# Applications of AI
# in experimental high energy physics

Gang LI

Institute of High Energy Physics, CAS

第 四 届 全 国 核 物 理 及 核 数 据 中 的 机 器 学 习 应 用 研 讨 会

2025.10.30-11.03 南华大学

# Disclaimers

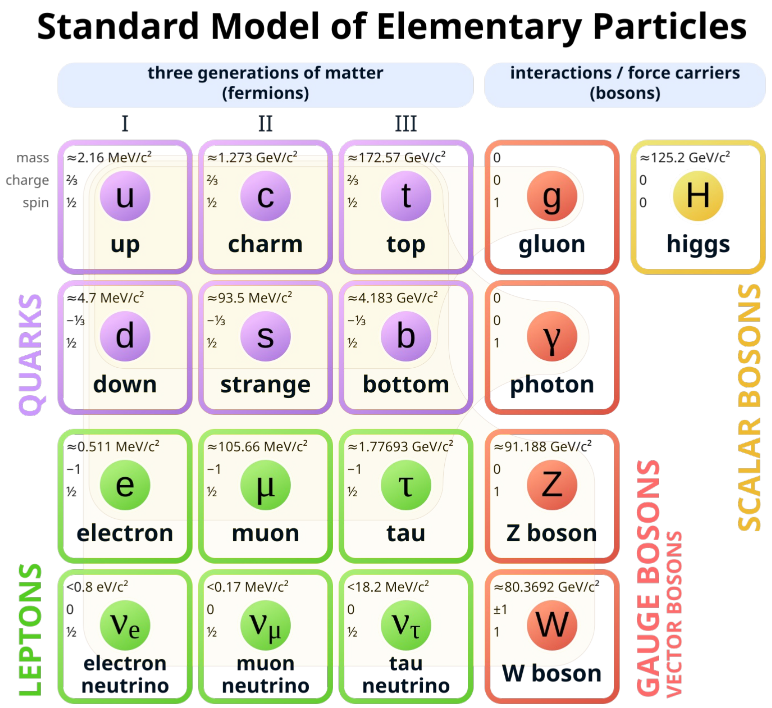
- This is a very personal review, mainly focus on offline data processes, highly biased
- Lots of things not covered

# Outline

- Introduction
- ML and  a few words on it
- Deep learning
- LLM
- Summary

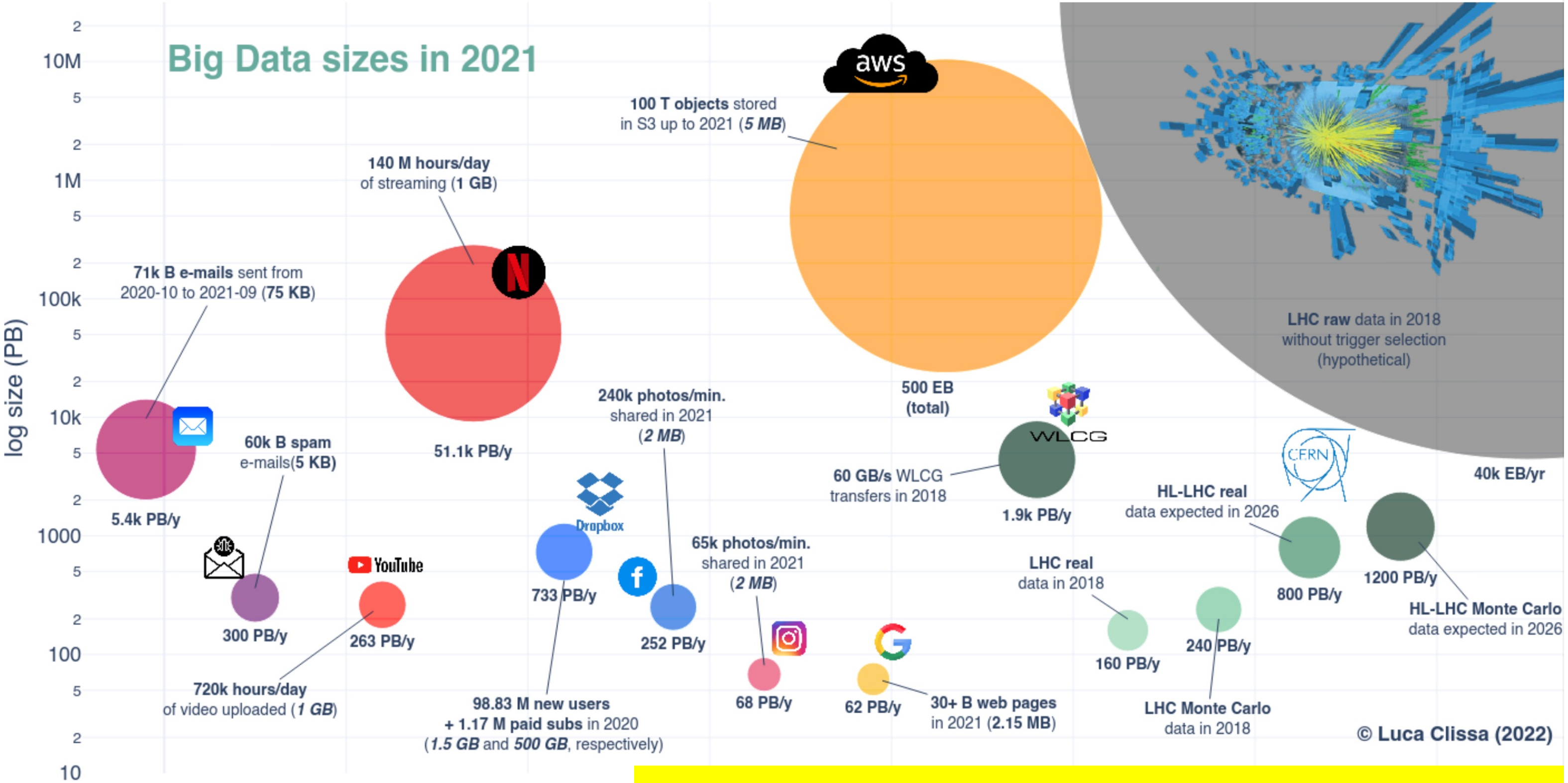
# Introduction

- The Standard Model (SM) theory was completed following the discovery of the Higgs boson in 2012.
- However, it remains merely an effective theory at low energies.
- The search for new physics beyond the SM is a top priority in particle physics.



