



ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

AI all'INFN

Daniele Bonacorsi (University of Bologna / INFN)



CCR workshop 2025 - La Biodola, Elba





ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

Qualche considerazione su **AI (ML/DL)** in HEP e all'INFN

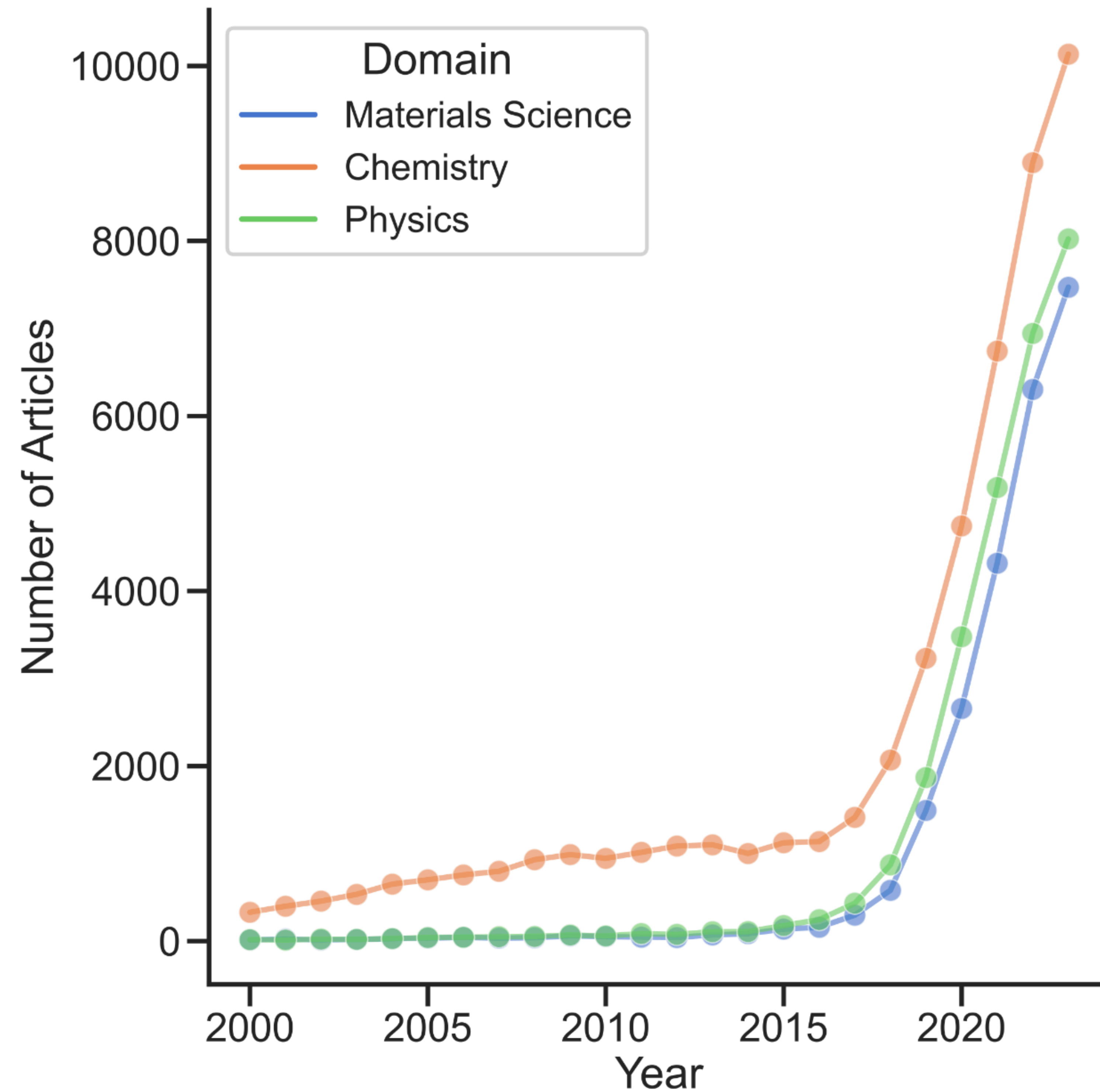
Daniele Bonacorsi (University of Bologna / INFN)



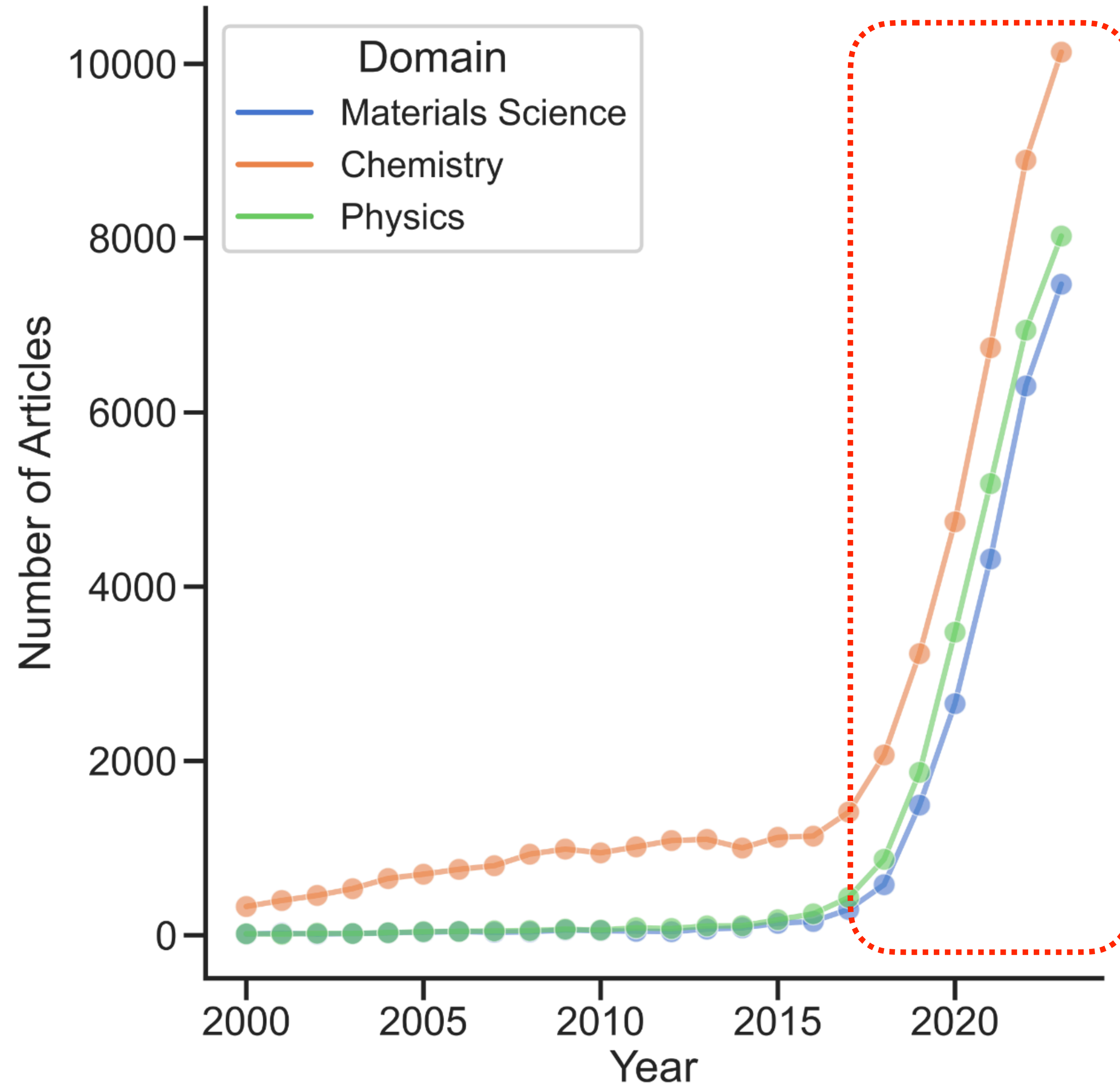
CCR workshop 2025 - La Biodola, Elba



ML-based publications in science



ML-based publications in science



We know **what** this is.

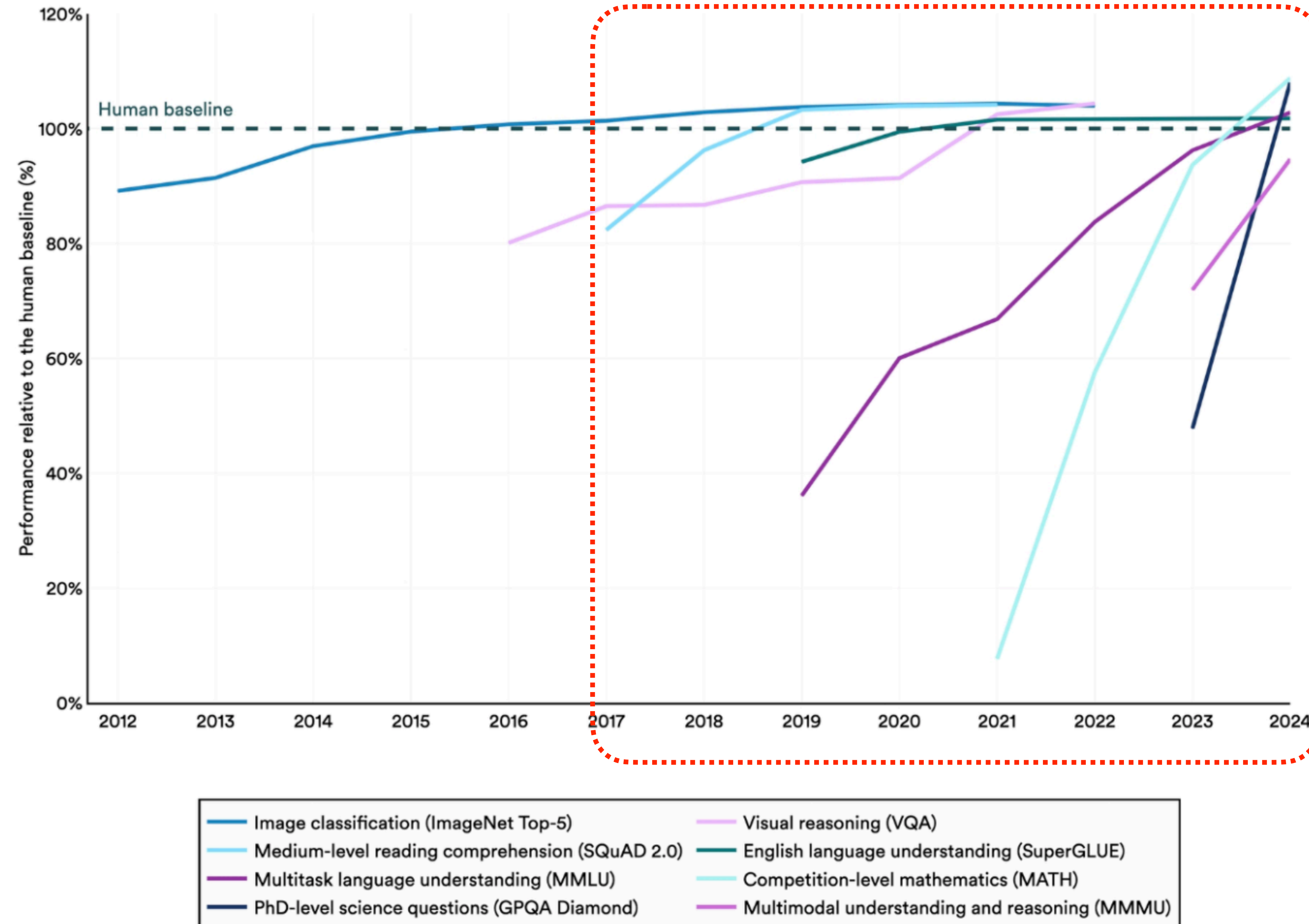
We know **why** it happened
at this point in time.

What we don't know:
Will it last? How will this plot
look in 10 years from now?

ML/DL vs humans

Select AI Index technical performance benchmarks vs. human performance

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



Same red box as previous slide..

Additional questions:

"Which ML" will we be talking about in 10 yrs from now?

(btw: are science and industry going towards the same directions?)

