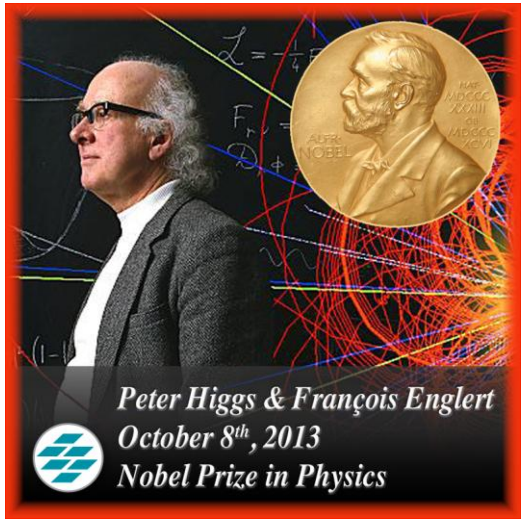
Prospects of AI application at BESIII

刘北江

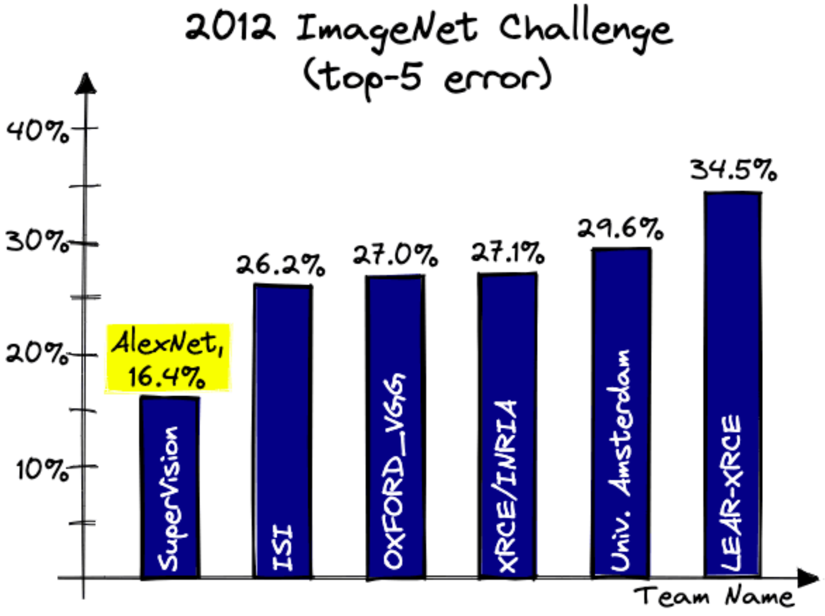
BESIII实验物理研讨会 2026.2.6-9

2012

The Higgs boson



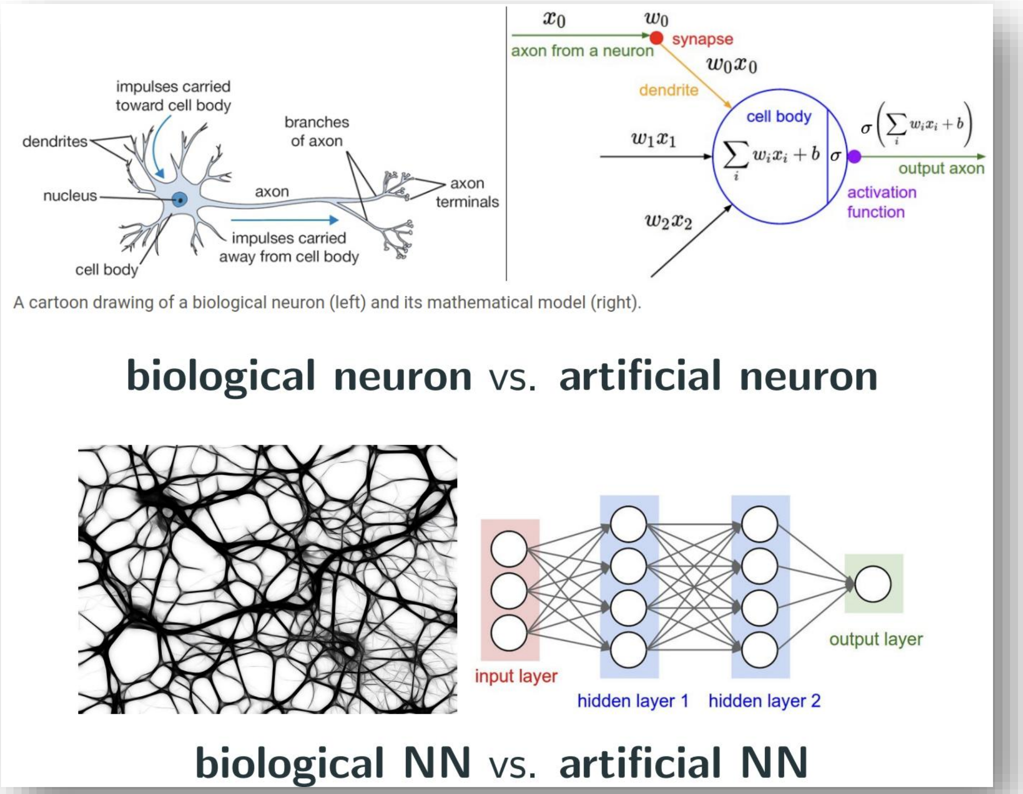
AlexNet

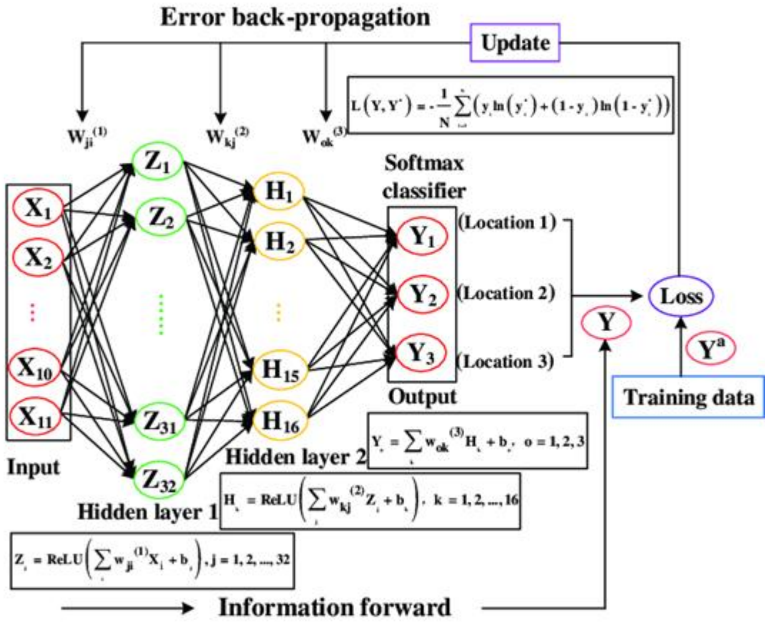


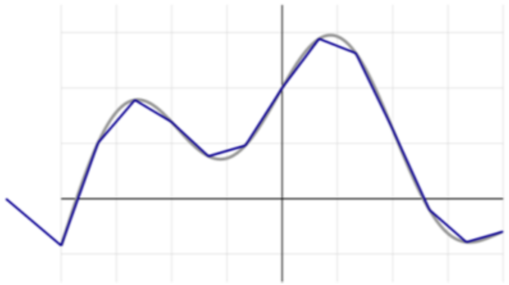
Machine Learning

Fundamental Belief: Universal Approximation Theorems

-- A neural network can effectively approximate a high-dim. function







- Differentiable programming
- Backward Propagation
- Gradient Descent Algorithm

Machine Learning

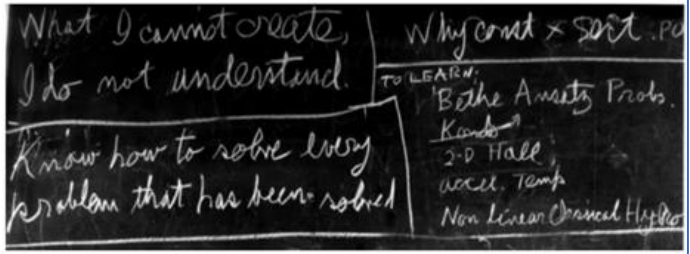
- Discriminative Learning : **prediction**

  function fitting $y = f(x)$

  conditional probability $\boxed{p_\theta(y|x)} \rightarrow p(y|x)$

- Generative Modelling : **understand**

  Joint probability distribution $\boxed{p_\theta(x, y)} \rightarrow p(x, y)$

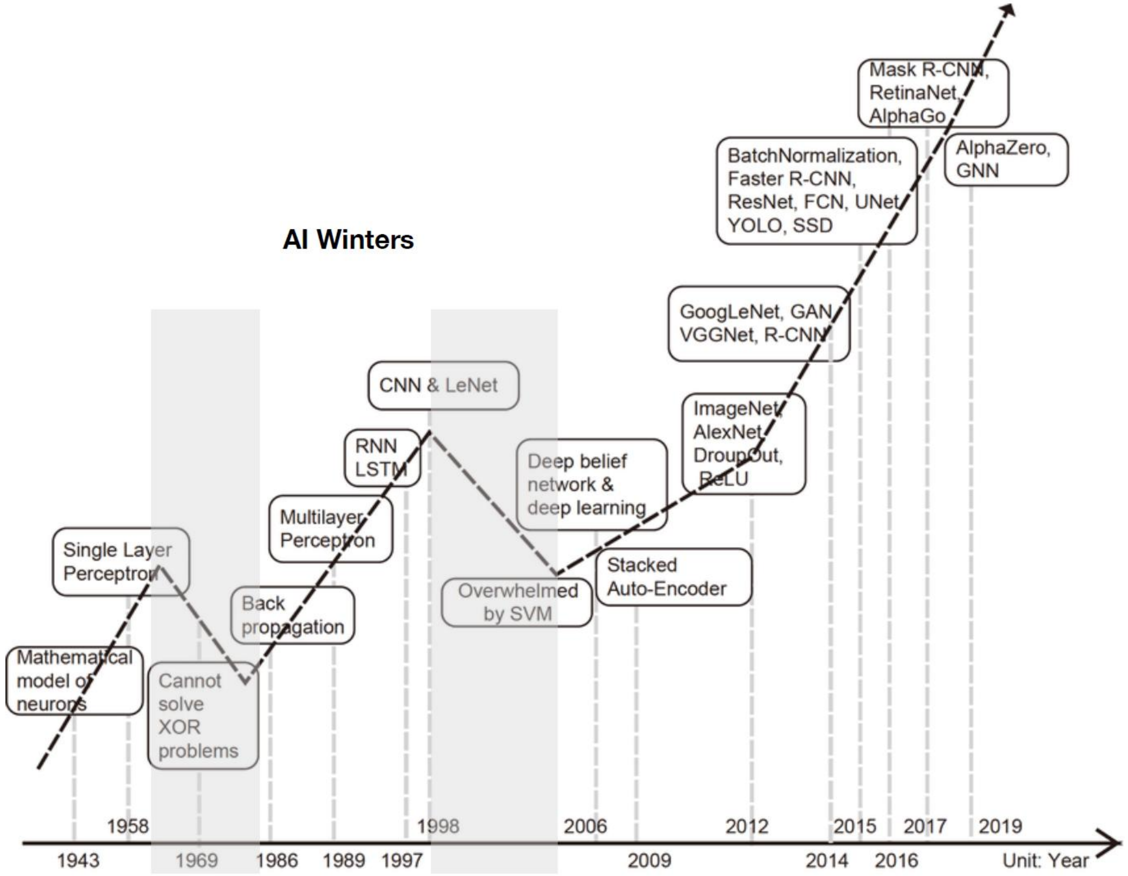


“What I can not create, I do not understand”

