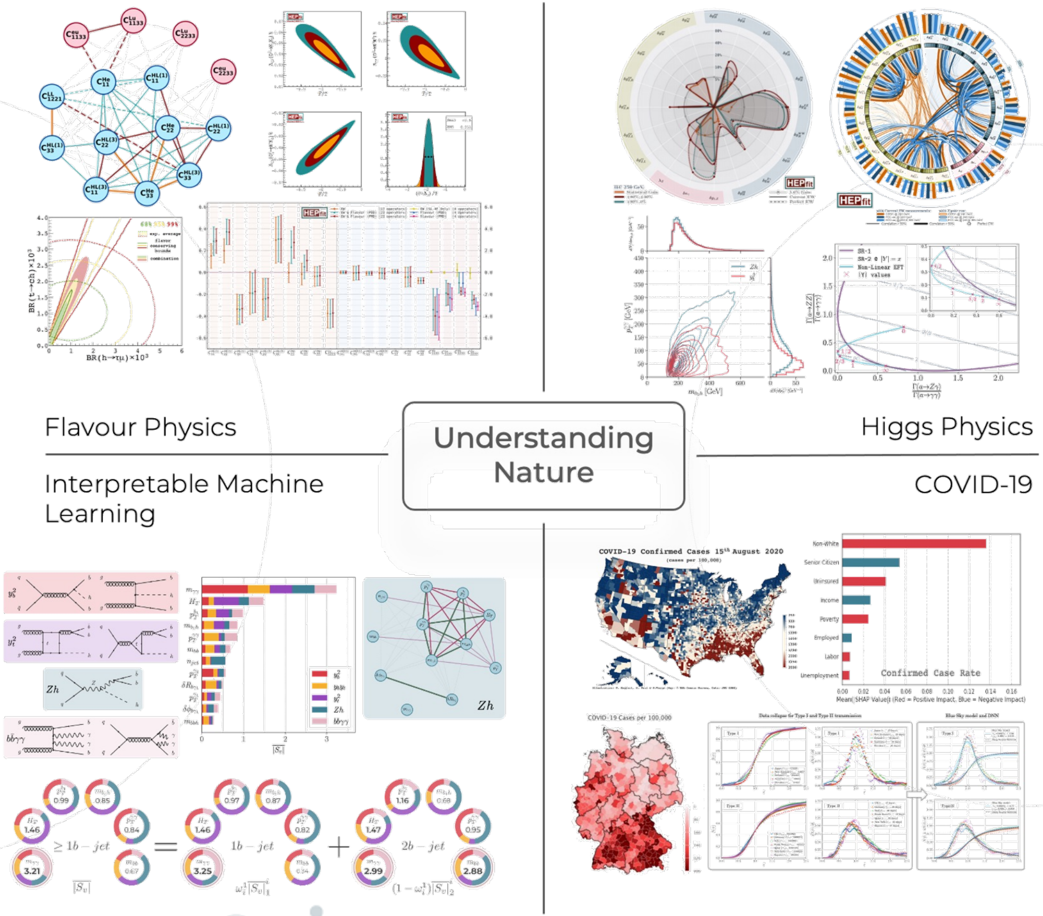
Machine Learning and Physics: the alliance of the Titans

Ayan Paul
The Institute of Experiential AI
Electrical and Computer Engineering, Northeastern University
CDNM, Brigham & Women's Hospital, Harvard Medical School

34th Rencontres de Blois on "Particle Physics and Cosmology" — May 19th 2023

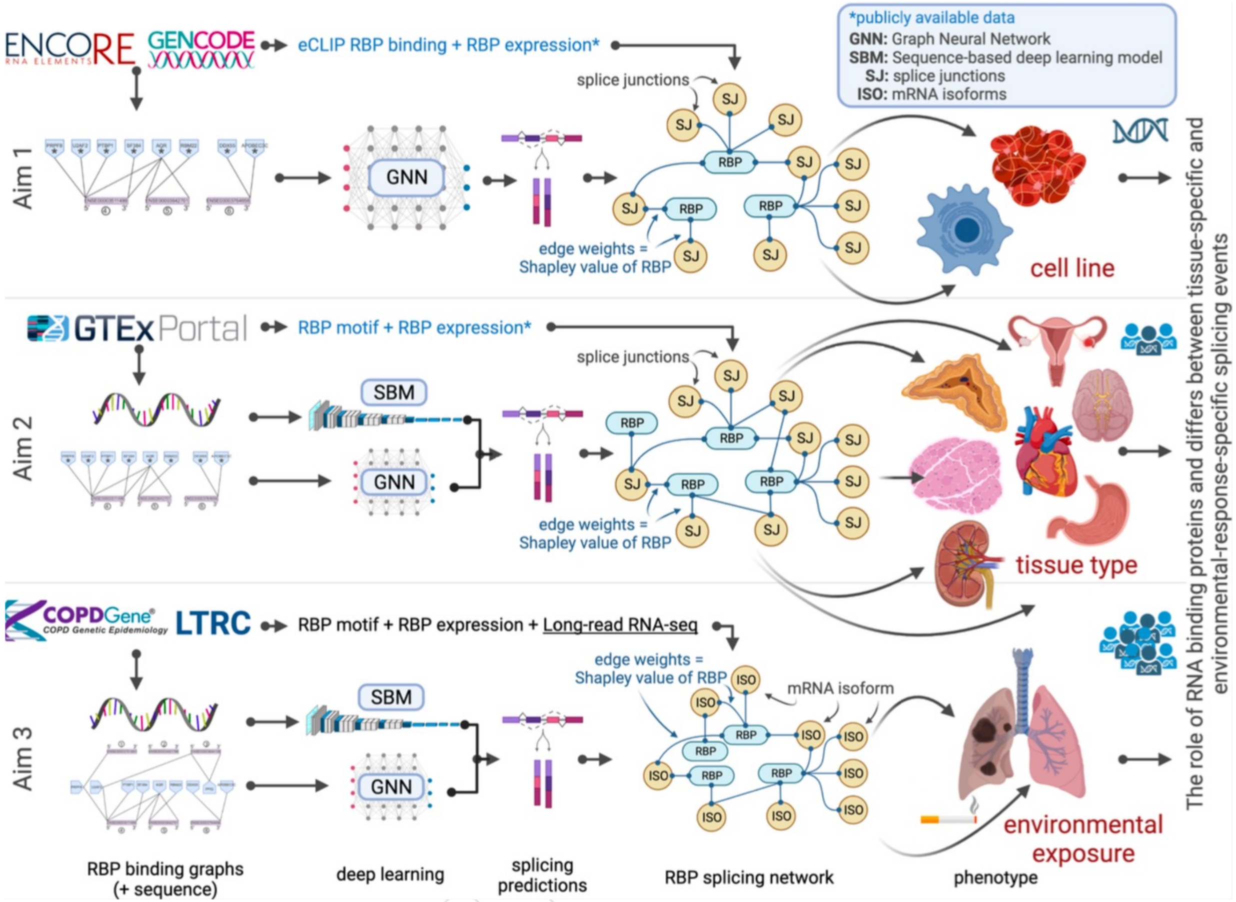
- PhD, Theoretical Particle Physics, 2012
    - University of Notre Dame du Lac, USA
- Postdoctoral Fellow, 2012 – 2017
    - INFN, Sezione di Roma, Italia
- Fellow, 2017 – 2022
    - Deutsches Elektronen-Synchrotron, Hamburg
- Senior Scientist, 2017 – 2022
    - Humboldt Universität zu Berlin
- Research Scientist, 2022 – Present
    - Northeastern University, Boston
    - Harvard Medical School, Boston
- Chief Scientific Officer, 2020 – 2022
    - Covis Inc, USA & Germany

PI – DESY Strategy Fund for Corona 2020
PI – VolkswagenStiftung Corona Response 2020



Application of Machine Learning to RNA Biology (genetics) for developing therapeutics to slow down or reverse terminal diseases.

So that we can live long enough to see the results from the future colliders.

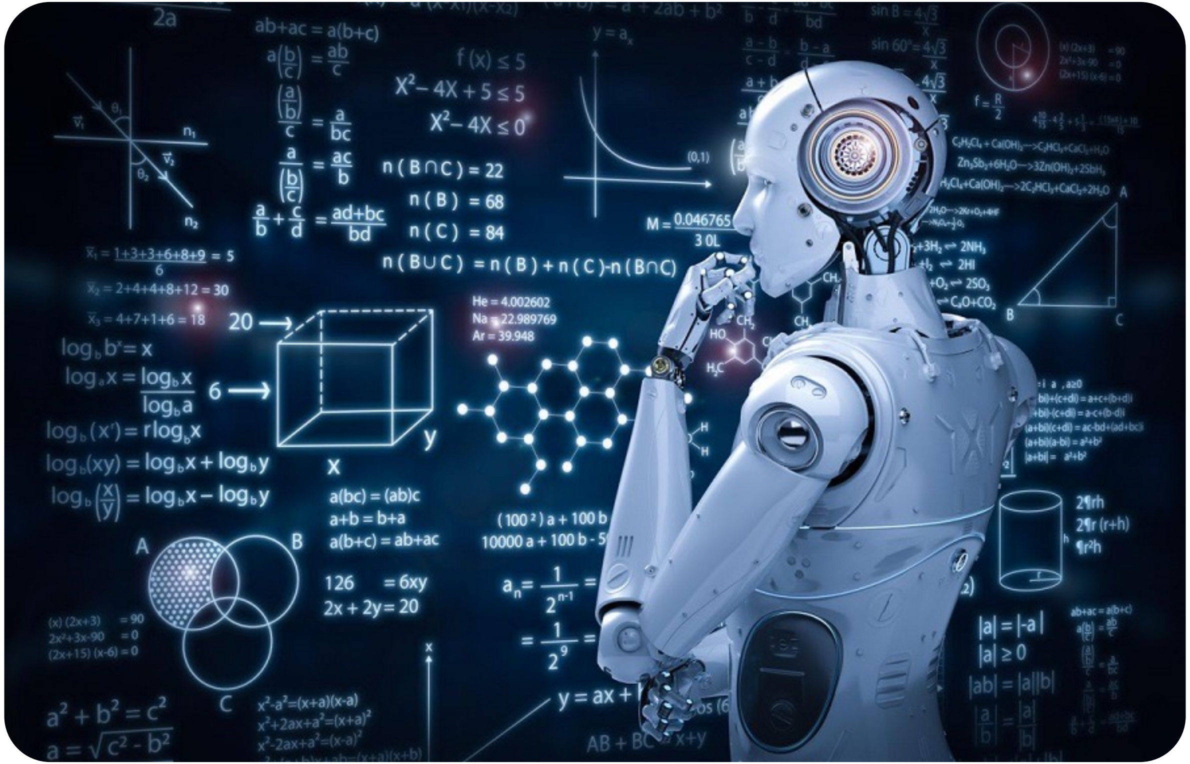


There is no Artificial Intelligence, yet.

Examples of Intelligence

Example of Sporadic Intelligence

Only in movies and books



The rest is glorified regression,
a.k.a. Machine Learning.

