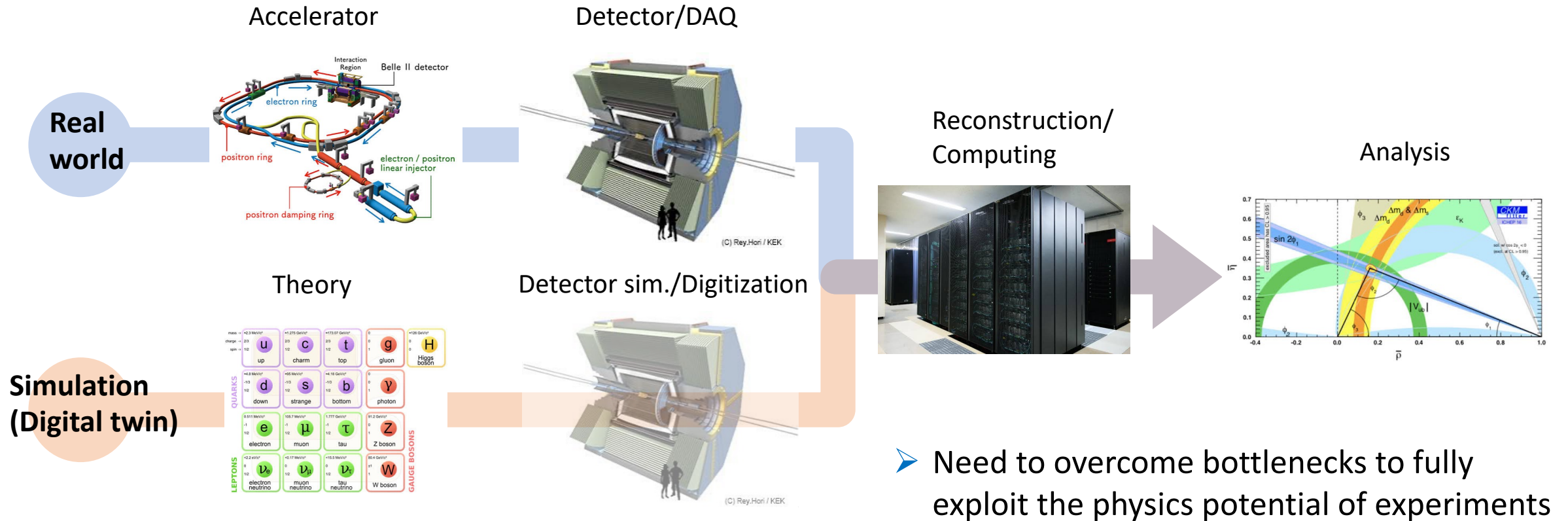


AI Research in Collider Physics in Japan

Tomoe Kishimoto, Computing Research Center, KEK

Collider Physics overview

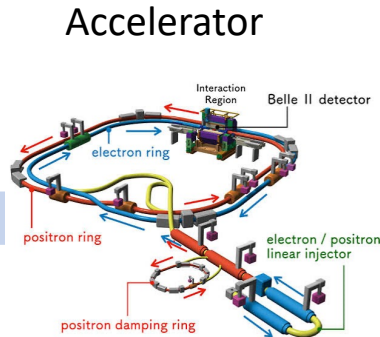


- Need to overcome bottlenecks to fully exploit the physics potential of experiments

Collider Physics overview

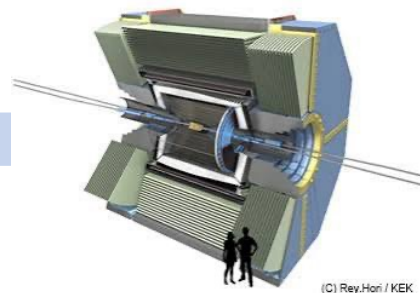
- See Maesaka-san/
Iwasaki-san talks

Real
world

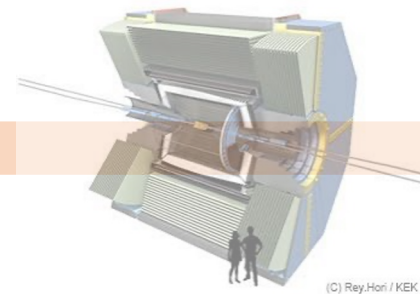


Accelerator

Detector/DAQ



Detector sim./Digitization

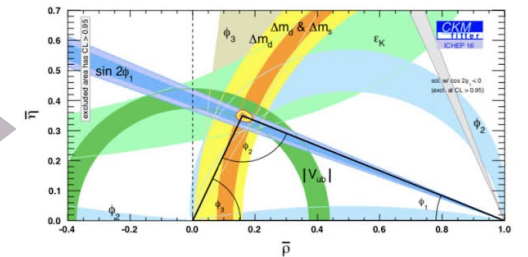


- Calibration
- Trigger/ML on FPGA
- Operations, etc

Reconstruction/
Computing



Analysis



- See Hashimoto-san
talk

Simulation
(Digital twin)

mass ~ 12 MeV/c ²	mass ~ 1.27 GeV/c ²	mass ~ 173.01 GeV/c ²	mass ~ 120 GeV/c ²
charge = 2/3	charge = 2/3	charge = 2/3	charge = 0
spin = 1/2	spin = 1/2	spin = 1/2	spin = 0
u	c	t	g
up	charm	top	gluon
mass ~ 4.2 MeV/c ²	mass ~ 4.18 GeV/c ²	mass ~ 4.18 GeV/c ²	mass ~ 125 GeV/c ²
charge = -1/3	charge = -1/3	charge = -1/3	charge = 0
spin = 1/2	spin = 1/2	spin = 1/2	spin = 0
d	s	b	γ
down	strange	bottom	photon
mass ~ 0.511 MeV/c ²	mass ~ 105.7 MeV/c ²	mass ~ 1.777 GeV/c ²	mass ~ 91.2 GeV/c ²
charge = -1	charge = -1	charge = -1	charge = 0
spin = 1/2	spin = 1/2	spin = 1/2	spin = 0
e	μ	τ	Z
electron	muon	tau	Z boson
mass ~ 0.511 MeV/c ²	mass ~ 105.7 MeV/c ²	mass ~ 1.777 GeV/c ²	mass ~ 80.4 GeV/c ²
charge = 0	charge = 0	charge = 0	charge = 0
spin = 1/2	spin = 1/2	spin = 1/2	spin = 0
ν_e	ν_μ	ν_τ	W
electron neutrino	muon neutrino	tau neutrino	W boson

- Fast simulation
- Detector design, etc

- Tracking/Vertexing
- Jet tagging
- Particle flow, etc

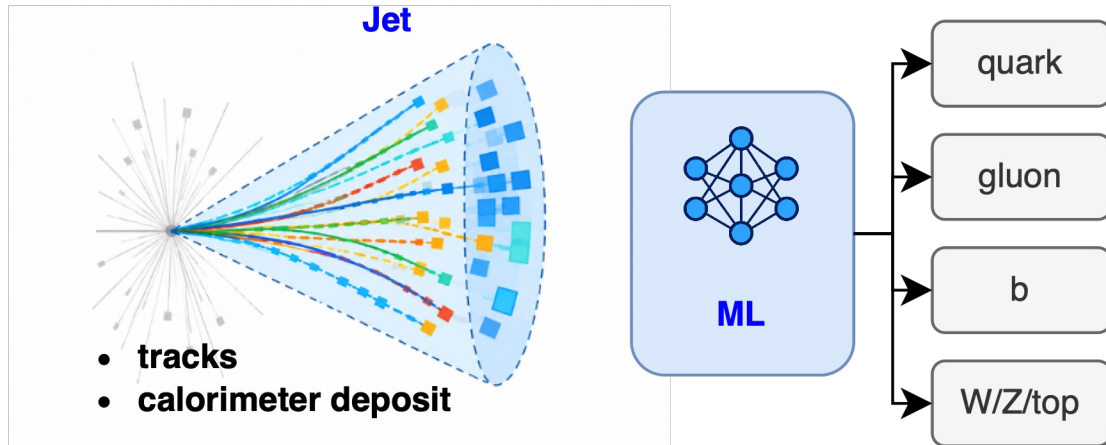
“AI” at every step of the workflow!