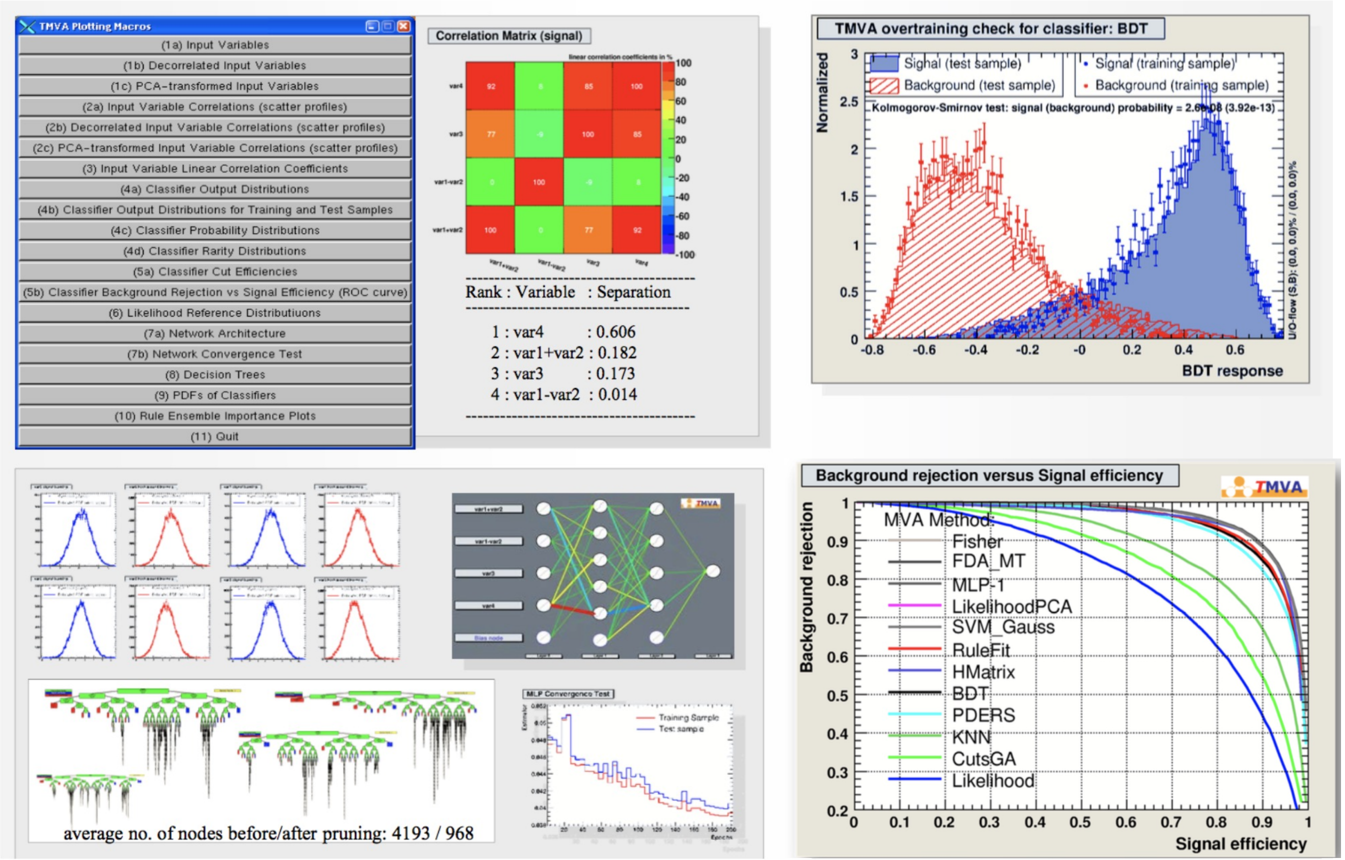
# ML in HEP: the huge data volume

Image credit



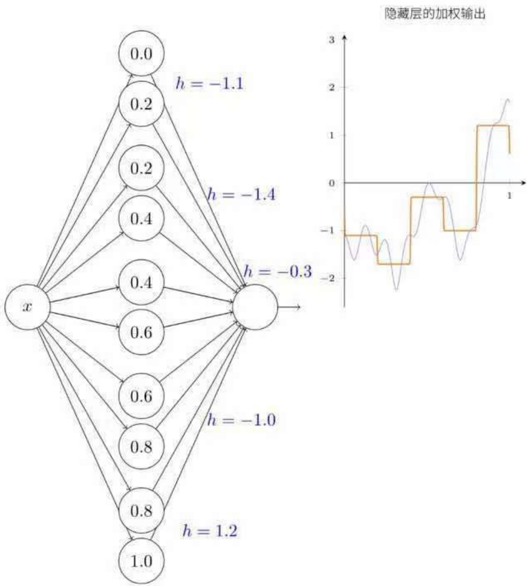

The best way: Big data + AI

# Toolkits for multivariate analysis TMVA more than 20 years

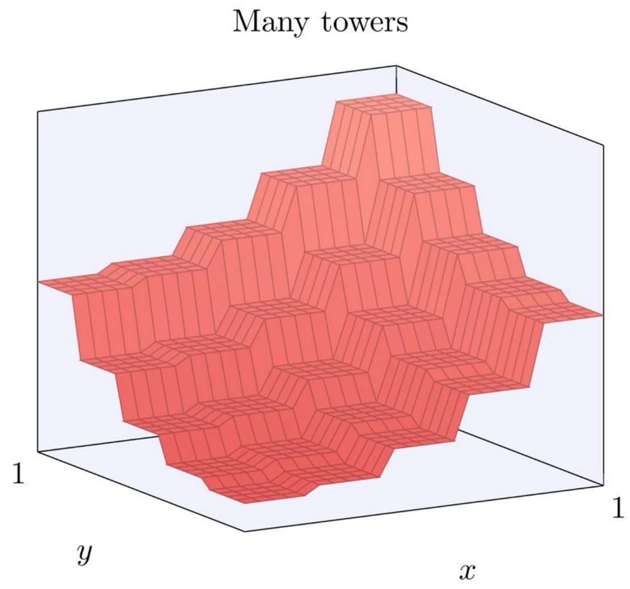


# What should we know about ML?

# Fact 1: Why it works?   Neural network as universal function approximator



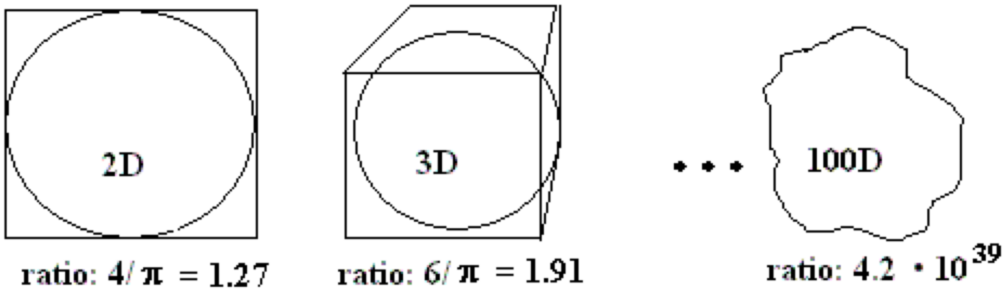

1D




2D

A notable fact about neural networks is that they can approximate a continuous function to any desired level of precision, provided that there are enough neurons in the hidden layers.

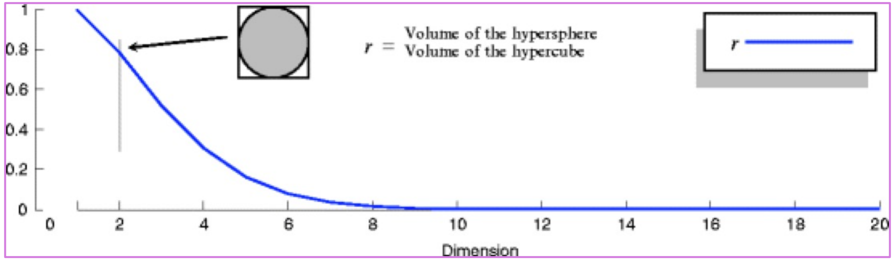
# Fact 2: Why it is difficult? Curse of dimensionality




$$\frac{A_{circle}}{A_{square}} = \frac{\pi}{4} \text{ for } d = 2$$

$$\frac{V_{sphere}}{V_{cube}} = \frac{\pi}{6} \text{ for } d = 3$$

$$\frac{V_{hypersphere}}{V_{hypercube}} = \frac{\pi^{d/2}}{d2^{d-1}\Gamma(d/2)} \to 0 \text{ as } d \to \infty$$




- When D=1: 100 evenly distributed points can sample a unit interval with a distance no greater than 0.01;
- When D=10: it requires $10^{20}$ sampling points to achieve the same sampling rate.
- Almost all points in high-D are isolated

On one hand: fortunately, most specific problems can be reduced in dimensionality!

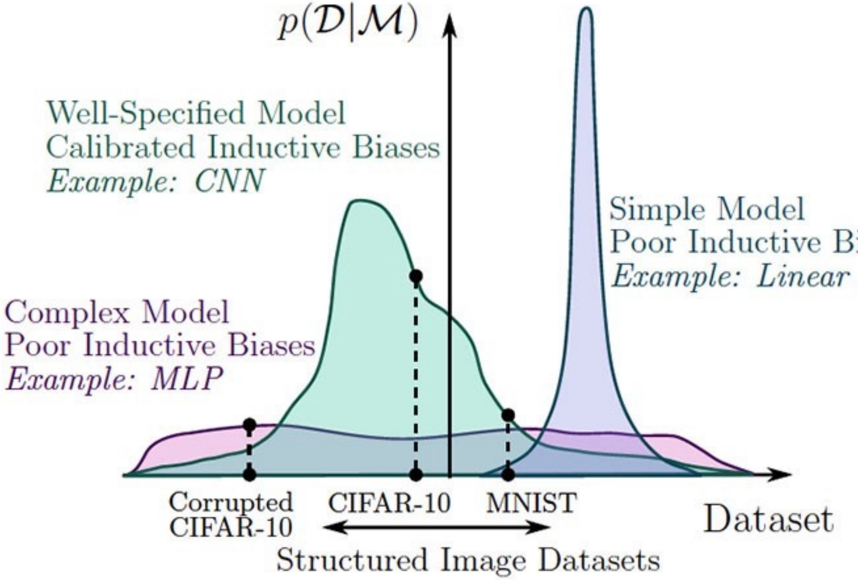
On the other hand:
Neural networks have demonstrated their ability to effectively address the dimension problem!

