Input on AI

Gregor Kasieczka, Stefano Giagu

ESPP Meeting, Venice
June 2025

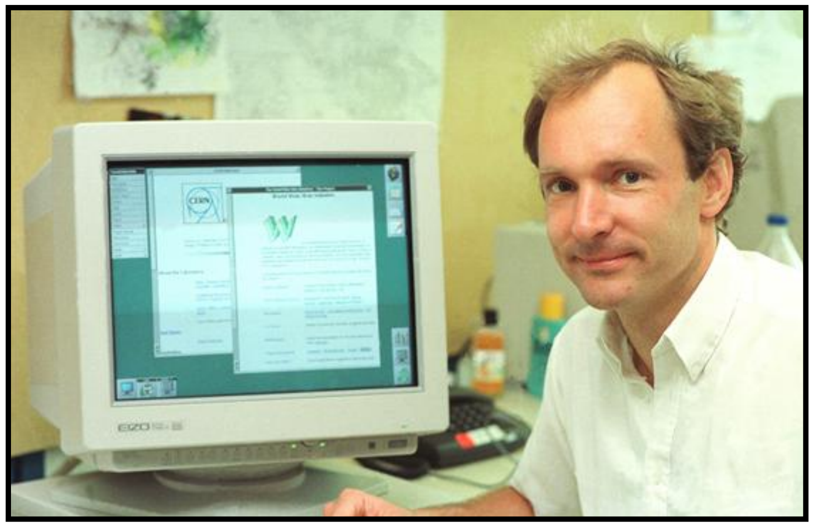
Universität Hamburg
DER FORSCHUNG | DER LEHRE | DER BILDUNG

CLUSTER OF EXCELLENCE
QUANTUM UNIVERSE

SAPIENZA
UNIVERSITÀ DI ROMA

Disclaimer: Thankful for input by everyone. Bias and opinions in this synthesis are my own.

Development



Past

WorldWideWeb: Proposal for a HyperText Project, Tim-Berners Lee et al **1990**

35 years ⟷

Present

Explosion of large-language models (LLMS) **2025**

35 years ⟷

Future

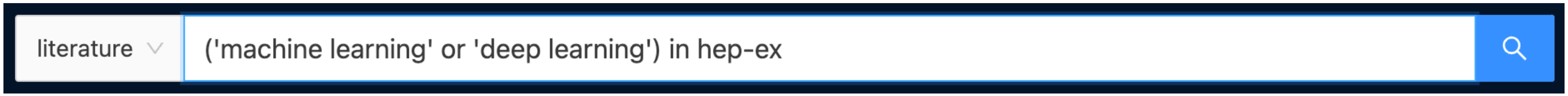
???, **2060**

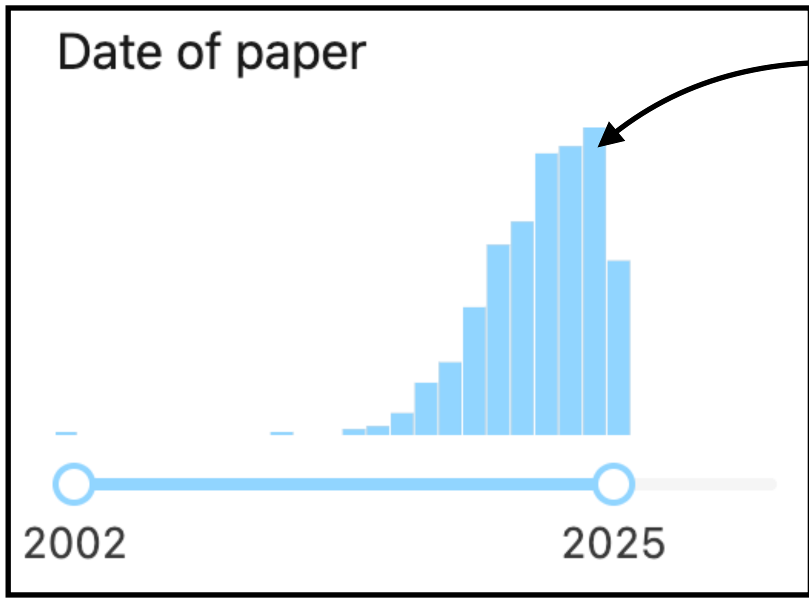
Cannot be too ambitious for the physics possiblities from developments in AI

