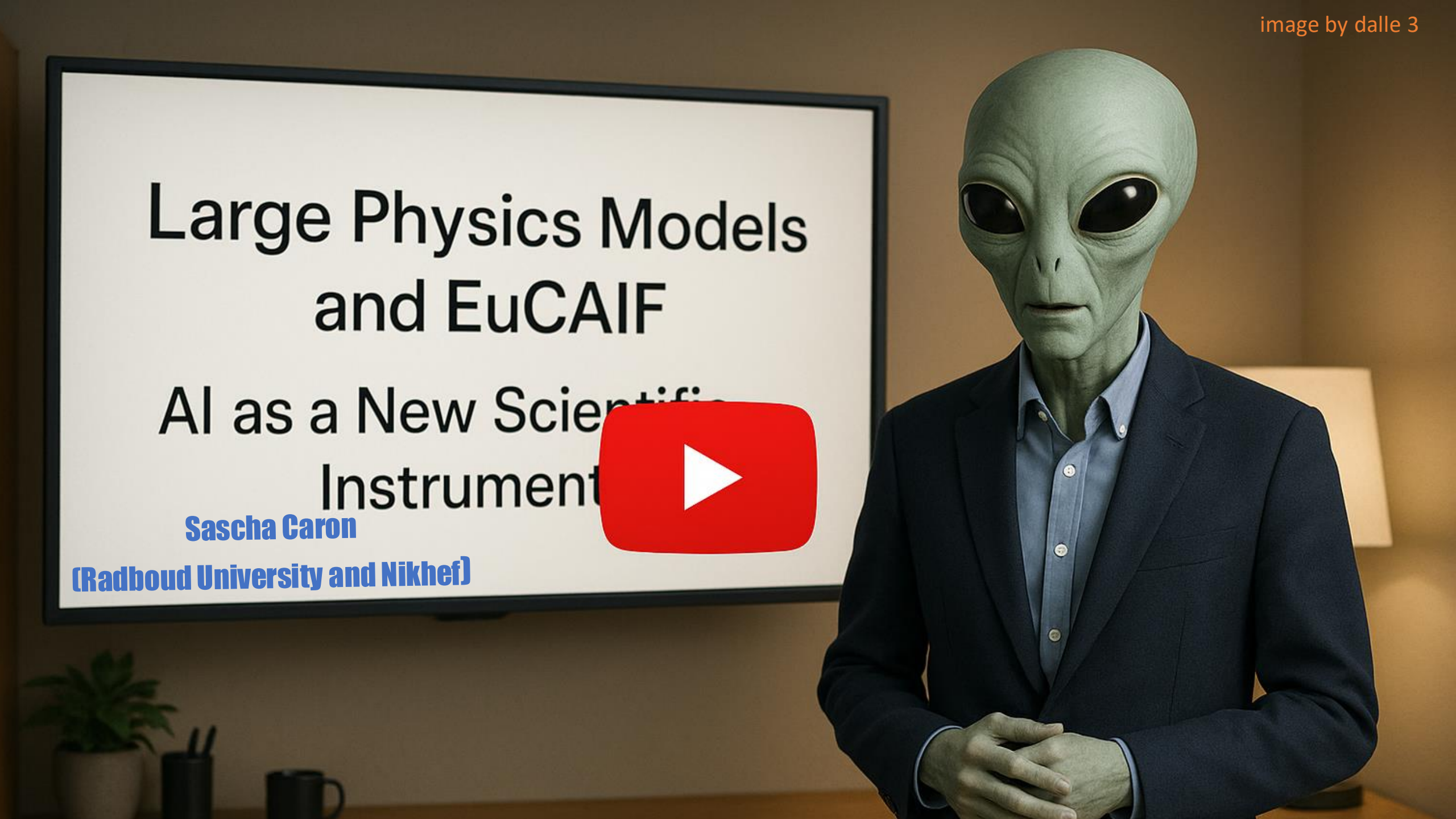
Created with DALL·E 3,

image by dalle 3



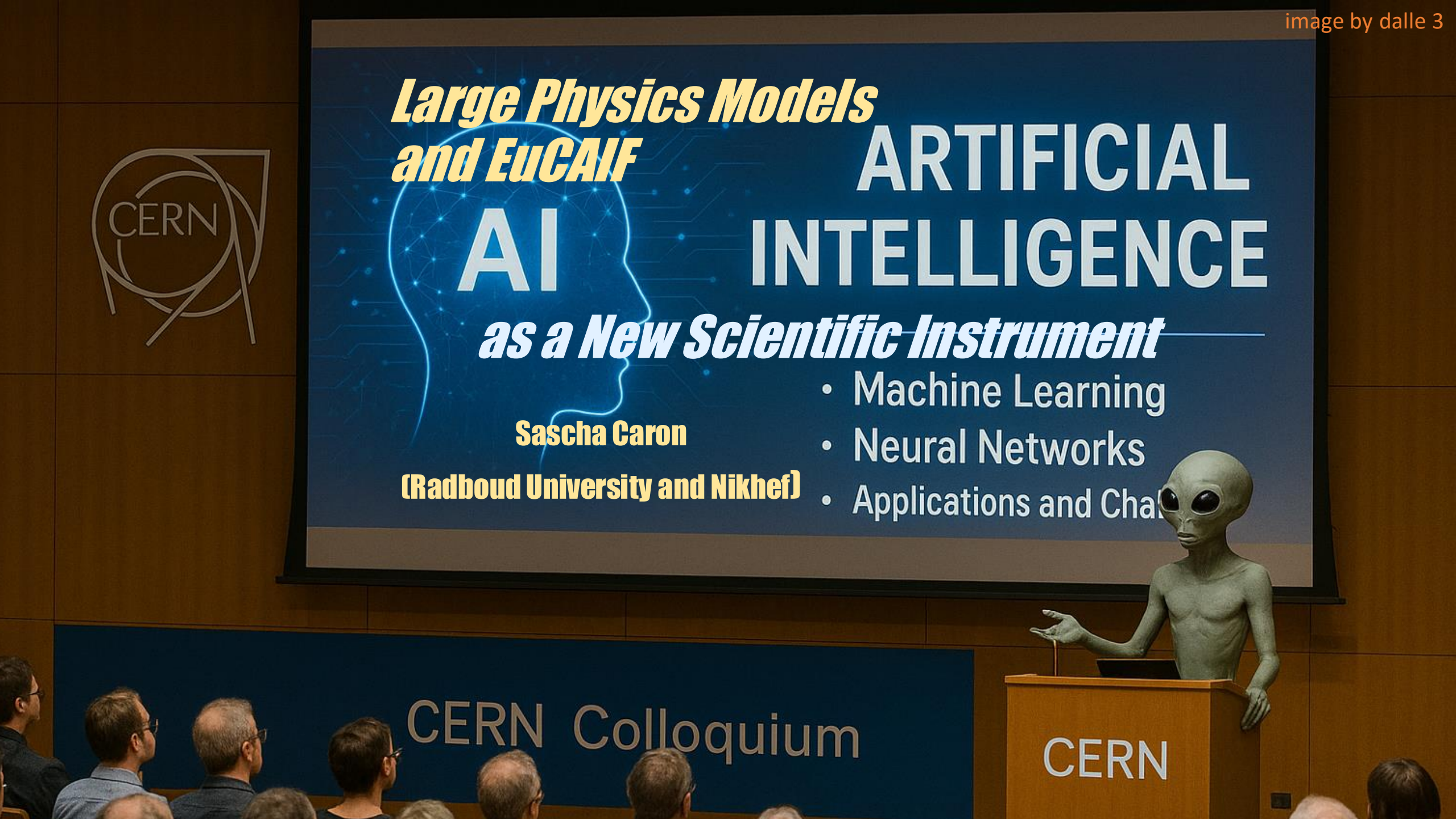


image by dalle 3

Who am I ?

Working on the search for "signals of new physics" for 20 years with automation / data analysis / data science / AI

➔What is ”new physics” ?

My aim: Establish ML/AI in fundamental physics

(via EuCAIF, darkmachines, Radboud AI, …)

Outline

I appologize for not mentioning your work or paper (too broad)

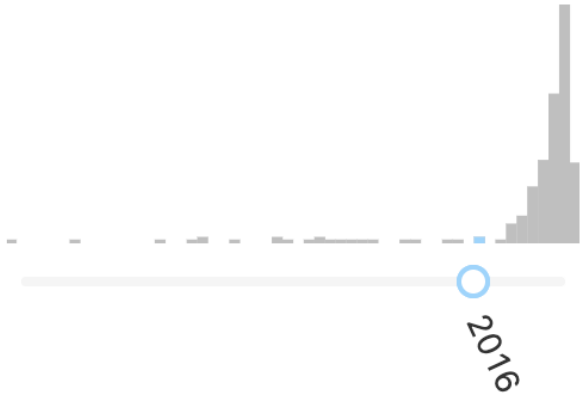
1. AI and HEP
2. Foundation Models (LHC raw example)
3. Text Models + benchmarking
4. Large Physics Models
5. EuCAIF

(*also for not showing our work on anomaly detection by Polina Moskvitina, gamma ray +astrophysics models, black hole simulation, DM, etc.)*

INSPIRE HEP | literature | find title AI | Help | Submit

Literature | Authors | Jobs | Seminars | Conferences | Data BETA | More...

Date of paper



Number of authors

- Single author 1
- 10 authors or less 2

2 results | cite all

Citation Summary | Most Recent

Oscillazioni dei neutrini nella materia e applicazione ai neutrini solari #1

Beatrice Moser (INFM, Padua) (Sep, 2016)

links | cite | claim | reference search | 0 citations

The BSM-AI project: SUSY-AI–generalizing LHC limits on supersymmetry with machine learning #2

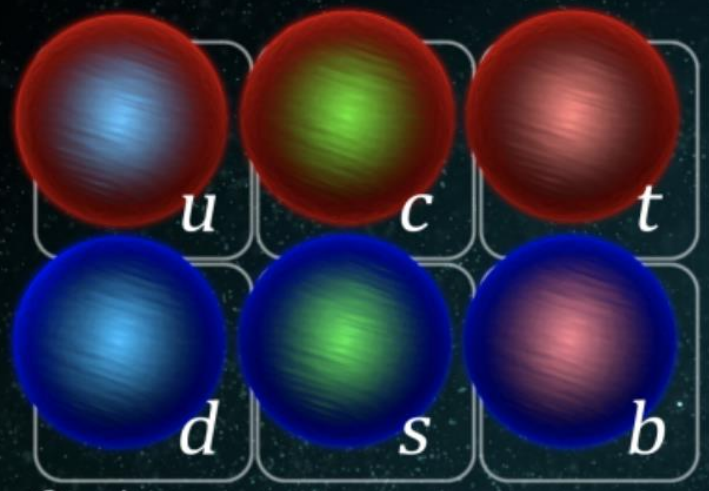
Sascha Caron (Nijmegen U., IMAPP and NIKHEF, Amsterdam), Jong Soo Kim (Madrid, IFT), Krzysztof Rolbiecki (Madrid, IFT and Warsaw U.), Roberto Ruiz de Austri (Valencia U., IFIC), Bob Stienen (Nijmegen U., IMAPP) (May 9, 2016)

Published in: *Eur.Phys.J.C* 77 (2017) 4, 257 • e-Print: 1605.02797 [hep-ph]

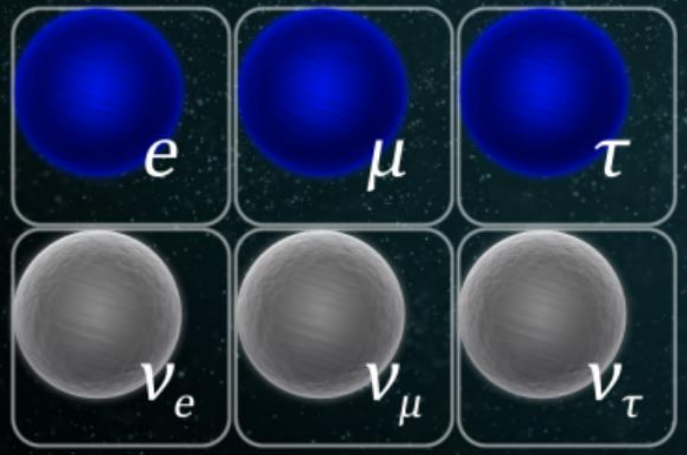
pdf | DOI | cite | claim | reference search | 70 citations

Fun fact: This appears to be among the earliest explicit uses of the term 'AI' in the title of a particle physics publication (2016), based on our literature search.

Standard Model



Quarks

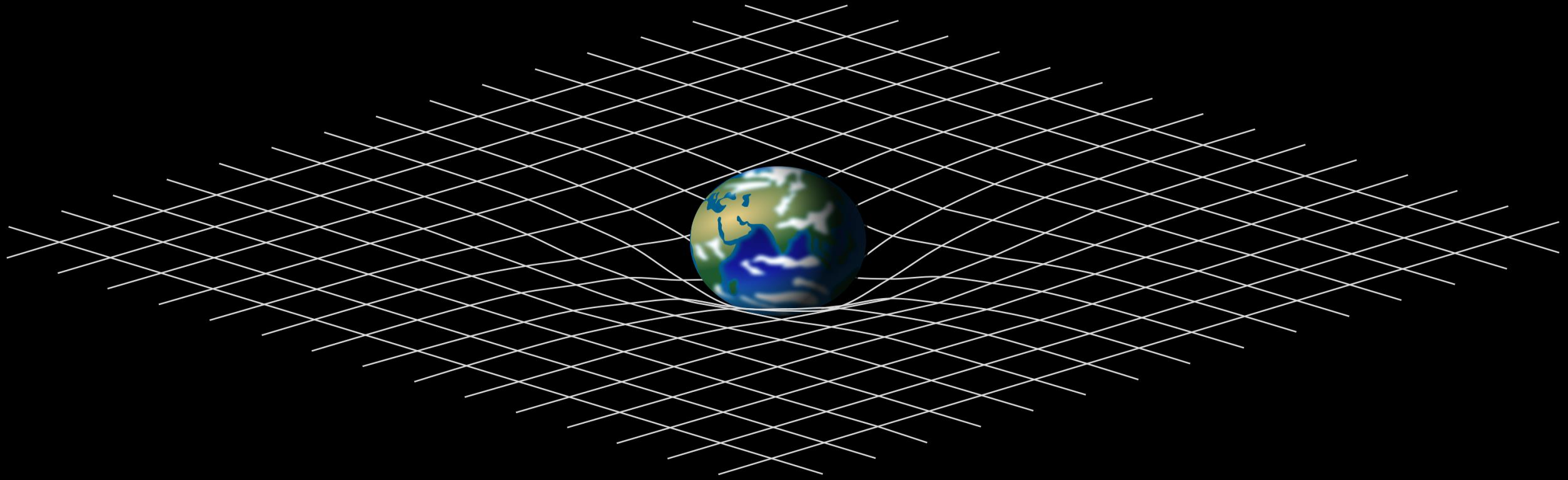


Leptons


Higgs boson

Forces

General Relativity



We know that this is wrong

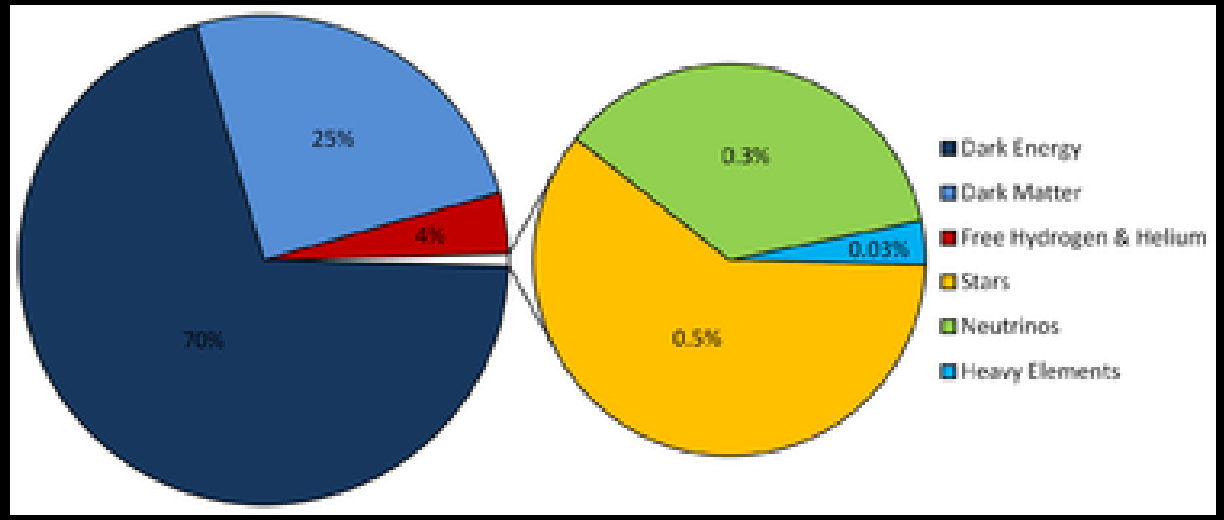
The Big Bang /Standard Model forces should have created (almost) equal amounts of matter and antimatter.

Why is there far more matter than antimatter in the universe?

What is Dark Matter/ Dark Energy ?

What is General Relativity doing at the Quantum Level ?

…

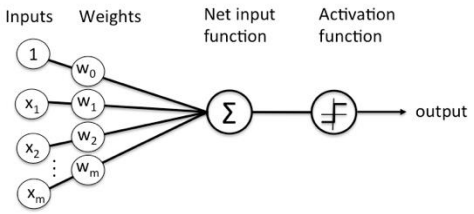


- **Maybe we don’t just need more data – we also might need better ideas**

- *Will AI be the tool to help us generate, test or organize them?*

The speed of AI (r)evolution

AI is evolving quicker than we are …
Deep Learning + generative AI changed the game.
HEP has been a user and developer for decades.

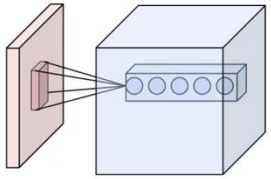


perceptron

Hopfield network

Backprop *(Hinton et al.)*

1st Workshop on AI In High energy & Nuclear Physics (AIHENP)



CNN

TMVA

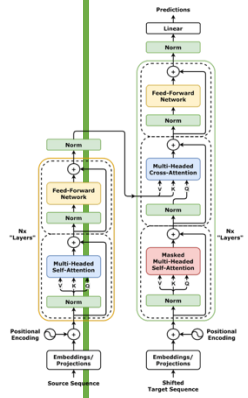
Start of LHC

”Deep Learning” (DBN, AlexNet)

-Higgs boson Kaggle challenge
-First deep learning Paper in HEP

GAN

IML working group at CERN



transformer

Particle Transformer

Chatgpt 3.5

New Collider?

Physics AI ?

1957 | 1982 | 1986 | **1990** | 1998 | **2010** | 2014 | 2017 | 2022 | 2040

Timetable for AI and HEP (with some examples of developments)

From AI in Fundamental Physics to Large Physics Models

2025:AI/Deep Learning (DL) in HEP

AI – unfolding

AI – inference

AI - event selection

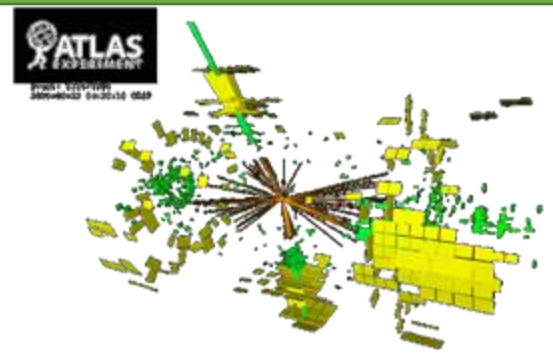
AI - reconstruction

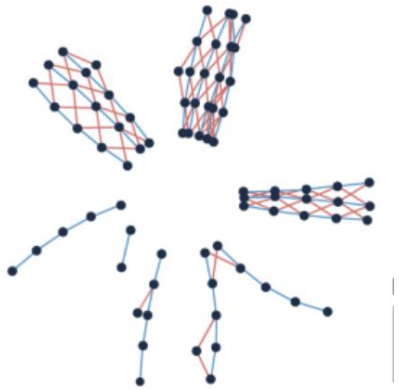
AI - tracking

AI - detector simulation

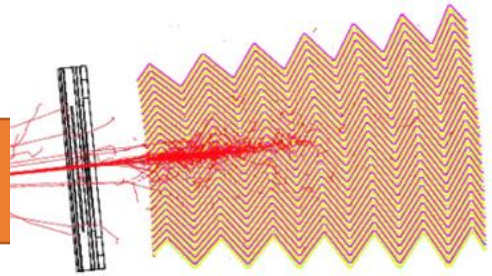
AI - event generation

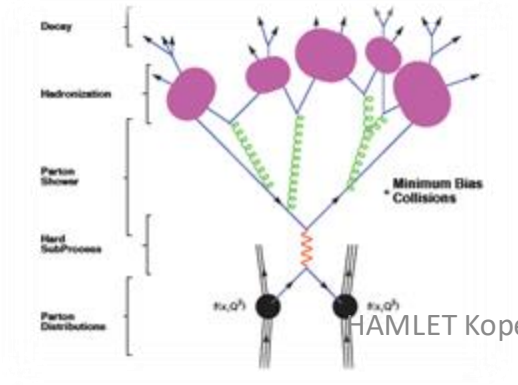
Energy and angles of reconstructed particles





Detector Simulator





AI - trigger

2025:AI/Deep Learning (DL) in HEP

AI – unfolding

AI – inference

AI - event selection

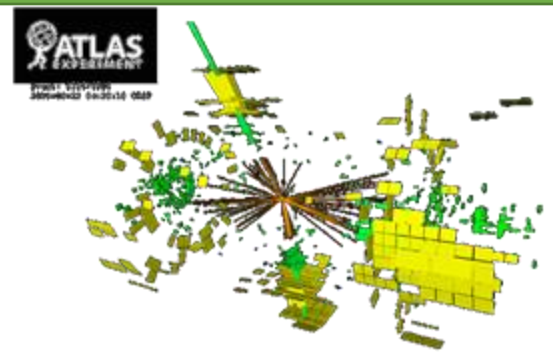
AI - reconstruction

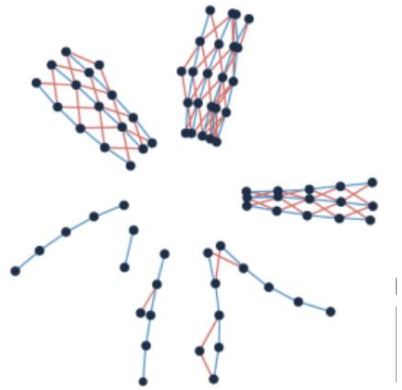
AI - tracking

AI - detector simulation

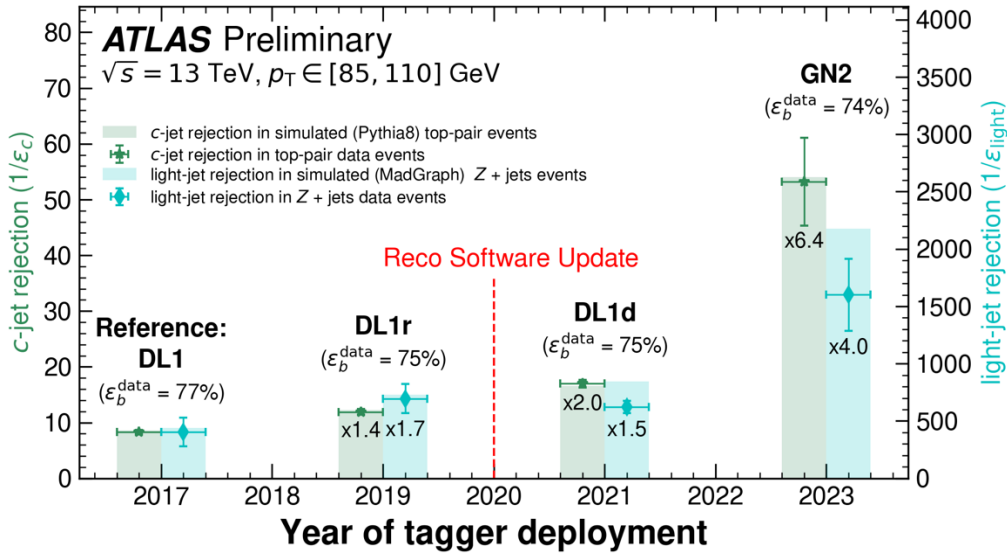
AI - event generation

Energy and angles of reconstructed particles

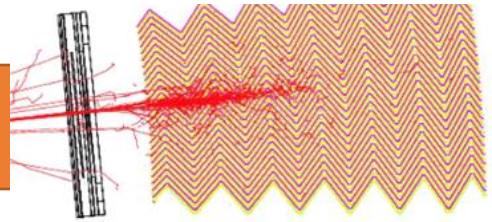


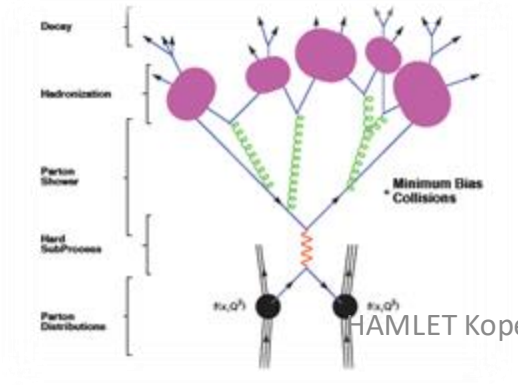


c & light jet rejection



Detector Simulator





AI - trigger

2025:AI/Deep Learning (DL) in HEP

AI – unfolding

AI – inference

AI - event selection

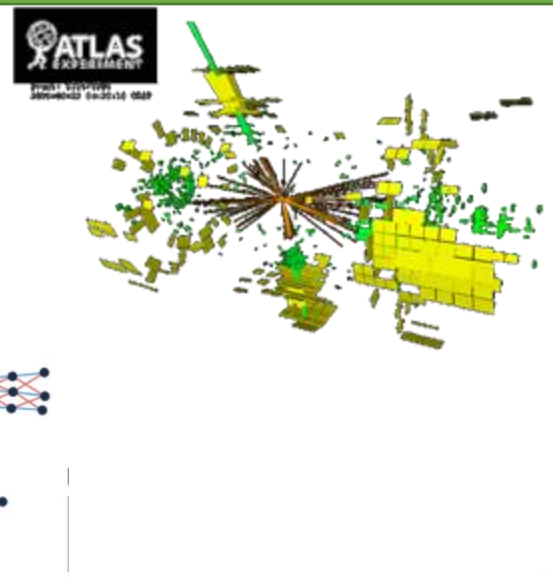
AI - reconstruction

AI - tracking

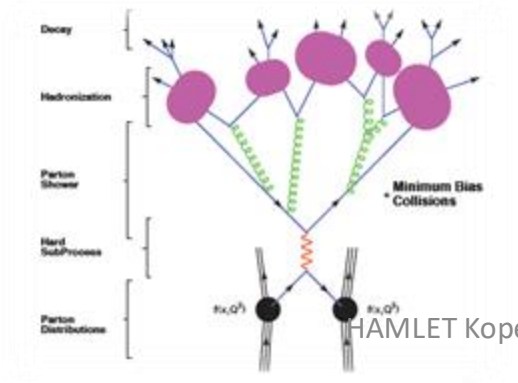
AI - detector simulation

AI - event generation

Energy and angles of reconstructed particles



Detector Simulator



Many other AI topics outside the standard analysis pipeline, e.g.