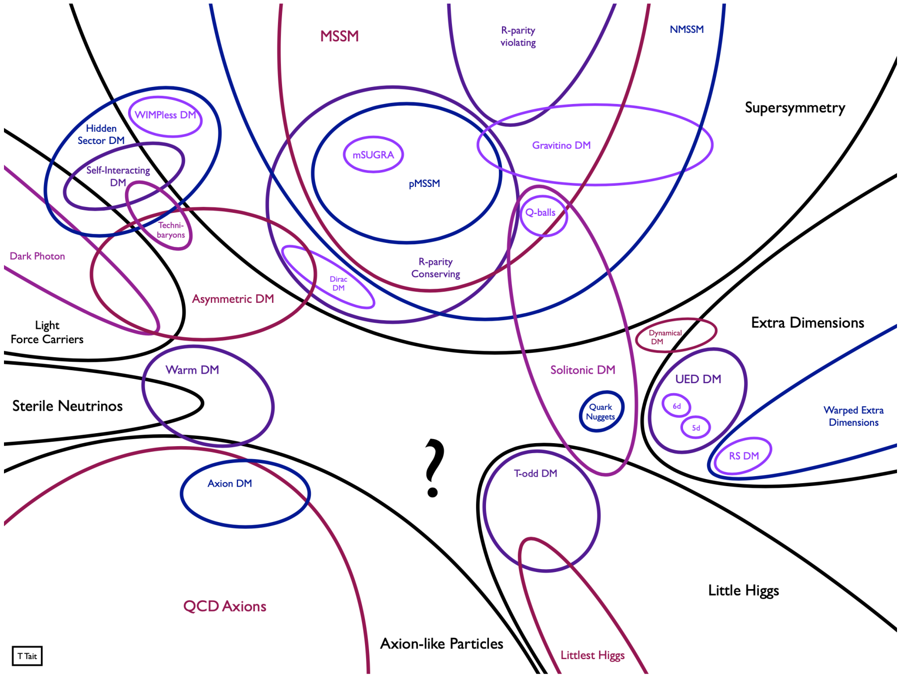
Does Physics need Machine Learning?

Warning: when you have only a hammer, everything looks like a nail.

Thanks Tim!

The sea of BSM theories:

- None have been realized at a few TeV
- No signatures of BSM have been found with either high energy or low energy observables
- There clearly seems to be no naturalness problem
- Nature does not mind and possibly advocates a large mass gap between EW and NP scales
- At the end of the day, regardless of what theories we conjure, data is the final judge



A roadmap for the future:

- We will have more data but incrementally better in terms of statistics (HL-LHC)
- We can have more data during the second half of the century but, really, do we have the patience or lifespan to wait?
- How do we make better use of data?
- How do we make better use of advanced statistical methods?
- Machine Learning (ML) is the bleeding edge of statistical learning algorithms. Can we leverage it effectively?

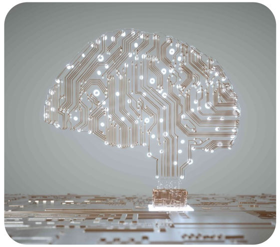
Is Machine Learning the new math that we need to find new physics

What is Machine Learning?

and can it be useful in understanding the laws of nature?

Some History: 1950 – 1960

really smart! → Can they think like us? →



If all the knowledge of the world could be converted to rules/laws
then, a computer can answer any question!

Sounds familiar?
hint: The Theory of Everything
Instead, lets call it the Rulebook of Everything or Artificial Intelligence

Turned out to be impossible: We are children collecting seashells when the vast ocean of knowledge lies before us unexplored

solution: learn from observations, the advent of data-driven statistical learning

Turing's requirement from an intelligent machine:

- **Natural Language Processing:** communicate in a human language
- **Knowledge Representation:** to store information it is exposed to
- **Automated Reasoning:** to use the stored information to answer questions or to draw new conclusions
- **Machine Learning:** to adopt to new circumstances and to detect and extrapolate patterns

Beyond Turing:

- **Computer Vision:** to perceive objects
- **Robotics:** to manipulate objects and move about

statistical learning has far outperformed rule-based algorithms in all these areas

Question: How many of these are useful in the kind of Physics we do?

Physics Goal: extract and process the underlying patterns in data to discover the laws of Nature

Question: How many of these are useful in the kind of Physics we do?

Turing's requirement from an intelligent machine:

- **Natural Language Processing:** sequence analysis, has not found a lot of applications
- **Knowledge Representation:** lower dimensional embedding, not often used
- **Automated Reasoning:** we wish!
- **Machine Learning:** has become the backbone of experimental analyses. A lot of it can be applied to phenomenology and theory

Beyond Turing:

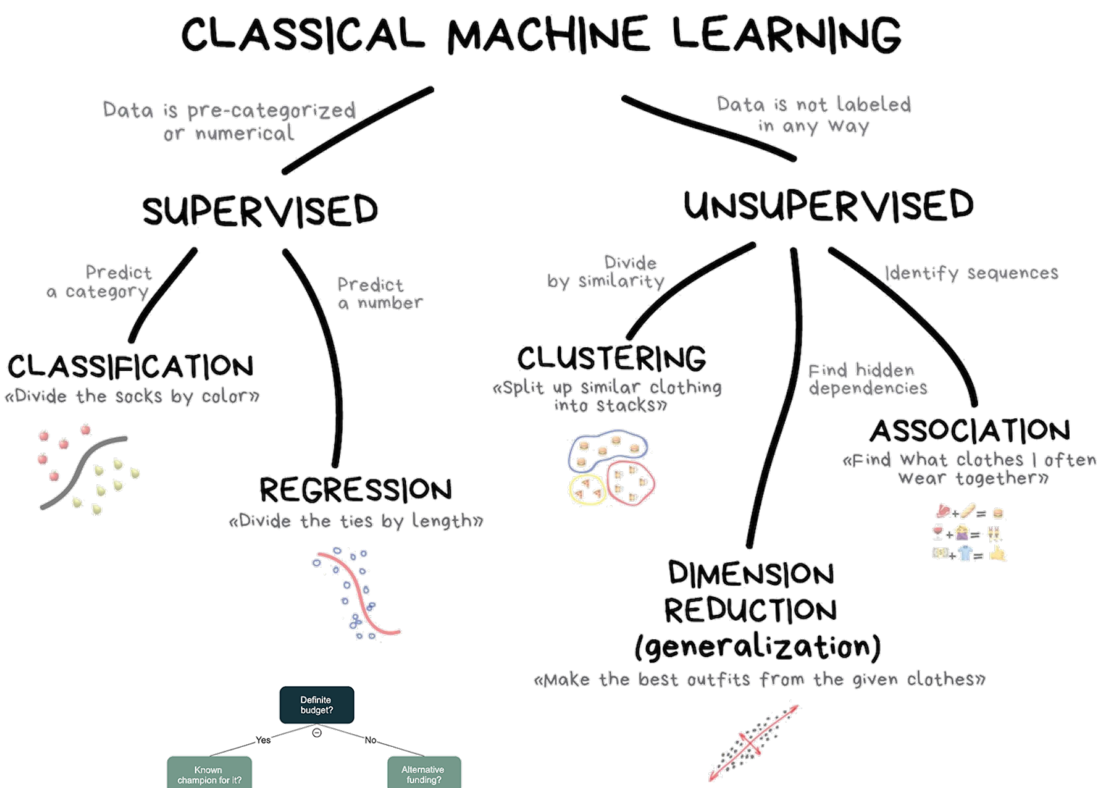
- **Computer Vision:** has found wide-spread application in experimental analyses
- **Robotics:** no, we still need to do our own jobs

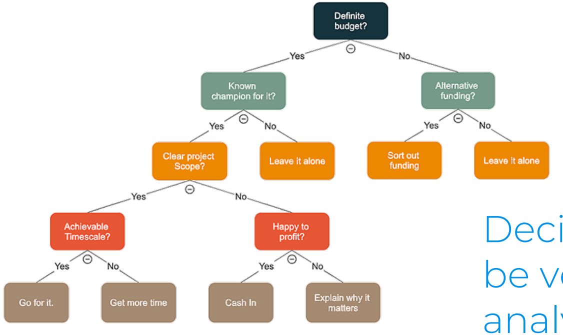
Physics Goal: extract and process the underlying patterns in data to discover the laws of Nature

Historically, classification and clustering have been used extensively in the analysis of experimental data

Classical methods are computationally less demanding and have contributed to almost all experimental discoveries in the past couple of decades

They are limited in their ability to map underlying patterns and quite limited in address several data types like images, sequences, non-Euclidean data, etc.

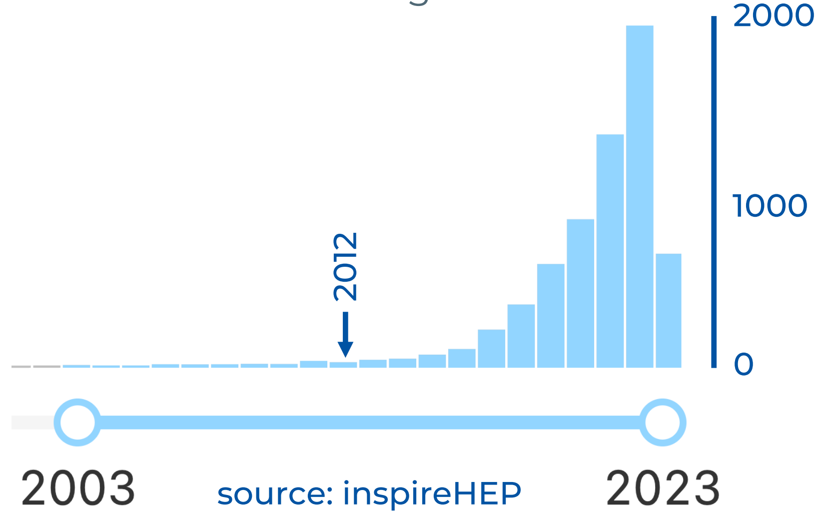




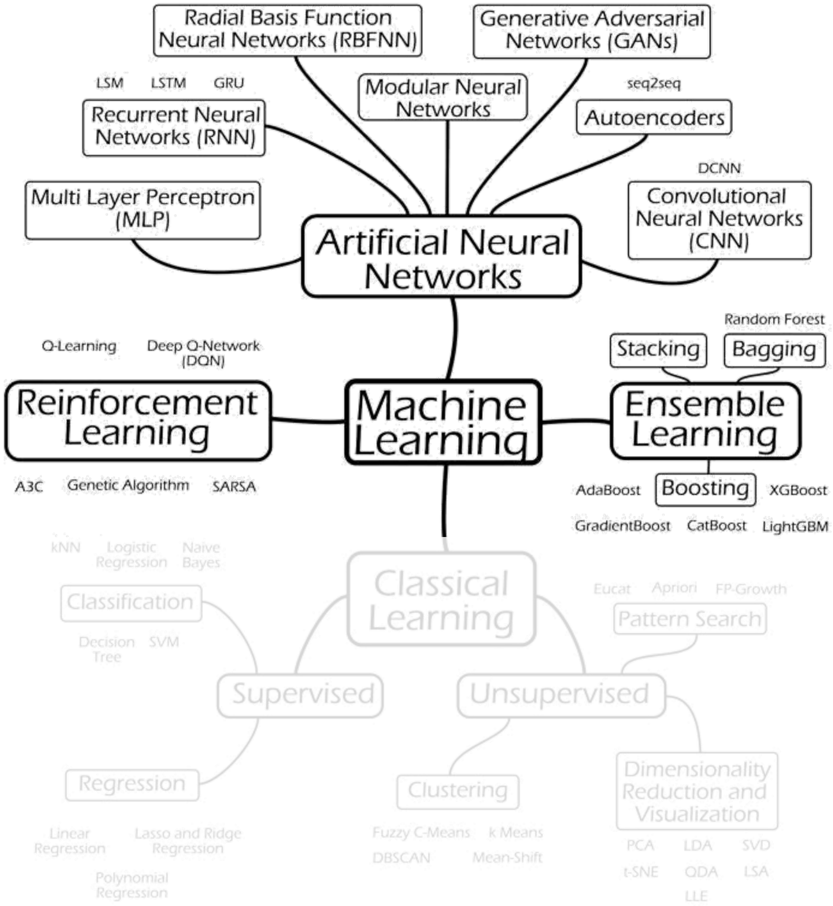
Decision trees turned out to be very useful for experimental analyses