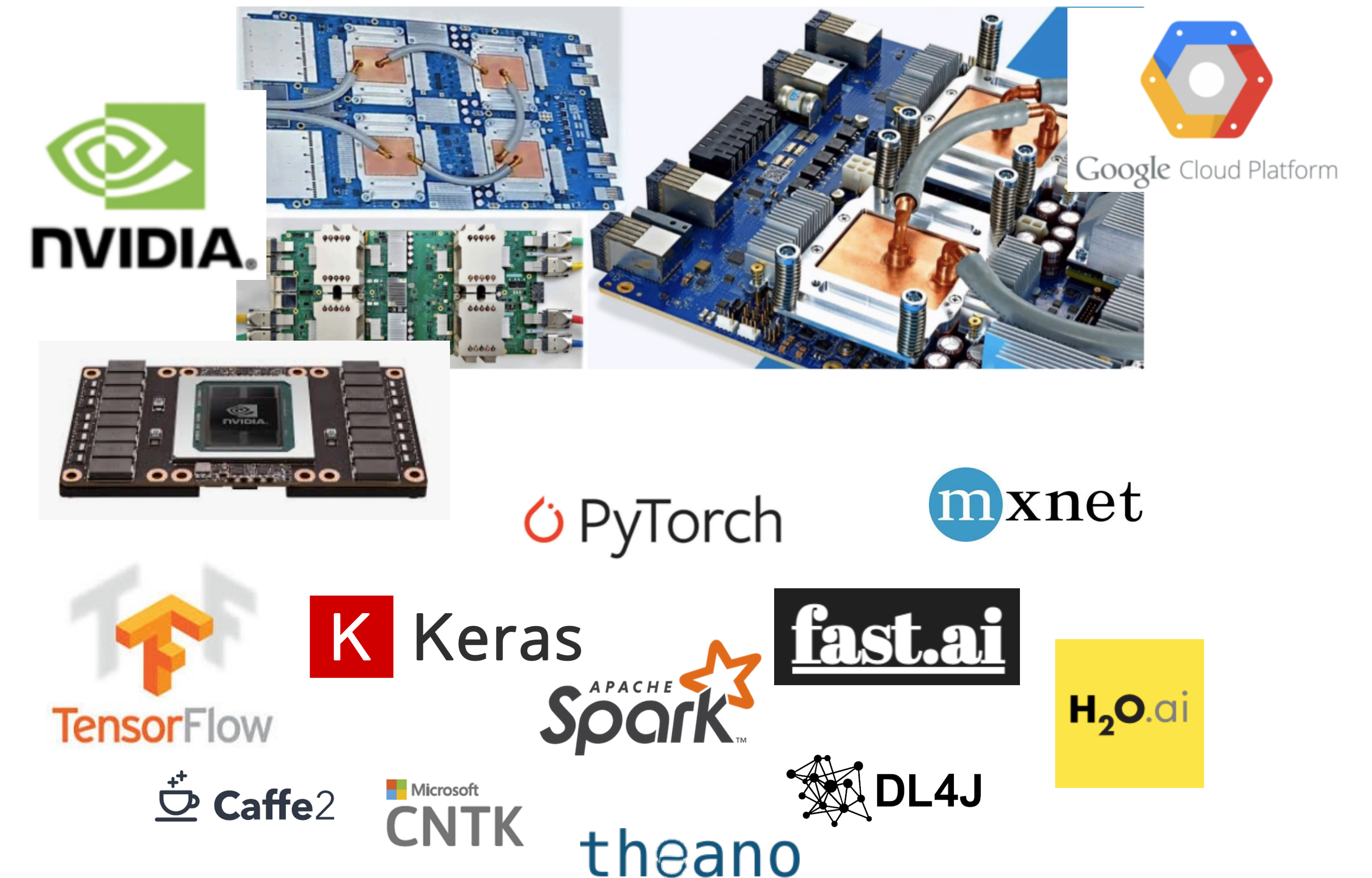
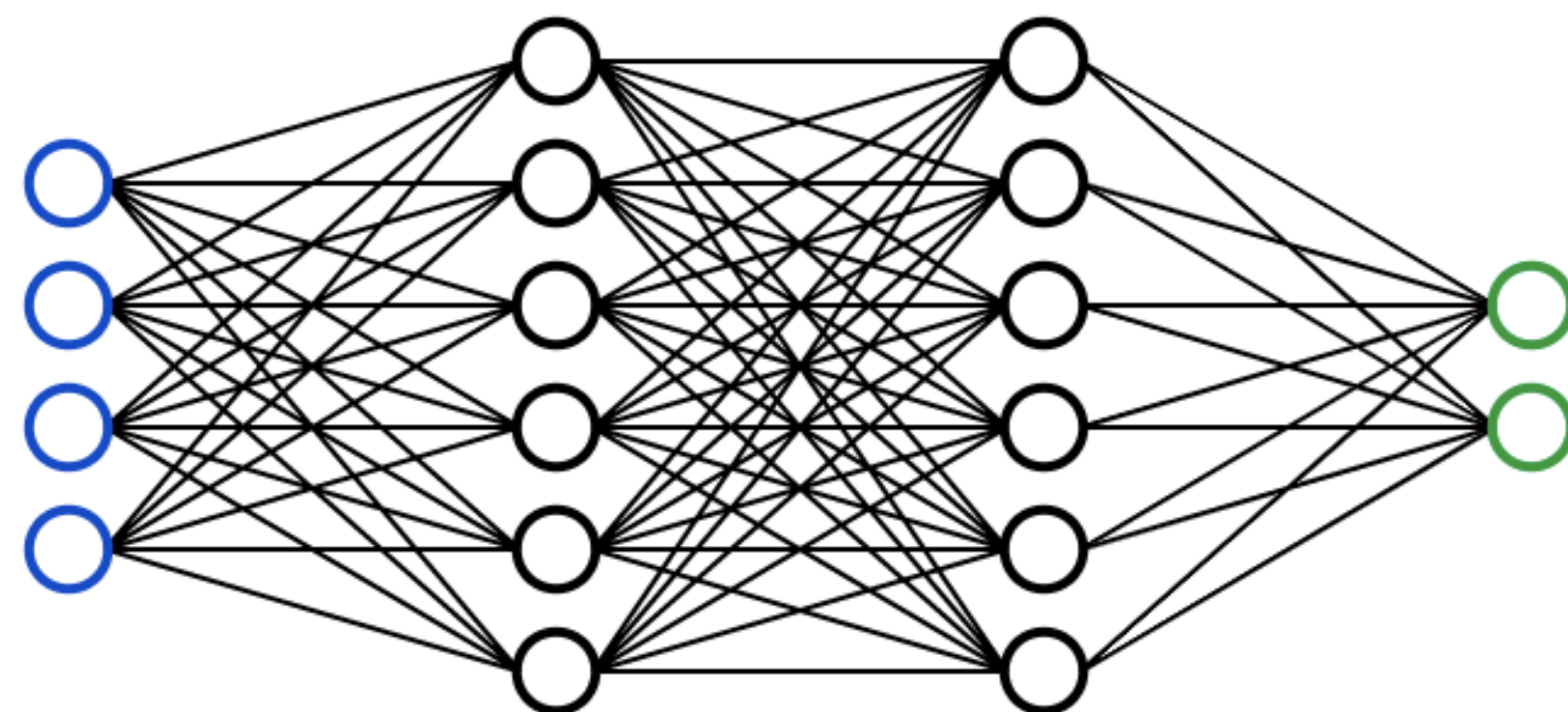
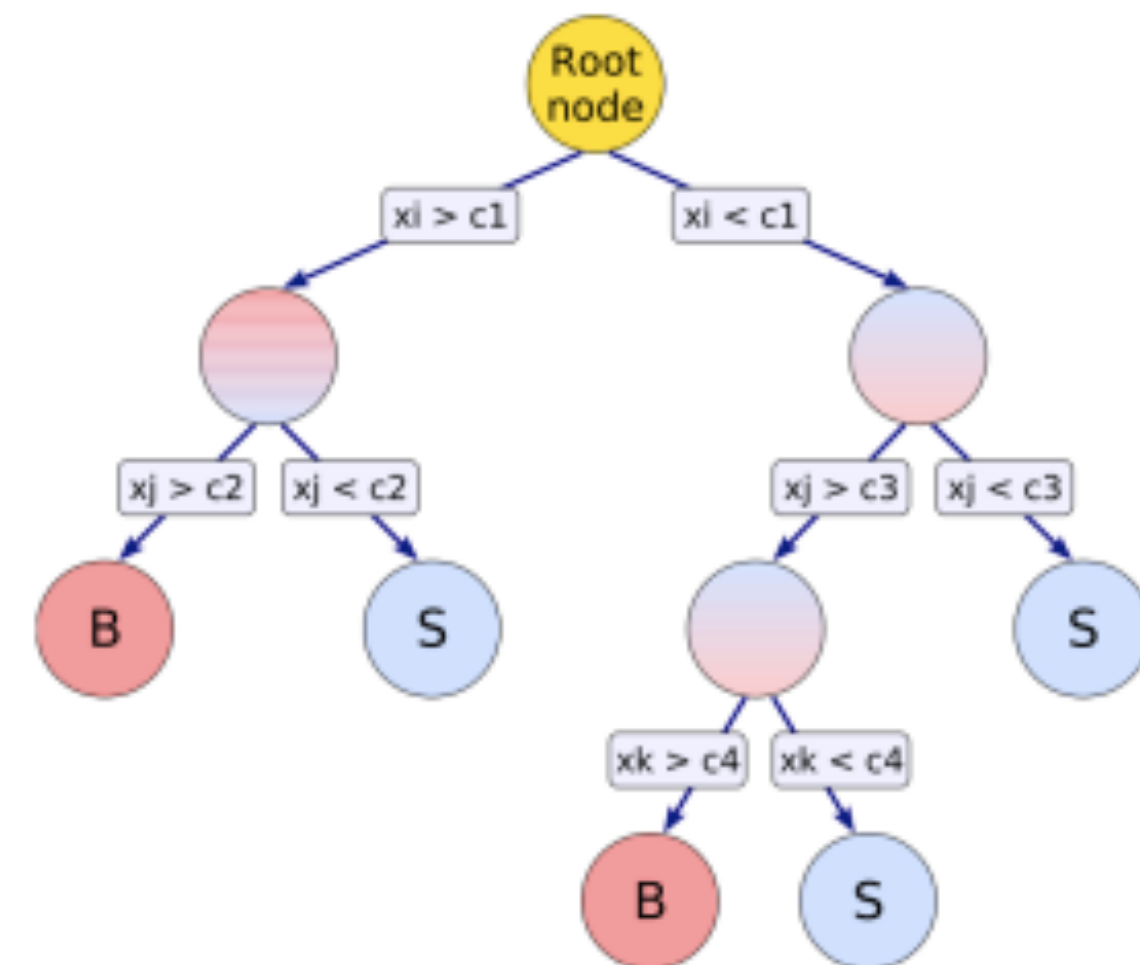
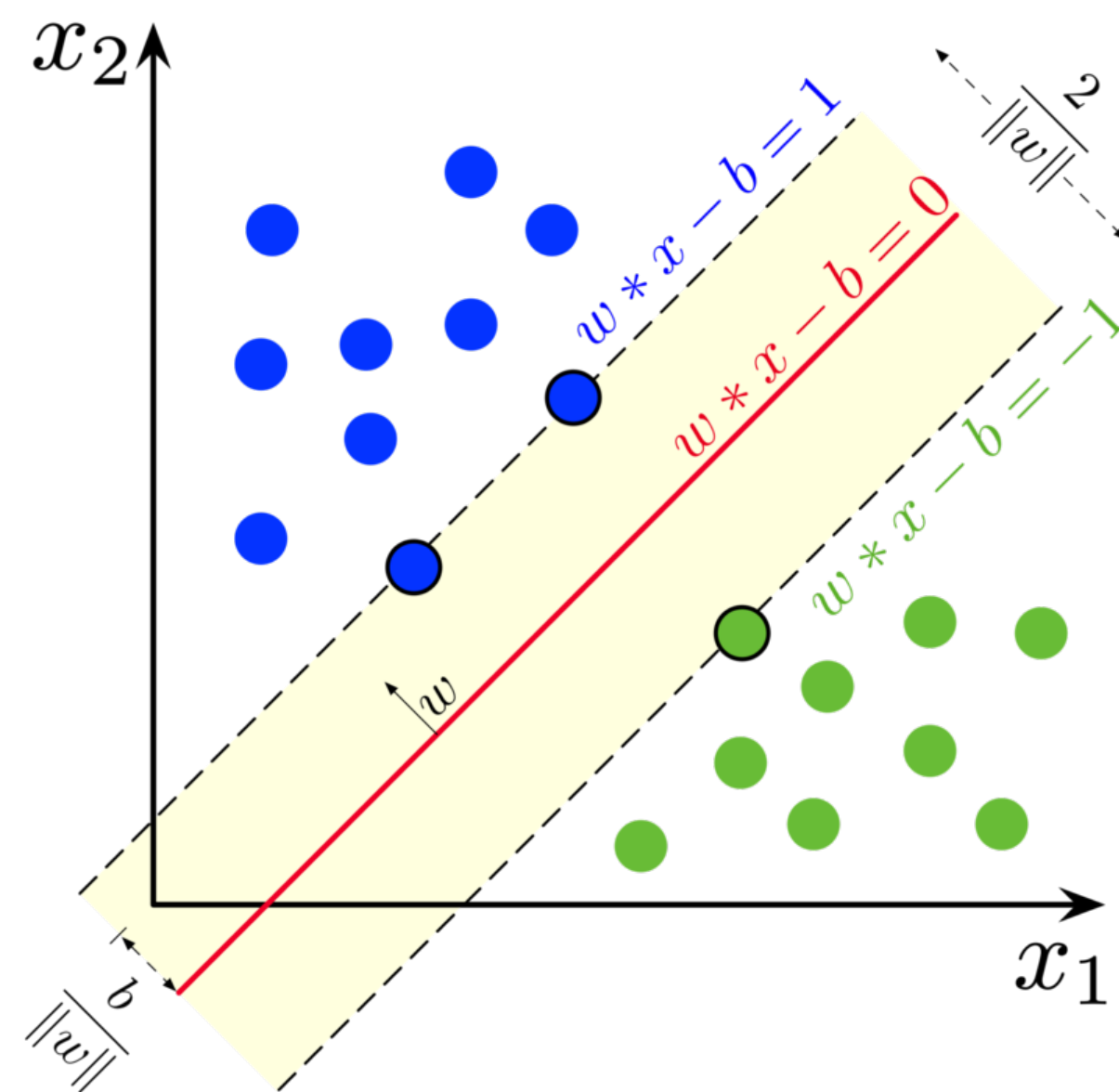
AI "winters": are we now living in an infinite "spring"?

Past "winters" of AI

- '60: **shallow NN** hard to train
- '90: Support Vector Machines (**SVM**), Boosted Decision Trees (**BDT**), ..
- 2000+: advanced **deep NN** architectures

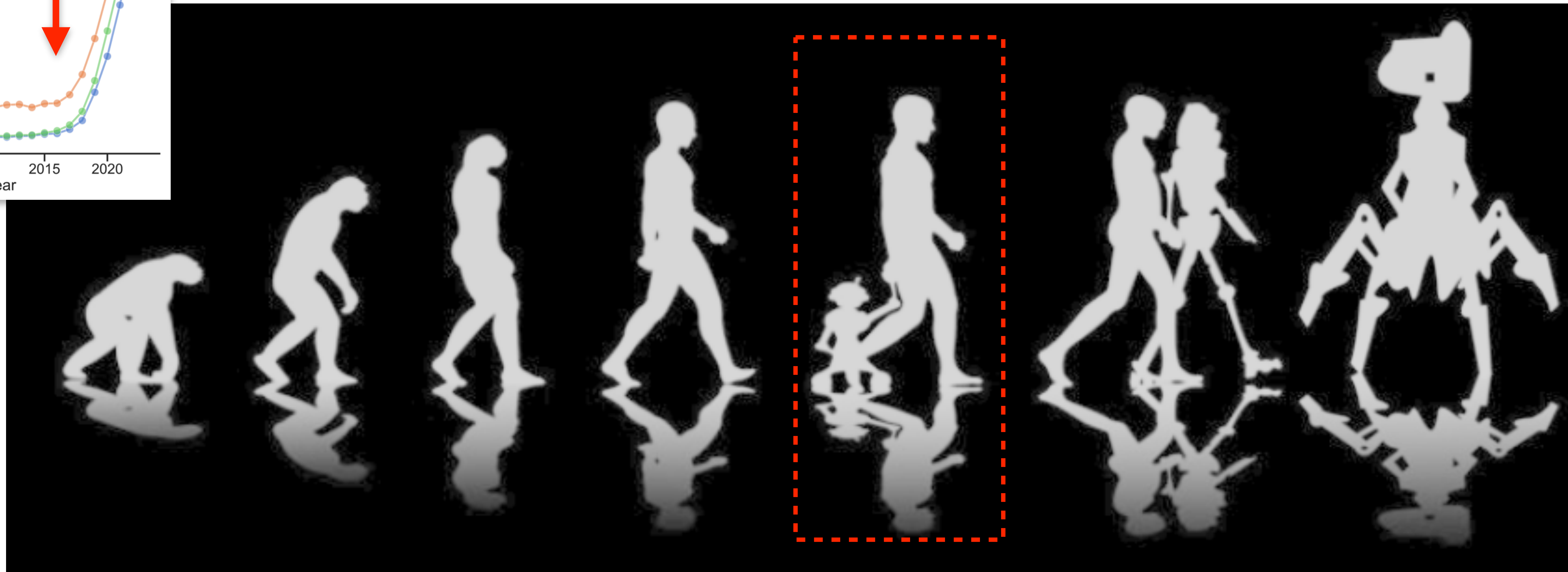
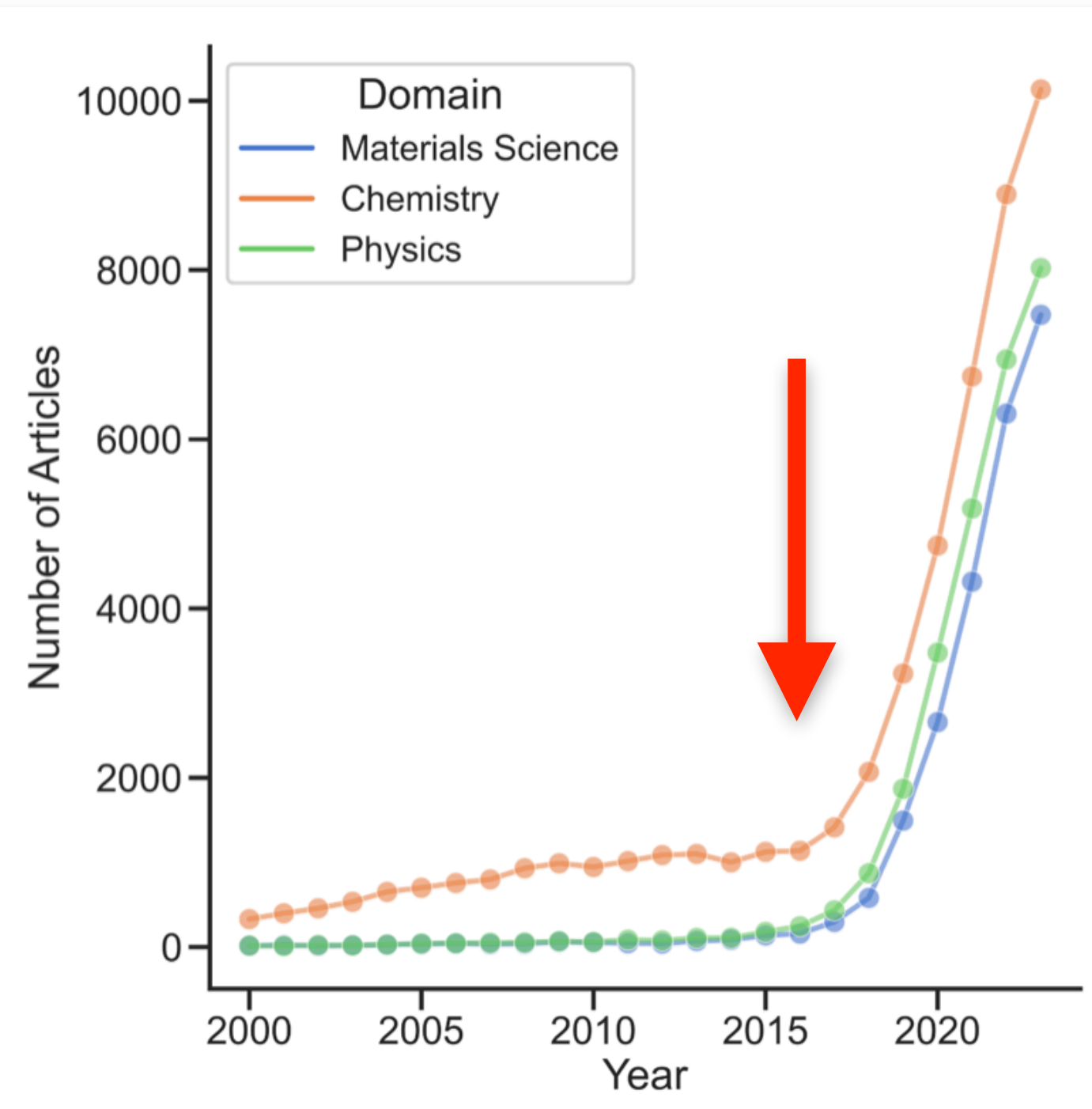
Current "**spring**" explainable by:

- "Big data"
- Technology + ML research
- (Accessibility)



ML/DL in HEP

“Traditional” ML



ML in HEP started by using domain knowledge to perform feature extraction/engineering