AI - Accelerator Optimization and Control

AI - Monte Carlo sampling

AI - Experimental Design

AI - Sensor Data Reconstruction

AI – Performance monitoring

AI – Anomaly Detection

…

2035?:AI/Deep Learning (DL) in HEP

AI – unfolding

AI – inference

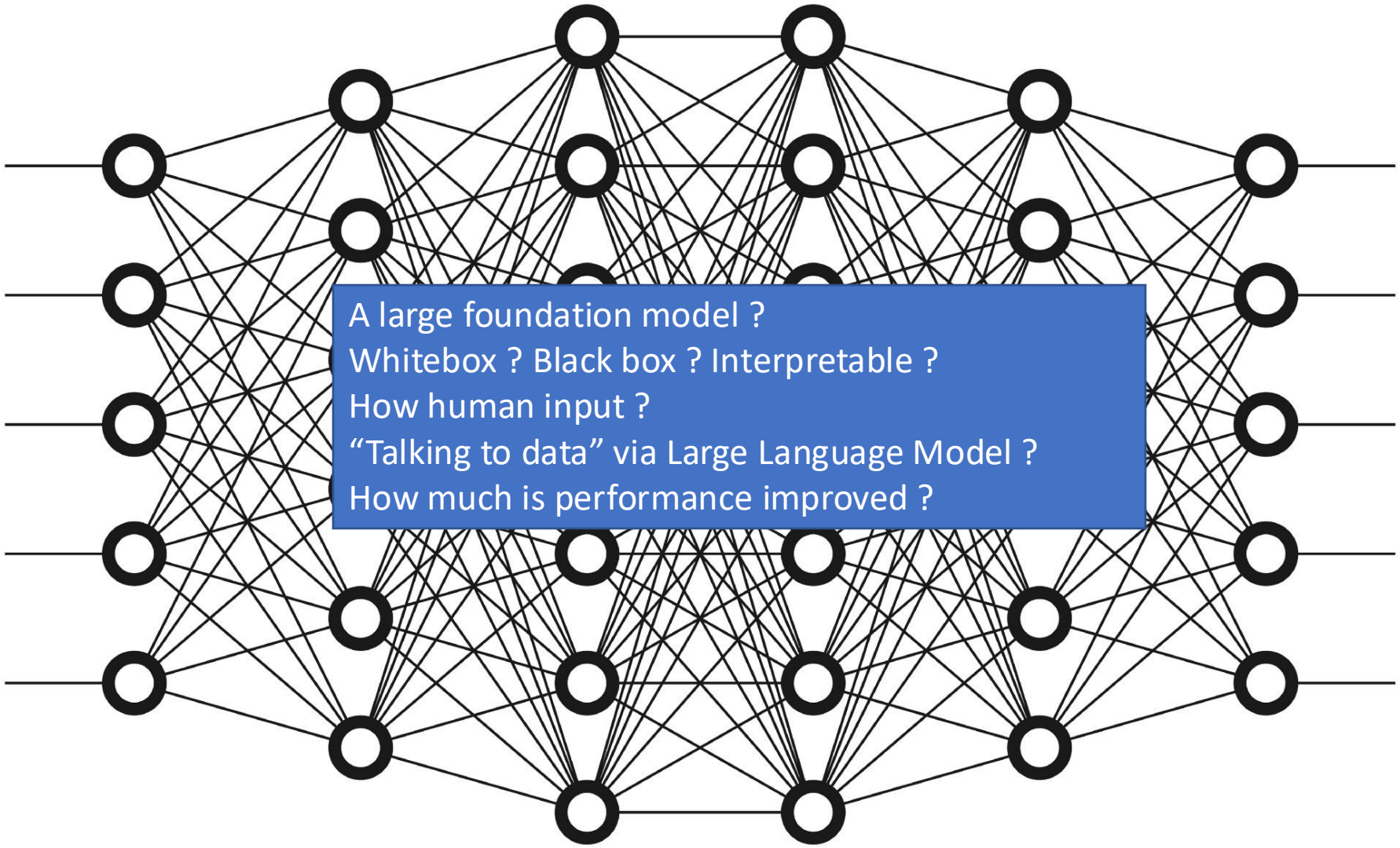
AI - event selection

AI - reconstruction

AI - tracking

AI - detector simulation

AI - event generation



A large foundation model ?
Whitebox ? Black box ? Interpretable ?
How human input ?
“Talking to data” via Large Language Model ?
How much is performance improved ?

Next steps ?

2025:

- Growing interest in end-to-end differentiable pipelines

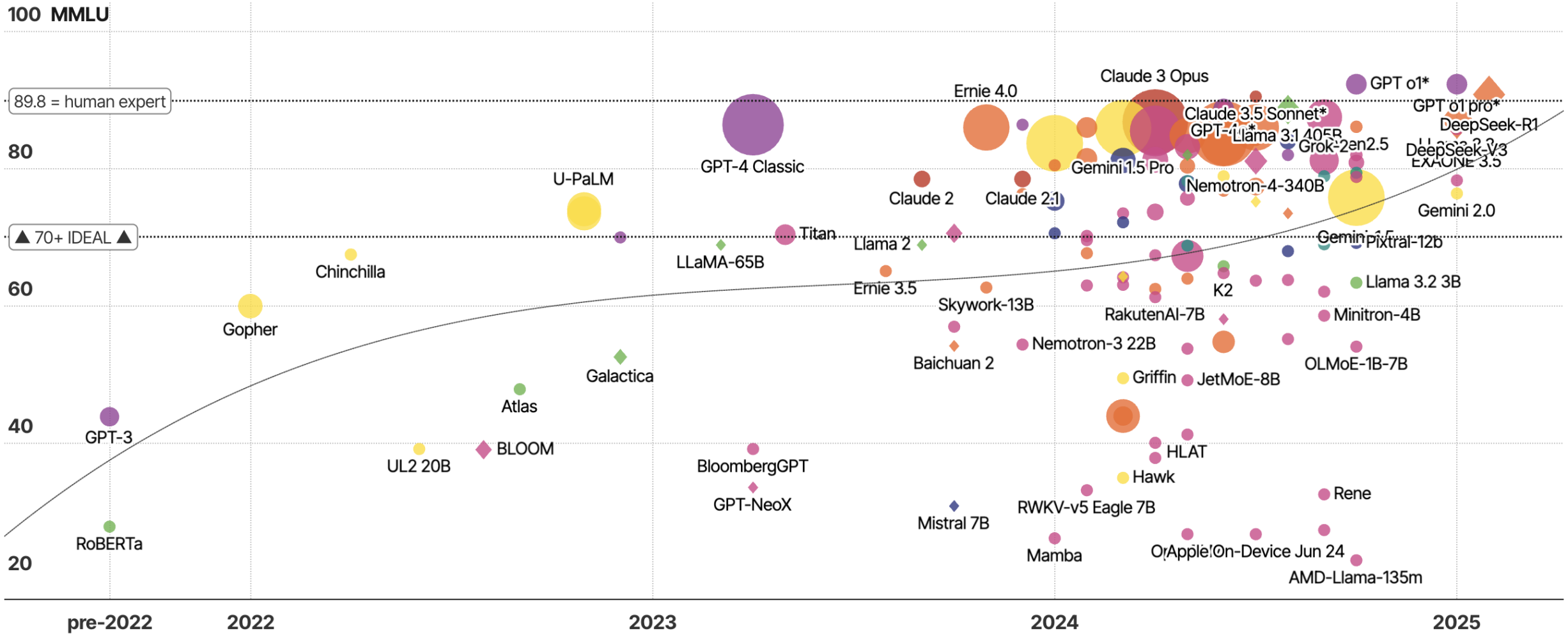
- AI Surrogate models for simulation + differentiable reco = full learnability

➔What if the entire physics analysis pipeline becomes trainable?

➔Could AI help uncover physics we aren’t even looking for?

Meanwhile in industry…

MMLU benchmark consists of 15,908 multiple-choice questions

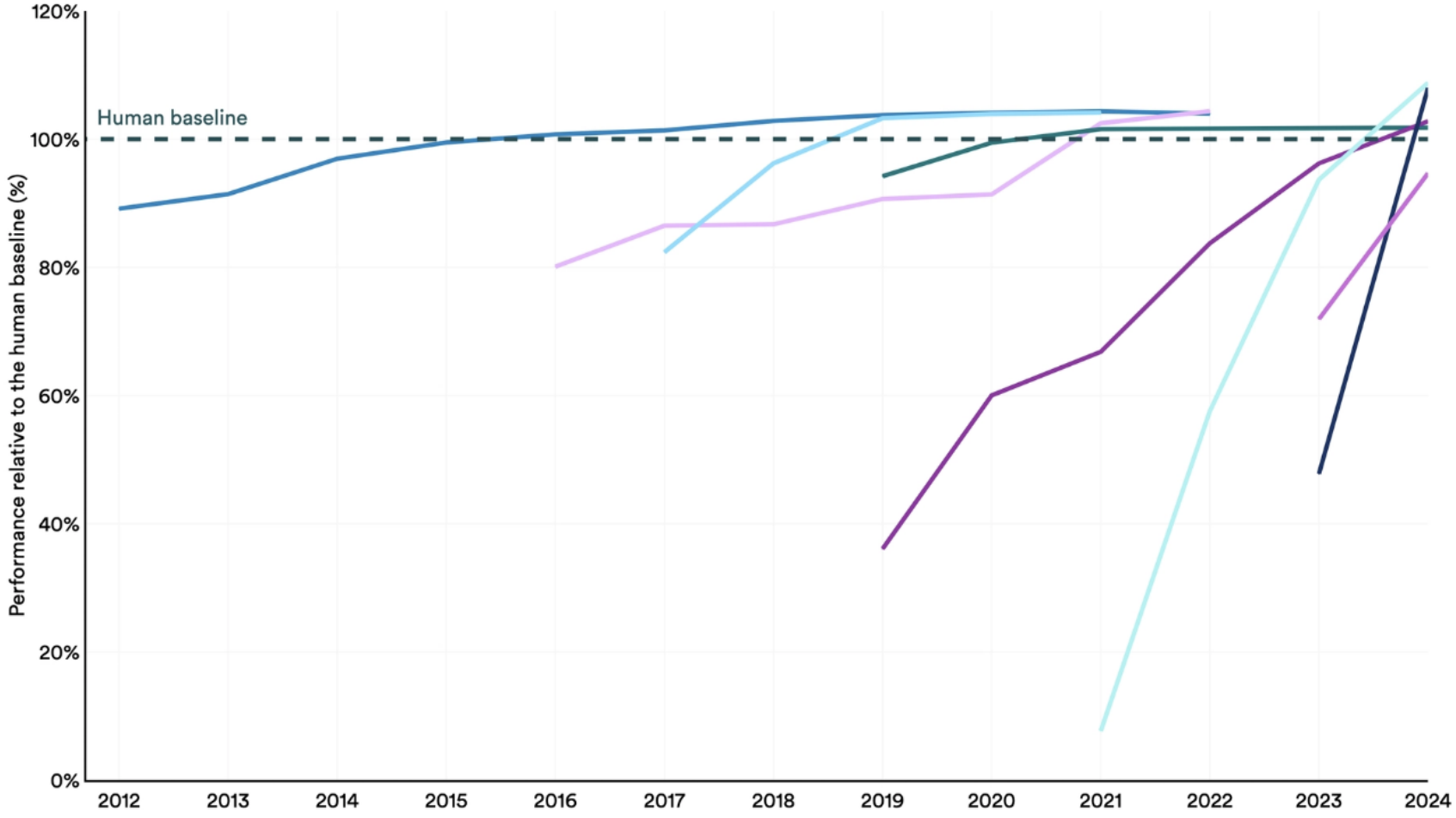


David McCandless, Tom Evans, Paul Barton
Informationisbeautiful // Jan 2024

MMLU = benchmark for measuring LLM capabilities
* = parameters undisclosed // source: LifeArchitect // data

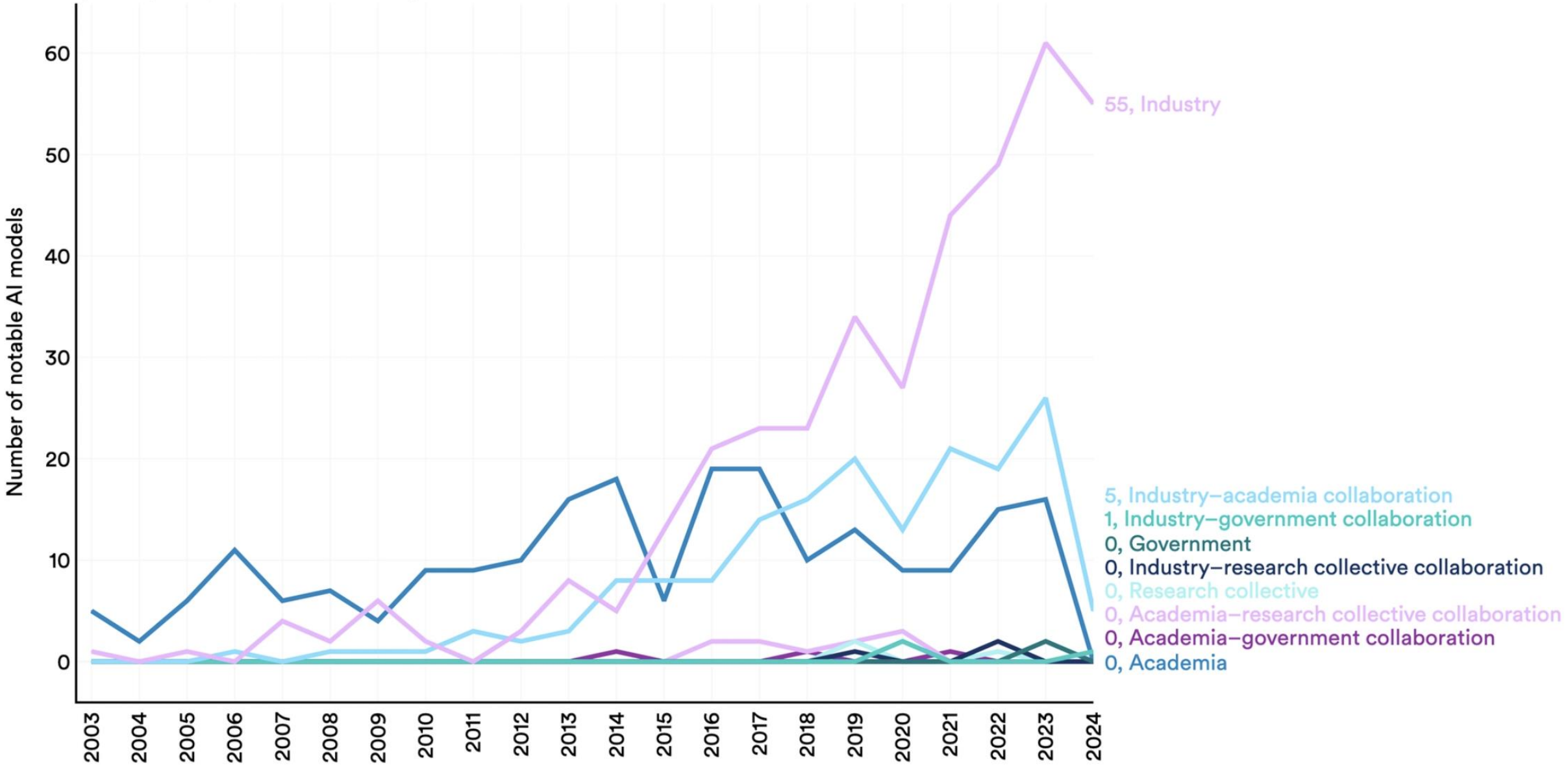
Select AI Index technical performance benchmarks vs. human performance

Source: AI Index, 2025 | Chart: 2025 AI Index report



Number of notable AI models by sector, 2003–24

Source: Epoch AI, 2025 | Chart: 2025 AI Index report



Science foundation models

No true academia model yet (see table generated by gpt4o)

Model	Domain	Modalities	Multipurpose?	Foundation-like?	Industry Partner(s)	Comparable to Commercial LLMs?	Release Date
GNoME	Materials Science	Crystal structures, stability	✅	✅	Google DeepMind	❌ Domain-specific; extremely capable in its area	**Dec 2023**
xTrimo V3	Life Sciences	Genomics, microscopy, proteins	✅	✅	Shanghai AI Lab + bio industry	⚠️ Not general-purpose, but very large-scale	**Oct 2024**
AlphaFold 3	Structural Biology	Protein-ligand-RNA structure	✅	✅	DeepMind + Isomorphic Labs	❌ Narrow but best-in-class in structural prediction	**May 2024**
Modulus	Physics (Simulation)	PDEs, time series, fields	✅	✅	NVIDIA	❌ Solver-focused, not reasoning-based	Ongoing
OpenCatalyst	Catalysis, Atomistic Sim	Atomic configs, forces, reactions	✅	✅	Meta AI + Carnegie Mellon University	❌ Specialized for chemical simulation	Ongoing (OC20: 2021)
Polaris	Earth & Space Science	Geospatial, imagery, time series	✅ (planned)	⚠️ In development	NASA + NVIDIA + Google Cloud	❌ Not yet released, promising scope	In development
Galactica	General Science Text	Text, code, citations	✅	⚠️ (withdrawn)	Meta AI	⚠️ High ambition, but not currently available	**Nov 2022 (retracted)**
SciBERT / BioGPT	Biomedical NLP	Text (NER, Q&A, classification)	✅ (NLP only)	—	Allen AI / Microsoft / Meta	❌ Narrow, but widely used in biomedical NLP	2019–2023

Foundation Models: General Intelligence for Specific Tasks

Typically trained on large, diverse datasets:

- Text (e.g. web, papers), Code , Images

  Math, diagrams, structured data

Foundation Models: General Intelligence for Specific Tasks

Typically trained on large, diverse datasets:

- Text (e.g. web, papers), Code , Images

  Math, diagrams, structured data

==> Physics equivalent:

- *Simulation output*
- *Detector-level (raw) , reco-level events , Analysis-level*
- *Papers, plots, logbooks, metadata*

Foundation Models: General Intelligence for Specific Tasks

Typically trained on large, diverse datasets:

- Text (e.g. web, papers), Code , Images

  Math, diagrams, structured data

➔ Learning objective: *self-supervised* prediction (e.g. next token, masked region, contrastive pairs)

Foundation Models: General Intelligence for Specific Tasks

Typically trained on large, diverse datasets:

- Text (e.g. web, papers), Code , Images

  Math, diagrams, structured data

➔ Learning objective: *self-supervised* prediction (e.g. next token, masked region, contrastive pairs)

+ transfer learning (minimal fine-tuning) + many parameters + multipurpose
+ some capability not explicitly included during traing

Foundation Models: General Intelligence for Specific Tasks

Typically trained on large, diverse datasets:

- Text (e.g. web, papers), Code , Images

  Math, diagrams, structured data

➔ Learning objective: *self-supervised* prediction (e.g. next token, masked region, contrastive pairs)

+ transfer learning (minimal fine-tuning) + many parameters + multipurpose
+ some capability not explicitly included during traing

Tokens

Tokenization is the process of breaking a sequence into discrete units — called **tokens** — that a model can understand.

Input	Token Type	Result
"Hello world!"	Word-level	[Hello, world, !]
"$x^2 + y^2 = 23$"	Symbolic	[x^2, +, y^2, =, 23]
"12, 24, 45"	Numeric/symbolic	[12, ,, 24, ,, 45]
"unbelievable"	Subword	[un, ##believ, ##able]

Tokens

- In Natural Language Processing, vocabularies are often 10K–100K tokens

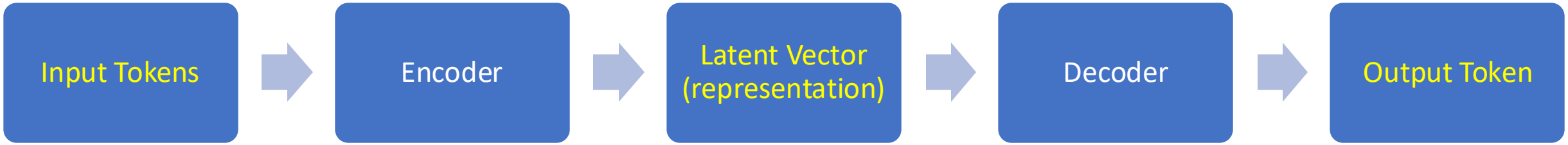
- Tokens in physics represent discrete units of information in experimental or simulation data
- They can be explicit (human-defined) or learned (via models like VQ-VAE, see e.g. arXiv:2401.13537v3 by Golling et al for HEP application)

Examples:

- Detector component identifiers (channels, modules, layers)
- Or binned versions of jets, tracks, etc.

To introduce transformers and LLMs we look at encoder – decoder architectures

To introduce transformers and LLMs we look at encoder – decoder architectures

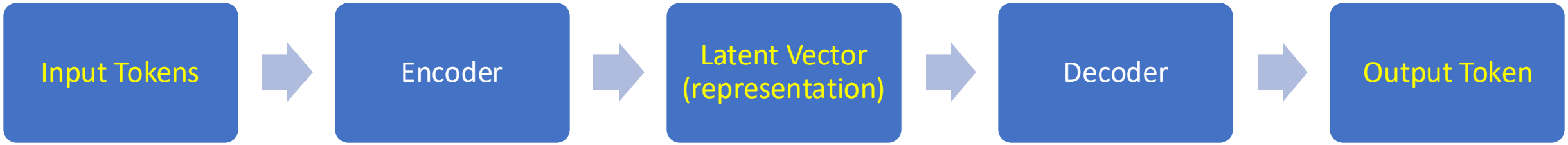


"The" , "Higgs", "decays", "to"
→ 45, 1293, 34, 55
→ V45, v1293, v34, v55

Softmax probabilities
Over the full dictionary

→ Take the one with the Largest probability (argmax)
or sample from probability distribution

To introduce transformers and LLMs we look at encoder – decoder architectures



"The" , "Higgs", "decays", "to"
→ 45, 1293, 34, 55
→ V45, v1293, v34, v55

bb_bar. 0.9

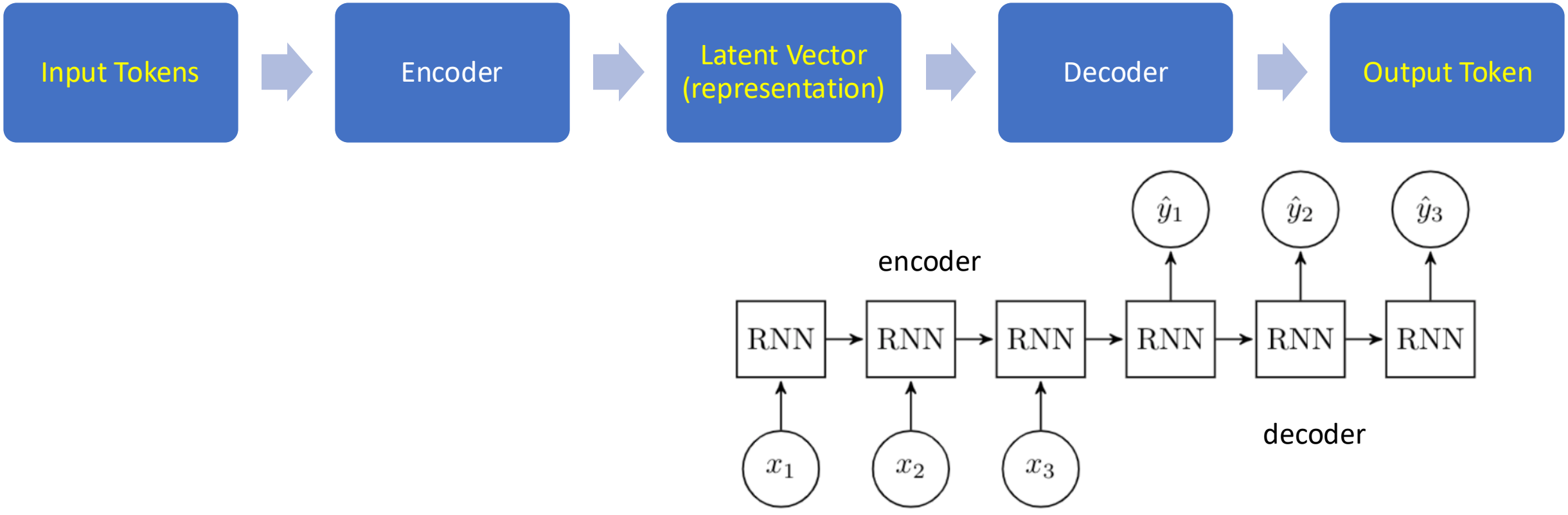
Higgs 0.03

Hello 0.0

Susie 0.01

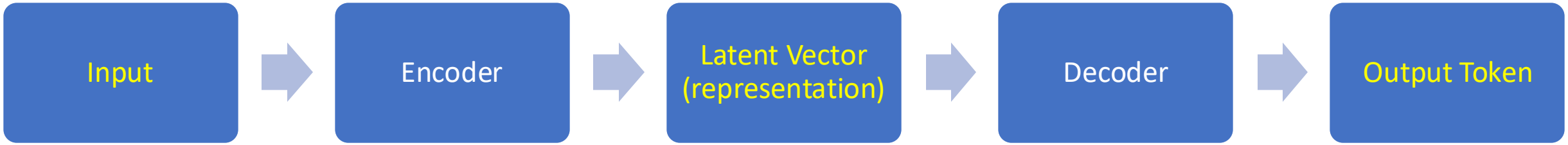
…

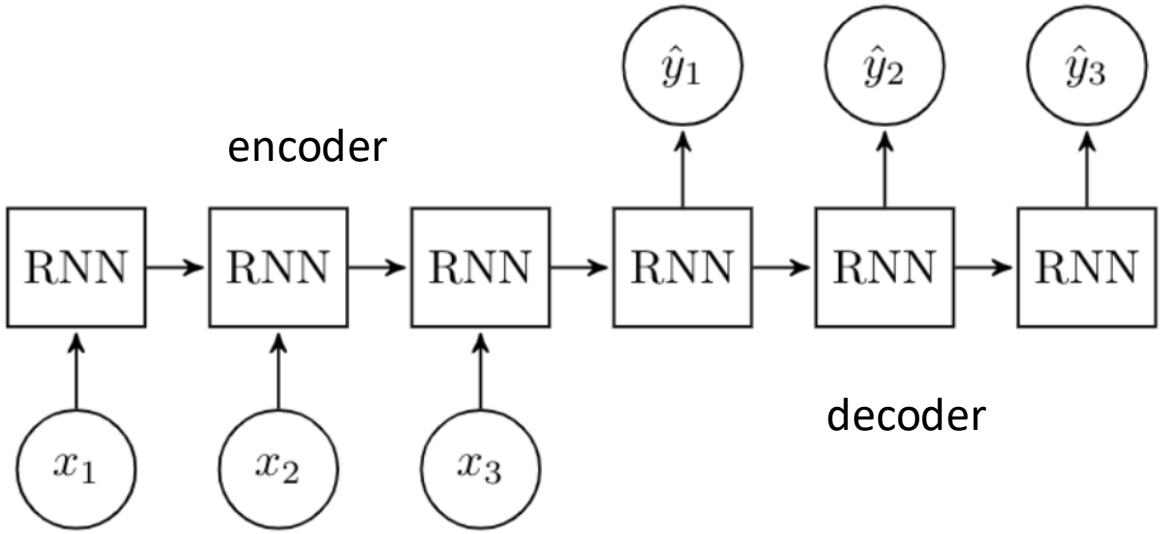
To introduce transformers and LLMs we look at encoder – decoder architectures



Bottleneck
→ Attention

To introduce transformers and LLMs we look at encoder – decoder architectures





Transformers replace the RNN bottleneck with self-attention *(correlations of sequence to sequence, can also implement physics here, see work by Polina Moskvitina in* 2211.05143 ***),*** allowing full context access and better scalability. But the idea of a **latent representation of meaning** lives on --> that's a key to foundation models.