加速器だから見える世界。

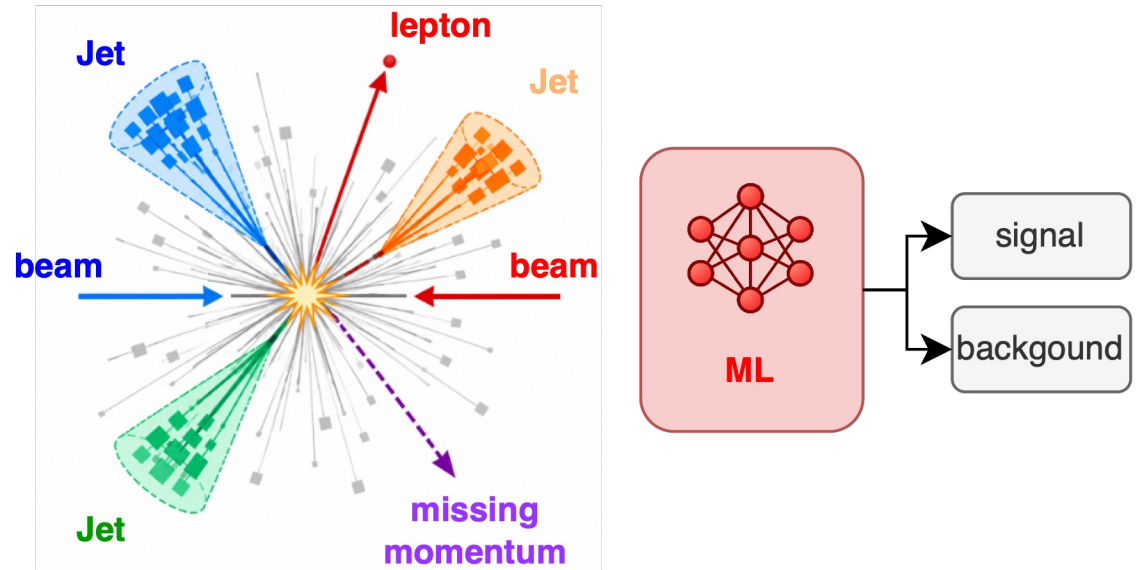


Typical classification tasks

JET CLASSIFICATION



EVENT CLASSIFICATION



→ Off course, there are many more tasks and applications

AI methods in HEP

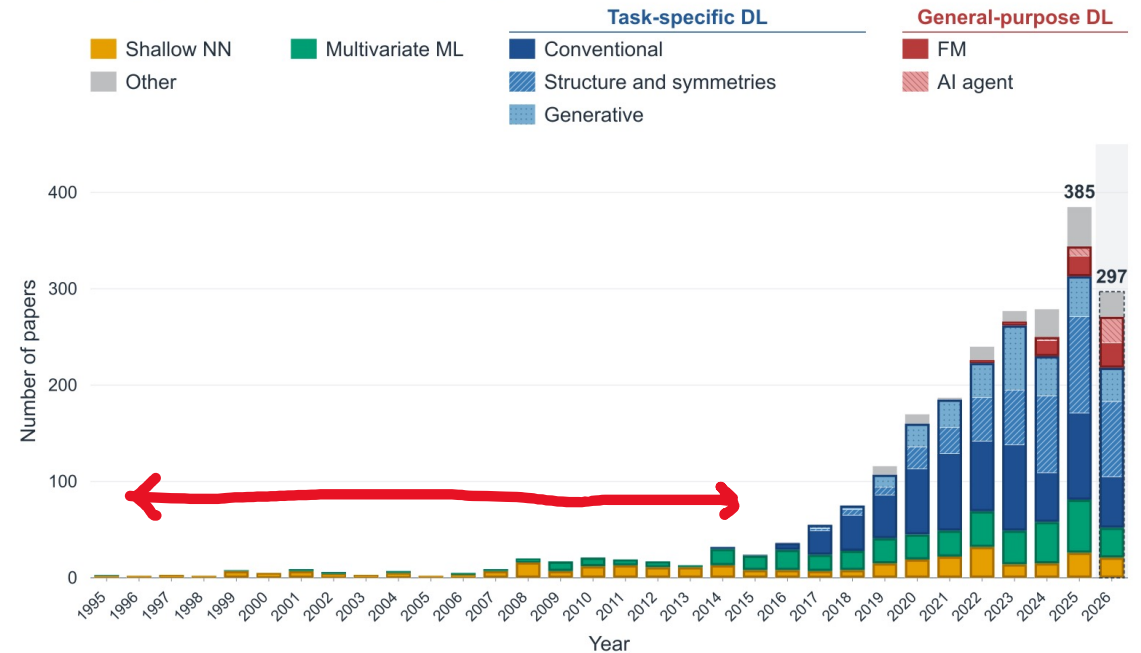
→ Tasks such as jet classification remain largely the same, but the AI methods evolve

- First-generation neural methods (1990s-)
 - Collider physics/HEP has been an early adopter of AI
 - Applying shallow neural networks to hand-engineered features
- Classical multivariate ML (growth in 2009)
 - Multivariate analysis using ensembles, kernel, and tree-based methods
 - Keywords: BDT, SVM, XGBoost, TMVA
 - **Remains a well-established and widely used methods**

arXiv papers (hep-ex)

Annual ML-related paper counts

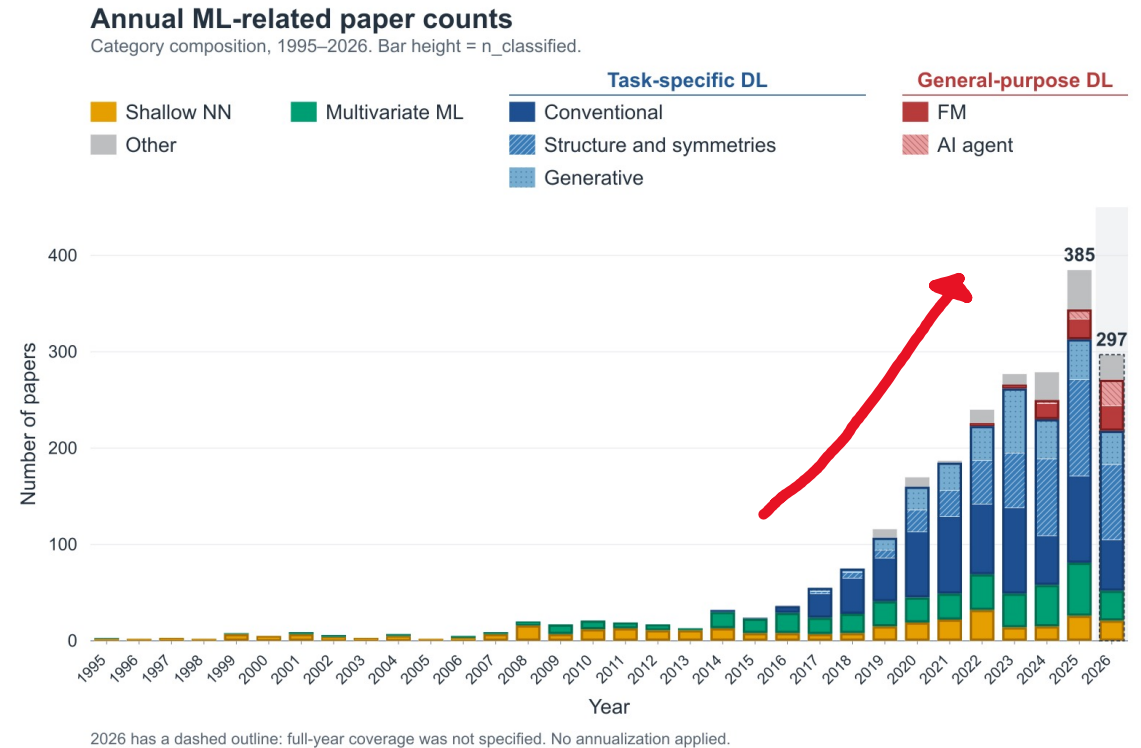
Category composition, 1995–2026. Bar height = n_classified.



