AI/ML for Collider Physics

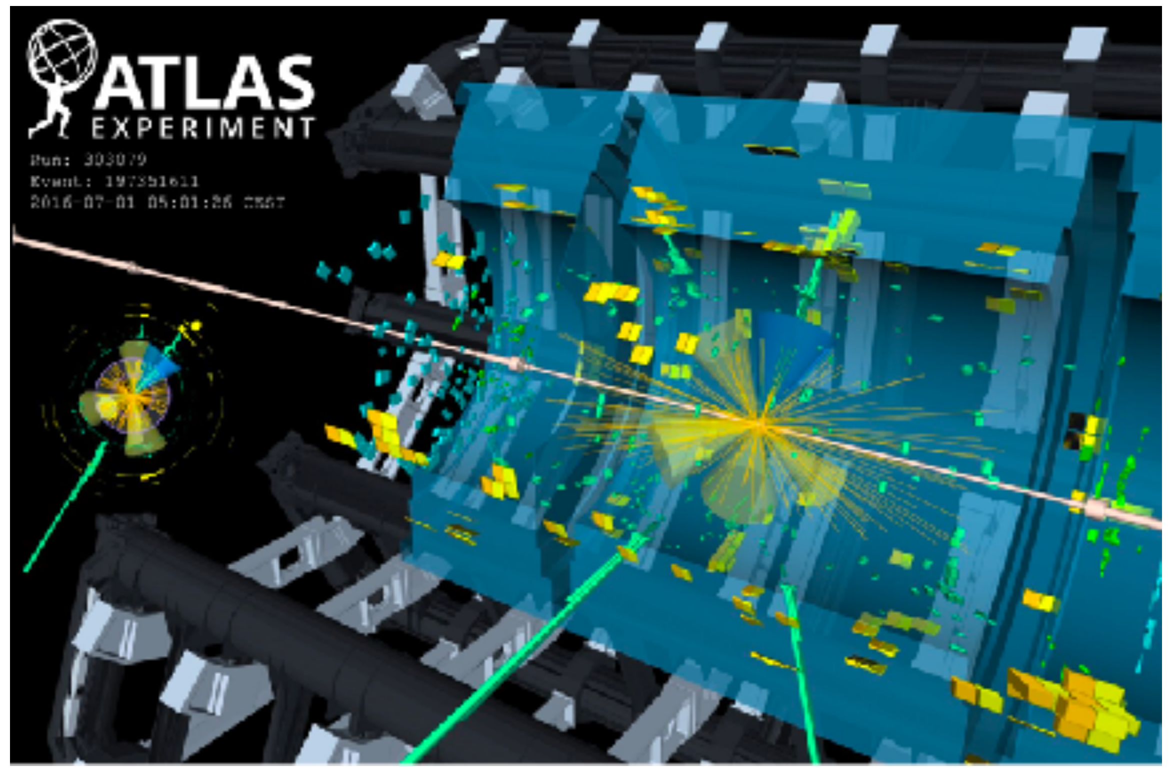
Lukas Heinrich, Bormio 2026


New directions in science are launched by new tools much more often than by new concepts.
- Freeman Dyson

The HEP Data Challenge

The fact that we can measure anything is, to me, remarkable

Data



O(100M) dimensional,
O(10m), O(100eV)

$x \longleftrightarrow \theta$

Theory

$$\mathcal{L} = -\frac{1}{4}F_{\mu\nu}F^{\mu\nu} + i\bar{\psi}\not{D}\psi + h.c. + \bar{\psi}_i y_{ij}\psi_j\phi + h.c. + |D_\mu\phi|^2 - V(\phi)$$

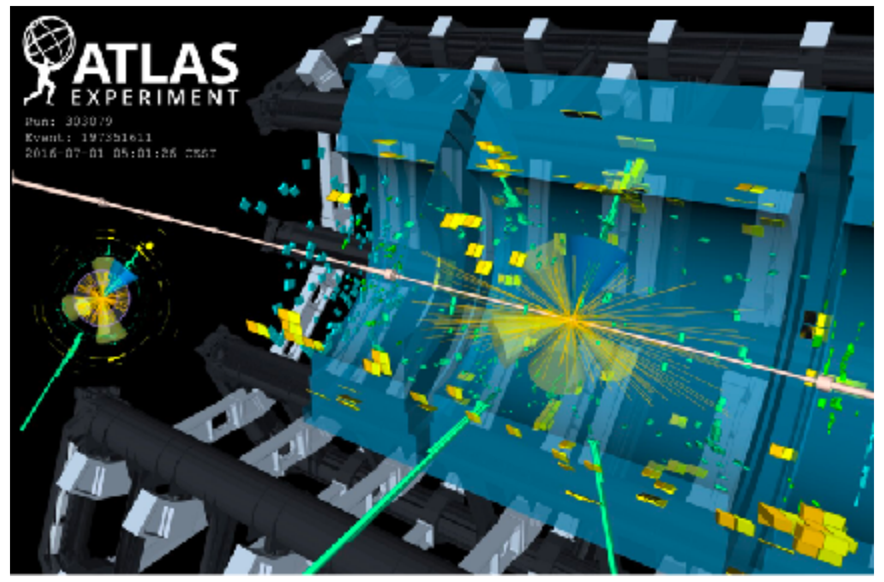
O(10) dimensional,
O(10^{-18} m), O(100 GeV)

