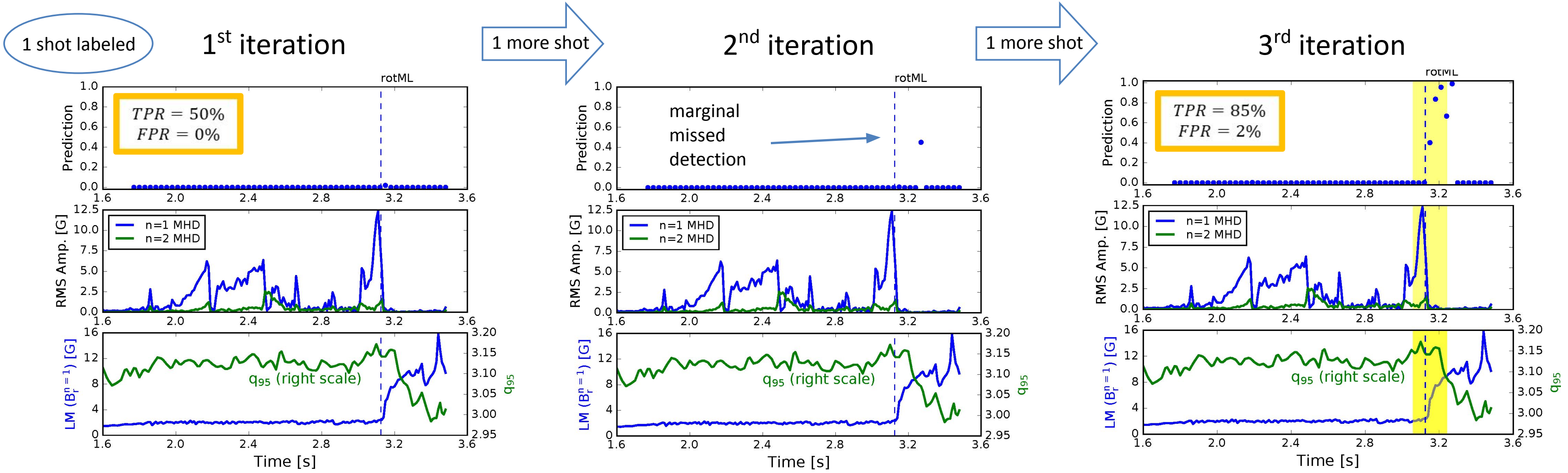


 Courtesy: K. Montes dissertation (2021)

Montes, Rea et al 2021 Nuclear Fusion **61** 026022

Can be applied to accelerate the construction of events databases

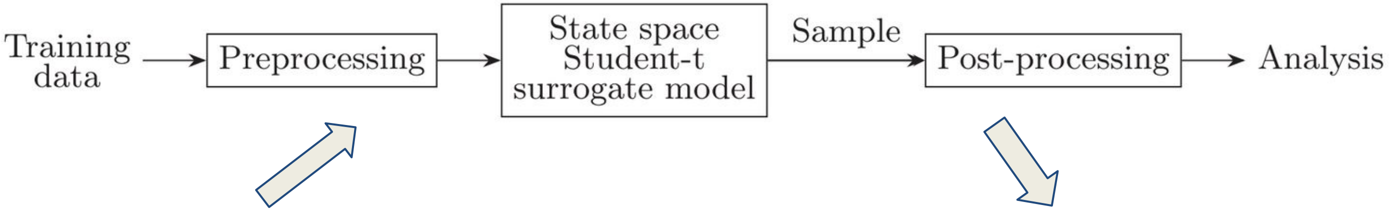
- Iteratively choose initially labeled shot from set of marginal detections
- Prediction quality on unlabeled shots improves consistently
  - Detected events in ~ 85% of hundreds of shots after manually analyzing just ~ 1%

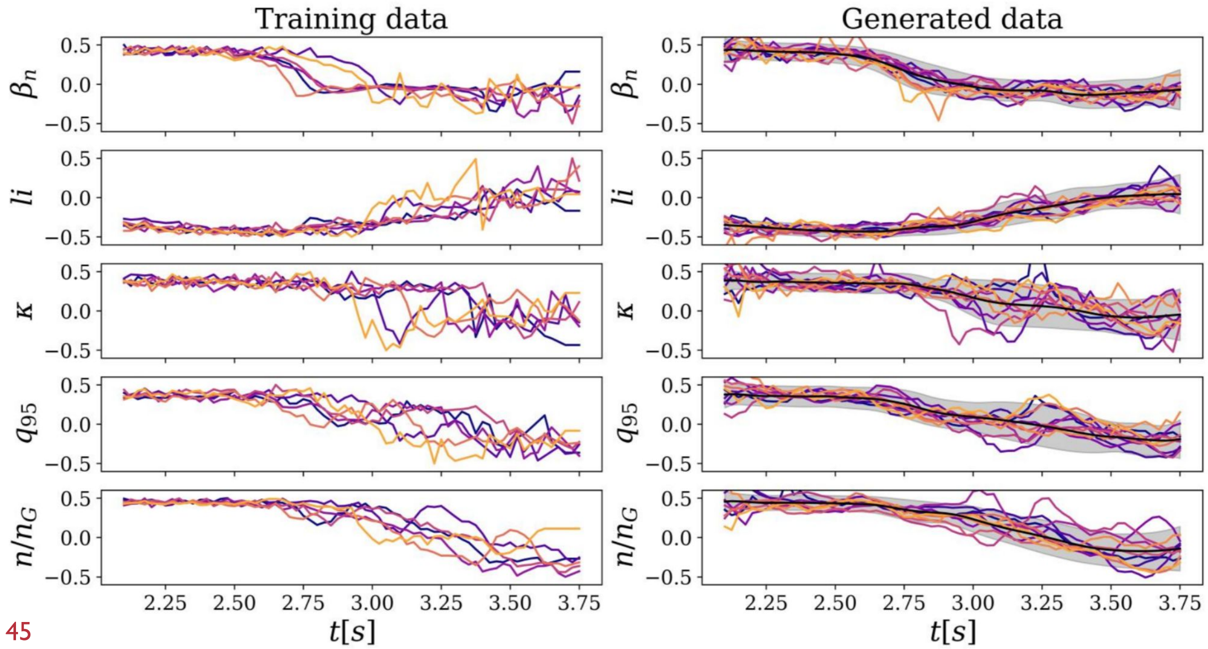


Courtesy: K. Montes dissertation (2021)

Montes, Rea et al 2021 Nuclear Fusion **61** 026022

Data augmentation to learn disruptive dynamics (I)

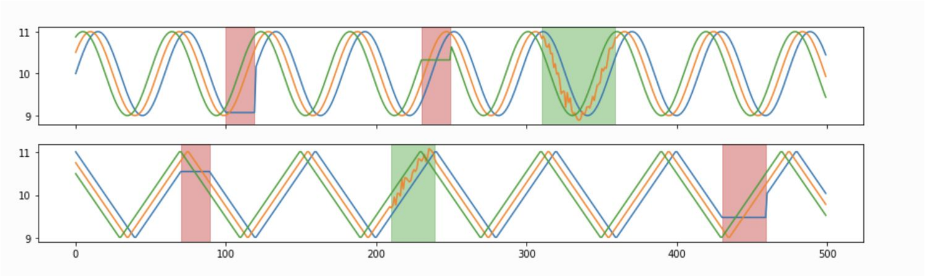


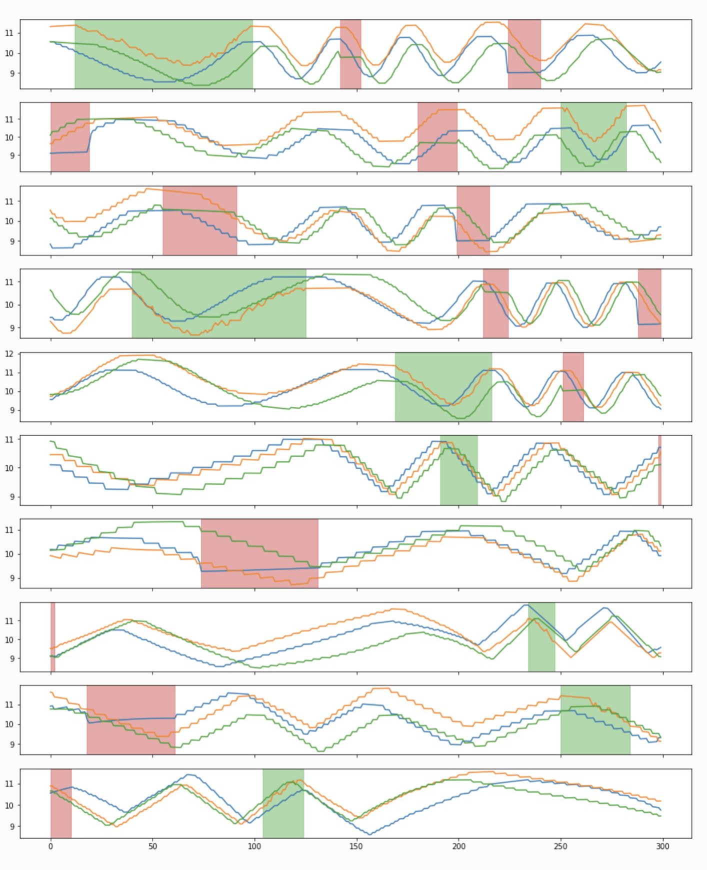


- **DL models are data-greedy**: need comprehensive training database to achieve satisfying and reliable results.
- Robust augmentation of the training database using **state space Student-T surrogate models**.