# Fact 3: No Free lunch theorem and bias-variance tradeoff

No free lunch theorem ( http://www.no-free-lunch.org )

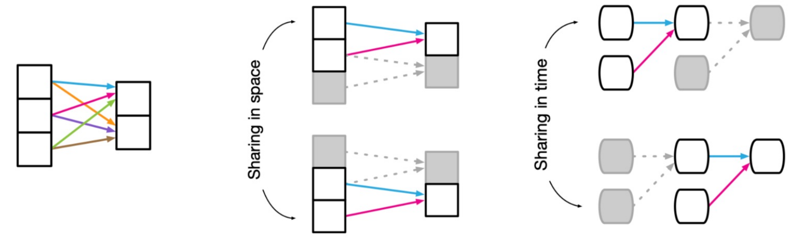
There is no single algorithm that is universally the best for all problems

Performance of a learning algorithm is problem-specific

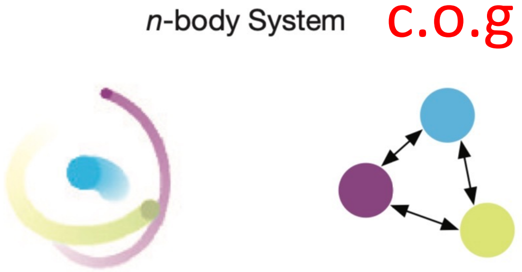



Why do some models perform well on certain datasets? Inductive bias

# Inductive bias of various algorithms




(a) Fully connected (b) Convolutional (c) Recurrent

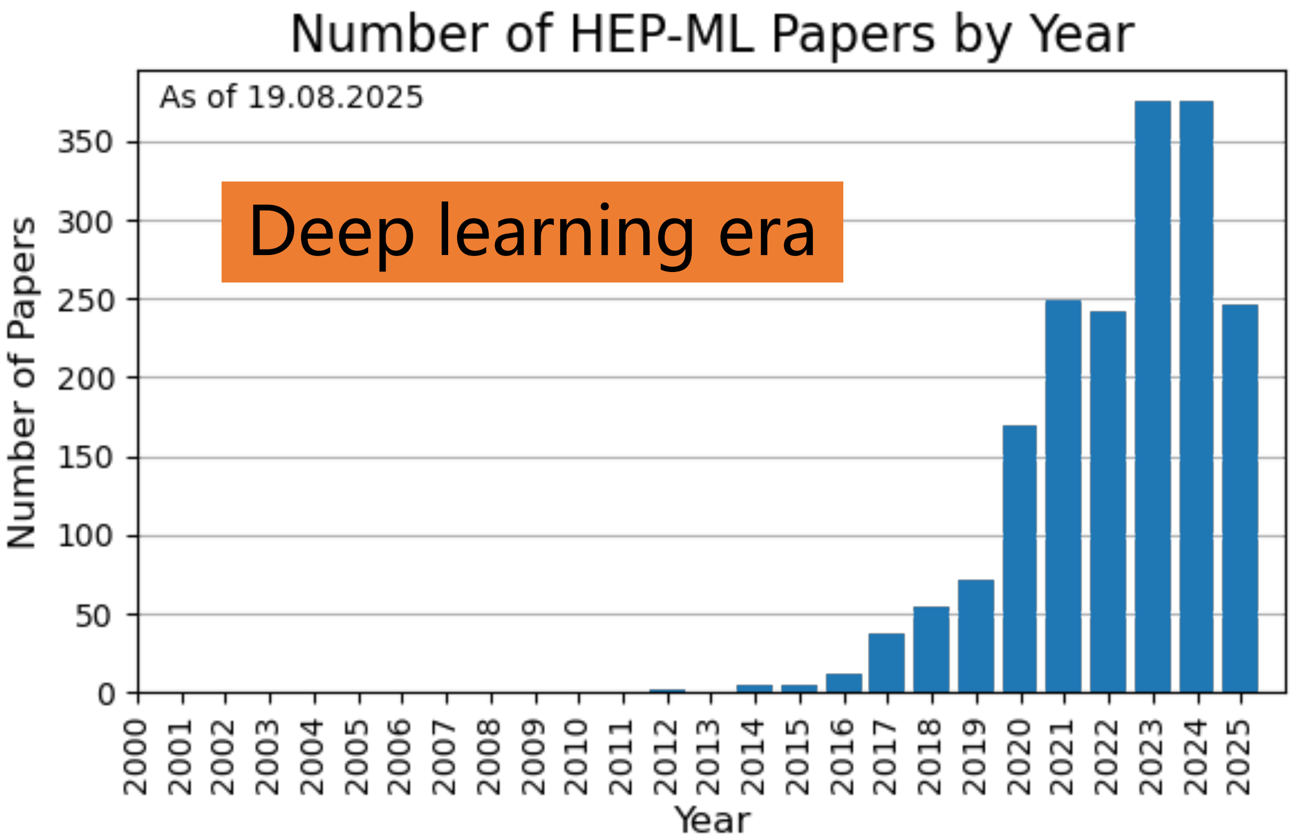



c.o.g

| Component | Entities | Relations | Rel. inductive bias | Invariance |
|---|---|---|---|---|
| Fully connected | Units | All-to-all | Weak | - |
| Convolutional | Grid elements | Local | Locality | Spatial translation |
| Recurrent | Timesteps | Sequential | Sequentiality | Time translation |
| Graph network | Nodes | Edges | Arbitrary | Node, edge permutations |