The times they are a-changing

We are now in the age of the universal function approximators (Deep Learning) and ensemble learning

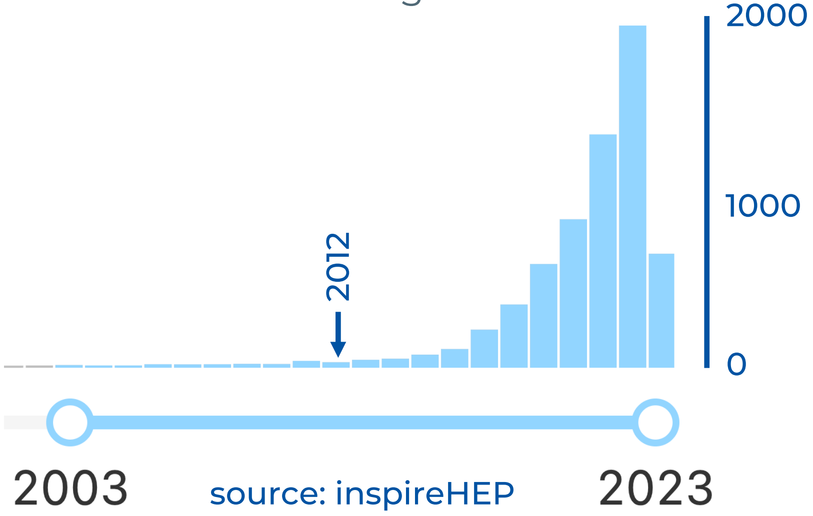


What happened in 2012?



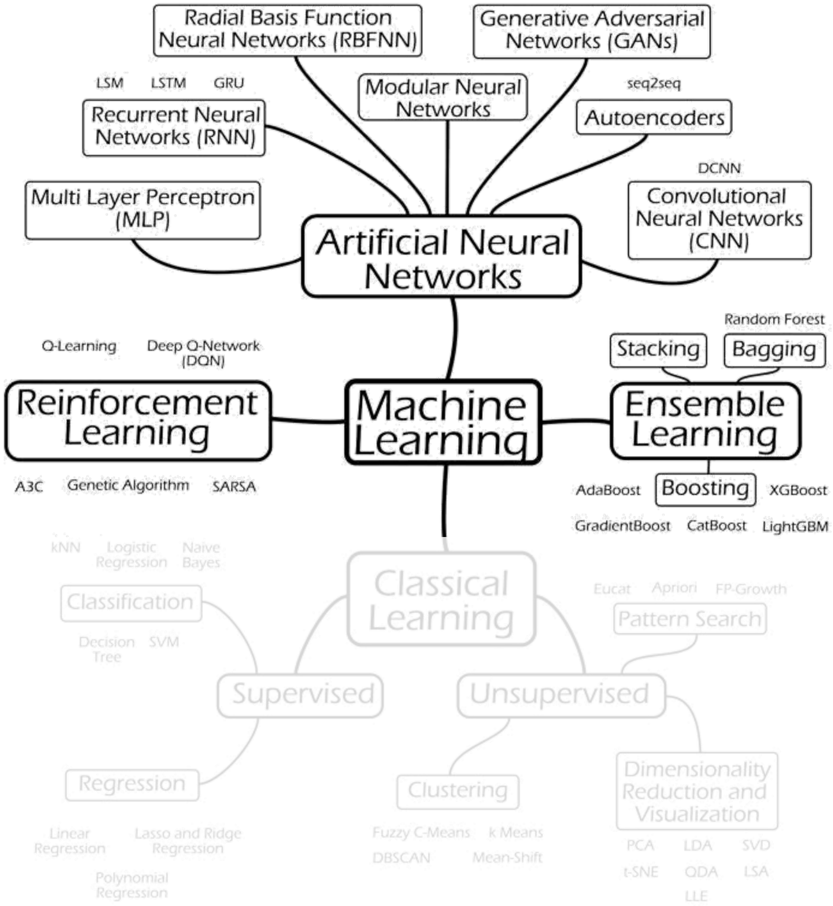
The times they are a-changing

We are now in the age of the universal function approximators (Deep Learning) and ensemble learning



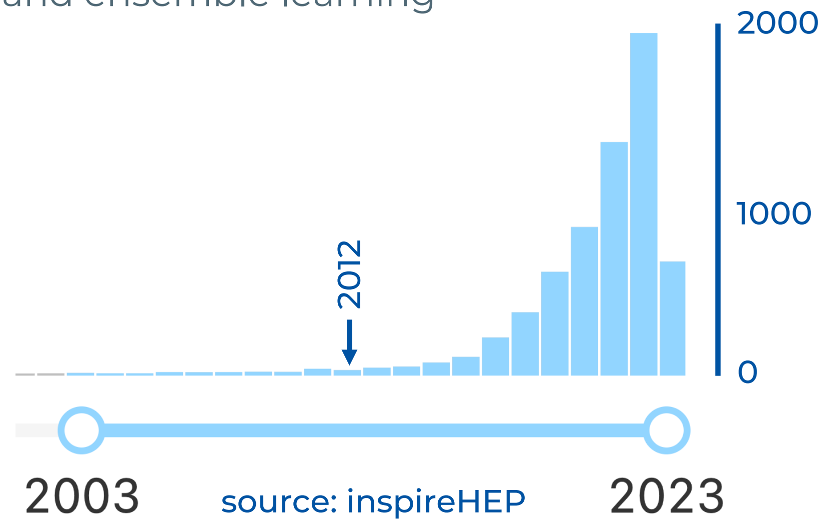
2012

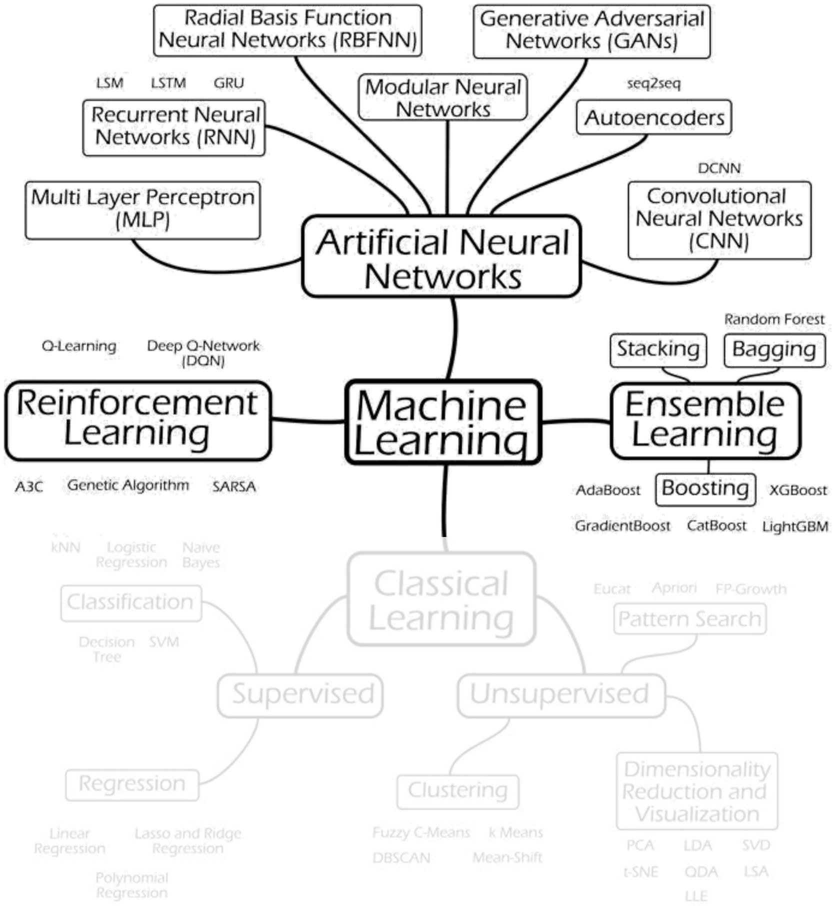
- Higgs discovered



The times they are a-changing

We are now in the age of the universal function approximators (Deep Learning) and ensemble learning