Rath, (...), Rea et al, "Data augmentation for disruption prediction via robust surrogate models" J. Plasma Phys. (2022), vol. 88, 895880502

Data augmentation to learn disruptive dynamics (II)

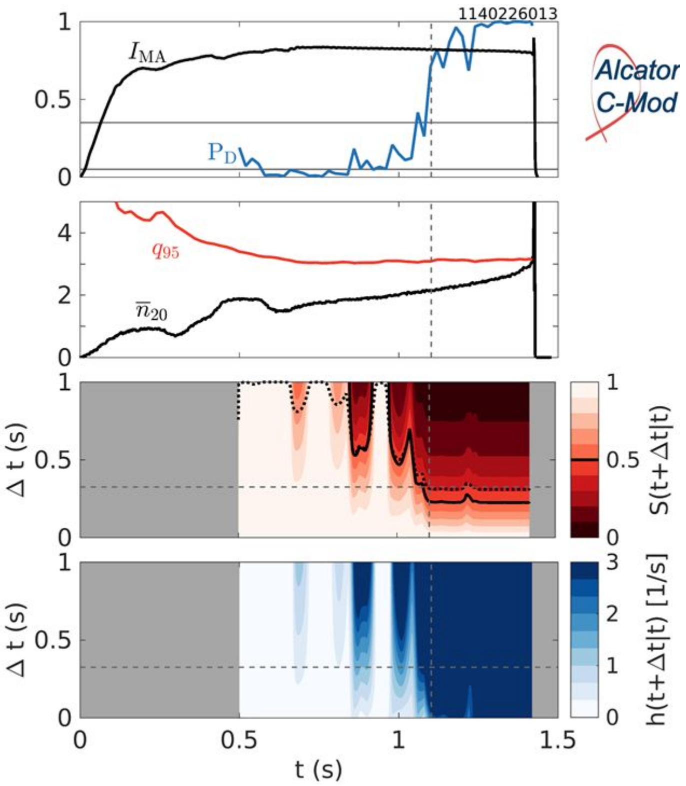




- random time warping 5 times in parallel,
- random crop subsequences with length 300,
- random quantize to 10-, 20-, or 30- level sets,
- with 80% probability , random drift the signal up to 10% - 50%,
- with 50% probability, reverse the sequence.

L. Spangher, C. Rea et al 2023

Predict "time-to-disruption" risk using classification probability



Any classification probability (P_D) cast between [0,1] can be used to:

- Predict the **future probability** of **plasma survival** $S(t+\Delta t\,|\,t)$ [1]
  or
- Model the **instantaneous hazard** [2,3] h=d ln$S$/dt to be used as probability generator.

Hazard function modeling connects dynamical systems and risk-aware control design by probability generation.

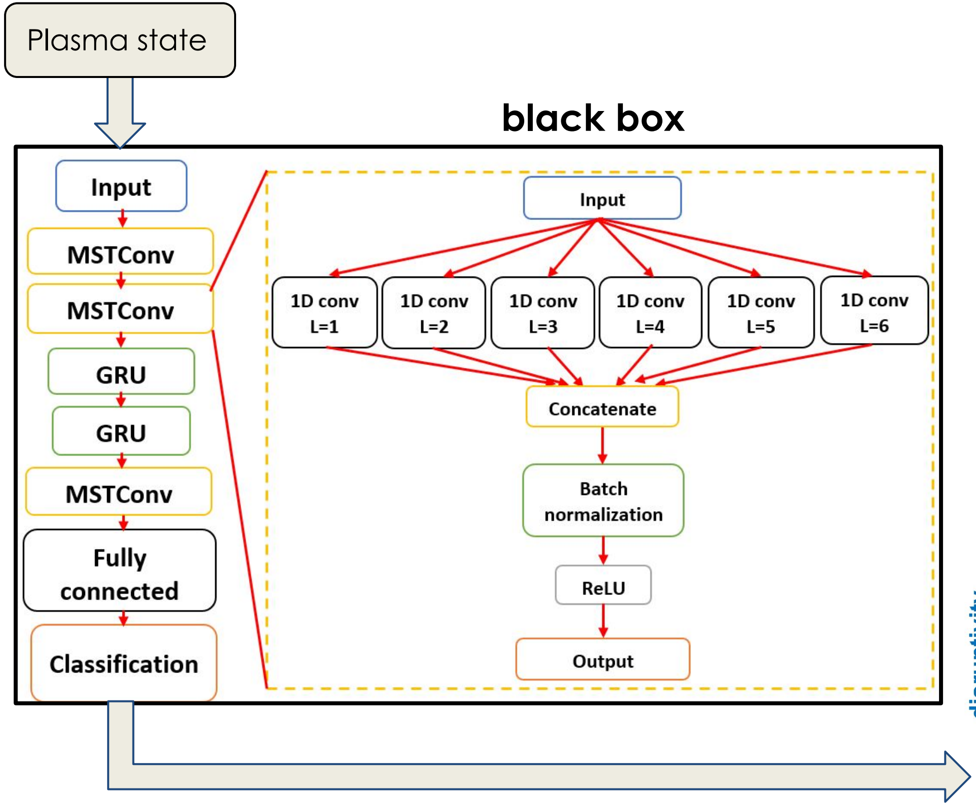
C-Mod data used as proof of concept to combine DPRF (or <u>any classifier</u>) disruptivity with survival analysis.

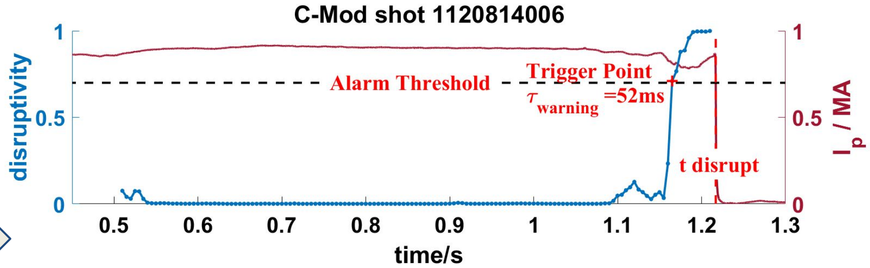
[1] RA Tinguely et al 2019 PPCF 61
[2] KEJ Olofsson et al 2018 PPCF 60
[3] KEJ Olofsson et al 2018 FED 146

Continued work by Z. Keith, C. Rea (MIT)

Hybrid Deep Learning predictor for cross-machine disruption prediction using time series data

not enough time:
J.X. Zhu, C. Rea et al, 2023 Nucl. Fusion 63 046009

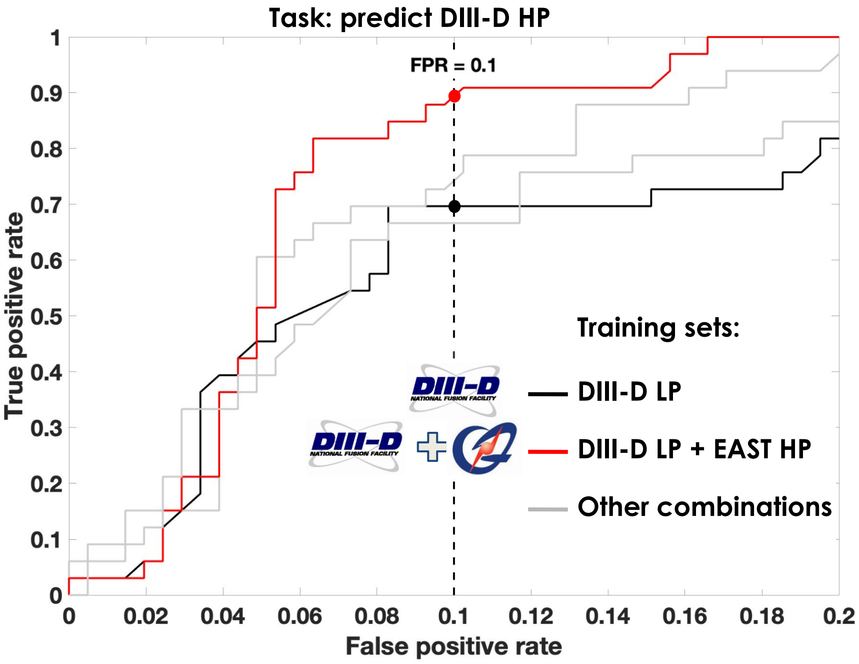




Adaptive strategies designed to optimize predictions across different fusion devices

Fusion

- **Adapt** current state-of-the-art **ML predictors** to different operational regimes across devices (DIII-D/EAST).
- **Implications for next-gen, yet-to-be-built devices!**
- Adaptive strategies:
  - ad-hoc design of **training sets to match target domain** by fully exploiting existing data[1],
  - **retrain predictors** after performance degradation[2].