CERN did play a role in large-scale development beyond HEP

CERN-GE-9407011-31; ChatGPT; Iron Man 2

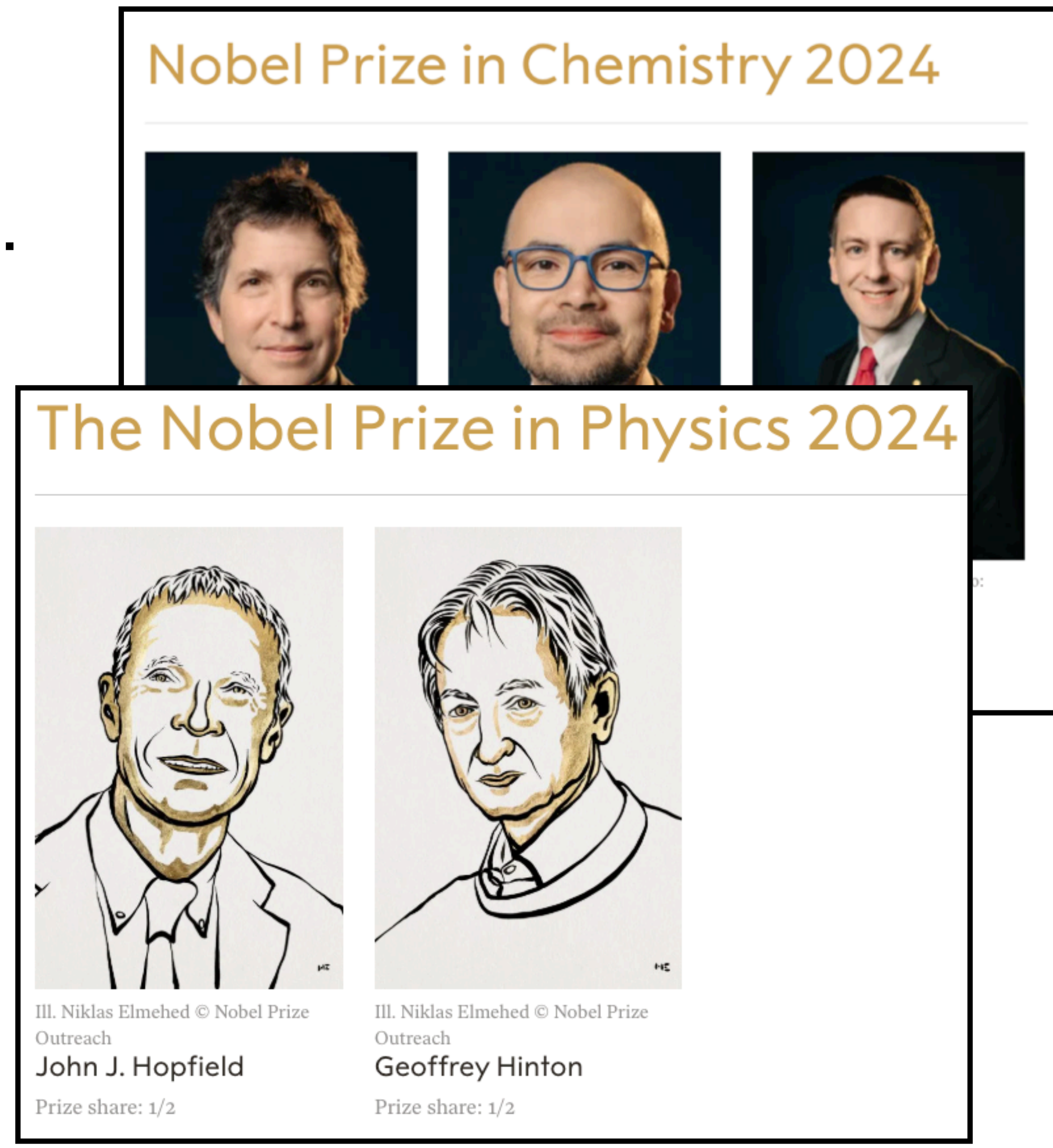
State of AI in Physics





200+ papers in 2024

Similar for nucl-ex, astro-ph, hep-ph, …



Rapid rise of AI in fundamental physics

Transition from concepts to applications

Recognition of AI work as Nobel-prize worthy

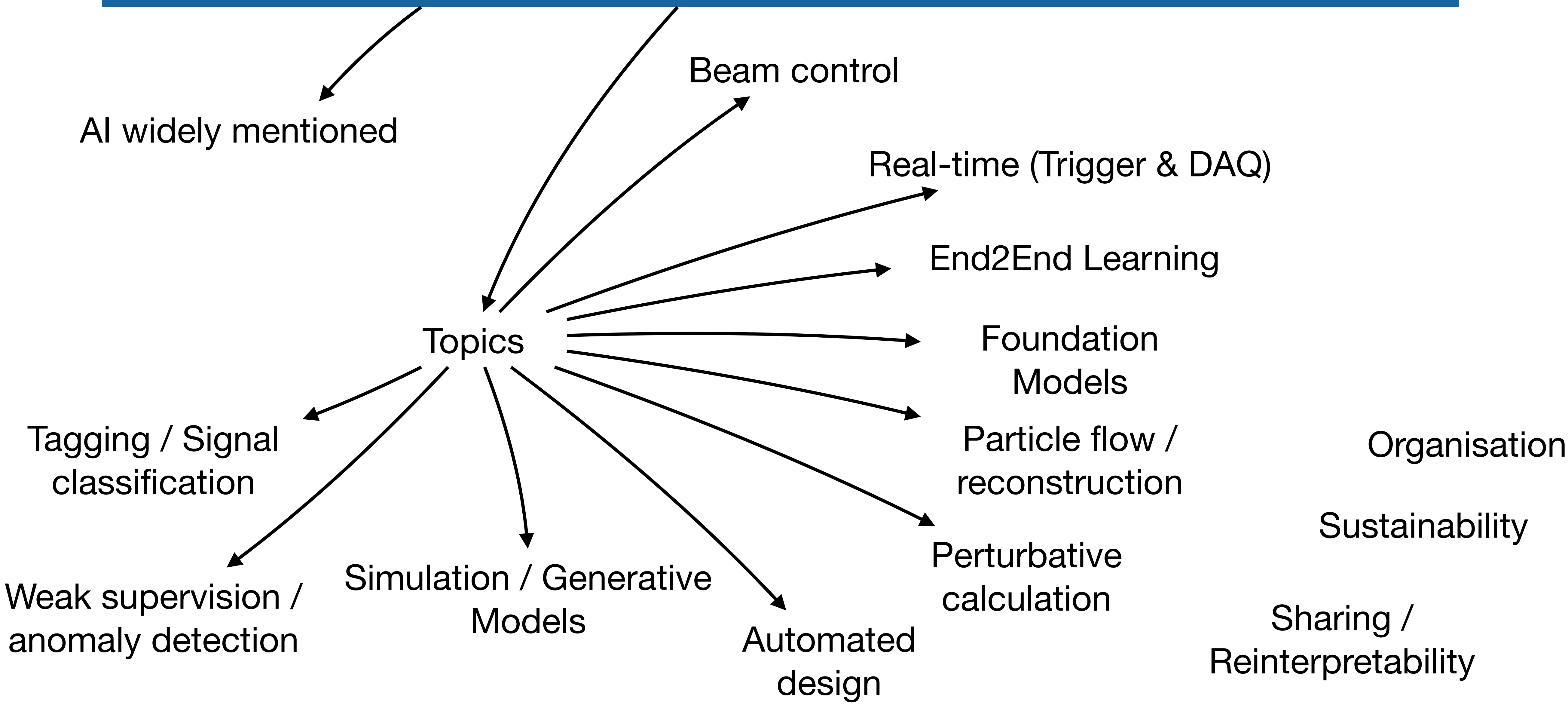
See also https://iml-wg.github.io/HEPML-LivingReview/