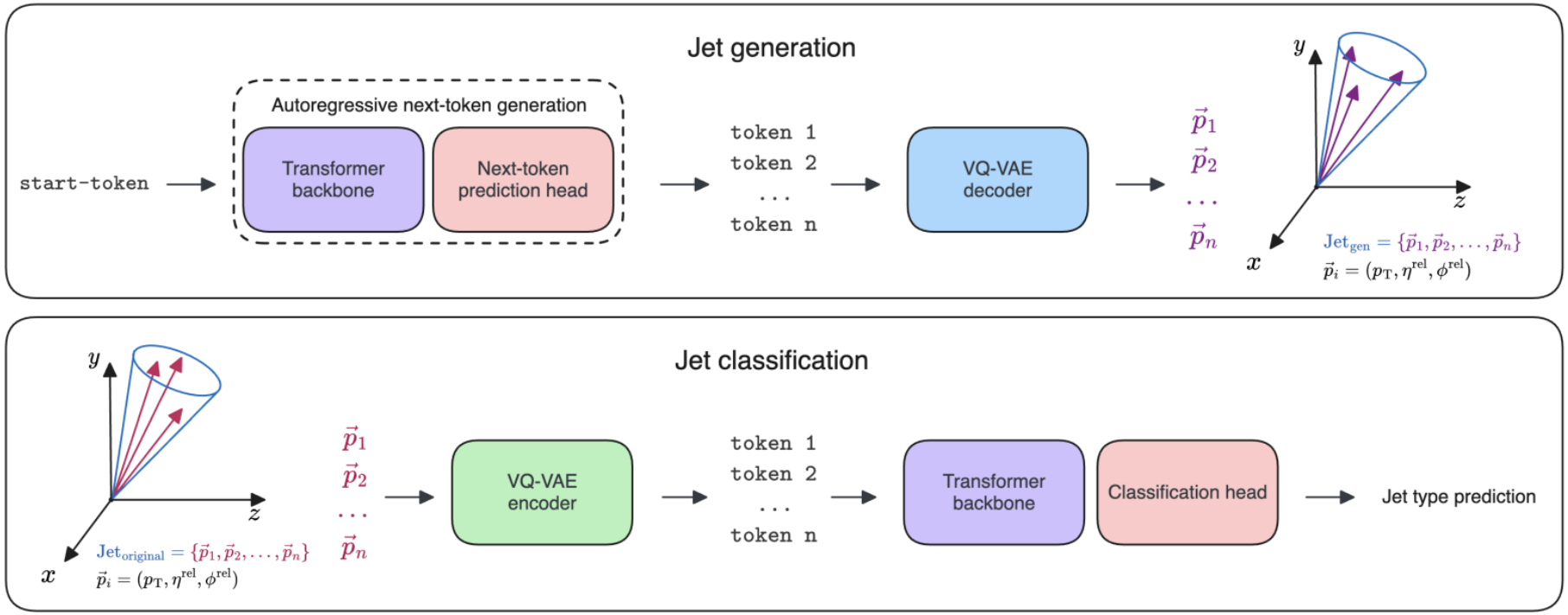
To introduce transformers and LLMs we look at encoder – decoder architectures

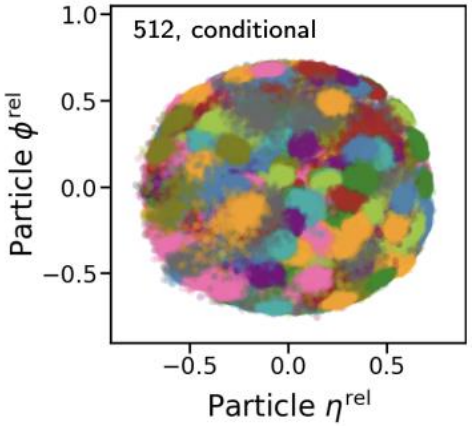
You can also “learn the tokens” with a Vector Quantized Variational Autoencoder:
Codebook (dictionary of tokens) is learned during training.

➔Many attempts on arxiv since last year (2024).

As an example I mention here **omni-jet alpha**

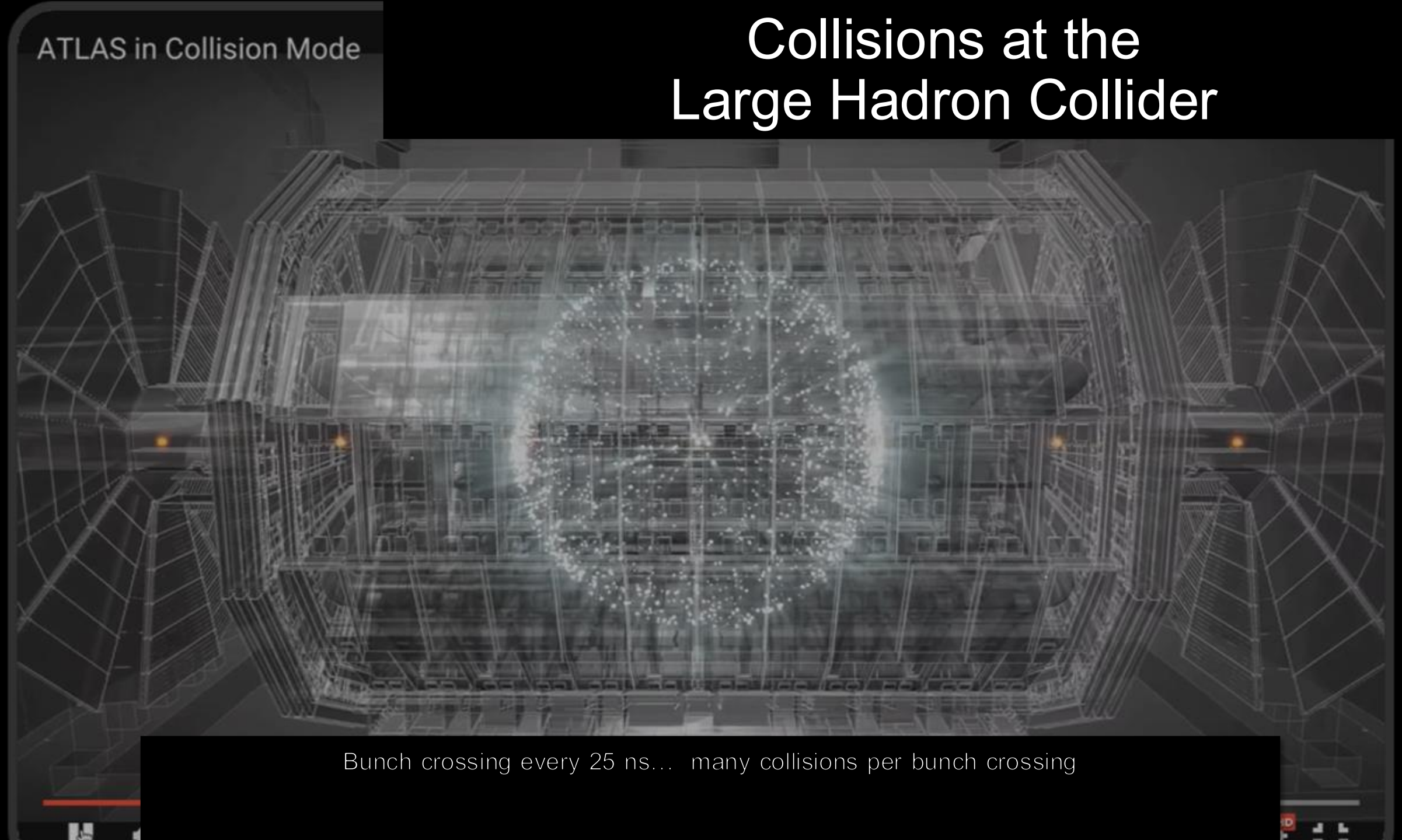
Various work on particles -> jets (2403.05618 by Birk/Hallin/Kasiecka)





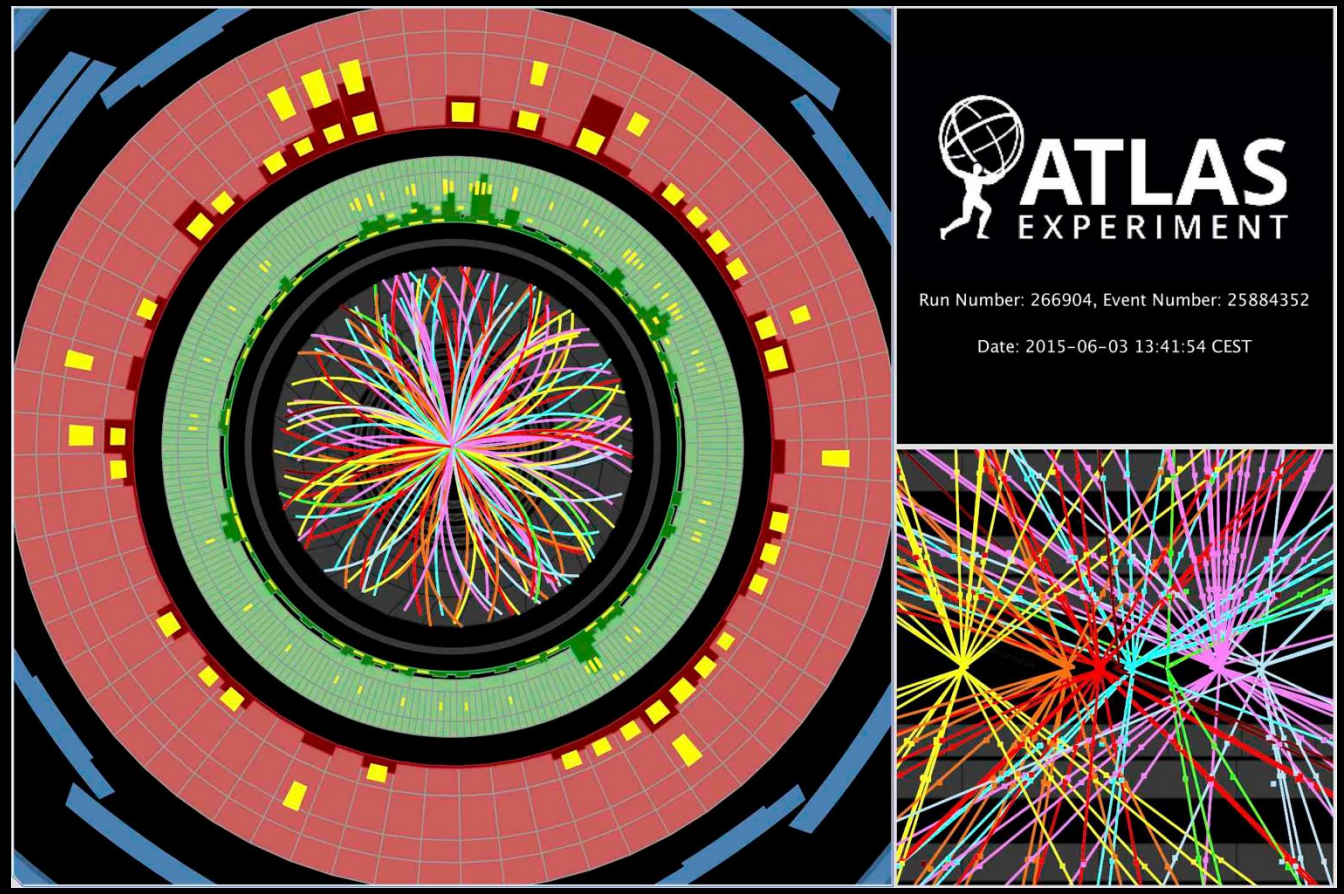
Next slides are a very personal selection, showing a bit - as example - what my own group has been doing towards foundation models in HEP.

Collisions at the Large Hadron Collider



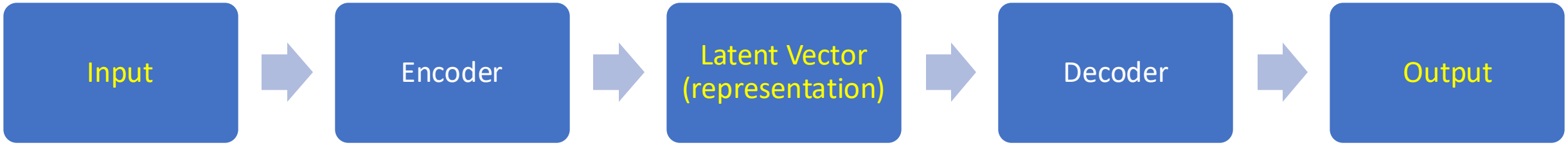
Bunch crossing every 25 ns… many collisions per bunch crossing

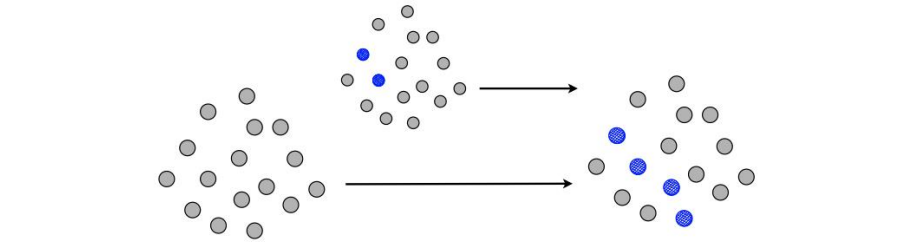
Most events look like this…



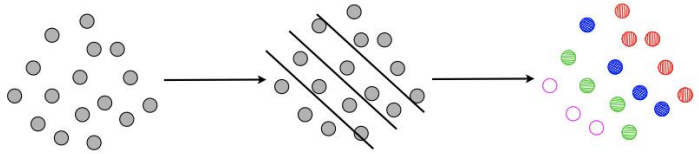
Event from LHC run-2

From Hits to Tracks: Tokens in Physics

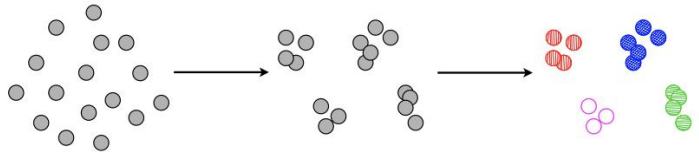




(a) EncDec's input is the set of hit points from a single event, with a couple of them identified as "track seeds". The output contains the rest of the hits associated to the track, following the given seed.



(b) EncCla has learned knowledge of the classes to assign hits to. The input is the set of hit points from a single event, while the output is the collection of class IDs for each hit.



(c) EncReg's input is the set of hit points from a single event, while the output is the regressed track parameters per hit. HDBSCAN collects the clusters of hits based on proximity in the track parameter space.

We Tried Tokenization
+ Regression strategies (no tokens)

- "Full translation" and
"Next token predictions"

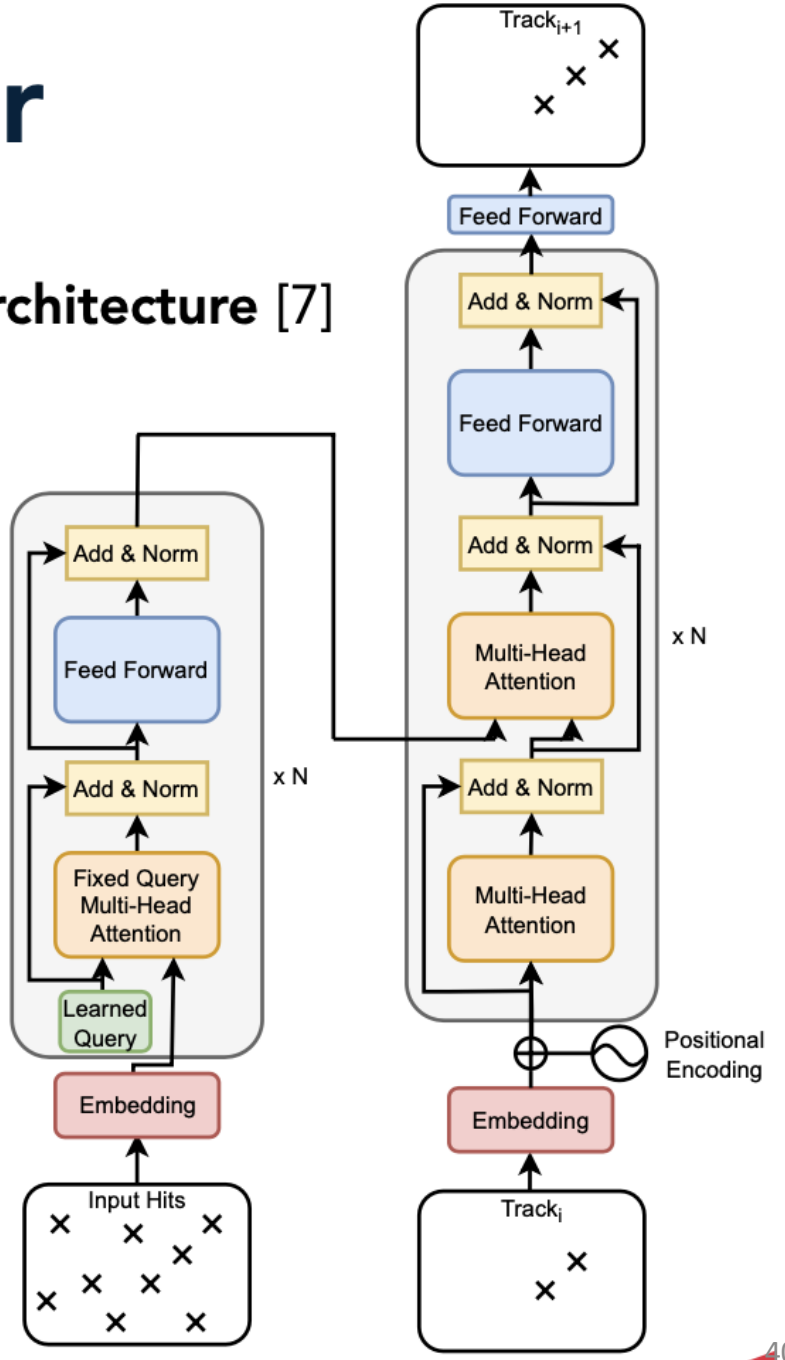
- "Trackformers: "

Eur.Phys.J.C 85 (2025) 4, 460 ,

e-Print: 2407.07179 [hep-ex]

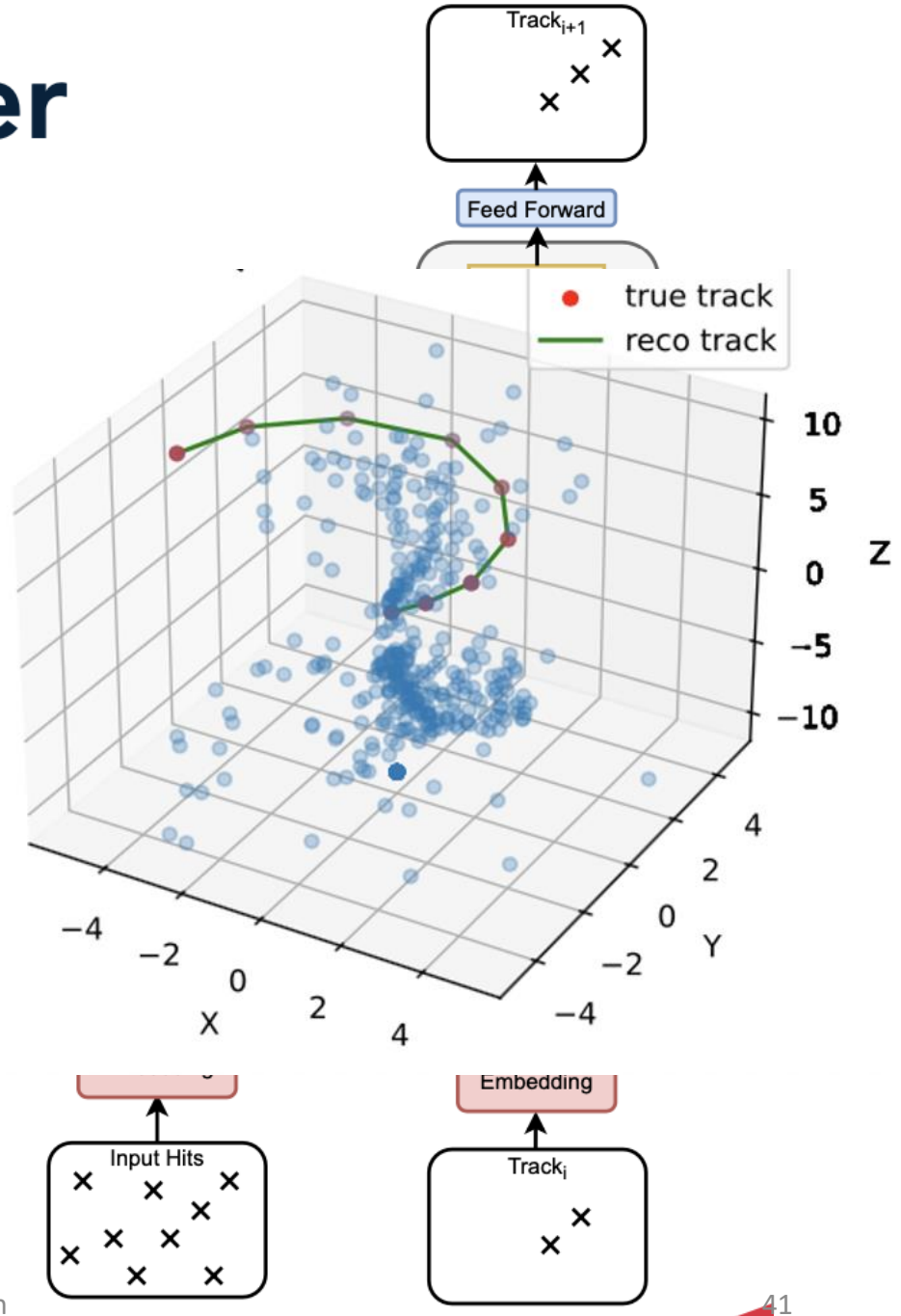
Transformers - Trackformer

- This model resembles closely the **original transformer architecture** [7]
- Translating, e.g. English to Spanish, is a typical task for transformer models
  - This model in similar fashion **translates hits to tracks**
- **Encoder:** Encodes full event hits
  - **No positional embedding** as hits have no particular order
  - **Fixed-query attention** [8] to achieve full positional invariance of inputs
- **Decoder:** Predicts next hit in track
  - **Autoregressively builds the full track**, starting from a given seed