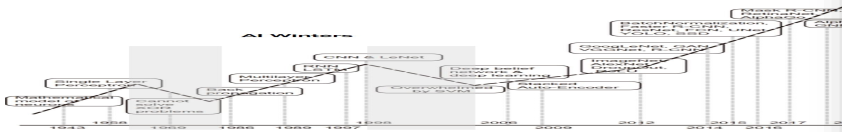
Slides courtesy to Prof. Kai Zhou

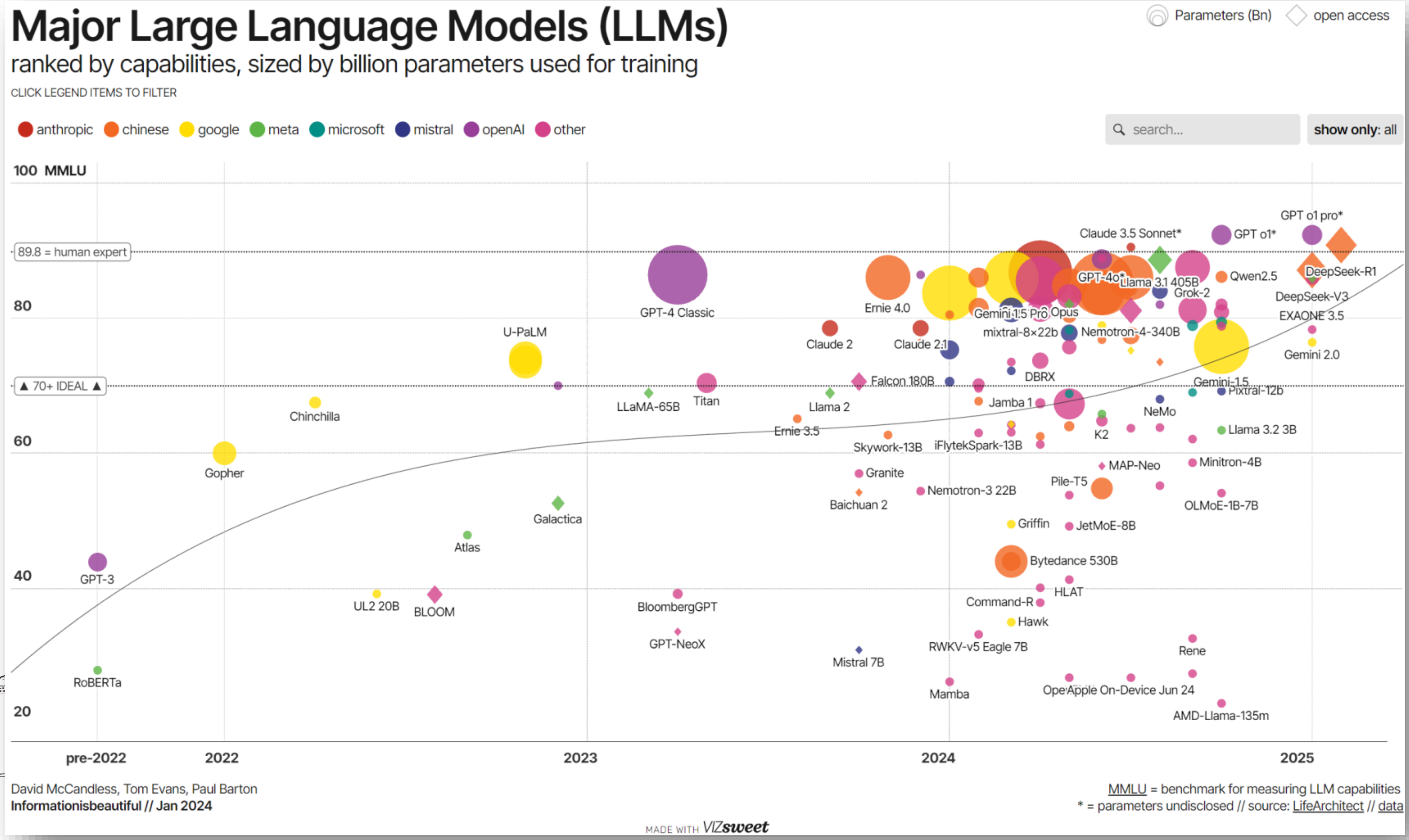
AI Generated

Machine Learning



Machine Learning





Machine Learning and HEP, a long story

https://iml-wg.github.io/HEPML-LivingReview/

Modern reviews

- Snowmass 2021 Computational Frontier CompF03 Topical Group Report: Machine Learning (2022)
- Artificial Intelligence and Machine Learning in Nuclear Physics [DOI] (2021)
- Machine Learning in the Search for New Fundamental Physics (2021)
- Modern Machine Learning and Particle Physics [DOI] (2021)
- Machine and Deep Learning Applications in Particle Physics [DOI] (2019)
- Machine learning and the physical sciences [DOI] (2019)
- Machine learning at the energy and intensity frontiers of particle physics (2018)
- Machine Learning in High Energy Physics Community White Paper [DOI] (2018)
- Deep Learning and its Application to LHC Physics [DOI] (2018)
- Jet Substructure at the Large Hadron Collider: A Review of Recent Advances in Theory and Machine Learning [DOI] (2017)

Specialized reviews

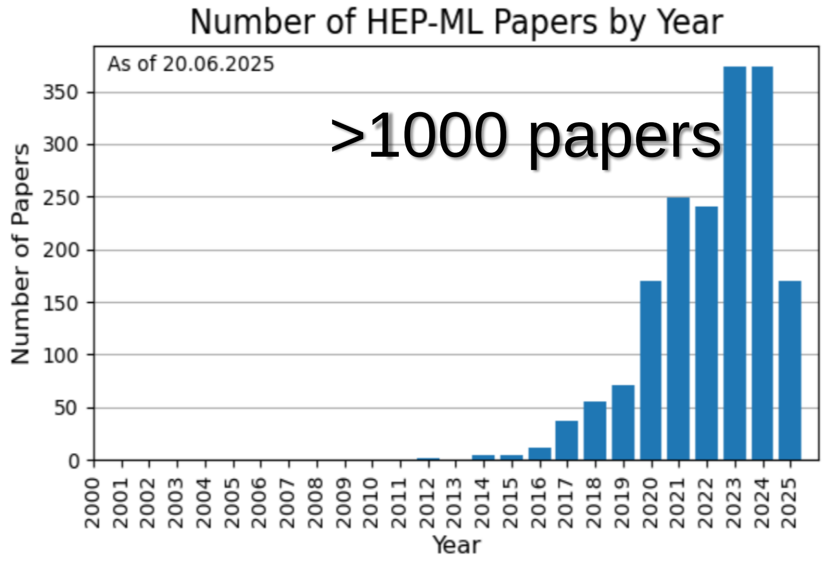
- Review of Machine Learning for Real-Time Analysis at the Large Hadron Collider experiments ALICE, ATLAS, CMS and LHCb (2025)
- Lecture Notes on Normalizing Flows for Lattice Quantum Field Theories (2025)
- What is AI, what is it not, how we use it in physics and how it impacts... you (2025)
- Strategic White Paper on AI Infrastructure for Particle, Nuclear, and Astroparticle Physics: Insights from JENA and EuCAIF (2025)
- Machine-learning approaches to accelerating lattice simulations [DOI] (2025)
- Run 3 performance and advances in heavy-flavor jet tagging in CMS [DOI] (2024)
- CaloChallenge 2022: A Community Challenge for Fast Calorimeter Simulation (2024)
- Exploring jets: substructure and flavour tagging in CMS and ATLAS [DOI] (2024)
- Novel machine learning applications at the LHC [DOI] (2024)

Computer Physics Communications 49 (1988) 429–448
North-Holland, Amsterdam

NEURAL NETWORKS AND CELLULAR AUTOMATA IN EXPERIMENTAL HIGH ENERGY PHYSICS

B. DENBY
Laboratoire de l'Accélérateur Linéaire, Orsay, France

Received 20 September 1987; in revised form 28 December 1987



Machine Learning and HEP, a long story

Tasks

1990	2000	2010	2020
Triggers Particle Identification Event selection …	Jet tagging Reconstruction Simulation …	Anomaly detection Simulation-based inference Foundation models	

Architectures

1990	2000	2010	2020
(shallow) ANNs	BDTs	DNNs	Transformers

Machine Learning and HEP, a long story

Tasks

Tri
Pa
Ev
…

1990

Architec

(shallow) A

nsformers

ML applications at BESIII

ICHEP2018 SEOUL
XXXIX INTERNATIONAL CONFERENCE ON high Energy PHYSICS
JULY 4 - 11, 2018
COEX, SEOUL

CHEP 2018
Sofia, Bulgaria

International Journal of Modern Physics A | Vol. 34, No. 35, 1930019 (2019) | Reviews

Machine and deep learning applications in particle physics

Dimitri Bourilkov

2.2.1. *Intensity frontier*

At the intensity frontier advanced detectors collect record amounts of luminosity at what would be considered "medium" energies by today's standards. One example is the Beijing Electron Positron Collider (BEPCII) running at center of mass energies 2.0–4.6 GeV. The BESIII experiment has collected record size data samples in this τ-charm region. Advanced ML techniques have been applied for many tasks.[27] One