Input to the European Strategy for Particle Physics - 2026 update

European Strategy Update

AI widely mentioned

Input to the European Strategy for Particle Physics - 2026 update



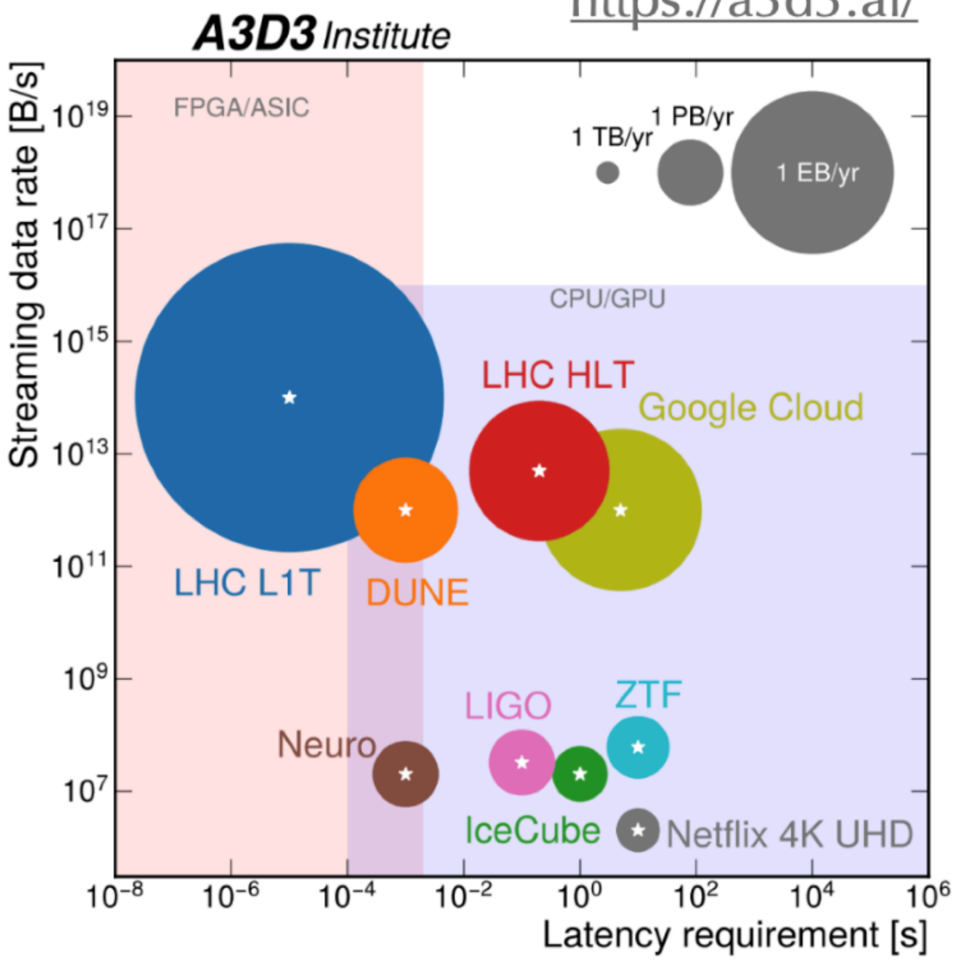
Physics Informed

Benchmarks

Will discuss key developments & strategy ahead

Data Taking

https://a3d3.ai/



Collider physics offers combination of data rate and latency requirements

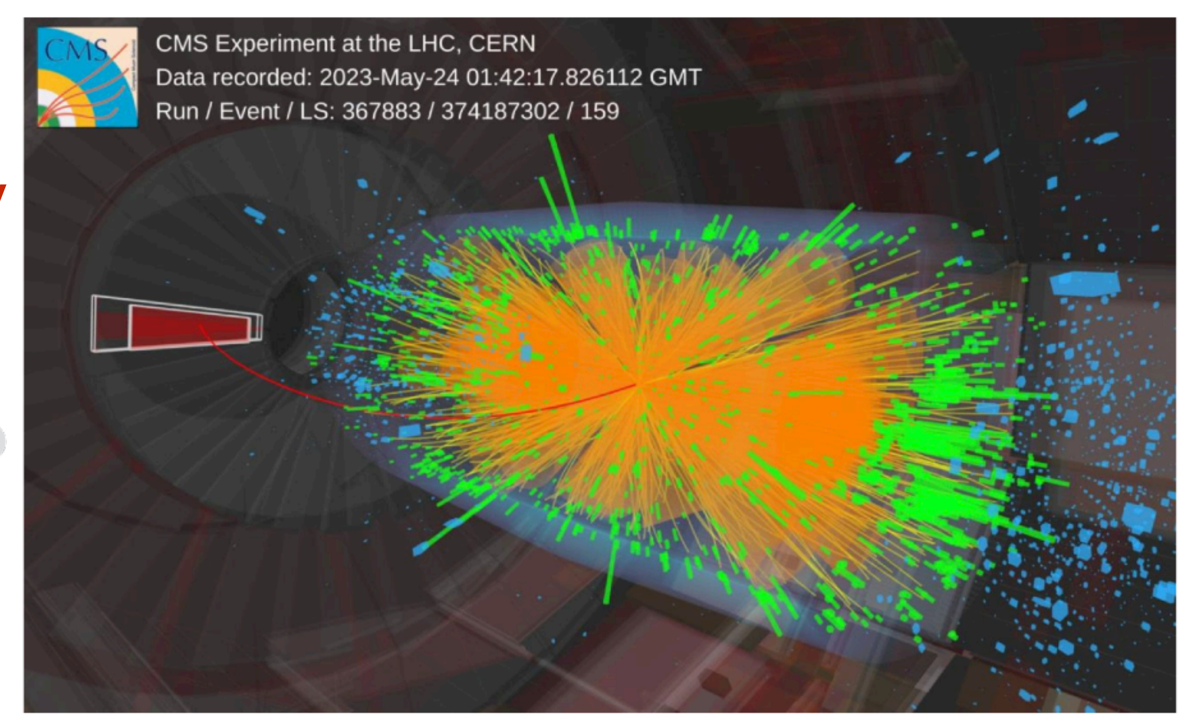
Relevant for data reduction and read-out