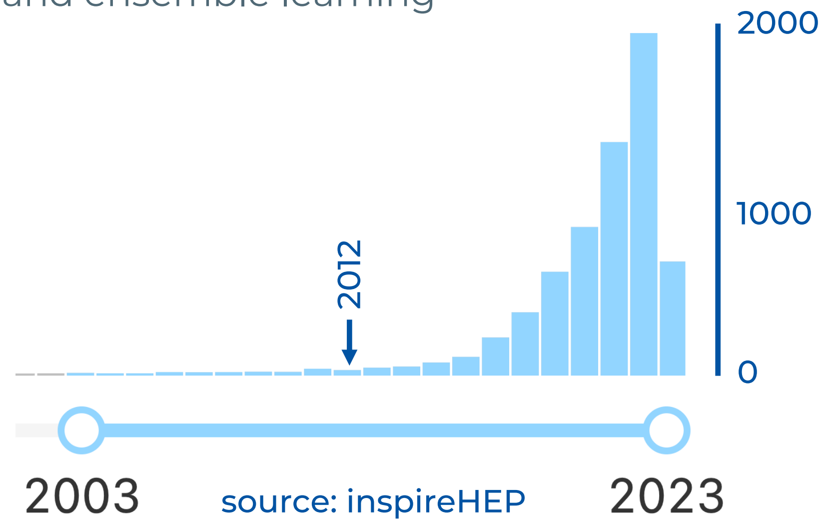


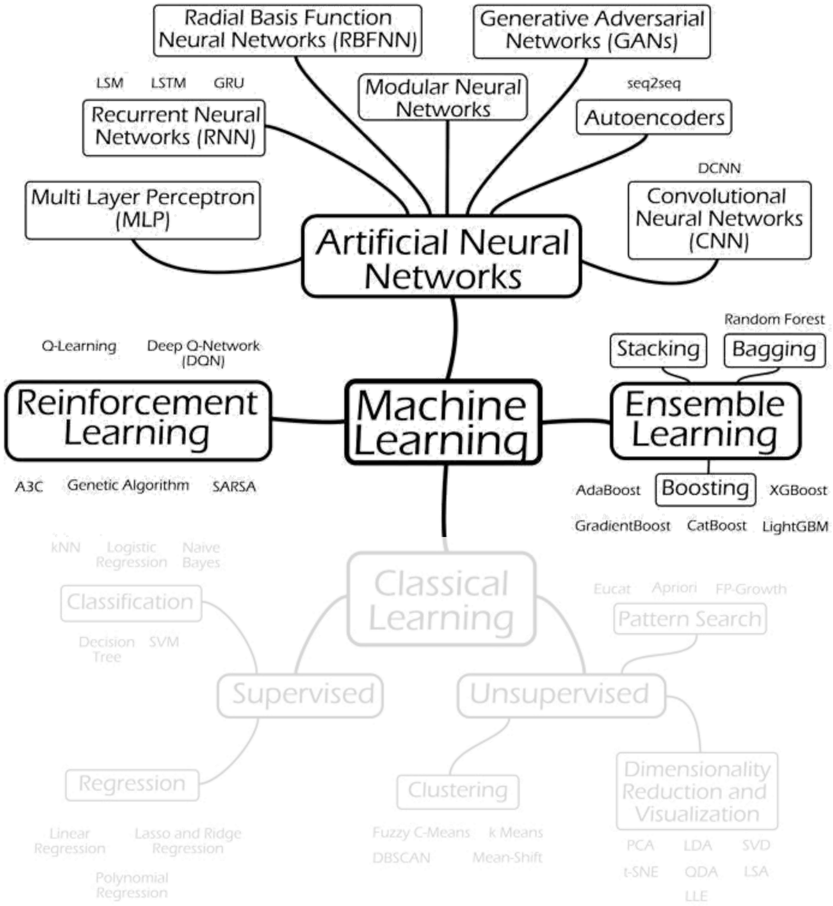
2012

- Higgs discovered
- Deep learning emerged as the leading statistical learning method

The times they are a-changing

We are now in the age of the universal function approximators (Deep Learning) and ensemble learning





2012

- Higgs discovered
- Deep learning emerged as the leading statistical learning method
- I completed my PhD degree.

So why did it take so long?

- Our labeled datasets were thousands of times too small.
- Our computers were millions of times too slow.
- We initialized the weights in a stupid way.
- We used the wrong type of non-linearity.

quoted from some of the giants of ML

```
Deep Learning =
Lots of training data + Parallel Computation + Scalable, smart algorithms
```

This summarizes what you need to build an analysis based on neural networks otherwise, use classical machine learning (like tree-based learners or clustering)

Can we do Physics with Machine Learning?

- Our labeled datasets were thousands of times too small.

— There are lots of data both experimental and synthetic (from simulations)
— There are methods that are not data hungry, but we rarely need them

- Our computers were millions of times too slow.

— Not anymore (but one needs to be not terrified of coding)
— Not all ML methods need large GPUs. A lot can be done on a laptop

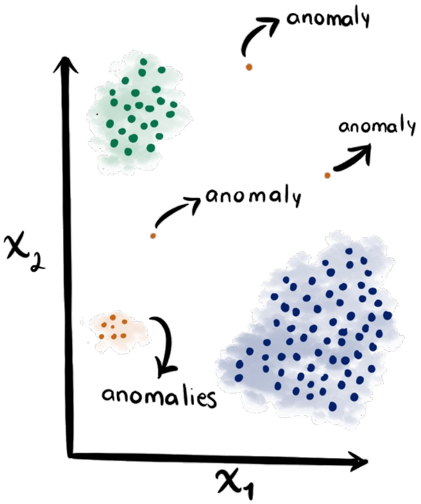
- We initialized the weights in a stupid way.

— Neural networks still depend on point estimation of the parameters and have a very large number of degenerate minima
— Neural networks also require a lot of tuning where all you can do is “hit and trial”

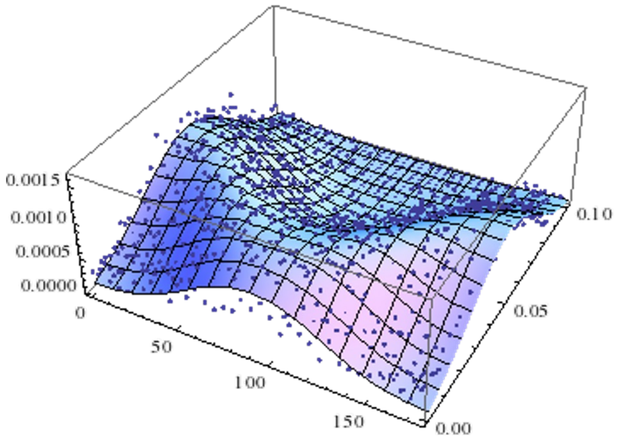
- We used the wrong type of non-linearity.

— This boils down to a choice of the basis function; something very familiar in physics
— Neural networks are nothing but nested functions whose coefficients (parameters) need to be estimated using likelihood minimization

Anomaly detection

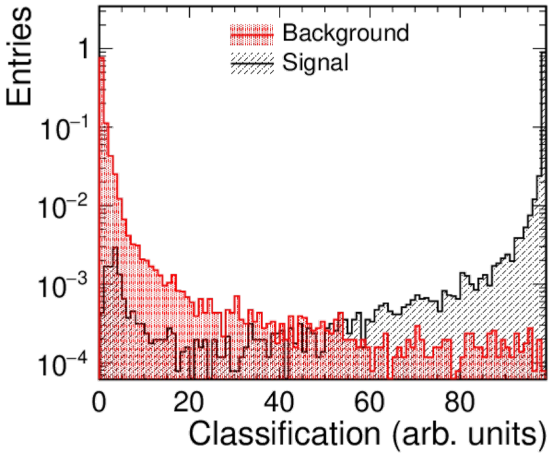


Regression

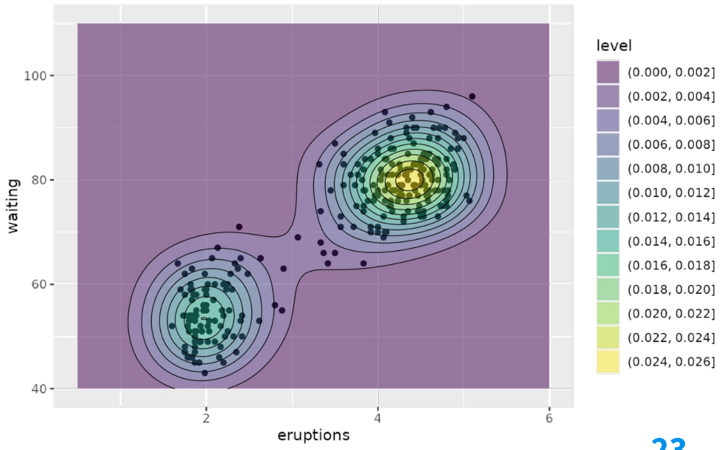


What can we do with machine learning?

Classification



Density estimation (sampling)



Two examples of Machine Learning

Introducing high-precision regressors and interpretable machine learning

High Precision Regressors for Monte Carlo Generators

Reducing the energy footprint of simulations for science

I do not think you can start with anything precise. You have to achieve such precision as you can, as you go along.

— Bertrand Russell

Based on:
F. Bishara, **A. Paul**, J. Dy, *High-Precision Regressors for Particle Physics*. Paper submitted to Nature Scientific Reports for peer-review.

F .Bishara, **A. Paul**, J. Dy, *Skip Connections for High Precision Regressors*. Machine Learning and the Physical Sciences, Workshop at the 36th Conference on Neural Information Processing Systems (NeurIPS 2022).

Can we speed up Monte Carlo event generators?

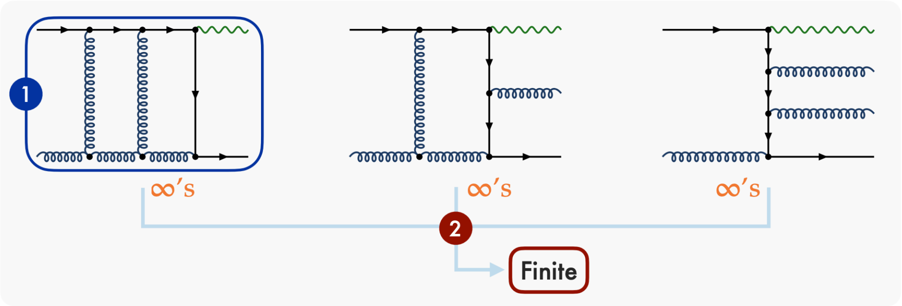
- Monte Carlo simulations of physics processes at particle colliders like the Large Hadron Collider at CERN take up a major fraction of the computational budget.
- A single data point can take seconds, minutes, or even hours to compute from first principles.
- $10^9 - 10^{12}$ data points are necessary per simulation; machine learning regressors can replace physics simulators to significantly reduce this computational burden.
- High-precision regressors are required that can deliver data with relative errors of less than 1% or even 0.1% over the entire domain of the function.
- **Goal: Significantly reduce the training and storage burden of Monte Carlo simulations at current and future collider experiments.**

Talk by Alexander Huss on May 16th

Main Challenges @ NNLO

$$\sigma \sim \int \mathrm{d}\Phi \, |\mathcal{M}|^2$$

cross section ~ phase space & scattering amplitude



complexity often quantified by the multiplicity ("#legs"): $2 \to n$

What if this part can be made much faster?

1 **amplitudes** & multi-loop integrals

- rapid growth in complexity with number of scales
  ⇝ kinematic invariants & particle masses (int./ext.)