What’s happening

Tasks

Tri
Pa
Ev
…

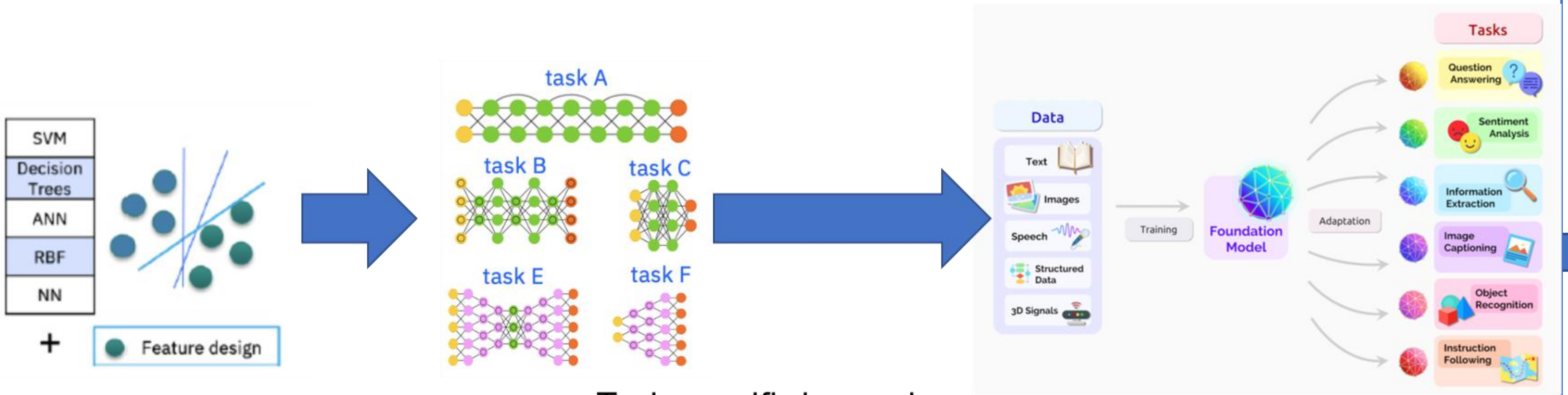
1990

Architec

(shallow) Ar

nsformers

More data, larger models

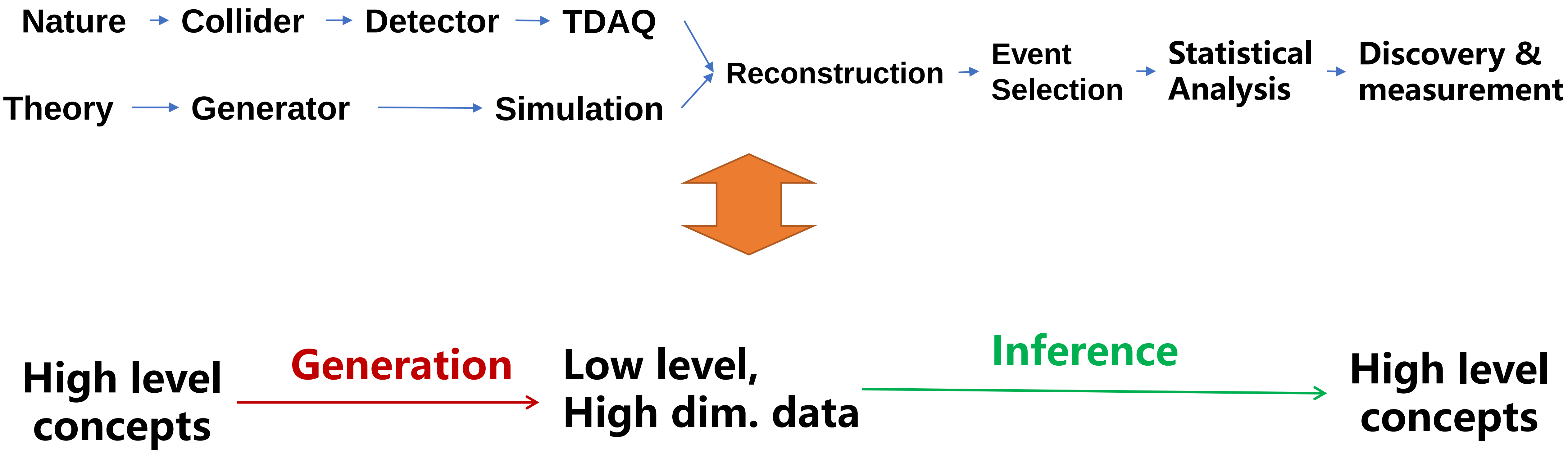


Task-specific hand crafted feature representations

Task-specific learned feature representations

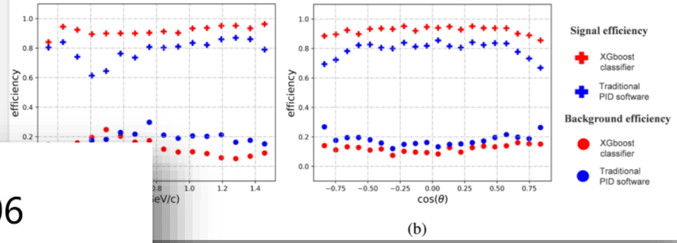
Generalizable & adaptable learned representations

AI+HEP analysis: a huge inverse problem

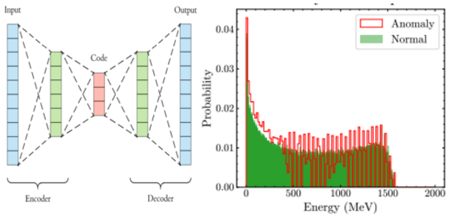


A few examples @ BESIII

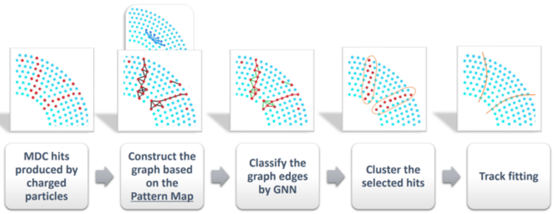
Particle identification
EPJ Web Conf. 295, 09027



Anormaly detection
EPJ Web Conf. 337, 01158



Track reconstrution
EPJ Web Conf. 295, 09006



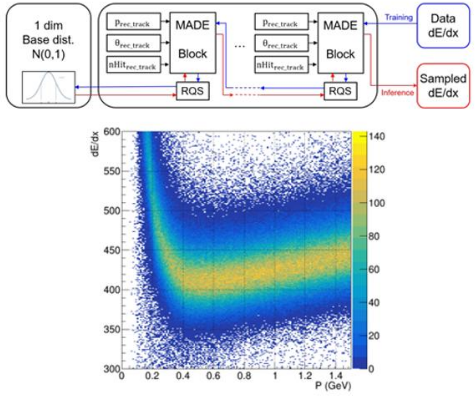
HEP:

Nature → Collider → Detector → TDAQ

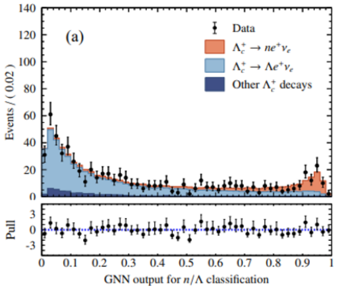
Theory → Generator → Simulation

Reconstruction → Event Selection → Statistical Analysis → Discovery & measurement

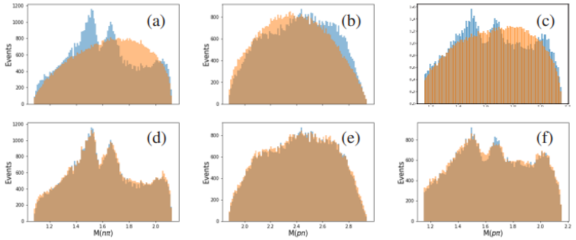
Fast simulation
NIM.A 1065,169544



Signal/Background separation
Nat Comm. 16, 681



Multi-dim. Background modeling
EPJ Web Conf. 214, 06033



AI:

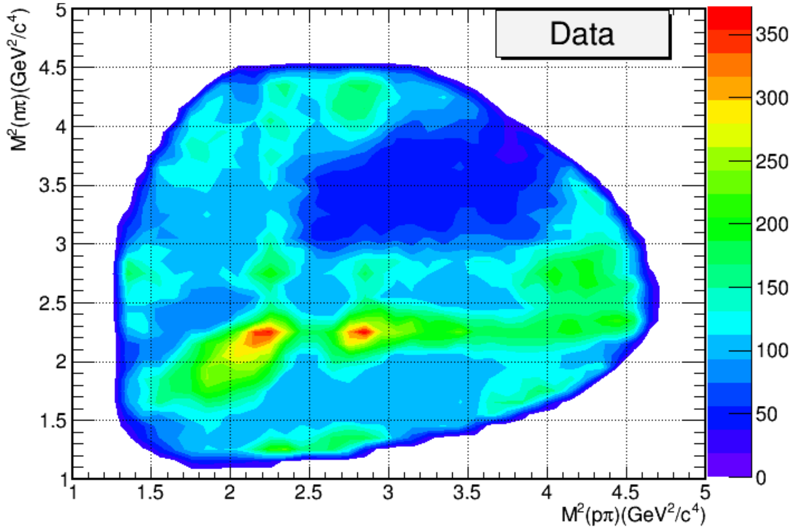
High level concepts Gene ... ve., ... im. data ... er ... **High level concepts**

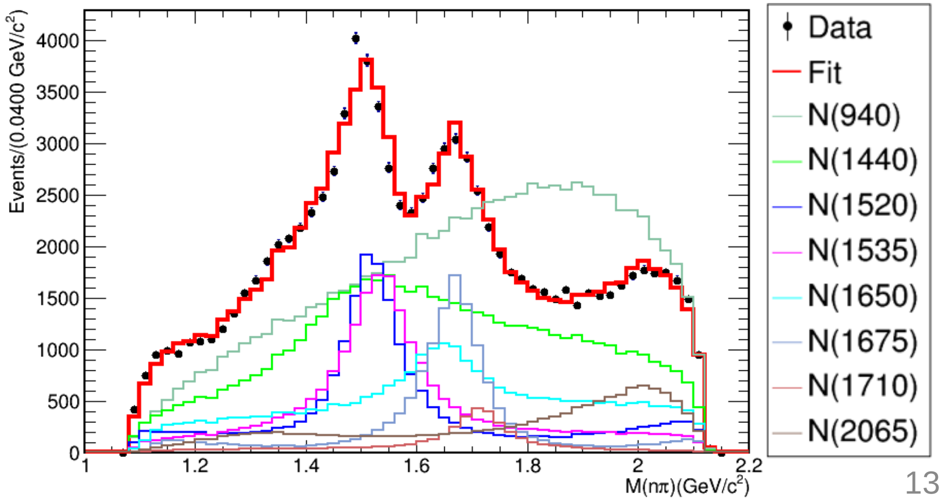
Declaimer: a random selection of published BESIII works

Example: Reweighting: to create data-like MC

$J/\psi \to \bar{p} n \pi^+$

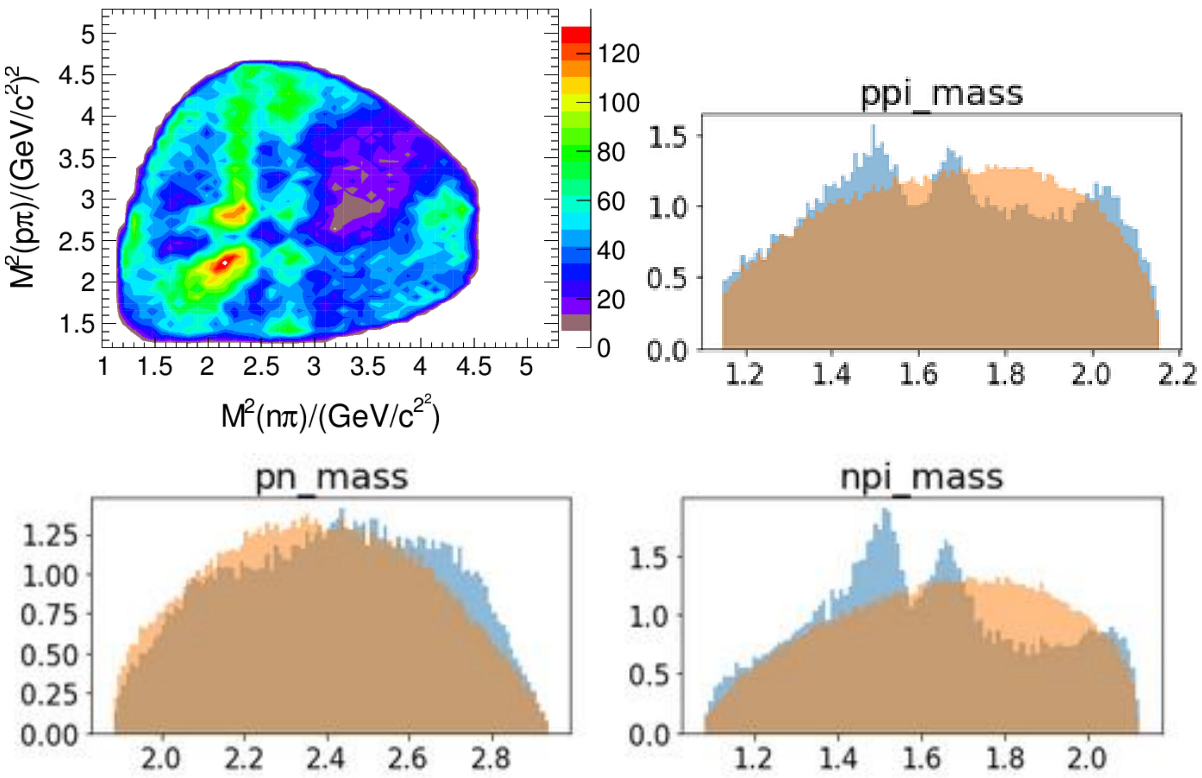
- Rich structures in data of hadron spectroscopy cannot be described by generic MC
- Amplitude analysis is mandatory to extract the resonances
- Data-like MC is useful to calculate efficiency or to estimated backgrounds