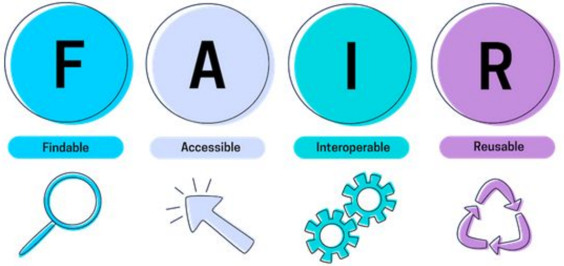


Adapted from [1]J.X. Zhu et al, NF (2021) 114005
[2]J. Vega et al., *Nat. Phys.* 18, 741–750 (2022)

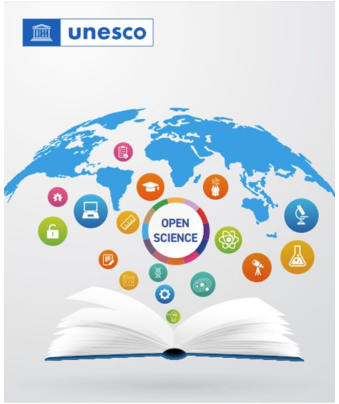
Domain adaptation

Existing current challenges in ML-applied research 😟 but also (!) <u>opportunities for future scientists</u> 🧑‍🎓

- trust in performance metrics → **missing benchmarks**
- trust in predictive output and learning → **model interpretation and explanation accuracy**
- prediction of out of distribution samples → **domain shifts, data shifts**
- integration with legacy architectures → **real-time vs offline implementations**
- **lack of labeled data** or of reliable (and automated) metadata extraction
- **uncertainty quantification**
- **open** and **FAIR** (!) **access** to data and models



M. Wilkinson, *et al.* The FAIR Guiding Principles for scientific data management and stewardship. *Sci Data* **3**, 160018 (2016)



UNESCO Recommendation on Open Science

DOE and International Agencies strongly support ML research to accelerate Fusion progress

- **2019 DOE-sponsored workshop** critical PROs identified
- **DOE Public Reusable Research Data** (PuRE) initiative

https://science.osti.gov/Initiatives/PuRe-Data

- **IAEA Coordinated Research Project** addressing cross-cutting issues

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	

D. Humphreys et al. "Advancing Fusion With Machine Learning" DOE Workshop (2020)

https://nucleus.iaea.org/sites/ai4atoms/ai4fusion/

Summary and conclusions

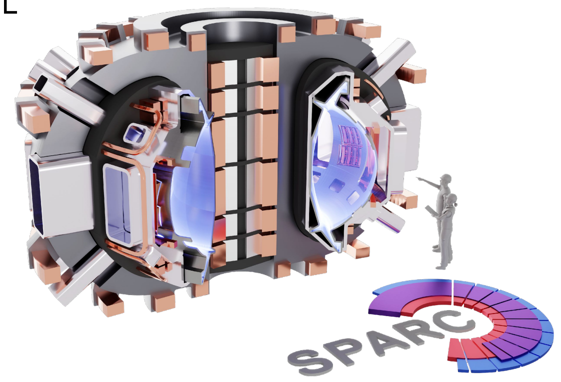
Plasma Science and Fusion Center

1. Fusion science and technology advancements accelerated by ML

2. ML black boxes can be explained
   - and accuracy does not prevent interpretability!

3. Enforce interpretability to reconcile with physics understanding

4. Fusion examples already out there employing
   a. Interpretable algorithms
   b. Explainable predictions
   c. Transfer/adaptive learning and statistical optimization
   d. Surrogate modeling for fast reconstructions
   …



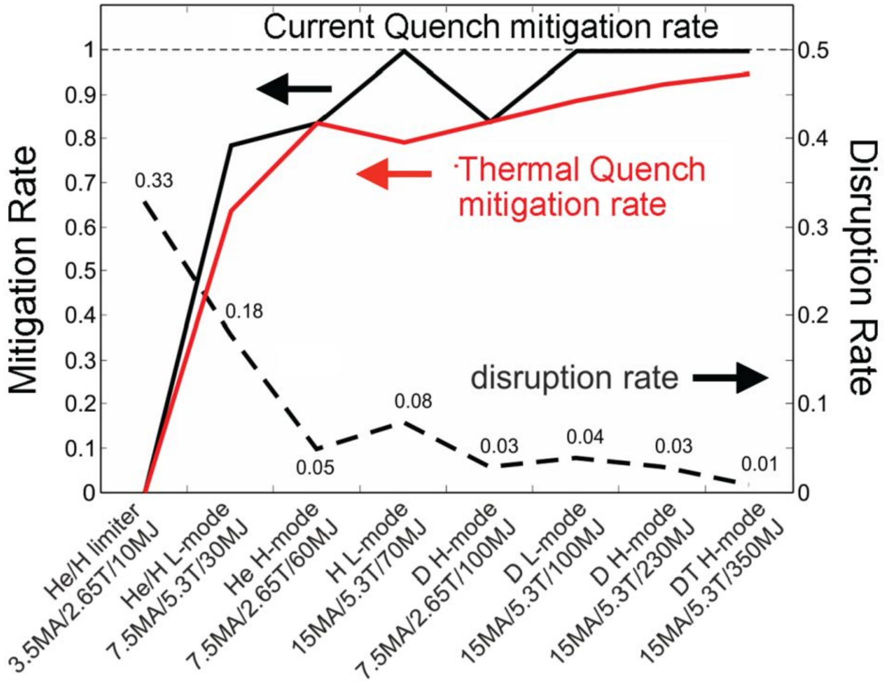
Credits: S. Mordijck

Long list of open questions and cross-cutting challenges,
but also **opportunities for future research, enabling change in the field!**

Reach out if interested in research opportunities at PSFC! crea@psfc.mit.edu

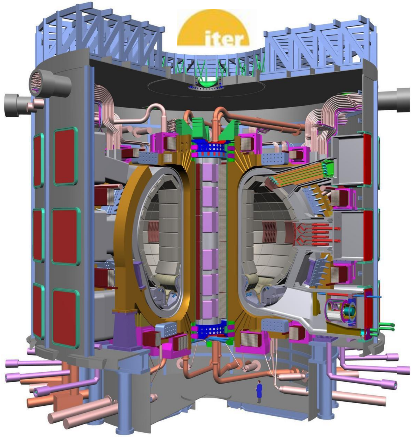
More stuff

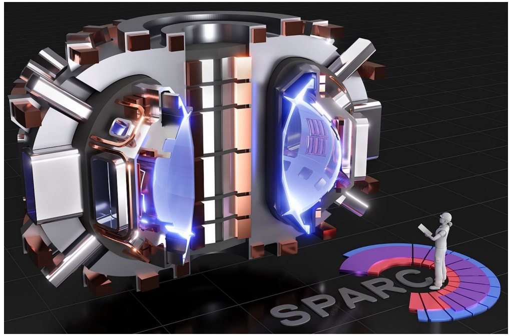
Next generation devices will not allow disruption rate > 1% at full performance



54

Lehnen M. et al 2016 "Plasma disruption management in ITER", 2016 IAEA Fusion Energy Conf. EX/P6-39
Reproduced at E. Strait et al Nucl. Fusion (2019) 112012



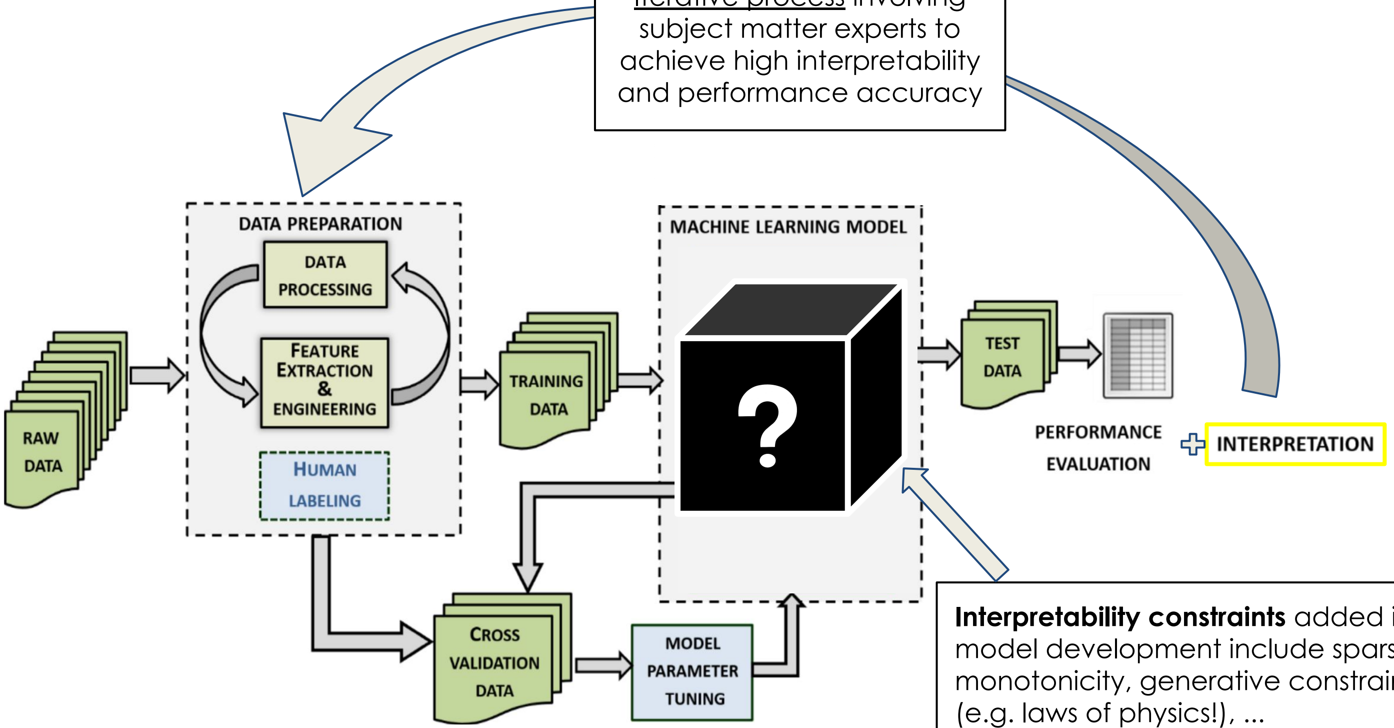


A simple, interpretable, and accurate model *should* exist, maybe (computationally) hard to develop

Supervised and interpretable ML classification workflow:

Adapted from A. Pau et al, Nuclear Fusion, 59(10):106017, 2019

Iterative process involving subject matter experts to achieve high interpretability and performance accuracy



Interpretability constraints added in model development include sparsity, monotonicity, generative constraints (e.g. laws of physics!), ...