2 **infrared subtractions** ↔ realistic setup (arbitrary cuts, observables, …)

- extract IR singularities in $\mathrm{d}\Phi$ without performing the integration
  ⇝ more difficult with more coloured legs (simpler if massive)

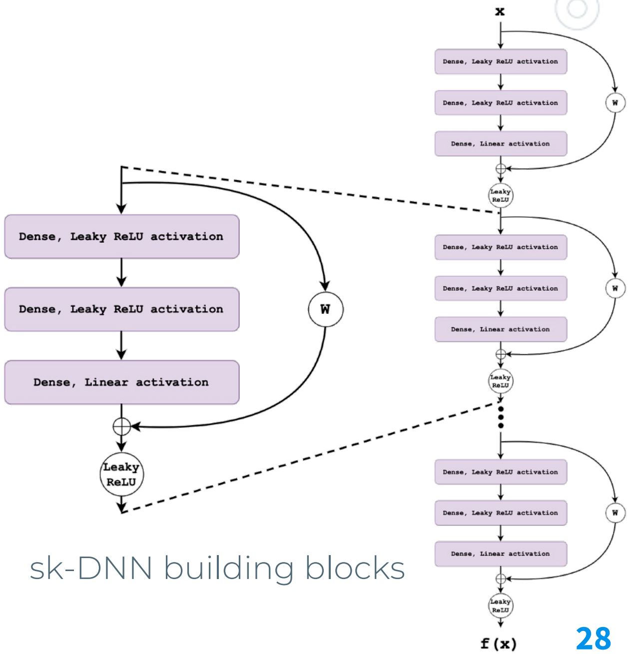
building surrogate regressors

$$F(\boldsymbol{x}) = f_{00}(\boldsymbol{x}) + \alpha\, f_{01}(\boldsymbol{x}) + \alpha^2 \underbrace{\{f_{11}(\boldsymbol{x}) + f_{02}(\boldsymbol{x})\}} + \ldots\,.$$

Needs a surrogate regressor since it is very slow to compute from first principles

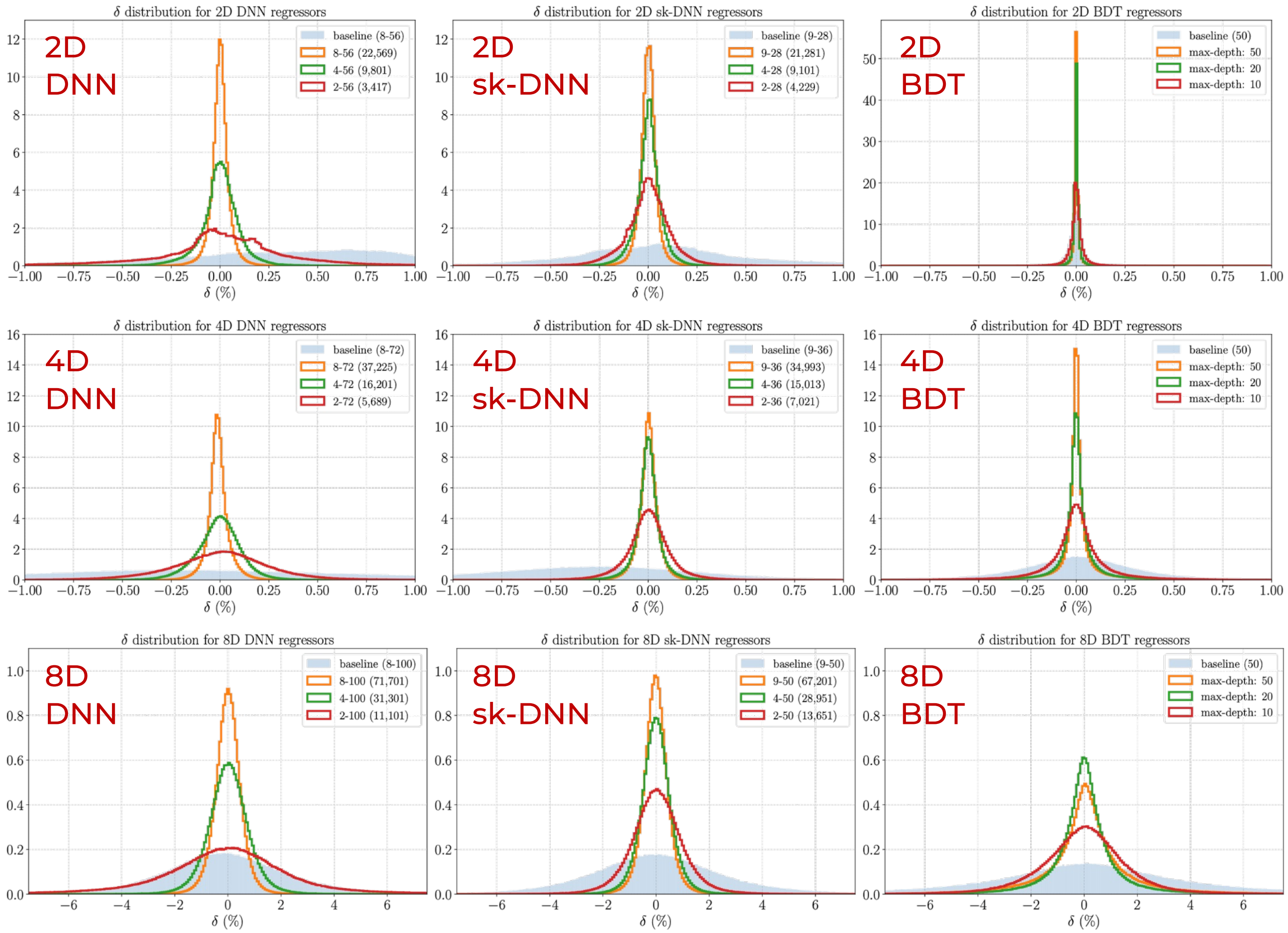
- Train boosted decision trees (BDT), Deep Neural Networks (DNN) and Deep Neural Networks with skip connections (sk-DNN) with simulated data from particle physics.
- We use 2-, 4-, and 8-dimensional (D) data to compare BDTs, DNN and sk-DNNs with the aim of reaching
- Aim: errors < **1% – 0.1%** over at least **90%** of the input feature space.
- Architectural decisions, training strategies and data volume.

A lot of generative models have been proposed but struggle to achieve even ~10% errors even with lower dimension kinematic space



sk-DNN building blocks

results



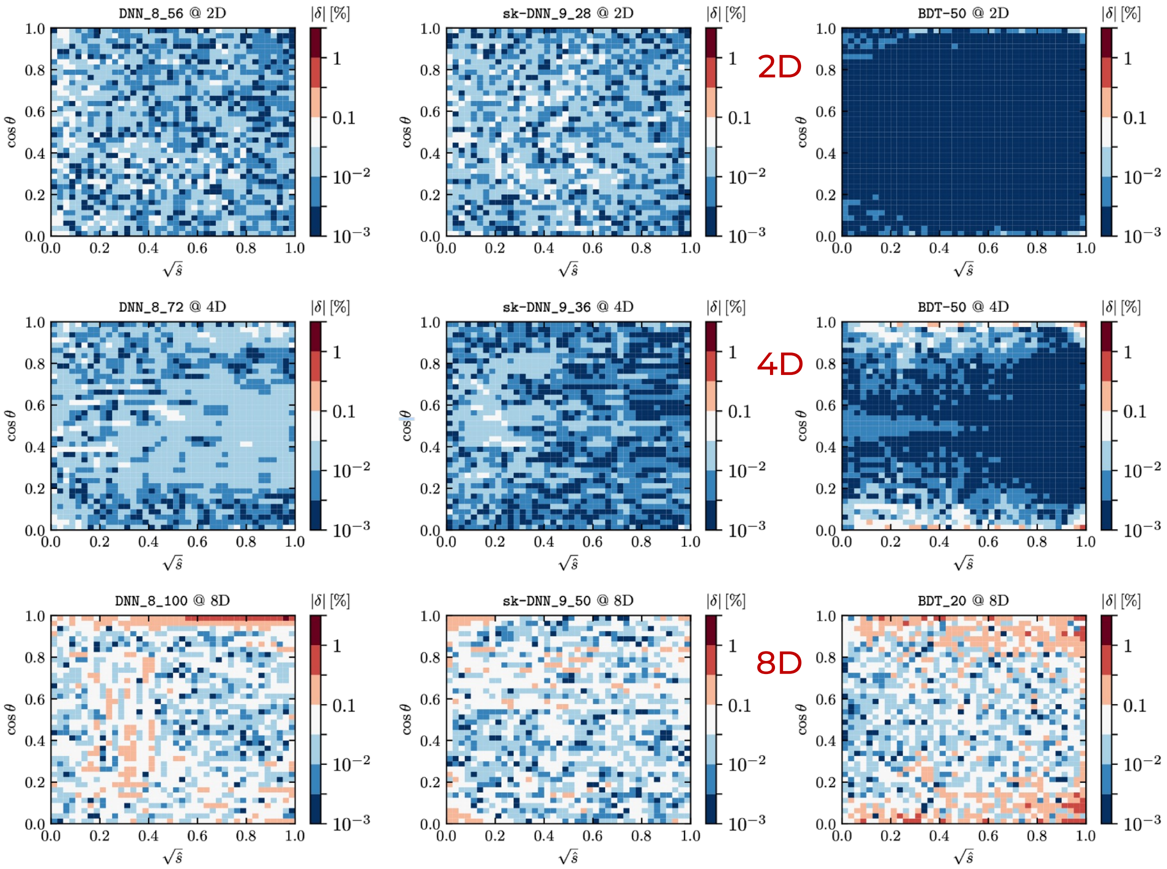
error quantification:

$$\delta = \frac{f(\boldsymbol{x})_{predicted} - f(\boldsymbol{x})_{true}}{f(\boldsymbol{x})_{true}}$$

The precision was achieved using:

- Symmetry properties of the physics process
- Exponential cooling of learning rate
- Physics informed normalization

summary



- BDTs outperform DNNs at lower dimensions.
- Fully-connected DNNs perform better at higher dimensions.
- sk-DNNs outperforms both BDTs and DNNs at 4D and 8D.
- sk-DNNs can outperform DNNs with a larger number of parameters.
- The regressors can provide precise predictions in 10^{-3} – 10^{-6} seconds << few seconds taken by MC simulation.

Code: https://github.com/talismanbrandi/high-precision-ml

Interpretable Machine Learning for Higgs Signals

Enhancing the power of particle collider searches

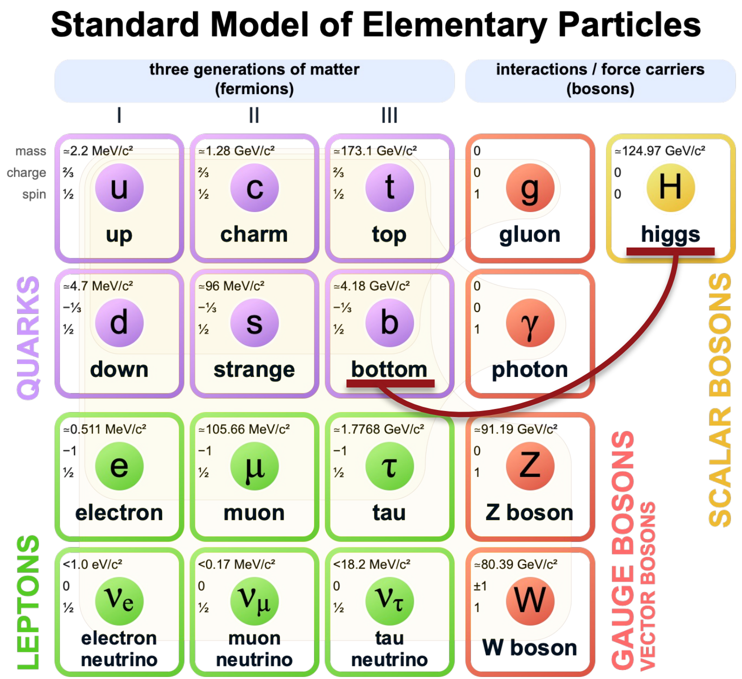
If you can't explain it simply, you don't understand it well enough.
– Albert Einstein