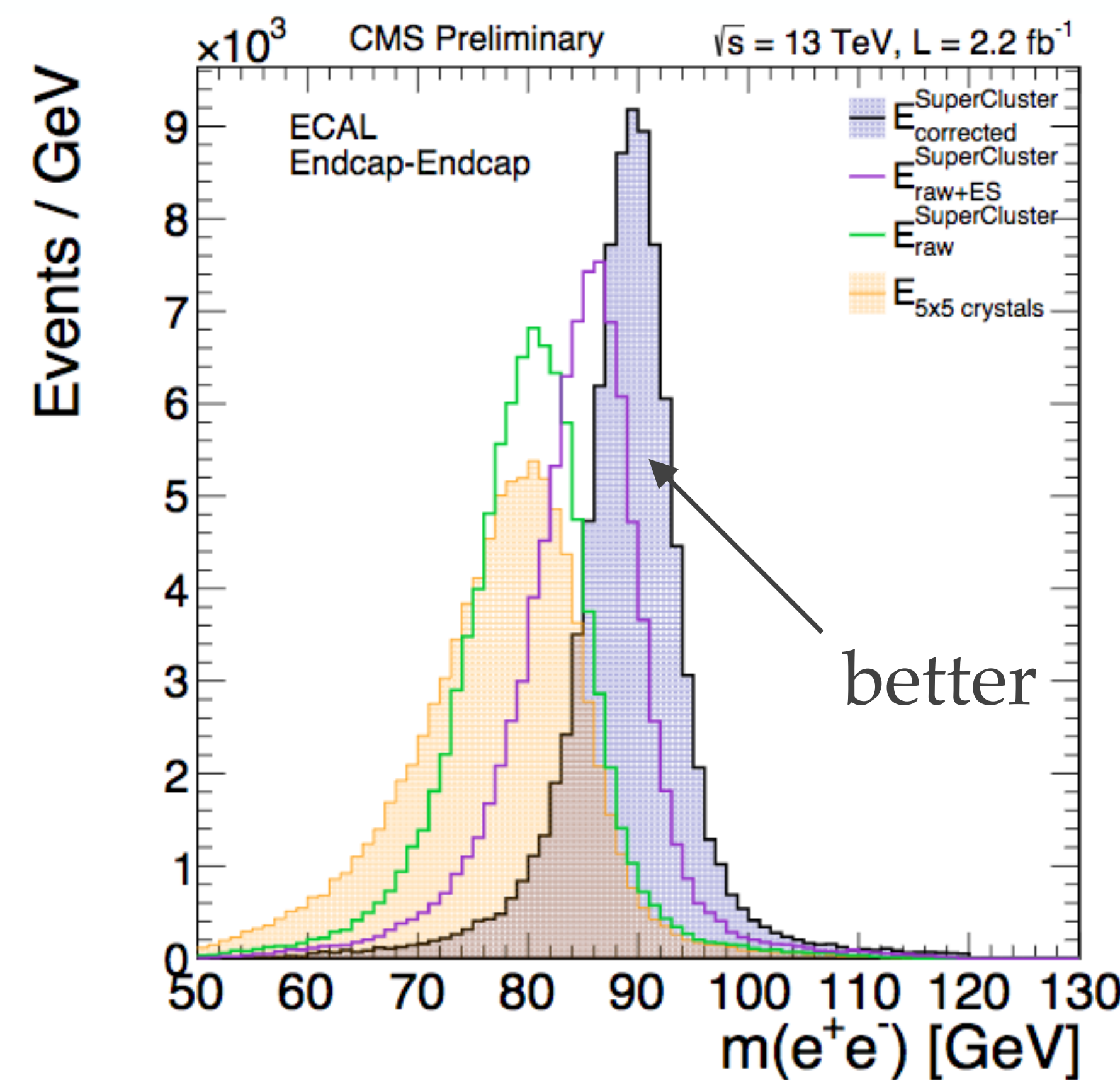
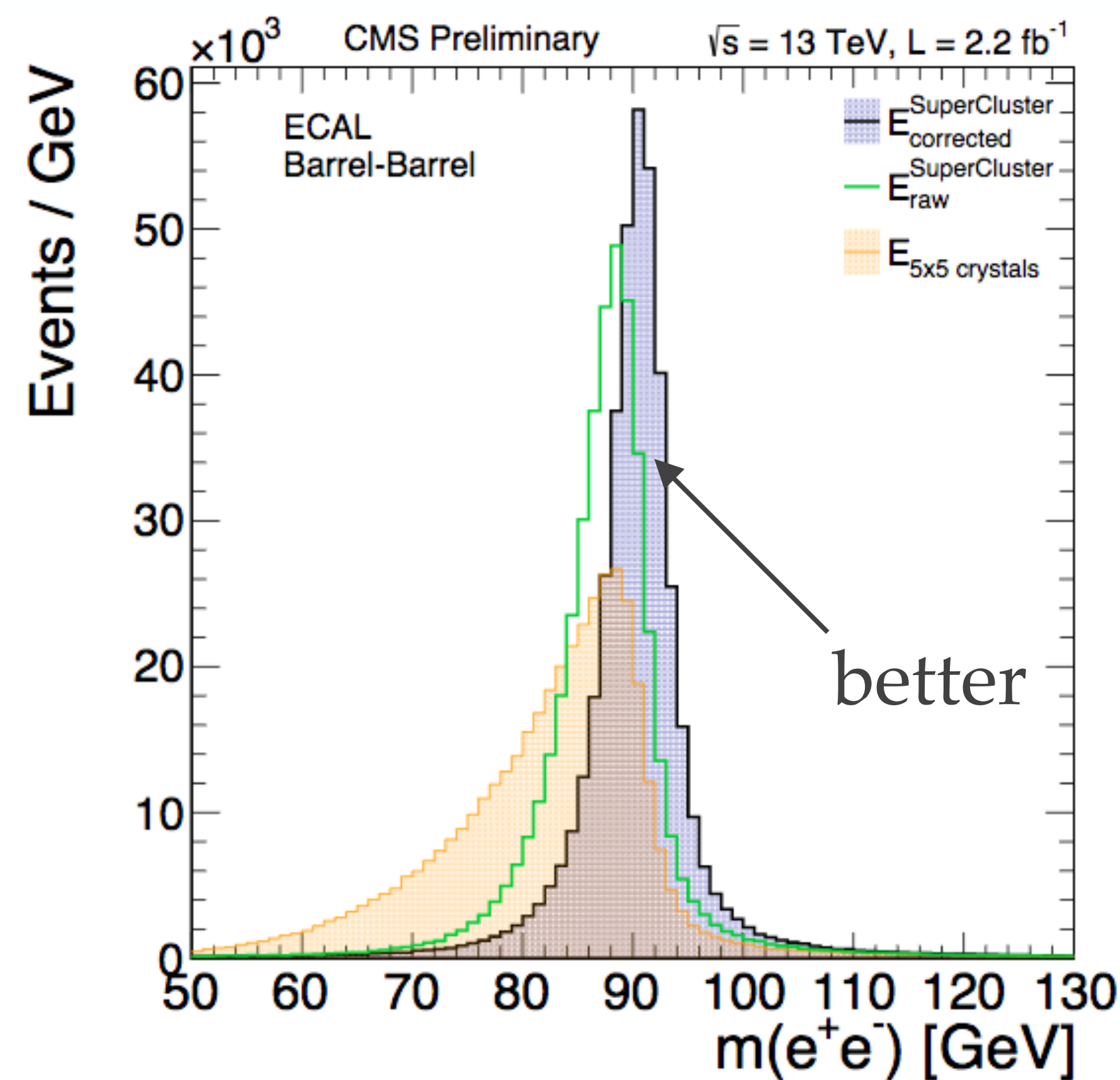
- HEP physicists design high-level features, and send them as input to traditional ML “shallow” algorithms

Particle id, energy resolution, and more..

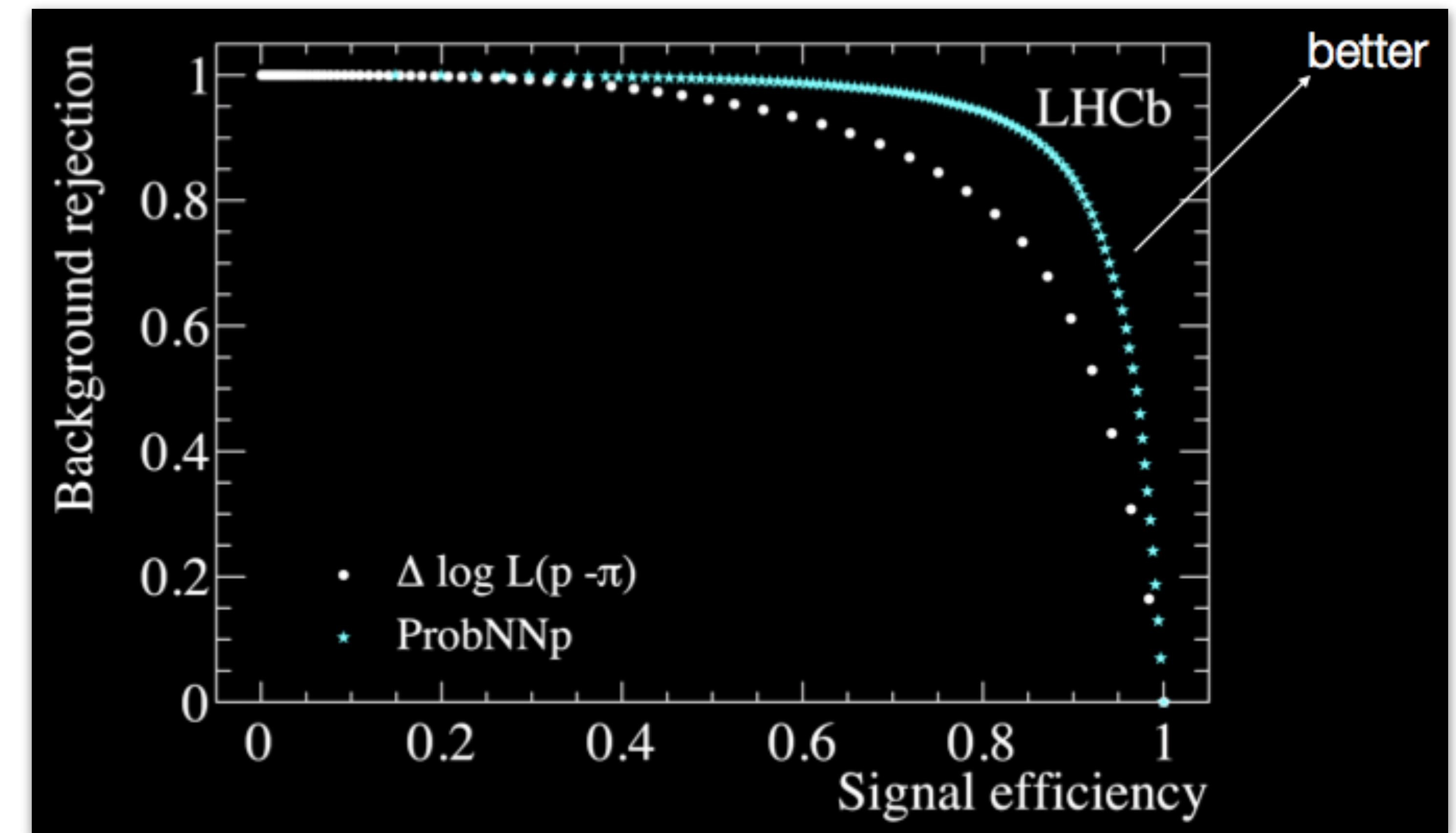
Using ML to improve the determination of particle properties is now commonplace in **all LHC experiments**

E.g. energy deposited in calorimeters is recorded by many sensors, which are clustered to **reconstruct the energy** of the original particle

- e.g. **CMS** trained **BDTs** to learn corrections using all information available in the various calorimeter sensors - thus resulting in a sizeable improvement in resolution



[2015 ECAL detector performance plots, CMS-DP-2015-057. Copyright CERN, reused with permission]



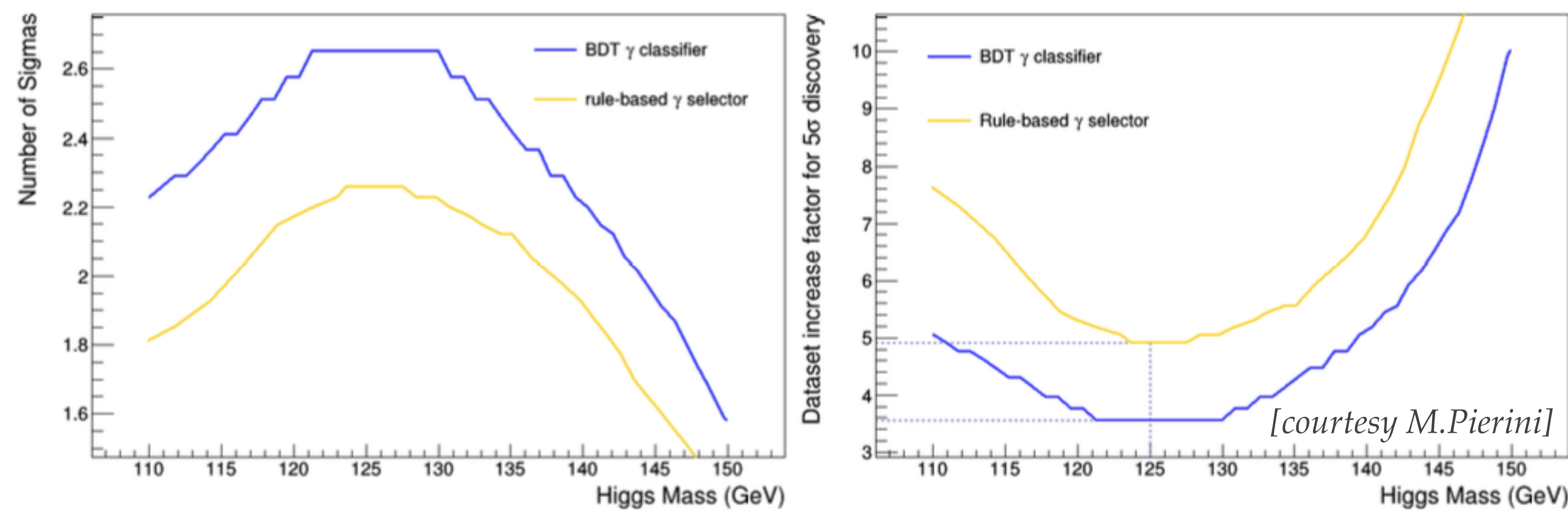
[courtesy: M.Williams]

Similarly, ML is commonly used to **identify particle types**

- e.g. **LHCb** used **NNs** trained on $O(30)$ features from all its subsystems, each of which is trained to identify a specific particle type
- ~3x less mis-ID bkg /particle. Further estimates indicated already that more advanced algorithms may reduce bkg by another ~50%

ML played a key role in the **discovery** of the Higgs boson

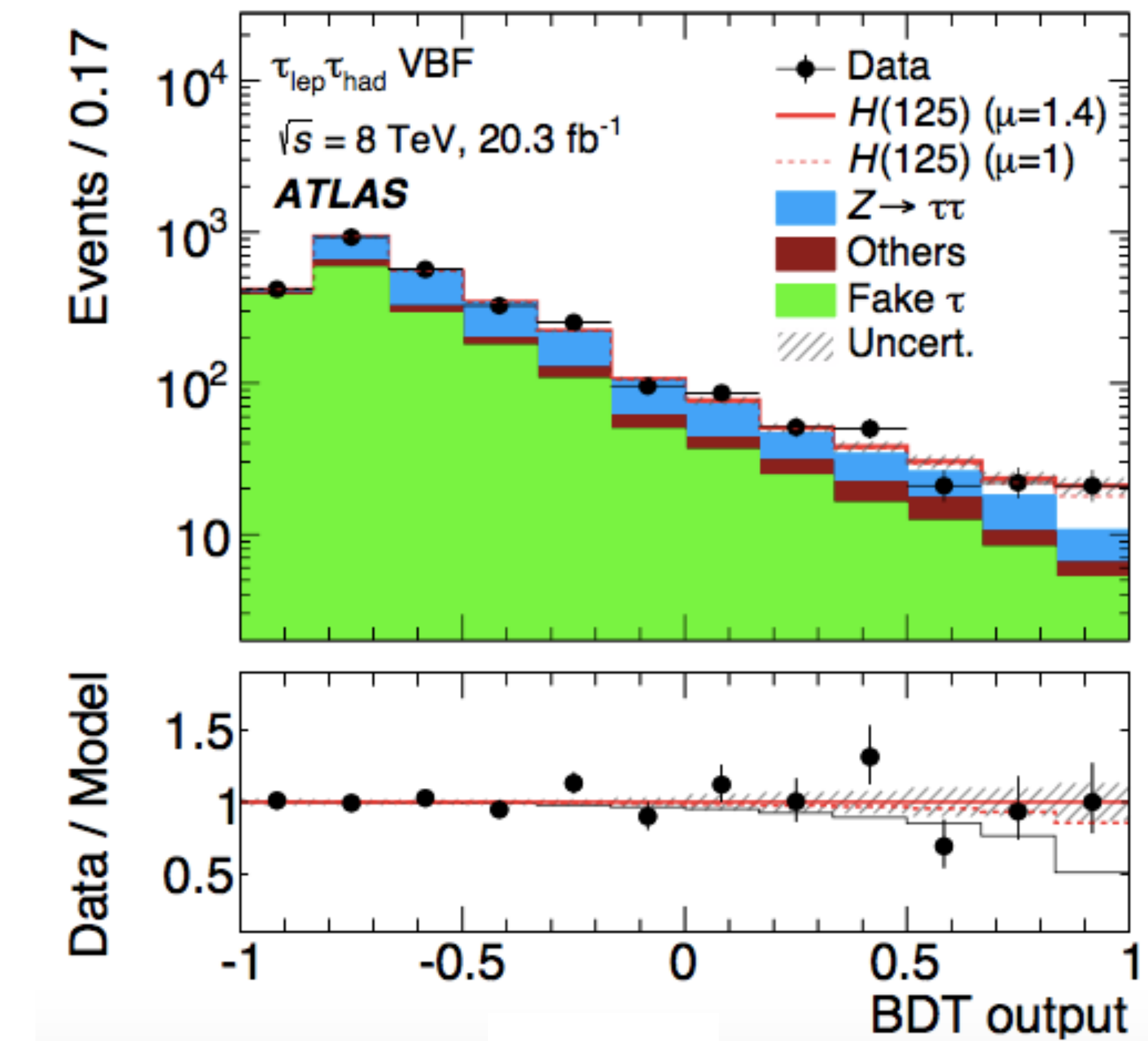
- especially in the diphoton analysis by **CMS**, where **BDTs** (used to improve the resolution and to select/categorise events)
- sensitivity increased by roughly the equivalent of collecting ~50% more data



We were not supposed to discover the Higgs boson **as early as 2012**

- Given how machine progressed, we expected discovery by end 2015 / mid 2016

We made it earlier thanks (also) to ML



ML impact also on the **study of Higgs properties**

→ e.g. analysis of τ leptons at LHC complex, as they decay before detection + loss of subsequently produced neutrinos + bkg from Z decays