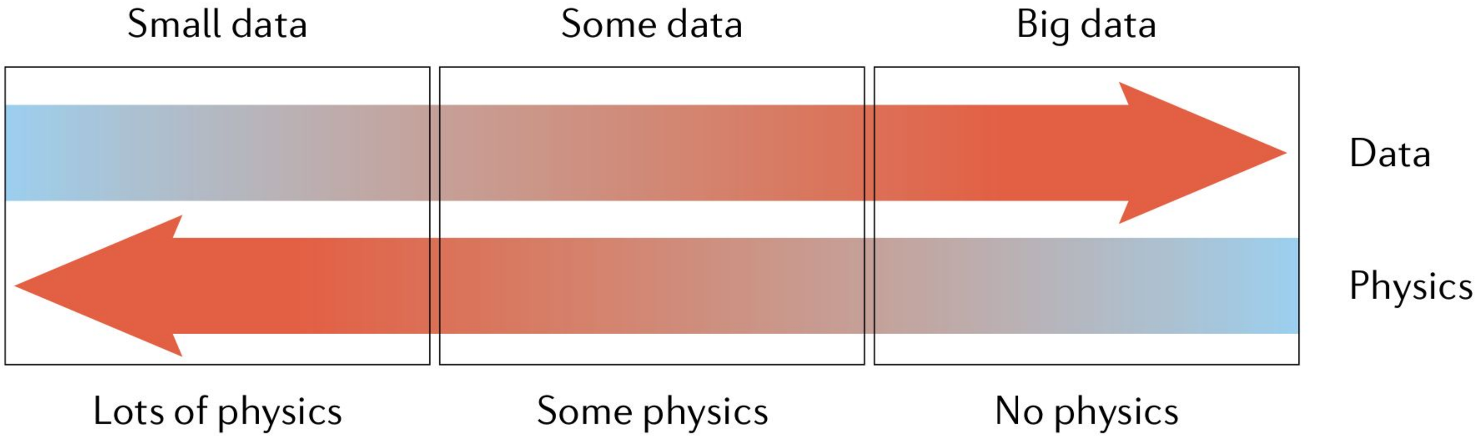
Examples of explainable models interpretable by design

Interpretable by design

- ❑ Dimensionality Reduction
- ❑ Physics Informed Neural Networks

Physics-informed machine learning seamlessly integrates data and governing physical laws

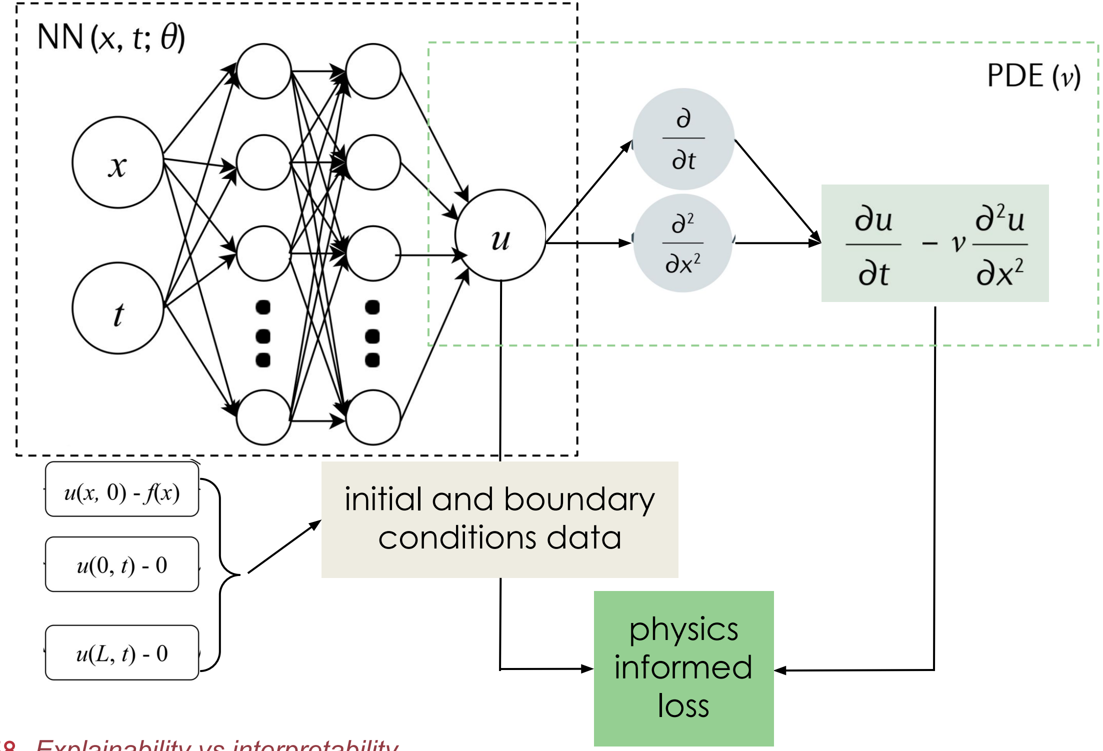
- NN and AutoDiff allow to design models with **partially missing physics (or data!)**
  - No need of domain adaptation or transfer learning.
  - Strong generalization, by enforcing/embedding physics constraints.
  - Can tackle high-dimensional problems.
  - Can address uncertainty due to physics, data, and learning models.



G.E. Karniadakis et al., 2021 Nat Rev Phys 3, 422–440

Physics Informed Neural Networks (PINNs) preserve interpretability through physics constraints

- PINN learns partial differential equations (PDEs) given initial and boundary conditions (I&BC): heat equation example.



- PINN training minimizes the PDEs residuals + I&BC, through **combined loss function** and **automatic differentiation**.
- **No need of labeled data, only generative constraints!**

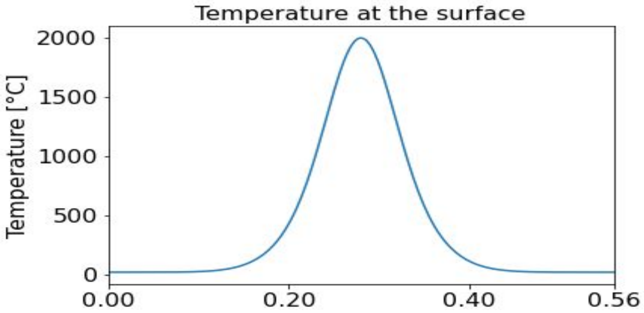
Adapted from:
C. Rudin et al., 2022 Stat. Surv. 16 1–85
G.E. Karniadakis et al., 2021 Nat Rev Phys 3, 422–440

PINN solves heat equation and computes heat flux on the top surface of W7-X divertor tiles

Gaussian top boundary condition constant in time $(t \in [0, 0.1]\ \mathrm{s})$

E. Aymerich et al., PSI-25 Poster

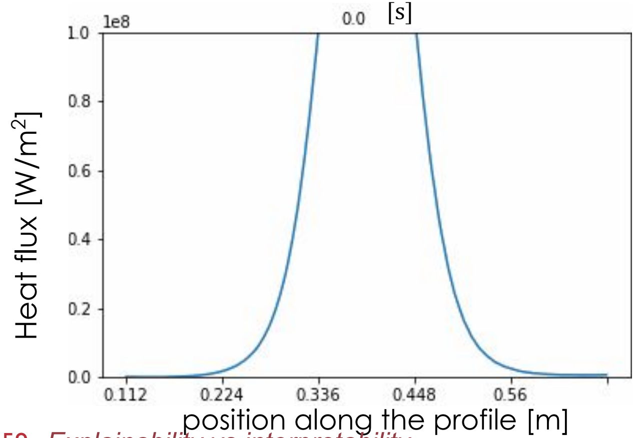
Fusion



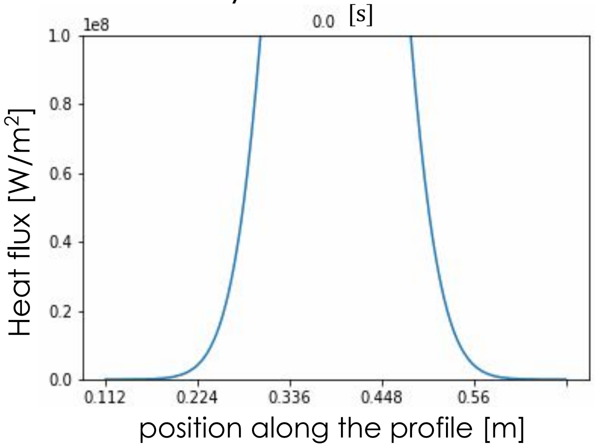
The PINN solves the equation and then computes the derivative on the top surface of the profile and estimate the normal heat flux thanks to automatic differentiation:

$$\boldsymbol{q} = -\nabla_n \boldsymbol{u}$$

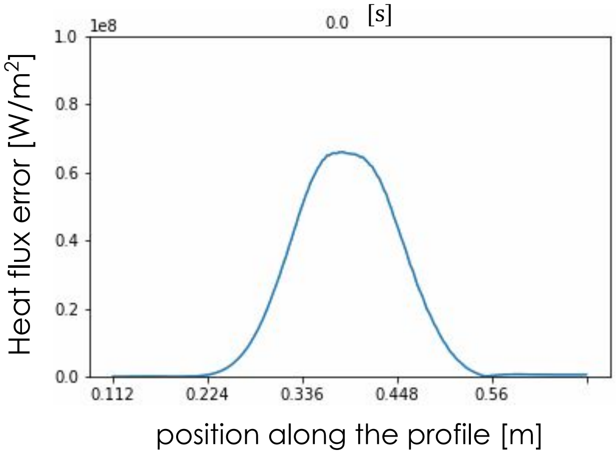
PINN



Physics simulation



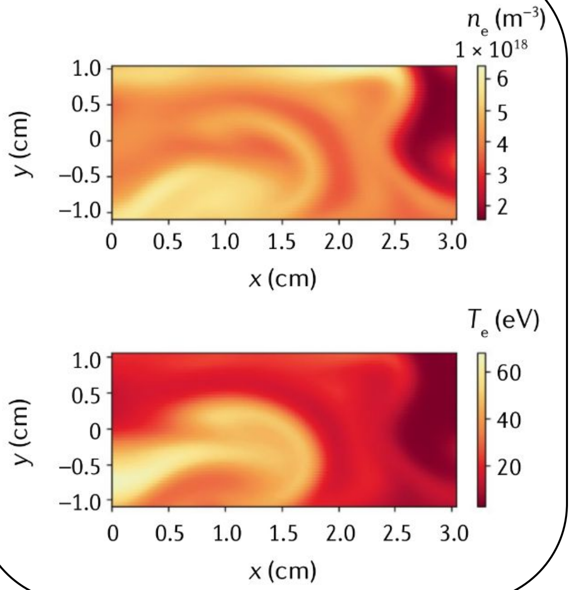
Error



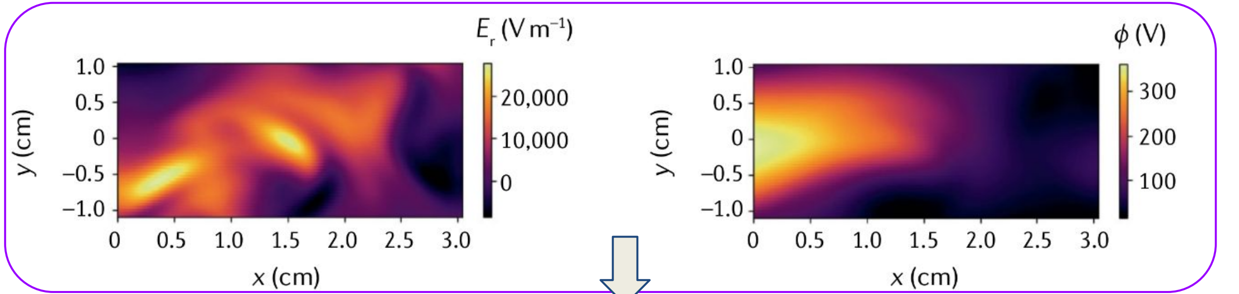
PINNs can accurately learn turbulent field dynamics consistent with theory, and from partial observations

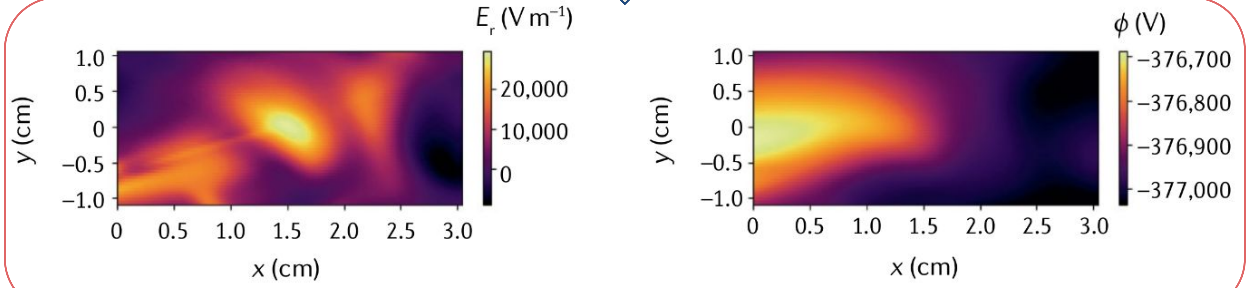
Fusion

Partial observations of Te, ne from one test discharge



Reference target solutions





PINN reconstructions

A. Mathews, et al, Phys. Rev. E 104, 025205 (2021)

Physics-informed machine learning: current limitations

- **Multiscale and multiphysics problems require further developments**.
  - High-frequency functions difficult to learn → F-principle or spectral bias.

- **PI ML involves highly non-convex optimization problems for complex loss functions**.
  - Need more robust algorithms and computational frameworks.
  - Meta-learning techniques to automate the design of best architectures?

- **Missing benchmarks** on openly available datasets from physics, chemistry, …

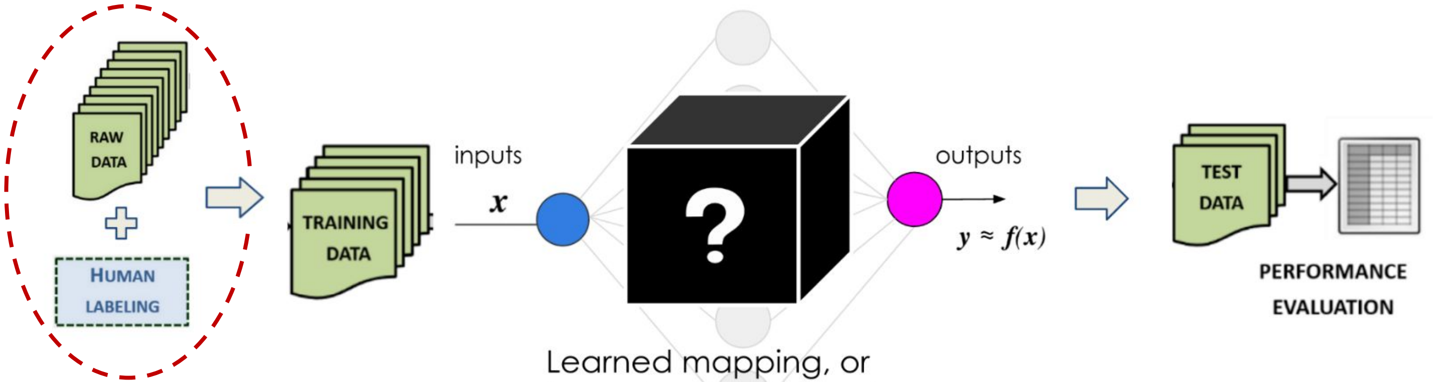
- **More research needed on the theoretical foundations of NN**.

G.E. Karniadakis et al., 2021 Nat Rev Phys 3, 422–440 and references therein

Domain adaptation and transfer learning: ML mapping from inputs to outputs, or learning to perform a task

Simplified supervised ML classification workflow:

Adapted from A. Pau et al, Nuclear Fusion, 59(10):106017, 2019



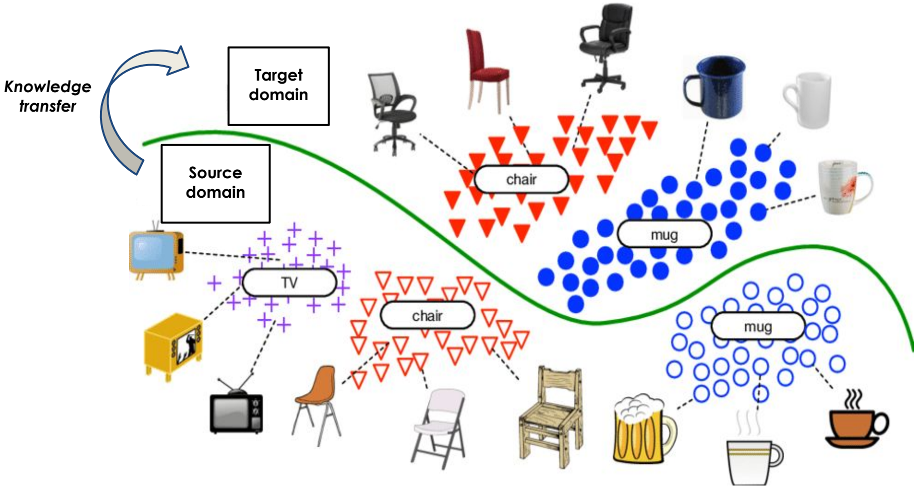
domain = data + labels

Learned mapping, or boundary between classes or **learning a task**

- Mapping from inputs to outputs through ML systems means to **learn to perform a task**.

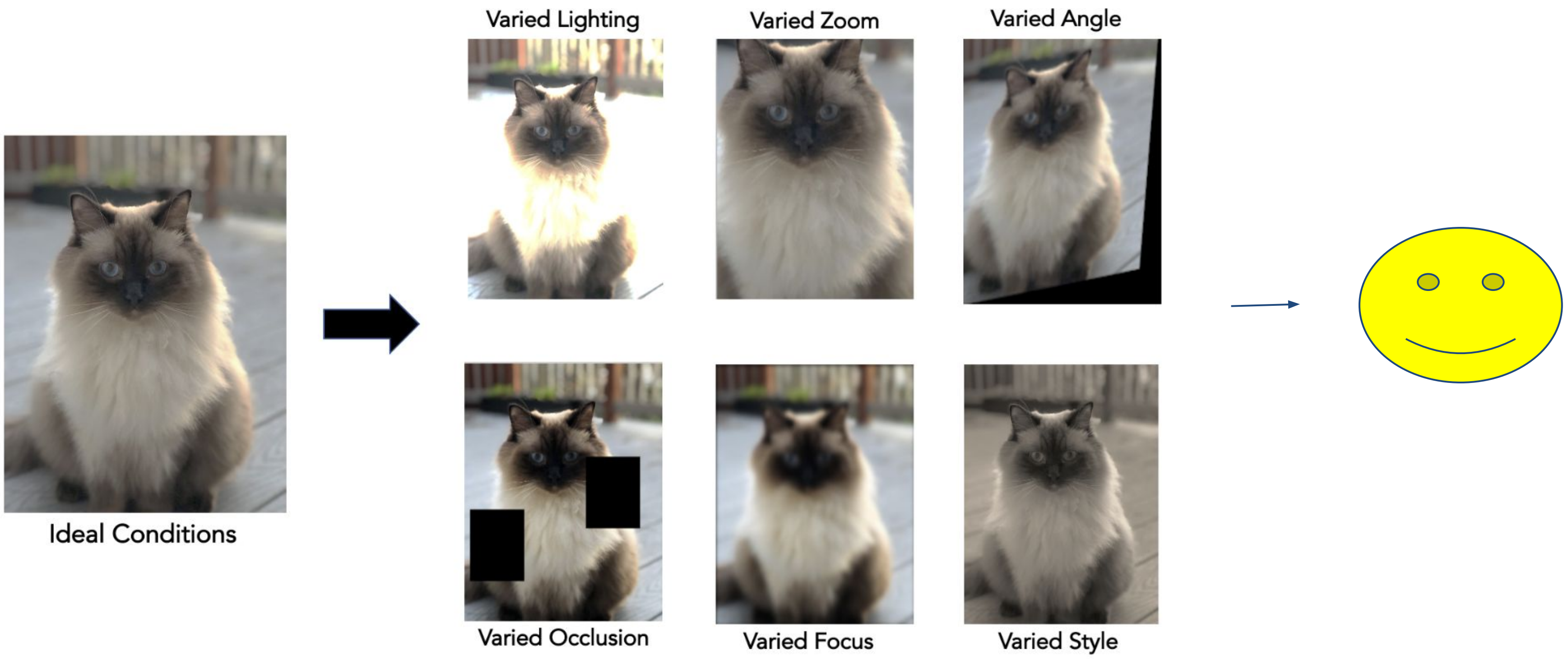
Learning a task heavily depends on dataset composition

- What if the collected data is **not an accurate reflection of the population**? Too limited, not accurately labeled, …
  - Learning *a general data representation* by finding common embeddings of source/target data!