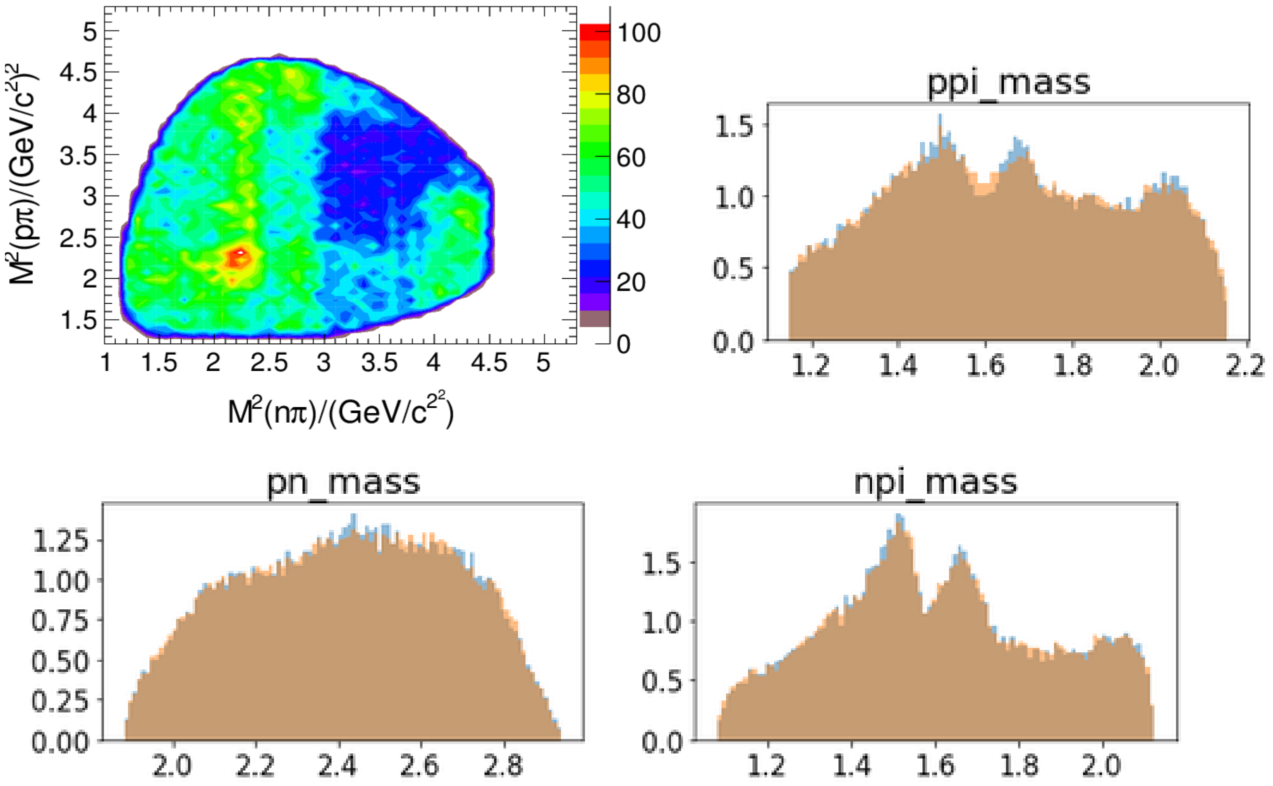


[EPJ Web Conf. 214, 06033]

Original distribution



After reweighting



- A classifier to distinguish Data and MC provide probabilities $p_{RD}(x)$ and $p_{MC}(x)$[1]

$$weight\ factor\ w(x) = \frac{f_{RD}(x)}{f_{MC}(x)} \sim \frac{p_{RD}(x)}{p_{MC}(x)}$$

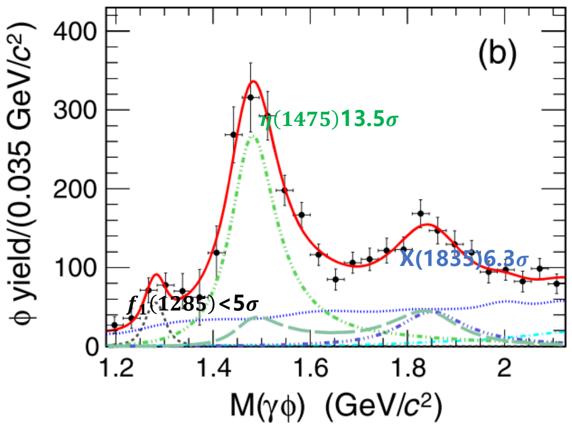
- We utilize this approach with XGBoost algorithm

[1] Martschei, D., Feindt, M., Honc, S., & Wagner-Kuhr, J. (2012). Advanced event reweighting using multivariate analysis. *Journal of Physics: Conference Series*, *368*(1), 012028.

$J/\psi \to \gamma\gamma\phi$, a $s\bar{s}$ flavor filter

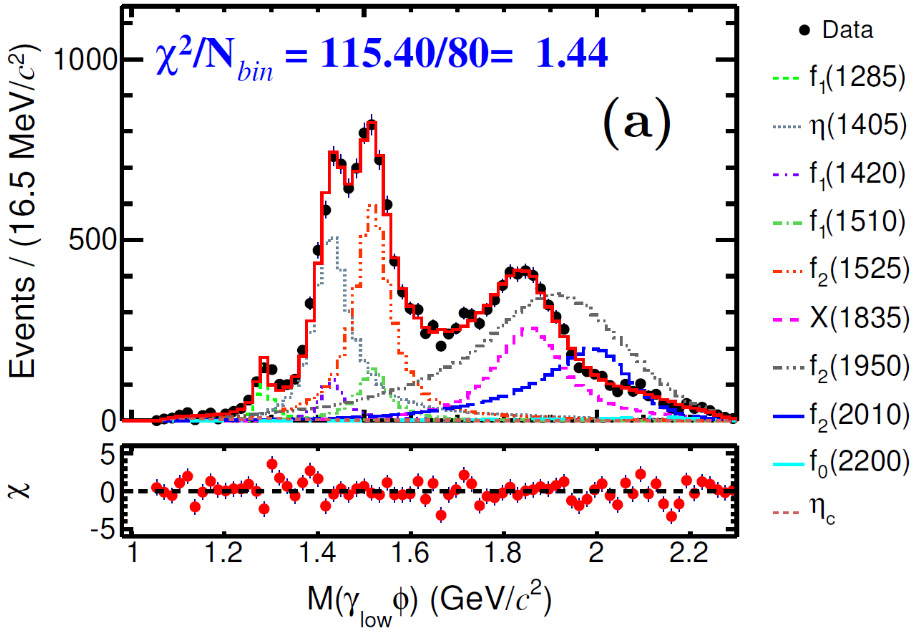
BESIII PhysRevD.111.052011(2025)

PR D97 051101 (2018)



Amplitude analysis with ML techniques for background subtraction

←Fit to mass spectrum



From the amplitude analysis,

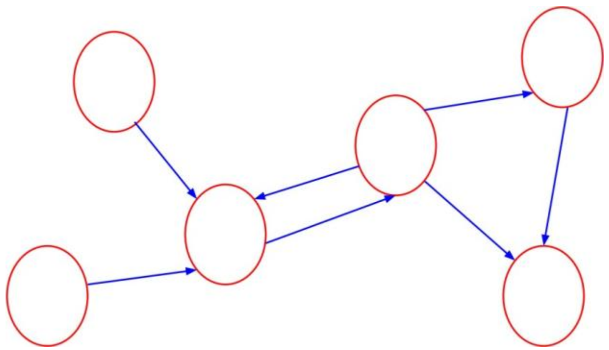
- $\eta(1405)$ is observed, while $\eta(1475)$ can not be excluded
- $X(1835) \to \gamma\phi$ suggests its assignment of η' excitation
- $\eta_c \to \gamma\phi$ are observed. The very first radiative decay mode of η_c
- Observation of $f_2(1950)$ and $f_0(2200) \to \gamma\phi$ unfavored their glueball interpretations [PRD 108, 014023, Sci.China Phys.Mech.Astron. 67 (2024) 11, 111012]

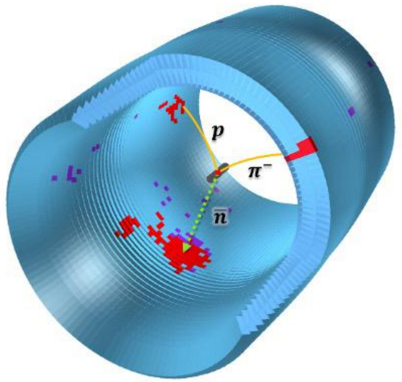
More than proof-of-concept, but in production:
PRL 134 (2025) 201902, PRL 136 (2025) 161902,
PRD 103 (2021) 012009, arXiv:2510.25111, arXiv:2501.04451, etc

Example: Hunting for $\Lambda_c \to n e^+ \nu$

[Nat. Comm. 16, 681]
Slides courtesy to Prof. Xiaorui Lv

- **Challenges:**
  - ✓ neutrino is missing in detection
  - ✓ dominant backgrounds from $\Lambda_c^+ \to \Lambda(\to n\pi^0)e^+\nu$, with ~10x yields than that of the pursuing signals
  - ✓ elusive neutron detection due to neutral charge and contaminations from the photon showers (& noises) in electro-magnetic calorimeter (EMC)





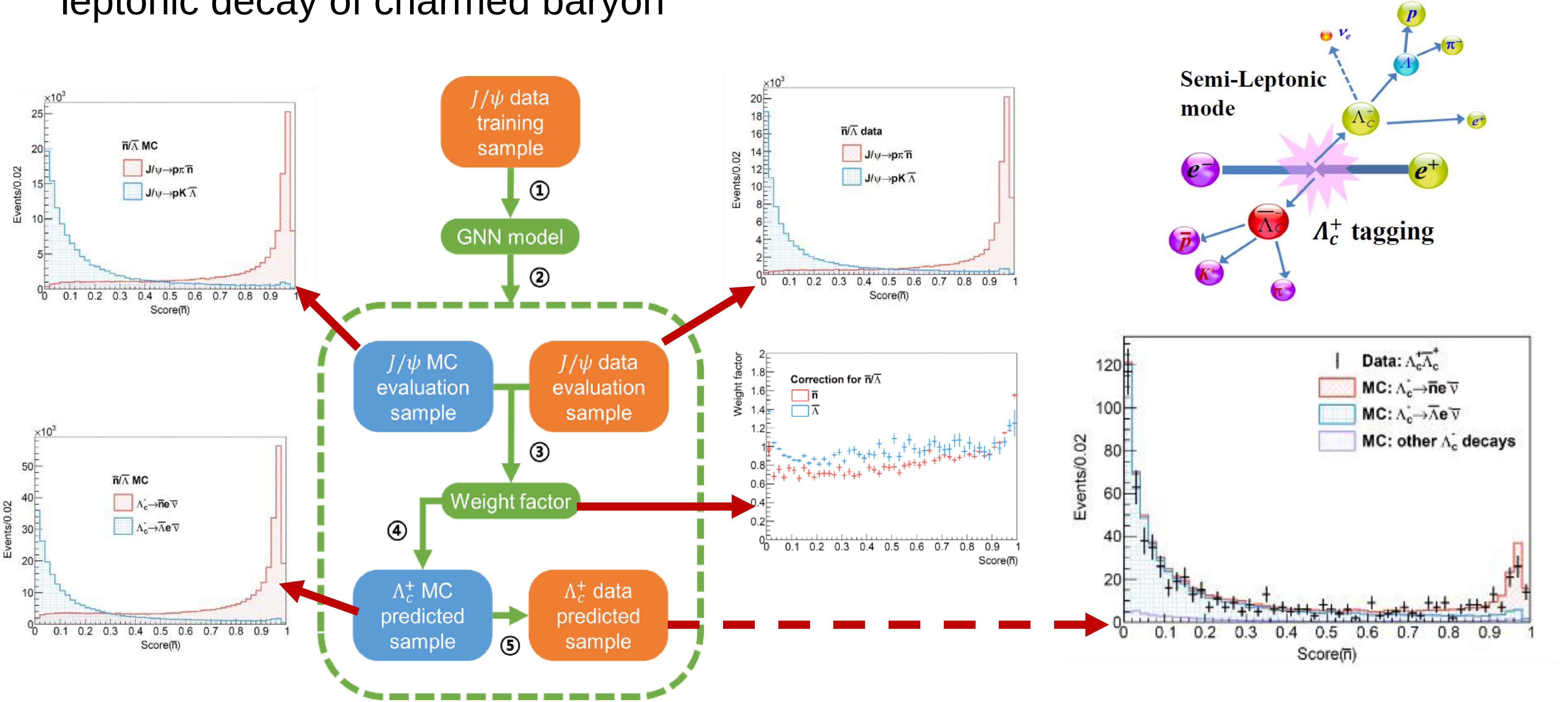
- Sharing of parameters across node and edge updates in the graph.
- Permutation invariance