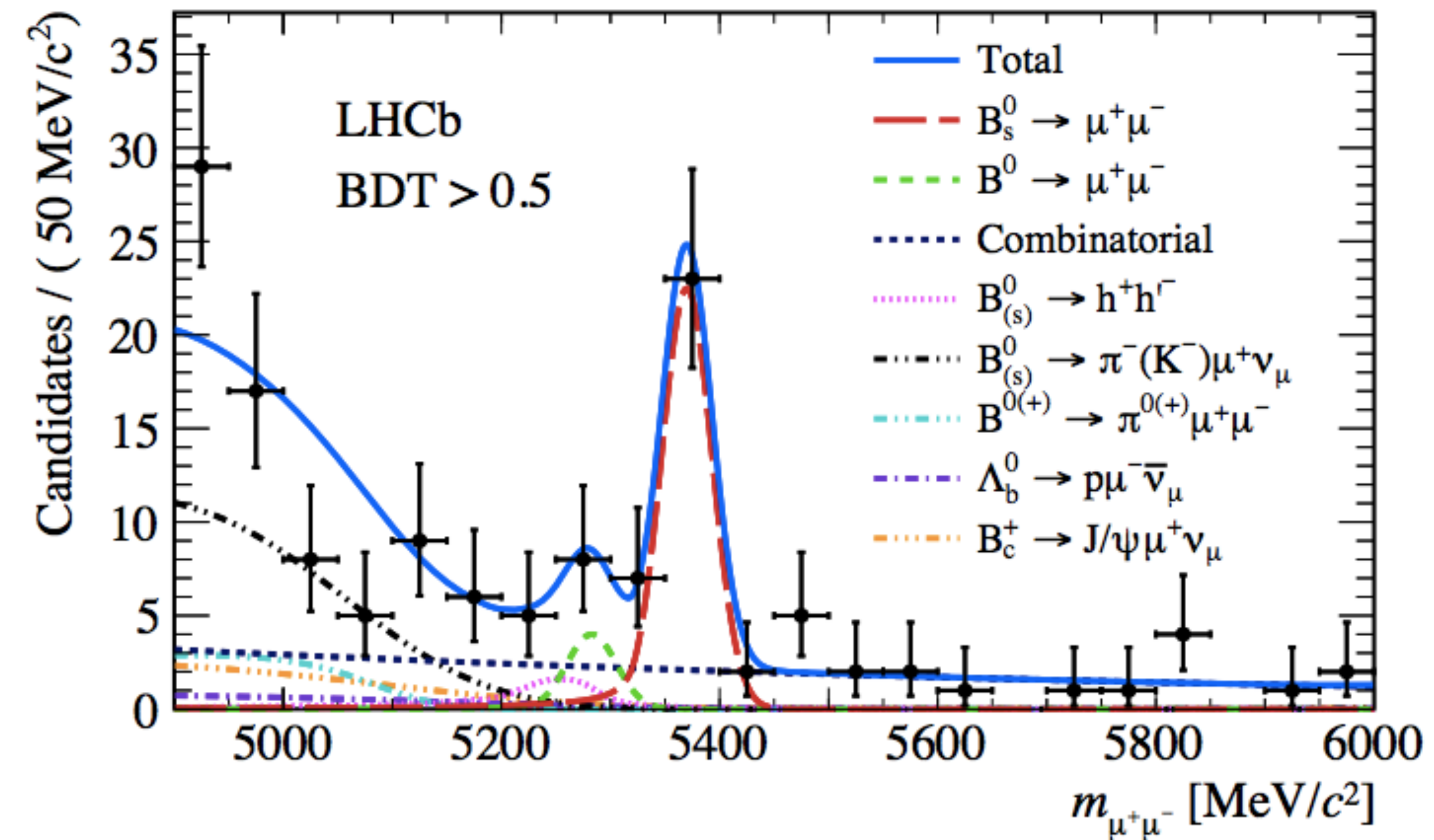
- e.g. **ATLAS** divided the data sample into 6 distinct kinematic regions, and in each a **BDT** was trained using 12 weakly discriminating features [1] → improved sensitivity by ~40% vs a non-ML approach

High-precision tests of the SM

CMS and **LHCb** were the first to find evidence for the $B^0_s \rightarrow \mu^+ \mu^-$ decay with a combined analysis [1] (as rare as $\sim 1 / 300$ billion pp collisions..)

- **BDTs** used to reduce the dimensionality of the feature space - excluding the mass - to 1 dimension, then an analysis was performed of the mass spectra across bins of BDT response
- decay rate observed is consistent with SM prediction with a precision of $\sim 25\%$, placing stringent constraints on many proposed extensions to the SM

Mass distribution of the selected $B^0 \rightarrow \mu^+ \mu^-$ candidates with $BDT > 0.5$ [2]



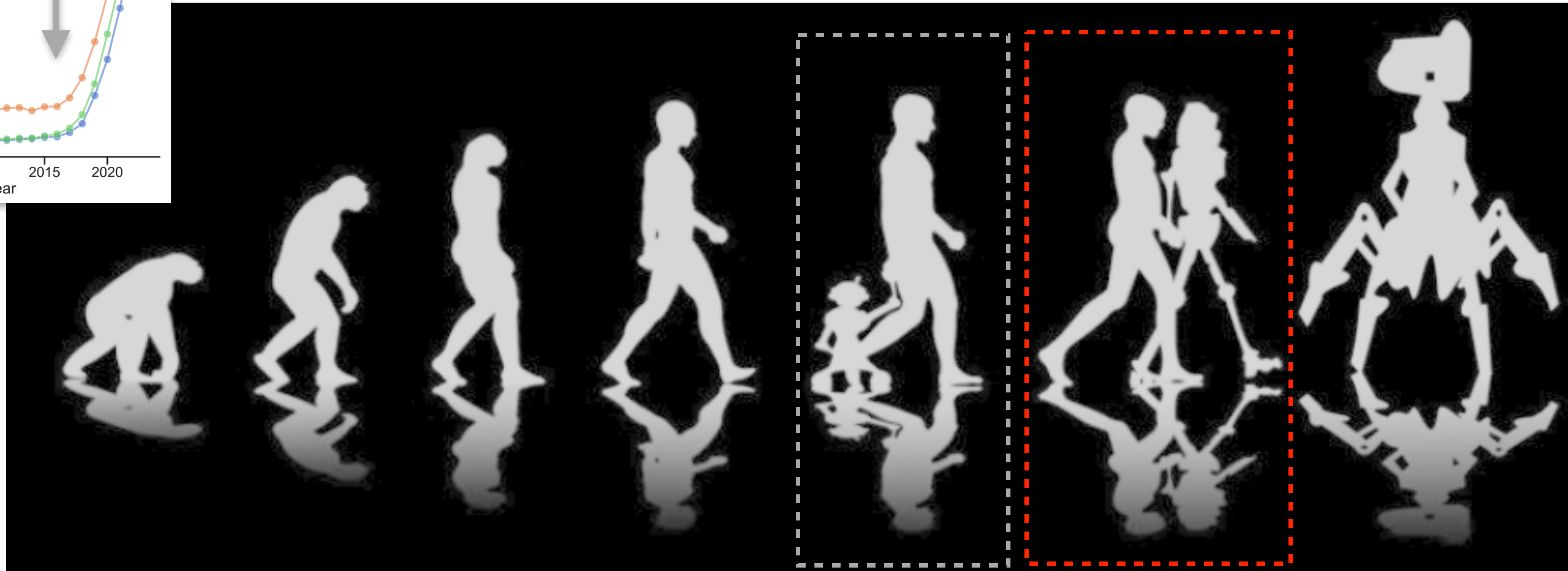
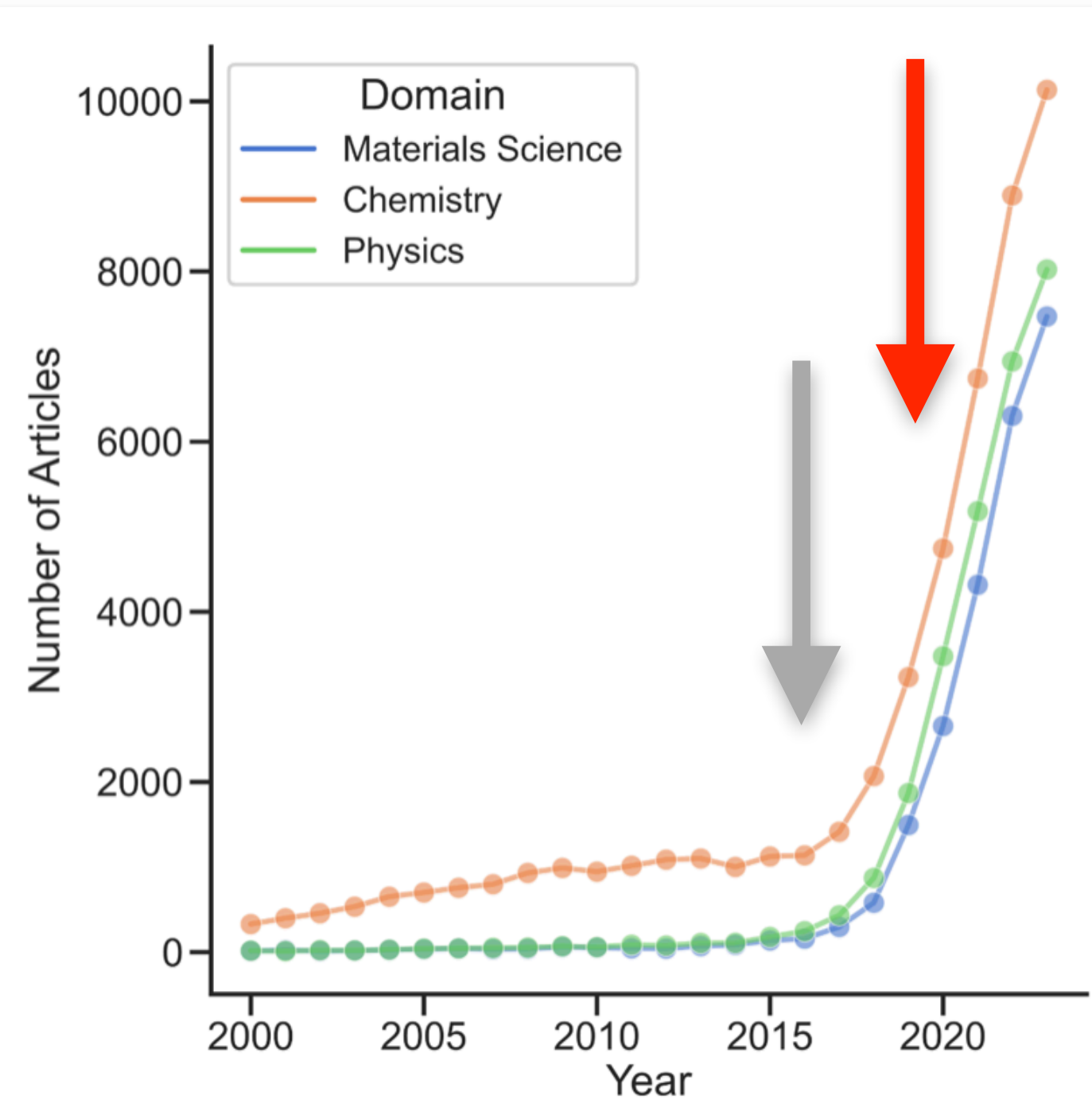
To obtain the same sensitivity without ML by LHCb as a single experiment would have required $\sim 4x$ more data

[1] Nature 522 68–72 (2015)

[2] Phys.Rev.Lett. 118 (2017) 19, 191801

ML/DL in HEP

“Traditional” ML Seeking DL solutions



Soon, ML (DL) in HEP started to seek for more advanced techniques, e.g. deep NNs

- **Use all the features space at its full dimensionality to train deep NN - no more manual feature engineering**

→ extract best from data, and do so by exploiting any architecture that might work for a given use-case (e.g. input from CV and NLP solutions..)

The zoo [1/2]

[DISCLAIMER: the list is not intended to be complete, and the classification is not rigorous but just for illustrative purposes]

Fully Connected Neural Networks (FCNNs / MLPs)

- Used in early applications (e.g. event classification, regression)
- Still widely used for tasks with structured tabular input (e.g. particle 4-vectors)
- *Examples: S/B discrimination, parameter estimation, ..*

Convolutional Neural Networks (CNNs)

- Suitable for image-like data: calorimeter hits, tracking detector layouts, jet images
- Benefit from local connectivity and translational invariance
- *Examples: jet tagging, energy deposition maps, neutrino detectors*

Recurrent Neural Networks (RNNs), then LSTMs and GRUs

- Handle sequential data, time-series, pulse shapes
- *Examples: Waveform analysis in neutrino or dark matter experiments, trigger streams*

Graph Neural Networks (GNNs)

- Represent events as graphs (e.g., hits, tracks, or particle interactions as nodes/edges)
- State-of-the-art for tracking, jet reconstruction, and physics object identification
- *Examples: Track finding, calorimeter clustering, particle flow*

Autoencoders (VAEs) (and Variational-AE)

- Used for anomaly detection and dimensionality reduction
- *Examples: Searching for rare or unknown physics events.*

Generative Adversarial Networks (GANs)

- Fast surrogate models for simulation (e.g., calorimeter shower generation).
- *Examples: Simulation acceleration, anomaly detection*

Transformers

- Originally from NLP, now extended to handle structured or variable-length (long!) inputs.
- Strong performance in classification and generative modelling, even in physics.
- *Examples: Event classification, generative modelling, scientific document parsing.*

Normalizing Flows

- Provide exact likelihood estimates and invertible mappings for complex distributions.
- *Applications: Simulation surrogates, likelihood-free inference, reweighting.*

Diffusion models

- Model data generation as reversing a diffusion process (progressive noise addition)
- *Examples: fast calorimeter and tracking simulation, anomaly detection, structured generation*

Physics-Informed Neural Networks (PINNs)

- Incorporate physical laws (e.g. PDEs) directly into the loss function
- *Examples: Solving PDEs in lattice QCD, fluid dynamics, or accelerator physics.*

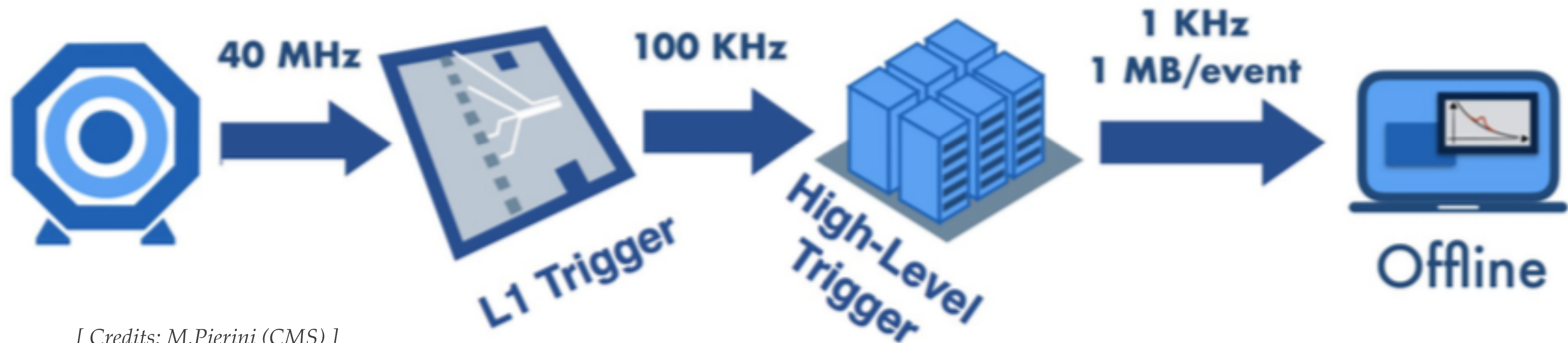
ML in **data acquisition** and **trigger**

- Bkg and trigger rate reduction
- Signal specific trigger paths
- Anomaly detection in data taking
- Unsupervised new physics mining

E.g. **LHC experiments' trigger** is a strong “driver” for **high-performances ML applications**

- Next-gen trigger systems → real-time reconstruction → real time analysis

Challenge is the trade-off between **algorithmic complexity** and the **performances** achievable under severe time constraints in inference

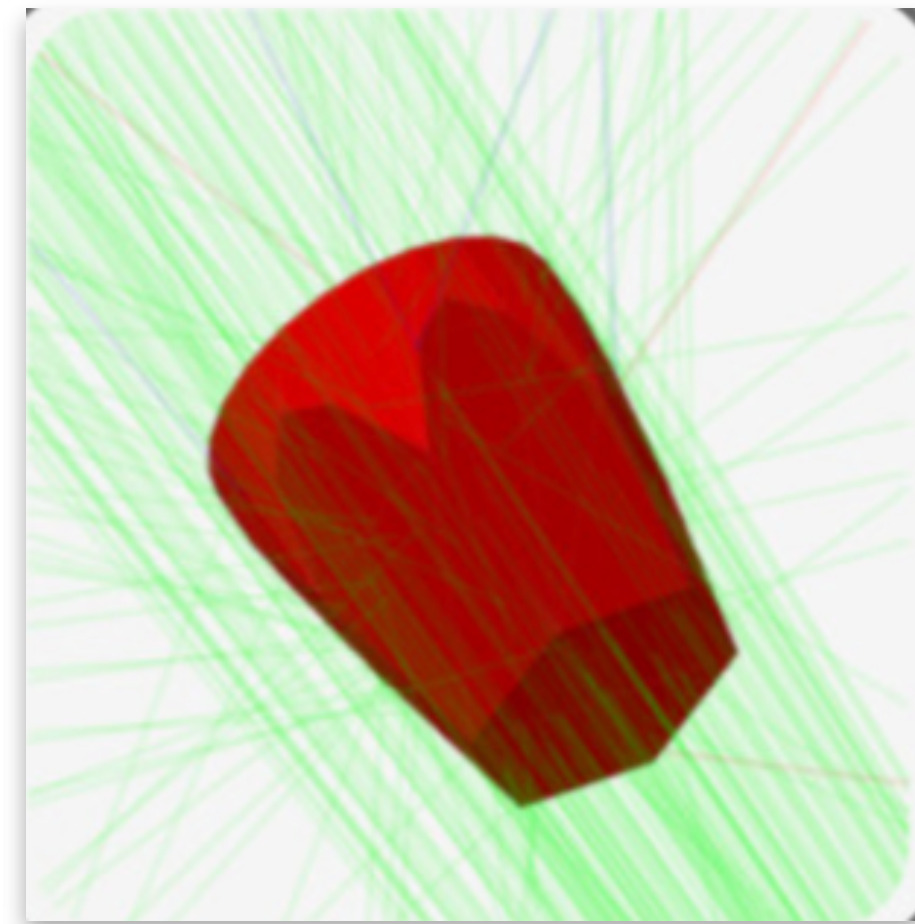


[Credits: M.Pierini (CMS)]