Need efficient ML on FPGA hardware for current and future experiments

More in Thea's talk

Already now: Exploring anomaly detection via autoencoders in CMS L1

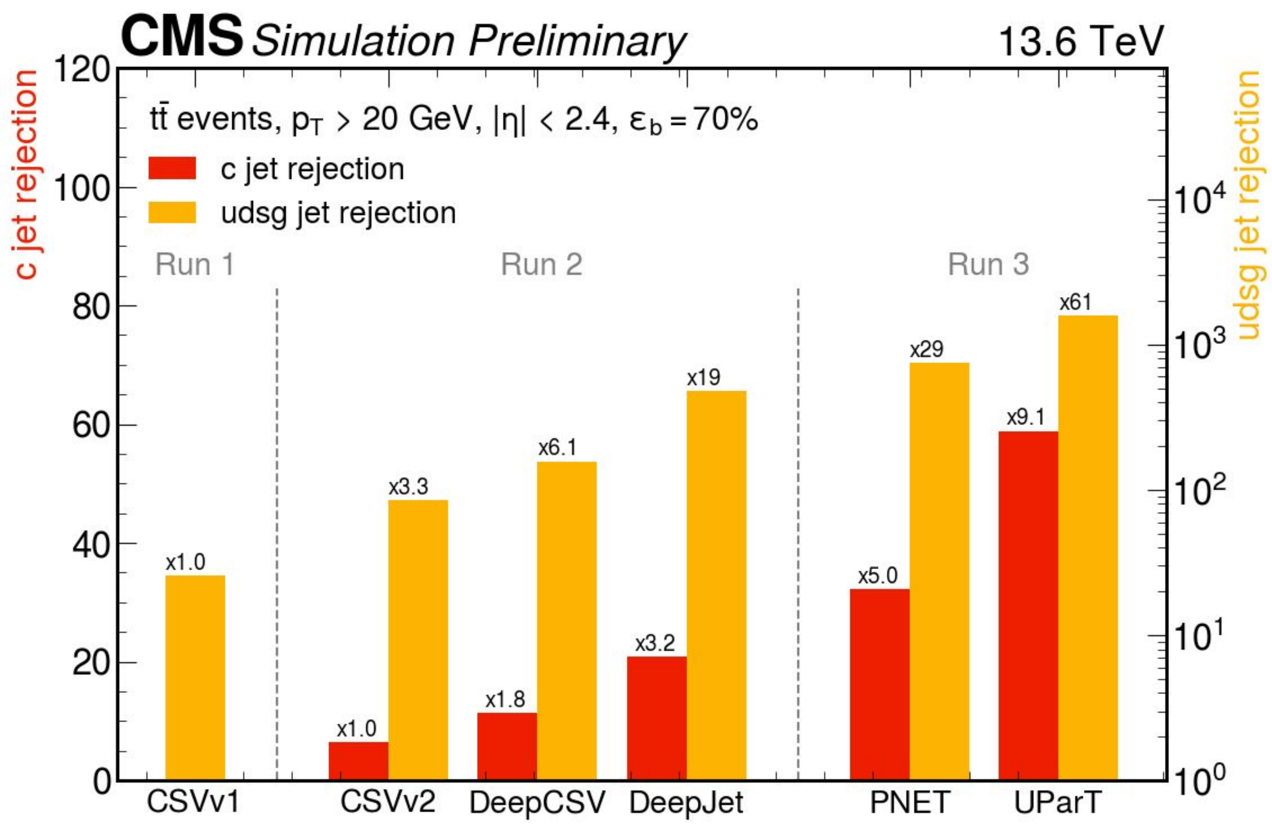
AXOL1TL hls4ml



CMS-DP-2024-059

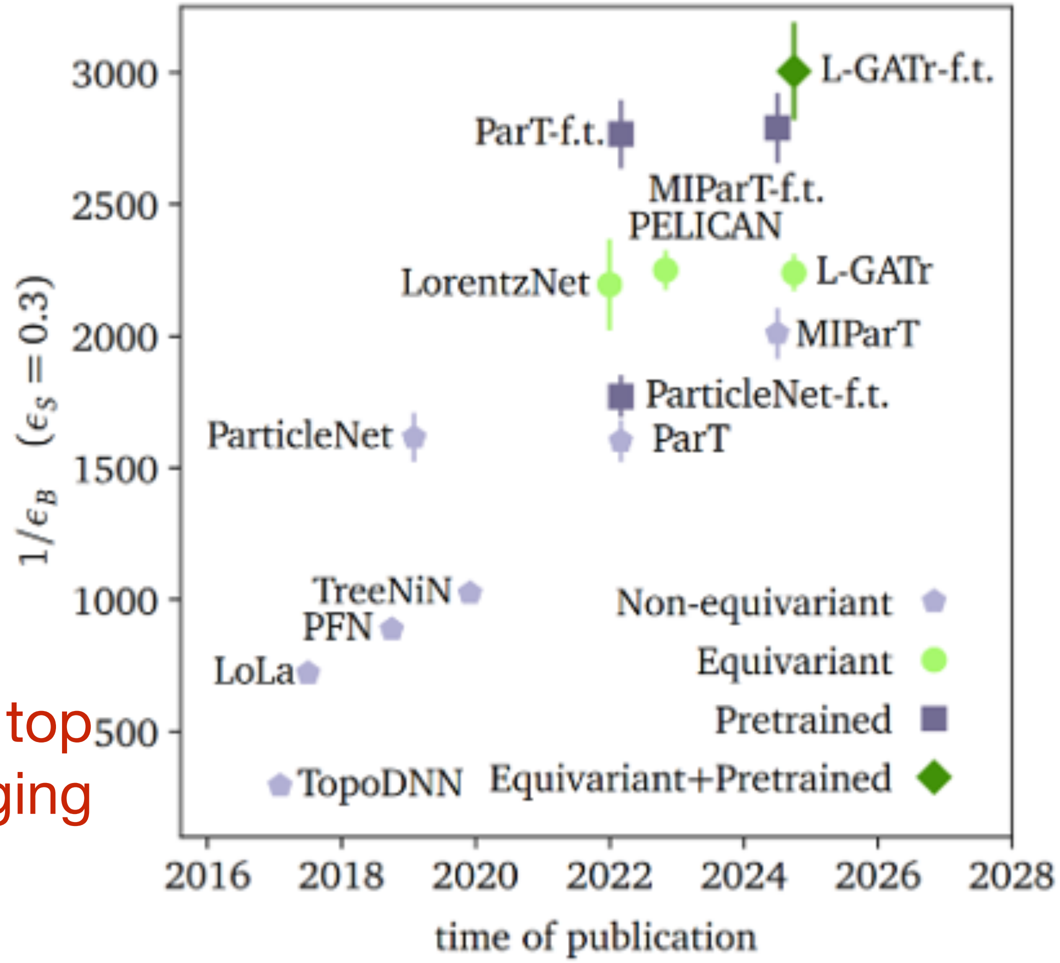
Tagging

- Determine source of a jet/event/..
- Wide-spread application of classification
- Flavour/charm tagging relevant Higgs factories



Flavour tagging

Boosted top quark tagging

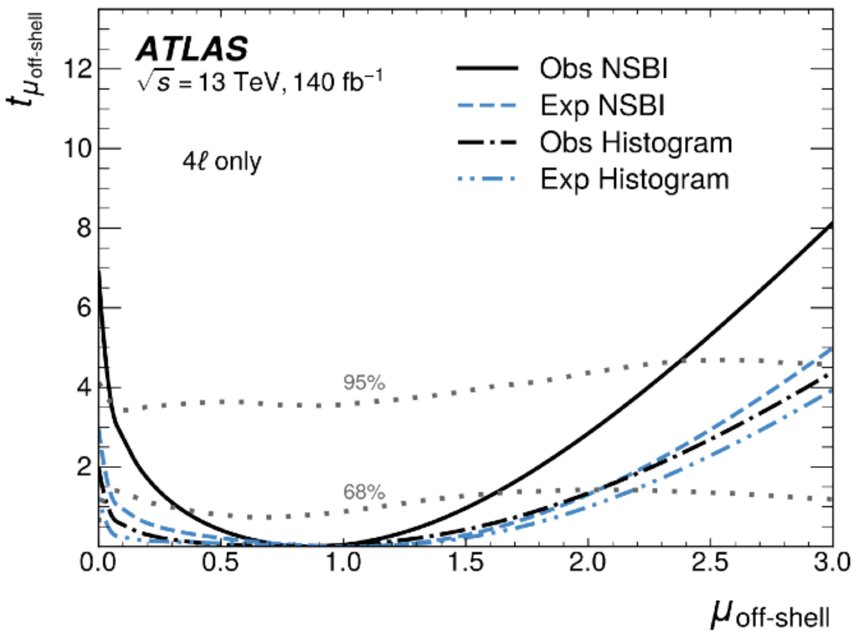


Orders of magnitude improvement in key physics application from better AI

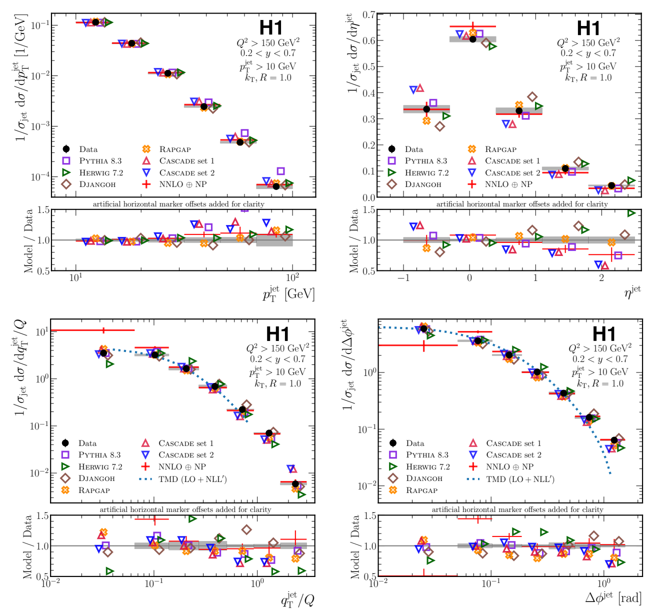
Direct pipeline from AI development to deployment

DP2024_066, 1902.09914, V Breso at ML4Jets, ..

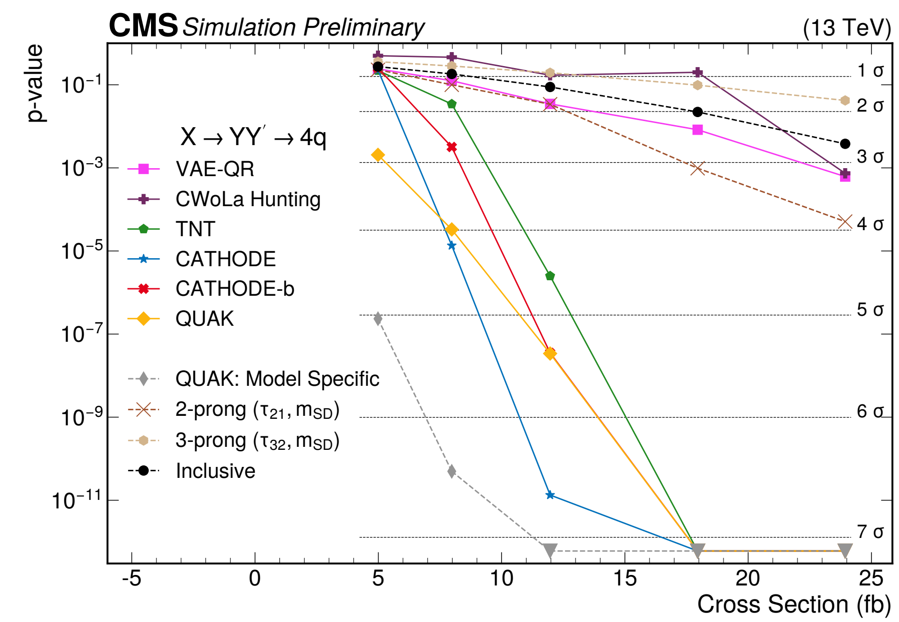
Analysis Paradigms



Simulation Based Inference



Unfolding



Anomaly detection

Unbinned high-dim shape information in statistical analysis improves sensitivity

Unfold in higher dimensions via reweighting and morphing

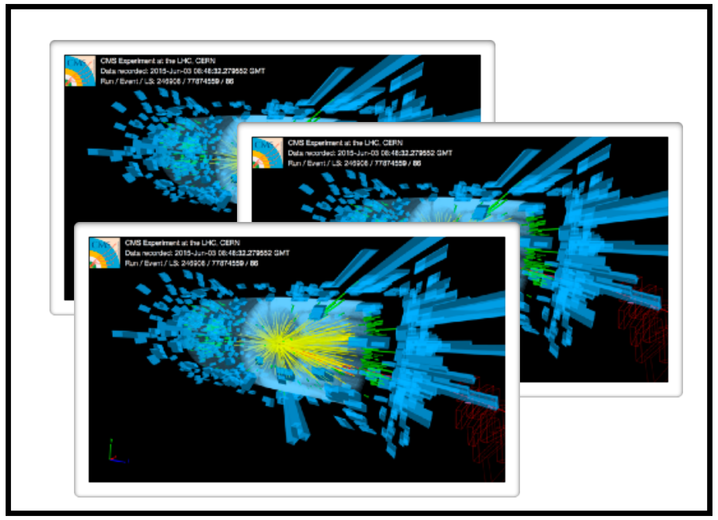
New searches with reduced model dependency