- Nearly unlimited labeled samples
- Structured data
- Clear training objectives

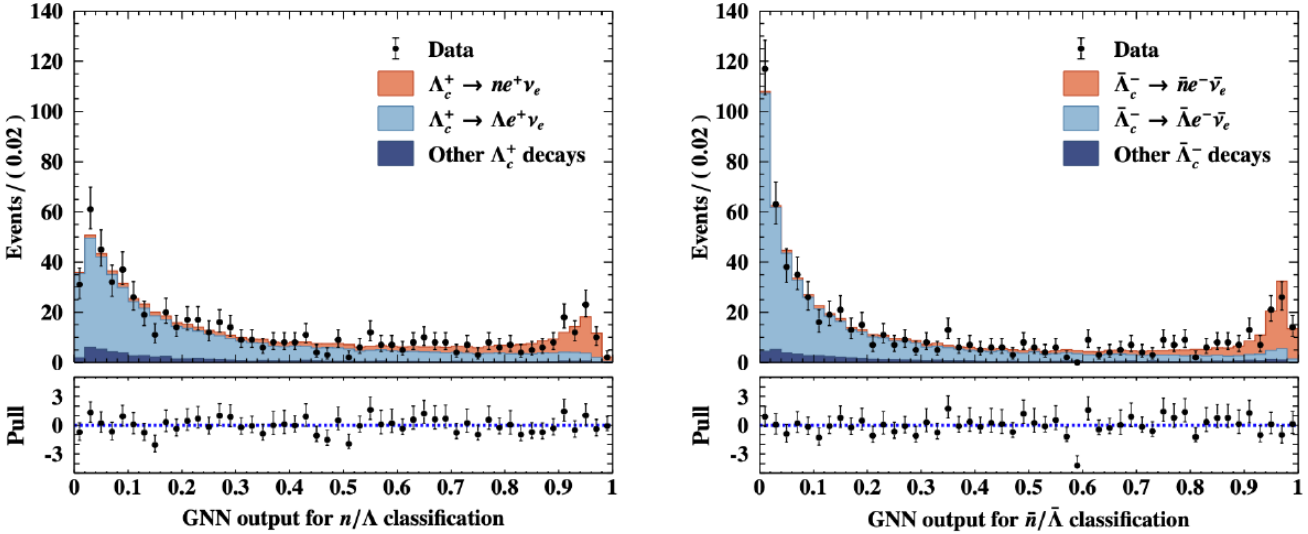
ParticleNet: Jet Tagging via Particle Clouds Phys.Rev.D 101 (2020) 5, 056019

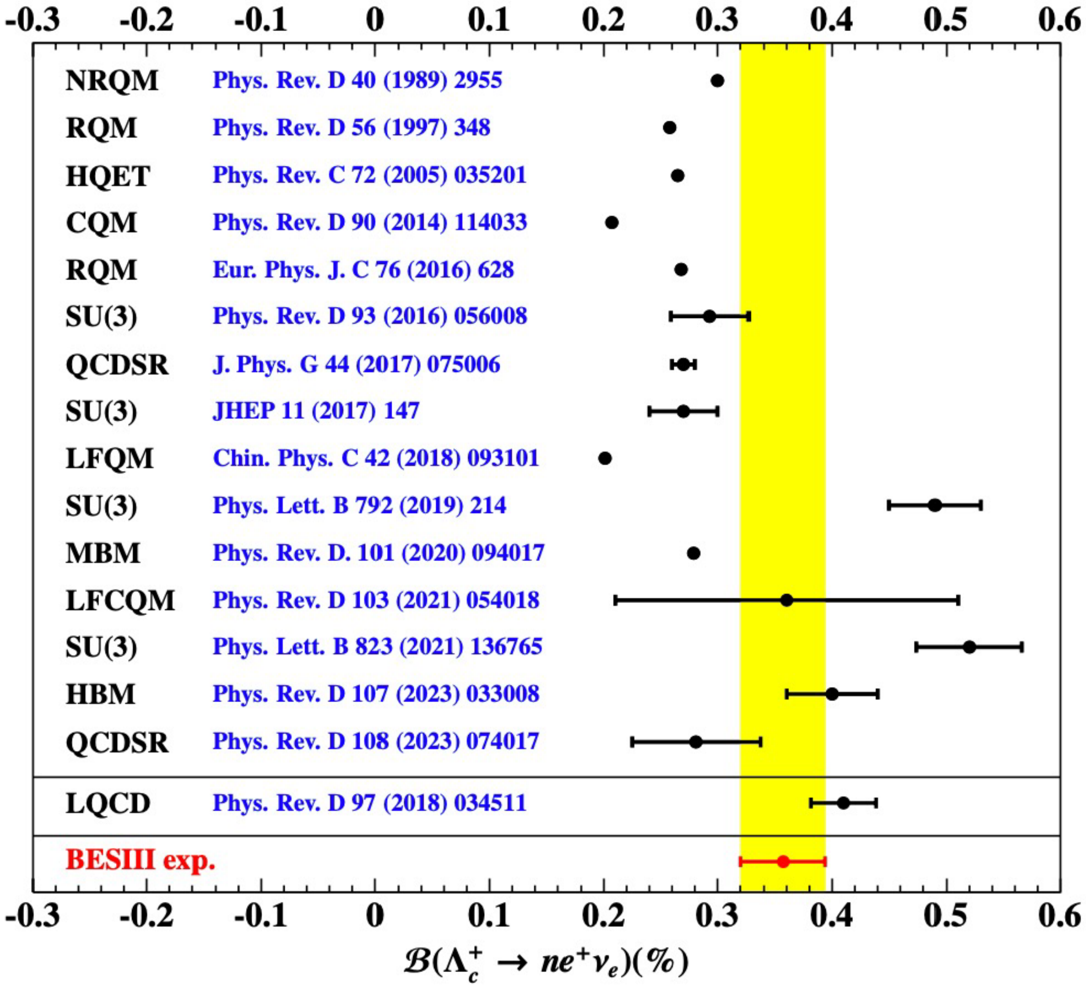
Machine-learning-driven observation of the elusive semi-leptonic decay of charmed baryon

[Nat. Comm. 16, 681]



[Nat. Comm. 16, 681]





$$\mathcal{B}(\Lambda_c^+ \to ne^+\nu_e) = (0.357 \pm 0.034_{\text{stat.}} \pm 0.014_{\text{syst.}})\%, \quad (>10\ \sigma)$$

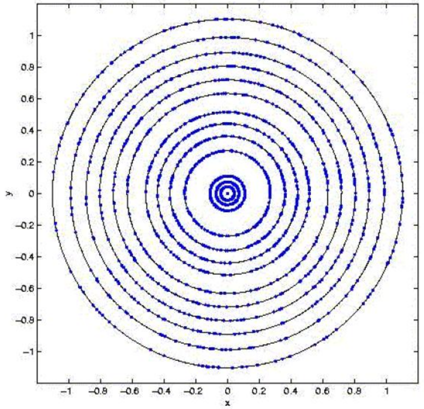
good control of systematics on GNN training

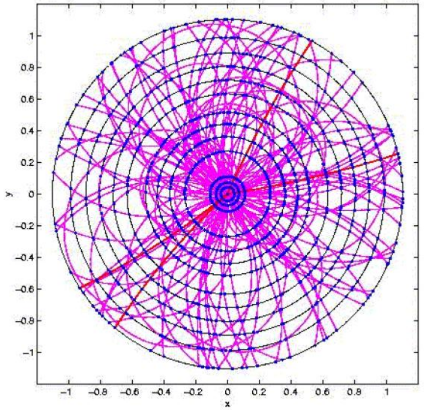
- **Model settings**: network weight initialization, batch processing sequence and dropout layer are randomly varied
- **Domain shift**: validation of independent control sample via $J/\psi \to \Sigma^+(n\pi^+)\overline{\Sigma}^-(\overline{p}\pi^0)$ and $J/\psi \to \Xi^-(\Lambda\pi^-)\overline{\Xi}^+(\overline{\Lambda}\pi^+)$

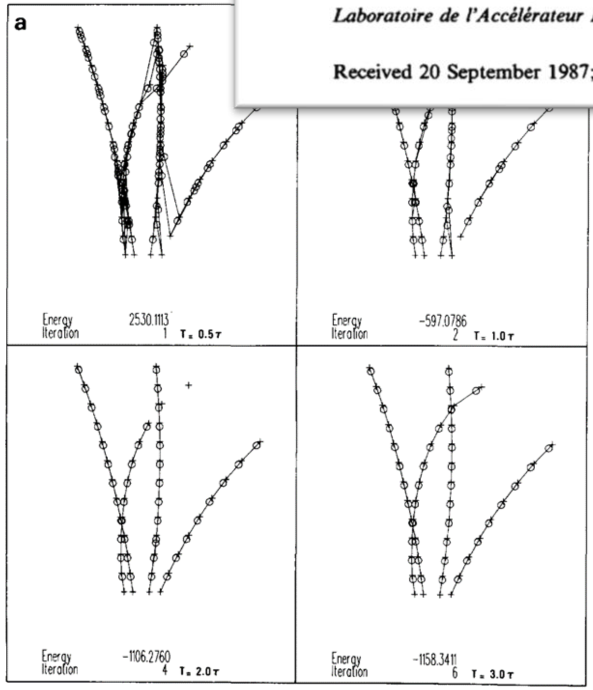
Where it begins

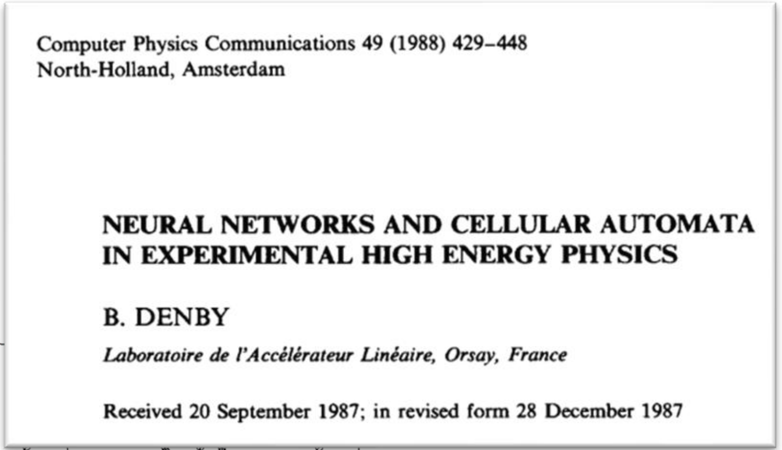
Tracks from bubble chambers and cloud chambers typically had to be inspected by eye. In this June 1984 image, Renee Jones, a bubble chamber scanner working at Fermilab, measures the details of the tracks, including length and curvature.