$$p(\theta|x) = \frac{\boxed{p(x|\theta)}}{p(x)}p(\theta)$$

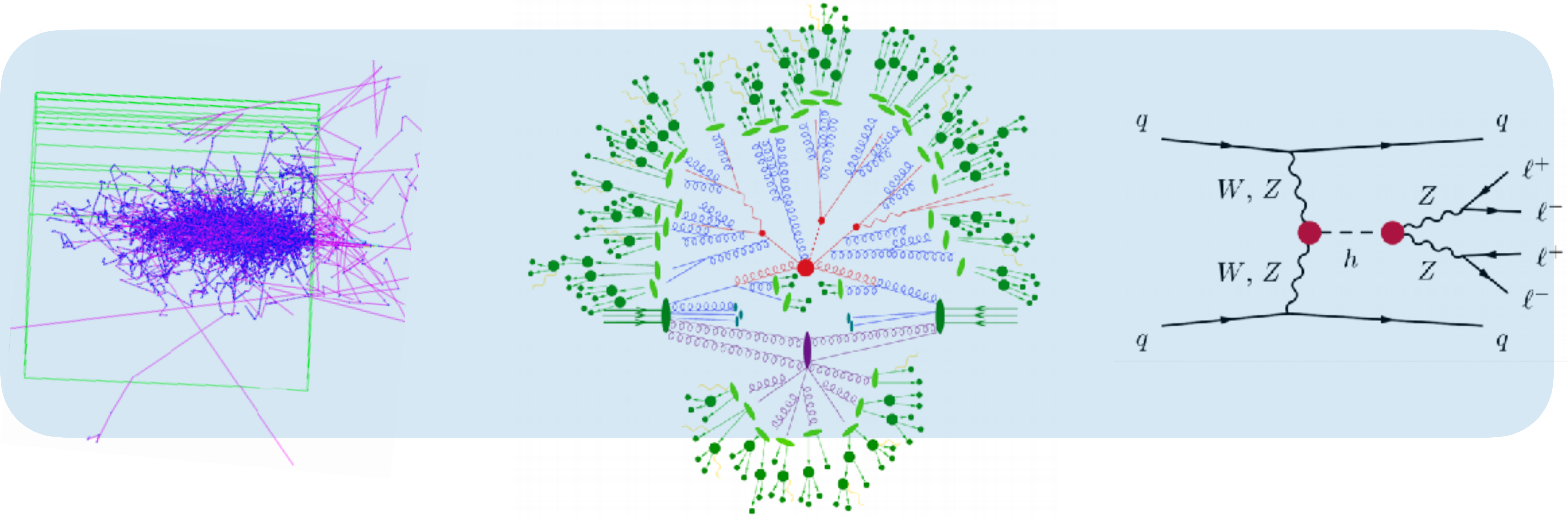
$$p(\Lambda_{\mathrm{BSM}}|x) \qquad p(\theta_{\mathrm{Higgs}}|x)$$

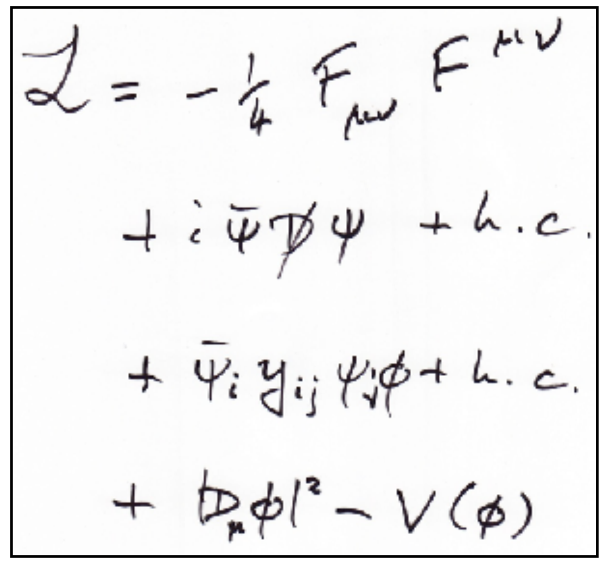
Bridging the Gap

Our current approach is a **triumph of domain knowledge**
→ we know a lot about how to go from QFT to Voltages