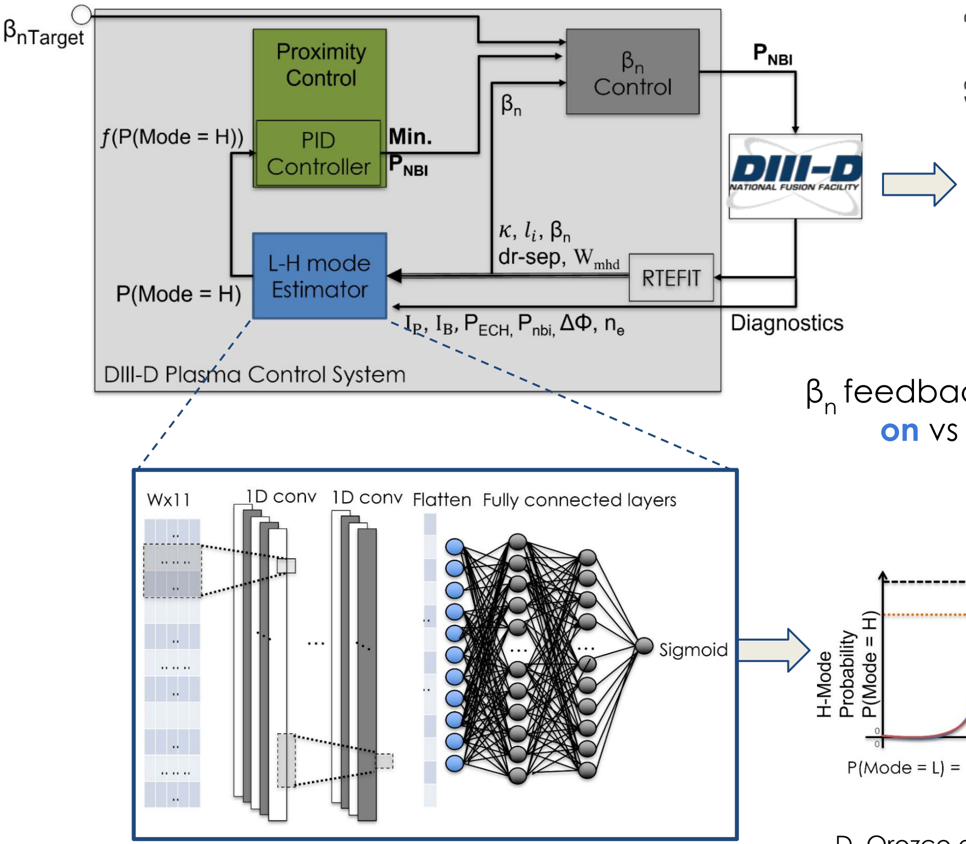
Domain adaptation and **transfer learning** designed to overcome **biased** dataset and/or **generalize** knowledge across **different tasks**.

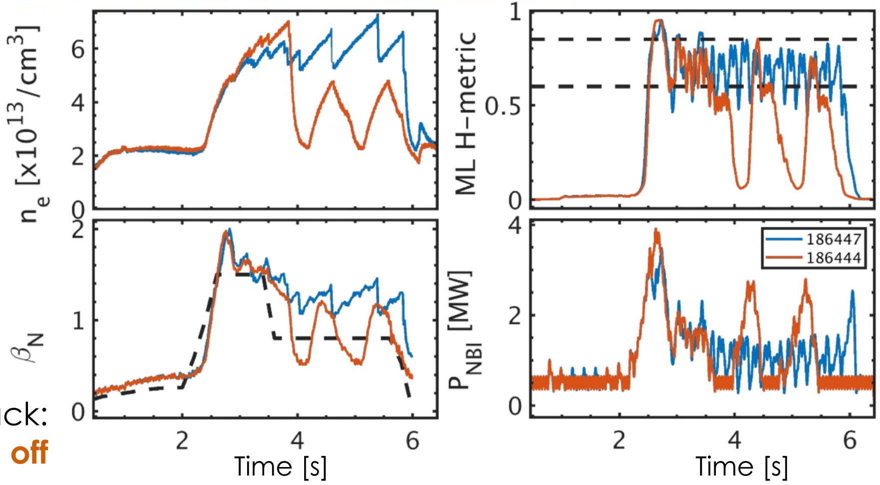
Data Augmentation: brief background



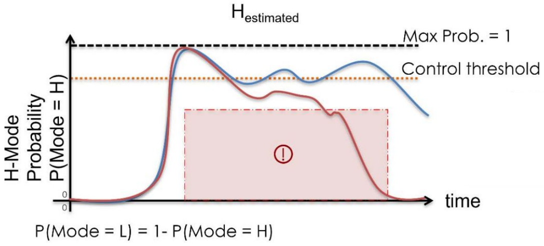
More (non-disruptive) examples:

Temporal Convolutional Neural Network predicts confinement probability 1ms in the future





β_n feedback: **on** vs **off**



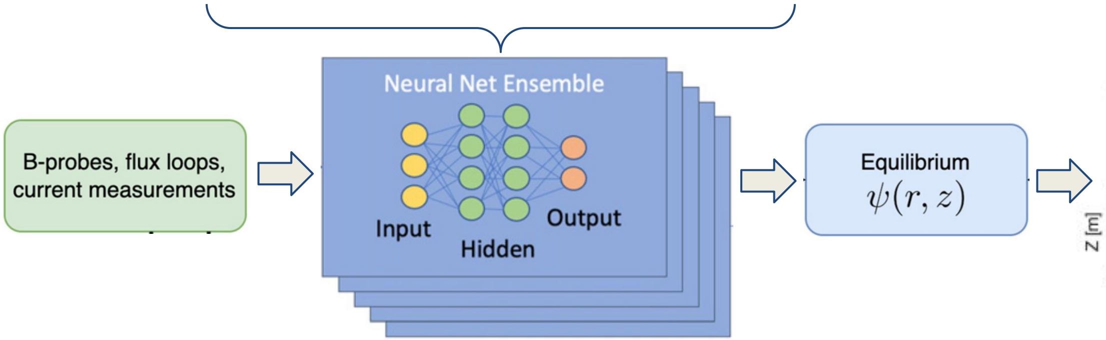
caution: limited expert-labeled dataset constraints model applicability

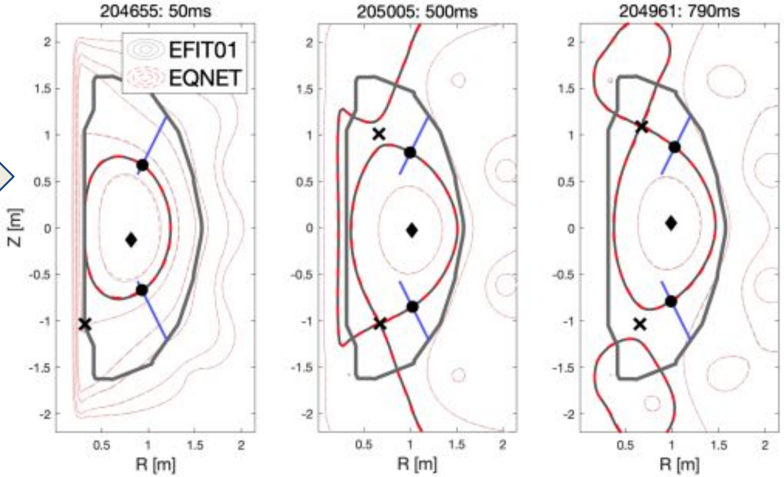
D. Orozco et al, IEEE TPS 2022

Neural networks accelerate equilibria reconstructions and profile evolution for shot planning and real-time control

caution: extrapolation to never seen equilibria or to other machines

$$\frac{\partial^2\psi}{\partial r^2} - \frac{1}{r}\frac{\partial\psi}{\partial r} + \frac{\partial^2\psi}{\partial z^2} = -\mu_0 r^2 \frac{dp}{d\psi} - \frac{1}{2}\frac{dF^2}{d\psi}$$



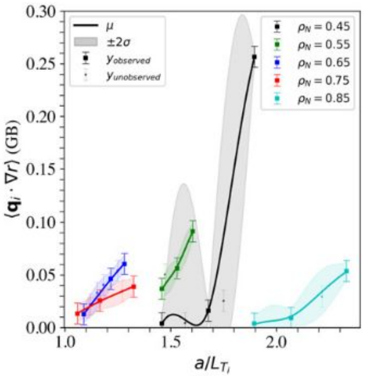


- NN learns free-boundary Grad-Shafranov solutions and reconstructs tokamak equilibria.