DAVID PARKER/SCIENCE SOURCE








Computer Physics Communications 49 (1988) 429–448
North-Holland, Amsterdam

NEURAL NETWORKS AND CELLULAR AUTOMATA IN EXPERIMENTAL HIGH ENERGY PHYSICS

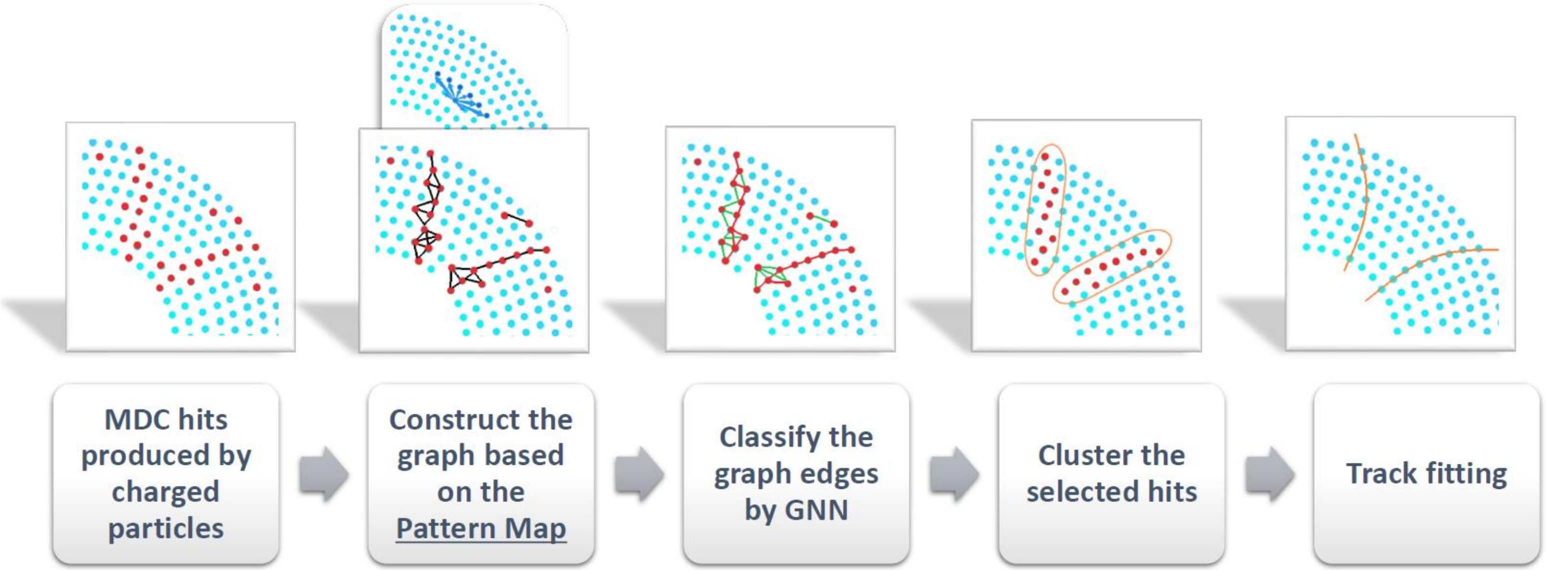
B. DENBY

Laboratoire de l'Accélérateur Linéaire, Orsay, France

Received 20 September 1987; in revised form 28 December 1987

Example: MDC tracking with GNN

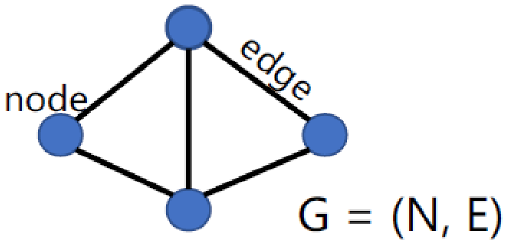
[EPJ Web Conf. 295, 09006]

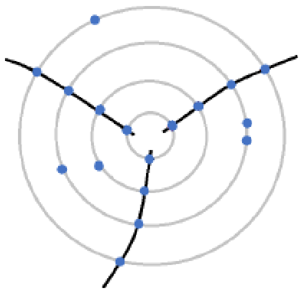


Graph: nodes, edges

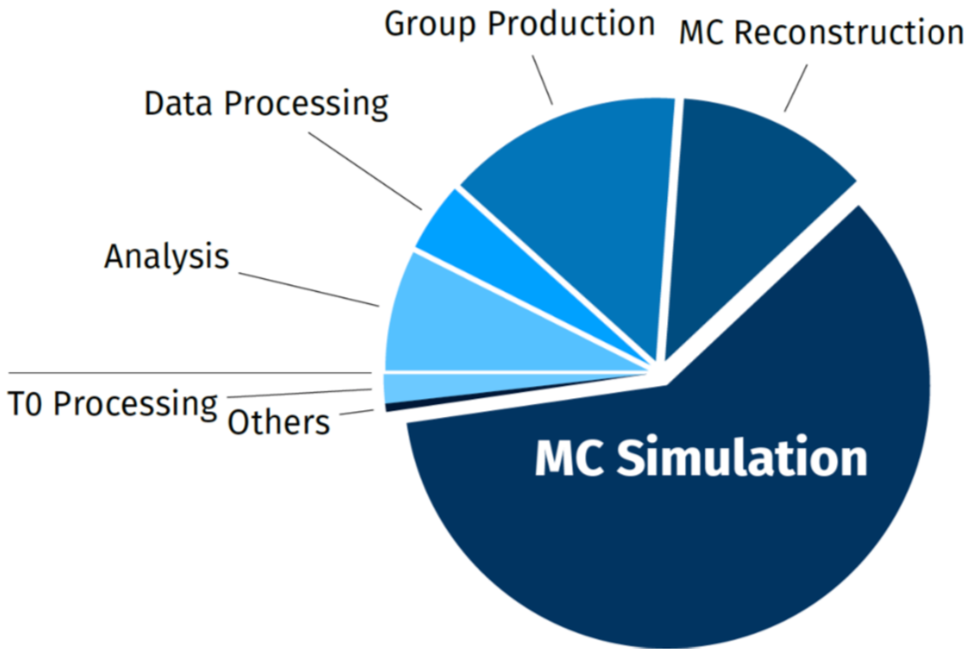
Graph → Track

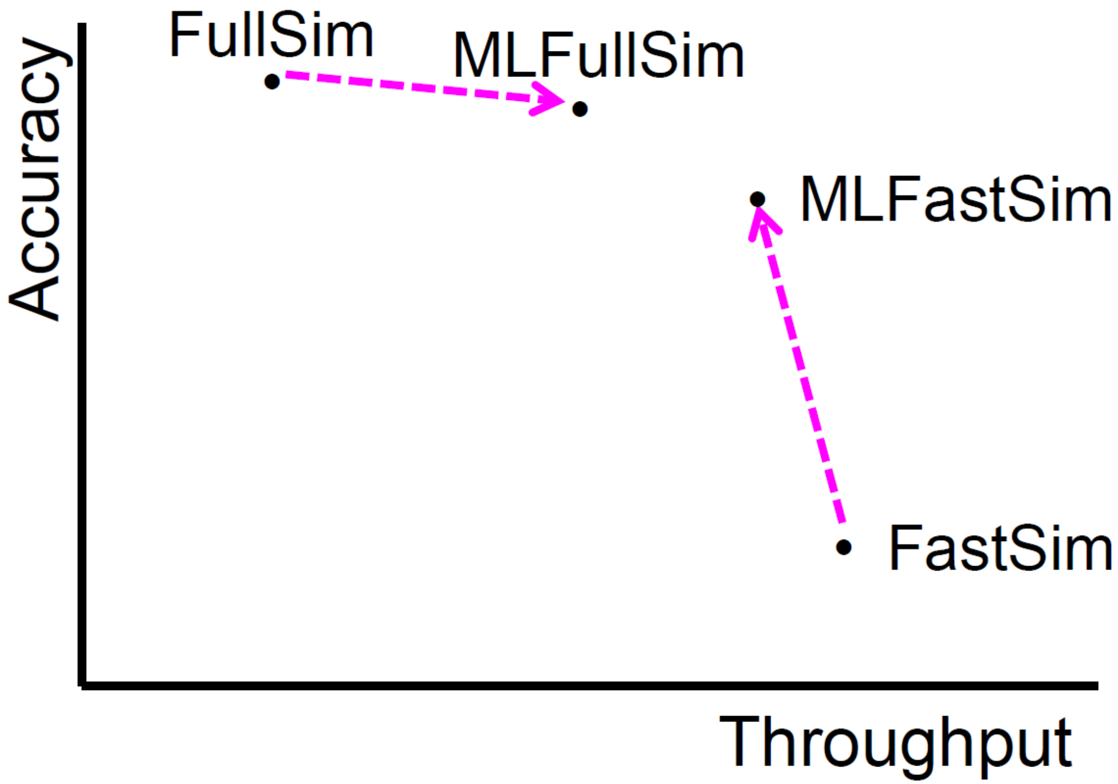
- Nodes → Hits
- edges → track segments



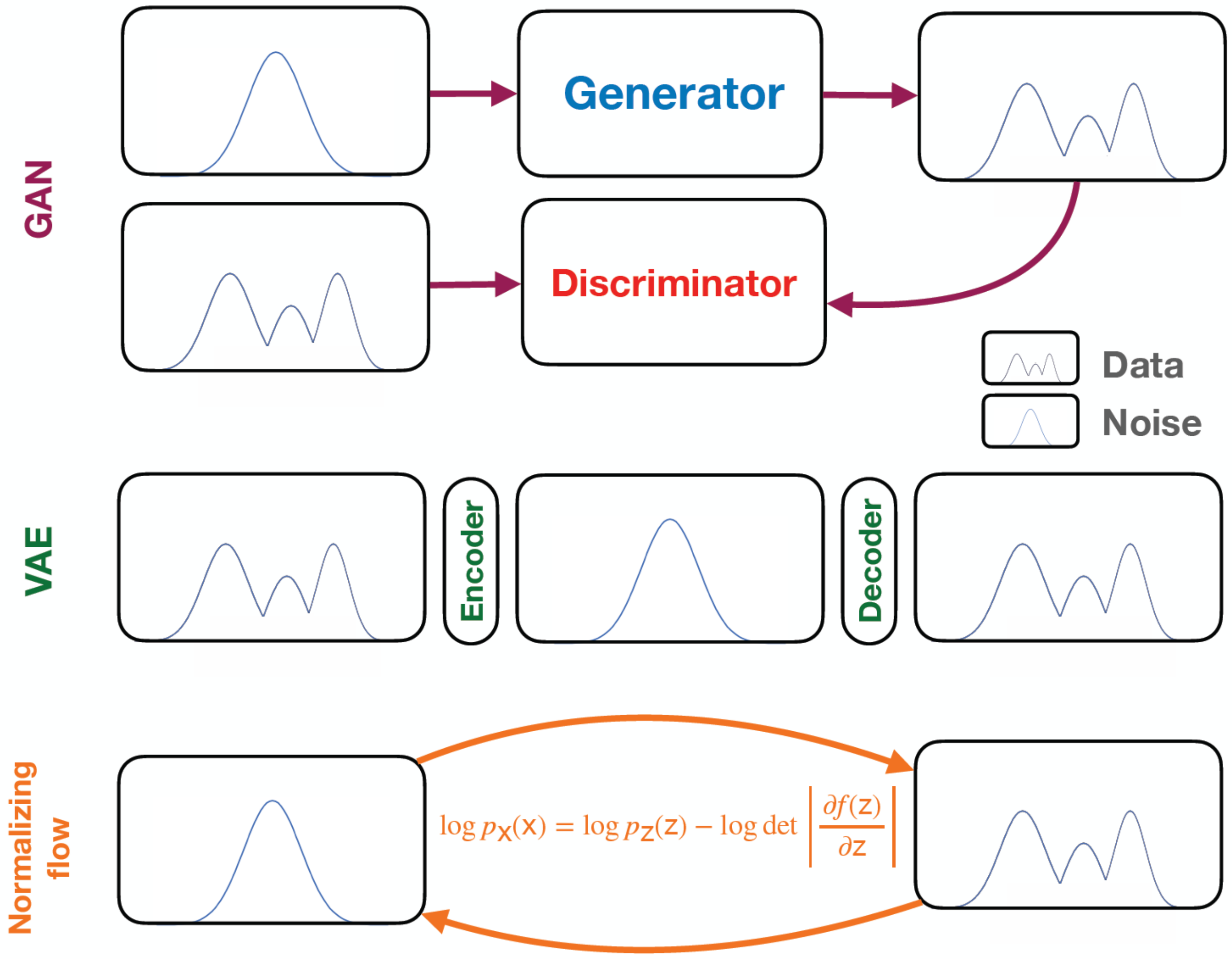


ML as Surrogate Models for simulation





Deep Generative Models



Learn p(x) implicitly

Learn p(x) explicitly

Example: A data-driven dE/dx simulation with normalizing flow

[NIM.A 1065, 169544]