x



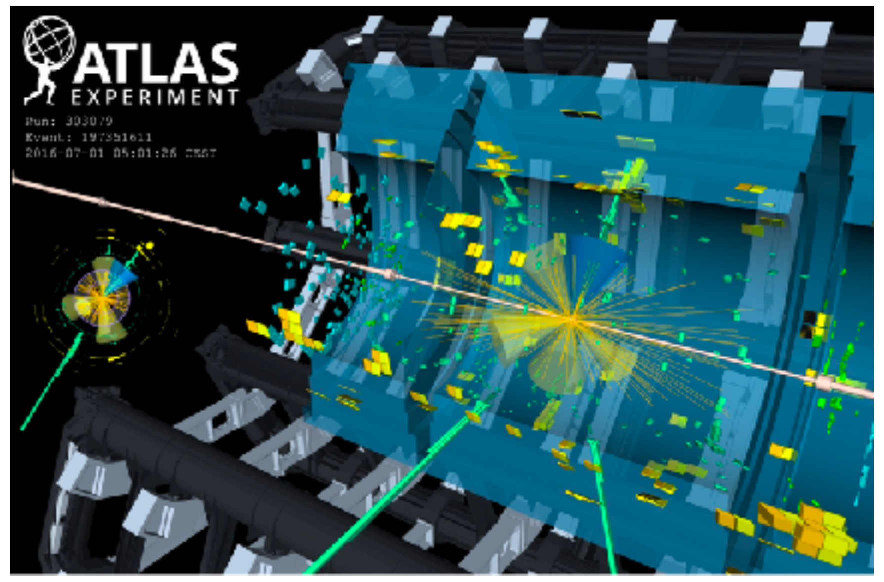


θ

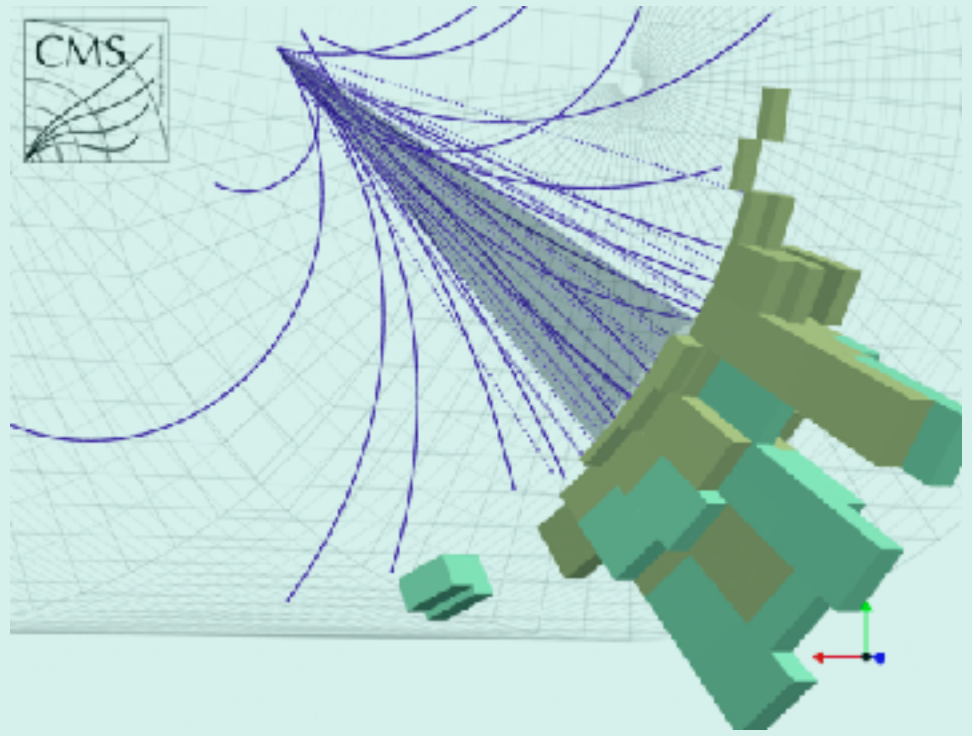
data-generating process: $x \sim p(x \,|\, \theta) p(\theta)$

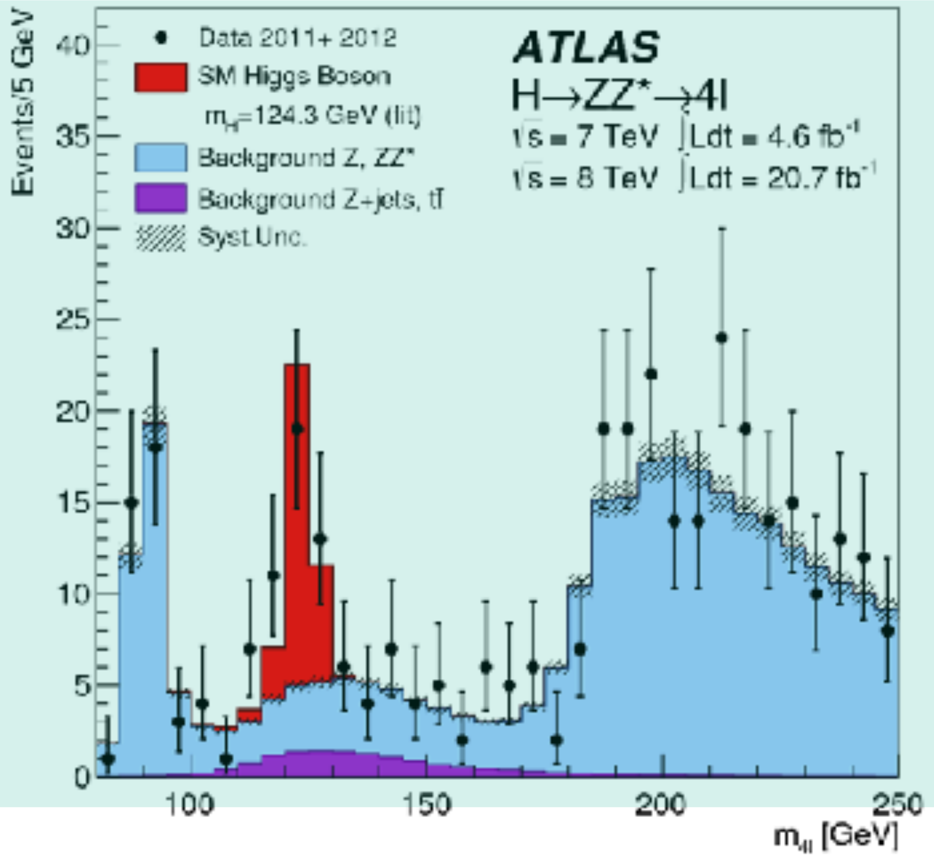
Bridging the Gap

Our current approach is a **triumph of domain knowledge**
→ we use it to "go backwards" from Voltages to Lagrangian



x





$$\mathcal{L} = -\frac{1}{4} F_{\mu\nu} F^{\mu\nu} + i\bar{\psi} \not{D} \psi + h.c. + \bar{\psi}_i y_{ij} \psi_j \phi + h.c. + |D_\mu \phi|^2 - V(\phi)$$