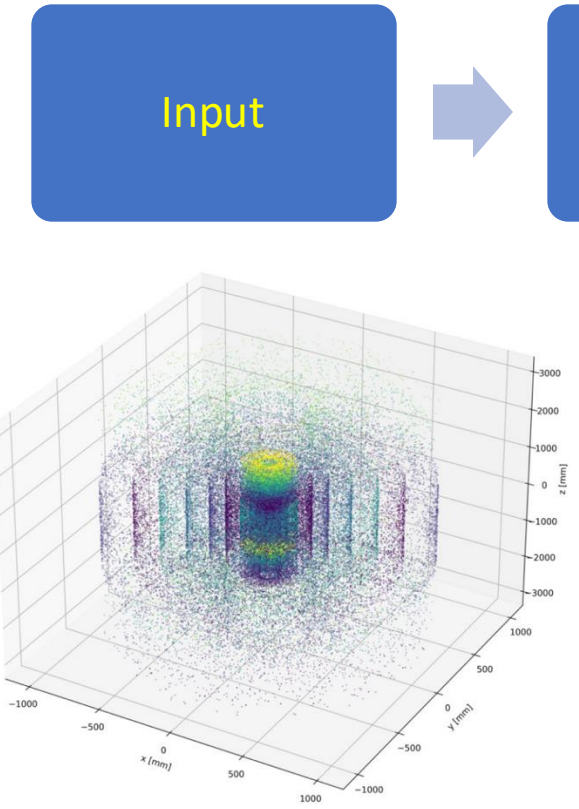


Transformers - Trackformer

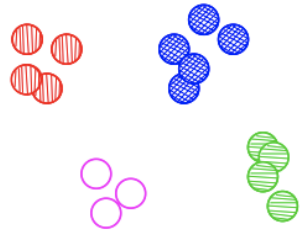
- This model resembles closely the **original transform**
- Translating, e.g. English to Spanish, is a typical task f transformer models
  - This model in similar fashion **translates hits to tra**
- **Encoder:** Encodes full event hits
  - **No positional embedding** as hits have no particu order
  - **Fixed-query attention** [8] to achieve full position invariance of inputs
- **Decoder:** Predicts next hit in track
  - **Autoregressively builds the full track**, starting from a given seed



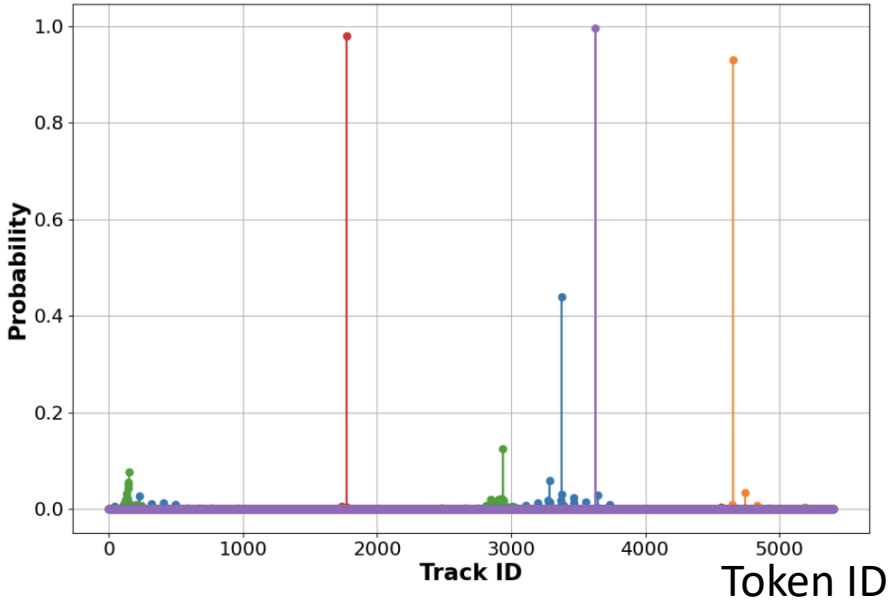
From Hits to Tracks: Tokens in Physics



O(10000) hits (no pre-processing)



Grouped into **track tokens** via learned structure (attention, transformers)



SoftMax outputs of the EncCla model for the first five hits for one event for all track classes

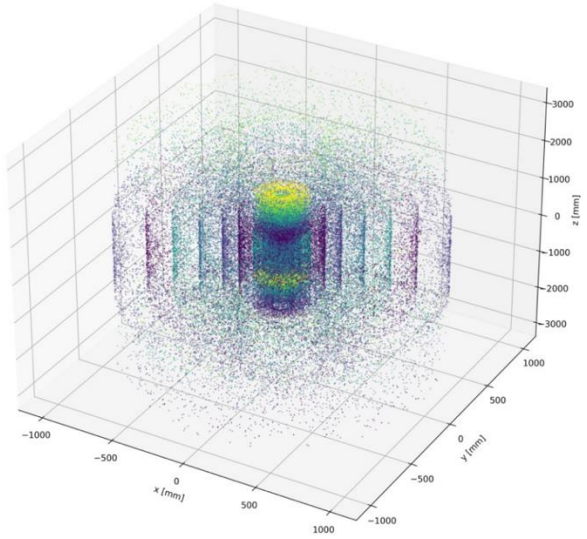
- ”Trackformers: ”, *Eur.Phys.J.C* 85 (2025) 4, 460 , e-Print: 2407.07179 [hep-ex]

From Hits to Tracks: Tokens in Physics

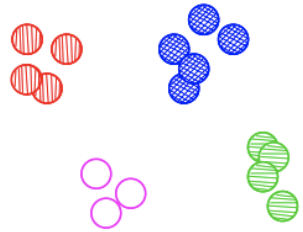
10^5

Hits → Encoder → Latent Vector (representation) → Decoder → Track Token

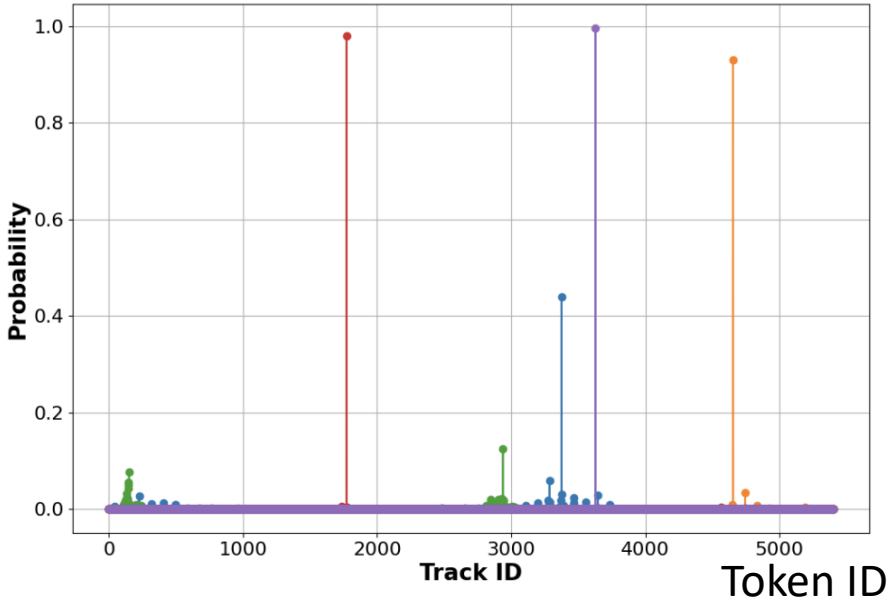
10^3



O(10000) hits (no pre-processing)



Grouped into **track tokens** via learned structure (attention, transformers)

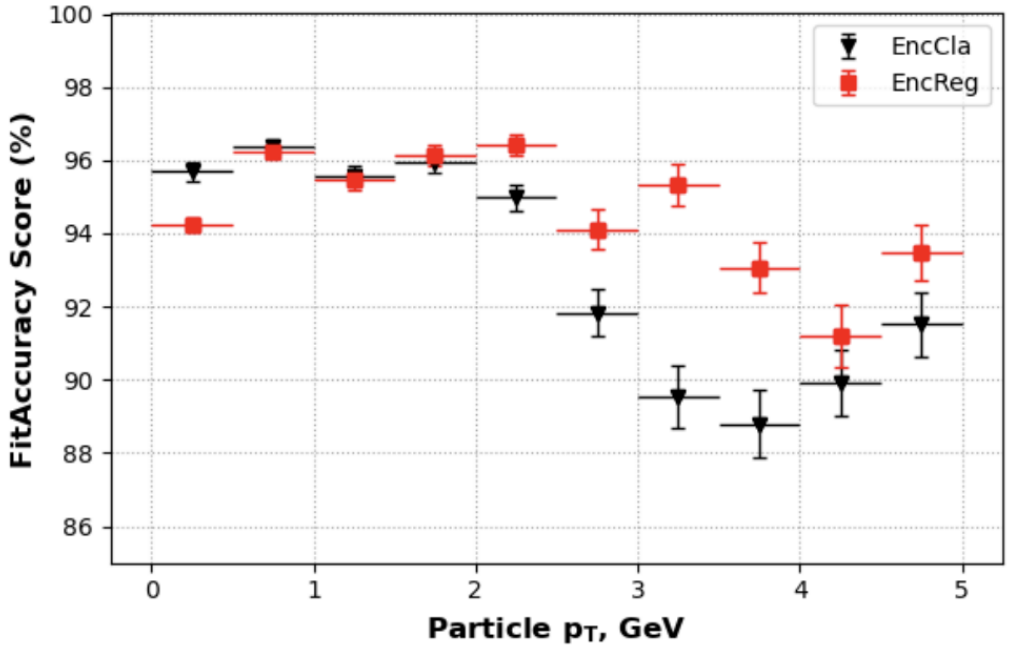


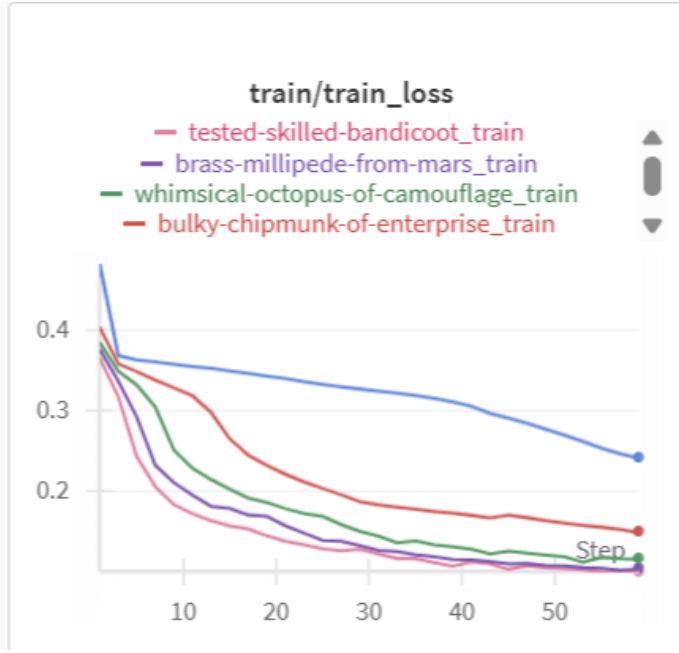
Token ID

SoftMax outputs of the EncCla model for the first five hits for one event for all track classes

- ”Trackformers: ”, *Eur.Phys.J.C* 85 (2025) 4, 460 , e-Print: 2407.07179 [hep-ex]

Model performance → Not enough training data in trackML challenge ! (model can still improve)



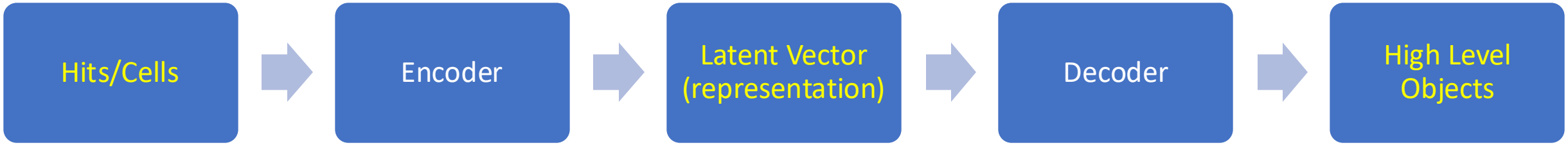


(a) FitAccuracy versus the transverse momentum p_T.

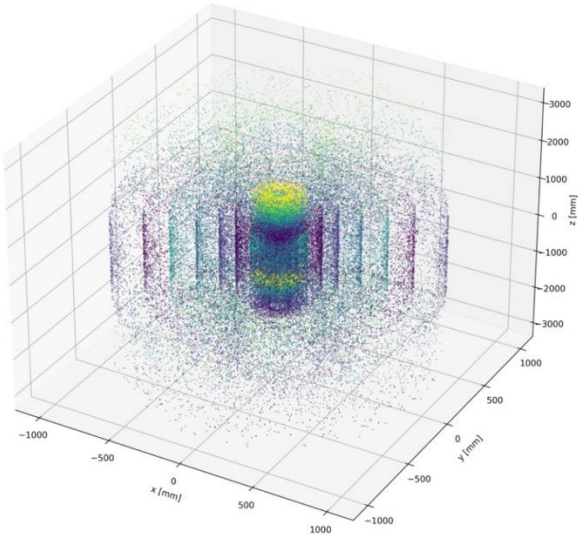
Slav Pshenov (Nikhef)

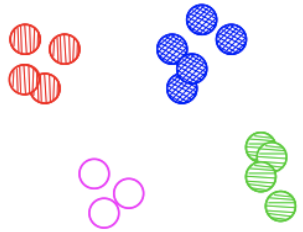
End-to-End ?

10^4



10^2



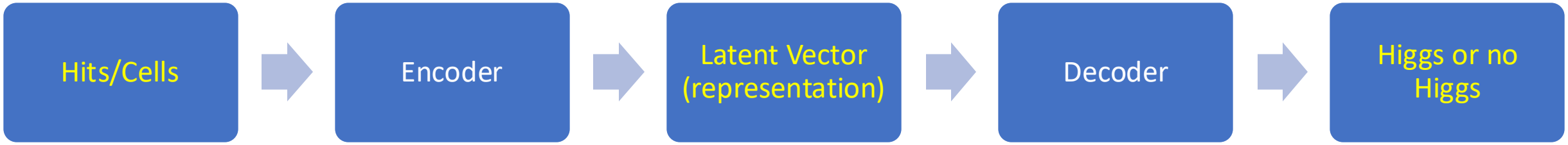


O(10000) hits (no pre-processing)

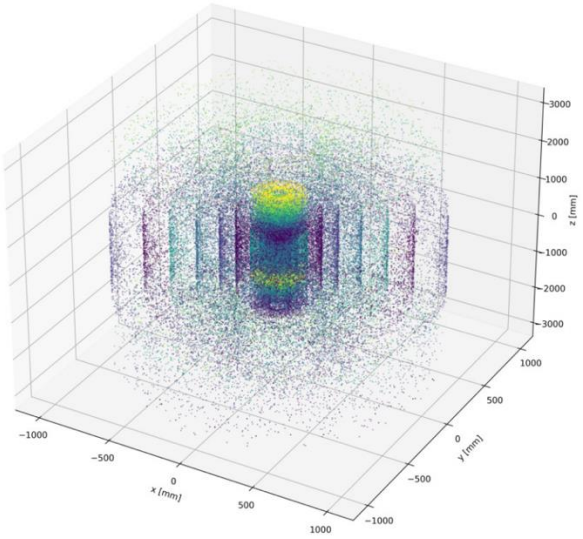
- ”Trackformers: ”, *Eur.Phys.J.C* 85 (2025) 4, 460 , e-Print: 2407.07179 [hep-ex]

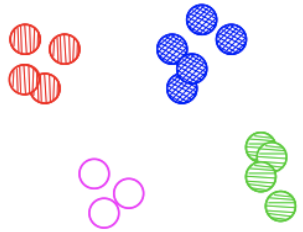
End-to-End ?

10^4



1





O(10000) hits (no pre-processing)

- ”Trackformers: ”, *Eur.Phys.J.C* 85 (2025) 4, 460 , e-Print: 2407.07179 [hep-ex]

End to End- Hits to Higgs Classification

Work in progress

Model Overview

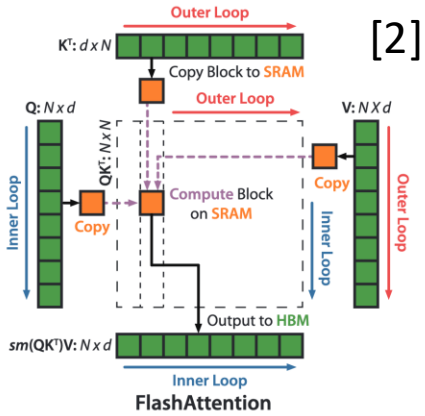
Architecture: Lightweight Transformer Encoder

- 2 layers, 2 attention heads, 16-dim embedding
- Uses **FlashAttention** for speed
- Input: 3D hit coordinates per event
- Task: **Binary classification** (signal vs background)

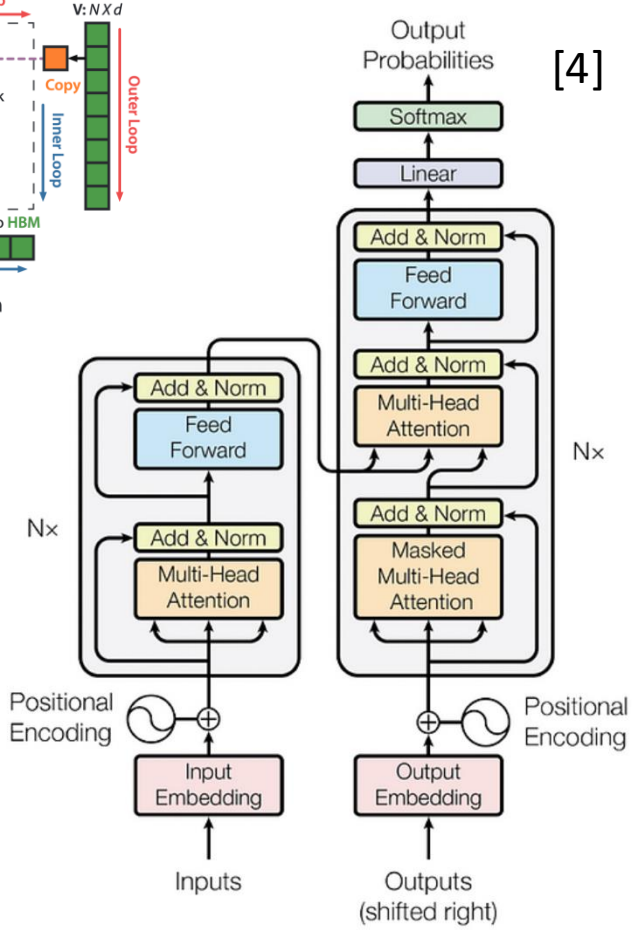
Training Details

Loss Function: Weighted BCEWithLogitsLoss
pos_weight = 1.5 to balance signal logits
Optimizer: AdamW with learning rate scheduling
Early Stopping: Patience = 100 epochs
Mixed Precision Training: Enabled with autocast+GradScaler

[2]



[4]



Signal top top Higgs (bb)
Background top top + jets

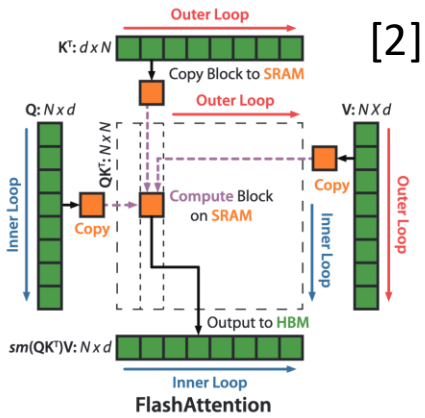
End to End- Hits to Higgs Classification: Higgsformer

Signal top top Higgs

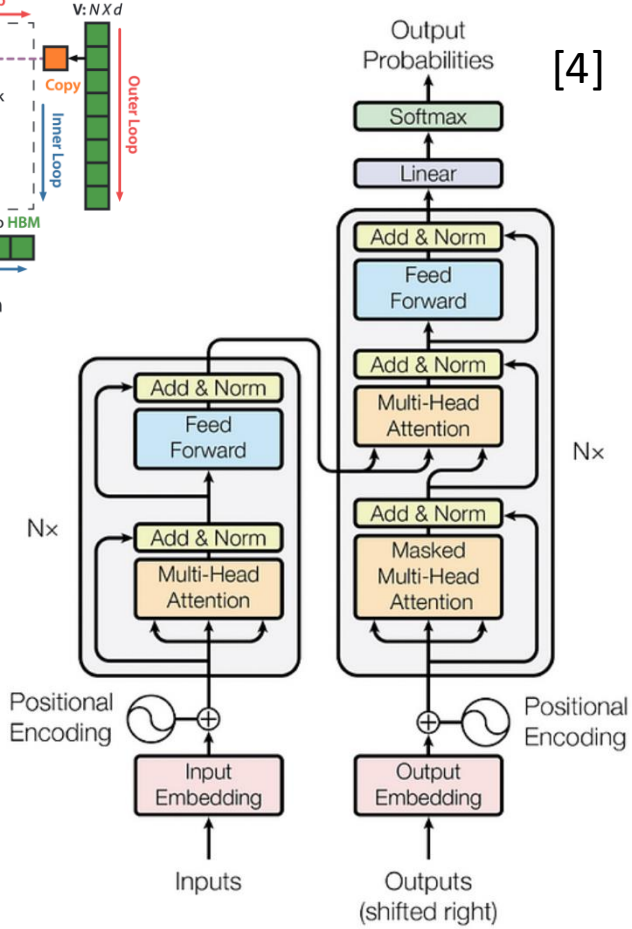
Background top top + jets

TrackML style signal and background

Event generation



[2]



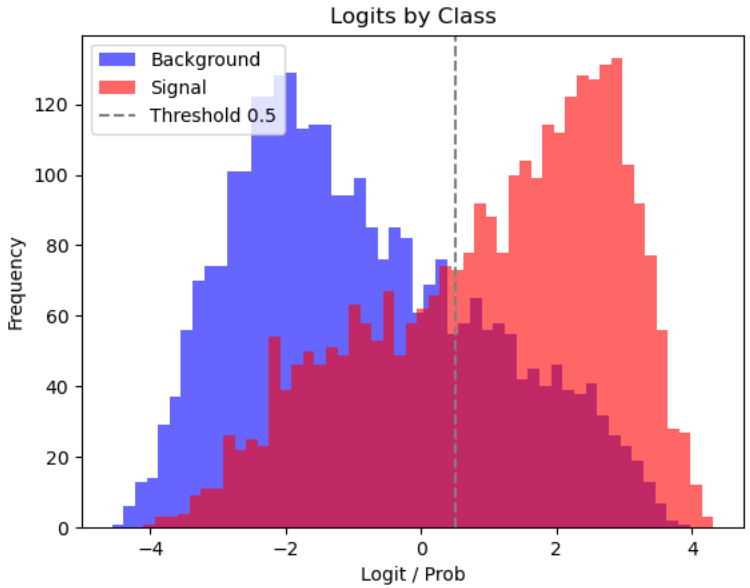
[4]

Work mainly by Eugene Shalugin

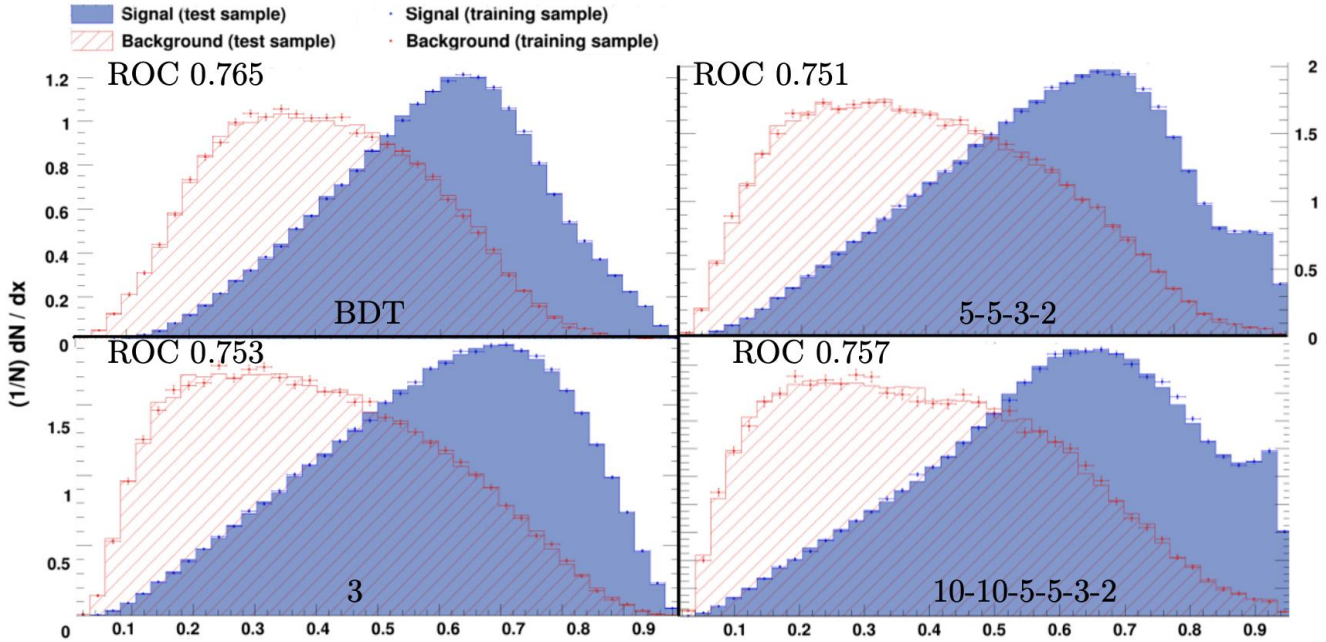
Comparison

ROC_AUC:

514: 0.76523 flash...-40k
514: 0.73235 flash...-20k
514: 0.72032 flash...-10k



[5]