AI methods in HEP

- Conventional DL (2014-)
 - Image data, time-series data, dimensionality reduction
 - Keywords: DNN, CNN, RNN, autoencoder
- Structure- and symmetry-aware DL (2017-)
 - Incorporating permutation invariance and other symmetries
 - Keywords: Deep Sets, GNN, point cloud, Transformer
 - **E.g.) Jet analysis has shifted from image-based approaches toward models such as Transformers**
- Generative scientific DL (2017-)
 - Fast event generation and detector simulation
 - Keywords: GAN, flow, diffusion, VAE

arXiv papers (hep-ex)



→ These methods have been primarily used for “specific tasks”, such as jet-tagging

General-purpose AI

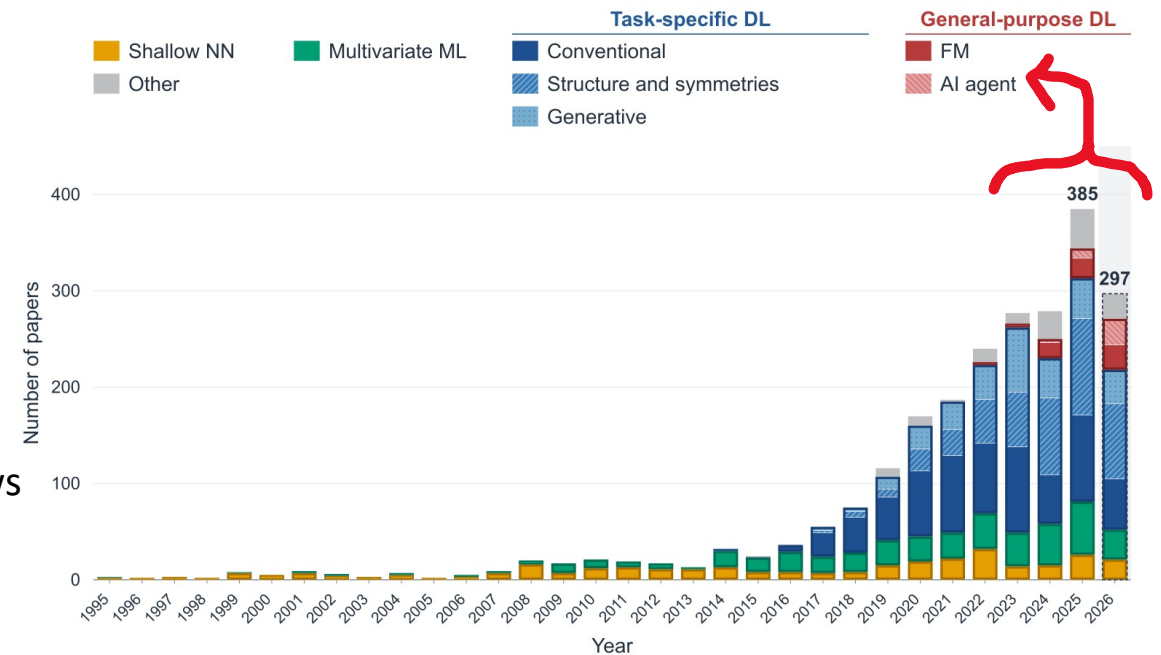
→ DL approaches that can be used across multiple tasks are also becoming a trend

- Foundation model (2023-)
 - Reusing large-scale pretrained model across multiple tasks, detectors, and datasets
- Language, agentic science (2024-)
 - Using LLMs to support and automate research workflows
 - Support reading papers, writing code, checking the result

arXiv papers (hep-ex)

Annual ML-related paper counts

Category composition, 1995–2026. Bar height = n_classified.



2026 has a dashed outline: full-year coverage was not specified. No annualization applied.

“Collider physics in Japan”

- AI in “Collider Physics in Japan”: Workshops and Initiatives (selected examples)
 - [ML@HEP](#) (ICEPP), 2022
 - ["ML at HEP workshop" in Japan 2024](#) (KEK), 2024
 - ["AI-Smart"](#), [MLPhYs A02 group](#), ... and many more (?)
- My AI research began with [a project at the Institute for AI and Beyond](#) (2019-)
 - U.Tokyo x Softbank, Leader: Junichi Tanaka (ICEPP)
 - Ongoing collaboration with KEK Computing Research Center (CRC) and ICEPP after the Beyond AI project (2023-)

→ Present lessons learned from the project and discuss recent work



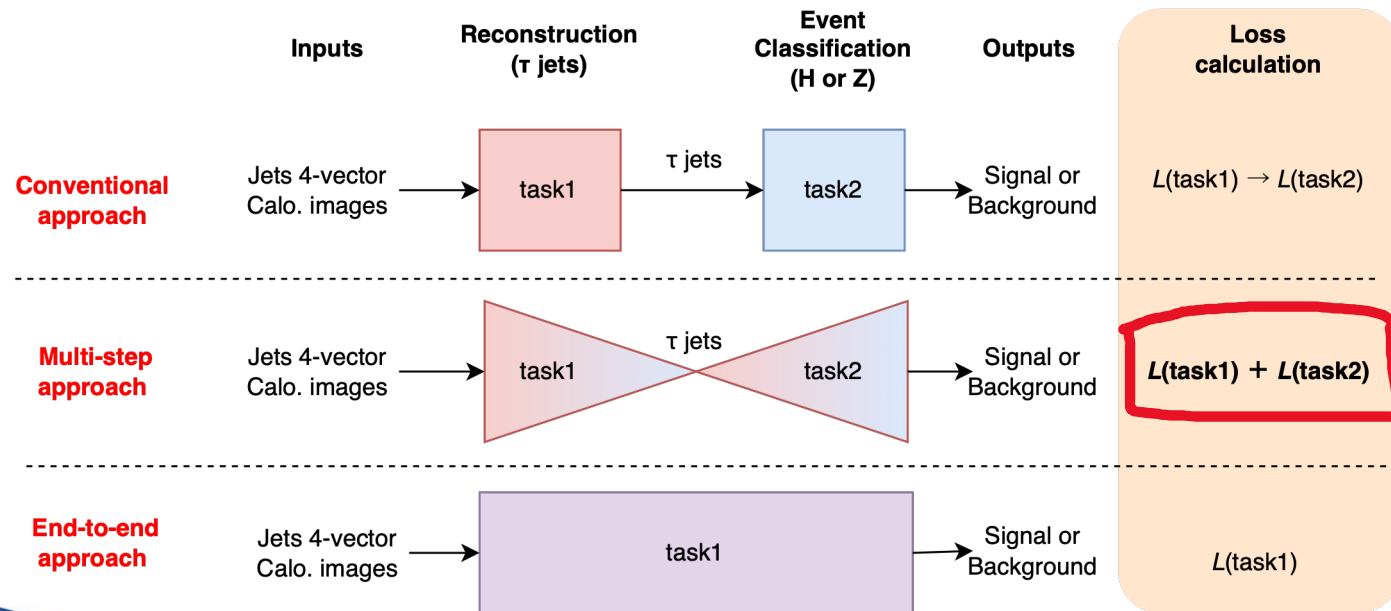
Topics:

1. Multi-step machine learning
2. Physics-motivated neural networks
3. Transfer learning
4. AI-agent

Topic 1: “Multi-step” machine learning

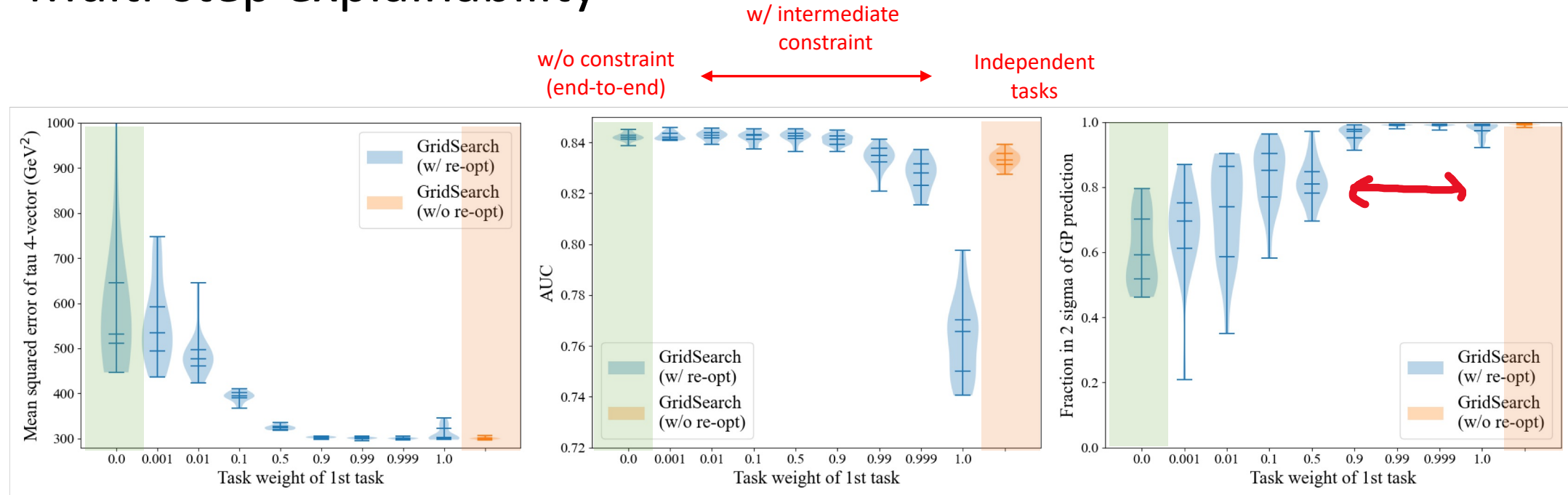
- 2019: many DL models (MLP, CNN, LSTM, etc) were actively discussed to solve a specific task
- The challenge we aimed to address is:
 - Does connecting the subtasks improve end-to-end performance?

e.g) τ jets reconstruction $\rightarrow$ Event classification (Z or H)



- We recognized the performance gain of end-to-end DL
- Explicit intermediate states offer additional benefits:
 - Visibility into the results of each subtask
 - Flexibility to use different models for each subtask

Multi-step explainability



(a) MSE of Task 1

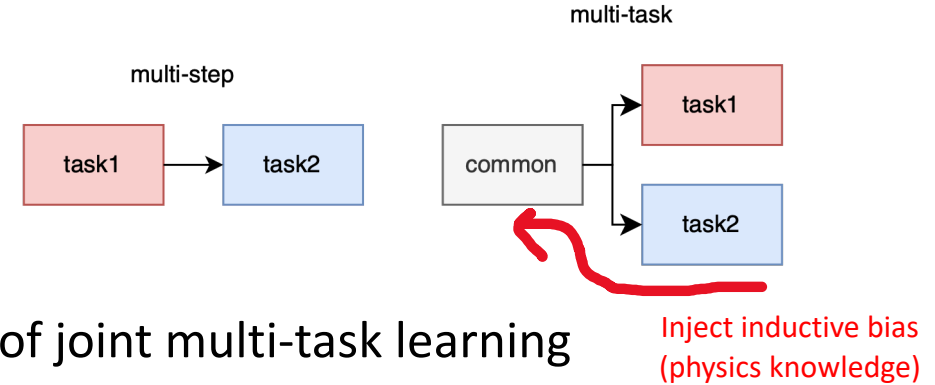
(b) AUC of Task 2

(c) Fraction of events matching with Gaussian process predictions

➤ Successfully trained two sequential tasks (multi-step):

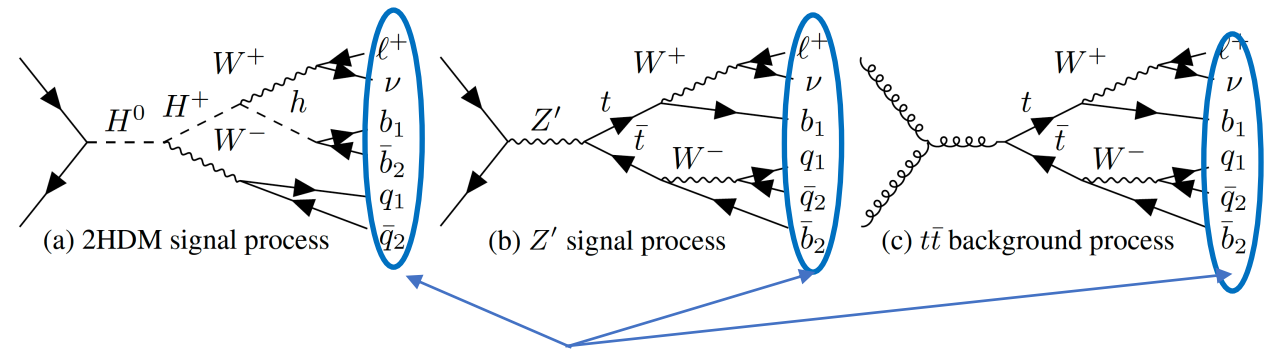
➤ Maintained overall performance, and **achieved physically consistent reconstruction results**

Topic 2: Physics-motivated NN



- Through this multi-step study, we recognized the effectiveness of joint multi-task learning
 - 2022: we therefore began exploring auxiliary tasks to introduce physics-informed inductive biases

- **Physics-motivated GNN learning:**
 - Structure-aware models such as GNNs were being actively explored
 - For event classification, the underlying Feynman-diagram structure is often known a priori



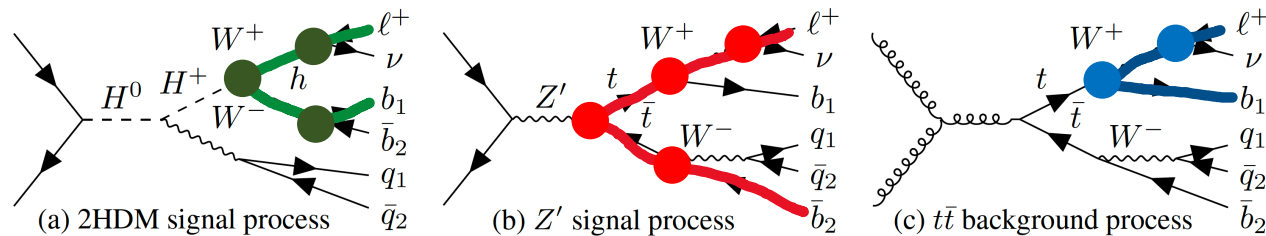
The final-state particles are the same, but the underlying Feynman-diagrams differ

→ Can an auxiliary task inject this physics knowledge into GNN training ?