arXiv:1806.01261

# Summary of the HEP ML Living Review



# Lots of tasks using AI

Classification

Anomaly detection

Unfolding

Statistical inference

… …

Particle identification

Tracking

calorimeter clustering

Fast simulation

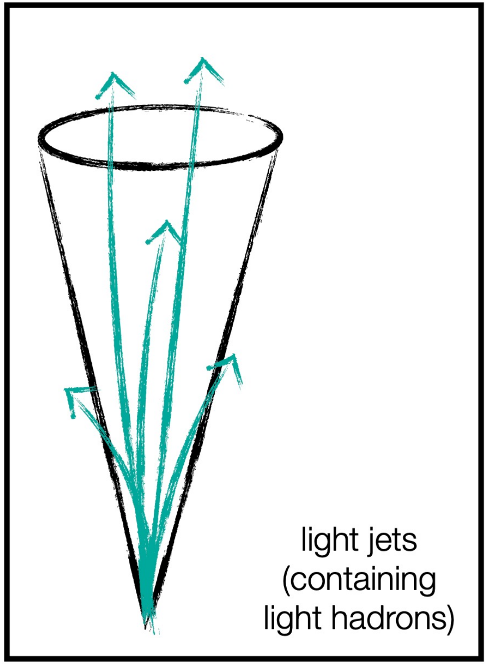
Trigger

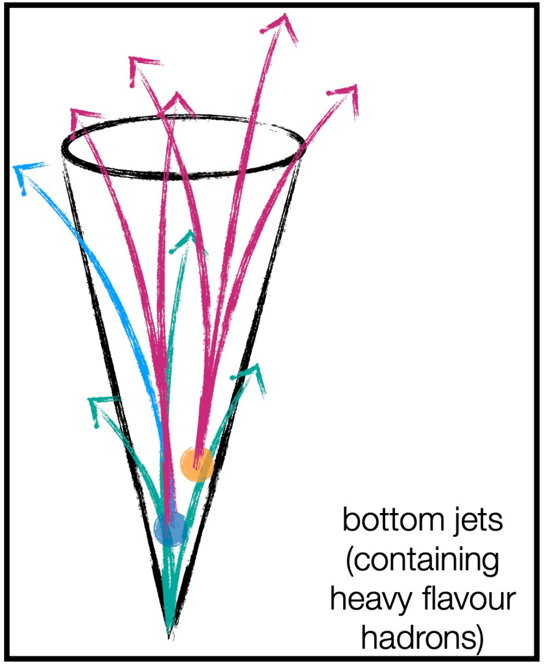
… …

# The most successful story starts from jets

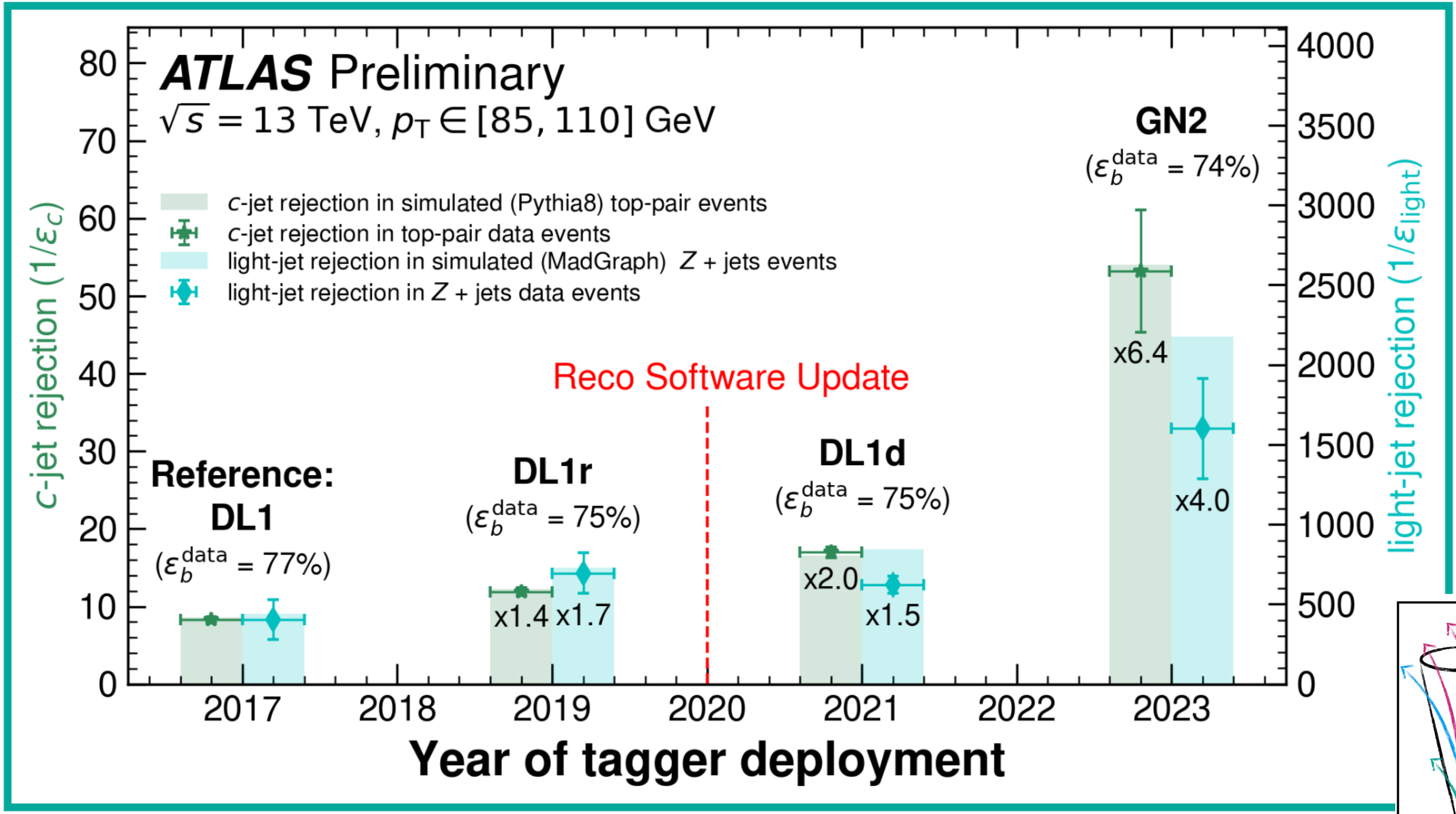
✓ Jet: a spray of a bunch particles
✓ Invariance: edge, node permutations

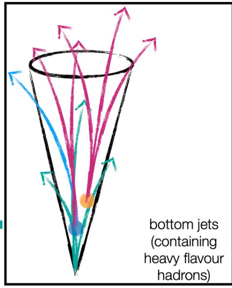
- **Jet Classification (jet tagger)**
- Jet Regression
- Jet Clustering



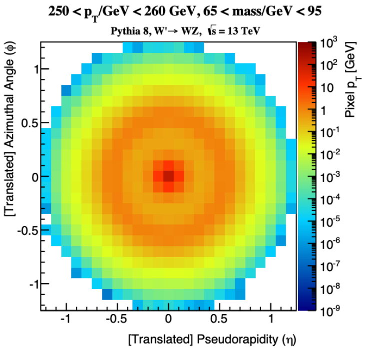



# Learn the History of ML…in ATLAS FTAG

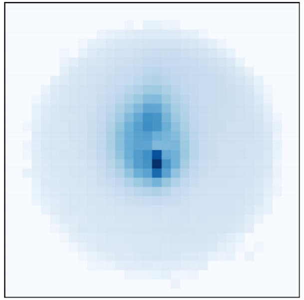


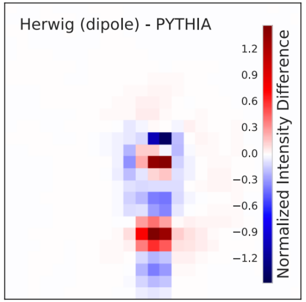
# Jet as images



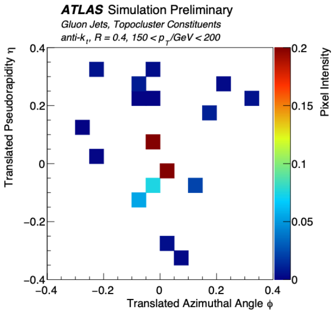

1511.05190



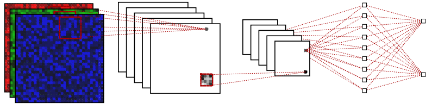
1603.09349




1609.00607




ATL-PHYS-PUB-2017-017



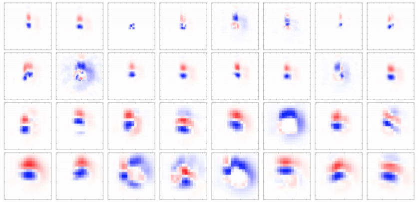
red = transverse momenta of charged particles