ML in Event **Simulation**

The production of simulated events (full/fast simulation) is extremely intense from the computation standpoint (up to the point it might impact the physics reach of the experiments). **ML can help to reduce such load**

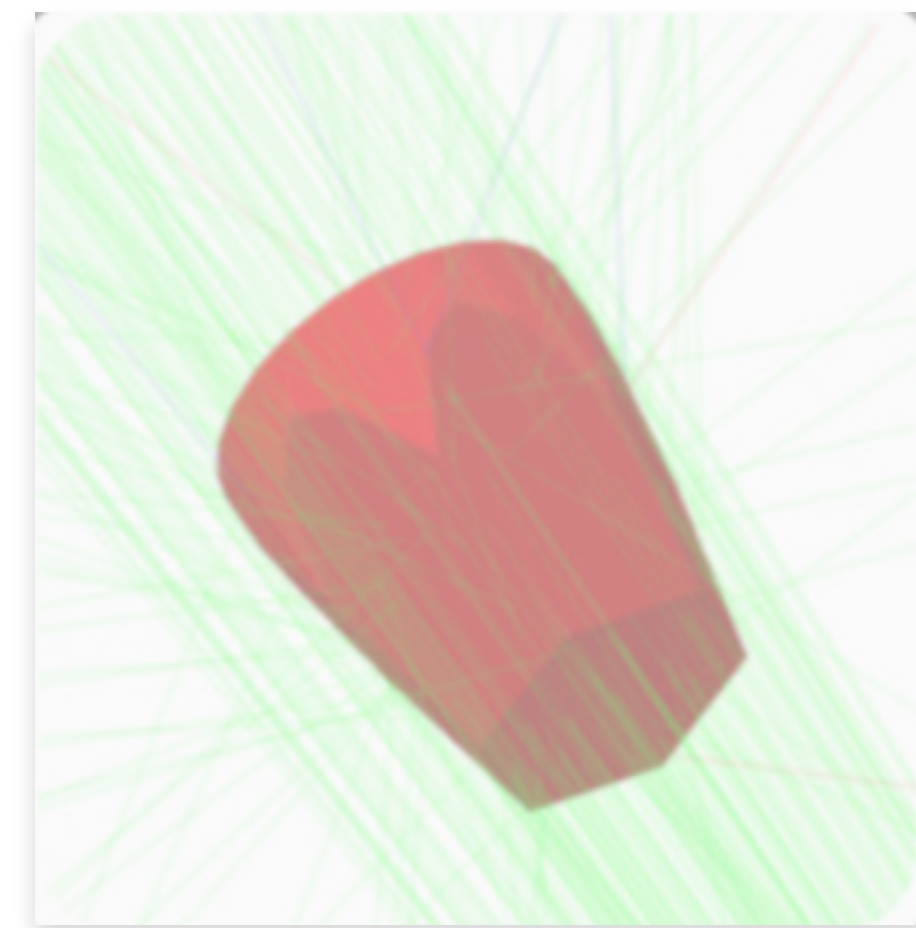
- Calorimeter shower surrogate simulator
- Analysis level simulator
- Pile-up overlay generator
- Monte Carlo integration
- ML-enabled fast-simulation
- Invertible full-simulation (probabilistic programming, ...)
- ...



ML in Event **Simulation**

The production of simulated events (full/fast simulation) is extremely intense from the computation standpoint (up to the point it might impact the physics reach of the experiments). **ML can help to reduce such load**

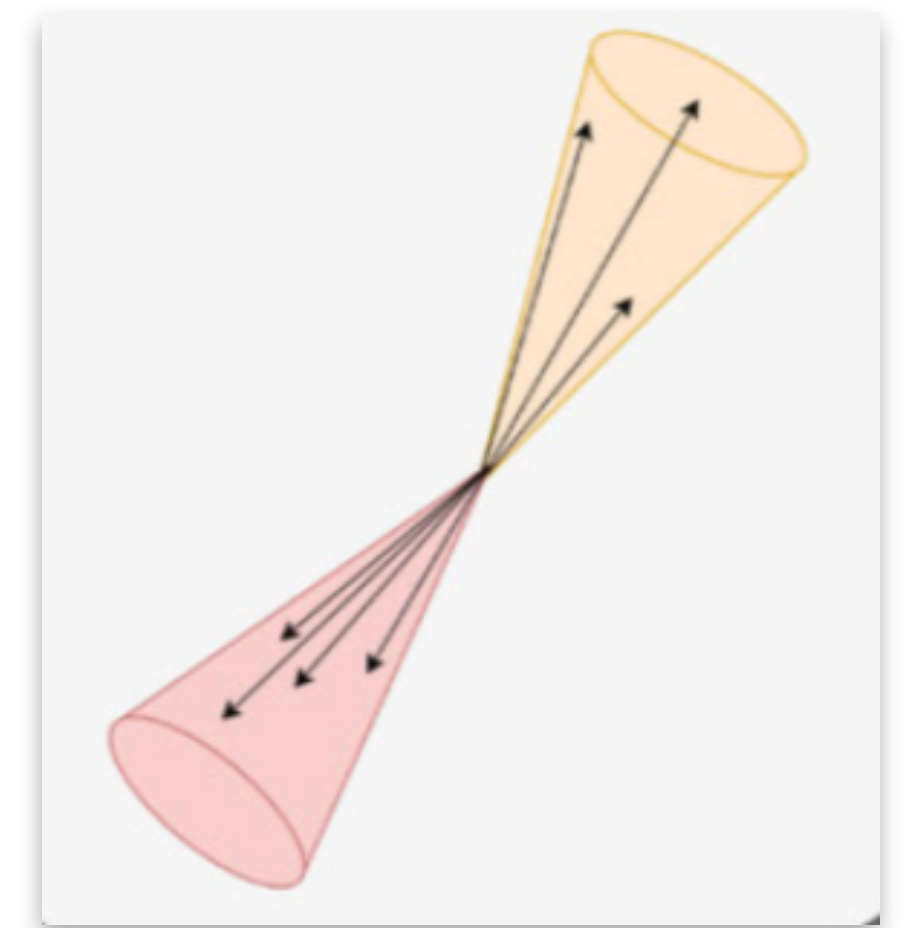
- Calorimeter shower surrogate simulator
- Analysis level simulator
- Pile-up overlay generator
- Monte Carlo integration
- ML-enabled fast-simulation
- Invertible full-simulation (probabilistic programming, ...)
- ...



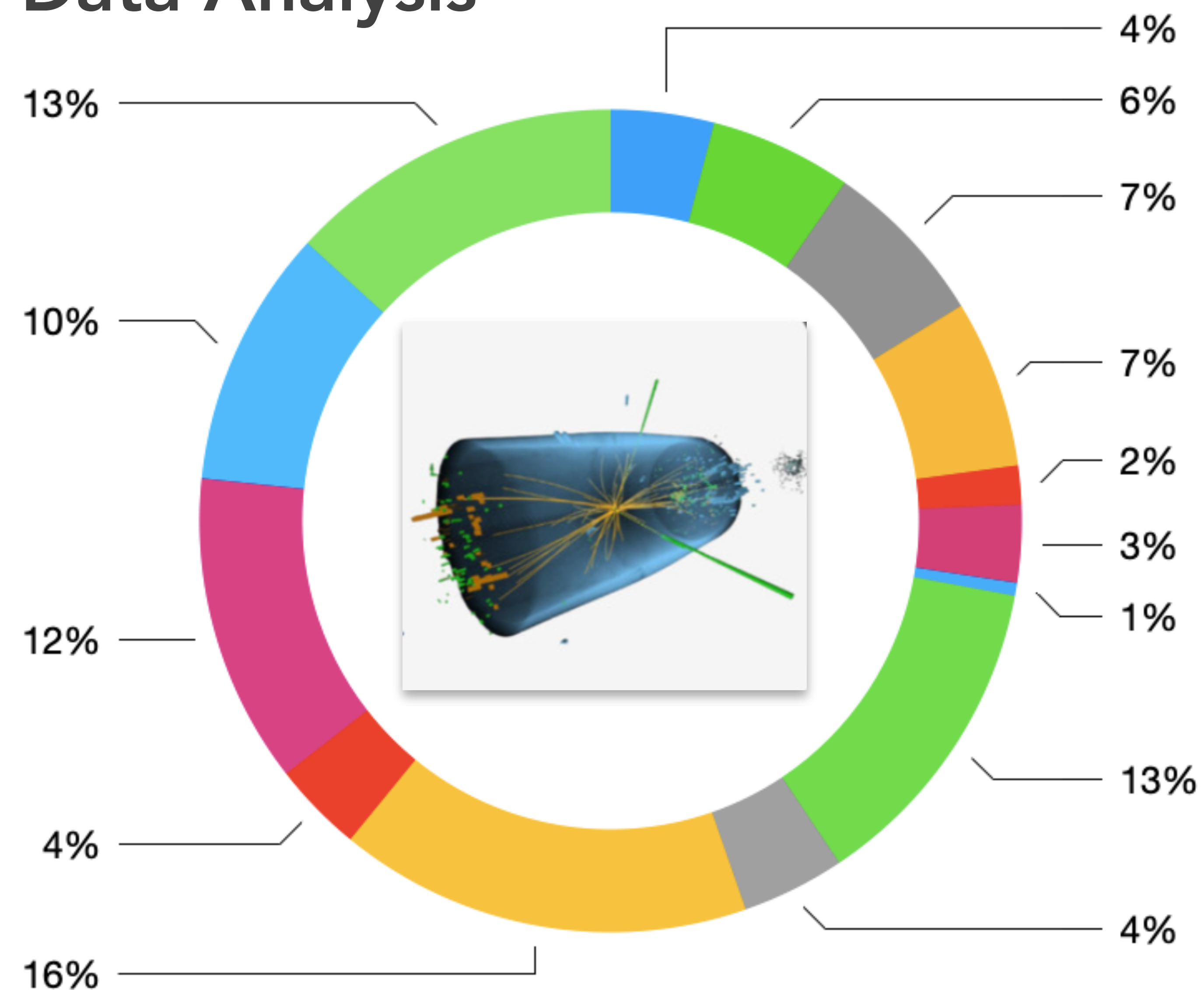
ML in Event **Reconstruction**

Online/offline reconstruction might be partially **replaced by surrogate models** (approximate → faster) or by **new algorithms** (that might offer unprecedented performances)

- Charged particle tracking (GraphNN, vertexing, ...)
- Calorimeter reconstruction (local, clustering, ...)
- Particle flow (GraphNN, ...)
- Particle identification (boosted jets, isolation, ...)
- Pileup mitigation
- Energy regression (end-2-end, ...)
- ...



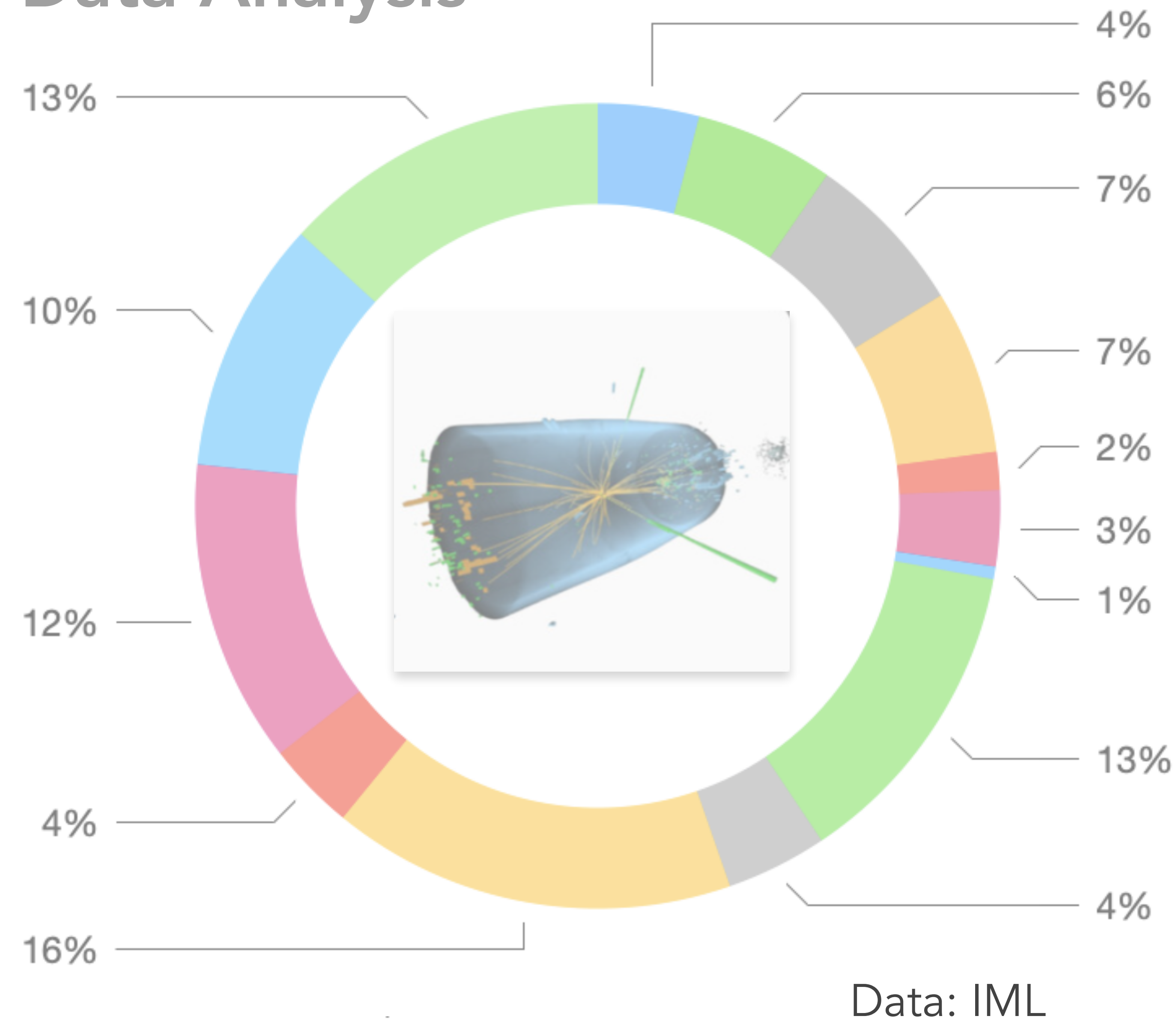
ML in Data Analysis



Data: IML

- W/Z tagging
- H→b bar
- quarks and gluons
- top quark tagging
- strange jets
- b-tagging
- flavor physics
- BSM particles and models
- particle identification
- neutrino detectors
- direct Dark Matter detectors
- cosmology, astroparticle, and cosmic ray physics
- tracking
- heavy ions / nuclear physics

ML in Data Analysis



- | | |
|----------------------------|--|
| ● W/Z tagging | ● particle identification |
| ● H→b bar | ● neutrino detectors |
| ● quarks and gluons | ● direct Dark Matter detectors |
| ● top quark tagging | ● cosmology, astroparticle, and cosmic ray physics |
| ● strange jets | ● tracking |
| ● b-tagging | ● heavy ions / nuclear physics |
| ● flavor physics | |
| ● BSM particles and models | |

ML in Computing Operations

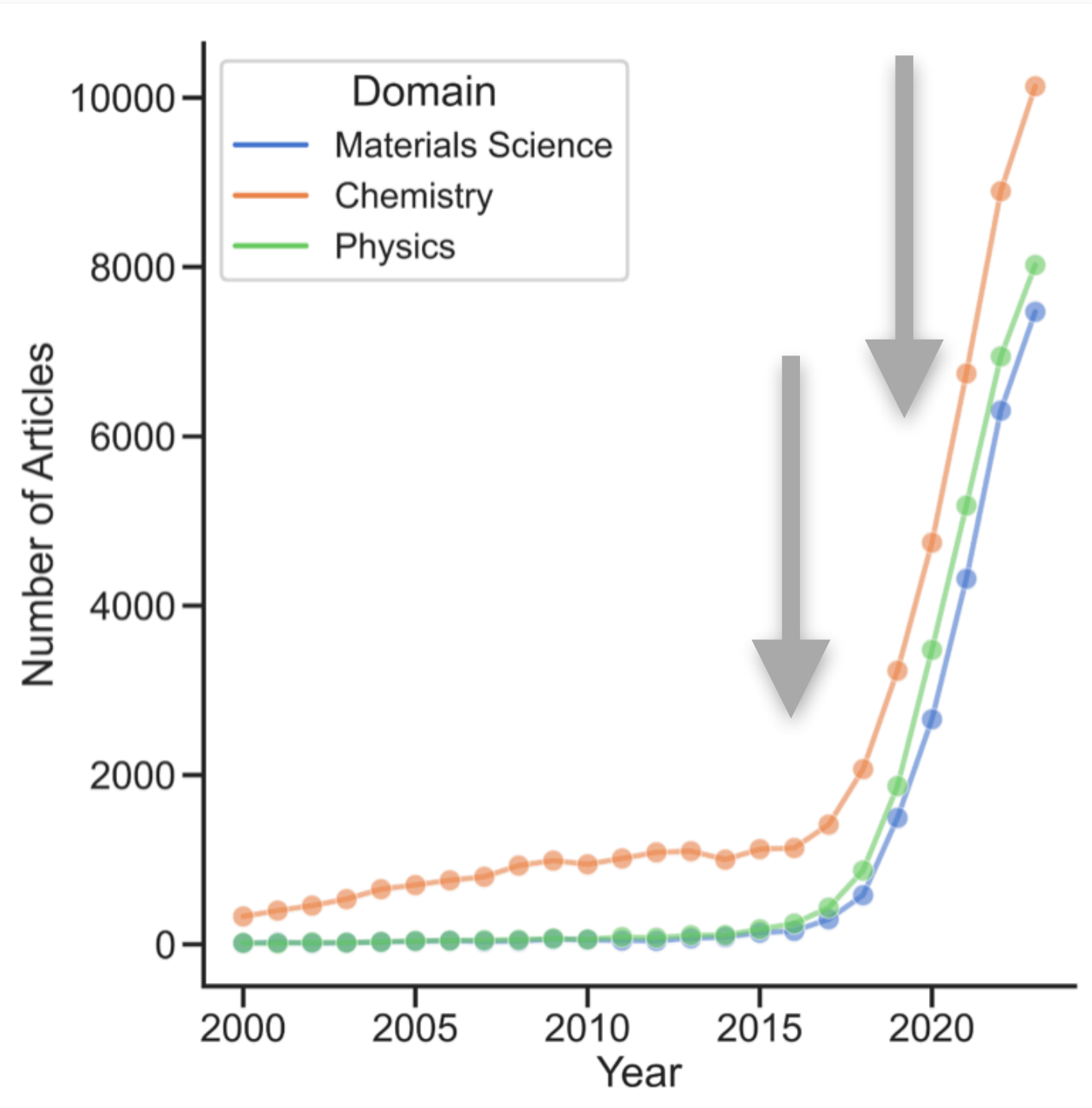
Application of ML to **non-event (meta-)data** might help to **increase efficiency** and **reduce the need of personpower** in Ops, e.g. automating specific tasks, developing intelligent/adaptive systems, ultimately acting on the full chain - from data collection to data analysis - and make it more agile

- Detector control
- Data quality monitoring
- Operational intelligence
- Predictive maintenance
- ...