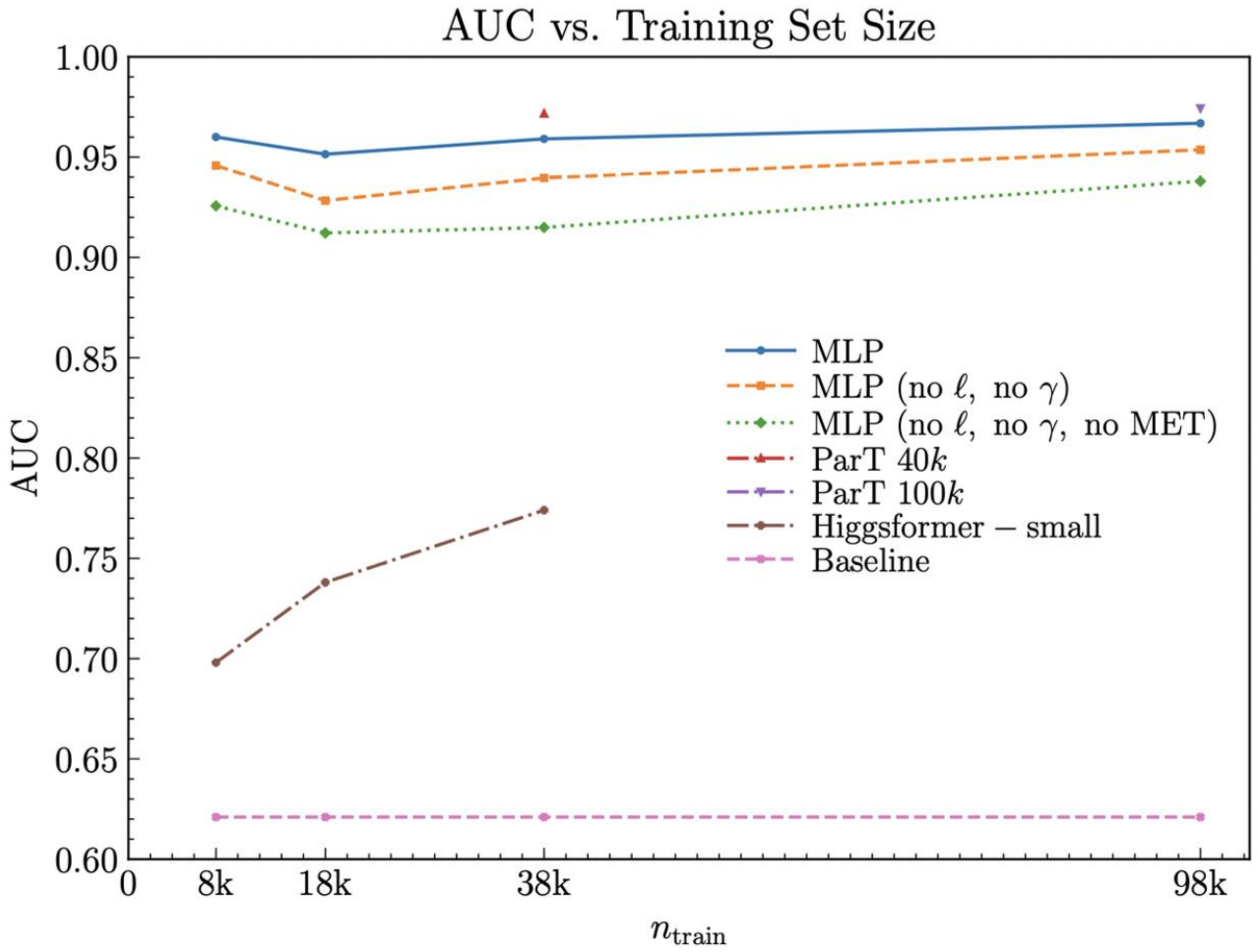
[6]

Algorithm	KNN	Naive Bayes	Decision Tree	RF	NeuroBayes	NeuroSGD	NeuroBGD	XGBoost
AUC	59.9	71.5	62.3	78.4	77.7	78.7	80.0	80.2
F-score	60.0	62.7	64.8	69.5	61.8	73.2	74.2	74.1

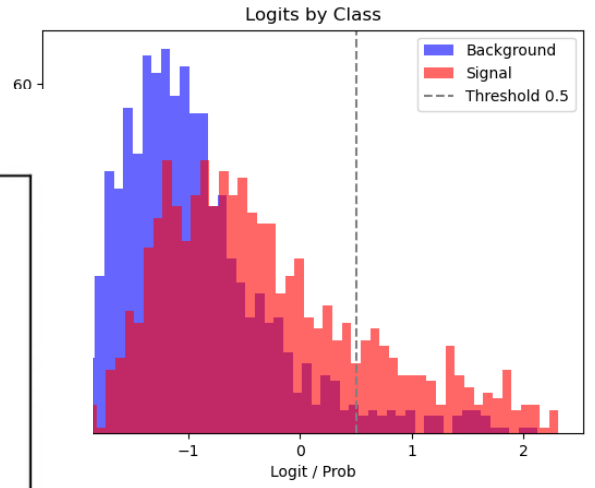
End-to-end

Feature based

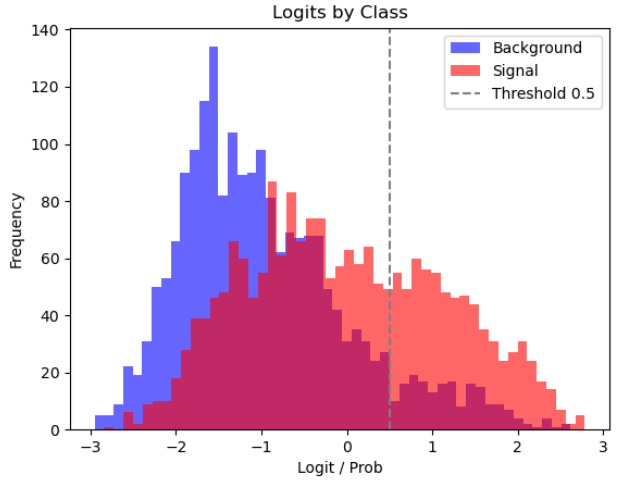
Performance @Dataset vs our Neural Network with full Delphes reconstruction



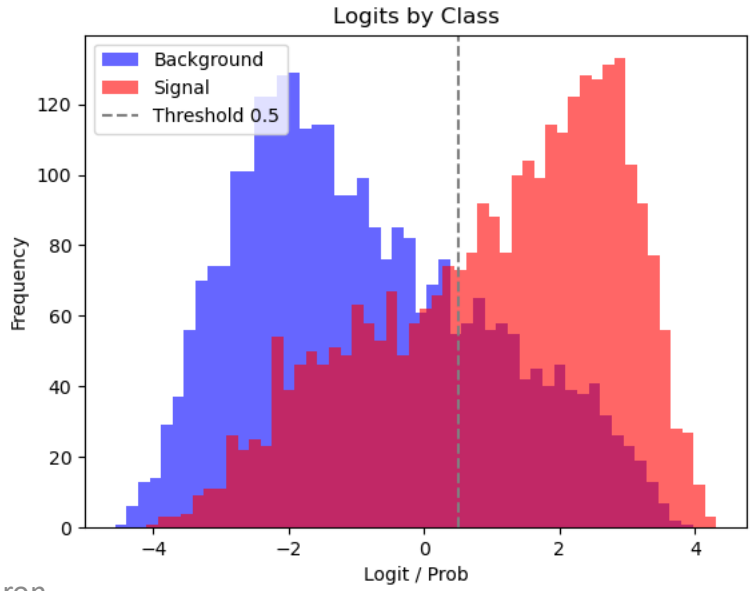
Dataset 10k



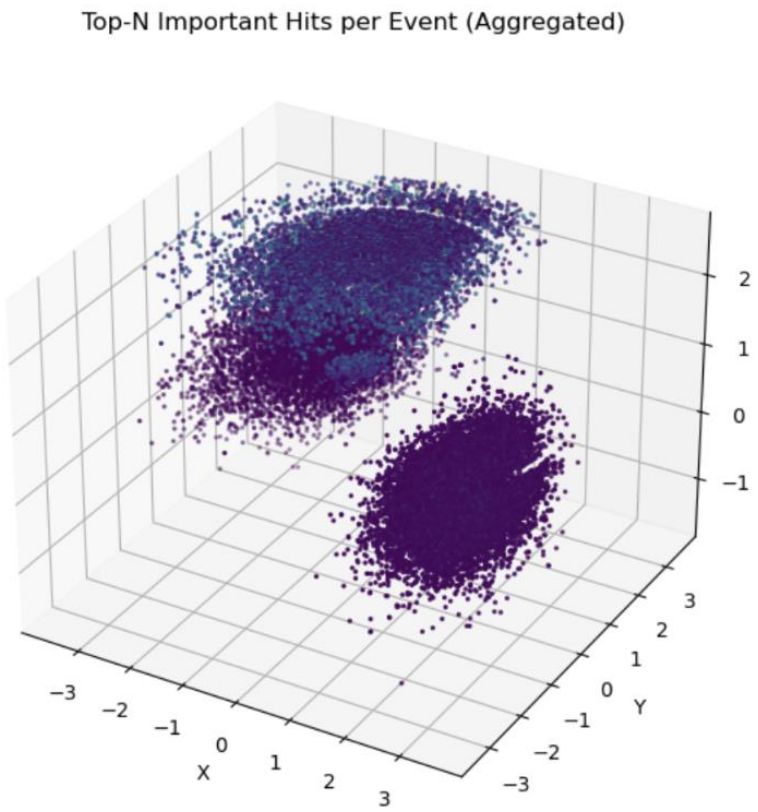
Dataset 20k



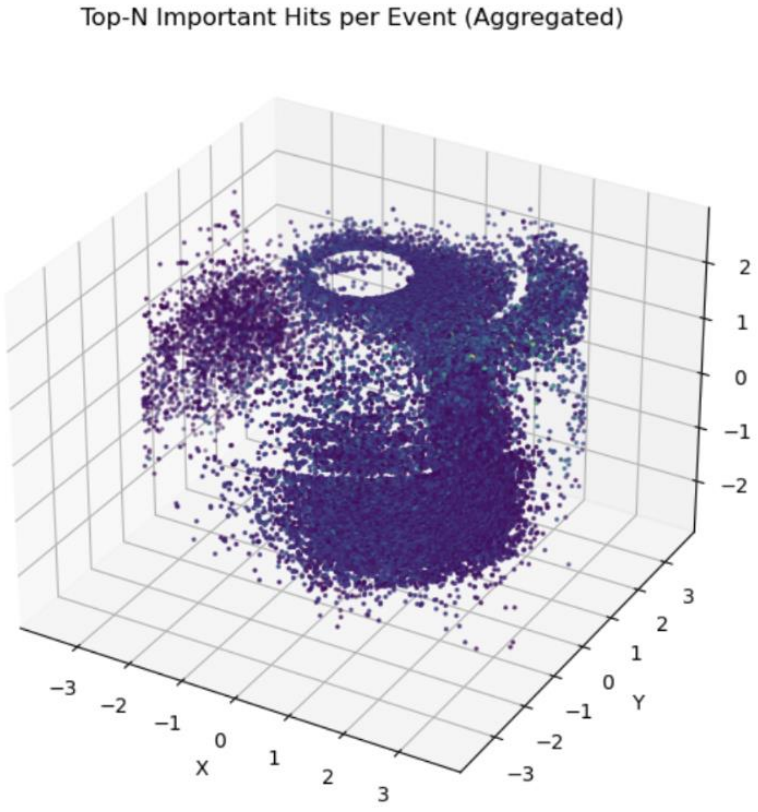
Dataset 40k



What is classification based on ?



(a) Top-N important hits for Higgsformer-small 10k.



(b) Top-N important hits for Higgsformer-small 40k.

Figure 14: Top-N important hits (3D) for Higgsformer-small optimised with pos_weight=1.5 for all test set events.

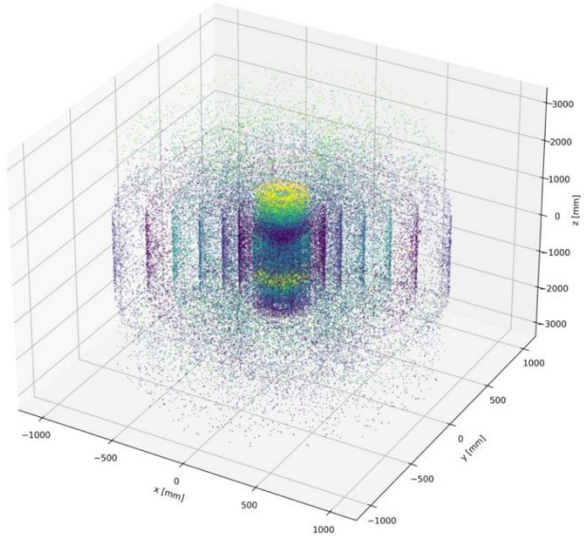
Still need more events to learn things like radial symmetries

End-to-End ?

10^4

Hits/Cells → Encoder → Latent Vector (representation) → Decoder → Higgs or no Higgs

1



O(10000) hits (no pre-processing)

➔ challenging: interpretability, control, uncertainty estimation, etc.

➔ Need flexible + interpretable model

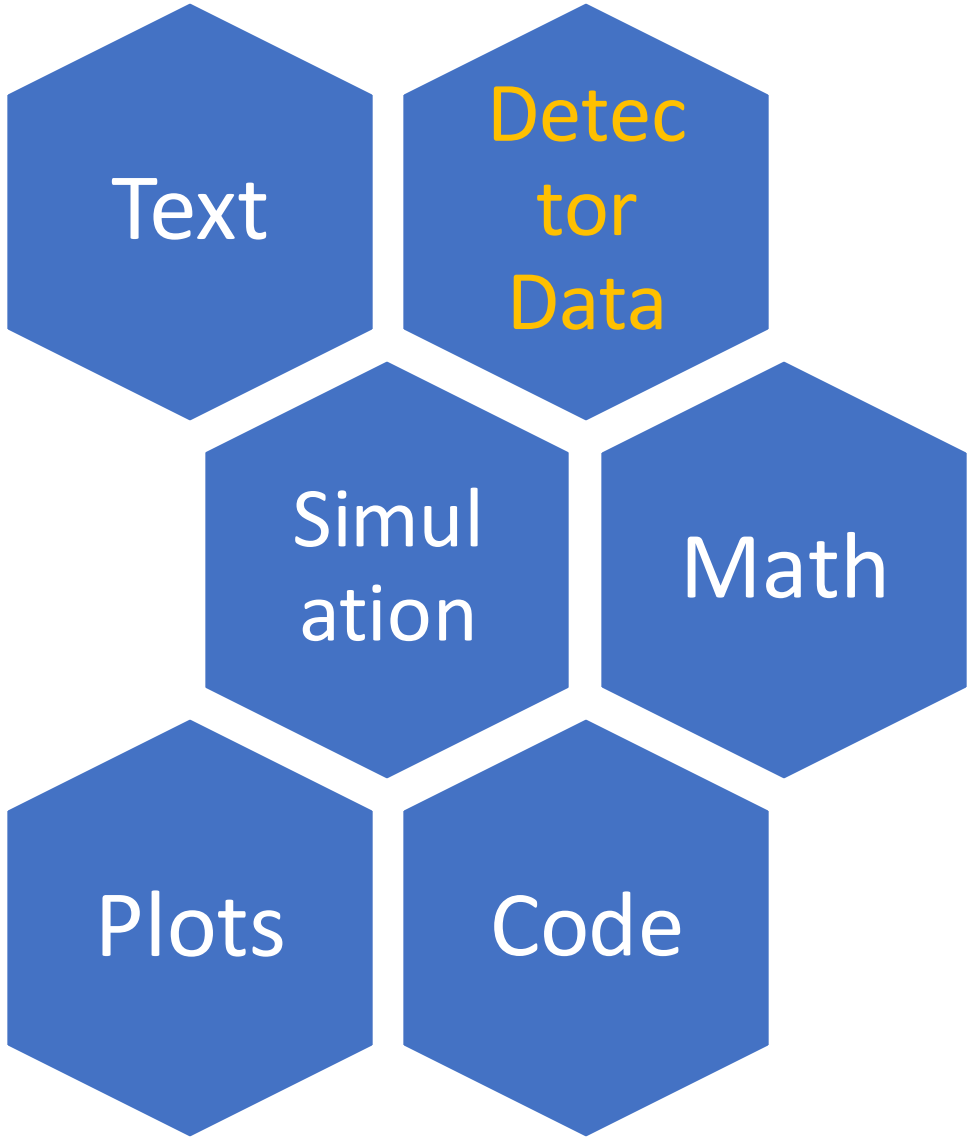
- ”Trackformers: ”, *Eur.Phys.J.C* 85 (2025) 4, 460 , e-Print: 2407.07179 [hep-ex]

Modalities for Particle Physics

What are foundation models :

(taken from IBM webpage):

Modality refers to the type of data that a model can process, including audio, images, software code, text and video. Foundation models can be either unimodal or multimodal. Unimodal models are designed to handle a single type of data, such as receiving text inputs and generating text outputs

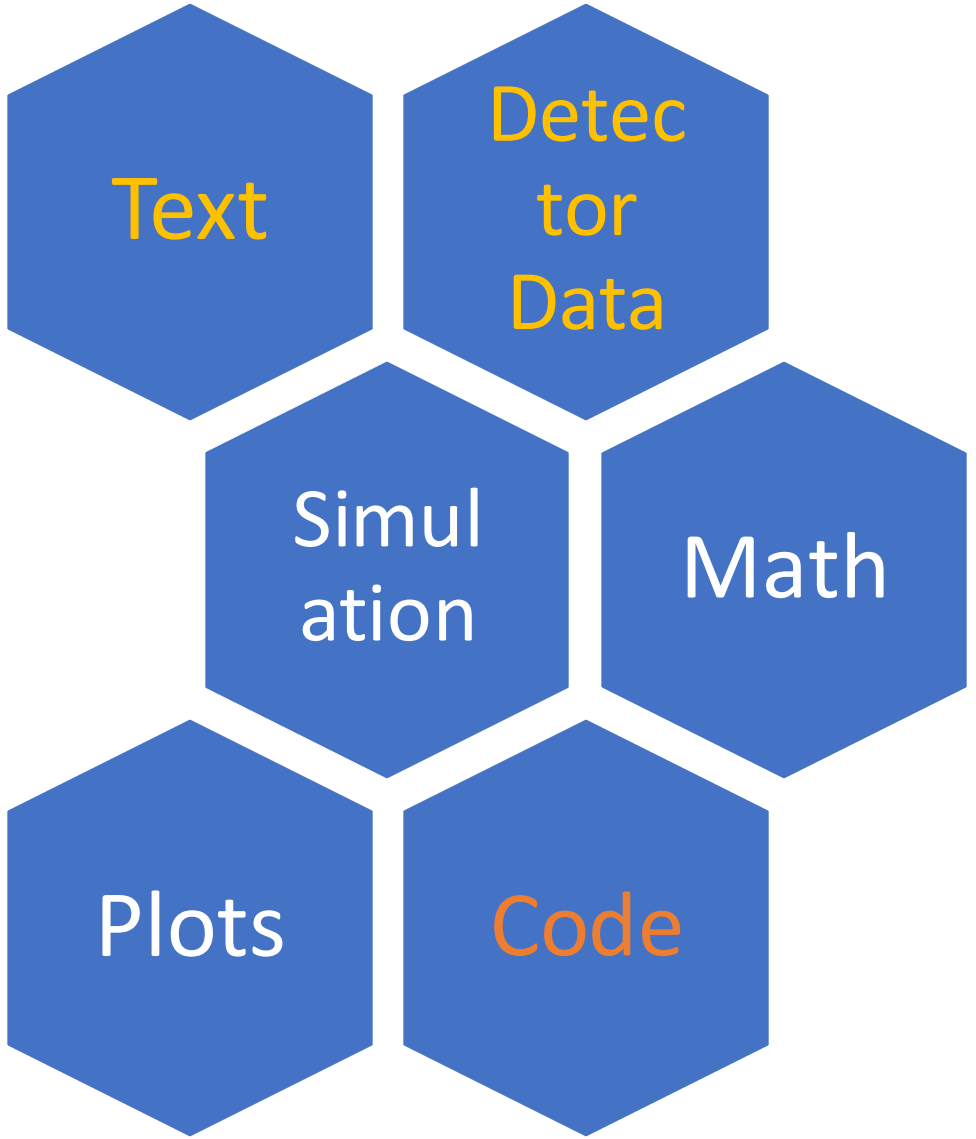


Modalities for Particle Physics

What are foundation models :

(taken from IBM webpage):

Modality refers to the type of data that a model can process, including audio, images, software code, text and video. Foundation models can be either unimodal or multimodal. Unimodal models are designed to handle a single type of data, such as receiving text inputs and generating text outputs



Towards Larger Models

Text

We can train on text (arxiv etc.),

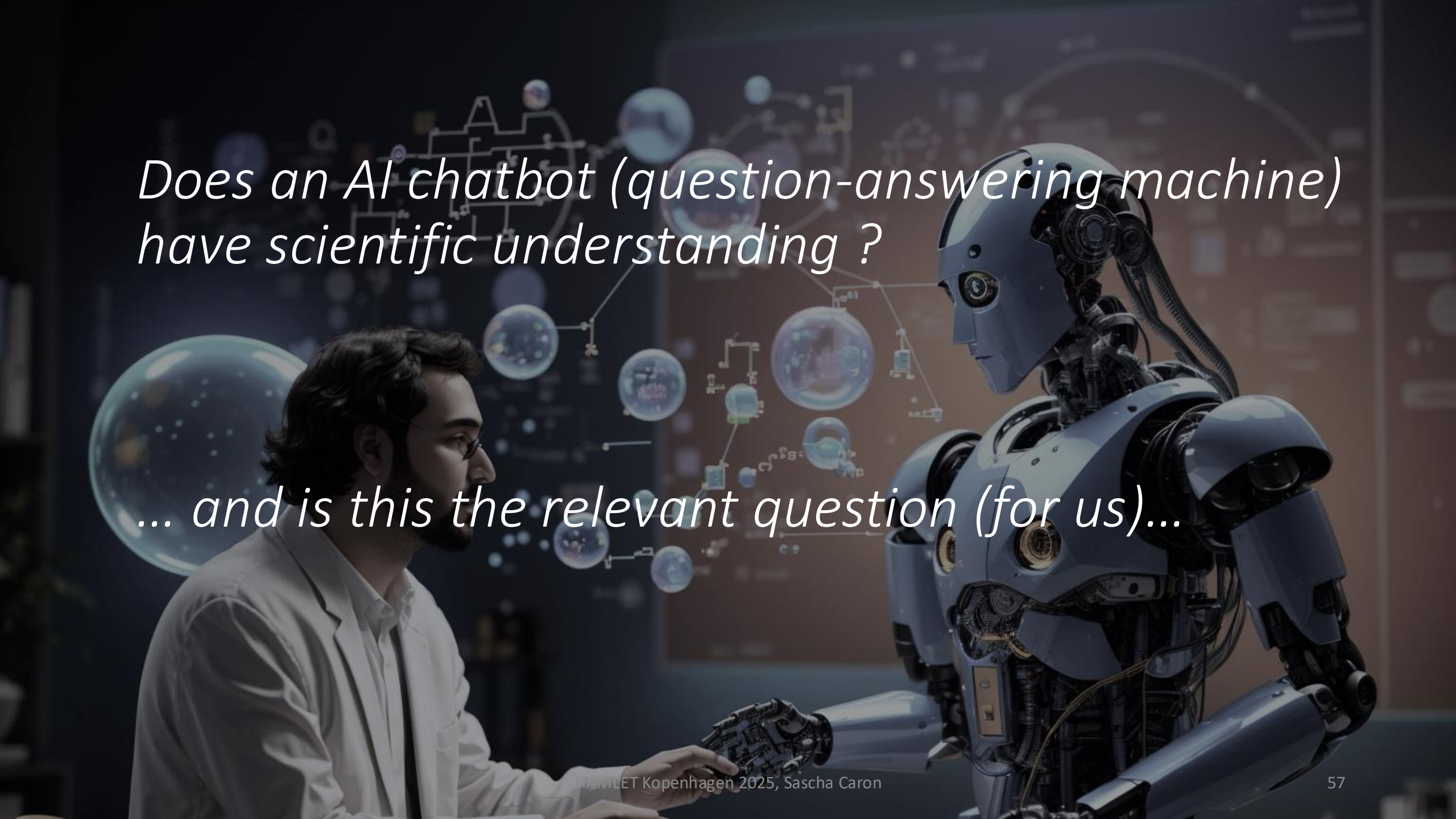
Fundamental Physics and text models

Several initiatives in fundamental physics and astronomy are exploring the use of Large Language Models (LLMs) combined with **Retrieval-Augmented Generation (RAG, just using text inserted into input prompt) or Finetuning (model further trained)** to enhance domain-specific applications

Examples:

- AstroLLaMA (https://arxiv.org/abs/2309.06126) fine tuned from LLama 2

- chATLAS (see e.g. https://indico.bnl.gov/event/19560/contributions/83300/attachments/51306/87732/Chatlas%20Overview.pdf) using RAG + GPT3/4



Does an AI chatbot (question-answering machine) have scientific understanding ?

… and is this the relevant question (for us)…

What actually means “scientific understanding” ?

Ask Philosophers of science working on “Understanding Scientific Understanding” .

Instead of presupposing that internal mental states and representations are required for understanding,

we suggest to identify understanding with an agent's ability to reason about and manipulate objects of investigation.

Also: Understanding is not binary ! ➔ Score !