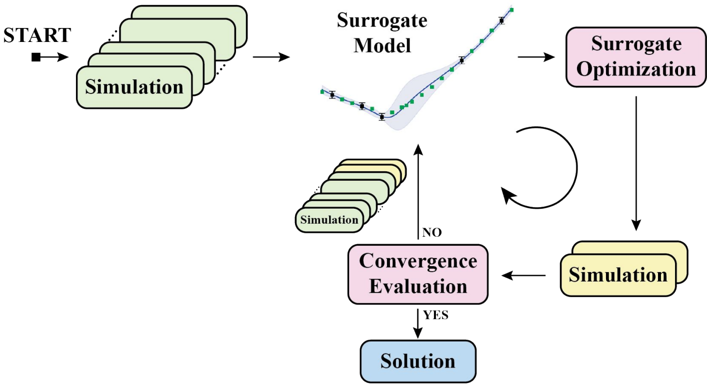
Adapted from J. Wai, M. Boyer, E. Kolemen, *Nucl. Fusion 62 (2022) 086042*

Gaussian Processes (GP) enable nonlinear simulations for performance prediction and gyrokinetic validation

- Few (10-20) simulations required to reach convergence, thanks to Bayesian Optimization (BO) workflow + GP surrogate modeling.
- Enabling profile predictions of unprecedented accuracy for:
  - ✔ Prediction of burning plasma performance (e.g. SPARC)
  - ✔ Validation of gyrokinetic codes (e.g. DIII-D)

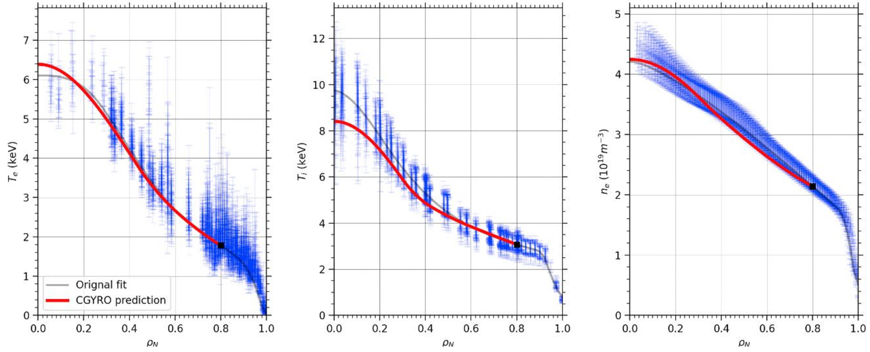


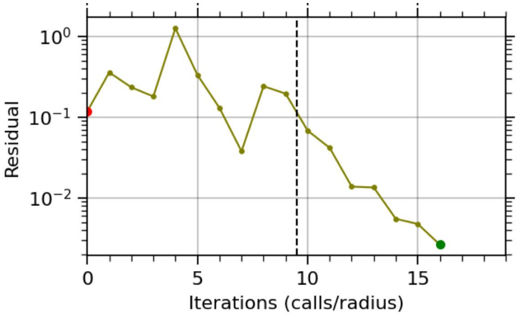
Standard BO workflow to find steady-state profile solutions



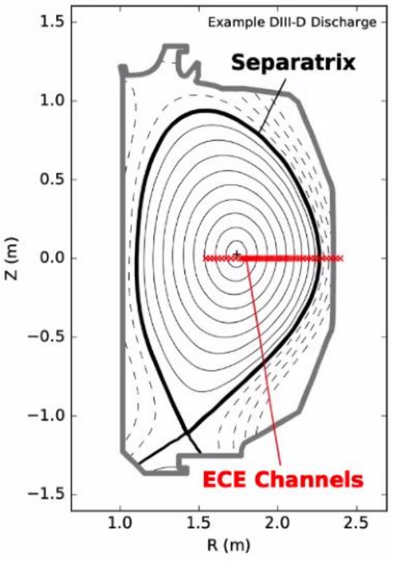
P. Rodriguez-Fernandez et al, Nucl. Fusion 62 (2022) 076036

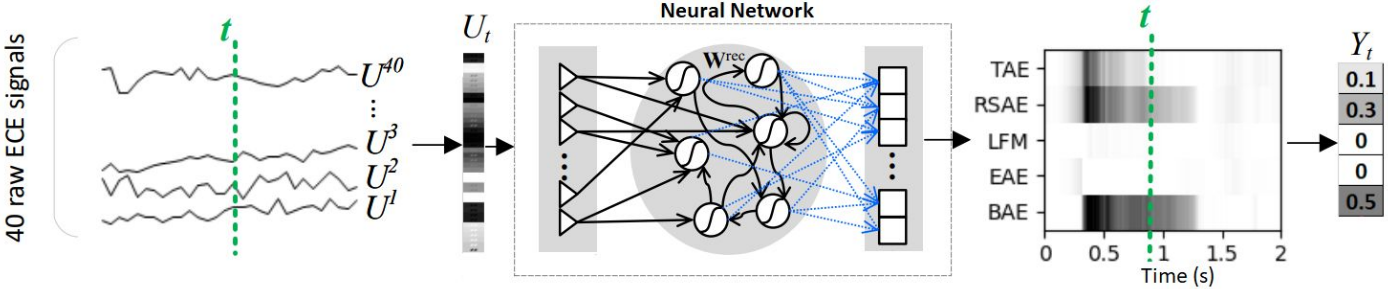
Open question: detangle local minima from unique physical solutions

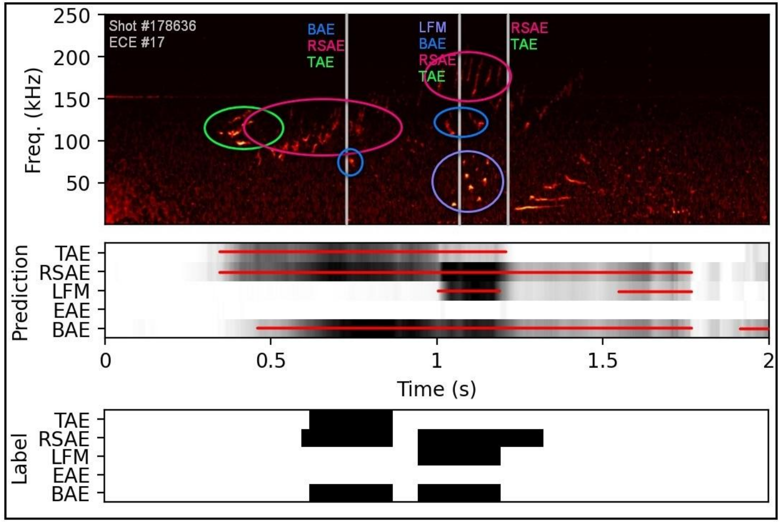




Raw ECE time series input data to Reservoir Computing Network to compute Alfvén Eigenmode score







Jalalvand et al 2022 Nucl. Fusion 62 026007

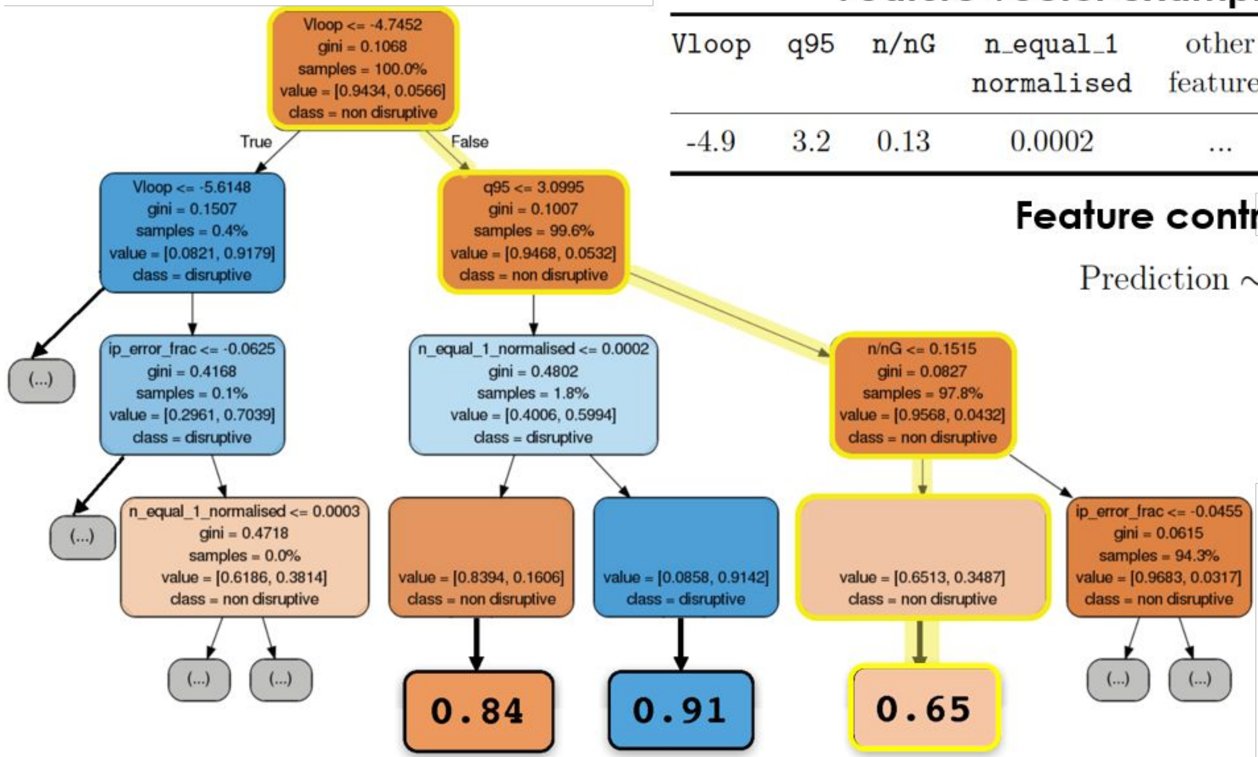
True Positive Rate: %91
False Positive Rate: %7

caution: how sensor failures affect ML workflow accuracy

Decision paths in RF trees provide local measures of explainability through information gain and loss

https://github.com/andosa/treeinterpreter, A. Saabas,
A. Palczewska et al., Integration of Reusable Systems (2014).