green = the transverse momenta of neutral particles

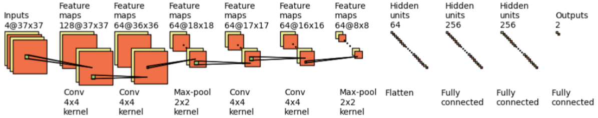
blue = charged particle multiplicity

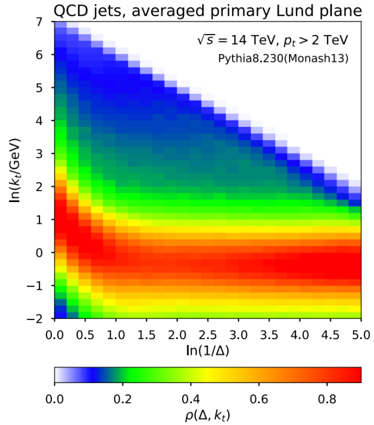
1612.01551

1701.08784

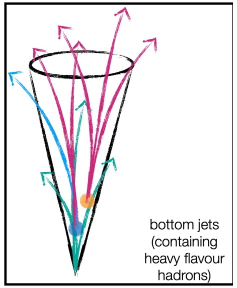


1803.00107



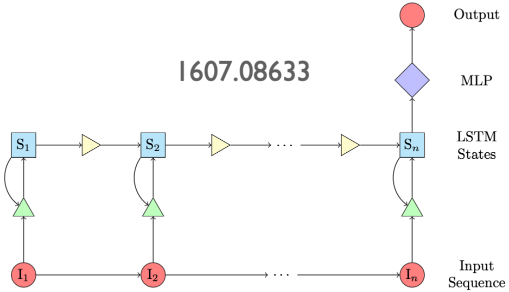



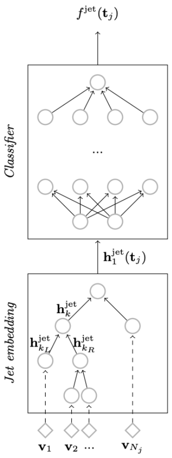

1807.04758



# Jet as sequence

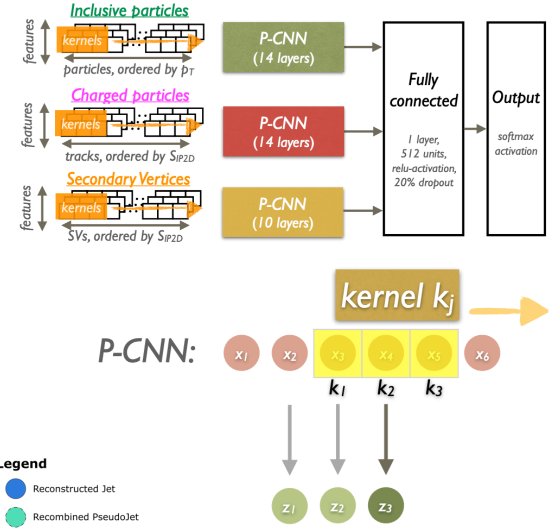
1607.08633

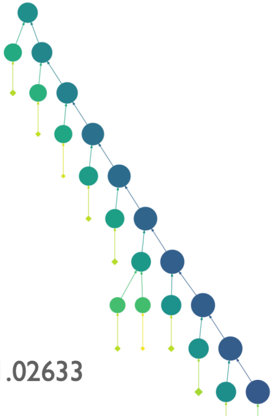


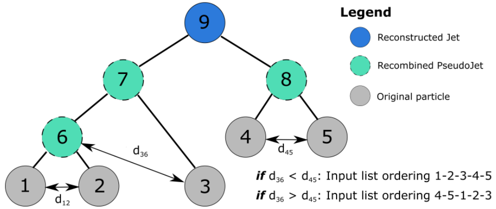

1702.00748

“DeepAK8”
CMS-DP-2017-049

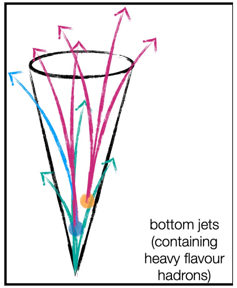


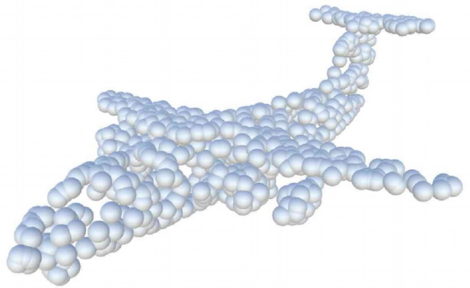
1711.02633

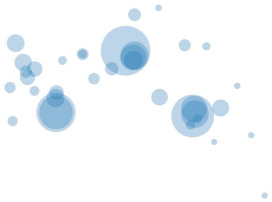



1711.09059



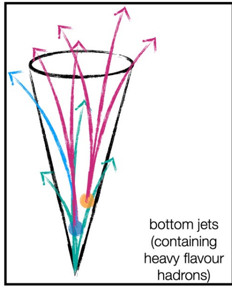
# Jet as sets or cloud





- Point cloud
  - points are intrinsically unordered
  - primary information:
    - 3D coordinates in the xyz space

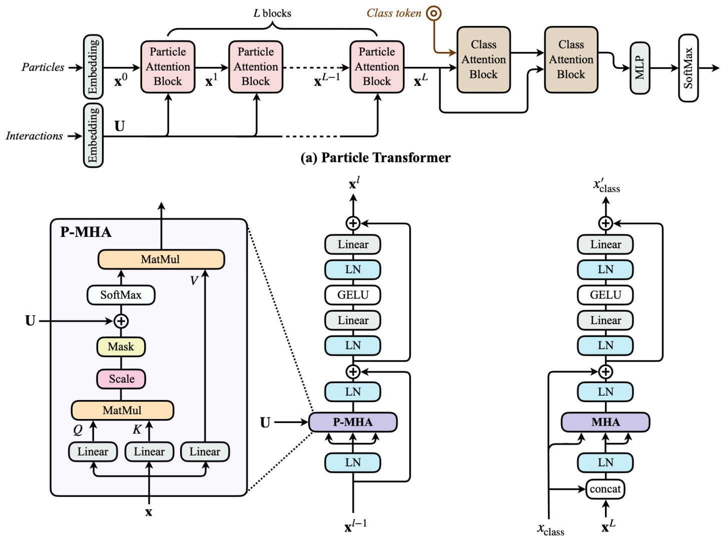
- Particle cloud
  - particles are intrinsically unordered
  - primary information:
    - 2D coordinates in the η-φ space
  - but also additional “features”:
    - energy/momenta
    - charge/particle type
    - track quality/impact parameters/etc.



# Particle Transformer

## ParT architecture

*H.Qu et al. arXiv:2202.03772, proceedings of 39th ICML, Vol.162*




(a) Particle Transformer

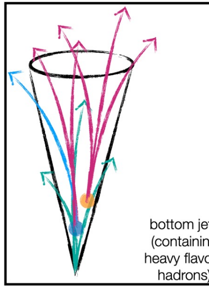
(b) Particle Attention Block

(c) Class Attention Block

| | Accuracy | AUC | $\mathrm{Rej}_{50\%}$ | $\mathrm{Rej}_{30\%}$ |
|---|---|---|---|---|
| P-CNN | 0.930 | 0.9803 | $201 \pm 4$ | $759 \pm 24$ |
| PFN | — | 0.9819 | $247 \pm 3$ | $888 \pm 17$ |
| ParticleNet | 0.940 | 0.9858 | $397 \pm 7$ | $1615 \pm 93$ |
| JEDI-net (w/ $\sum O$) | 0.930 | 0.9807 | — | 774.6 |
| PCT | 0.940 | 0.9855 | $392 \pm 7$ | $1533 \pm 101$ |
| LGN | 0.929 | 0.964 | — | $435 \pm 95$ |
| rPCN | — | 0.9845 | $364 \pm 9$ | $1642 \pm 93$ |
| LorentzNet | 0.942 | 0.9868 | $498 \pm 18$ | $2195 \pm 173$ |
| ParT | 0.940 | 0.9858 | $413 \pm 16$ | $1602 \pm 81$ |
| ParticleNet-f.t. | 0.942 | 0.9866 | $487 \pm 9$ | $1771 \pm 80$ |
| **ParT-f.t.** | **0.944** | **0.9877** | $\mathbf{691 \pm 15}$ | $\mathbf{2766 \pm 130}$ |