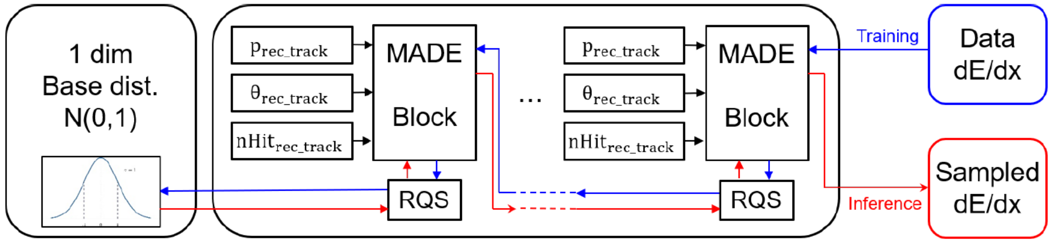


Figure 1. Schematic view of the Normalizing Flow model. The training process follows bl
arrows, inference process (or simulation) follows red arrows.

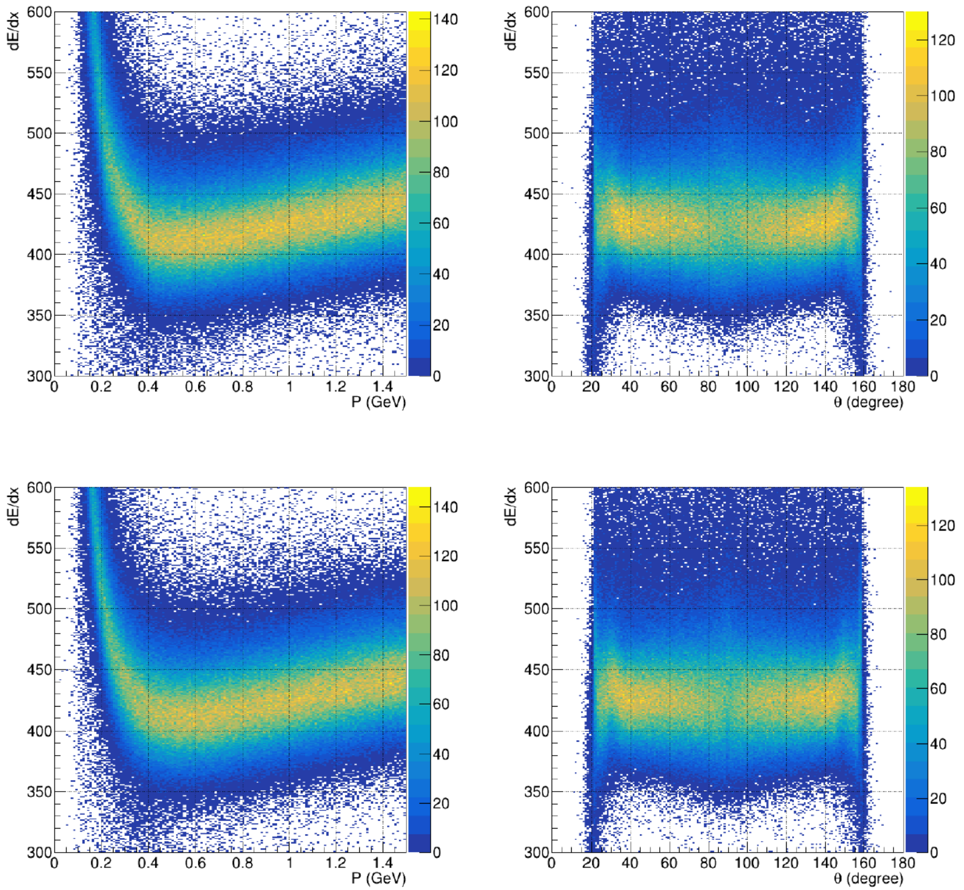


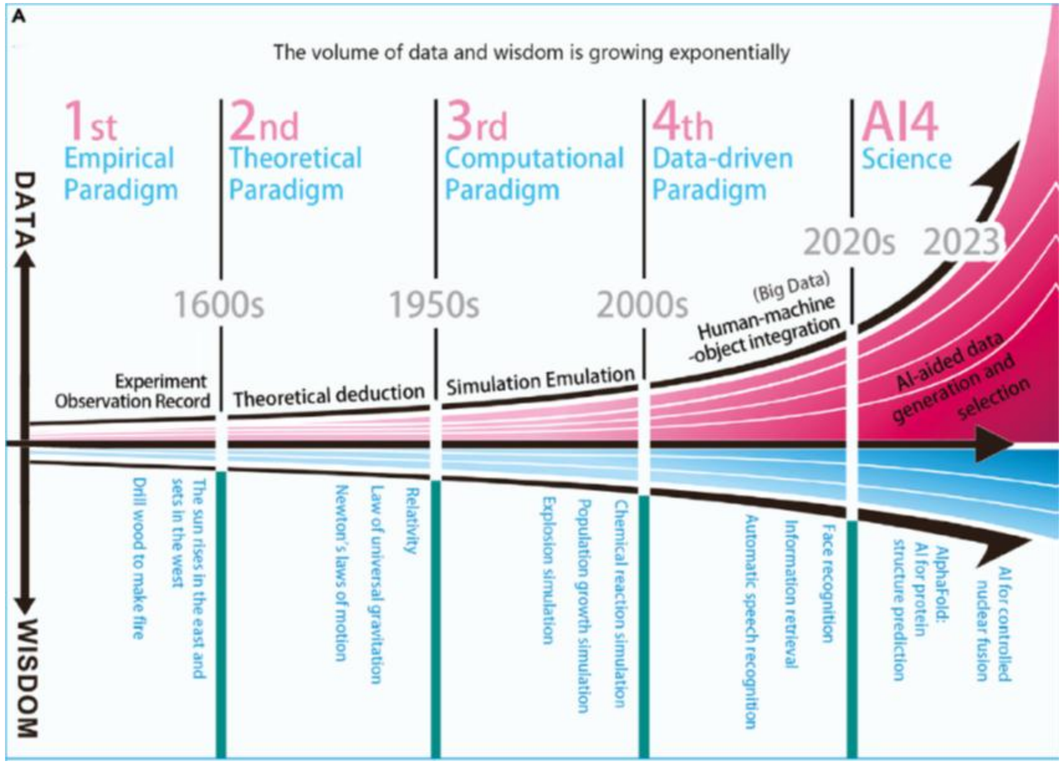
Figure 2. The dE/dx distribution of π^+. The left (right) plots are dE/dx versus momentum (θ). The top (bottom) plots for simulated (data).

What's next

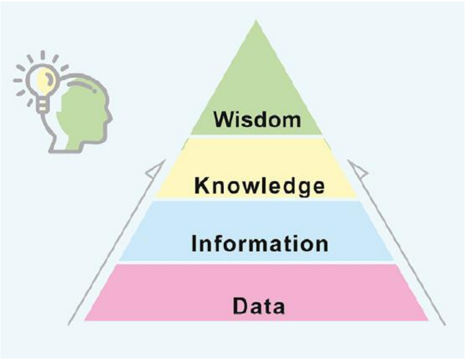
Tasks

Tri
Pa
Ev
…

1990



The Innovation 4(6): 100525



High dim.,
Huge,
Structured,
HEP Data sets
→ ?

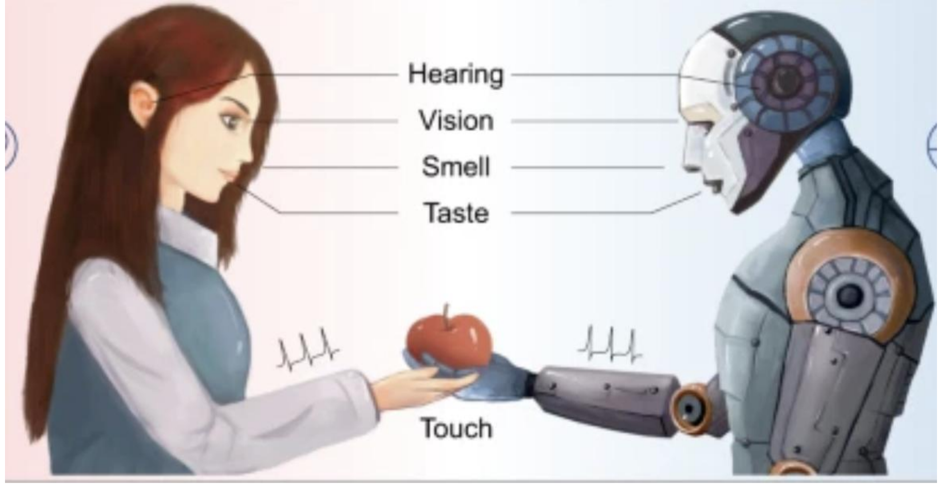
Architec

(shallow) Ar

…nsformers

Data with many forms and modalities

Sparse, Heterogeneous, Multi-scale measurements



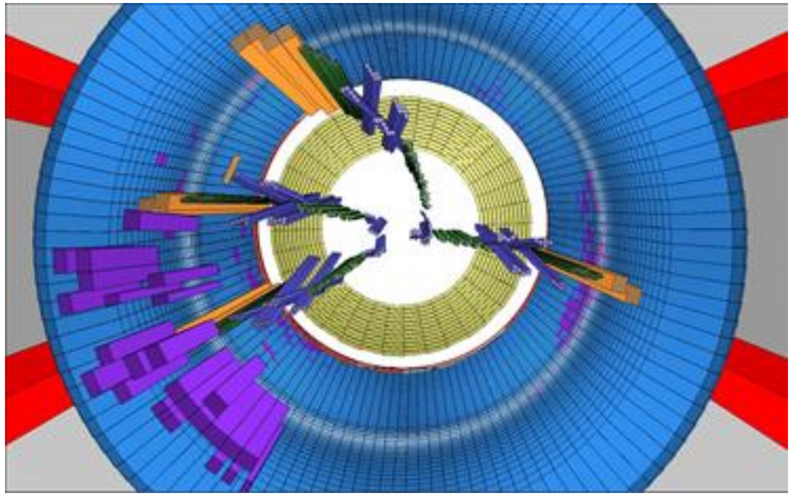
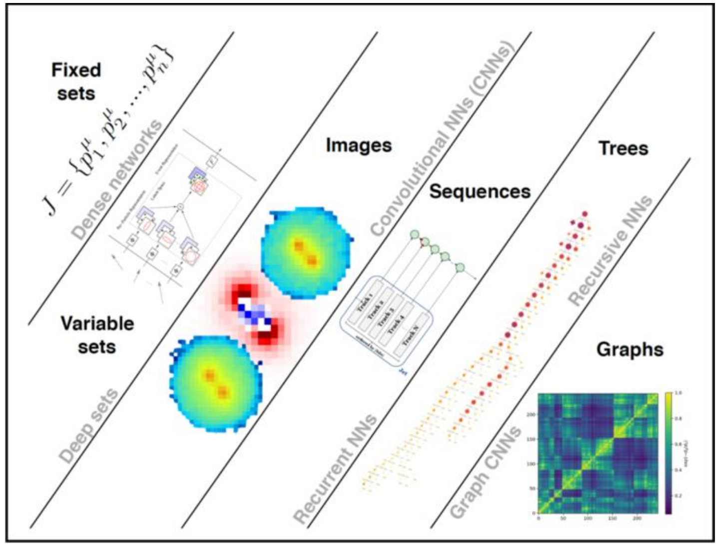