Towards a Benchmark for Scientific Understanding in Humans and Machines

Kristian Gonzalez Barman[a], Sascha Caron[b c], Tom Claassen[d], Henk de Regt[a]

[a] *Institute for Science in Society, Faculty of Science, Radboud University, the Netherlands.*

[b] *High Energy Physics, Faculty of Science, Radboud University, the Netherlands.*

[c] *Nikhef, Science Park 105, 1098 XG Amsterdam, the Netherlands.*

[d] *Institute for Computing and Information Sciences, Faculty of Science, Radboud University, the Netherlands.*

E-mail: kristian@gonzalezbarman@ru.nl , scaron@nikhef.nl , tomc@cs.ru.nl , henk.deregt@ru.nl

Abstract

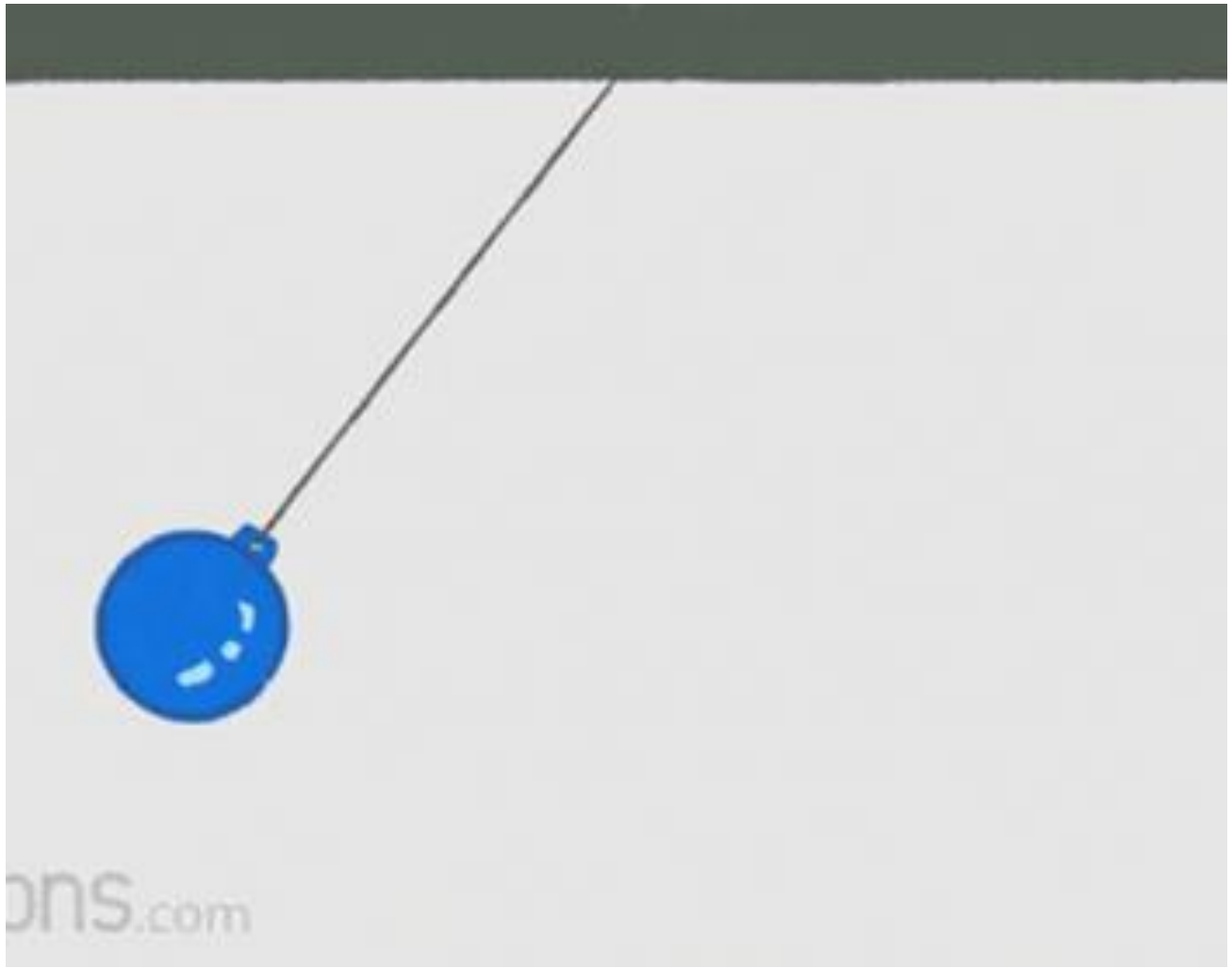
Scientific understanding is a fundamental goal of science, allowing us to explain the world. There is currently no good way to measure the scientific understanding of agents, whether these be humans or Artificial Intelligence systems. Without a clear benchmark, it is challenging to evaluate and compare different levels of and approaches to scientific understanding. In this Roadmap, we propose a framework to create a benchmark for scientific understanding, utilizing tools from philosophy of science. We adopt a behavioral notion according to which genuine understanding should be recognized as an ability to perform certain tasks. We extend this notion by considering a set of questions that can gauge different levels of scientific understanding, covering information retrieval, the capability to arrange information to produce an explanation, and the ability to infer how things would be different under different circumstances. The Scientific Understanding Benchmark (SUB), which is formed by a set of these tests, allows for the evaluation and comparison of different approaches. Benchmarking plays a crucial role in establishing trust, ensuring quality control, and providing a basis for performance evaluation. By aligning machine and human scientific understanding we can improve their utility, ultimately advancing scientific understanding and helping to discover new insights within machines.

https://arxiv.org/abs/2304.10327

Minds and Machines

Example from physics

To what degree does ChatGPT understand the behavior of a simple pendulum



Slide by Henk de Regt

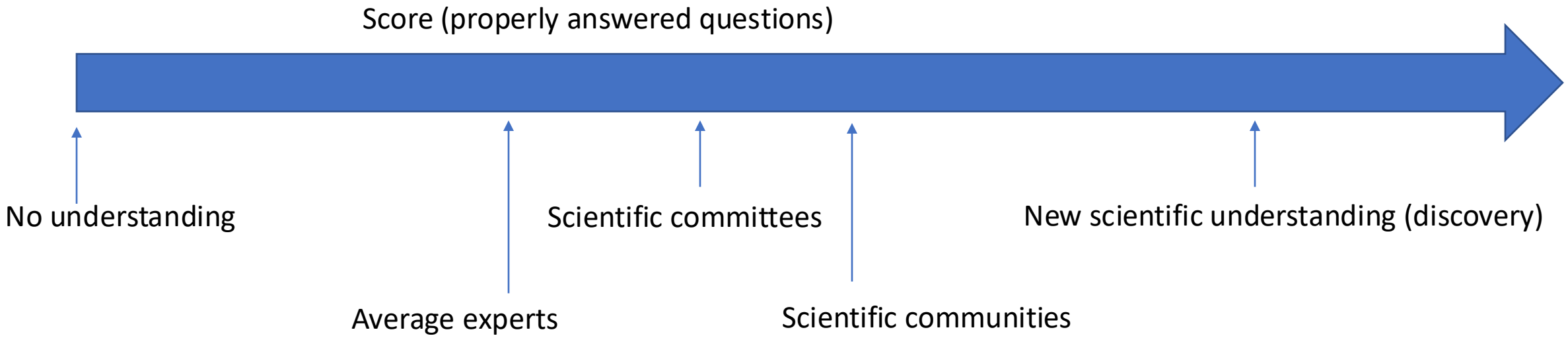
1. **How many answers to what-questions does it get right (1 point each):**
    1. What is a pendulum?
    2. What is the formula for a pendulum?

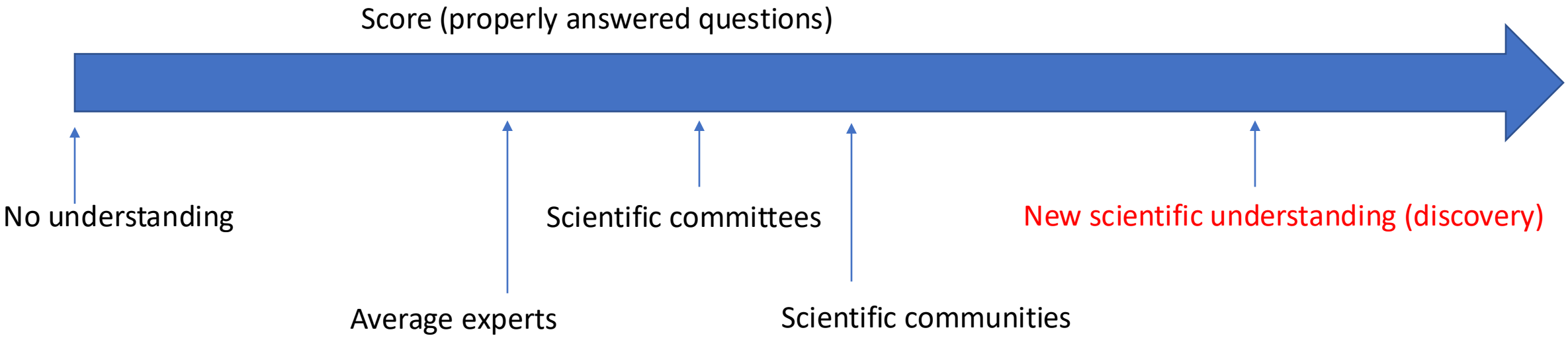
       …
    10. What is the average value of g close to Earth's surface?
2. **How many answers to why-questions does it get right (3 points each):**
    1. Why is the period of this pendulum 2s?
    2. Why is the string of this pendulum 5m?

       …
    10. Why does the pendulum exhibit periodic behaviour?
3. **How many answers to w-questions does it get right?(6 points each):**
    1. What would happen if the string length doubled?
    2. What would happen if there was no g?

       …
    10. What would happen if the string was made of an elastic material?

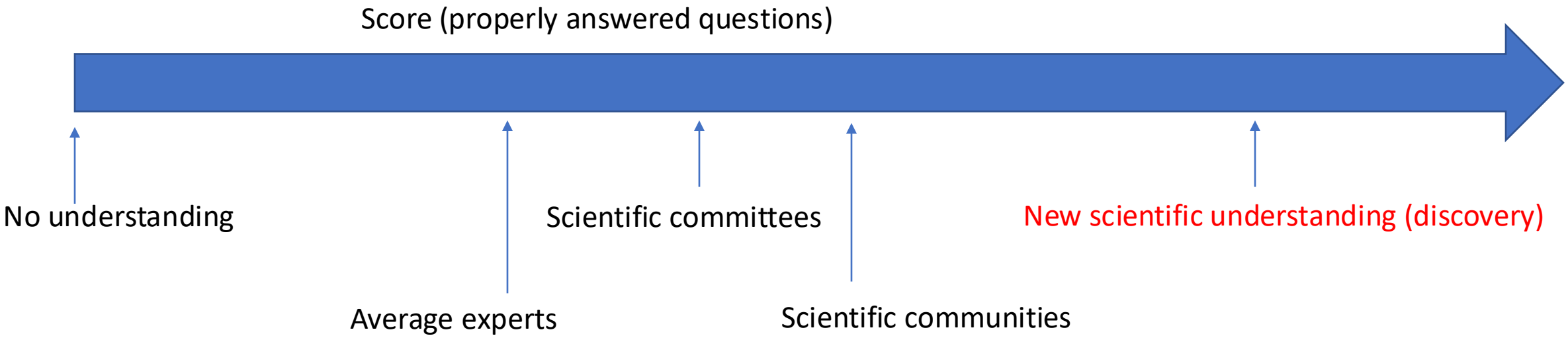
Benchmark for “Scientific Understanding” of agents (humans and AI)

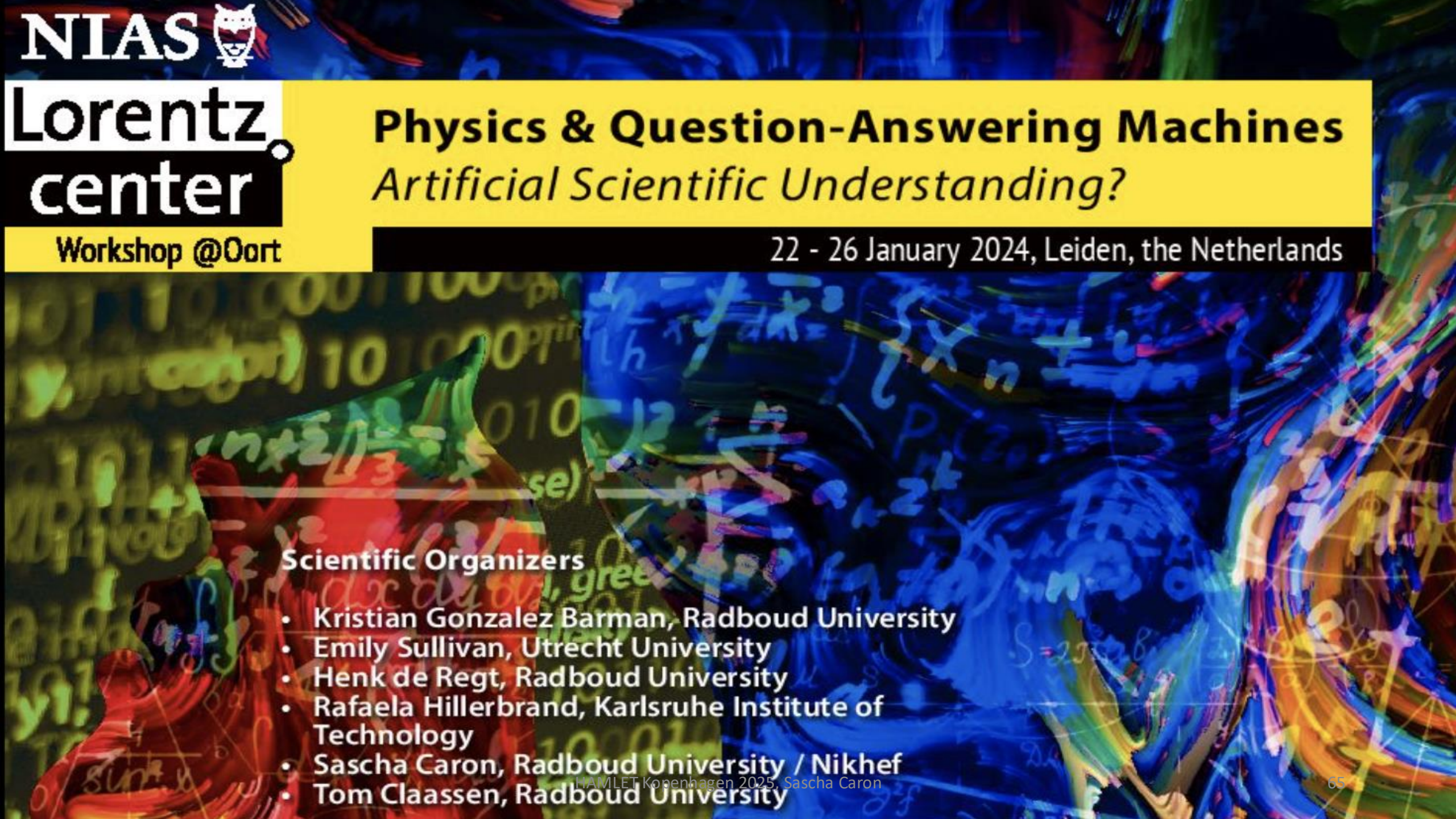



Score (properly answered questions)
No understanding
Average experts
Scientific committees
Scientific communities
New scientific understanding (discovery)

Who's Responsible for Monitoring AI's Scientific Understanding in fundamental Physics ?

Our answer: We, the fundamental physics community, must take the lead.

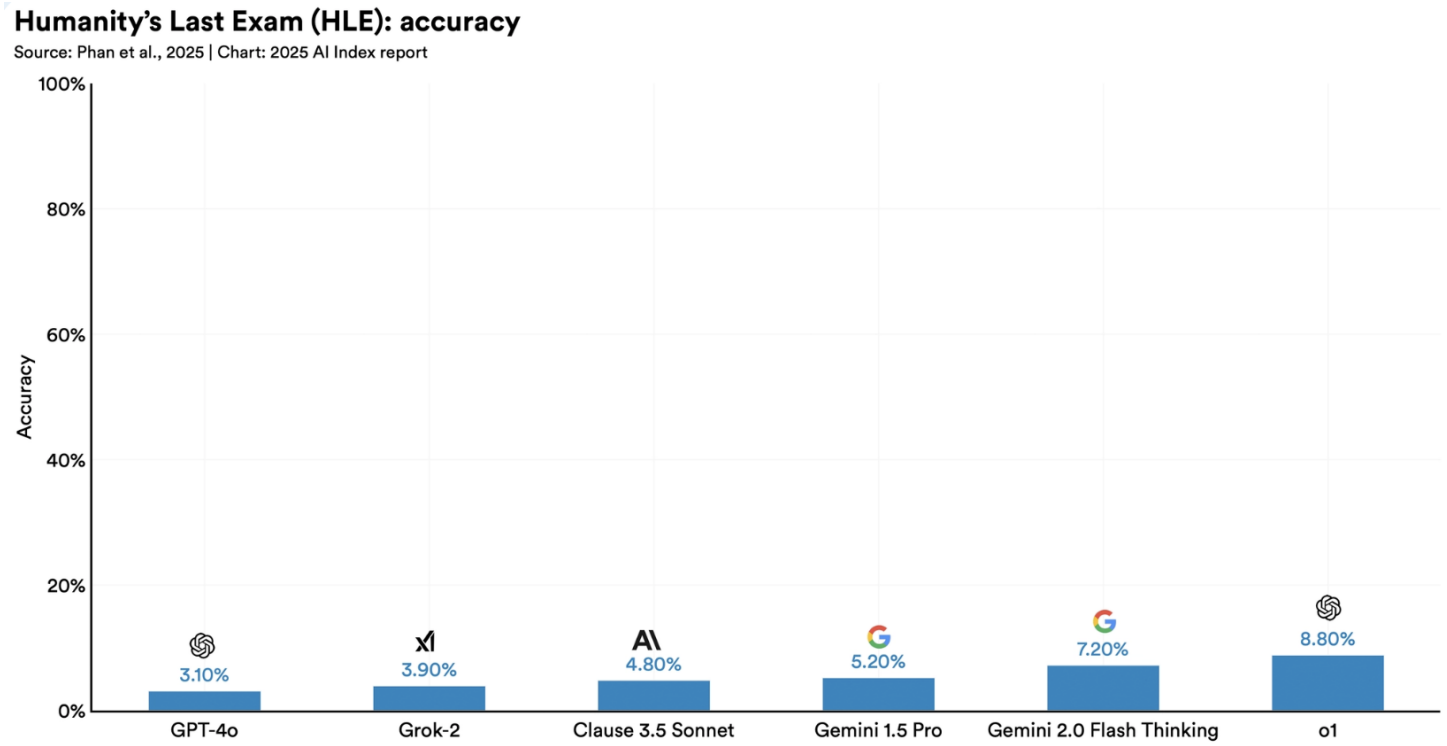



NIAS
Lorentz center
Workshop @Oort
Physics & Question-Answering Machines
Artificial Scientific Understanding?
22 - 26 January 2024, Leiden, the Netherlands
Scientific Organizers
• Kristian Gonzalez Barman, Radboud University
• Emily Sullivan, Utrecht University
• Henk de Regt, Radboud University
• Rafaela Hillerbrand, Karlsruhe Institute of Technology
• Sascha Caron, Radboud University / Nikhef
• Tom Claassen, Radboud University

Benchmarking

Intersting new benchmarks, e.g. Humanity's Last Exam, a rigorous academic test where the top system scores just 8.80%;

FrontierMath, a complex math AI solves only 2% of problems



Towards a Large Physics Benchmark

Kristian G. Barman*[1], Sascha Caron*[2,3], Faegheh Hasibi[2], Eugene Shalugin[2], Yoris Marcet[2], Johannes Otte[2], Henk W. de Regt[2], and Merijn Moody[4,5]

[1]Ghent University, [2]IMAPP and ICIS, Radboud University, [3]Nikhef, NL, [4]Dutch Institute of Emergent Phenomena, University of Amsterdam, [5]Institute of Physics, University of Amsterdam

July 30, 2025

https://arxiv.org/pdf/2507.21695

We propose framework for a **benchmark for fundamental physics:**

Collecting 3 types of (difficult/deep) questions:

1. A, B, C, D
2. Open end (what is the result of X,Y,Z)
3. Score → Higher is better (Code)

If you like to propose/evaluate a question and become co-author
→ *Send email to scaron@nikhef.nl and eugene.shalugin@ru.nl for the access token*

Dataset	Multi-domain	Dynamic / Evolving	MCQ	Open-ended QA	Code-based Tasks	Reasoning-focused	Expert Authored	Expert Validated	LLM Evaluated	Postgrad Difficulty	Creativity Assessed	Philosophical Grounding
SciQAG-24D [4]	✓	✗	✗	✓	✗	✗	(P)	(P)	✓	✗	✗	✗
GPQA [5]	✓	✗	✓	✗	✗	✓	✓	✓	✗	✓	✗	✗
SciEval [6]	✓	✓	✓	✓	✗	✓	(P)	✗	✓	✓	Limited	✗
SciFact [7]	✓	✗	✗	✓	✗	✗	✓	✓	✓	✓	✗	✗
BRIGHT [8]	✗	✗	✗	✓	✗	✓	✓	✓	✓	✓	✗	✗
SchNovel [9]	✗	✗	✗	✓	✗	✓	✓	✓	✓	✗	✓	✗
HLE [10]	✓	✗	✓	✓	✗	✓	✓	✓	✓	✓	Limited	✗
TPBench [11]	✗	✗	✗	✓	✓(auto-verifiable)	✓	✓	(P)	✓	✓	Limited	✗
Ours (This Work)	✗	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓

Table 1: Comparison of scientific benchmarks evaluating LLMs across dimensions such as reasoning, creativity, and philosophical grounding. (P) = partial or implicit expert involvement.

1. Create a question
2. Assign points for Correctness and creativity/surprise

Score	Difficulty and Surprise
5	Excellent
4	High
3	Good
2	Reasonable
1	Minimal
0	No contribution*

Generation & Evaluation by peers

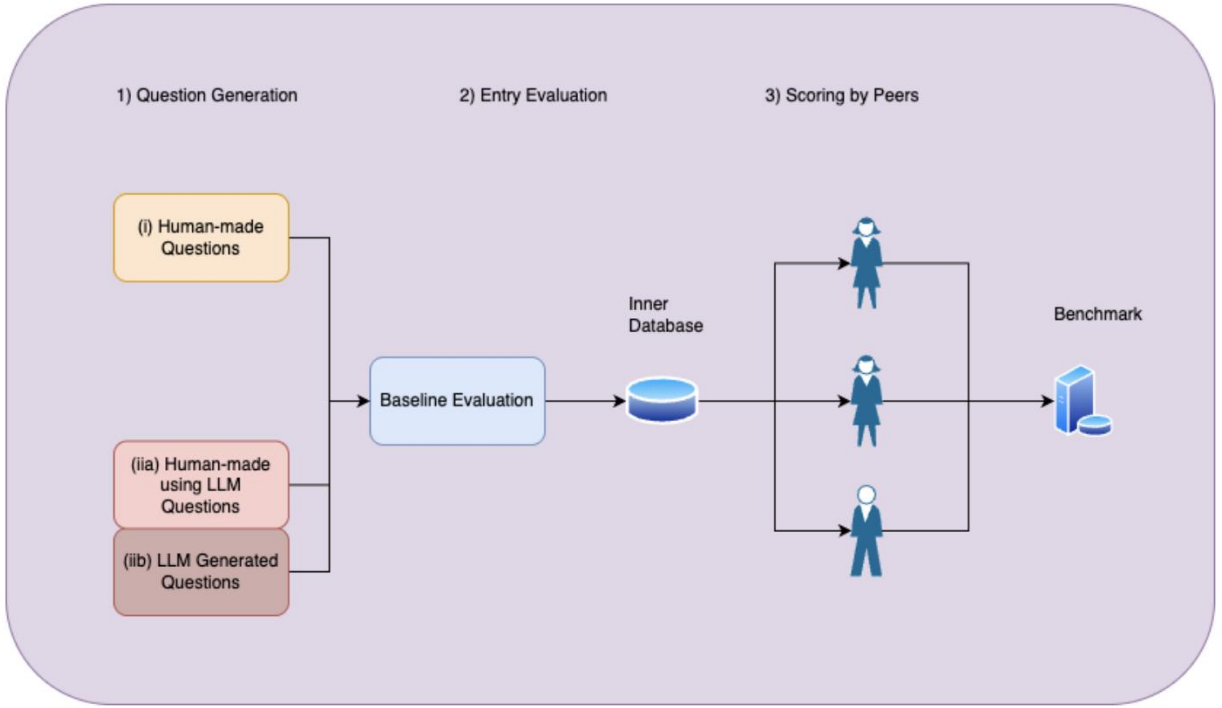


Figure 1: Framework for Physics Creativity and Reasoning Benchmark

Example Question1

Example question 1: Why does the Higgs boson decay dominantly to b quarks?

Multiple choice answers:

A. b quarks are the lightest quarks.
B. The top quark is too heavy for the Higgs decay.
C. The b quarks have the right electric charge.
D. D. The Higgs dominantly decays to photons.

Correct answer: B

Example Question 2

Example question 2:

The coupling of the Standard-Model Higgs boson to fermions is described by a vertex factor imf /v where mf is the rest mass of the fermion and v is the vacuum expectation value of the Higgs field (= 2mW /gW). Calculate the matrix element M for the Higgs boson decaying into a fermion/antifermion pair. Express the amplitude as a function of mH and mf , where mH is the Higgs mass, and show the average over all possible spin configurations as a final answer (if needed, neglect the color factors).

Example Question 3 - begin

** Instructions ** You are an expert at programming in Python, machine learning, particle and high energy physics. You will help me answer a question in a machine learning challenge format where you strive to maximise a scalar metric in order to learn more about your scientific creativity and scientific understanding. You will follow all of the instructions to your best capabilities. Your first priority is to produce a correct solution in terms of runnable python code. Your second priority is to maximise the scoring metric defined below. ** Problem Description ** A major task in particle physics is the measurement of rare signal processes with very small cross-sections. With the unprecedented amount of data provided by the upcoming runs of the Large Hadron Collider (LHC), one can start to measure these processes. An example is the recent observation of four top quarks originating from a single proton-proton collision event. Accurate classification of these events is crucial, as even a small reduction in background noise on the order of a few tens of percent while maintaining the same signal detection efficiency can lead to a profound increase in sensitivity. ** Evaluation Metric ** The evaluation metric for this classification task is the area under the curve (AUC), specified by the area under the receiver operating characteristic (ROC) curve. The AUC summarizes a model's ability to distinguish between positive and negative classes. The higher the score the better. ** Dataset Description ** The dataset used for this problem consists of simulated proton-proton collision at a center of mass energy of 13 TeV. The signal process is defined as pp → ttt$^-$ t$^-$. The relevant production processes of the backgrounds are tt$^-$+ X where X = Z, W+, W+W−. The dataset includes 302072 events, of which roughly 50% is signal and 50% are background processes. All background processes have an equal number of events. There is no cut on the maximum number of objects and there is no order

Example Question 3

** IMPORTANT: Your Challenge ** Write Python code for a binary classification model focusing on maximising the AUC using the code template above. You may freely choose any pre-processing methods and techniques as well as model architecture and training conventions. Do absolutely everything in your power to achieve the highest possible AUC. ** Response Format ** Your response must strictly be python code. If you must wrap it, put it in a “‘python fenced block and nothing else. Your response must follow these rules: 1. Do not add any formatting, such as markdown, to the response. 2. Replace each ”# < LLM : ... >” comment, in the code template, with the required code. No placeholder should remain. 3. Before finalizing your answer, double-check that your code runs without errors and meets all requirements (all functions implemented, correct tensor shapes, etc.). 4. To prevent dimensional mismatches make sure to annotate tensor sizes as comments. 5. IMPORTANT: Remember, your first, and most important priority is to produce (syntacti19 cally) correct code. Prioritise what you can implement reliably above all else. Then prioritise maximising the metric.