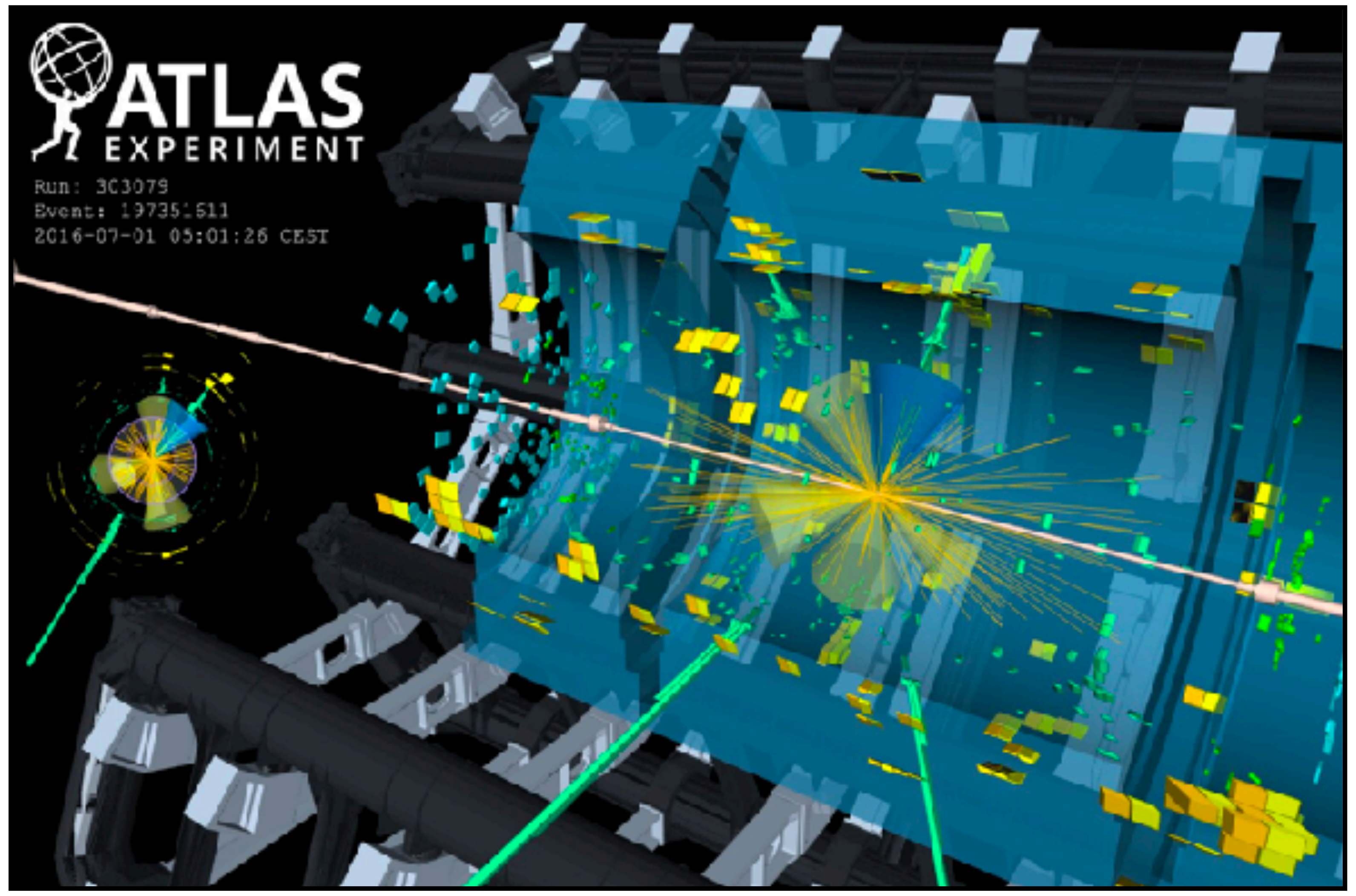
θ

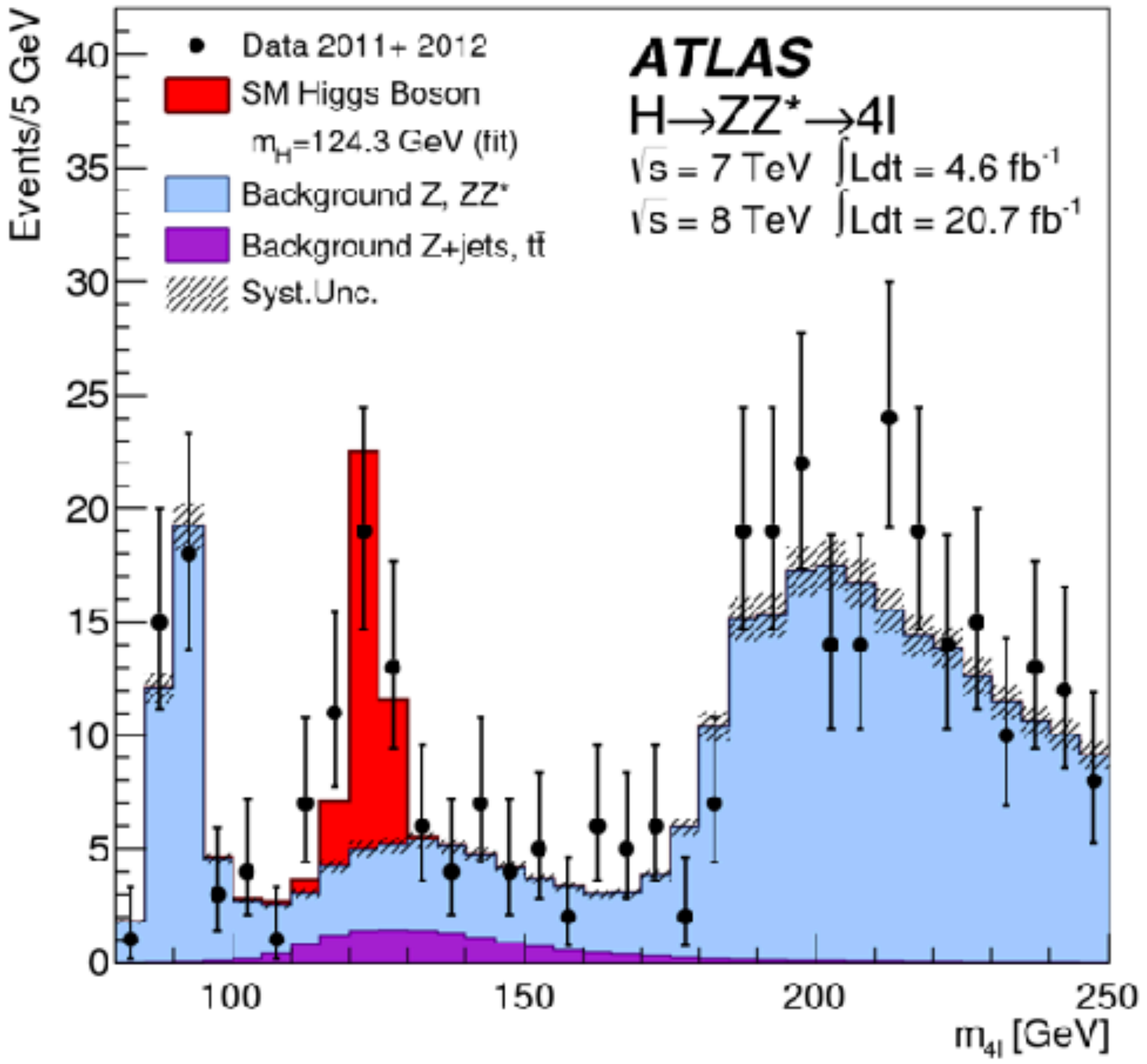
inference process: $p(\theta \,|\, x)$

Bridging the Gap

This approach works so well it generated a Nobel Prize

Frixione yesterday: We know what we're doing!