Auxiliary task labels

- Simulation data were produced:
 - **Signal:** Two-Higgs-doublet model (2HDM), a heavy neutral particle, **Background:** $t\bar{t}$
- The auxiliary tasks is designed to learn the particles decays:

Feynman diagrams



Auxiliary task labels

	l	ν	b_1	b_2	q_1	q_2
l	0	1	3	3	4	4
ν	1	0	3	3	4	4
b_1	3	3	0	1	4	4
b_2	3	3	1	0	4	4
q_1	4	4	4	4	0	1
q_2	4	4	4	4	1	0

(a) 2HDM signal process

	l	ν	b_1	b_2	q_1	q_2
l	0	1	2	4	5	5
ν	1	0	2	4	5	5
b_1	2	2	0	3	4	4
b_2	4	4	3	0	2	2
q_1	5	5	4	2	0	1
q_2	5	5	4	2	1	0

(b) Z' signal process

	l	ν	b_1	b_2	q_1	q_2
l	0	1	2	4	5	5
ν	1	0	2	4	5	5
b_1	2	2	0	3	4	4
b_2	4	4	3	0	2	2
q_1	5	5	4	2	0	1
q_2	5	5	4	2	1	0

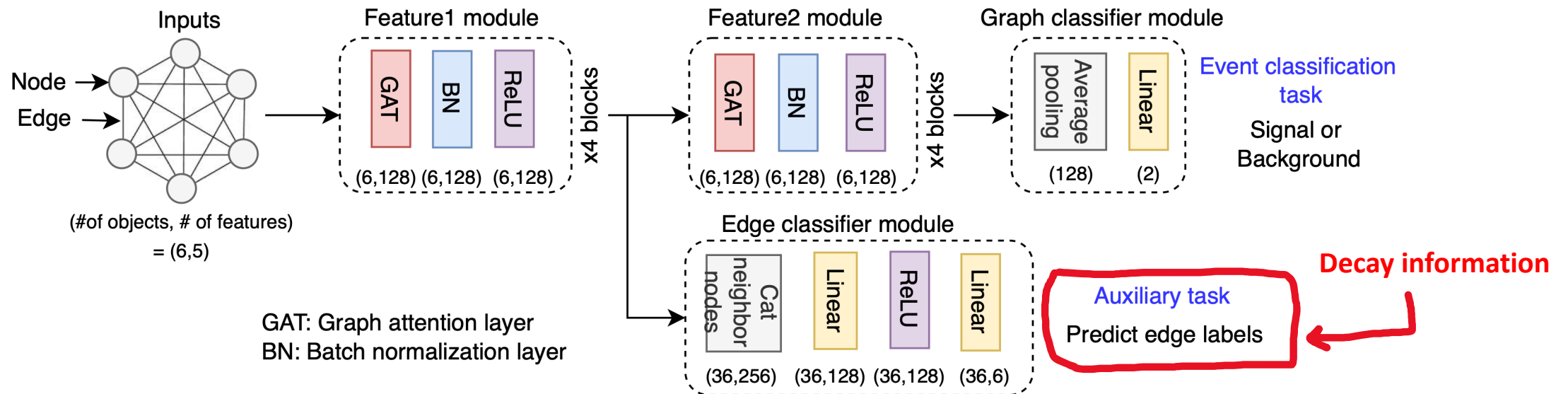
(c) $t\bar{t}$ background process

- The label is defined by counting decaying points between observed objects
- Truth information allows us to track parent particles

➔ These labels are learned as GNN edge labels

Decay-aware neural network

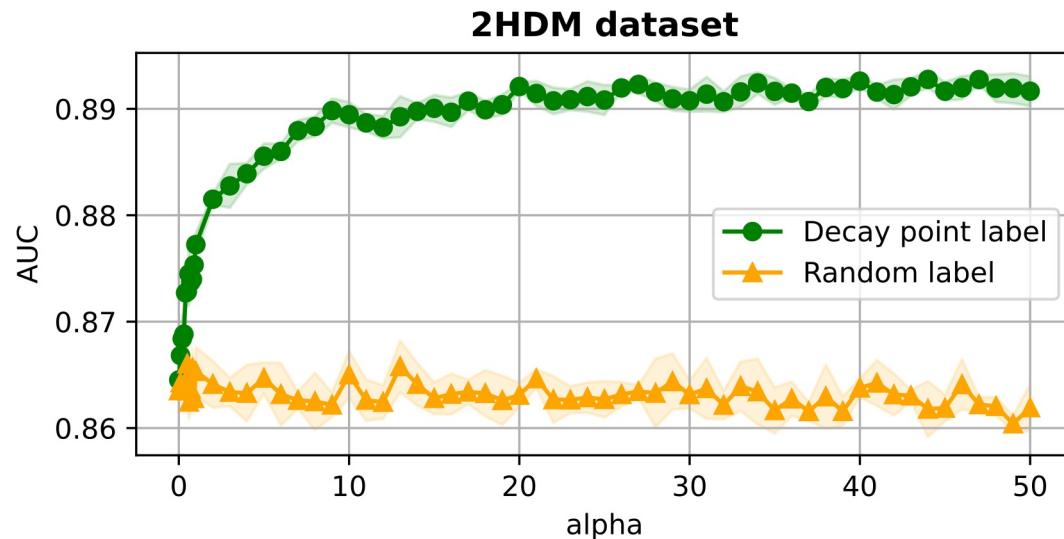
➤ Main task: Event classification (graph-level)



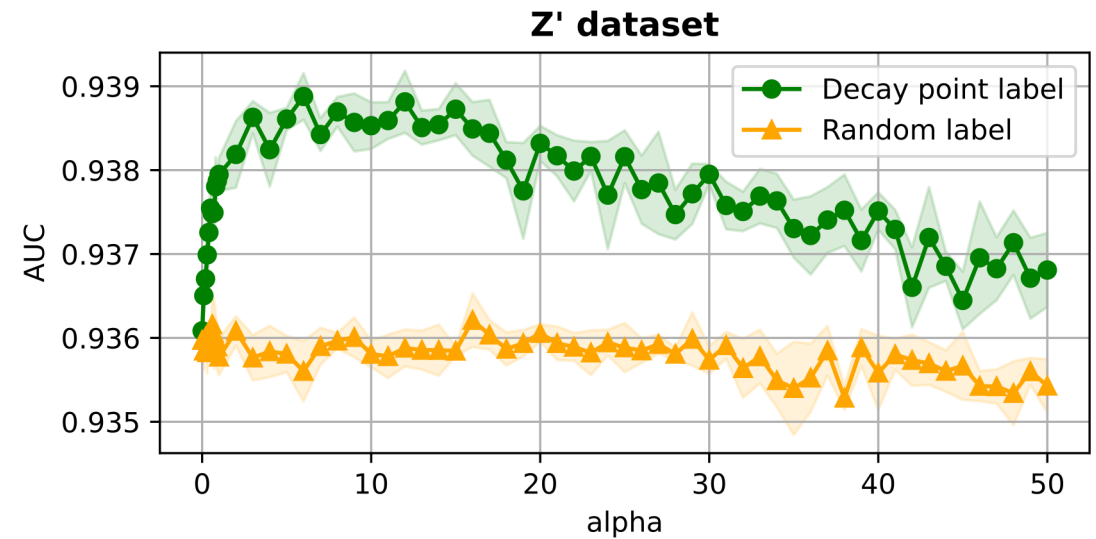
$$L = L_{\text{graph}} + \alpha \cdot L_{\text{edge}} \quad (\alpha \text{ scales the auxiliary task})$$

→ A normal event classification is performed if $\alpha=0$

Multi-task learning: experiments



Normal event classification $\longrightarrow$ Adding decay information



- The decay-aware neural network significantly improved event classification performance
- Learning features shared across multiple tasks also proved effective