The-state-of-the-arts

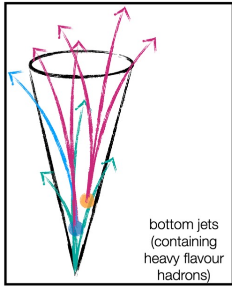
Multi-head attention



# The most performant

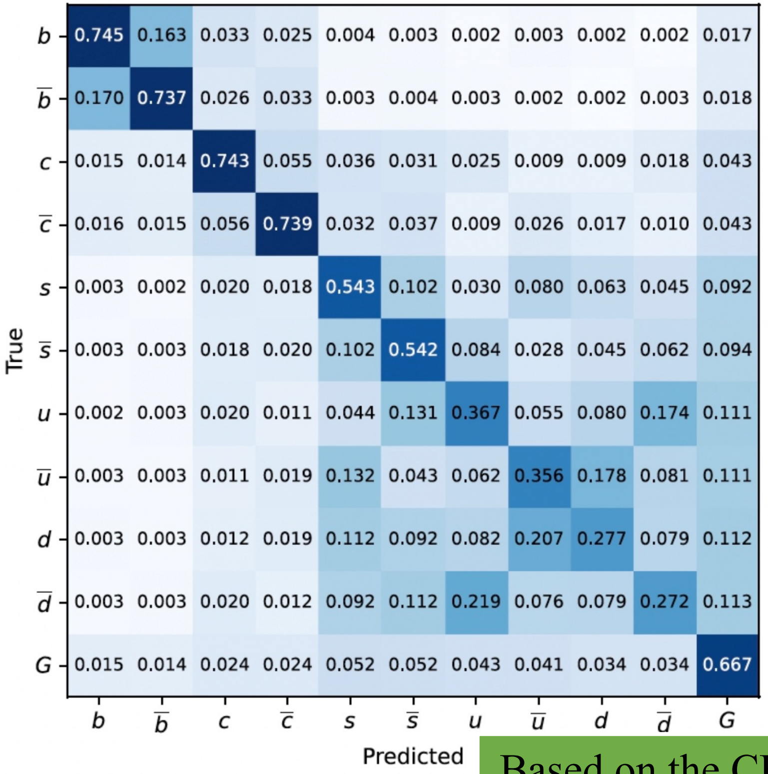
| | Accuracy | AUC | $\mathrm{Rej}_{50\%}$ | $\mathrm{Rej}_{30\%}$ |
|---|---|---|---|---|
| P-CNN | 0.930 | 0.9803 | $201 \pm 4$ | $759 \pm 24$ |
| PFN | — | 0.9819 | $247 \pm 3$ | $888 \pm 17$ |
| ParticleNet | 0.940 | 0.9858 | $397 \pm 7$ | $1615 \pm 93$ |
| JEDI-net (w/ $\sum O$) | 0.930 | 0.9807 | — | 774.6 |
| PCT | 0.940 | 0.9855 | $392 \pm 7$ | $1533 \pm 101$ |
| LGN | 0.929 | 0.964 | — | $435 \pm 95$ |
| rPCN | — | 0.9845 | $364 \pm 9$ | $1642 \pm 93$ |
| LorentzNet | 0.942 | 0.9868 | $498 \pm 18$ | $2195 \pm 173$ |
| ParT | 0.940 | 0.9858 | $413 \pm 16$ | $1602 \pm 81$ |
| ParticleNet-f.t. | 0.942 | 0.9866 | $487 \pm 9$ | $1771 \pm 80$ |
| **ParT-f.t.** | **0.944** | **0.9877** | $\mathbf{691 \pm 15}$ | $\mathbf{2766 \pm 130}$ |