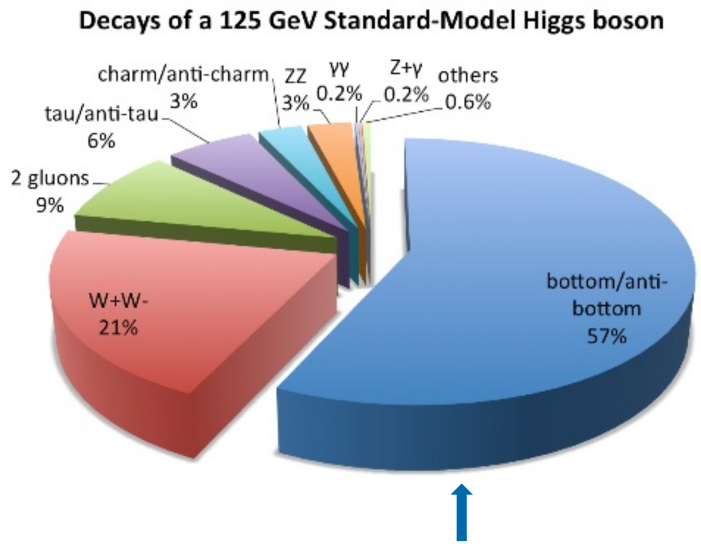
Based on:
C. Grojean, **A. Paul**, and Z. Qian, *Resurrecting bbh with kinematic shapes*. JHEP**04** (2021) 139. DOI: 10.1007/JHEP04(2021)139

Related to:
C. Grojean, **A. Paul**, and Z. Qian, I. Strümke, *Lessons on interpretable machine learning from particle physics*. Nature Review Physics **4**, 284–286 (2022). DOI: 10.1038/s42254-022-00456-0

L. Alasfar, R. Gröber, C. Grojean, **A. Paul**, and Z. Qian, *Machine learning the trilinear and light-quark Yukawa couplings from Higgs pair kinematic shapes*. JHEP**11** (2022) 045. DOI: 10.1007/JHEP11(2022)045

Higgs couplings and $b\bar{b}h$ with $h \to \gamma\gamma$

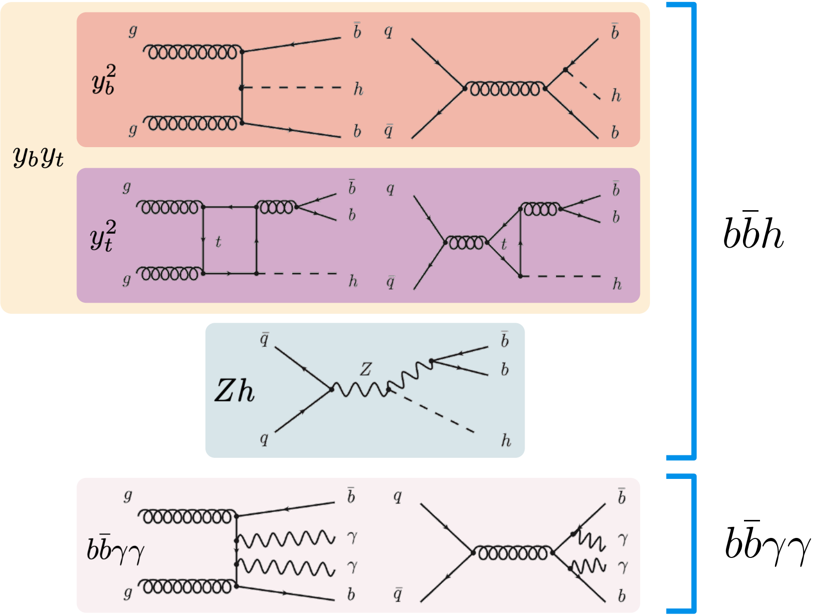




This is what we are looking for.

It took 40 years to find the Higgs.
So, measuring one of its decay modes precisely is a very challenging task!

tiny signals embedded in multiple large backgrounds



Traditional cut-based analysis cannot separate the different $b\bar{b}h$ contributions – no y_b sensitivity at HL-LHC

signals

Channel	LO σ (fb)	NLO-k-fact	6 ab^{-1}[#evt]	$2b$-jets[%]
y_b^2	0.0648	1.5	583	7.7%
$y_b y_t$	-0.00829	1.9	-95	4.0%
y_t^2	0.123	2.5	1,840	12%
Zh	0.0827	1.3	645	21%
$\sum b\bar{b}h$	0.262	-	2,970	-
$b\bar{b}\gamma\gamma$	12.9	1.5	116,000	14%

$b\bar{b}h$ background

$b\bar{b}\gamma\gamma$ background

Signal to background ratio ~ 1:250

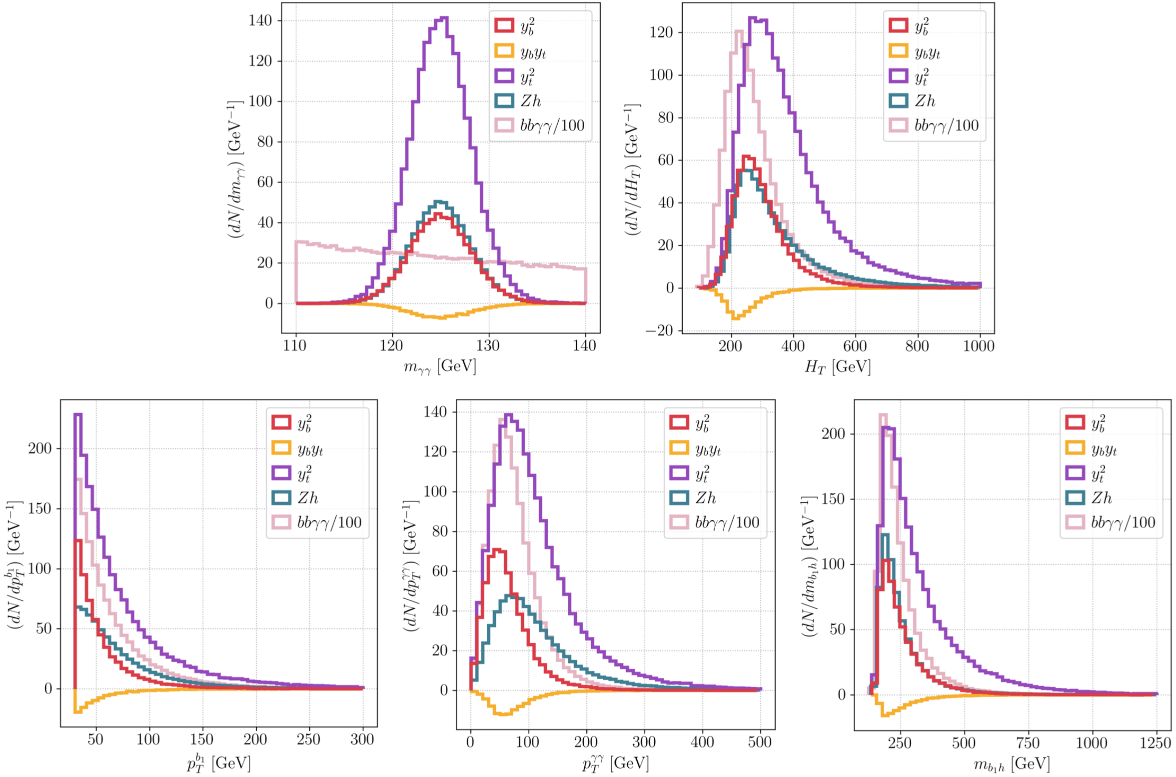
building an interpretable framework

Understanding differences in shapes

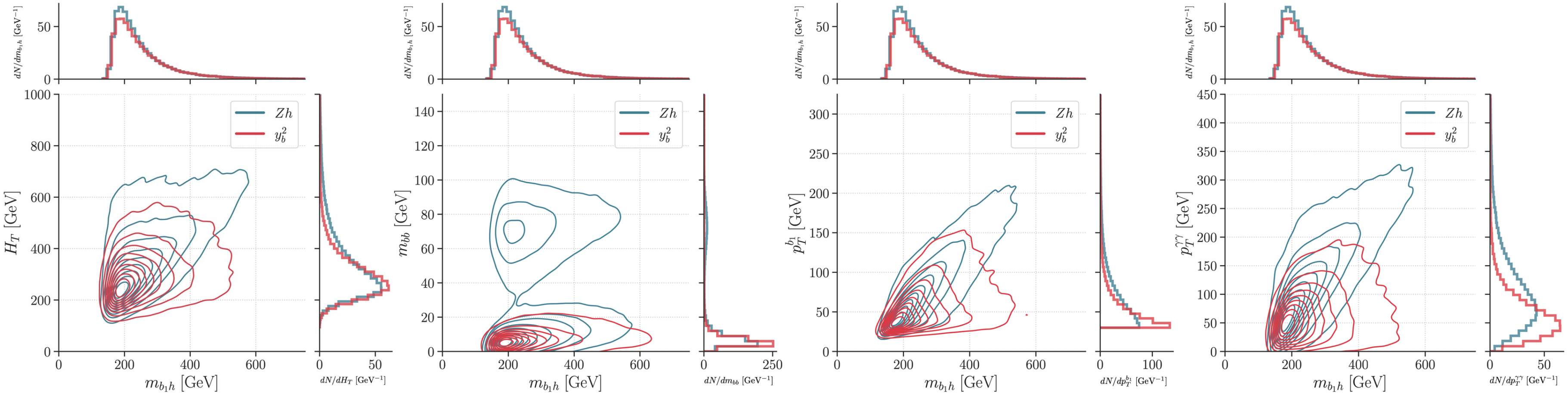
- $p_T^{b_1},\ p_T^{b_2},\ p_T^{\gamma_1},\ p_T^{\gamma\gamma},$
- $\eta_{b_{j1}},\ \eta_{b_{j2}},\ \eta_{\gamma_1},\ \eta_{\gamma\gamma},$
- $n_{bjet},\ n_{jet},\ \Delta R^{b\gamma}_{\min},\ \Delta\phi^{bb}_{\min},$
- $m_{\gamma\gamma},\ m_{bb},\ m_{b_1 h},\ m_{b\bar{b}h},\ H_T.$

The choice of variables is important:

- Momenta four vectors are not easily interpretable
- Kinematic variables are interpretable but there is no clear “complete set”



the devil is in the correlation

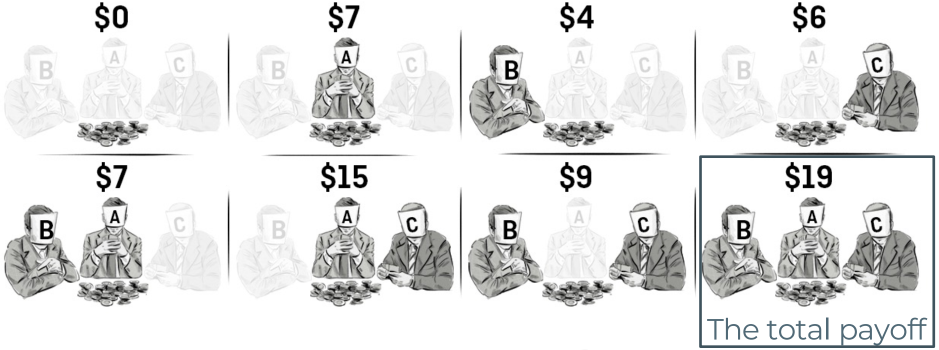


- Cut-based analyses start to falter with multivariate correlations – difficult to visualize and interpret
- Machine learning algorithms excel at multivariate analyses