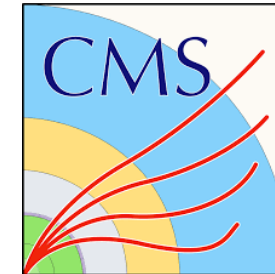
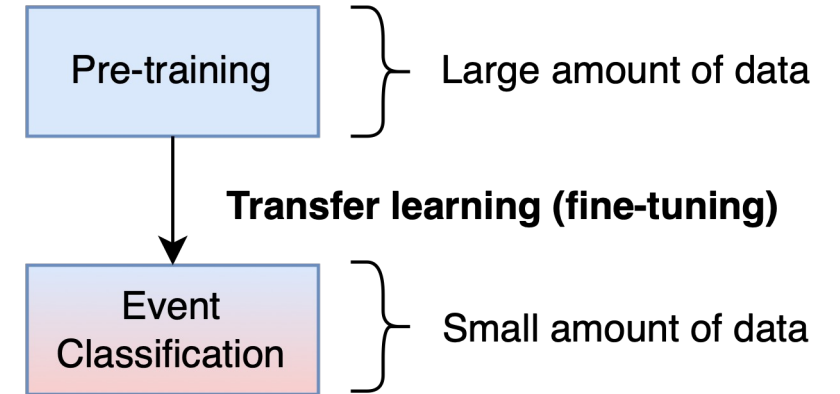
$\rightarrow$ Extended Transfer learning study..

Topic 3: Transfer learnings

- Transfer learning is a method for reusing common features
- 2023: A pre-training strategy was studied for its feasibility in efficiently training event-classification tasks

→ A potential approach to save computational resources and introduce a foundation model

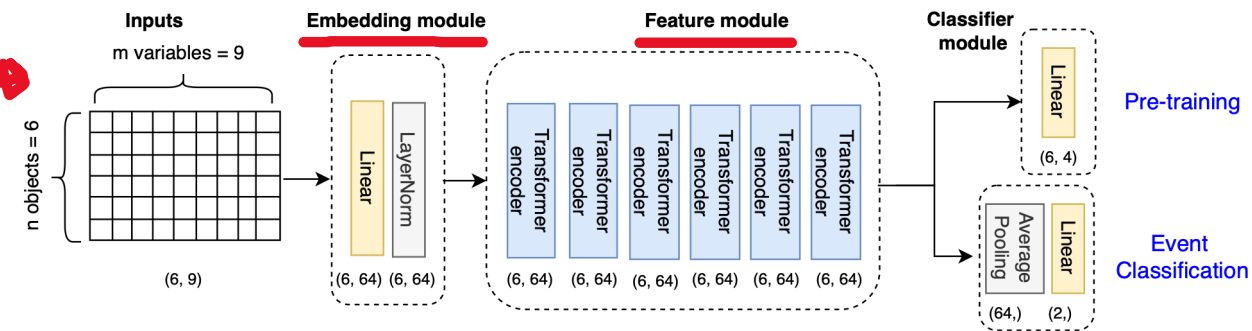
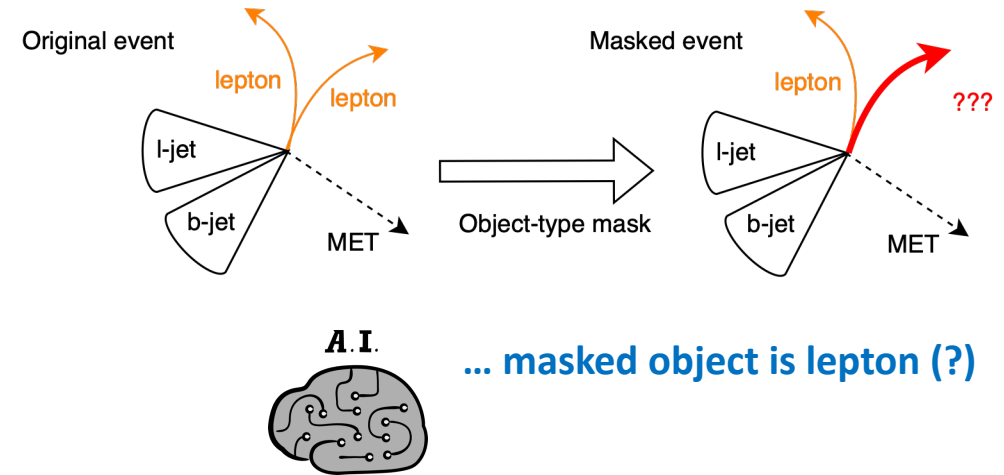
- **Real collision data** were used for pre-training as novel approach
 - No need to generate a large amount of MC simulations
 - Mitigate the bias introduced by selecting an arbitrary physics process



- CMS Open Data
- 13 TeV collision data recorded in 2015

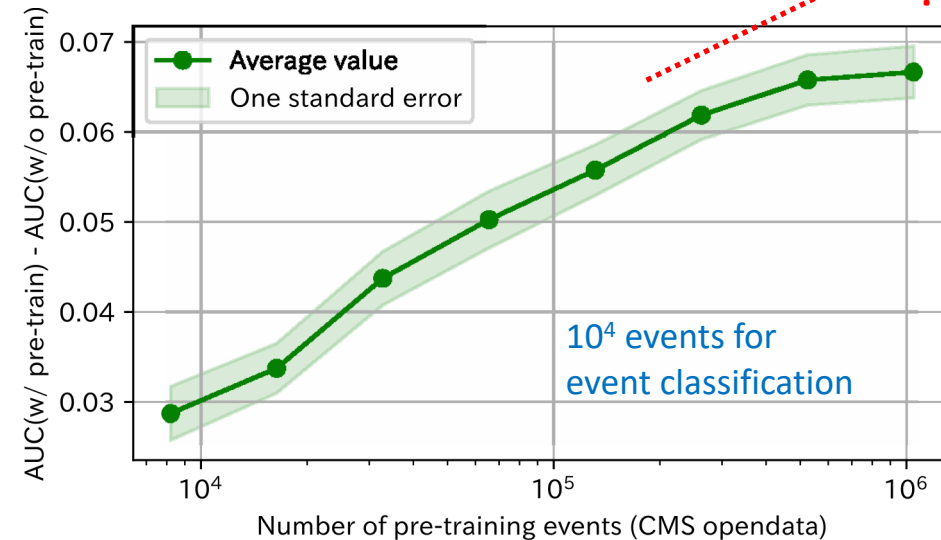
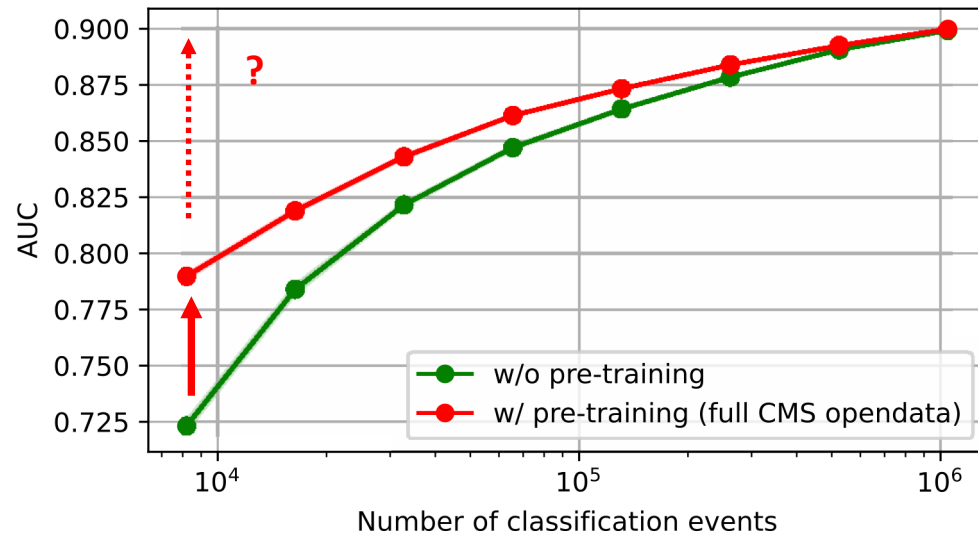
Pre-training strategy

- **Self-supervised learning** was employed to handle the unlabeled collision data
- Strategy:
 - Object-types (lepton, b-jet, l-jet, MET) were randomly masked when preparing each mini-batch
 - DL model is trained to predict the masked object-types
- Transformer encoder was employed
 - ~ 1.7 M trainable parameters
 - Weight parameters of embedding and feature modules are transferred and fine-tuned



Scaling law

Event classification: 2HDM vs ttbar



- Pre-training improves performance when the # of events available for event classification is small
- The scaling behavior encourages a pre-training with a larger data
 - However, this was not thoroughly evaluated because of limited computational resources...