Figure from Nature Comm. 12, 1120 (2021)

Challenge of Representations

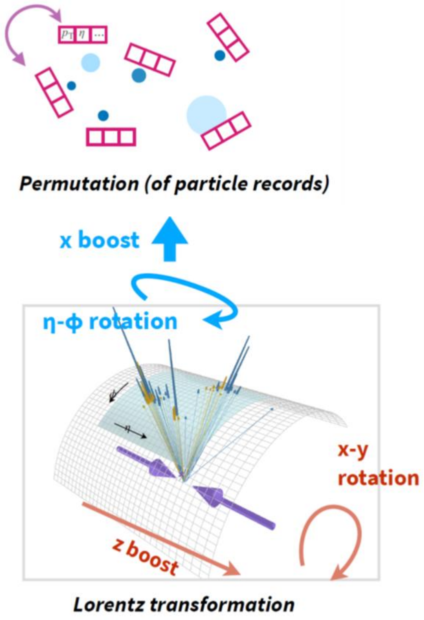
Rich streams of data to be combined in a uniform and distilled representation serving a large variety of downstream tasks
→ Foundation models

Industrial



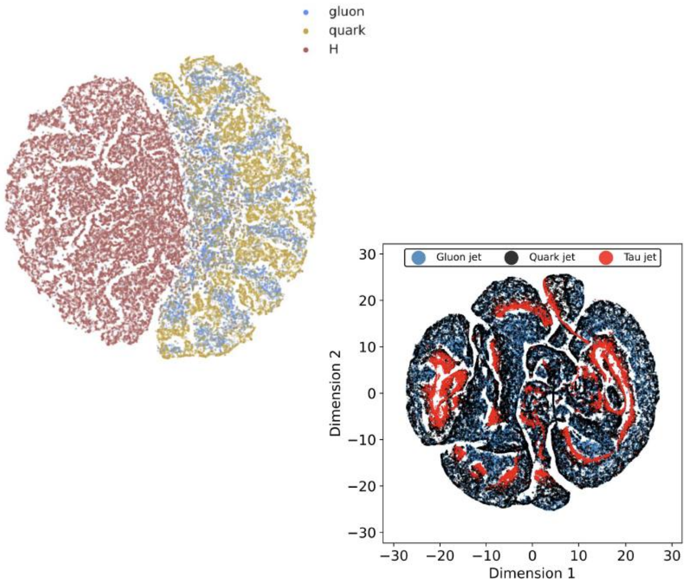
courtesy to B. Nachman

Physics informed



courtesy to C. Li

Self-supervised Learning



arXiv:2403.07066, 2503.11632

The Power of Scale

Should artificial intelligence be interpretable to humans?

Matthew D. Schwartz

Nature Reviews Physics **4**, 741–742 (2022) | Cite this article

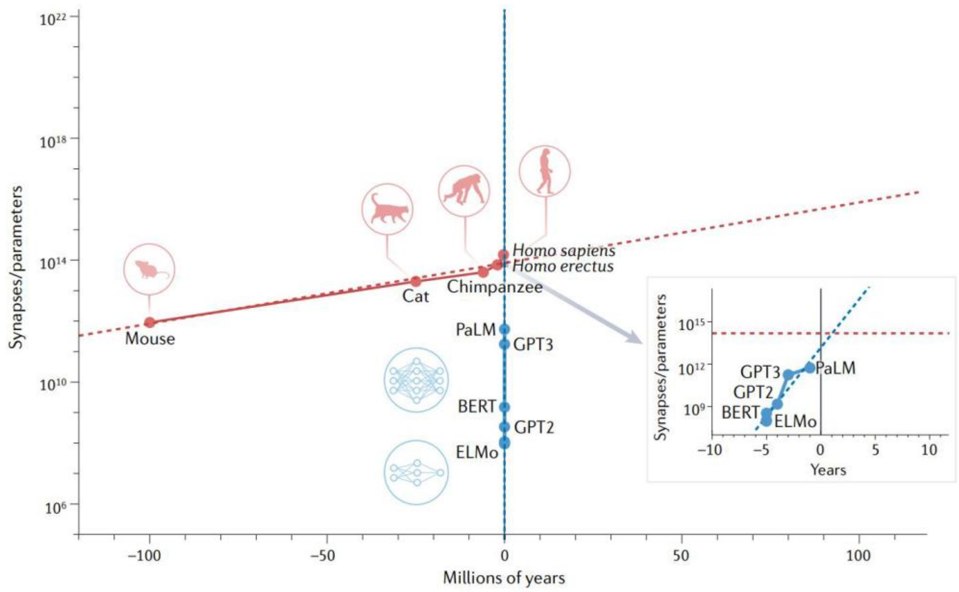


Fig. 1 | The evolution of biological and artificial intelligence takes place on dramatically different timescales. Any hope of interpreting and understanding AI will exponentially fade. Some example data points are highlighted in the evolution of biological (red) and artificial (blue) intelligence. The dashed lines represent the linear regression to these points. The acronyms in the figure are: Pathways Language Model (PaLM), Embeddings from Language Model (ELMo), Bidirectional Encoder Representations from Transformers (BERT), Generative Pre-trained Transformer (GPT).

With massive corpus data and large models, ChatGPT

- A success in “Predicting The Next Word”
- More over, a compression of world’s knowledge

With massive HEP data and large models, ???

What can we get from an effective compression?

Powered by LLM:
From discriminative tools to generative agents

- Recap

Machine Learning

Fundamental Belief: Universal Approximation Theorems

-- A neural network can effectively approximate a high-dim. function

- AI models to approximate the patterns in data

- AI models to approximate the patterns of human behavior

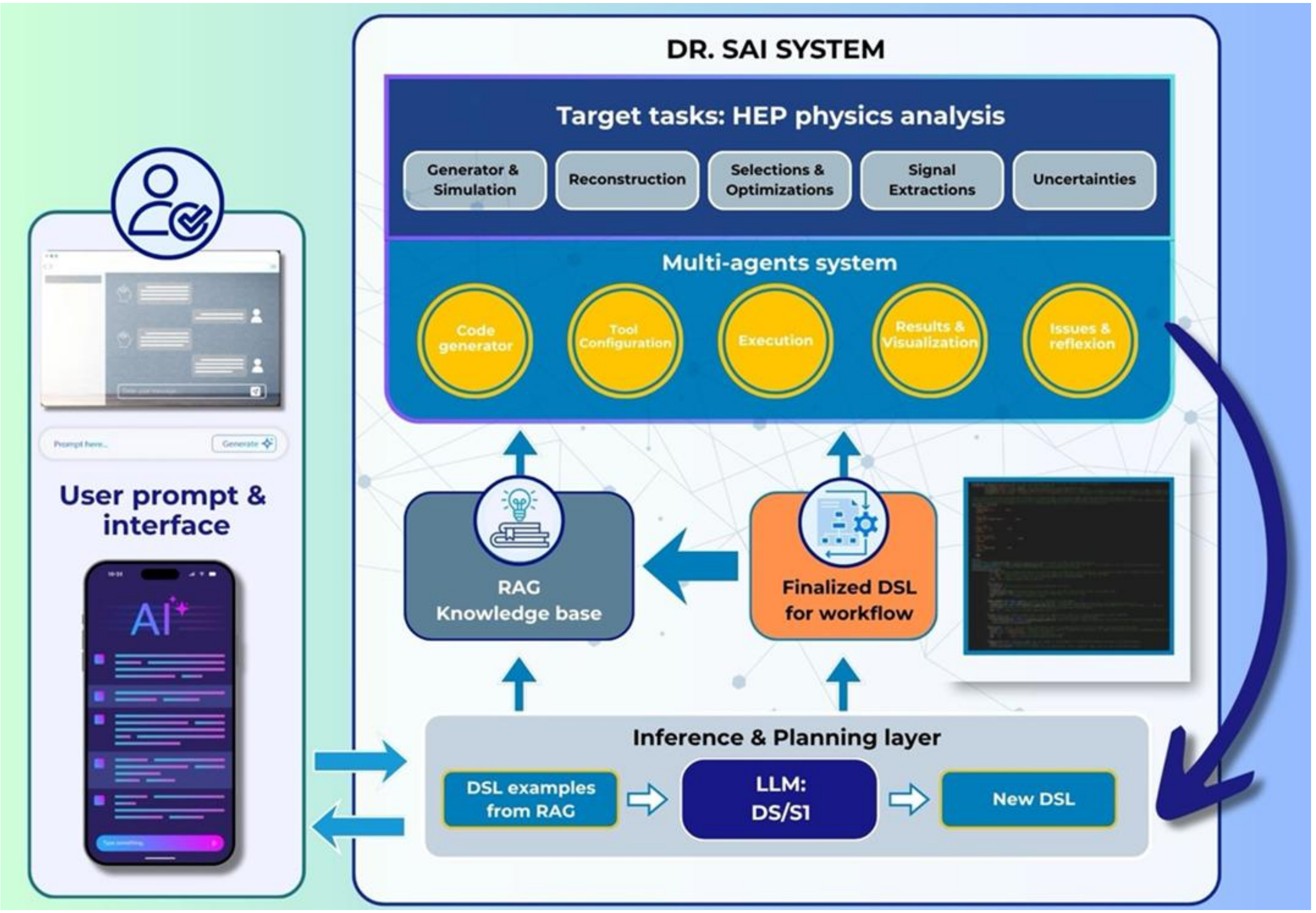
Powered by LLM:
From discriminative tools to generative agents

What are the 5 steps to AGI? From OpenAI

	Name	Description
Level 1	Chatbots	AI with natural conversation language abilities
Level 2	Reasoners	AI's with human-levels of problem solving across a broad range of topics
Level 3	Agents	AI systems that can take actions independently or from human instruction
Level 4	Innovators	AI that can aid in the invention of new ideas and contribute to human knowledge
Level 5	Organizations	AI that is capable of doing all of the work of an organization independently

- Enabling scientists to focus on domain science via automation
- Ultimately increasing efficiency and accelerating scientific discovery

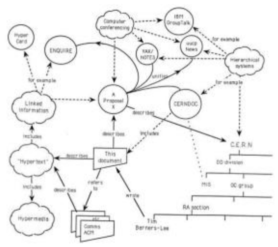
Dr.Sai: Advancing LLM agents for autonomous HEP data analysis



Integrated workflows to accelerate discovery

Outlook

WWW
(1989)

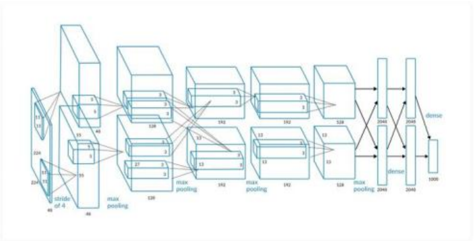


+10 years

Google founded
(1999)

+2 decades

Rising of DNN
(2012)



+10 years

LLM taking off
(2022)

+2 decades

Now

?

BESIII
(2009)

+ 1x years

+2 decades