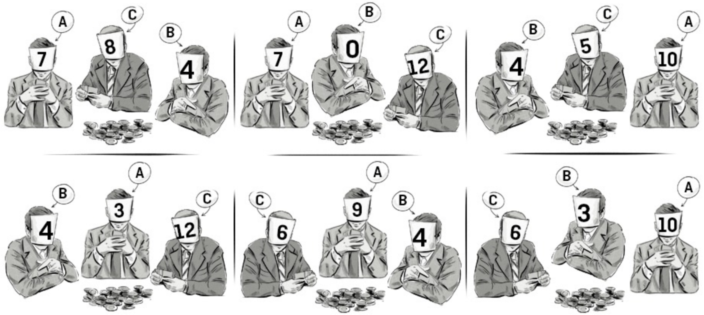
- Machine learning algorithms are essentially black-boxes – not good for understanding the underlying dynamics

a cooperative game

The value of each player and each combination of players



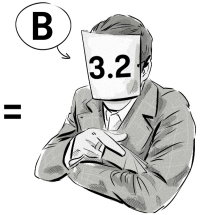
The value of the player in each game



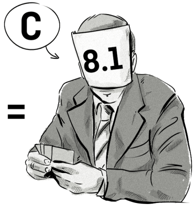
Marginalized values

(7+7+10+3+9+10) / 6 =

(4+0+4+4+4+3) / 6 =



(8+12+5+12+6+6) / 6 =



The most important player

L. S. Shapley, *Notes on the n-Person Game-II: The Value of an n-Person Game* (1951).

some useful properties of Shapley values

For a game $\mathcal{G} = (\mathcal{K}, \nu)$ with a set $\mathcal{K}$ of players and a payoff ν:

Dummy Player: A player that doesn't contribute to any subset of players must receive zero attribution

$$\phi_k(\nu) = 0.$$

Efficiency: Attributions must add to the total gain

$$\sum_{k \in \mathcal{K}} \phi_k(\nu) = \nu\left(\mathcal{K}\right).$$

Symmetry: Symmetric players must receive equal attribution

$$\nu(\mathcal{A} \cup \{k\}) = \nu(\mathcal{A} \cup \{i\}) \implies \phi_k(\nu) = \phi_i(\nu).$$

Linearity: Attribution for the (weighted) sum of two games must be the same as the (weighted) sum of the attributions for each of the games

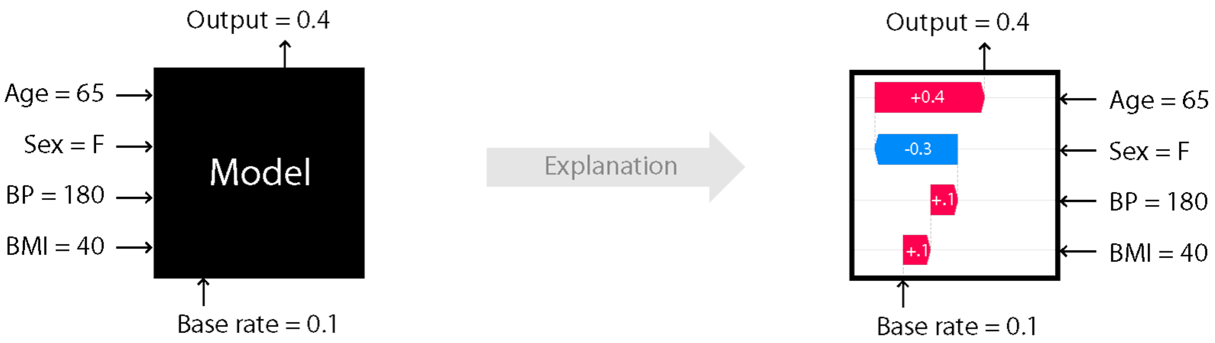
$$\phi_k\left(\nu + \omega\right) = \phi_k(\nu) + \phi_k(\omega) \qquad \forall\, k \in \mathcal{K}.$$

Shapley Additive exPlainers

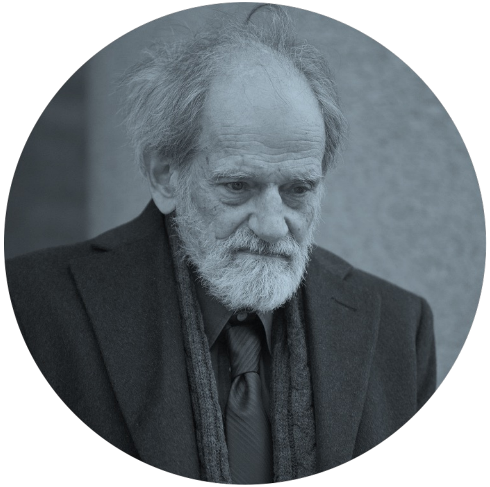
SHAP

Black-box model

Interpretable model



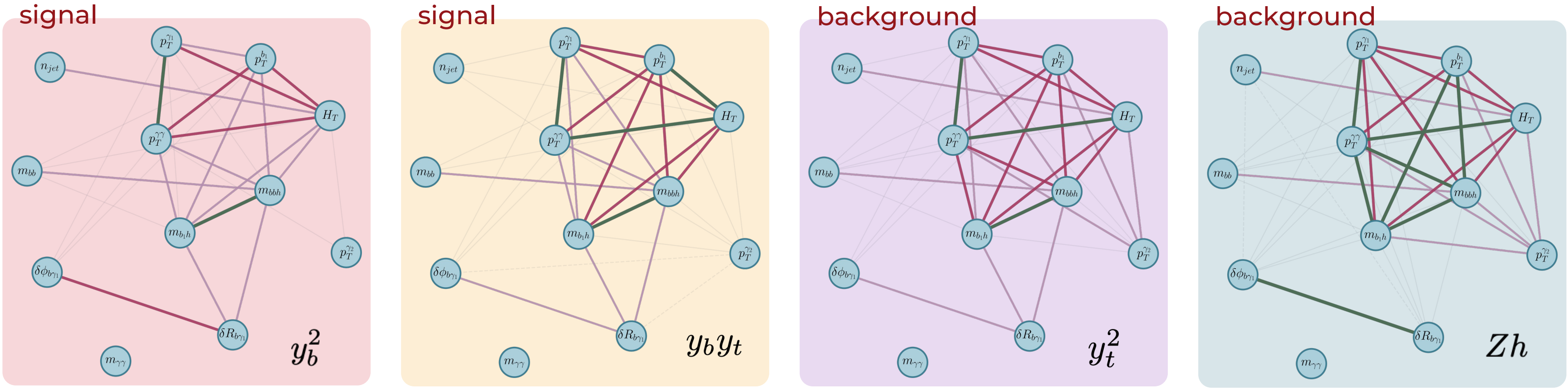
Local interpretation: event by event

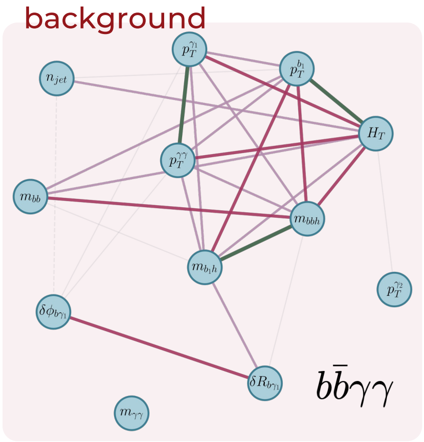


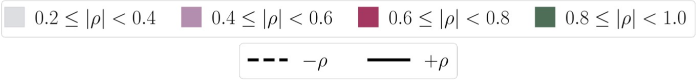
Lloyd S. Shapley
Nobel Laureate 2012

S. M. Lundberg et al., *From local explanations to global understanding with explainable AI for trees.* [Nature Machine Intelligence 2, 56–67]() (2020)

cooperation in Physics



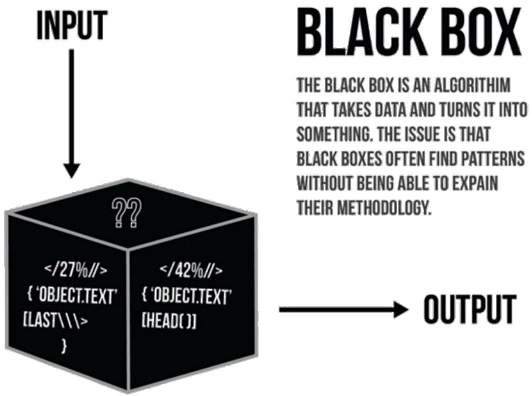


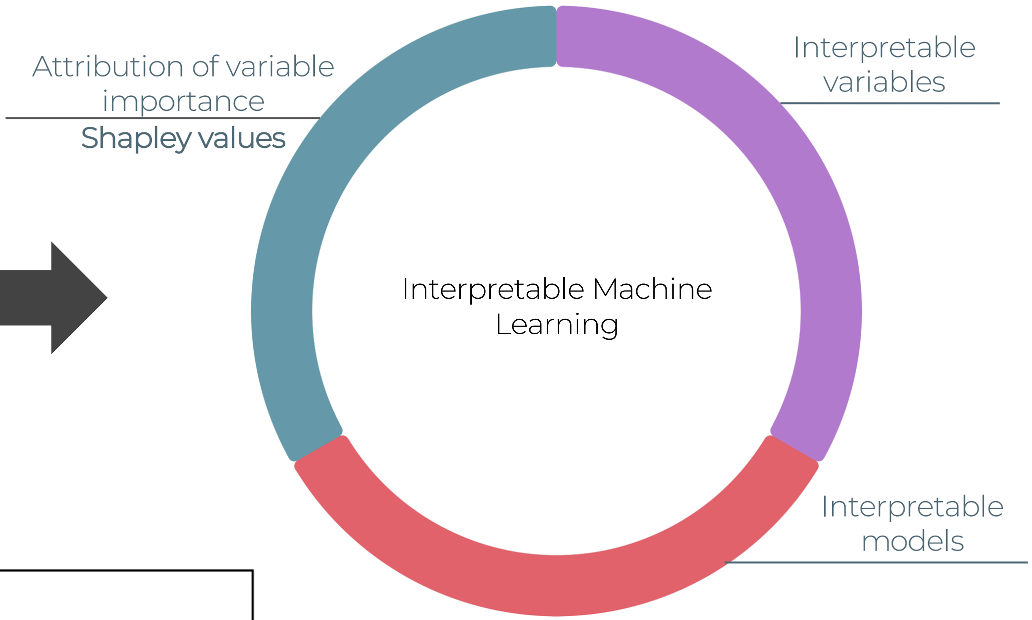


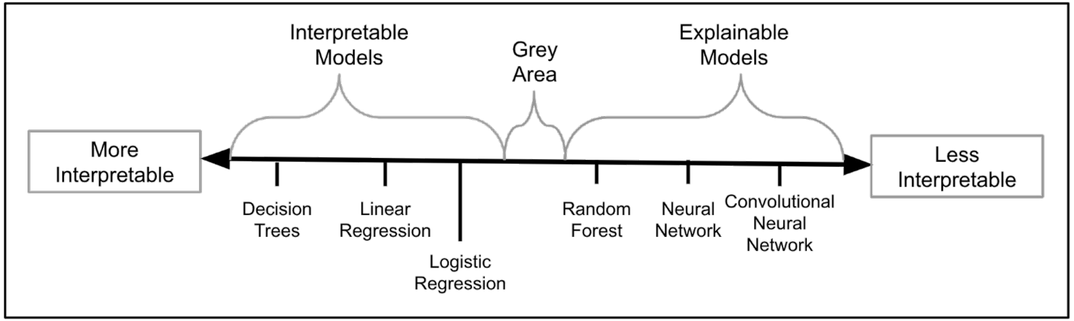
multivariate inherits correlations!