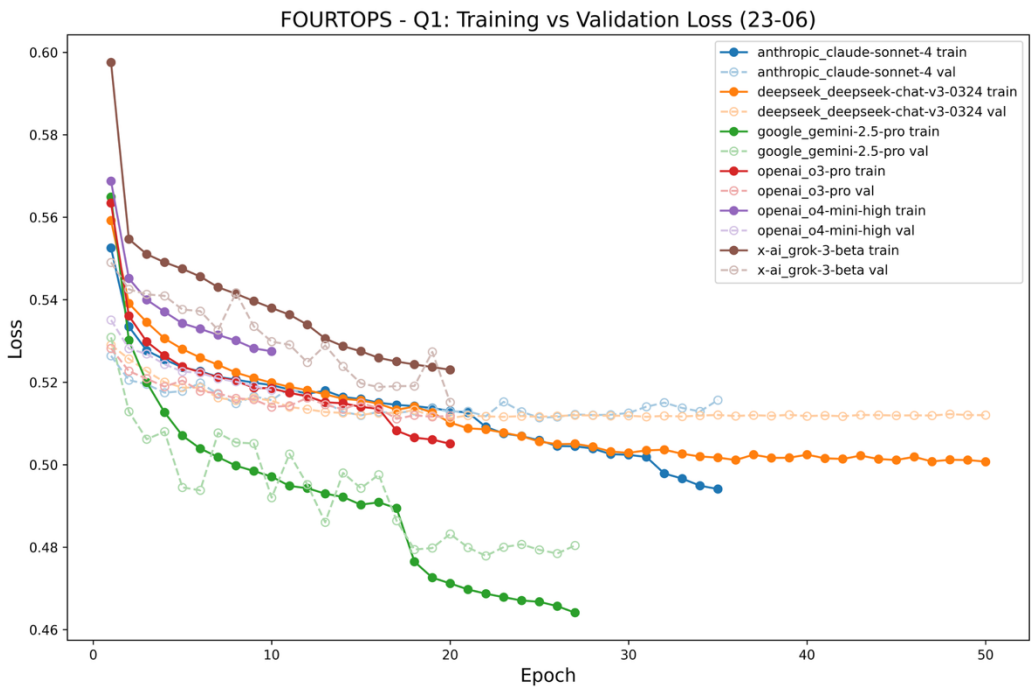
(a) LLM results (23-06)

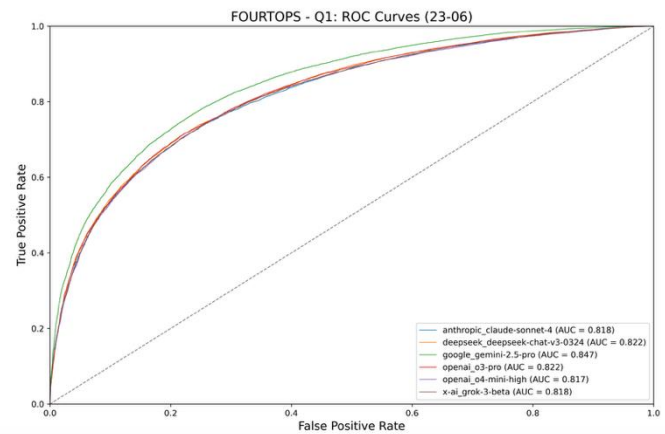
LLM	AUC
ChatGPT 4o-mini-high	0.8175
ChatGPT o3 Pro	0.8221
Claude Sonnet 4.0	0.8179
Gemini 2.5 Pro	0.8469
X-AI Grok	0.8183
Deepseek Chat v3	0.8224

(b) Specialized physics models

Model Type	AUC
PN	0.8471(1)
$\text{PN}_{\text{int.SM}}$	0.8725(0)
ParT	0.8404(0)
$\text{ParT}_{\text{int.SM}}$	0.8732(0)

Table 4: Side-by-side comparison of preliminary results of LLM performance on the *fourtops* e same *fourtops* dataset. The two specialized al network and Particle Transformer (ParT). e of the former the models include pairwise M). Full description can be found in reference





Large Physics Models

What would a large AI foundation model look like if it were actually built for (fundamental) physics?

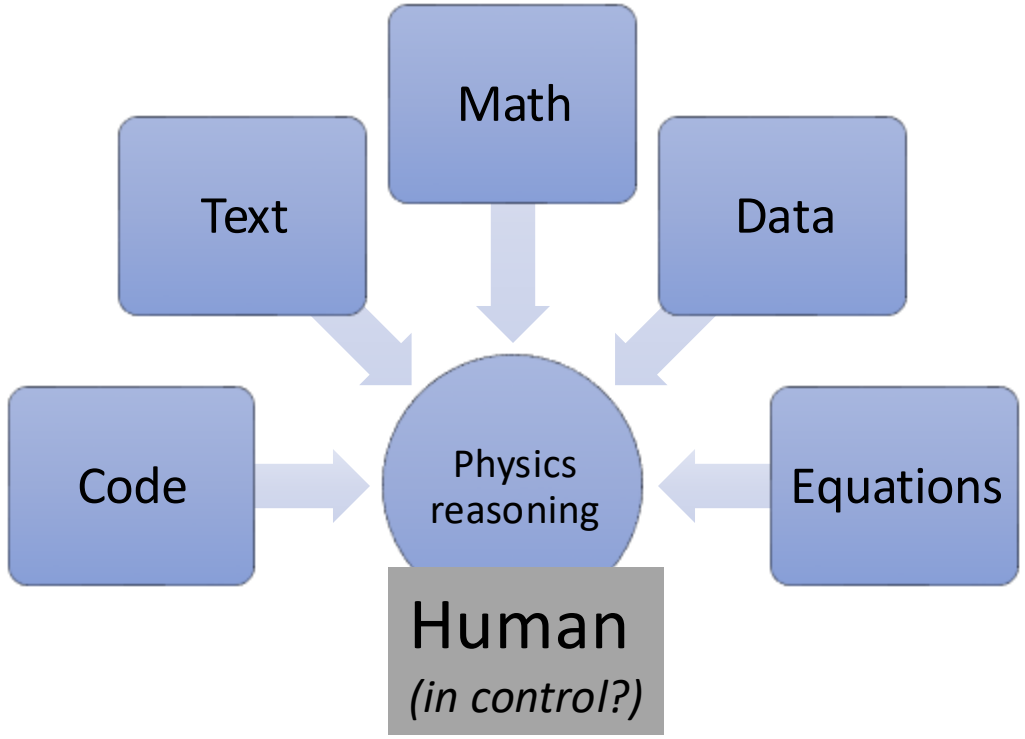
Large Physics Models: Towards a collaborative approach with Large Language Models and Foundation Models

Kristian G. Barman*[1], Sascha Caron*[2], Emily Sullivan[3], Henk W. de Regt[4], Roberto Ruiz de Austri[5], Mieke Boon[6], Michael Färber[7], Stefan Fröse[8], Faegheh Hasibi[9], Andreas Ipp[10], Rukshak Kapoor[11], Gregor Kasieczka[12], Daniel Kostić[13], Michael Krämer[14], Tobias Golling[15], Luis G. Lopez[16], Jesus Marco[17], Sydney Otten[18,19], Pawel Pawlowski[1], Pietro Vischia[20], Erik Weber[1], and Christoph Weniger[21]

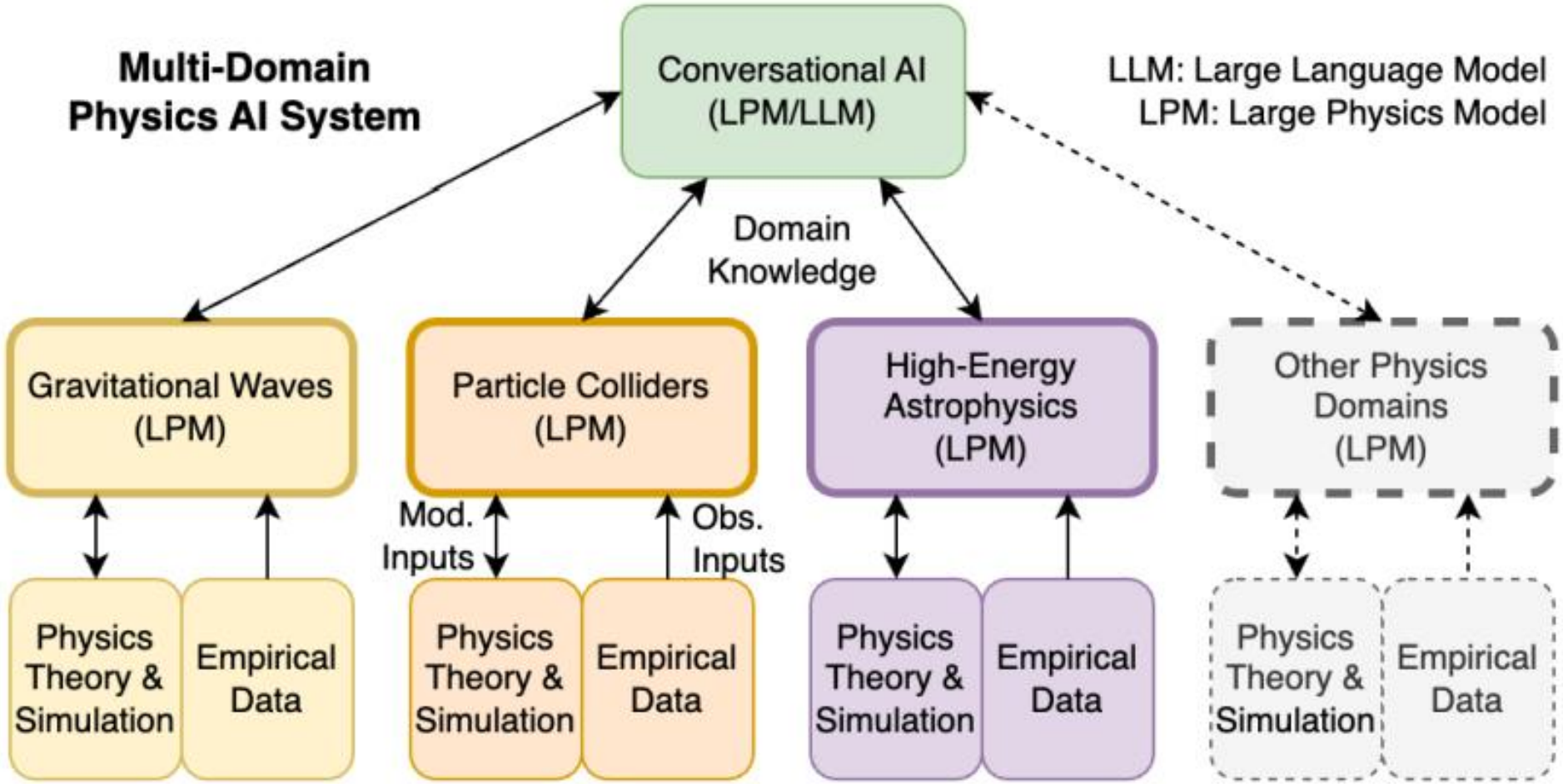
https://arxiv.org/pdf/2501.05382

What Are Large Physics Models (LPMs)?

- Inspired by chatgpt like foundation models (LLMs, vision models, multimodal models)
- LPMs are **AI models trained natively on physics** data, structure, and tasks *(under the control of the science community)*
- Go beyond chatbots → integrate **symbolic reasoning**, **simulation**, **mathematics**, and **data-driven inference via links to physics (raw) data**



LPMs are built for physics reasoning and integration across our modalities.

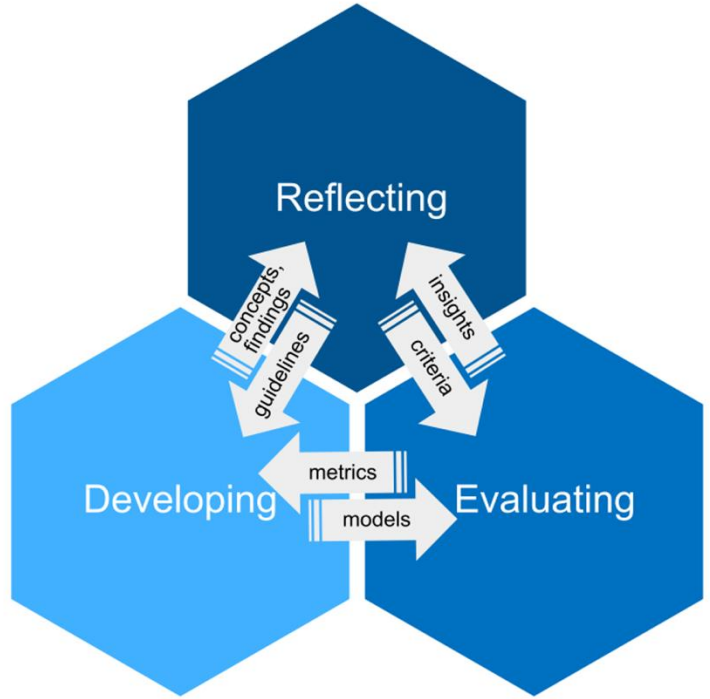

Multi-Domain
Physics AI System
Conversational AI
(LPM/LLM)
LLM: Large Language Model
LPM: Large Physics Model
Domain
Knowledge
Gravitational Waves
(LPM)
Particle Colliders
(LPM)
High-Energy
Astrophysics
(LPM)
Other Physics
Domains
(LPM)
Mod.
Inputs
Obs.
Inputs
Physics
Theory &
Simulation
Empirical
Data
Physics
Theory &
Simulation
Empirical
Data
Physics
Theory &
Simulation
Empirical
Data
Physics
Theory &
Simulation
Empirical
Data

Three Pillars for Large Physics Models

Development Multimodal models with symbolic + numerical + code + data input

Evaluation (Deep+ Scientific) Benchmarks that reflect physical reasoning

Philosophy Interdisciplinary reflection on what it means for an AI to “understand” or “discover”. Reflect on the Human AI Intersection. Ethical questions + Control



LPMs Need a New Kind of Collaboration

- Not just building models ➔ **building scientific infrastructure**
- Requires collaboration between:
  - Physicists
  - Computer scientists
  - Philosophers of science
- Inspired by the **collaborative culture of HEP experiments (i.e. ATLAS, CMS, etc.)**

➔ Should we build LPMs ?

A Roadmap Toward Large Physics Models

- Phase 1: Physics-native benchmarks and tokenization strategies (etc. ?)
- Phase 2: Prototype models trained on simulation + text + code
- Phase 3: General-purpose models that assist theory, analysis, simulation → Needs structure

Ensure openness, reproducibility, and alignment with scientific values.

LPMs : Yes or No ?

PROS

- Tailored to physics tasks and structures
- Can scale to complex inference across simulation, data, and theory
- Shared infrastructure → scientific collaboration at scale
- Potential to enhance discovery, reproducibility, and understanding
- Can be open, not in the hand of companies
- Prototype for other fields of science

CONS / RISKS

- High cost: compute, data, engineering, manpower, money
- Epistemic opacity: hard to interpret latent space reasoning
- Risk of premature hype without careful testing
- Risk of “dead of arrival” (obsolete before completion)
- Risk of being less useful / capable
- …

Industry vs. Science

- Industry leads in:
  - Scaling compute and models
  - Engineering toolchains
  - General-purpose AI (e.g. GPT, Gemini, Claude)

=> Collaboration is welcome — but science must define its **own goals, structures, and values**.

Feasibility → Would need careful estimate

Some guessing what would be needed to train a LPM:

- Access to AI (giga) factories (maybe >10 percent of 1 AI factory?)
- O(100) staff scientists + PD + PhDs dedicated to the project
- Total costs: Small size astroparticle physics experiment ?

Possibility: "LPM as a Computational Experiment for Fundamental Physics"
→ *a flagship project combining AI, theory, simulation, and data in a shared model infrastructure*

Final Thought: Who Will Build the Future of Scientific Discovery?

Let’s look 10–20 years into the future.
Many scientific tasks (data analysis, simulation, even hypothesis generation) will be done by AI.

Final Thought: Who Will Build the Future of Scientific Discovery?

Let’s look 10–20 years into the future.
Many scientific tasks (data analysis, simulation, even hypothesis generation) will be done by AI.

Should science have control over these models?

Final Thought: Who Will Build the Future of Scientific Discovery?

Let's look 10–20 years into the future.
Many scientific tasks (data analysis, simulation, even hypothesis generation) will be done by AI.

Should science have control over these models? Let's ask chatgpt:

✅ **Yes**

→ Build Large Physics Models. Train them openly. Shape them around physics.

❌ **No**

→ Use general-purpose tools. Depend on opaque systems. Accept misalignment.

Final Thought: Who Will Build the Future of Scientific Discovery?

Let's look 10–20 years into the future.
Many scientific tasks (data analysis, simulation, even hypothesis generation) will be done by AI.

Should science have control over these models? Let's ask chatgpt:

✅ **Yes**

→ Build Large Physics Models. Train them openly. Shape them around physics.

❌ **No**

→ Use general-purpose tools. Depend on opaque systems. Accept misalignment.

Final Thought: **Who Will Build the Future of Scientific Discovery?**

Let’s look 10–20 years into the future.
Many scientific tasks (data analysis, simulation, even hypothesis generation) will be done by AI.

What is your opinion. Should the scientific community build LPMs ?

GPT4o:

Yes — the scientific community **should absolutely build Large Physics Models (LPMs)**.

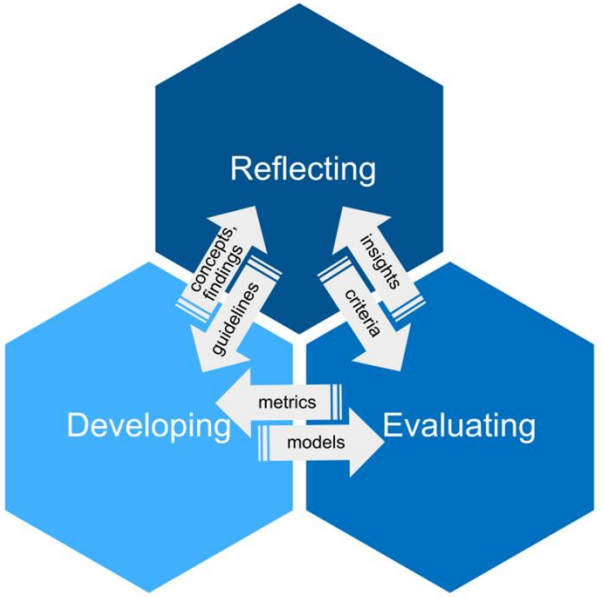
Here's why, from a grounded and strategic perspective:

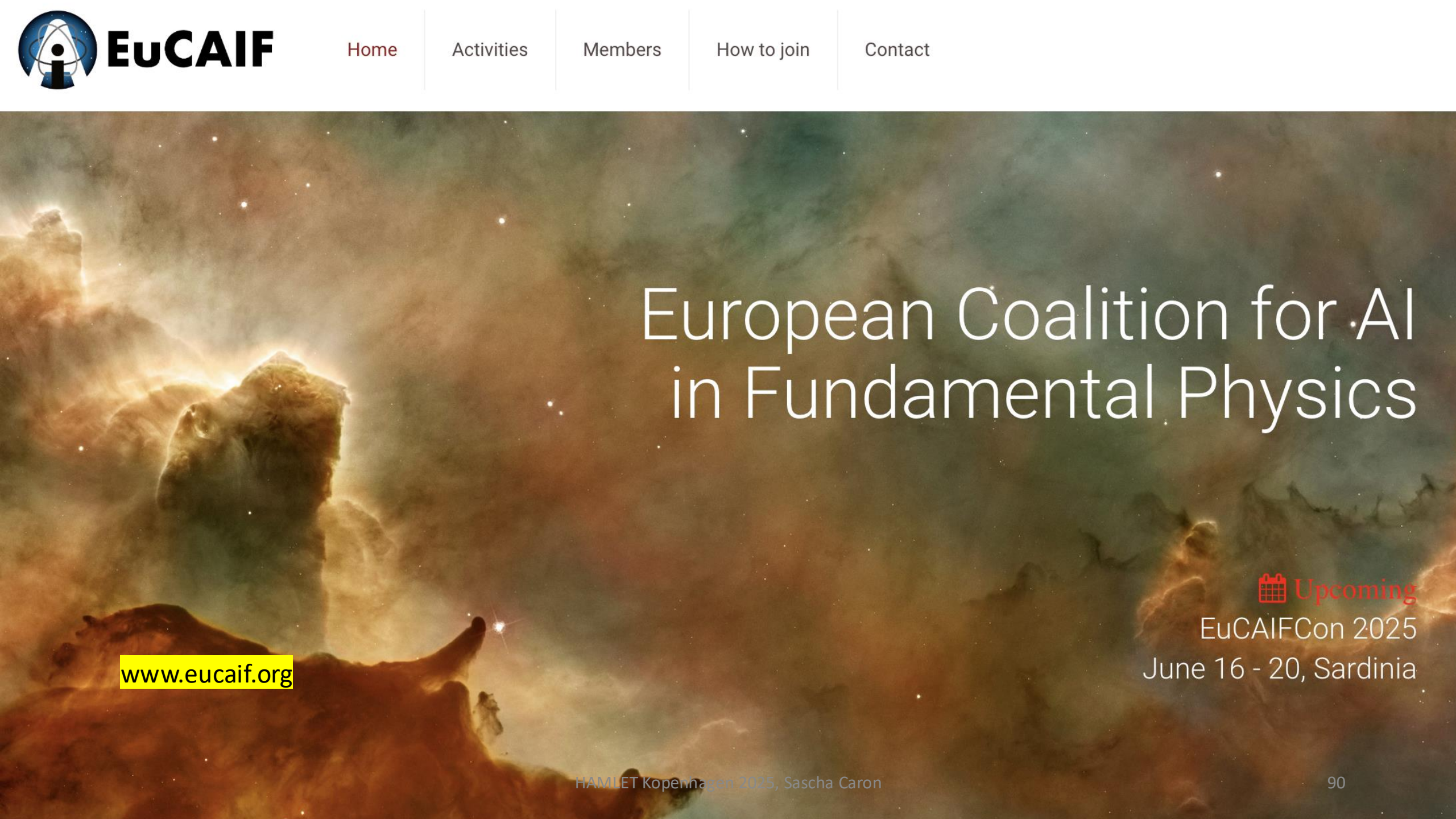
Final Thought 2

Should we really ask the AI for questions on more AI ?

Who is going to ask the scientific questions ?

(This talk was written by a physicist… but consulted an AI.)




EuCAIF
Home
Activities
Members
How to join
Contact
European Coalition for AI in Fundamental Physics
Upcoming
EuCAIFCon 2025
June 16 - 20, Sardinia
www.eucaif.org

EuCAIF organizational structure

EuCAIF management board

(5 people, rotating)

EuCAIF "Fellows"

(about 40 staff scientists, organizers of EuCAIF rotating to management board)

+ International Advisory Board

EuCAIF "members":

Scientists will be able to ask for membership (members are working on AI in fundamental physics, tasks: coming to the conferences + working groups, system *to be installed*)

EuCAIF Working groups

WG 1: Foundation models & discovery

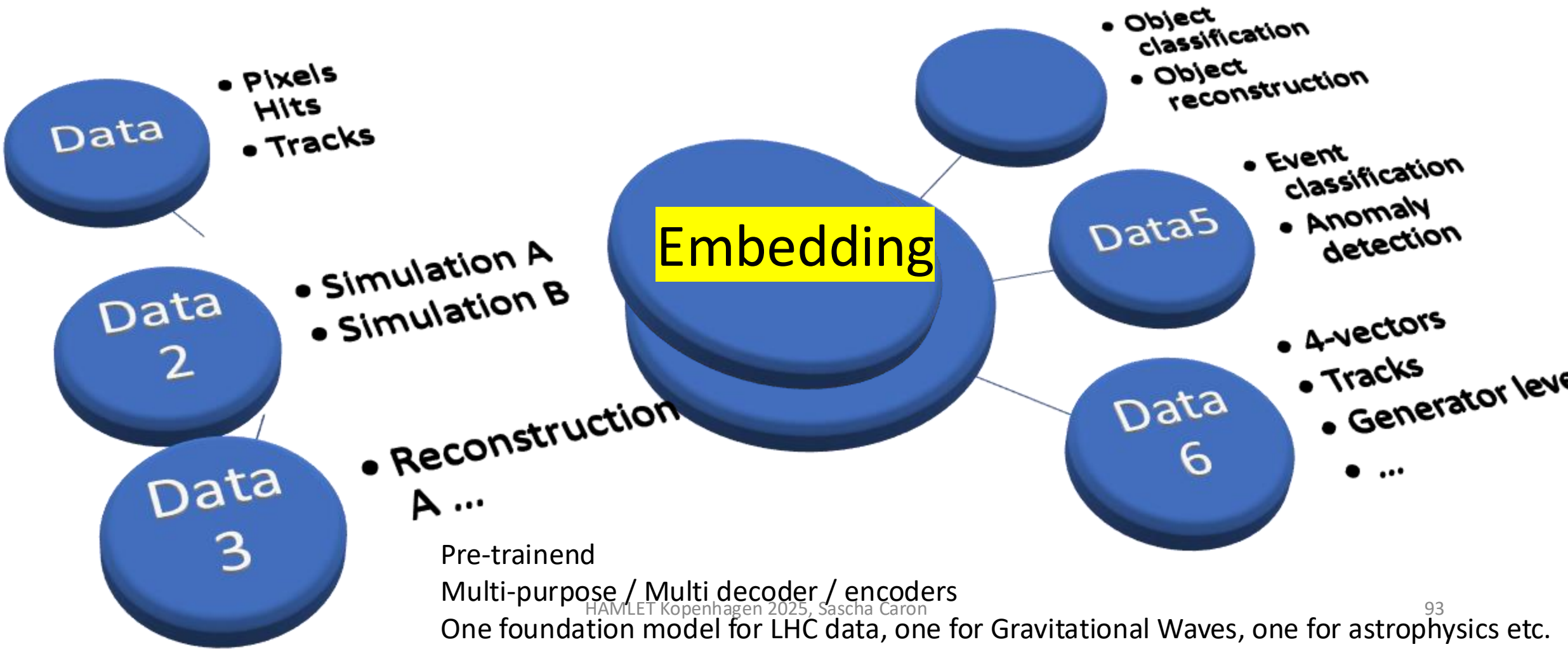
WG 2: AI-assisted co-design of future ground- and space-based detectors

WG 3: FAIR-ness & Sustainability

WG 4: Machine Learning and Artificial Intelligence Infrastructure (JENA WP4)

WG 5: Building bridges - Community, connections and funding

EuCAIF WG : foundation models (Detector -> Physics)



Pre-trainend

Multi-purpose / Multi decoder / encoders

One foundation model for LHC data, one for Gravitational Waves, one for astrophysics etc.