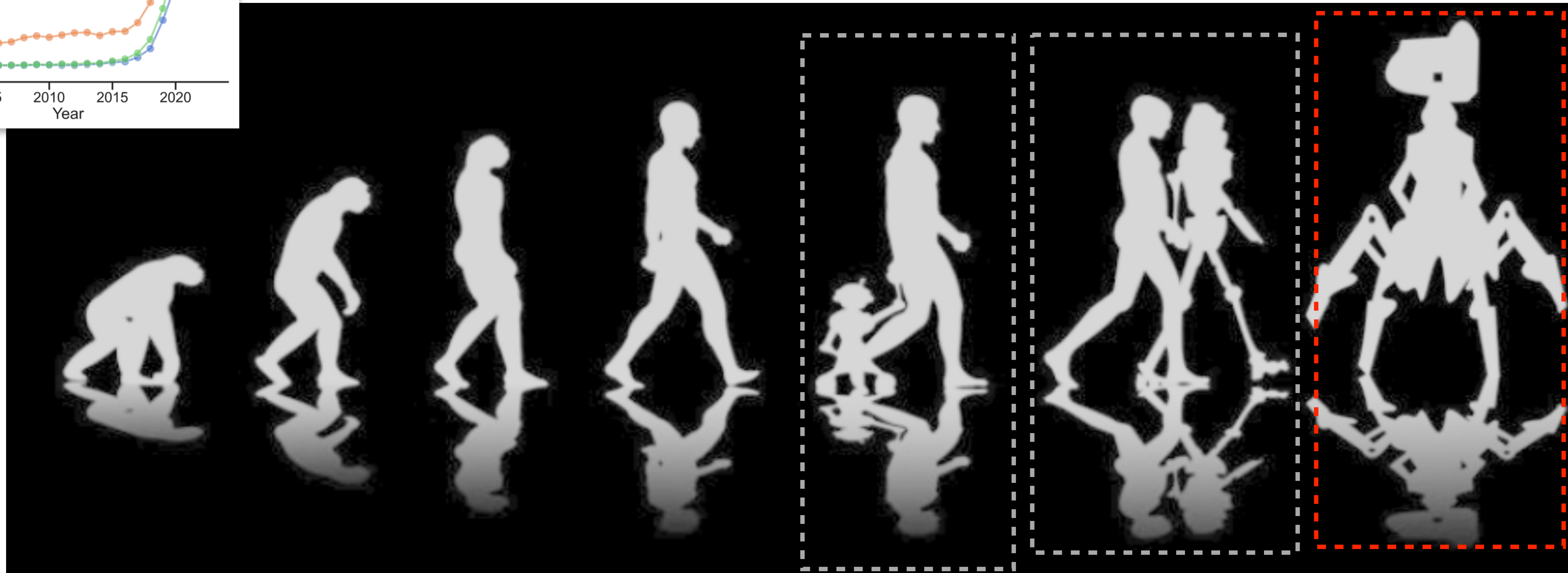
ML/DL in HEP

?

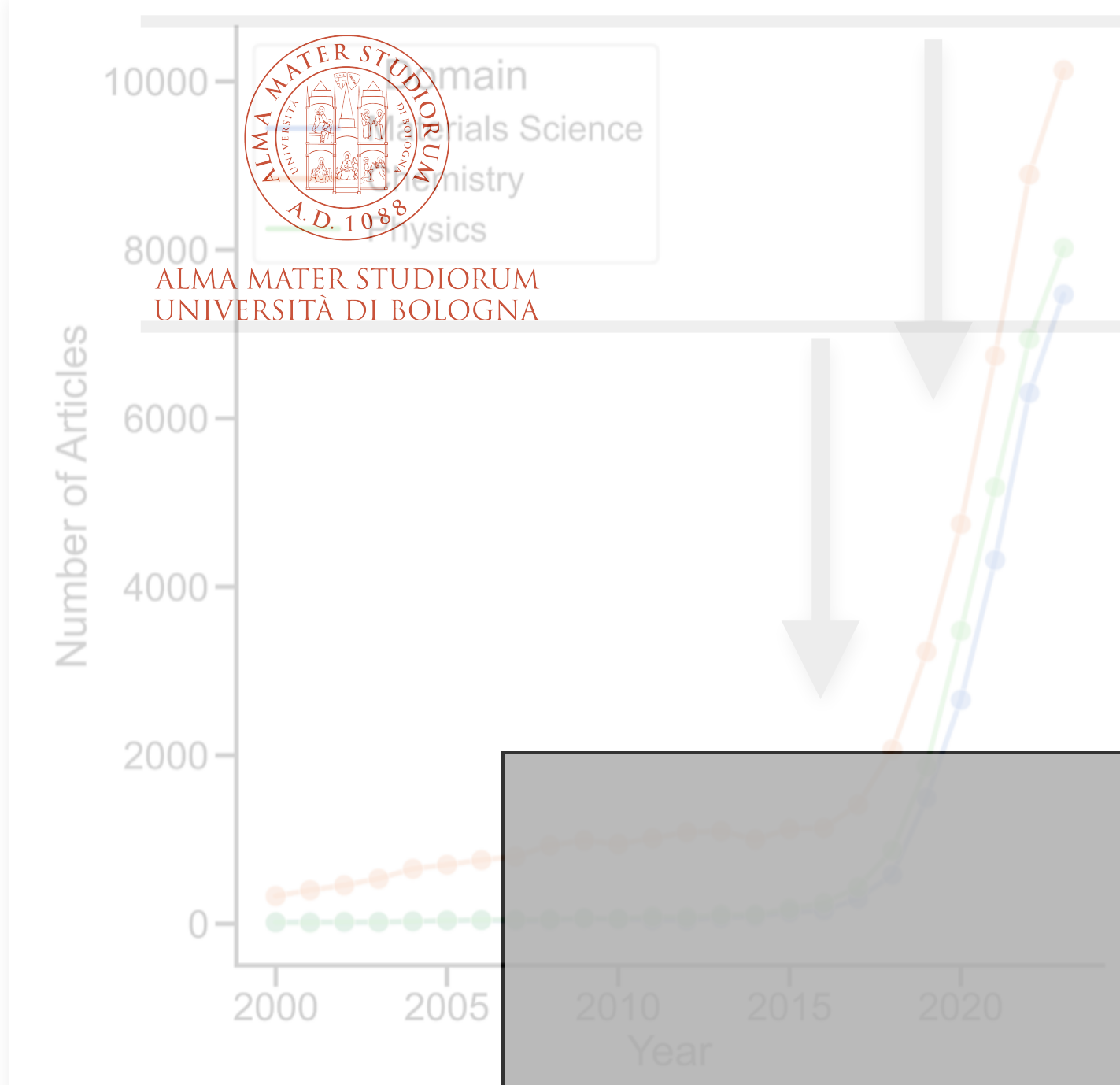


“Traditional” ML Seeking DL solutions

What next?



?



ML/DL in HEP



“Traditional” ML Seeking DL solutions

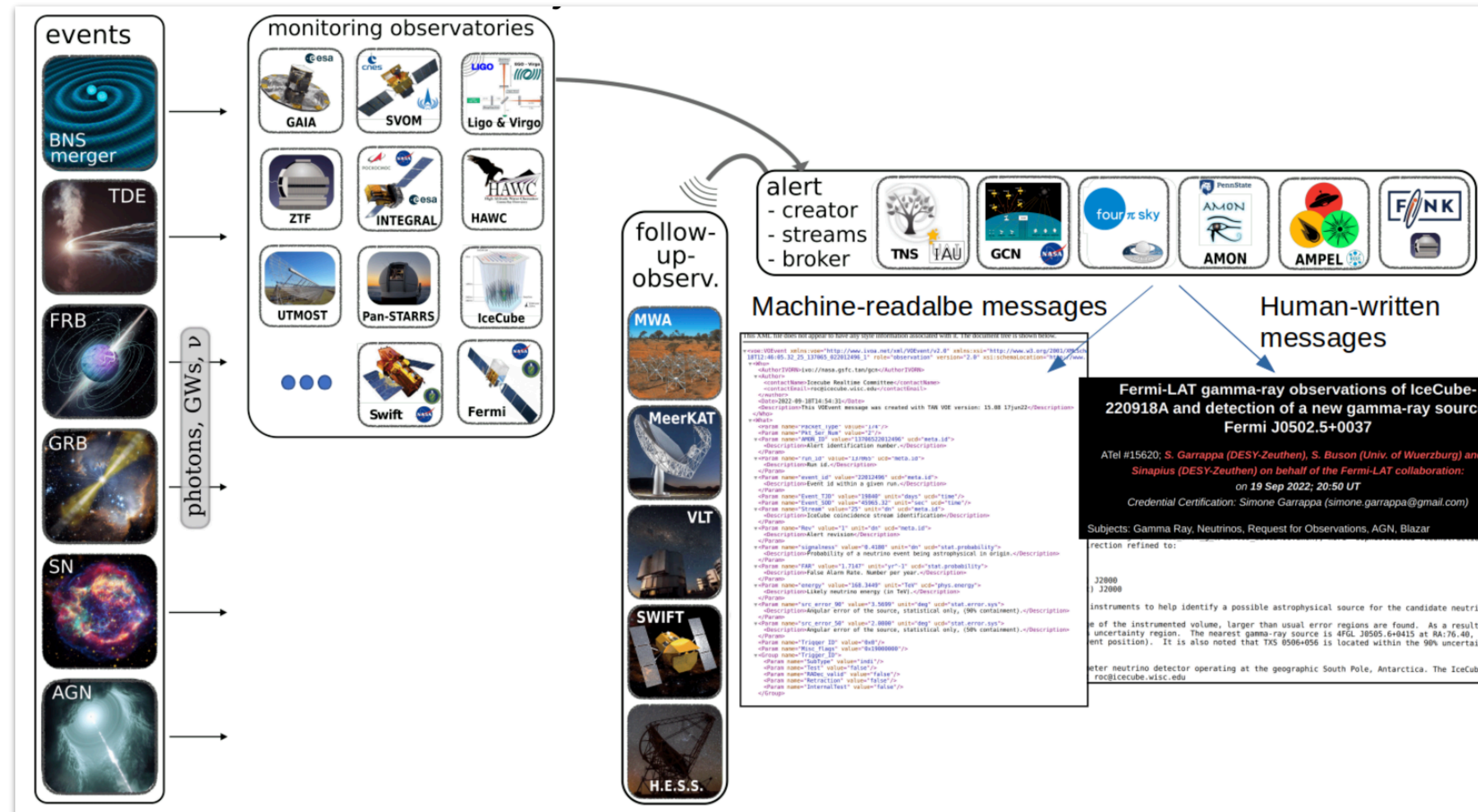
What next?



Of course... → a mention to **(second-generation) Large Language Models (LLM)**

LLMs for Multimessenger Astronomy

Observations in astronomy:



LLMs for Multimessenger Astronomy

*Description: Referred to by ATel #: 8706, 8718, 8783, 8789 On Jan 14, 2016, the Large Area Telescope (LAT) observed strong **gamma-ray emission** from a new source. The best-fit location of this **gamma-ray source** (RA=8.91 deg, Dec=61.52 deg, J2000.0) has a 95% containment radius of 0.08 deg. This source is not in any published LAT atalog and in the past has not been detected by AGILE or EGRET. The closest candidate counterpart is the **radio source** **87GB 003232.7+611352**.*

An example text based on
ATel messages
(astronomerstelegam.org)

■ object name; ■ the type of the object or physical phenomena; ■ event type

Goal is to build an information extraction system, i.e. recognise a list of predefined concepts (celestial objects, astronomical facilities, physical properties, people, organisation etc.) from a text and produce LLM-generated event summaries based on the parameters of each event

A Cycle 6 ALMA **obs** proposal for 4 epochs of stand-alone **Telescope** Atacama Compact Array **wlght** Band 7 observations of variables in **obj** Serpens identified by the **Survey** JCMT Transient Survey has been accepted (PI: **Person** Logan Francis, project code **Grant** 2018.1.00917.S). These observations will complement results from the contemporaneous Transient Survey by observing at **Wavelength** 850 μ m with a resolution of 38 (compared to the 146 resolution of the **tel** JCMT), sufficient to reach the scale of the inner envelopes (~ 1500 au) of protostars in **obj** Serpens.

From entity recognition...

... to semantic relationship extraction

Celestial Object MASTER OT J105440.86-391319.0 - PSN in PGC600519 (2.9"E,7"N) MASTER-SAAO auto-detection system (Lipunov et al. , " MASTER Global Robotic Net " , Advances in Astronomy , 2010 , 30L)
discovered OT source at (RA , Dec) = 10h 54 m 40.86s , -39d 13 m 19s.0 on 2016 - 10 - 08 02:56:46.019UT with m_OT=16.5 . This PSN is in 2.9"E,7"N from the center of PGC600519 and is seen on 8 images . We have reference image without OT on 2015 - 02 - 27 22:35:30UT with m_lim=19.8 Spectral observations are required The discovery and reference images are
<http://master.sai.msu.ru/static/OT/MASTEROTJ105440.86-391319.0.jpg> **Celestial Object** MASTER OT J160918.30 - 333456.7 discovery - bright possible dwarf nova outburst , ampl>5.4 m MASTER-SAAO

“*Cosmosage*”, a general-purpose AI-assistant specialised in answering questions about cosmology (based on Mistral-7B-v0.1)

- training dataset: arXiv papers, astro textbooks, physics textbooks, wikipedia

user:
 What came before the Big Bang?

Popular science

cosmosage:
 The concept of what came before the Big Bang is a topic of debate and speculation in the realm of cosmology. The Big Bang theory states that our universe began approximately 13.8 billion years ago from a singularity, an infinitely dense and hot point. At this point, space, time, matter, and energy were all created simultaneously. Scientists cannot explain with any certainty why the Big Bang happened, so any speculation about what came "before" is pure theoretical musings. The Big Bang is widely accepted as accounting for everything we have been able to observe about the universe, but it doesn't explain itself. There are many theories and models that attempt to explain the pre-Big Bang universe, such as the cyclic model, the ekpyrotic model, and the string theory landscape. However, these theories are still under development and have not been proven to be accurate descriptions of the actual events that took place before the Big Bang.
ChatGPT 3.5 gives a similar quality answer.

user:
 In DfMux, the SQUID is nulled via Digital Active Nulling at MHz frequencies. How is the SQUID kept nulled at low frequency?

Detailed Instrumentation

cosmosage:
 In DfMux, the SQUID is kept nulled at low frequency by using a low-frequency flux-locked loop (FLL). The FLL is a feedback system that maintains the SQUID in its nulled state by adjusting the input current to cancel out the flux signal. This is necessary because the SQUID is sensitive to magnetic fields, and the magnetic field generated by the input current can change the bias flux, causing the SQUID to lose its null state. The FLL helps to maintain the null state by continuously adjusting the input current to counteract any changes in the bias flux.
ChatGPT 3.5 says [... passive nulling ... without requiring active intervention...] which is exactly the wrong answer.

user:
 Why does a cross-spectrum not suffer from noise bias?

Data analysis principles

cosmosage:
 An auto spectrum suffers from noise bias because the noise is correlated with itself, leading to an overestimation of the signal. In contrast, a cross spectrum does not have this noise bias because the noise in different maps is uncorrelated. Therefore, cross spectra are often preferred for cosmological analysis because they provide an unbiased estimate of the signal.
ChatGPT 3.5 says [... cross-spectra are less prone to noise due to cancellation effects ...] which is incorrect.

user:
 What is the current best measure of the optical depth to reionization?

Current cosmological parameter constraints

cosmosage:
 The current best measure of the optical depth to reionization is $\tau=0.054\pm0.007$, as determined by the Planck Collaboration et al. (2018). This value provides strong evidence for reionization at high redshift and helps constrain models of cosmic reionization.
ChatGPT 3.5 talks about future measurements but does not mention a numerical value or even mentions Planck.