Topic 4: AI-agent

- “AI-scientist” may be a dream we all share..
 - Recent advances in LLMs have already made many of these capabilities a reality
- As an initial milestone, we set the goal of enabling AI to reproduce published results

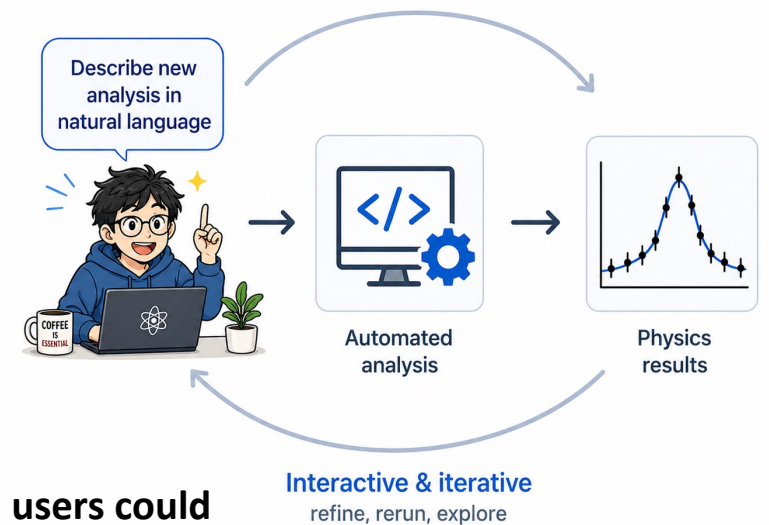
Goal: Build a framework that automatically generates an analysis environment and code from HEP publications, and verifies the reproducibility of the results.

Reliability & feedback

- Improve the reliability of published results
- Provide constructive feedback to authors (missing/ambiguous descriptions)

Toward analysis automation

- **Once paper→code is established, users could describe new analyses in natural language**
- Long-term vision: interactive, automated HEP analysis workflows



Paper reproduction

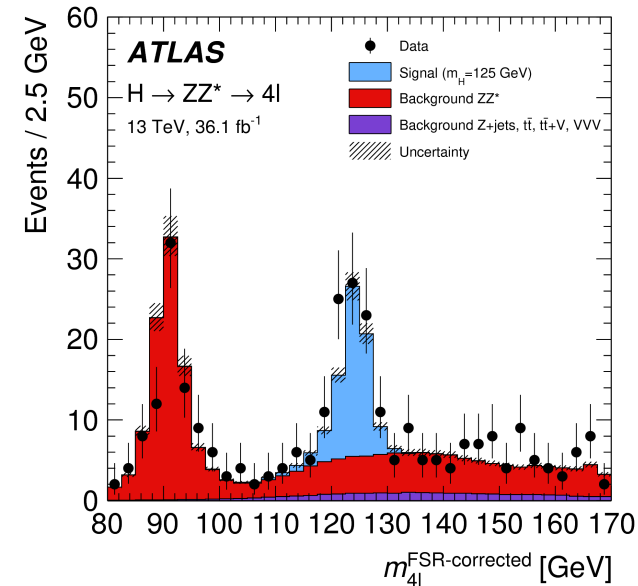
➤ Target analysis

- ATLAS $H \rightarrow ZZ^* \rightarrow 4l$ (arXiv: [1708.02810](https://arxiv.org/abs/1708.02810))
- DATA: ATLAS Open Data, pp collisions (2015-2016)
- Why this paper?
 - Physics significance (Higgs boson)
 - Should be publicly reproducible via Open Data
 - Non-trivial: selection details are delegated to references

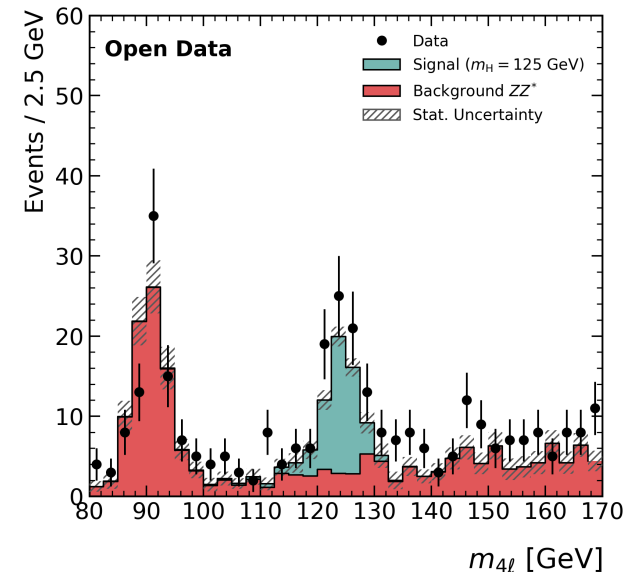
➤ Ground truth

- We manually reproduced the analysis ([baseline code](#))
- **28 events/objects selection criteria** extracted from the paper

Paper



Human-reproduced



→ The distributions agree well. This baseline is the reference for all downstream evaluations

ATLAS skills

- Collider physics analyses rely on implicit knowledge that LLMs have not yet fully learned

- MC scaling, applying Good Run List, applying trigger matching, etc...

→ **Commercial LLMs appear to address these issues through web search (?)**

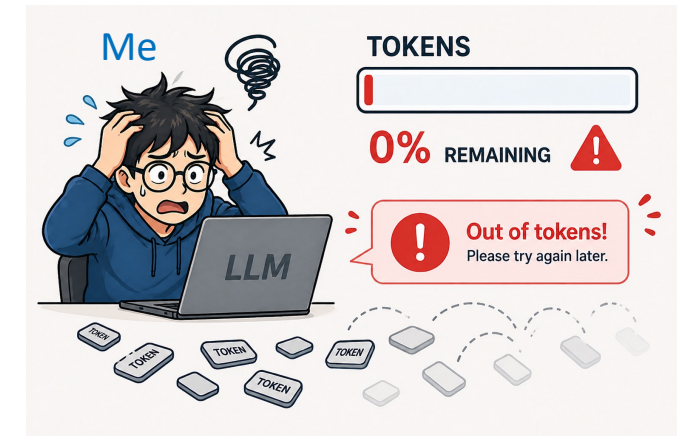
- Our AI-scientist is expected to work 24 hours / 365 days

- Token usage challenges, internet accesses (security concern)

- **Open-weight models combined with locally stored knowledge offer one possible solution**