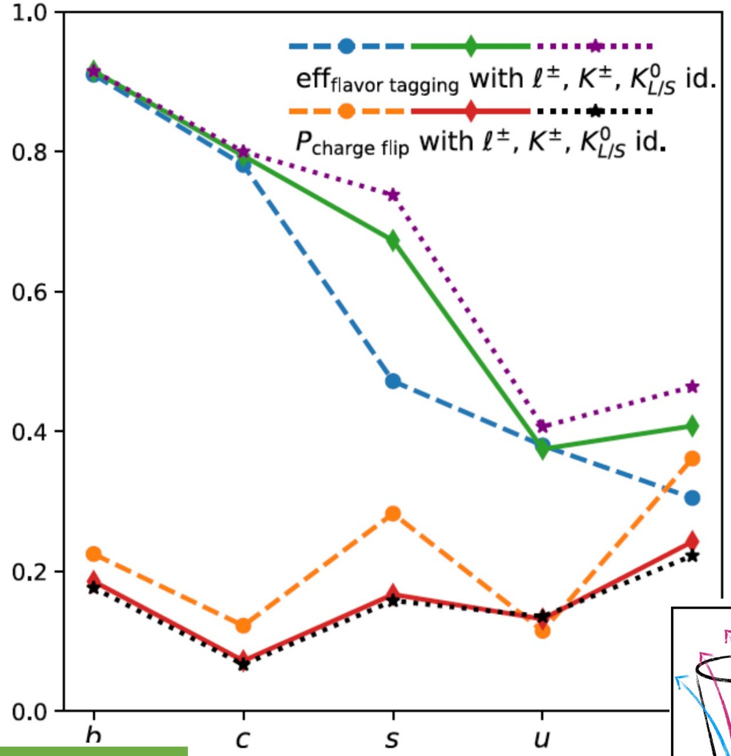
On top quark tagging dataset



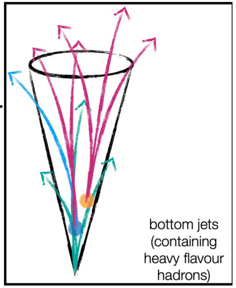
# All 11 types of jets

PRL *132, 221802 (2024)*



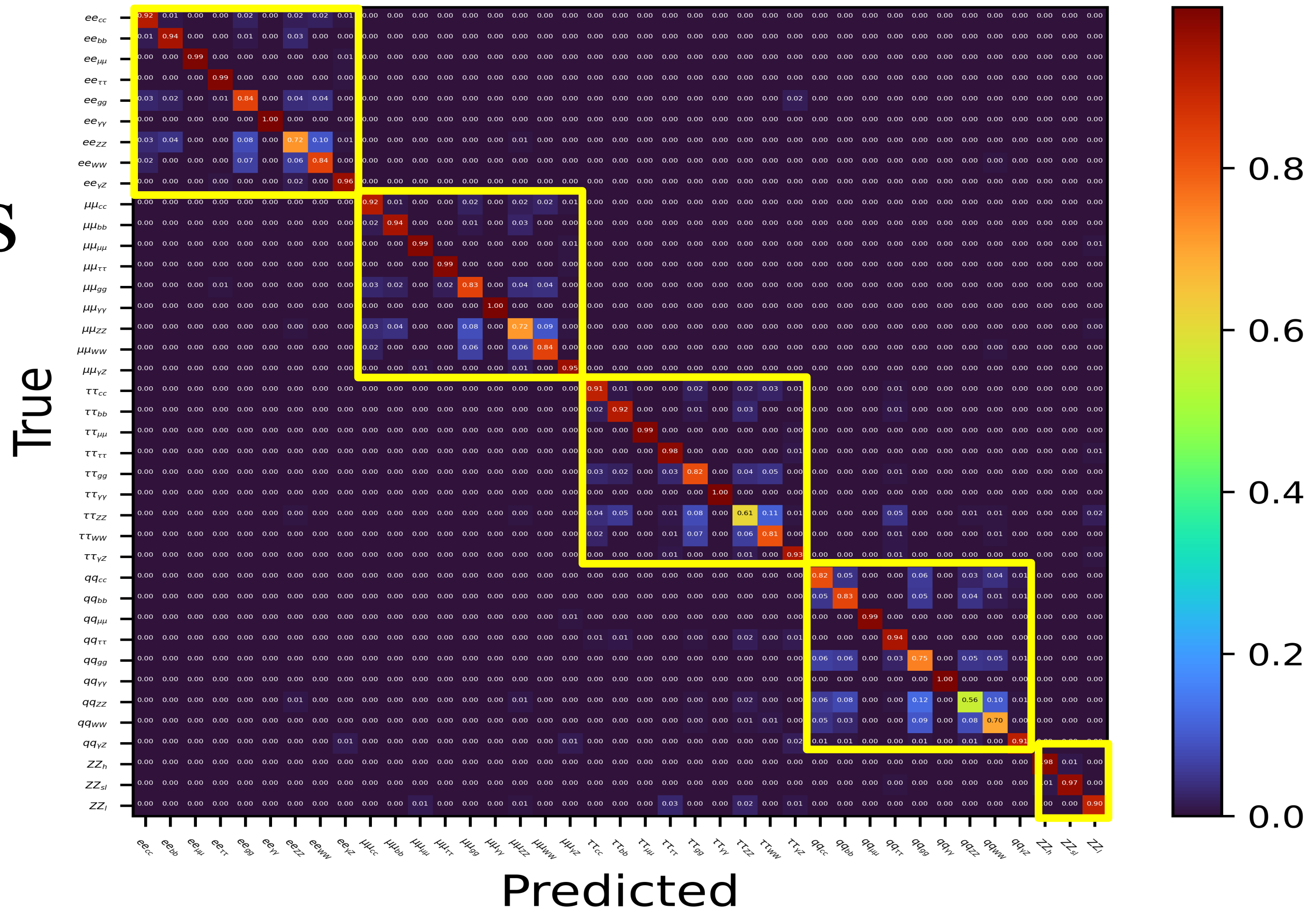

Based on the CEPC simulation



# Not only jets

# Multi-classification of event

39 classes



# Sophon at CMS

hep-ex:2405.12972

| Major types | Index range | Label names |
|---|---|---|
| Resonant jets: $X \to$ 2 prong | 0–14 | $bb$, $cc$, $ss$, $qq$, $bc$, $cs$, $bq$, $cq$, $sq$, $gg$, $ee$, $\mu\mu$, $\tau_{\mathrm{h}}\tau_e$, $\tau_{\mathrm{h}}\tau_\mu$, $\tau_{\mathrm{h}}\tau_{\mathrm{h}}$ |
| Resonant jets: $X \to$ 3 or 4 prong | 15–160 | $bbbb$, $bbcc$, $bbss$, $bbqq$, $bbgg$, $bbee$, $bb\mu\mu$, $bb\tau_{\mathrm{h}}\tau_e$, $bb\tau_{\mathrm{h}}\tau_\mu$, $bb\tau_{\mathrm{h}}\tau_{\mathrm{h}}$, $bbb$, $bbc$, $bbs$, $bbq$, $bbg$, $bbe$, $bb\mu$, $cccc$, $ccss$, $ccqq$, $ccgg$, $ccee$, $cc\mu\mu$, $cc\tau_{\mathrm{h}}\tau_e$, $cc\tau_{\mathrm{h}}\tau_\mu$, $cc\tau_{\mathrm{h}}\tau_{\mathrm{h}}$, $ccb$, $ccc$, $ccs$, $ccq$, $ccg$, $cce$, $cc\mu$, $ssss$, $ssqq$, $ssgg$, $ssee$, $ss\mu\mu$, $ss\tau_{\mathrm{h}}\tau_e$, $ss\tau_{\mathrm{h}}\tau_\mu$, $ss\tau_{\mathrm{h}}\tau_{\mathrm{h}}$, $ssb$, $ssc$, $sss$, $ssq$, $ssg$, $sse$, $ss\mu$, $qqqq$, $qqgg$, $qqee$, $qq\mu\mu$, $qq\tau_{\mathrm{h}}\tau_e$, $qq\tau_{\mathrm{h}}\tau_\mu$, $qq\tau_{\mathrm{h}}\tau_{\mathrm{h}}$, $qqb$, $qqc$, $qqs$, $qqq$, $qqg$, $qqe$, $qq\mu$, $gggg$, $ggee$, $gg\mu\mu$, $gg\tau_{\mathrm{h}}\tau_e$, $gg\tau_{\mathrm{h}}\tau_\mu$, $gg\tau_{\mathrm{h}}\tau_{\mathrm{h}}$, $ggb$, $ggc$, $ggs$, $ggq$, $ggg$, $gge$, $gg\mu$, $bee$, $cee$, $see$, $qee$, $gee$, $b\mu\mu$, $c\mu\mu$, $s\mu\mu$, $q\mu\mu$, $g\mu\mu$, $b\tau_{\mathrm{h}}\tau_e$, $c\tau_{\mathrm{h}}\tau_e$, $s\tau_{\mathrm{h}}\tau_e$, $q\tau_{\mathrm{h}}\tau_e$, $g\tau_{\mathrm{h}}\tau_e$, $b\tau_{\mathrm{h}}\tau_\mu$, $c\tau_{\mathrm{h}}\tau_\mu$, $s\tau_{\mathrm{h}}\tau_\mu$, $q\tau_{\mathrm{h}}\tau_\mu$, $g\tau_{\mathrm{h}}\tau_\mu$, $b\tau_{\mathrm{h}}\tau_{\mathrm{h}}$, $c\tau_{\mathrm{h}}\tau_{\mathrm{h}}$, $s\tau_{\mathrm{h}}\tau_{\mathrm{h}}$, $q\tau_{\mathrm{h}}\tau_{\mathrm{h}}$, $g\tau_{\mathrm{h}}\tau_{\mathrm{h}}$, $qqqb$, $qqqc$, $qqqs$, $bbcq$, $ccbs$, $ccbq$, $ccsq$, $sscq$, $qqbc$, $qqbs$, $qqcs$, $bcsq$, $bcs$, $bcq$, $bsq$, $csq$, $bce\nu$, $cse\nu$, $bqe\nu$, $cqe\nu$, $sqe\nu$, $qqe\nu$, $bc\mu\nu$, $cs\mu\nu$, $bq\mu\nu$, $cq\mu\nu$, $sq\mu\nu$, $qq\mu\nu$, $bc\tau_e\nu$, $cs\tau_e\nu$, $bq\tau_e\nu$, $cq\tau_e\nu$, $sq\tau_e\nu$, $qq\tau_e\nu$, $bc\tau_\mu\nu$, $cs\tau_\mu\nu$, $bq\tau_\mu\nu$, $cq\tau_\mu\nu$, $sq\tau_\mu\nu$, $qq\tau_\mu\nu$, $bc\tau_{\mathrm{h}}\nu$, $cs\tau_{\mathrm{h}}\nu$, $bq\tau_{\mathrm{h}}\nu$, $cq\tau_{\mathrm{h}}\nu$, $sq\tau_{\mathrm{h}}\nu$, $qq\tau_{\mathrm{h}}\nu$ |
| QCD jets | 161–187 | $bbccss$, $bbccs$, $bbcc$, $bbcss$, $bbcs$, $bbc$, $bbss$, $bbs$, $bb$, $bccss$, $bccs$, $bcc$, $bcss$, $bcs$, $bc$, $bss$, $bs$, $b$, $ccss$, $ccs$, $cc$, $css$, $cs$, $c$, $ss$, $s$, others |

*resulting to 4-prong or 3-prong*

***All final states!***

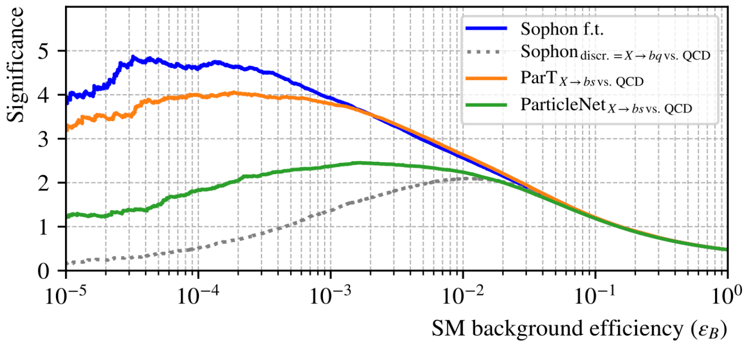
~200 classes

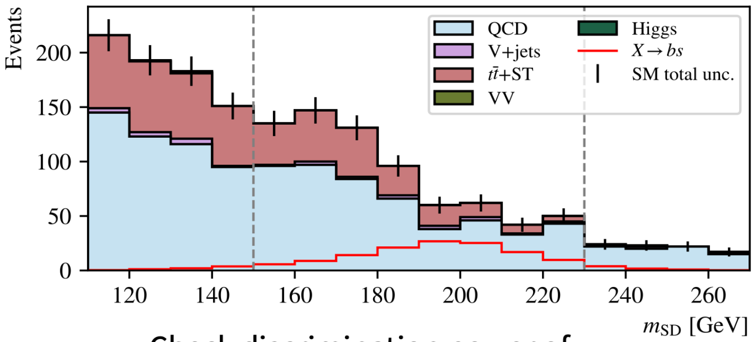
# Sophon: performance benchmark

arXiv:2405.12972

Search significance:

$$Z = \sqrt{2((s+b)\log(1+s/b) - s)}$$




- Check discrimination power of X (200 GeV) → **bs** signal vs. all backgrounds

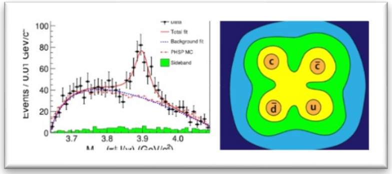
## ***Transfer learning ability***

(adapt it to a brand new task)

- **Sophon** (training on 188 classes) reaches the best performance **after fine-tuned (via transfer learning)**
- **ParT** and **ParticleNet** for binary X→bs vs QCD classification: they reveal the best performance we can reach in the experiment now