Substantial improvements and qualitatively new approaches done already **now**

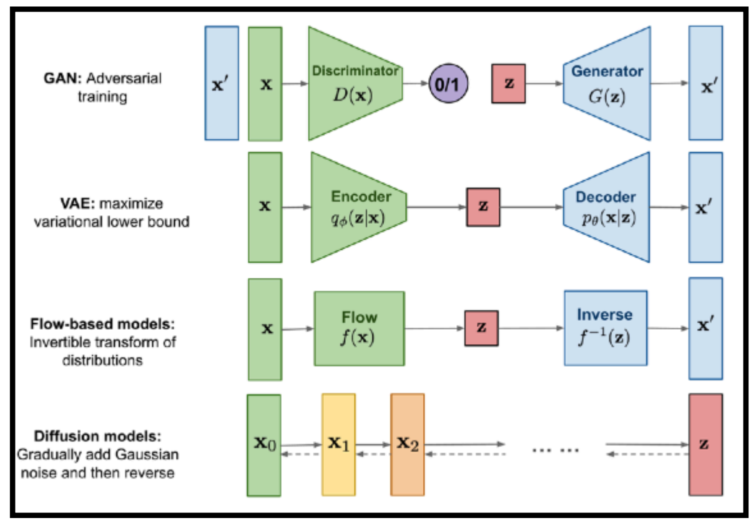
1911.09107; 2108.12376; 2412.01600, 2412.01548; 2412.03747

Simulation

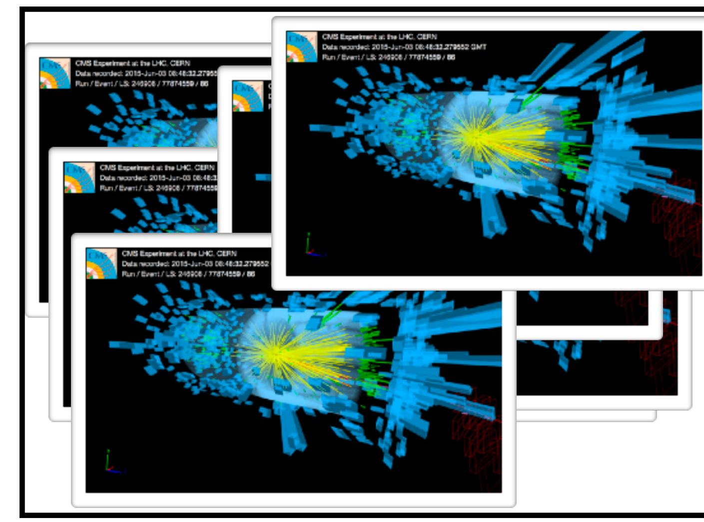
1. Simulation or collider data as input



2. Train generative surrogate



3. Sample



Fast Calorimeter Simulation Challenge 2022

Hardware	Simulator	NFE	Batch Size	Time / Shower [ms]	Speed-up
CPU	Geant4			3914.80 ± 74.09	×1
	CaloClouds	100	1	3146.71 ± 31.66	×1.2
	CaloClouds II	25	1	651.68 ± 4.21	×6.0
	CaloClouds II (CM)	1	1	84.35 ± 0.22	×46
GPU	CaloClouds	100	64	24.91 ± 0.72	×157
	CaloClouds II	25	64	6.12 ± 0.13	×640
	CaloClouds II (CM)	1	64	2.09 ± 0.13	×1873

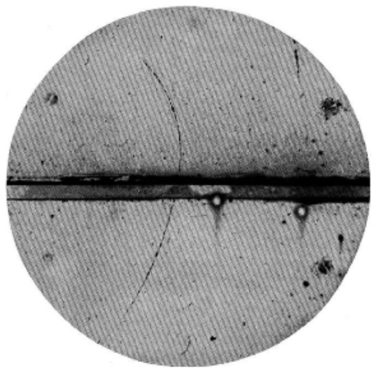
1712.10321, 2109.02551, 2309.05704, 2410.21611, …

Broad effort to speed-up / bypass simulation at detector (e.g. calorimeters) and theory (event generators MadNIS/ Sherpa/..) level

~2 orders of magnitude speed-up on same hardware achieved

Essential for simulation needs of next generation experiments

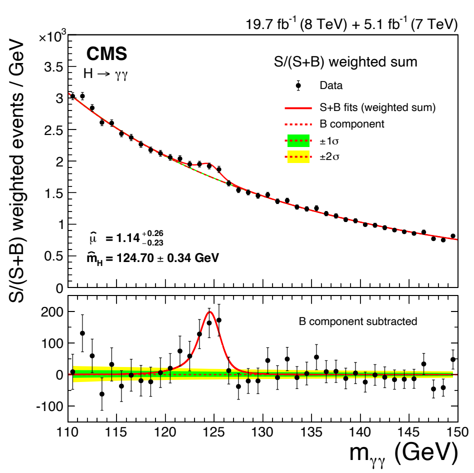
End to end



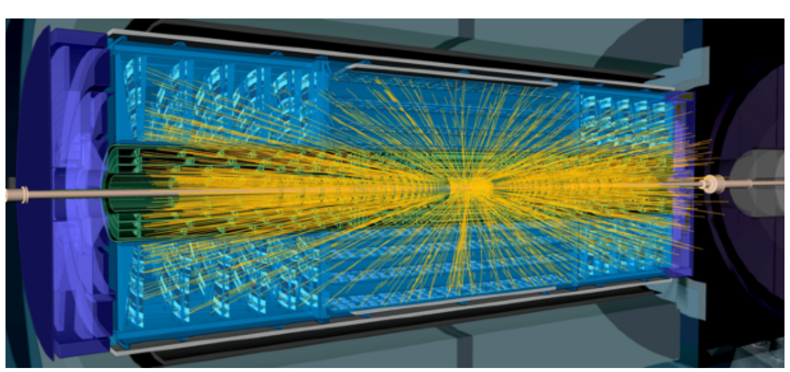
1930s — Positron discovery: Single event

Channel:	SVX
observed	27 tags
expected background	6.7 ± 2.1
background probability	2×10^{-5}

1990s — Top discovery: Multiple events, cut & count



2010s — Higgs discovery: Histogram & shape information



20??s — X discovery: Holistic processing of low-level information

Separately optimising

…detector calibrations, particle flow objects, jet reconstruction, pile-up supression, object tagging, background rejection, signal selection…

loses information at each step

Future analysis will be event-level and end-to-end for optimal sensitivity

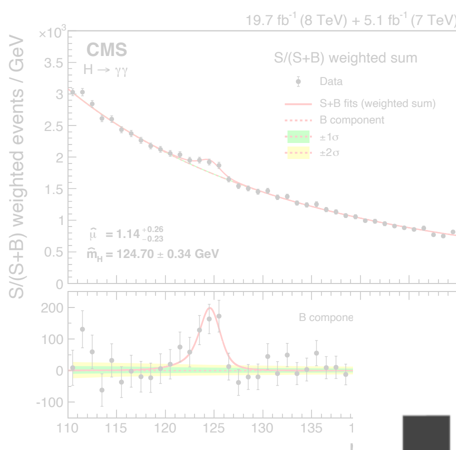
Jets and other objects for human consumption only