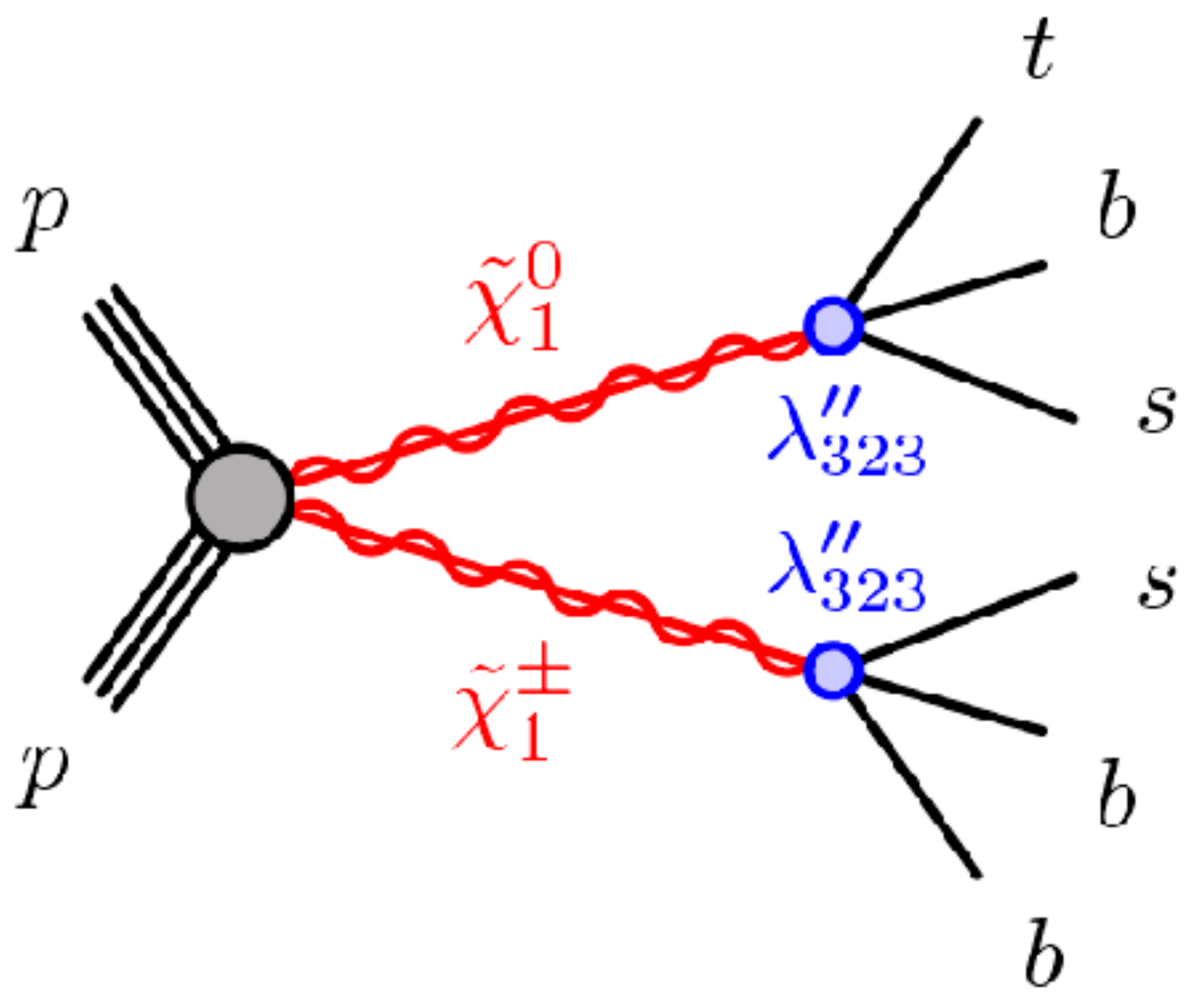


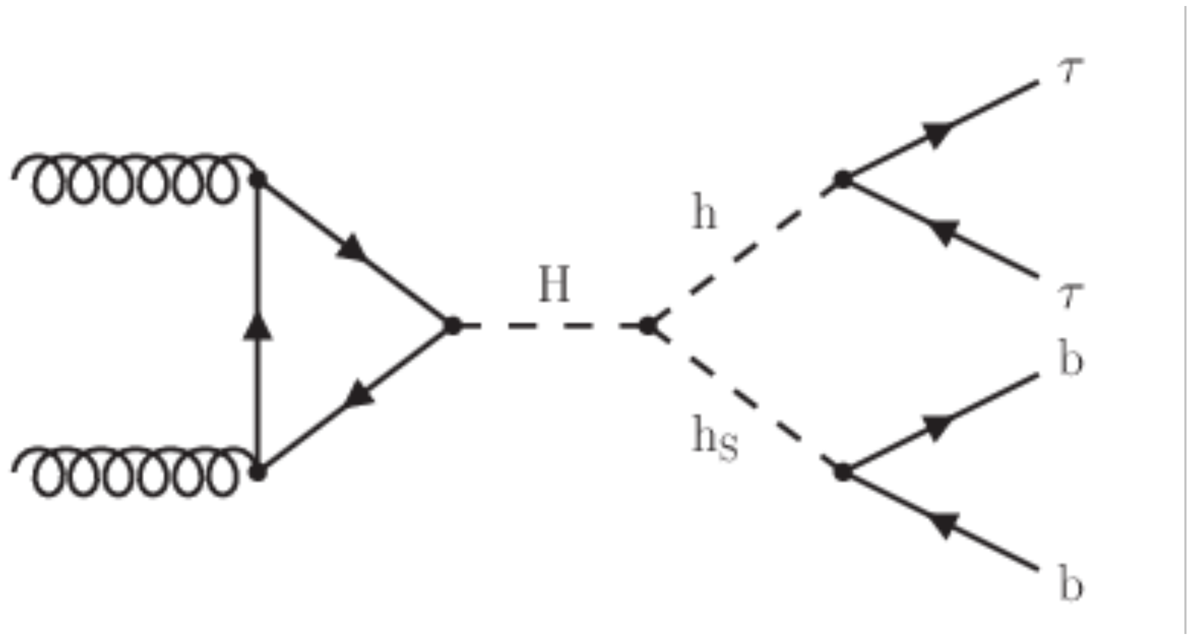


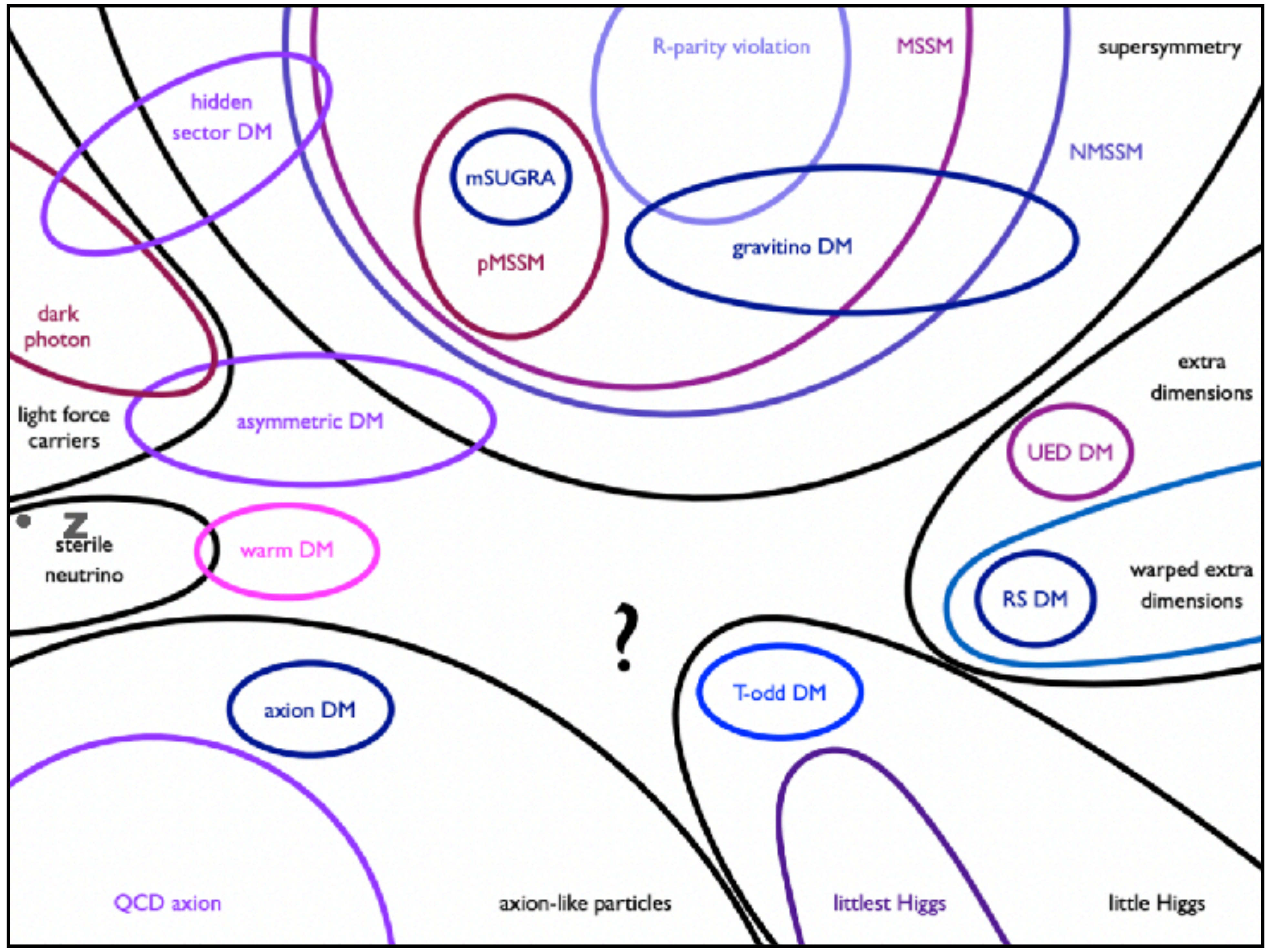
$\mathbb{R}^{10^8} \longrightarrow \mathbb{R}^1 \longrightarrow$

What's not to like? What are we missing?

What’s the Problem







Good but, not good enough

Will our analysis techniques enable discovery in e.g. fully hadronic states?