A LLM-based AI-assistant for a CERN experiment



"ChATLAS" a **prototype LLM project in a LHC experiment (ATLAS) at CERN** (as of end 2023)

Data **gathering** part is interesting (data **chunking** and data **retrieval** not described here)

- Docs: twiki (>2k), sw docs (>500), e-groups/emails archive (>10k), indico meetings' agendas incl. attached slides and minutes (>440k), Mattermost, Jira tickets, experiment' papers and internal notes (>66k)
- Either HTML or scraped into markdown

Many open challenges:

- highly heterogeneous data
- ensure that collaboration DBs are accessible and exportable; websites should live on a git repo; pubs should be saved as latex, and compiled separately; discussion forums should have anonymisation options... Estimates indicate that this would have saved ~1 yr of data wrangling
- Hallucinations are a real problem
- Not many gpu-hrs, but many expert-hrs, needed for any high-quality fine-tuned AI assistant



Educational Outreach with **AI-Assisted CERN Open Data**



opendata
CERN

Help About ▾

Explore more than **five petabytes**
of open data from particle physics!

Search

search examples: [collision datasets](#), [keywords:education](#), [energy:7TeV](#)

Explore

- [datasets](#)
- [software](#)
- [environments](#)
- [documentation](#)

Focus on

- [ATLAS](#)
- [ALICE](#)
- [CMS](#)
- [LHCb](#)
- [OPERA](#)
- [PHENIX](#)
- [Data Science](#)

<https://opendata.cern.ch/>

An example: a Higgs analysis guide

Define a training goal

ATLAS Open Data Higgs analysis guide

By Philipp Gadow

Guides on Higgs boson data analysis with ATLAS Open Data.

What is the Higgs boson and how can I measure its m... Show me how to plot the invariant mass distribution ...

How do I read ROOT files for my analysis? How can I fit the mass peak of the Higgs and how can...

Message ATLAS Open Data Higgs analysis guide...

ChatGPT can make mistakes. Consider checking important information.

Get draft code for it

```
sorted_indices = np.argsort(event_E[::-1])
if len(sorted_indices) < 2: # Check if there are at least two photons
    continue # Skip events with fewer than two photons
idx1, idx2 = sorted_indices[:2] # Indices of the two photons with largest energy

# Extract the leading and subleading photon four-momentum components
E1, E2 = event_E[idx1], event_E[idx2]
eta1, eta2 = event_eta[idx1], event_eta[idx2]
phi1, phi2 = event_phi[idx1], event_phi[idx2]

# Convert eta, phi, E to px, py, pz
px1,
```

Message ATLAS Open Data Higgs analysis guide...

ChatGPT can make mistakes. Consider checking important information.

Get explanations, refine, learn

you are calculating the invariant mass using these ene

GeV (gigaelectronvolts), this could be the source of discrepancy leading to an unexpected range or empty plots.

To fix this, you should convert the energies from MeV to GeV before performing the invariant mass calculation. This can be done by dividing the energy values by 1000:

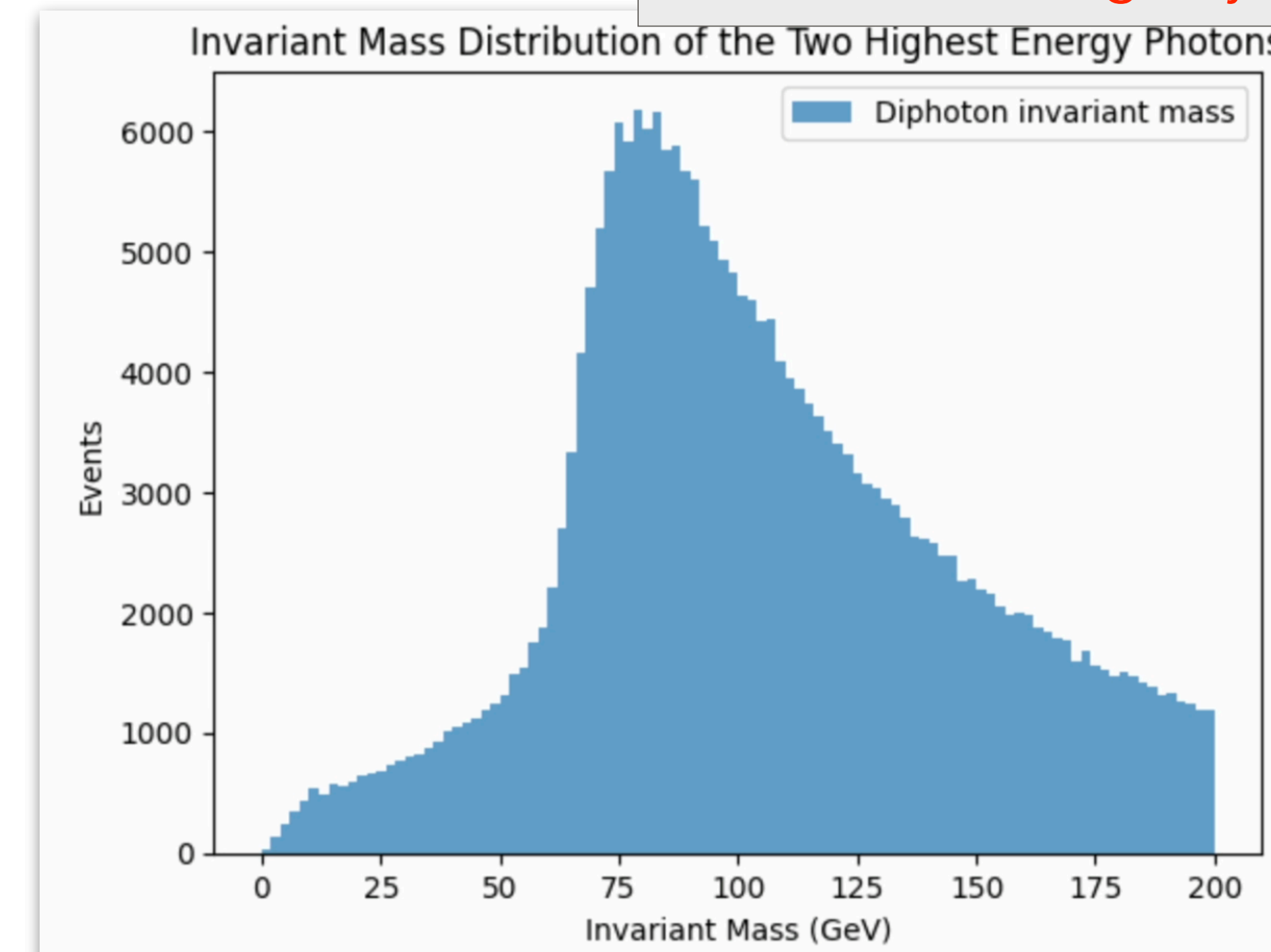
```
python
E1, E2 = event_E[idx1] / 1000., event_E[idx2] / 1000. # Convert from MeV to GeV
```

Incorporate this conversion into your code where you define `E1` and `E2` and then plot the invariant mass distribution again. This adjustment should help

Message ATLAS Open Data Higgs analysis guide...

ChatGPT can make mistakes. Consider checking important information.

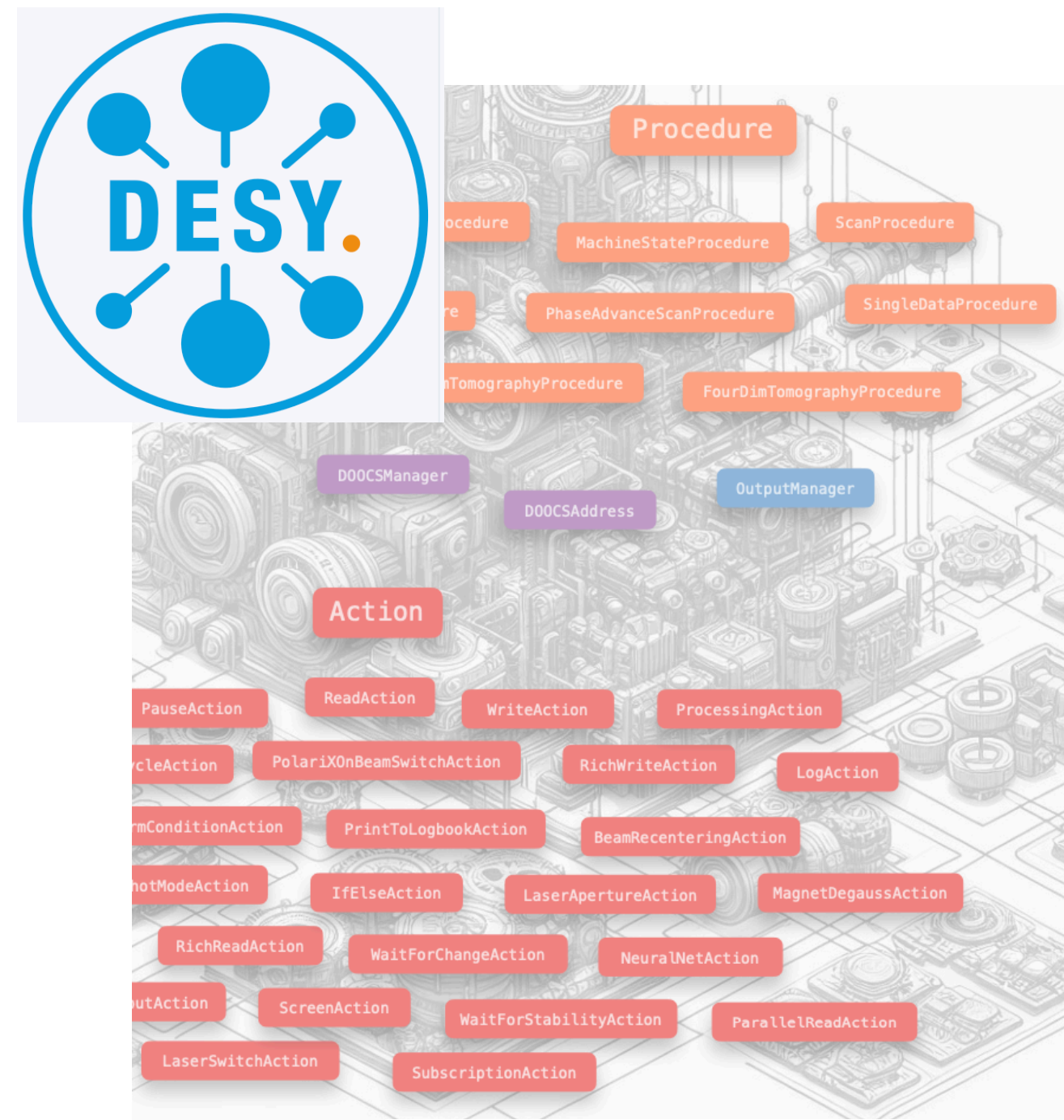
Reach the training objective



A variety of projects..

Plenty of work in progress on LLMs, showing potential towards natural language driven autonomous particle accelerators

- Attempts with GPT 3.5 Turbo, Megadolphin, Vicuna 7B 16K, Mistral 7B, Mixtral 7x8B, Starling-LM, GPT 4 Turbo, GPT4, Orca 2 7B, Orca 2 13B , Llama 2 70B, Falcon 180B, ..
- Constant seek for (and tests with) better models, better prompting, ...



GAIA (@**DESY**): a General AI-assistant for Intelligent Accelerator Ops

- Experimental “procedures” defined as a collection of high-level “actions” in a Control System e.g. for managing machine pre-sets
- Exploring a LLM (mixtral:8x7b-instruct-v0.1-q8_0 with 32k context size), agent implemented in Python using the langchain module, prompting based on ReAct (as a combination of chain-of-thought prompting and information injection via “actions”)

EPA project (@**CERN**), AccGPT, etc..



- EPA = Efficient Particle Accelerator project
- AccGPT = accelerating science via a chatbot for knowledge retrieval for CERN specific content

A possible (?) path



[Credits: Sasha Caron]

A possible (?) path

OK, the past..



[Credits: Sasha Caron]

A possible (?) path

OK, the past..

OK, the present..



[Credits: Sasha Caron]

A possible (?) path

OK, the past..

OK, the present..

Yes, work in progress..



[Credits: Sasha Caron]

A possible (?) path

OK, the past..

OK, the present..

Yes, work in progress

Ok, well.. maybe..

BDTs

Deep
Learning

Generative
Models

Foundation
Models

Large
Physics
Models

[Credits: Sasha Caron]

A possible (?) path

OK, the past..

OK, the present..

Yes, work in progress

Ok, well.. maybe..

Wait, what?!

BDTs

Deep
Learning

Generative
Models

Foundation
Models

Large
Physics
Models

[Credits: Sasha Caron]

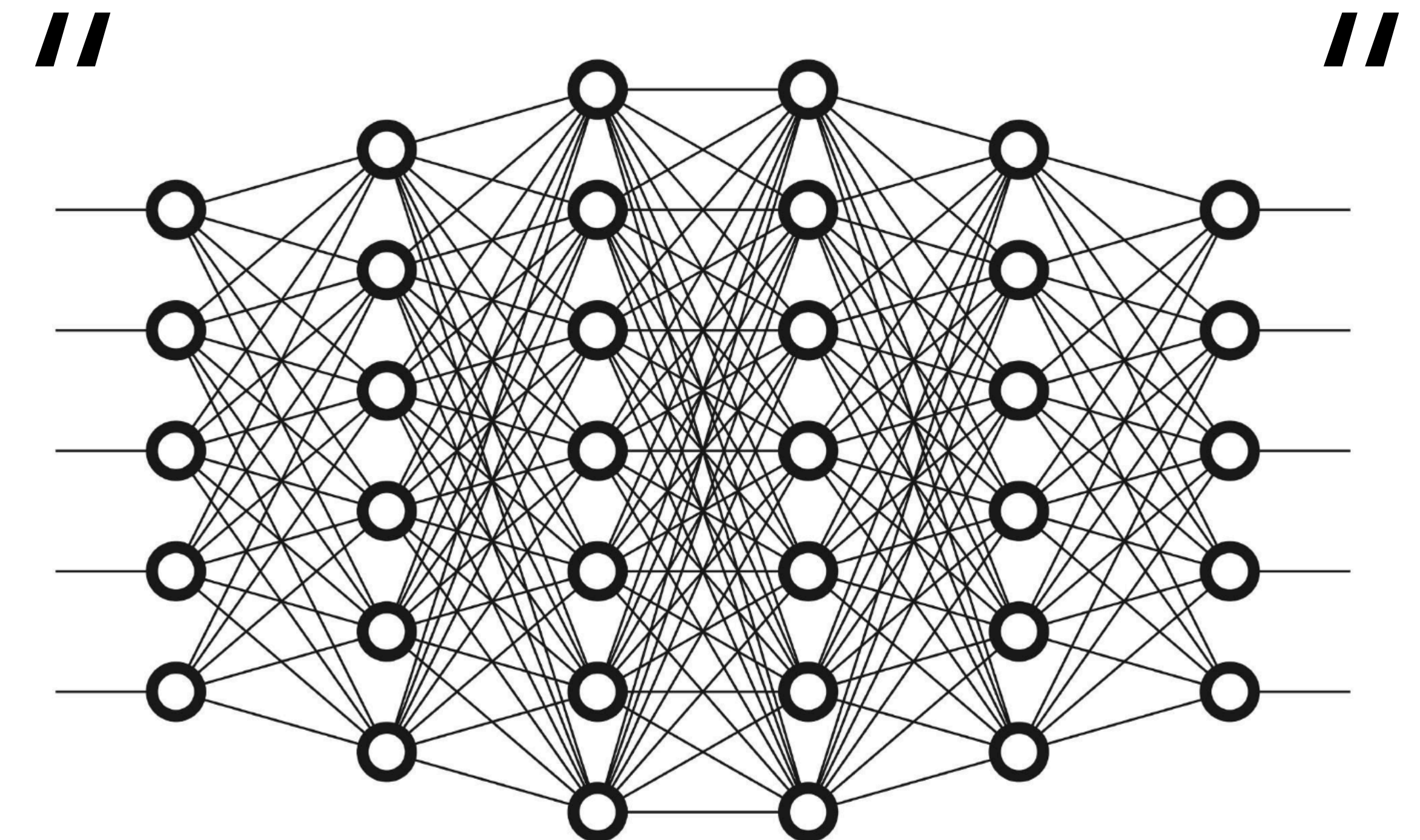
A possible (?) path



Is this a possible path that extrapolates the past to a “possible” future?