End to end

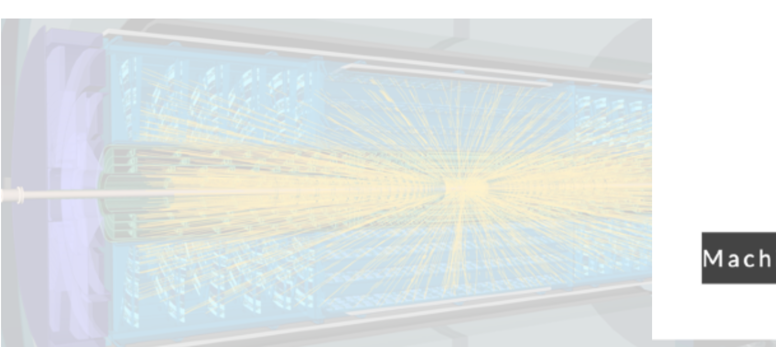
1930s — Positron discovery: Single event

Channel:	SVX
observed	27 tags
expected background	6.7 ± 2.1
background probability	2×10^{-5}

1990s — Top discovery: Multiple events, cut & count



2010s — Higgs discovery: Histogram & shape information



MODE

Machine-learning Optimized Design of Experiments

Separately optimising

…detector calibrations, particle flow objects, jet reconstruction, pile-up supression, object tagging, background rejection, signal selection…

loses information at each step

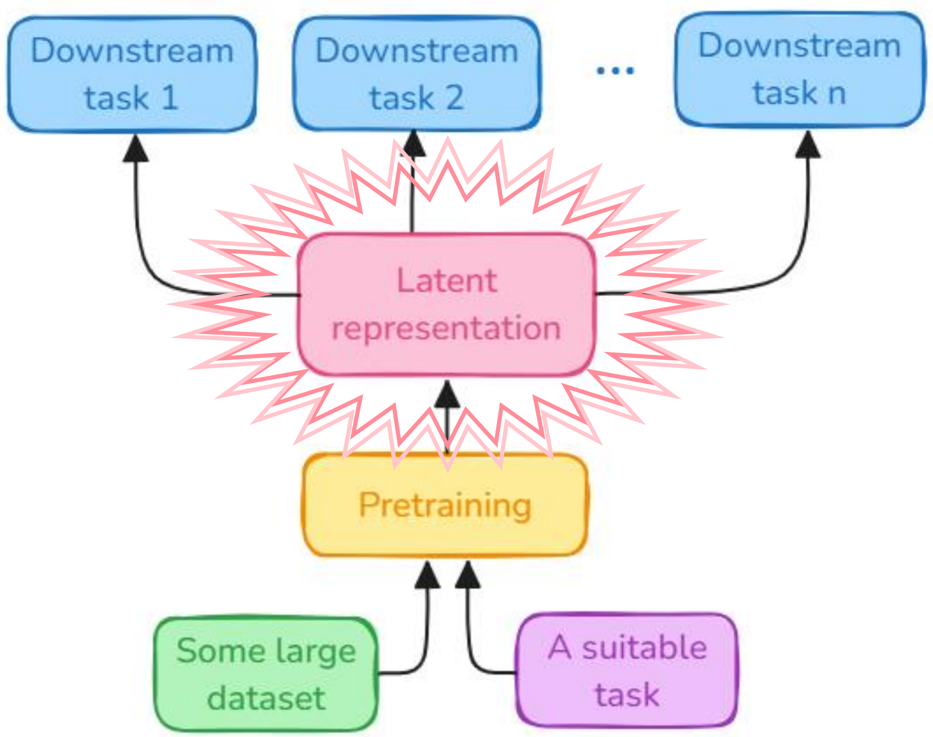
Future analysis will be event-wise and end-to-end for optimal sensitivity

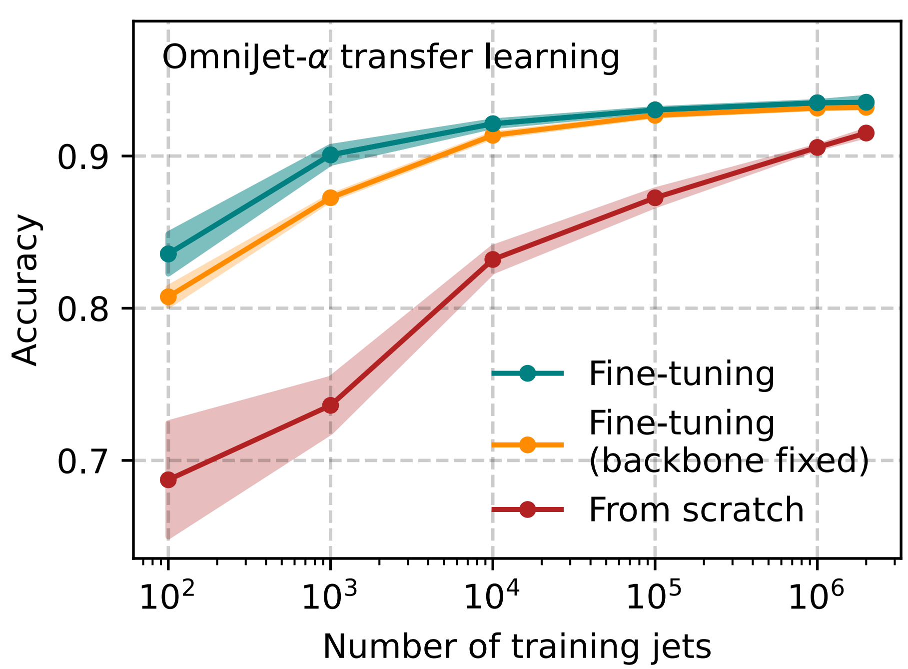
Jets and other objects for human consumption only

One step further: Differentiable programming and surrogates allow end-to-end optimsation of detector design

https://mode-collaboration.github.io/; 2203.13818; 2204.01681, ..

Foundation models





Similarity across many AI problems in fundamental physics

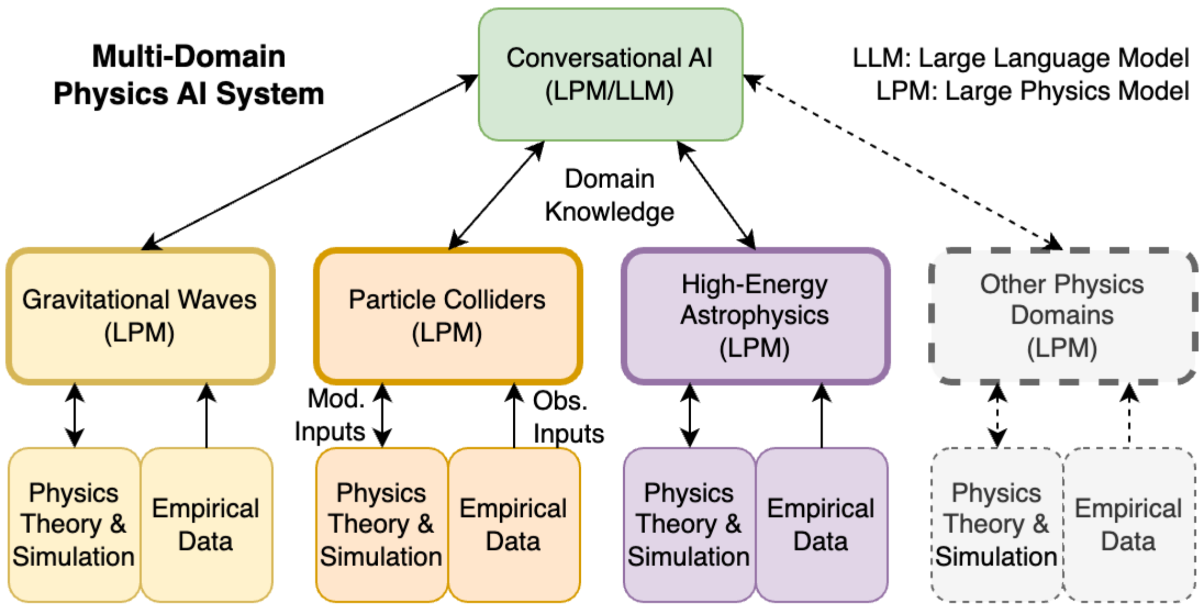
Build multi-dataset multi-task foundation models that can be adapted via fine-tuning to new problems