Much too Slow

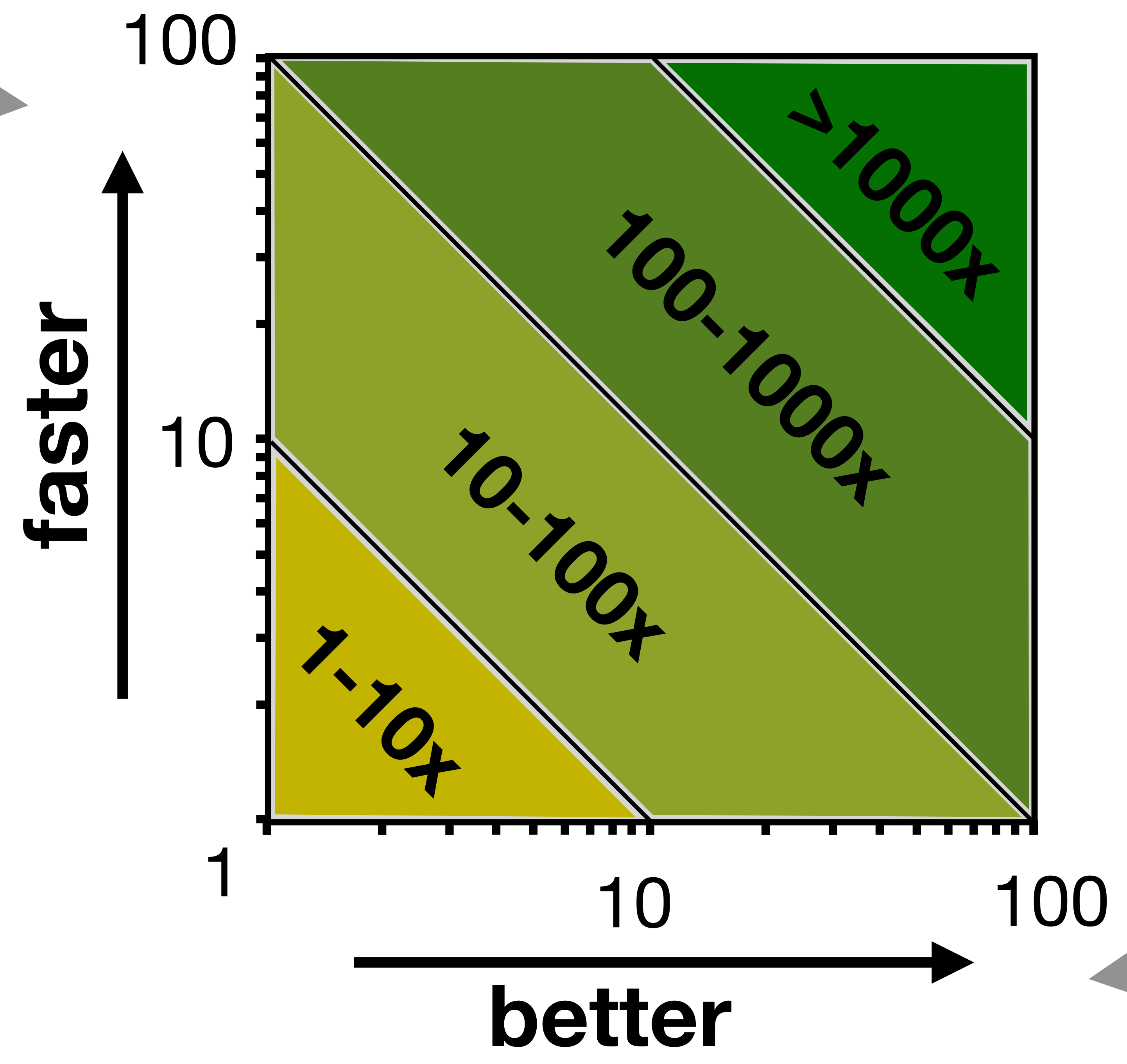
Only scratched the surface of the BSM because a single analysis takes O(years)

What we need

Two complementary - **equally valid** - methods for fundamentally change the field

Can we accelerate the time it takes to assess e.g. a BSM theory to days or hours?

e.g. by rapidly re-optimizing existing analyses to new signals?



How much more is there to gain in terms of analysis techniques?

AI might be the tool to get us there (→ Dyson)

It’s not necessarily a pipe dream

faster

AlphaFold’s breakthrough was not discovery of something new.

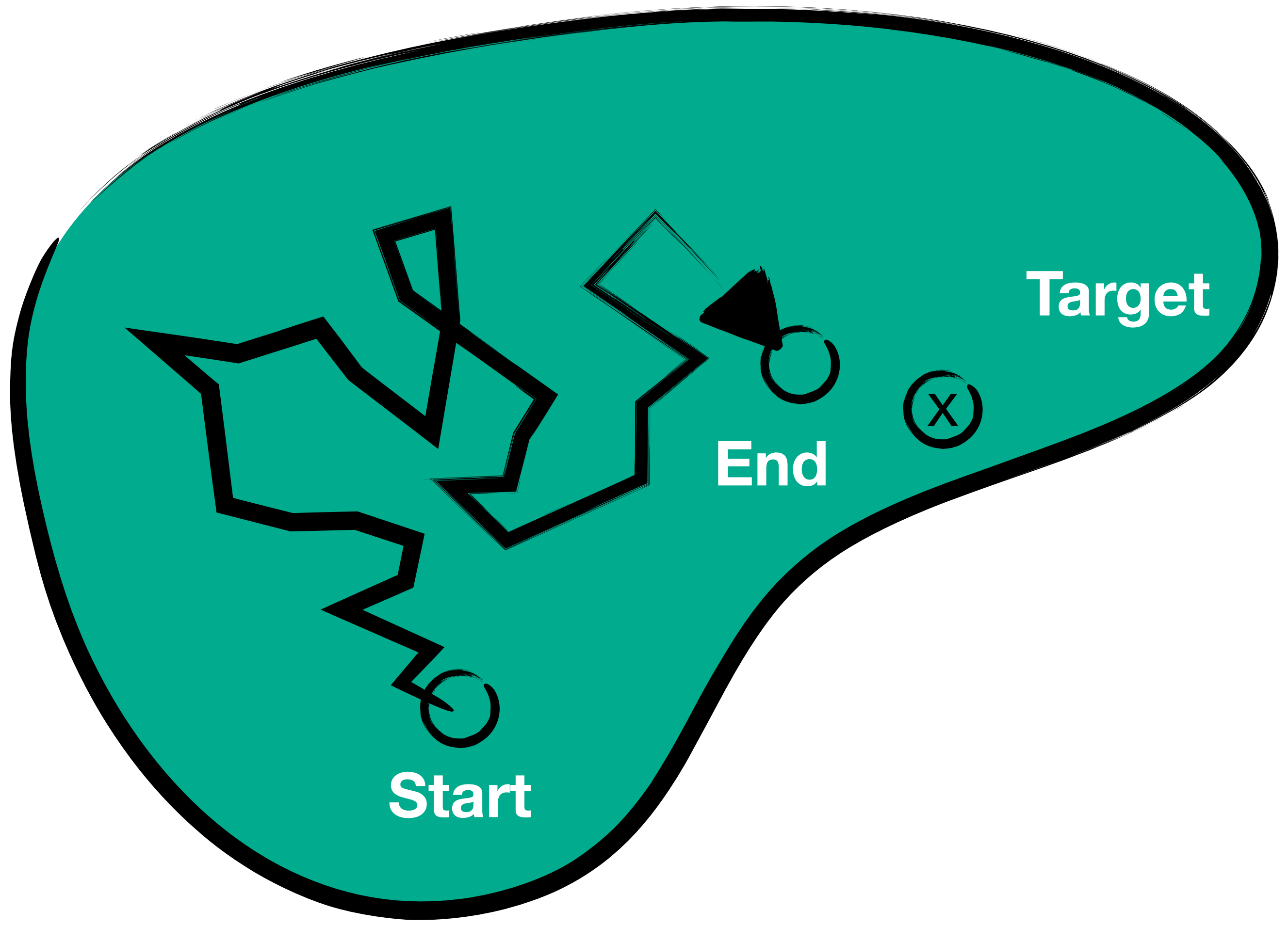
It was the **massive acceleration** of a extremely labor- and compute-intensive process from ~thesis to ~seconds.