# Large Language Model

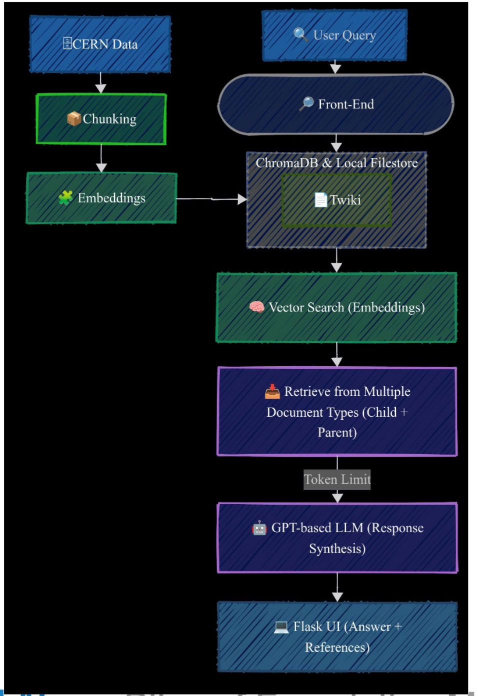
# 用于BESIII物理分析的“Dr.Sai”智能体

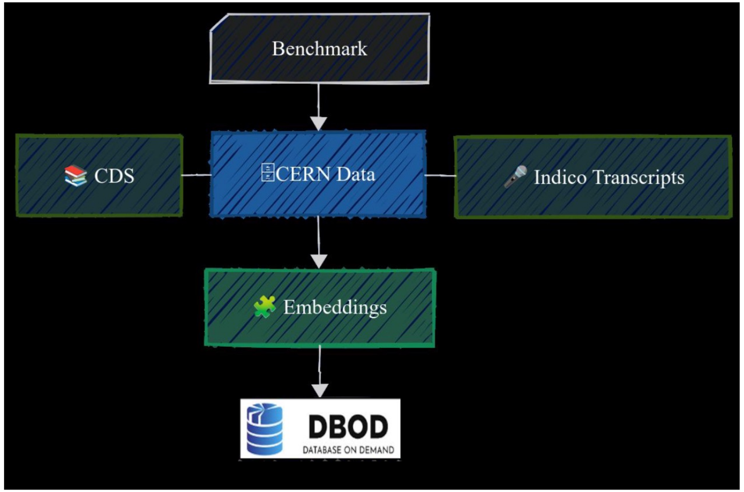



**当前目标：“赛博士”智能体重新发现Zc(3900)粒子**

# CHATLAS: copilot of ATLAS experiment

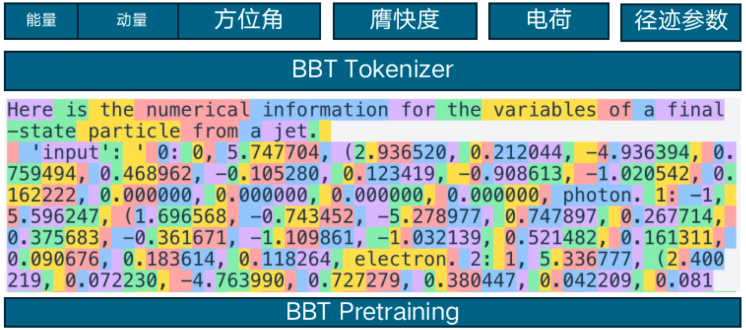
## What's New? - Key Improvements



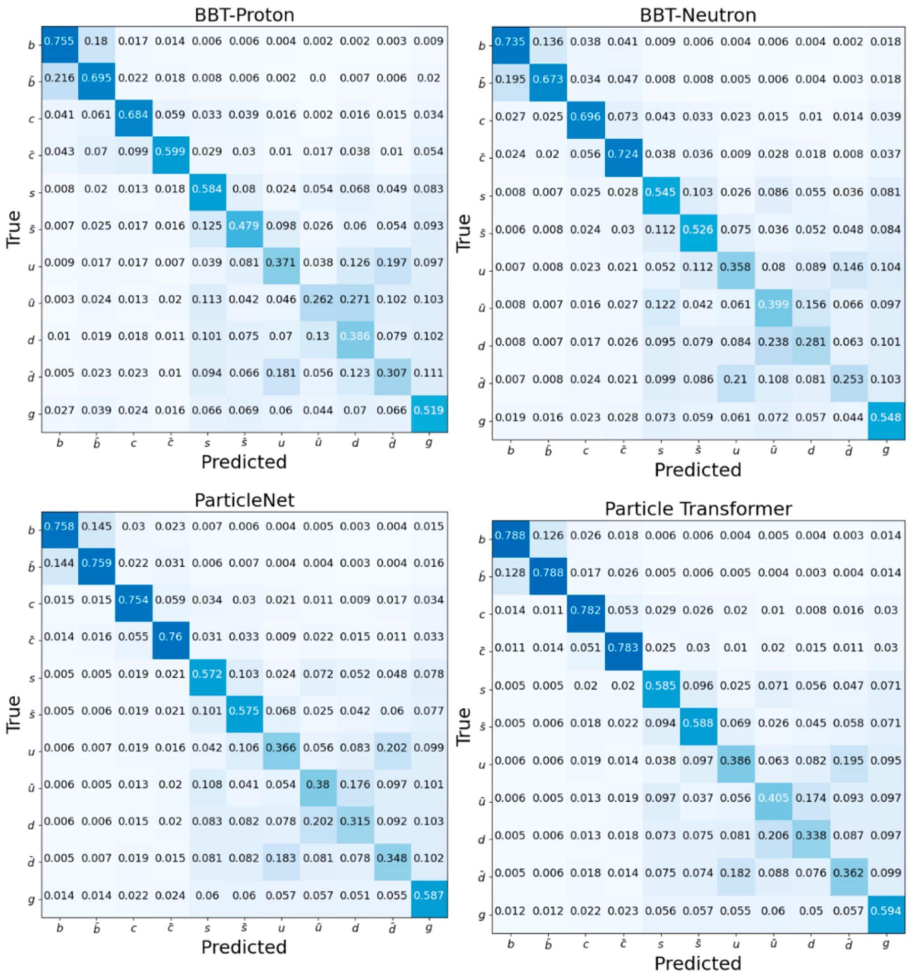



# Supersymmetry Technologies

Feed data to LLM directly with some sophisticated tokenizer




The results comparable with the-state-of-the-art models



# Summary

- Experimental particle physics is accumulating ever-increasing volume of data
- AI (machine learning) is essential for the search for the BSM physics
- Traditional methods remain powerful and offer greater interpretability
- Deep learning shows promising  performance when proper invariance is injected
- Systematics uncertainties remain challenging
- LLMs simultaneously used as a copilot or directly employed to process data.

# Extras

iDcyTr: 3357
N_track: 16
N_charged: 7

What we observed with noise

| Track | pdgid | motheridx | px | py | pz | E | mass |
|---|---|---|---|---|---|---|---|
| 0 | 443 | -1 | 0.0340672 | 0 | 0.000446902 | 3.09708 | 3.09689 |
| 1 | -10213 | 0 | 0.0765072 | -0.47771 | 0.269094 | 1.31691 | 1.19489 |
| 2 | 211 | 0 | 0.394086 | 0.18534 | -0.484988 | 0.666594 | 0.13957 |
| 3 | 331 | 0 | -0.436526 | 0.29237 | 0.216341 | 1.11358 | 0.957719 |
| 4 | 223 | 1 | -0.0961185 | -0.266653 | -0.100157 | 0.837144 | 0.781304 |
| 5 | -211 | 1 | 0.172626 | -0.211057 | 0.369251 | 0.479762 | 0.139571 |
| 6 | 111 | 4 | -0.204614 | 0.199897 | -0.159925 | 0.35443 | 0.134977 |
| 7 | 22 | 4 | 0.108496 | -0.46655 | 0.059768 | 0.482714 | -1.49012e-08 |
| 8 | 22 | 6 | -0.0179737 | 0.0829306 | -0.00473804 | 0.0849881 | 6.38483e-09 |
| 9 | 22 | 6 | -0.18664 | 0.116967 | -0.155187 | 0.269442 | 5.26836e-09 |
| 10 | 211 | 3 | 0.00706431 | 0.143894 | -0.0392698 | 0.204395 | 0.13957 |
| 11 | -211 | 3 | -0.0112446 | -0.0309255 | -0.00346318 | 0.14344 | 0.139571 |
| 12 | 221 | 3 | -0.432345 | 0.179402 | 0.259074 | 0.765748 | 0.547853 |
| 13 | 22 | 12 | 0.0240193 | -0.0330553 | -0.0133468 | 0.0429851 | 6.46917e-09 |
| 14 | -211 | 12 | -0.321773 | 0.258329 | 0.289116 | 0.522819 | 0.139571 |
| 15 | 211 | 12 | -0.134592 | -0.0458715 | -0.0166951 | 0.199944 | 0.13957 |

What we want to know in a event