Learn underlying joint representation

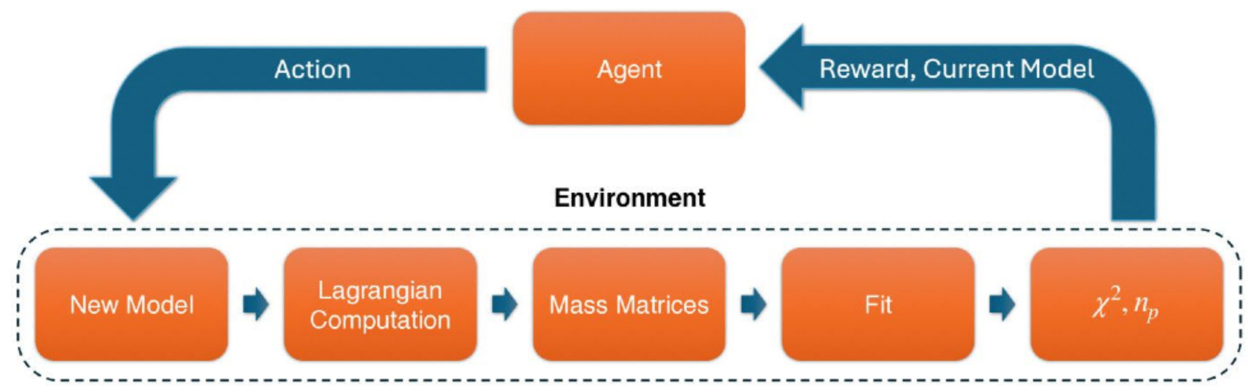
Improve:
performance (*higher max. accuracy)*
data efficiency (*lower number events needed)*
scientific turn-around (*lower time to result)*

2403.05618, … ; e.g. A Hallin's overview at EUCAIFCon '25

Large-physics models



Language models as glue between domain experts



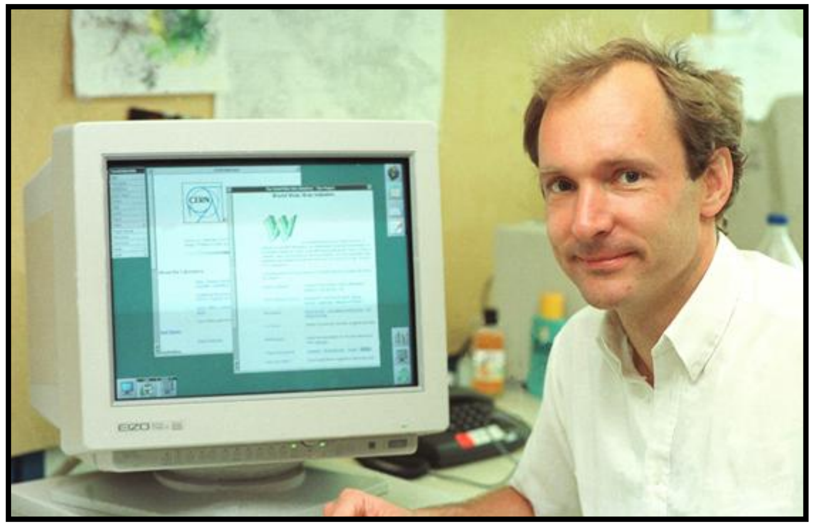
Agent-based hypothesis generation & iteration

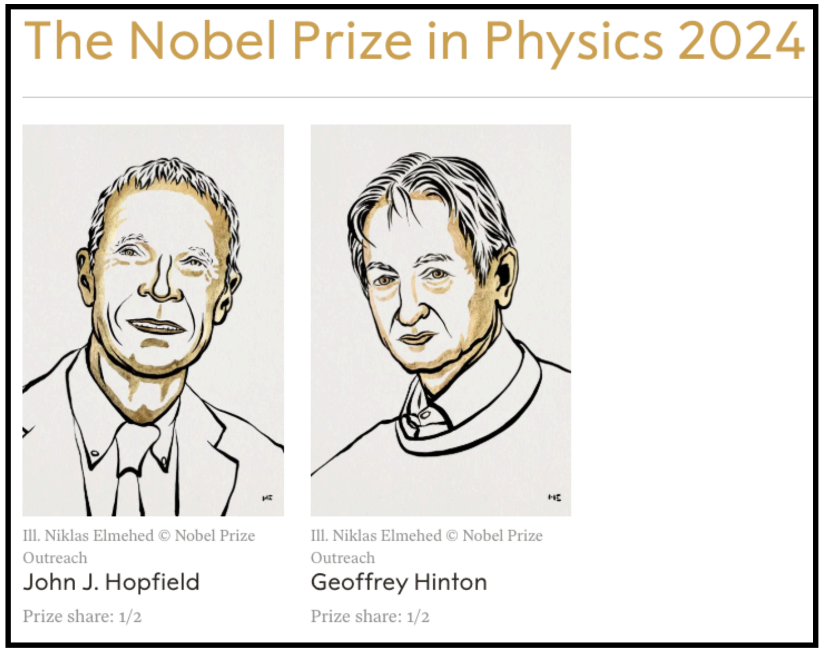
Advances in large language models (LLMs) allow a new degree of autonomous operation

Leverage for information systems, accelerator control, automated analysis design, …