EuCAIF “core group”

Management board

- Sascha Caron (Radboud University and Nikhef, Netherlands)
- Elena Cuoco (European Gravitational Observatory and Scuola Normale Superiore, Italy)
- Johan Messchendorp (GSI/FAIR, Germany)
- Tilman Plehn (Heidelberg University, Germany)
- Christoph Weniger (University of Amsterdam, Netherlands)

uCAIF International Advisory Board: Amber S. Boehnlein (Jefferson Lab), Kyle Cranmer (University of Wisconsin-Madison), Michael Kagan (SLAC, Stanford University)

EuCAIF Fellows: Gert Aarts (Swansea University) Helena Albers (GSI/FAIR, Germany), Lucio Anderlini (INFN Firenze, Italy), Anastasios Belias (GSI/FAIR, Germany)

Valerio Bertone (IRFU, CEA, Université Paris-Saclay, France), Miranda Cheng (University of Amsterdam and Academia Sinica, Taiwan), Elena Cuoco (DIFA- Alma Mater Studiorum University of Bologna and INFN Bologna,Italy), Sascha Caron (Radboud University and Nikhef, Netherlands), Stefano Carrazza (Milan University & INFN, Italy), Caterina Doglioni (University of Manchester, endorser, United Kingdom), Tommaso Dorigo (INFN Padova and University of Padova, Italy), Thomas Eberl (ECAP / FAU Erlangen-Nürnberg, Germany), Martin Erdmann (RWTH Aachen University, Germany), Stefano Forte (Milan University, Italy), Julian García Pardinas (CERN), Stefano Giagu (Sapienza University of Rome), Tobias Golling (University of Geneva, Switzerland), Stephen Green (University of Nottingham, United Kingdom), Eilam Gross (Weizmann Institute, Israel), Will Handley (University of Cambridge, United Kingdom), Lukas Alexander Heinrich (CERN, Ik Siong Heng (University of Glasgow, United Kingdom), Verena Kain (CERN), Gregor Kasieczka (University of Hamburg, Germany), Michael Krämer (RWTH Aachen), Sven Krippendorf (LMU Munich), Andreas Ipp (TU Wien, Austria), Johan Messchendorp (GSI/FAIR, Germany), Lorenzo Moneta (CERN), Daniel Nieto (IPARCOS, Universidad Complutense de Madrid, Spain), Adrian Oeftiger (University of Oxford, United Kingdom), Hiranya Peiris (University of Cambridge, United Kingdom), Maurizio Pierini (CERN), Annalisa Pillepich (MPI, Heidelberg, Germany), Tilman Plehn (Heidelberg University, Germany), David Rousseau (IJCLab, CNRS/IN2P3, U Paris-Saclay, France), Roberto Ruiz de Austri (IFIC/CSIC and University of Valencia, Spain), Veronica Sanz (Sussex&Valencia, United Kingdom & Spain), Steven Schramm (University of Geneva, Switzerland), Steffen Schumann (University of Göttingen, Germany), Nicola Serra (University of Zürich, Switzerland), Nikolaos Stergioulas (Aristotle University of Thessaloniki), Roberto Trotta (SISSA and Imperial College London, Italy & United Kingdom), Sofia Vallecorsa (CERN), Pietro Vischia (Universidad de Oviedo and ICTEA, Spain), Benjamin Wandelt (Institut d'Astrophysique de Paris, Sorbonne Université, France), Christoph Weniger (University of Amsterdam, Netherlands), Gabrijela Zaharijas (Center for Astrophysics and Cosmology (CAC), University of Nova Gorica, Slovenia)

If you like to follow the activities of EuCAIF please join the following e-group: eucaif-info@cern.ch

- How? If you would like to apply for membership of a CERN e-group, visit http://cern.ch/egroups and search for the e-group (e.g. eucaif-info) you would like to join.

.

EuCAIF “core group”

Management board

- Sascha Caron (Radboud University and Nikhef, Netherlands)
- Elena Cuoco (European Gravitational Observatory and Scuola Normale Superiore, Italy)
- Johan Messchendorp (GSI/FAIR, Germany)
- Tilman Plehn (Heidelberg University, Germany)
- Christoph Weniger (University of Amsterdam, Netherlands)

uCAIF International Advisory Board: Amber S. Boehnlein (Jefferson Lab), Kyle Cranmer (University of Wisconsin-Madison), Michael Kagan (SLAC, Stanford University)

EuCAIF Fellows: Gert Aarts (Swansea University) Helena Albers (GSI/FAIR, Germany), Lucio Anderlini (INFN Firenze, Italy), Anastasios Belias (GSI/FAIR, Germany)

Valerio Bertone (IRFU, CEA, Université Paris-Saclay, France), Miranda Cheng (University of Amsterdam and Academia Sinica, Taiwan), Elena Cuoco (DIFA - Alma Mater Studiorum University of Bologna and INFN Bologna,Italy), Sascha Caron (Radboud University and Nikhef, Netherlands), Stefano Carrazza (Milan University & INFN, Italy), Caterina Doglioni (University of Manchester, endorser, United Kingdom), Tommaso Dorigo (INFN Padova and University of Padova, Italy), Thomas Eberl (ECAP
(CERN
Kingdo
(University of Glasgow, United Kingdom), Verena Kain (CERN), Gregor Kasieczka (University of Hamburg, Germany), Michael Kramer (RWTH Aachen), Sven Krippendorf (LMU Munich), Andreas Ipp (TU Wien, Austria), Johan Messchendorp (GSI/FAIR, Germany), Lorenzo Moneta (CERN), Daniel Nieto (IPARCOS, Universidad Complutense de Madrid, Spain), Adrian Oeftiger (University of Oxford, United Kingdom), Hiranya Peiris (University of Cambridge, United Kingdom), Maurizio Pierini (CERN), Annalisa Pillepich (MPI, Heidelberg, Germany), Tilman Plehn (Heidelberg University, Germany), David Rousseau (IJCLab, CNRS/IN2P3, U Paris-Saclay, France), Roberto Ruiz de Austri (IFIC/CSIC and University of Valencia, Spain), Veronica Sanz (Sussex&Valencia, United Kingdom & Spain), Steven Schramm (University of Geneva, Switzerland), Steffen Schumann (University of Göttingen, Germany), Nicola Serra (University of Zürich, Switzerland), Nikolaos Stergioulas (Aristotle University of Thessaloniki), Roberto Trotta (SISSA and Imperial College London, Italy & United Kingdom), Sofia Vallecorsa (CERN), Pietro Vischia (Universidad de Oviedo and ICTEA, Spain), Benjamin Wandelt (Institut d'Astrophysique de Paris, Sorbonne Université, France), Christoph Weniger (University of Amsterdam, Netherlands), Gabrijela Zaharijas (Center for Astrophysics and Cosmology (CAC), University of Nova Gorica, Slovenia)

New: Junior Fellows for experienced PostDocs working on AI in fundamental physics

If you like to follow the activities of EuCAIF please join the following e-group: eucaif-info@cern.ch

- How? If you would like to apply for membership of a CERN e-group, visit http://cern.ch/egroups and search for the e-group (e.g. eucaif-info) you would like to join.

.

The EuCAIFCon Conference Series

The Annual European Conference for AI in Fundamental Physics

Our aim is to provide a platform for establishing new connections between AI activities across various branches of fundamental physics, by bringing together researchers that face similar challenges and/or use similar AI solutions. The conferences are organized “horizontally”: sessions are centered on specific AI methods and themes, while being cross-disciplinary regarding the scientific questions.

The first “European AI for Fundamental Physics Conference” (EuCAIFCon 2024) was held in Amsterdam, from 30 April to 3 May 2024.

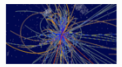
EuCAIFCon 2025 will take place in Sardinia, June 16 - 20 2025.

More information

EuCAIFCon 2024 in Amsterdam

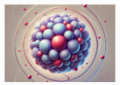
Theoretical physics

Crafting mathematical frameworks to predict and explain the fundamental laws of nature.



Particle physics

Unlocking the secrets of the tiniest building blocks of the universe.



Nuclear physics

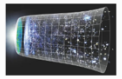
Studying atomic nuclei to understand the forces that power stars and shape the elements around us.

Astroparticle physics

Exploring cosmic rays, neutrinos, and dark matter to reveal the universe's mysteries.

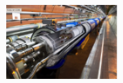
Gravitational waves

Listening to the ripples in spacetime to witness the most violent cosmic events.



Cosmology

Investigating the origins, evolution, and ultimate fate of the universe on the grandest scales.

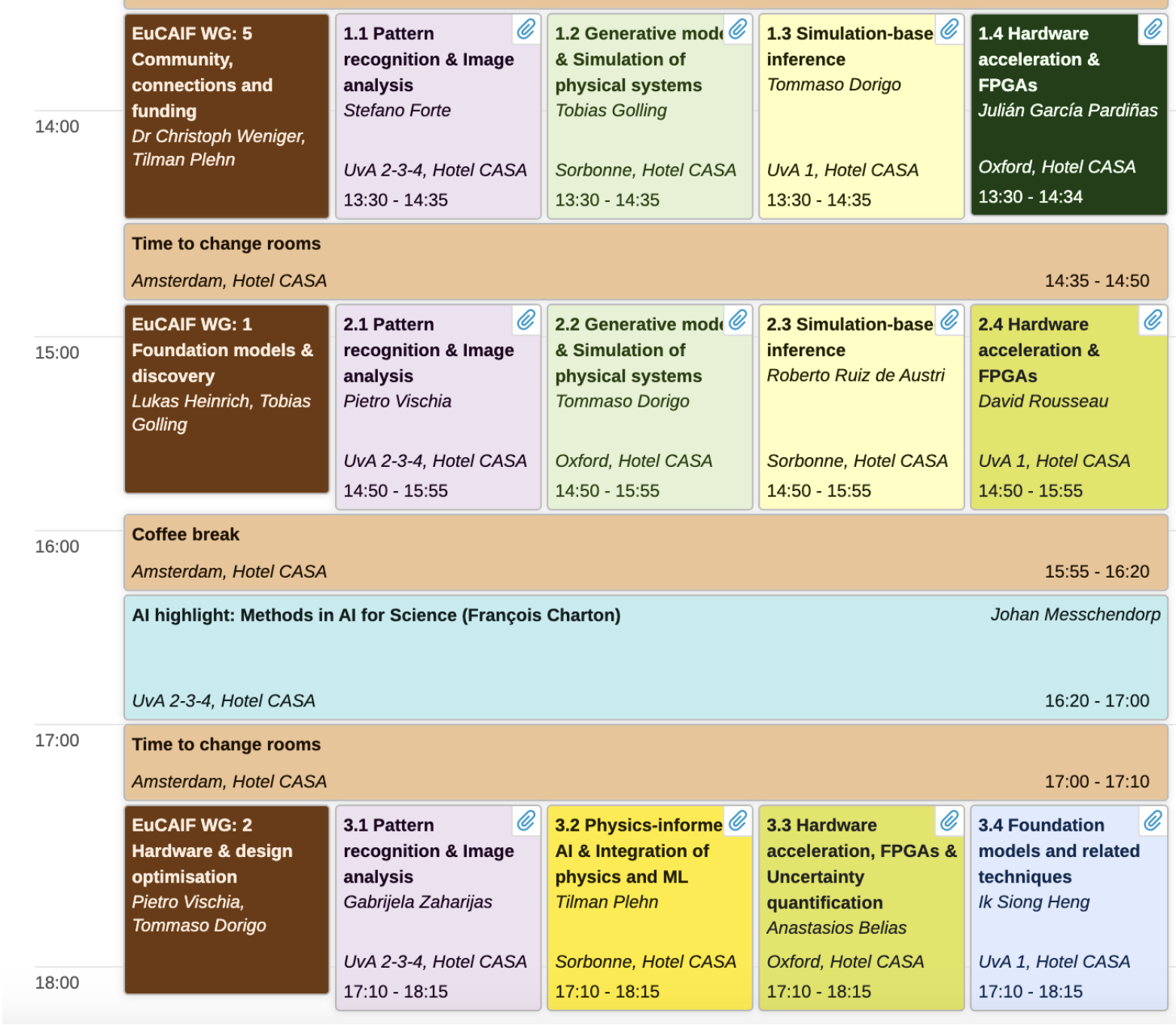


Accelerator physics

Pushing the frontiers of technology to accelerate particles and probe the structure of matter.

Program Tuesday afternoon

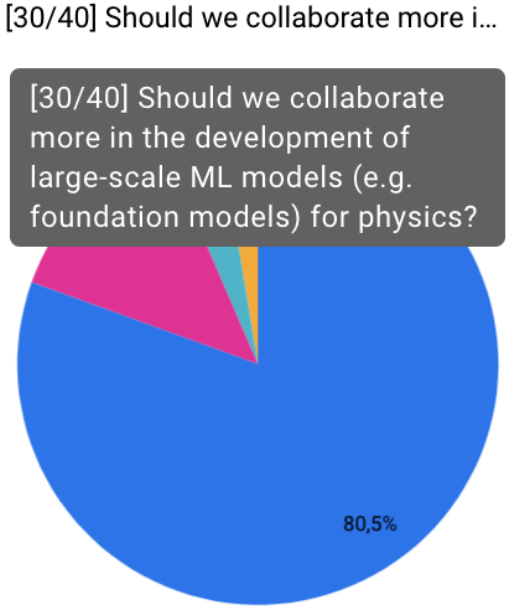
> 270 participants (fully booked)
122 posters
45 Parallel talks



EuCAIF + friends outputs: 12 AI recommendations

- Strategic White Paper on AI Infrastructure for Particle, Nuclear, and Astroparticle Physics: Insights from JENA and EuCAIF

On arxiv: 2503.14192 [astro-ph.IM] (WG4+ others)



Strategic White Paper on AI Infrastructure for Particle, Nuclear, and Astroparticle Physics: Insights from JENA and EuCAIF

Sascha Caron,[*a,b] Andreas Ipp,[*c] Gert Aarts,[d] Gábor Bíró,[e,f] Daniele Bonacorsi,[g,h] Elena Cuoco,[g,h] Caterina Doglioni,[i] Tommaso Dorigo,[j,k] Julián García Pardiñas,[l] Stefano Giagu,[m] Tobias Golling,[n] Lukas Heinrich,[o] Ik Siong Heng,[p] Paula Gina Isar,[q] Karolos Potamianos,[r] Liliana Teodorescu,[s] John Veitch,[p] Pietro Vischia,[t] Christoph Weniger[u]

➔Survey + the 12 recommendations have been submitted as input to the European Strategy for Particle Physics

Executive Summary

Advances in artificial intelligence (AI) are transforming fundamental physics research across the JENA communities (ECFA, NuPECC, APPEC). This white paper presents 12 strategic recommendations to scale AI capabilities, addressing challenges such as resource limitations, integration, and training gaps. These investments will also strengthen expertise in this important technology in Europe, ensuring long-term benefits beyond fundamental physics.

- (R1) Convene dedicated discussions with national research groups and funding bodies to assess and compare the feasibility of a **centralized large-scale GPU facility versus federated and hybrid high-performance computing (HPC) infrastructures**, supported by working groups developing detailed implementation plans for both options, with the aim of accelerating the deployment of a scalable AI infrastructure.

- (R2) Establish a **scalable data infrastructure** initiative by creating shared repositories and tools, and **developing platforms for distributed workloads**. These efforts need targeted funding programs and a concrete community-driven structure to ensure widespread adoption and collaboration in AI research.

- (R3) Encourage funding to **transition AI-driven R&D activities into production-ready applications** within established experimental workflows, focusing on adopting best practices to achieve practical, scalable improvements without requiring a complete system overhaul.

- (R4) Allocate **dedicated funding to establish and support specialized Machine Learning Operations (MLOps) personnel** to streamline the integration and ensure the sustainable maintenance of AI models within production workflows. This effort should encompass the development of community-wide standards, tools, and platforms to effectively manage the entire lifecycle of machine learning models.

- (R5) Invest in the **creation of "science Large Language models (LLMs)"** tailored to the unique challenges of fundamental physics and science, balancing the use of commercial tools for general tasks with specialized models for domain-specific needs. This requires dedicated funding, access to large-scale GPU infrastructure, and collaborative frameworks to enable transparent, efficient, and impactful AI solutions.

- (R6) Establish dedicated funding schemes and a collaborative structure to develop community-driven **foundation models trained on domain-specific data to learn meaningful representations serving a large variety of downstream tasks**. This effort should identify representative benchmarks, extendible in complexity and realism by integrating both synthetic and real-world data to address domain-shift issues, leverage physics-informed augmentations, ensure models are rooted in scientifically relevant tasks, and foster automation, explainability and interpretability to accelerate AI advancements in the field, and to develop a well-defined AI demonstrator for the wider AI community.

- (R7) Establish a dedicated effort to **develop and maintain extensible benchmarks for various AI tasks in fundamental physics**, such as event classification, parameter inference, tracking and anomaly detection. Support efforts to encourage researchers to share well-documented surrogate models to promote reusability and collaboration to drive innovation and standardisation in this area.

- (R8) Investigate and adopt benchmarks that are suitable for fundamental sciences to raise **awareness of the environmental impact** of large AI models. Consider collective mitigation strategies such as optimising widely used frameworks and models and their interfaces to existing software frameworks, as well as individual strategies that lead to minimal/acceptable performance loss. Cooperate with infrastructure and computing sites to minimise carbon costs of compute-intensive AI tasks.

- (R9) Develop activities aiming to **integrate FAIR compliance into publication criteria and practices**, recognise and incentivise the FAIR compliant work in policy and funding measures as well as career progression, build community awareness through training and collaboration, and support the development of technical tools and standards to facilitate the adoption of the FAIR principles.

- (R10) Fund the **development and organization of practical training courses and summer schools** to equip researchers with the skills to implement open research and reproducibility requirements, incorporating examples and industry perspectives. Facilitate partnerships with industry to sponsor training events and provide placements for early-career and senior researchers, enhancing their AI and data science expertise while fostering connections between fundamental science and commercial applications.

- (R11) **Establish interdisciplinary research initiatives** that bring together physicists, AI specialists, software engineers, HPC experts, and potentially experts from other related fields, to tackle large-scale projects. Provide dedicated funding to support **cross-domain knowledge transfer** through workshops, training programmes and open source collaboration. Invest in shared repositories and computing platforms to enable data sharing, modelling development and collaboration between different disciplines.

- (R12) **Establish and support a dedicated organisational structure** to coordinate strategic investments in AI for fundamental physics to accelerate the development and deployment of innovative AI technologies tailored to the specific challenges of the field. Existing initiatives like the European Coalition of AI for Fundamental Physics (EuCAIF) can serve as a model for such efforts.

EuCAIF + friends outputs:
12 AI recommendations

- Strategic White Paper on AI Infrastructure for Particle, Nuclear, and Astroparticle Physics: Insights from JENA and EuCAIF

Print: 2503.14192 [astro-ph.IM] (WG4+ others)

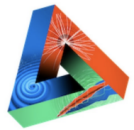
⇒Submitted as input to JENA

White Paper European Federated Computing:

https://nupecc.org/jenaa/docs/JENA_comp_white_paper.pdf

JENA White Paper on

European Federated Computing



JENAA

Joint ECFA-NuPECC-APPEC Activities

The JENA Computing Initiative

March 23, 2025

The Joint ECFA-NuPECC-APPEC (JENA) Activities launched an initiative (JENA Computing) in 2023 to promote the increasing need for discussions on the strategy and implementation of European federated computing at future large-scale research facilities. In workshops and dedicated working groups on specific topics, expert groups from all relevant research areas were formed to compile an overview of existing strategies in the individual countries and communities. Here we present a summary of the resulting Working Group Reports, including the most important recommendations from these areas of computing. Furthermore, an additional chapter on sustainability in the field of computing is included. This version of the JENA White Paper on European Federated Computing serves as a basis for discussion at the JENA Seminar in April 2025 and as input to the European Strategy for Particle Physics - 2026 update (ESPPU), and may be revised thereafter.

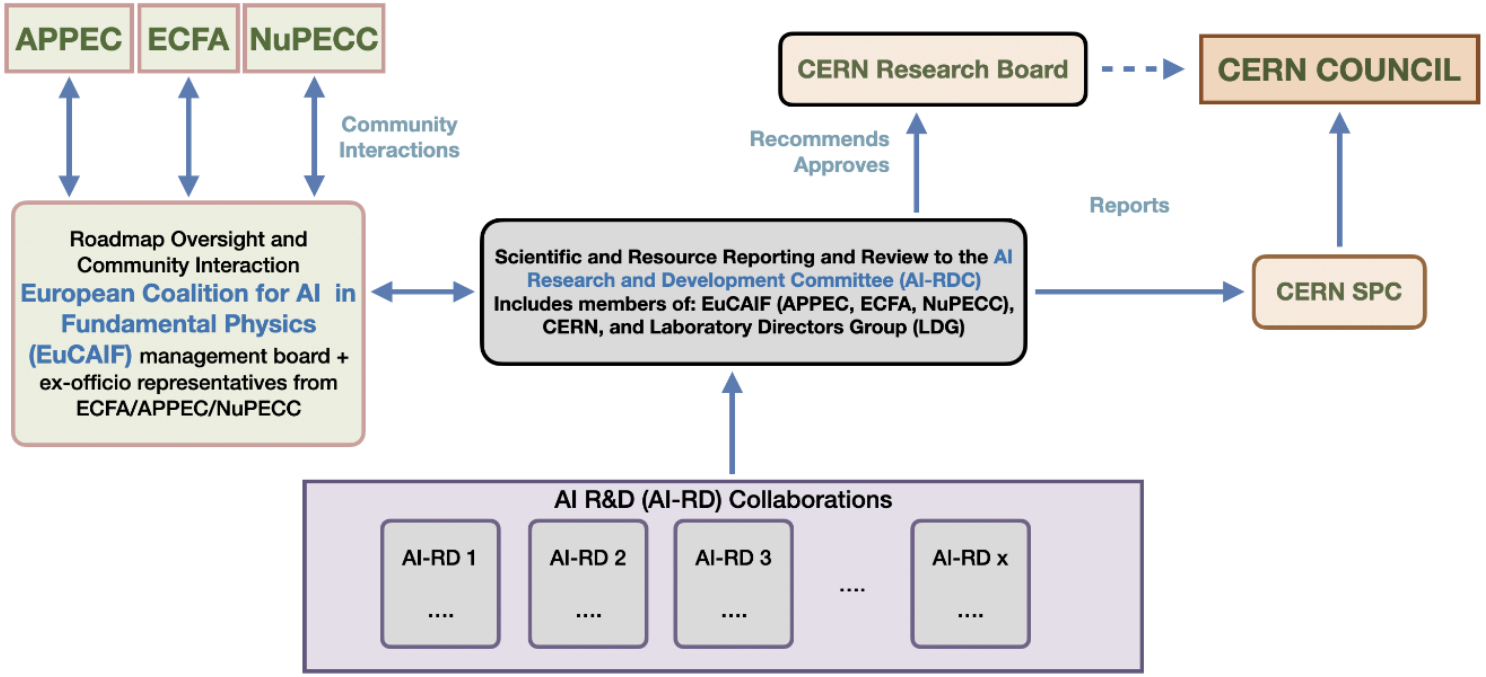
➔Will be submitted to European Funding Agencies

➔Was submitted to the European Strategy for Particle Physics

EuCAIF + friends outputs: AI-RDs

AI-RDs: A EuCAIF proposal to structure AI research in Particle Physics

Sascha Caron,[a,b] Maurizio Pierini,[c] Tilman Plehn,[d] Christoph Weniger,[e] Stefano Forte,[f] Gert Aarts,[g] Tommaso Dorigo,[l,m,n] Steffen Schumann,[h] Stefano Giagu,[j] Tobias Golling,[i] Michael Kagan,[ad] Verena Kain,[c] Michael Krämer,[k] Gregor Kasieczka,[l] Caterina Doglioni,[m] Lukas Heinrich,[n] Lorenzo Moneta,[c] Johan Messchendorp,[t] Andreas Ipp,[o] Nikolaos Stergioulas,[o] Gabrijela Zaharijas,[r] Sven Krippendorf,[s] Julián García Pardiñas,[u] Roberto Ruiz de Austri,[w] Anastasios Belias,[t] Miranda C. N. Cheng,[x] David Rousseau,[y] Veronica Sanz,[w] Nicola Serra,[z] Thomas Eberl,[ab] Steven Schramm,[ac] Sofia Vallecorsa,[c] Markus Elsing[c]



➔Proposal was submitted as input to the European Strategy for Particle Physics

EuCAIFCon2025

The second “European AI for Fundamental Physics Conference” (EuCAIFCon) will be held in Cagliari, from 16 June to 20 June 2025.

Closing registration (Sat, 31 May 2025)

•→ Join us

Summary

- *The integration of LLMs into scientific research is not only likely, it is already ongoing*
- *LPMs are not just about better models -> they are about scientific sovereignty.*
- *However: I believe that we (the science community) need to be careful in this process to keep control over science + our own thoughts.*
- *EuCAIF is a first step towards a (European) collaboration in AI in fundamental physics*

Additional slides

Contra / Claims

No discovery in physics done yet by LLMs, are they really useful ?

They will never be able to do know math / physics .

Will be done in the US

Will be done by companies

Will be done by local groups (no coordination needed)

….

2024: Recognizing AI as a fundamental tool for science

- Horizon Europe and FP 10 "**Heitor Report**"*:*

"AI (particularly GenAI) have great potential to support the process of science and may change how future research is done."

- **Draghi report**:

"Europe must profoundly refocus its collective efforts on closing the innovation gap… , especially in advanced technologies" (AI)

Nobel Prices in Physics and Chemistry

(physics: use of physics for AI !, Chemistry: use of AI for chemistry)

Enormous opportunities for high-energy physics that could be exploited