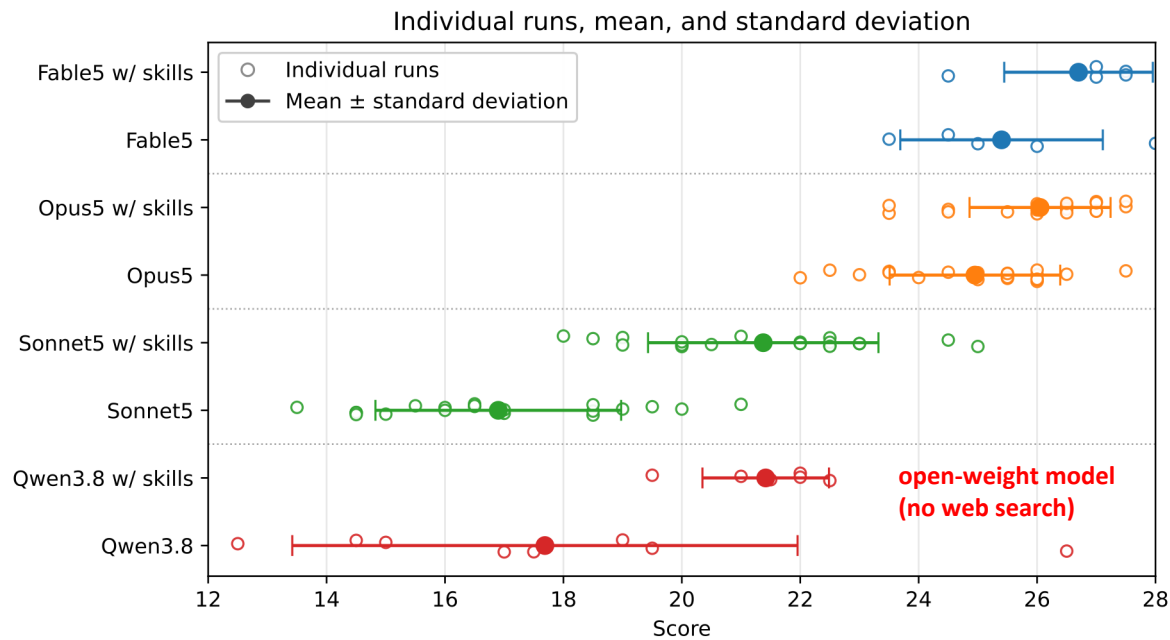
- We wrote [a small set of documents](#) (~1.5k lines) as a Claude code skill

- General information about the analysis: how to apply MC scaling, the meaning of the stored variables, etc

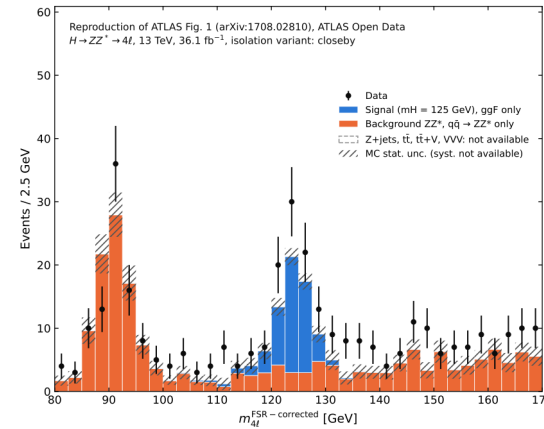


Paper reproduction results

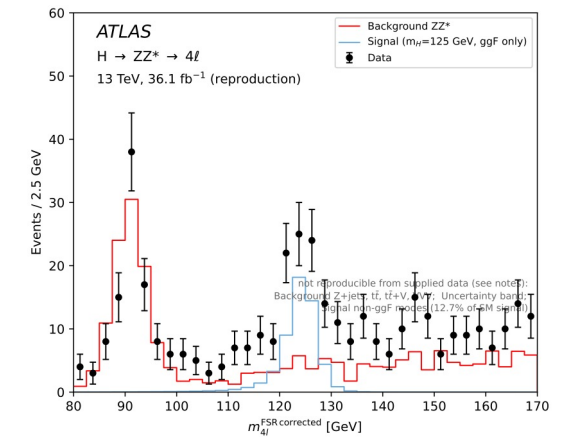
- Provide the paper (PDF) in the prompt, and ask the model to generate analysis code that reproduces the figure
- Evaluate the code by the number of criteria it satisfies (out of 28)



Fable w/ skills



Qwen w/ skills



- Better models, better scores...
- The ATLAS skill improved the scores as expected, even with commercial models
- Open-weight model + skills shows promise
 - However, a powerful GPU is needed for reasonable response times

Future work of AI-agent

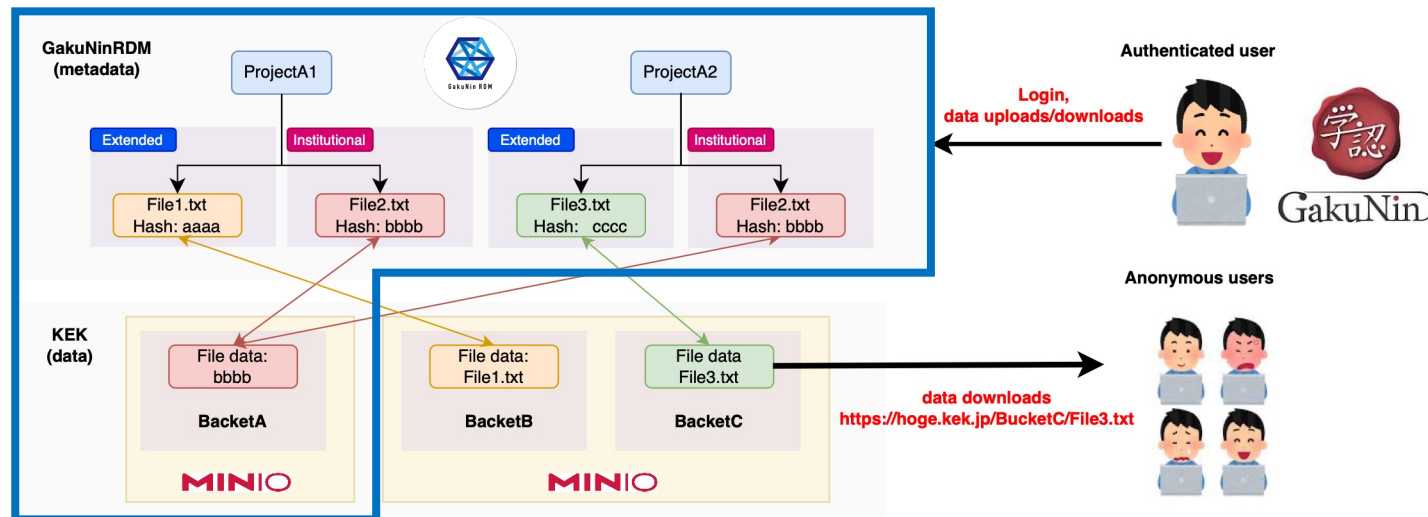
- How to provide local information is under consideration
 - We currently provide the information as Markdown text, but excessive detail can consume a large fraction of the context window
- Toward analysis automation:
 - Collider physics has traditionally relied on highly accurate simulations
 - Thus, we can simulate new-physics signals and quantitatively evaluate whether an AI-agent can discover them
 - Then apply the AI-agent to real data to search for new physics

Open data

- Open data is crucial for advancing AI research
- KEK CRC (including me) is developing “the infrastructure” for preserving and publishing open data related to KEK research projects
 - On-premise S3 storage operated by KEK CRC
 - Integrated with [GakuNinRDM](#), developed by National Institute of Informatics (NII)

System design under consideration

Closed collaboration



Summary

- AI is used throughout the collider physics workflow from detectors and reconstruction to simulation and analysis
- Physics knowledge improves learning. For example, decay-aware auxiliary tasks provide classification performance
- Pre-training on real collision data is promising, especially when labels samples are limited
- AI-agents and open-data infrastructure are key steps toward reliable and more automated analyses

Backup