Fusion



Feature vector example for inference

Vloop	q95	n/nG	n_equal_1 normalised	other features	prediction
-4.9	3.2	0.13	0.0002	...	~ 0.65

Feature contribution breakdown

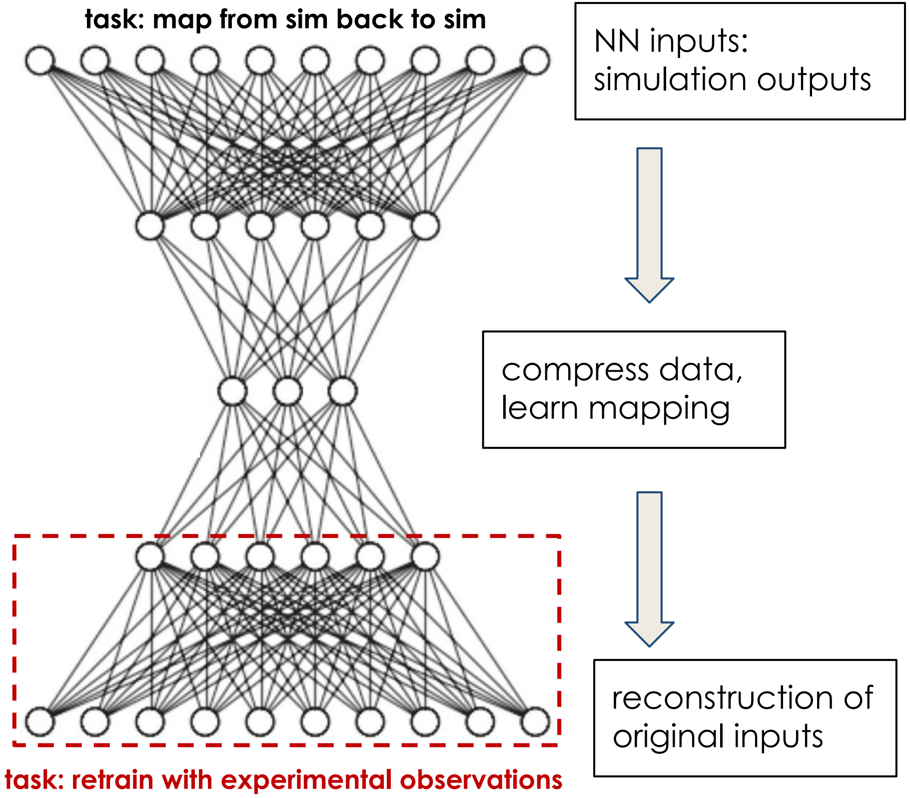
$$
\begin{aligned}
\text{Prediction} \sim 0.651 \sim\ & 0.943 \quad \text{(bias term)} \\
& + 0.003 \quad (\text{gain, } \texttt{Vloop}) \\
& + 0.010 \quad (\text{gain, } \texttt{q95}) \\
& - 0.305 \quad (\text{loss, } \texttt{n/nG})
\end{aligned}
$$

contributions, or information gain (+) and loss (-)

Predictions for forest of *M* trees can be **decomposed** in the *K* **contributions** from each evaluated input feature:

$$F(x) = \frac{1}{M}\sum_{m=1}^{M} \text{bias}_m + \sum_{k=1}^{K}\left(\frac{1}{M}\sum_{m=1}^{M} \text{contrib}_m(x,k)\right)$$

Source domain (simulations) allows to learn how to reconstruct target data (experiments)



Fusion

- **Large datasets** built through inexpensive but possibly inaccurate **simulations**.
- Networks (autoencoders) trained to learn **mapping** between **sim-to-sim** inputs to outputs.
- Mapping then transferred to new task → **learning corrective transformation mapping sim-to-exp** → **transfer learning**!

Adapted from Humbird et al., PoP 28, 042709 2021

Transfer learning

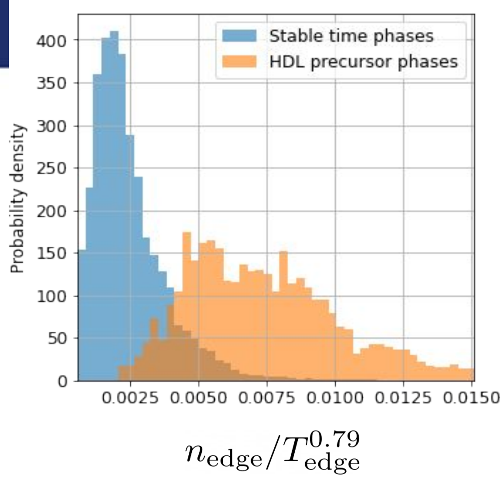
Improving instability avoidance with data-driven analysis of the "density limit"

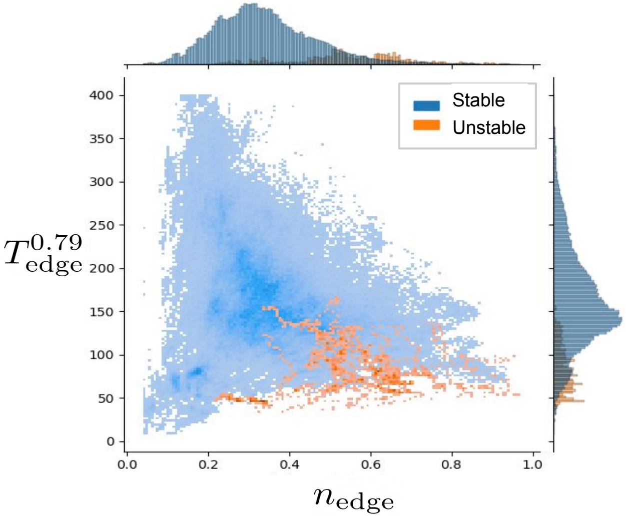
- Density limit critical for ITER, but standard "Greenwald Limit" not sufficiently reliable
- Our study identified novel equation

$$\frac{n_{\text{edge}}}{T_{\text{edge}}^{0.79}} = 3 \cdot 10^{-3}$$

outperforms Greenwald, other benchmarks, in predicting density limit at DIII-D

- Expanding to multi-machine database





DIII-D NATIONAL FUSION FACILITY

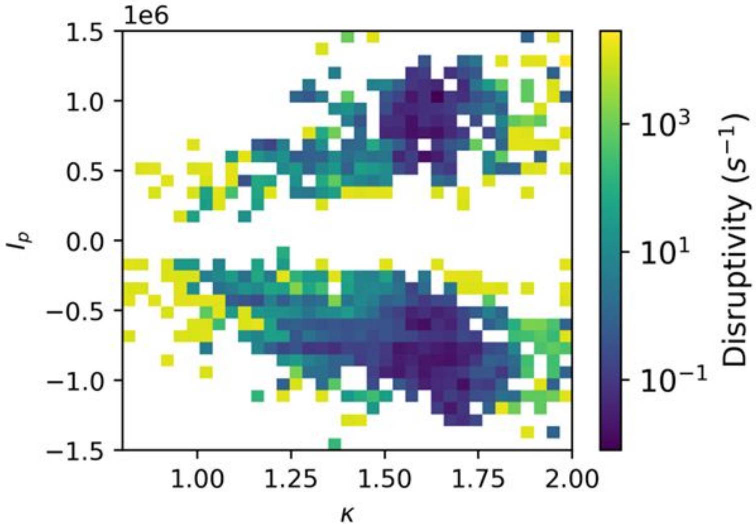
Andrew Maris, maris@mit.edu

Disruptions modeled as Poisson processes with characteristic time 1/disruptivity (1/s):

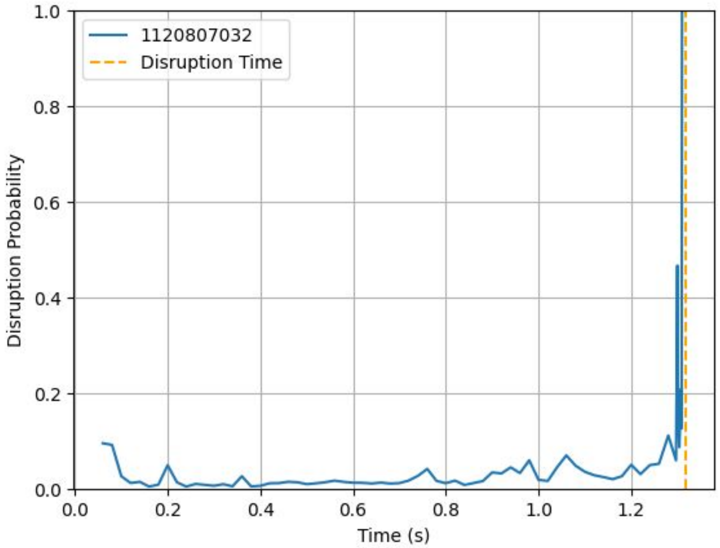
$$P(k \text{ disruptions in } \Delta t|d) = \frac{(d\Delta t)^k e^{-d\Delta t}}{k!} \quad \Longrightarrow \quad P_D(\Delta t|d) = 1 - e^{-d\Delta t}$$

probability of disruption in the next *Δt (s)*

P. Kaloyannis (EPFL), C. Rea, 2023 Master's thesis



Pulse planning parameter space optimization.



Real time nonlinear boundary avoidance and disruption prediction.