- Variables “cooperate” to bring the outcome
- Outcome can be a measurable quantity or a probability of being of a certain kind
- This covers both regression and classification

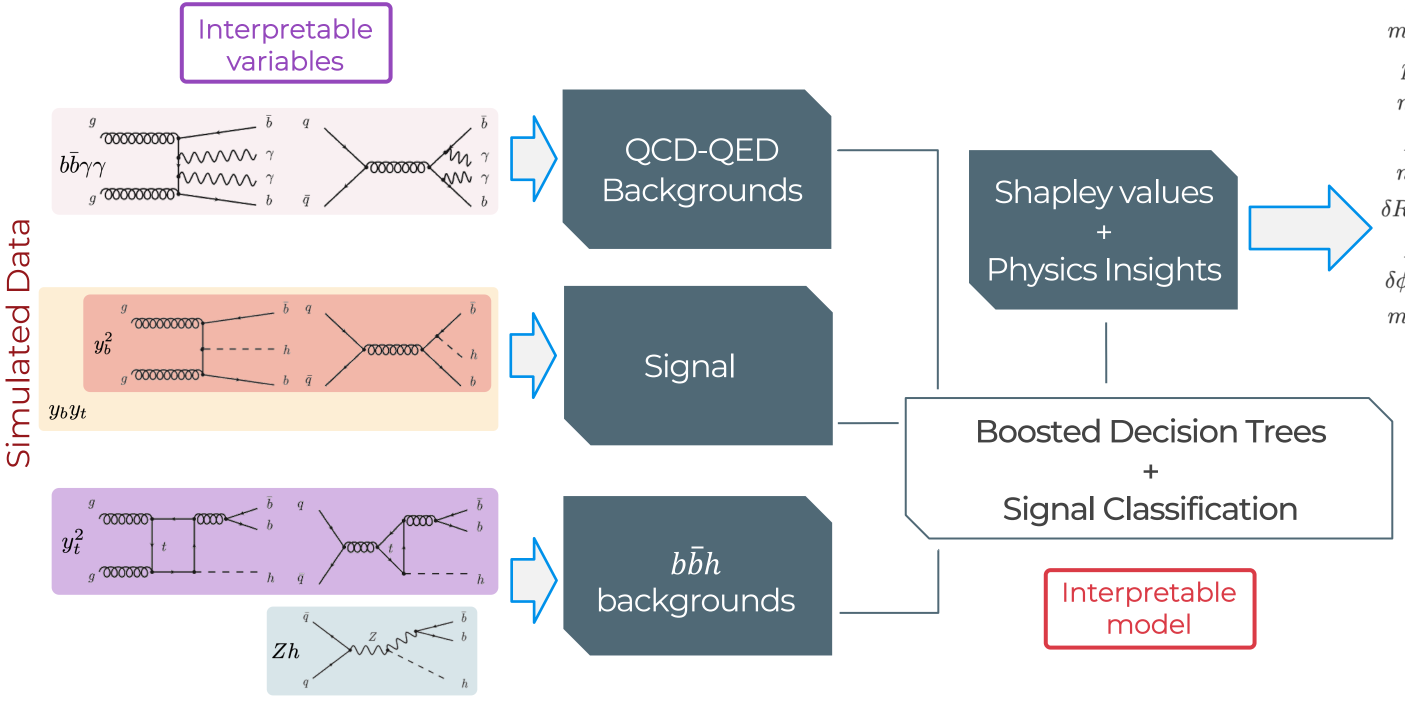
the transition to interpretability

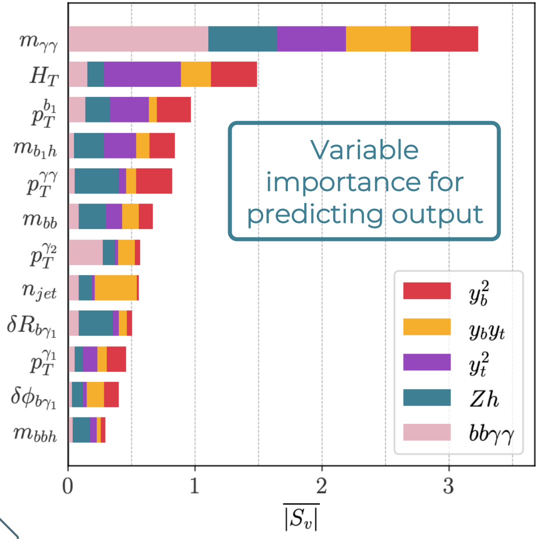


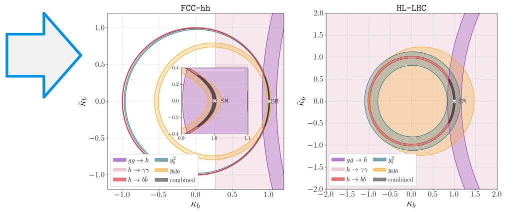




the interpretable framework





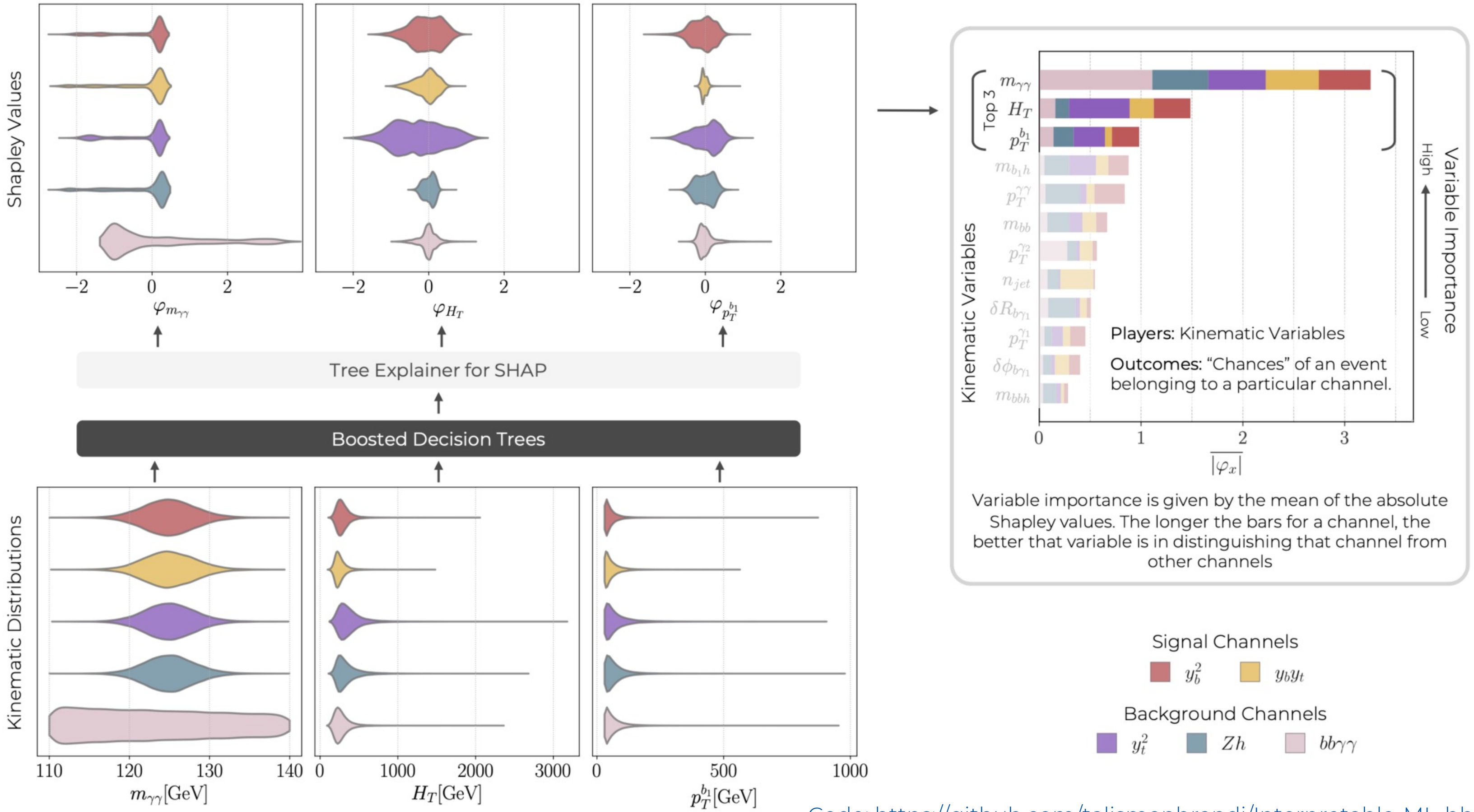


Measurement of Yukawa couplings

- The ML model gives a 60% improvement over a traditional statistical analyses
- The ML model is doing what it is supposed to do from the Shapley values: Trust in ML

Shapley values tell us why the 5 channels can be separated although the kinematic distributions are very similar for 4 of them

12 variable machine learning assisted analysis for classifying 5 particle-production channels



Code: https://github.com/talismanbrandi/Interpretable-ML-bbh

Summary

- Sorry, generative Large Language Models will not write your papers anytime soon.
- Machine learning is good for approximating underlying multivariate distributions and mapping them to downstream tasks like classification, regression, sample generation etc.
- What ML models you use depends completely on the data type and the task at hand. Sometimes, classical methods outperform deep learning.
- Several statistical analyses can be converted to statistical learning analyses by simply rethinking how we do the analyses.
- Interpretable Machine Learning can help us demystify the ML workflow converting a blackbox into an explainable model.
- This is just the tip of the tip of the iceberg. (no type here)

For a wide range of ML papers in HEP: https://iml-wg.github.io/HEPML-LivingReview

“Give me the liberty to know, to utter, and to argue freely according to [con]science, above all liberties.”

— John Milton, Areopagitica

Thank you!