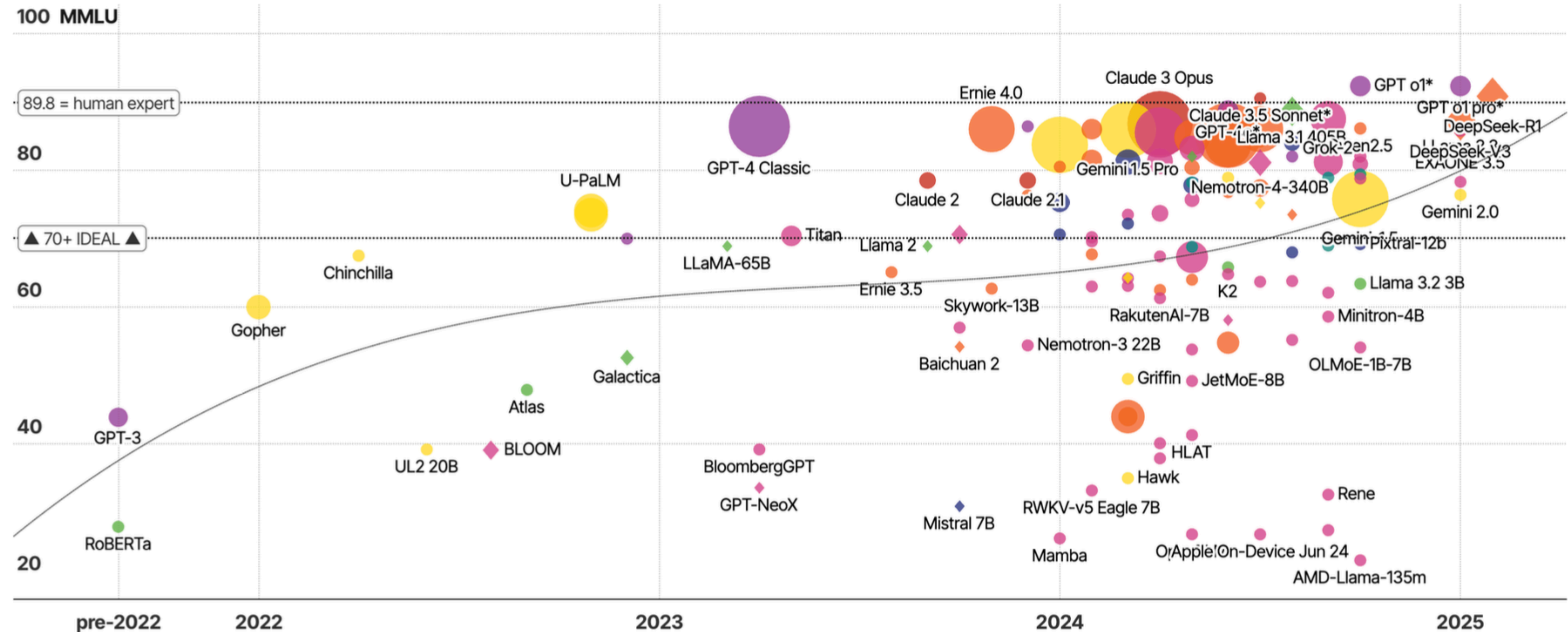
Foreseeing “**one big trainable unit**”, that just goes end to end, and we get rid of more and more of the traditional pipelines we are confident with? What if e.g. the entire physics data “analysis” pipeline becomes trainable, e.g. all the experiment code becomes an end-to-end differentiable pipeline that can be adjusted for a goal?

How will we treat data? Will a large foundation model be a black box, or will it be interpretable? If the latter, through human-in-the-loop? Will we want to talk to our data via large language models? What about performance gains? What about scientific rigour and reproducibility?



Industry, in the meantime..

Industry towards extremely large scale, multi-purpose models.. → **foundation models**



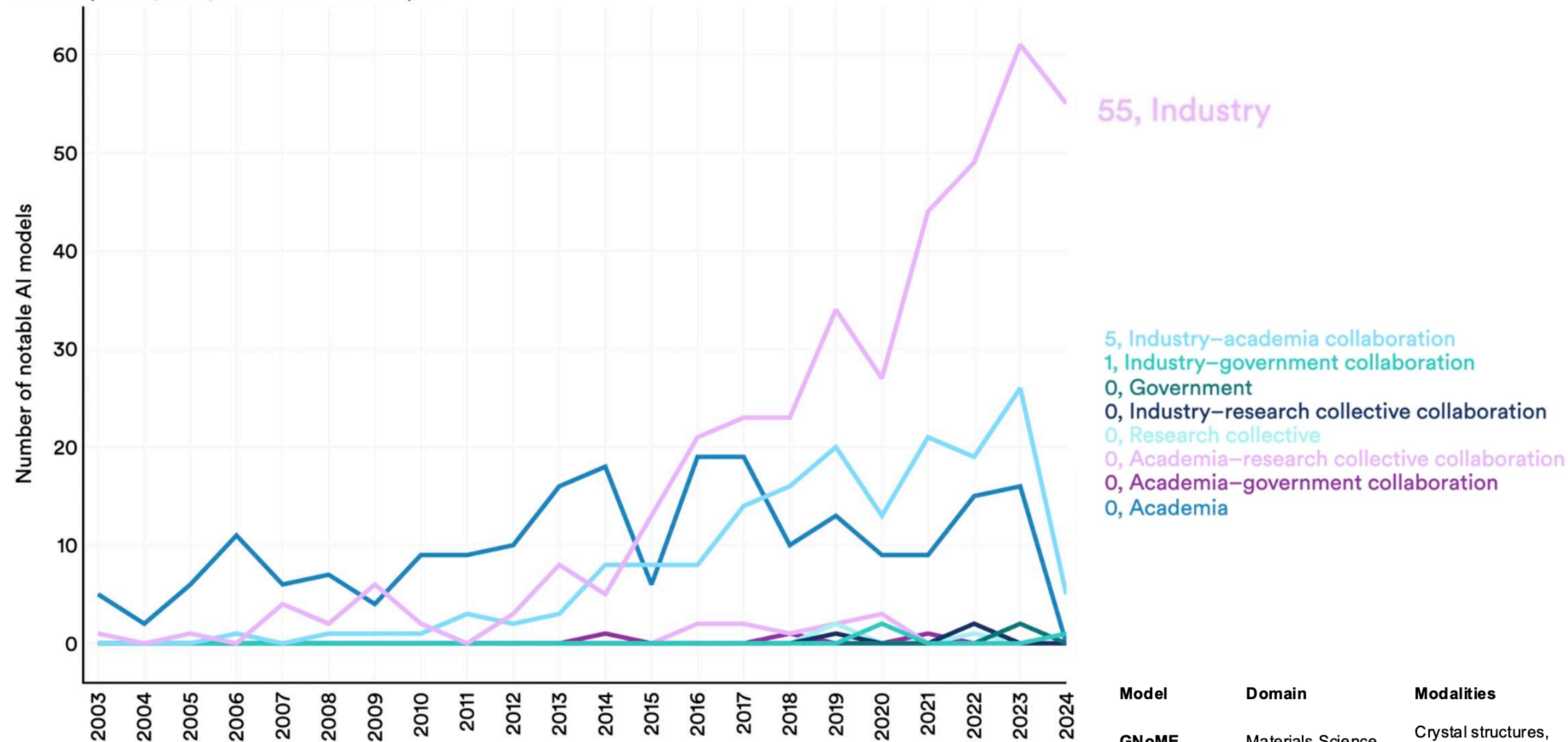
David McCandless, Tom Evans, Paul Barton
Informationisbeautiful // Jan 2024

MMLU = benchmark for measuring LLM capabilities
* = parameters undisclosed // source: [LifeArchitect](#) // [data](#)

Industry, in the meantime..

Number of notable AI models by sector, 2003–24

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



Notably, no true full-academia model (yet?)

Do you foresee particle physics to appear in this table at some time? →

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



A Large “fundamental Physics” foundation Model?

A foundation model in general:

- A large-scale ML model trained on broad and diverse data, at scale, mainly with self-supervised learning objectives, designed to be adaptable to a wide range of downstream tasks with “minimal” fine-tuning
 - ❖ In general: training on text (e.g. web, papers), audio, video, code, images, math, structured data, ..

A foundation model for fundamental physics (**LPM**)?

- Training on large and diverse datasets within a given scientific domain
 - ❖ In HEP: detector-level raw data, simulation-level data, reco-level data, analysis-level papers/plots/logbooks/docs, metadata, ..
- + transfer learning (minimal fine-tuning) + many parameters + multipurpose + some capability not explicitly included during training..

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

Kristian G. Barman^{*1}, Sascha Caron^{*2}, Emily Sullivan³, Henk W. de Regt⁴, Roberto Ruiz de Austri⁵, Mieke Boon⁶, Michael Färber⁷, Stefan Fröse⁸, Faegheh Hasibi⁹, Andreas Ipp¹⁰, Rukshak Kapoor¹¹, Gregor Kasieczka¹², Daniel Kostić¹³, Michael Krämer¹⁴, Tobias Golling¹⁵, Luis G. Lopez¹⁶, Jesus Marco¹⁷, Sydney Otten^{18,19}, Pawel Pawlowski¹, Pietro Vischia²⁰, Erik Weber¹, and Christoph Weniger²¹

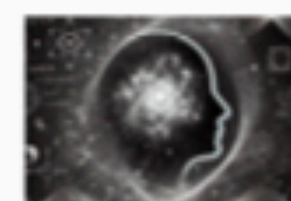
PROs

- Tailored to physics tasks and structures
- Scaling 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:

- 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

EuCAIFCon 2024 in Amsterdam



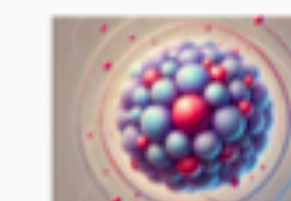
Theoretical physics

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.



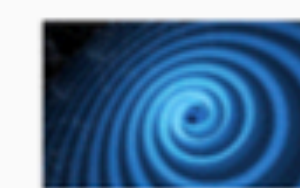
Nuclear physics

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



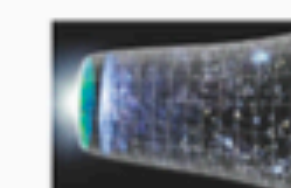
Astroparticle physics

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.



CERN colloquium, data science, 2025, Sascha Caron

European Coalition for AI in Fundamental Physics

 **Upcoming**

EuCAIFCon 2025

June 16 - 20, Sardinia

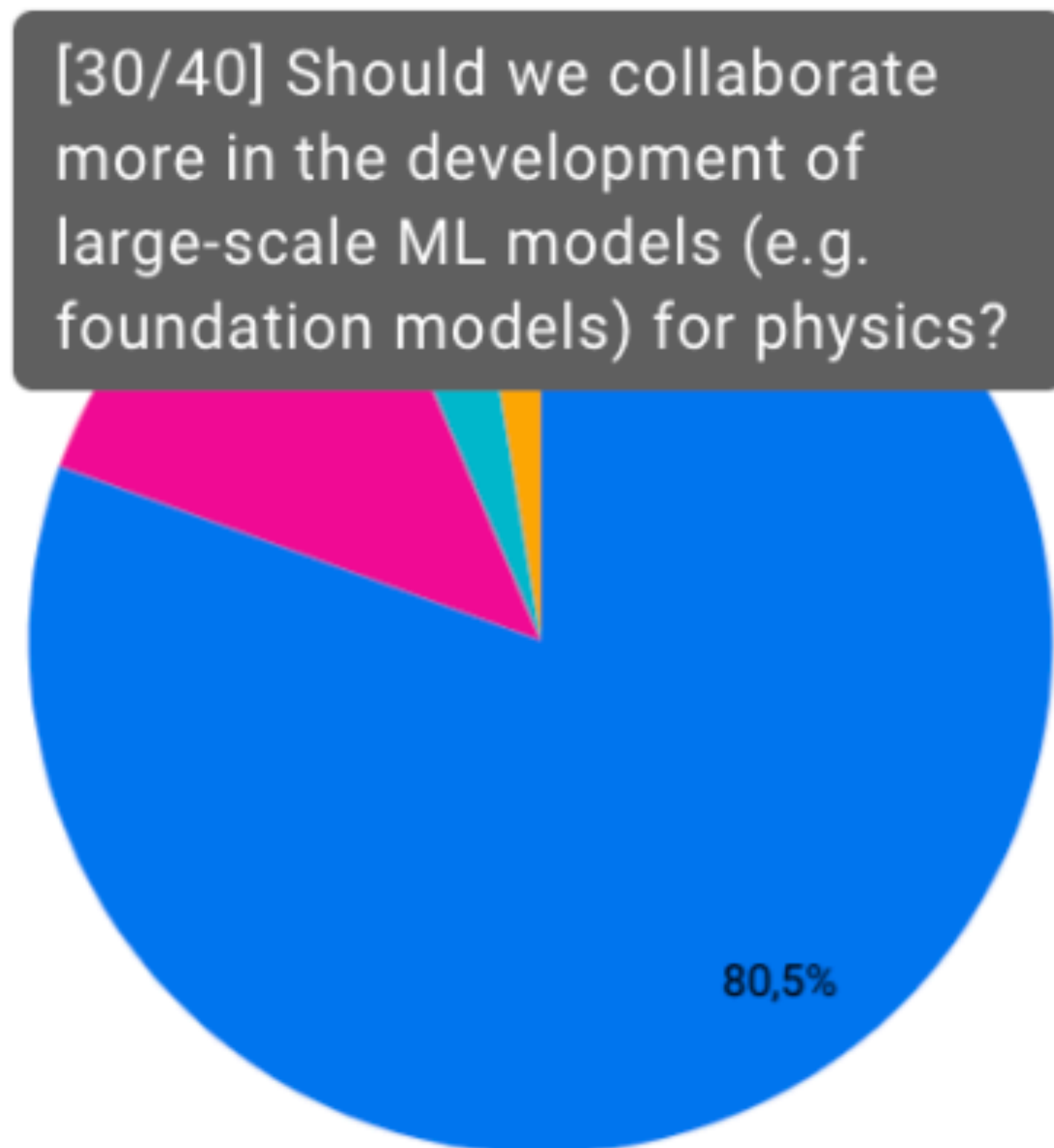
www.eucaif.org

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)

[30/40] Should we collaborate more i...



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
Karlo 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](#)

(my personal) reorganisation of the quoted recommendations

Strategic
Coordination



**Scalable AI
Infrastructure**

From R&D to
Production with
MLOps

Interdisciplinary
Collaboration

Centralised GPU-based facilities vs. federated/
hybrid HPC tailored for AI workloads

Invest in shared data platforms, and tools to
support distributed training and inference

Scientific LLMs &
Foundation Models

Training and Industry
Collaboration

Ensure long-term sustainability and collaboration
through structured, community-driven initiatives

FAIR Principles &
Open Science

Sustainable AI
Practices

Benchmarks &
Reusability



(my personal) reorganisation of the quoted recommendations



Strategic
Coordination

Scalable AI
Infrastructure



**From R&D to
Production with
MLOps**

Interdisciplinary
Collaboration

Mature AI prototypes must get funding and support to become production-ready tools

Establish dedicated MLOps personnel dedicated to model deployment, reproducibility, and integration into large-scale physics workflows

Scientific LLMs &
Foundation Models

Training and Industry
Collaboration

Promote standardisation of tools, APIs, and lifecycle management across the community

Benchmarks &
Reusability

FAIR Principles &
Open Science

Sustainable AI
Practices



(my personal) reorganisation of the quoted recommendations



Strategic
Coordination

Scalable AI
Infrastructure

From R&D to
Production with
MLOps

Interdisciplinary
Collaboration

Invest in science-focused large language models.

Train foundation models using real and synthetic physics data, incorporating domain knowledge



**Scientific LLMs &
Foundation Models**

Training and Industry
Collaboration

Govern the use of general-purpose commercial models w.r.t alternatives that prioritise transparency, automation, and interpretability in model development and evaluation

Benchmarks &
Reusability

FAIR Principles &
Open Science

Sustainable AI
Practices



(my personal) reorganisation of the quoted recommendations



Strategic
Coordination

Scalable AI
Infrastructure

From R&D to
Production with
MLOps

Interdisciplinary
Collaboration

Create standardised, extensible benchmarks for
key physics ML tasks (e.g., classification,
tracking, ..)

Scientific LLMs &
Foundation Models

Training and Industry
Collaboration

Encourage community-wide model sharing to
enhance reproducibility and innovation



**Benchmarks &
Reusability**

FAIR Principles &
Open Science

Sustainable AI
Practices



(my personal) reorganisation of the quoted recommendations



Strategic
Coordination

Scalable AI
Infrastructure

From R&D to
Production with
MLOps

Interdisciplinary
Collaboration

From red-AI to green(er)-AI

Promote community awareness of carbon footprint and a culture of monitoring and mitigating environmental impact of large models

Scientific LLMs &
Foundation Models

Training and Industry
Collaboration

Coordinate with HPC centres towards use of energy-efficient hardware and scheduling, and encourage optimisation of software frameworks and energy use in training

Benchmarks &
Reusability

FAIR Principles &
Open Science



**Sustainable AI
Practices**



(my personal) reorganisation of the quoted recommendations



Strategic
Coordination

Scalable AI
Infrastructure

From R&D to
Production with
MLOps

Interdisciplinary
Collaboration

Integrate FAIR (Findable, Accessible, Interoperable, Reusable) principles as standards into publishing and research practices, and develop tools to simplify their adoption

Scientific LLMs &
Foundation Models

Training and Industry
Collaboration

Incentivise FAIR-compliant work in funding decisions and career advancement, promote open science and data sharing through training and policy alignment.



**FAIR Principles
& Open Science**

Sustainable AI
Practices

Benchmarks &
Reusability



(my personal) reorganisation of the quoted recommendations



Strategic
Coordination

Scalable AI
Infrastructure

From R&D to
Production with
MLOps

Interdisciplinary
Collaboration

Ramp up in hands-on AI training courses, summer schools, etc tailored to physicists and researchers, both newcomers and upskilling, that cover topics beyond “traditional”, i.e. include in curricula reproducibility, FAIR principles, MLOps, AI infrastructures, etc

Scientific LLMs &
Foundation Models



**Training and Industry
Collaboration**

Partner with industry for co-hosted events, internships, specific trainings, even tool development

Benchmarks &
Reusability

FAIR Principles &
Open Science

Sustainable AI
Practices



(my personal) reorganisation of the quoted recommendations



Strategic
Coordination

Scalable AI
Infrastructure

From R&D to
Production with
MLOps



**Interdisciplinary
Collaboration**

Launch goal-oriented collaborative efforts bringing together physicists, ML researchers, HPC experts, engineers and foster mutual learning.

Create research programs involving physicists, AI experts, HPC engineers, etc.

Enable cross-domain innovation via shared platforms and open source.

Scientific LLMs &
Foundation Models

Training and Industry
Collaboration

Benchmarks &
Reusability

FAIR Principles &
Open Science

Sustainable AI
Practices



(my personal) reorganisation of the quoted recommendations



Strategic Coordination

Scalable AI Infrastructure

From R&D to Production with MLOps

Interdisciplinary Collaboration

Form a permanent structure to coordinate AI strategy in fundamental physics.

Use initiatives like EuCAIF as governance models?

Scientific LLMs & Foundation Models

Training and Industry Collaboration

Benchmarks & Reusability

FAIR Principles & Open Science

Sustainable AI Practices