2501.05382, 2506.08080, 2506.14757

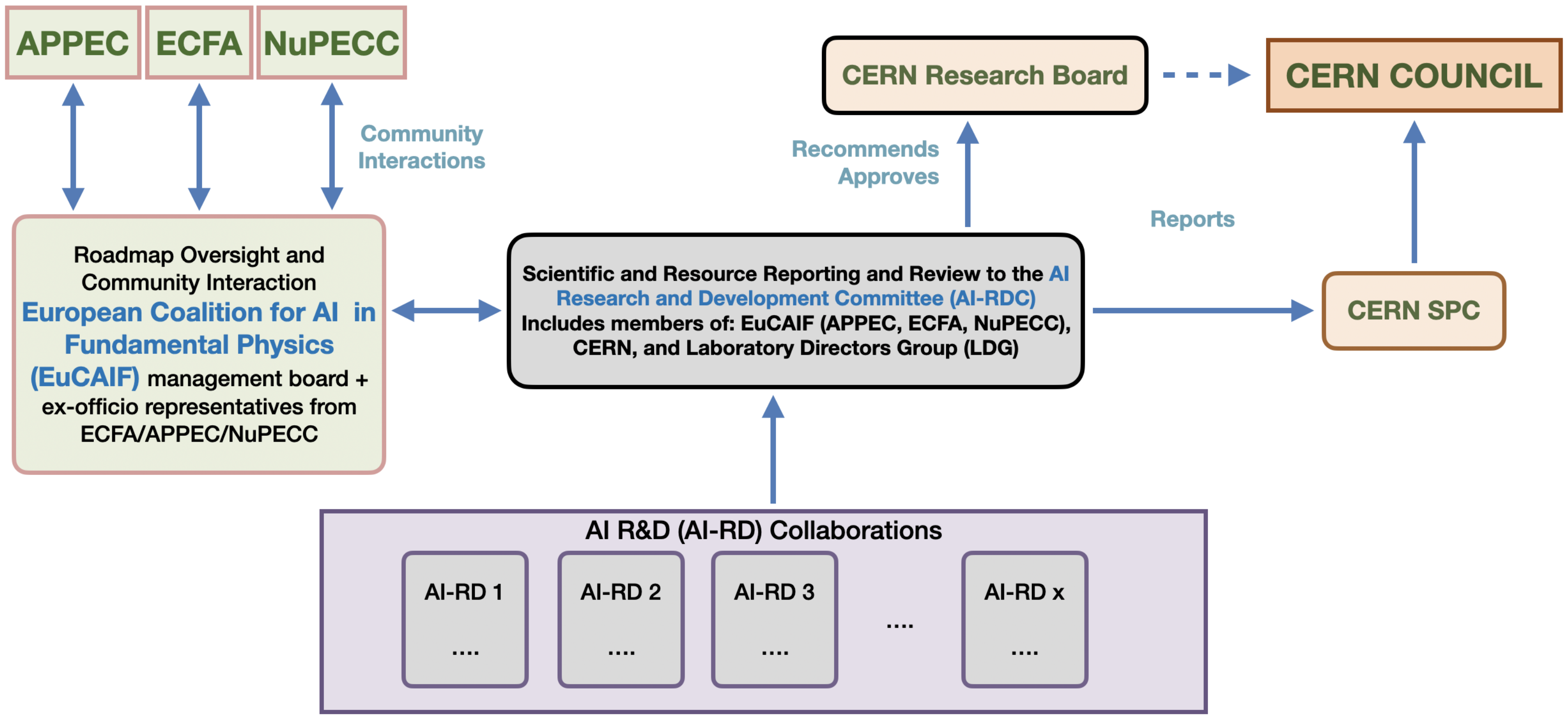
Physics for AI





- Problems from fundamental physics historically driving force for informatics
- Chance to regain with future collider strategic push
- Topics of:
  real-time challenge
  multi-source uncertainty quantification
  ultra-rare signals
  complex detector data
  …

How to get there?



- Form and coordinate AI R&D collaborations
  - Organise research
  - Central provider of tools
  - Training and career development
- AI-RDs as nucleus for a future flagship AI effort

EuCAIF

Examples:
AI for Data Processing
AI for Detector and Accelerator Control
AI for Detector Optimization
AI for Event Reconstruction
AI for Analysis
AI for Generation and Simulation
AI for Theory
AI for HPC Usage
AI for Documentation and Education

See also CERN: https://edms.cern.ch/ui/file/3234243/2/RCS_AI_Program-v2-2.pdf →Expect workshop in September; https://eucaif.org/; JENA recommendations: 2503.14192

Environment

Federal Ministry
of Research, Technology
and Space

AI and basic research in the natural sciences

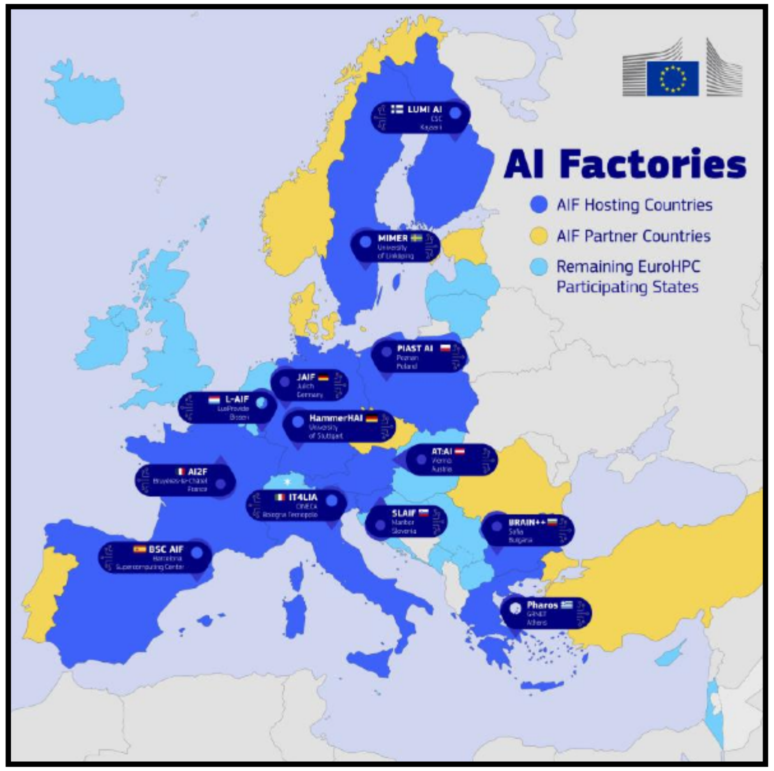
Artificial intelligence (AI) is of central importance for basic research in the context of the framework programme "Exploration of the Universe and Matter – ErUM".[1] It opens up entirely new dimensions of research in terms of speed, precision, and scalability, thereby enabling scientific questions to arise that were previously unthinkable. At the same time, it is indispensable for handling unprecedented amounts of data, such as those generated by large-scale research infrastructures. The data-intensive nature of basic research in the natural sciences plays a key role in driving the development of new AI algorithms and would not be possible without AI methods. Scientific requirements for data analysis and statistics are fostering the creation of innovative methods, particularly in the field of explainable AI, where results must be traceable and interpretable.

At the same time, AI systems greatly benefit from the high-quality, non-personal datasets provided by ErUM disciplines—such as those from particle collisions or sky surveys—which offer ideal training resources. These models can detect patterns even in low-resolution data, a capability that is also highly relevant for fields like cancer diagnostics. The close interplay between fundamental research and AI development not only advances science but also yields domain-specific AI solutions that are particularly relevant for industrial applications.

To fully leverage the potential of AI for basic research in the natural sciences in Europe, the following priorities are essential:

- AI funding should specifically promote interdisciplinary approaches and foster the exchange of solutions across disciplines to accelerate cross-learning and unlock synergies (for example, funding initiatives from Big Data to Smart Data in ErUM-Data or Mathematics for Innovations). Particularly promising are versatile AI approaches with the potential to accelerate multiple research domains simultaneously.
- Furthermore, initiatives focused on networking (e.g. EuCAIF at European level, or ErUM-Data-Hub at national level) and on developing explainable AI should be strengthened. Establishing structured and reliable pathways for the transfer of research results into commercial applications is also crucial and should be strategically addressed.

German input to European AI strategy highlights particles collisions



Build conditions for future "AI Factory" at CERN

Link to future European large-scale AI initiatives

Politics expects leadership on AI in science from us
→Benefit from CERN for AI discussion

Opportunity from possible future collider project as first AI first machine

See also e.g. input by CERN: https://ec.europa.eu/info/law/better-regulation/have-your-say/initiatives/14547-A-European-Strategy-for-AI-in-science-paving-the-way-for-a-European-AI-research-council_en

Conclusions

- Substiantial responses on AI
- Breakthroughs in AI directly translate to better physics
  - Qualitative: Powerful new analysis paradigms
  - Quantitative: Multiple order-of-magnitude improvements
- Chance to build future generation experiments AI first
  - End-to-end designed for sensitivity
  - Consider in planning physics reach
  - Be open to continued innovation
- Boosts overall future collider effort
  - Our community is expected to lead in AI
  - Attract innovations and AI funding
- Start now with substantial effort
  - Form AI R&D collaborations
  - Use HL-LHC as testbed

Thank you!