Analogy:
structure prediction = single BSM model
“drug discovery” = “landscape scan”

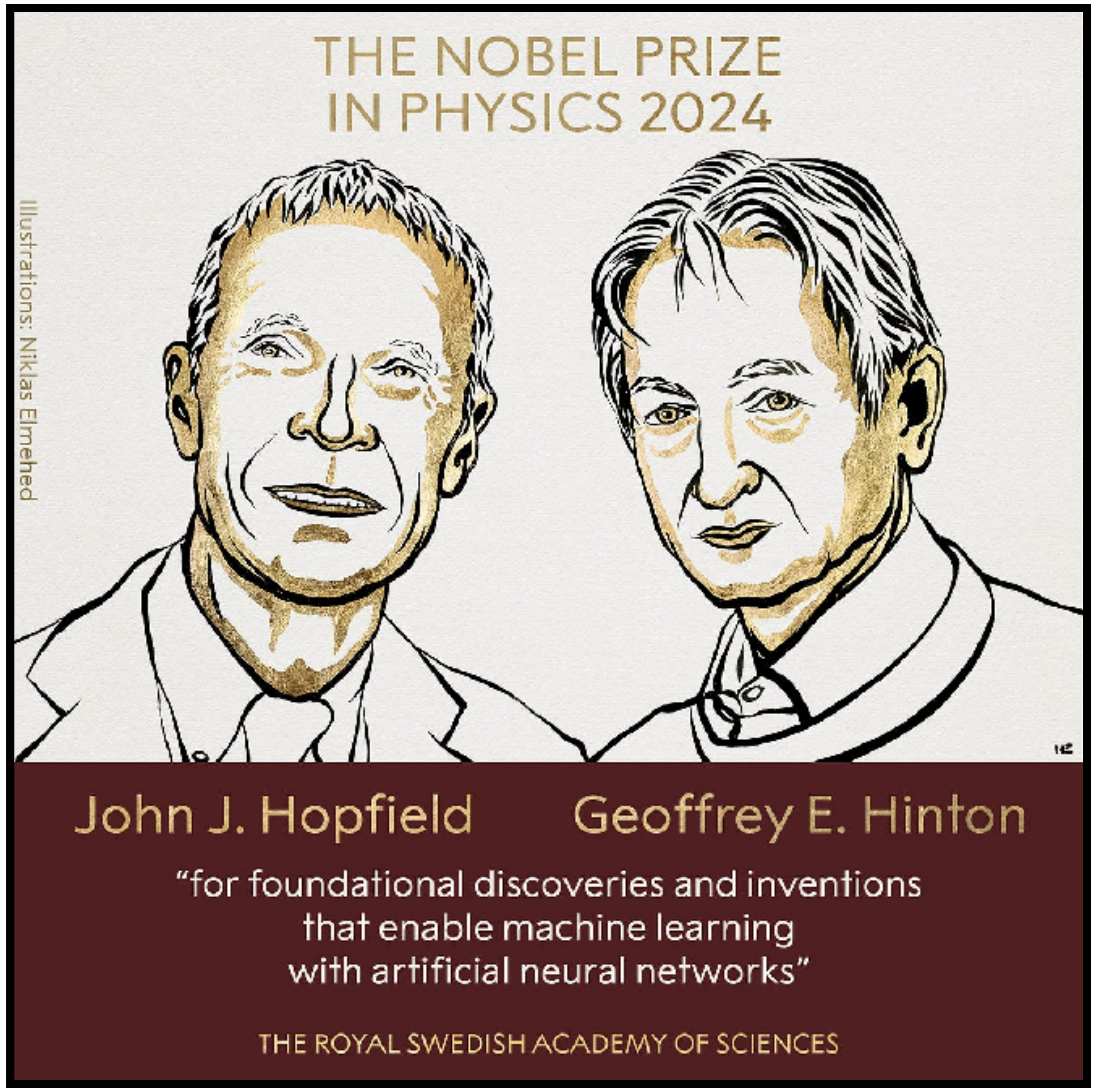
It’s not necessarily a pipe dream

Hinton is most famous for proving that neural networks can be trained efficiently in high (now trillions!) dimensions

→ Key: efficient gradient estimation
(Automatic Differentiation / Backpropagation)



Training =
Algorithm Search

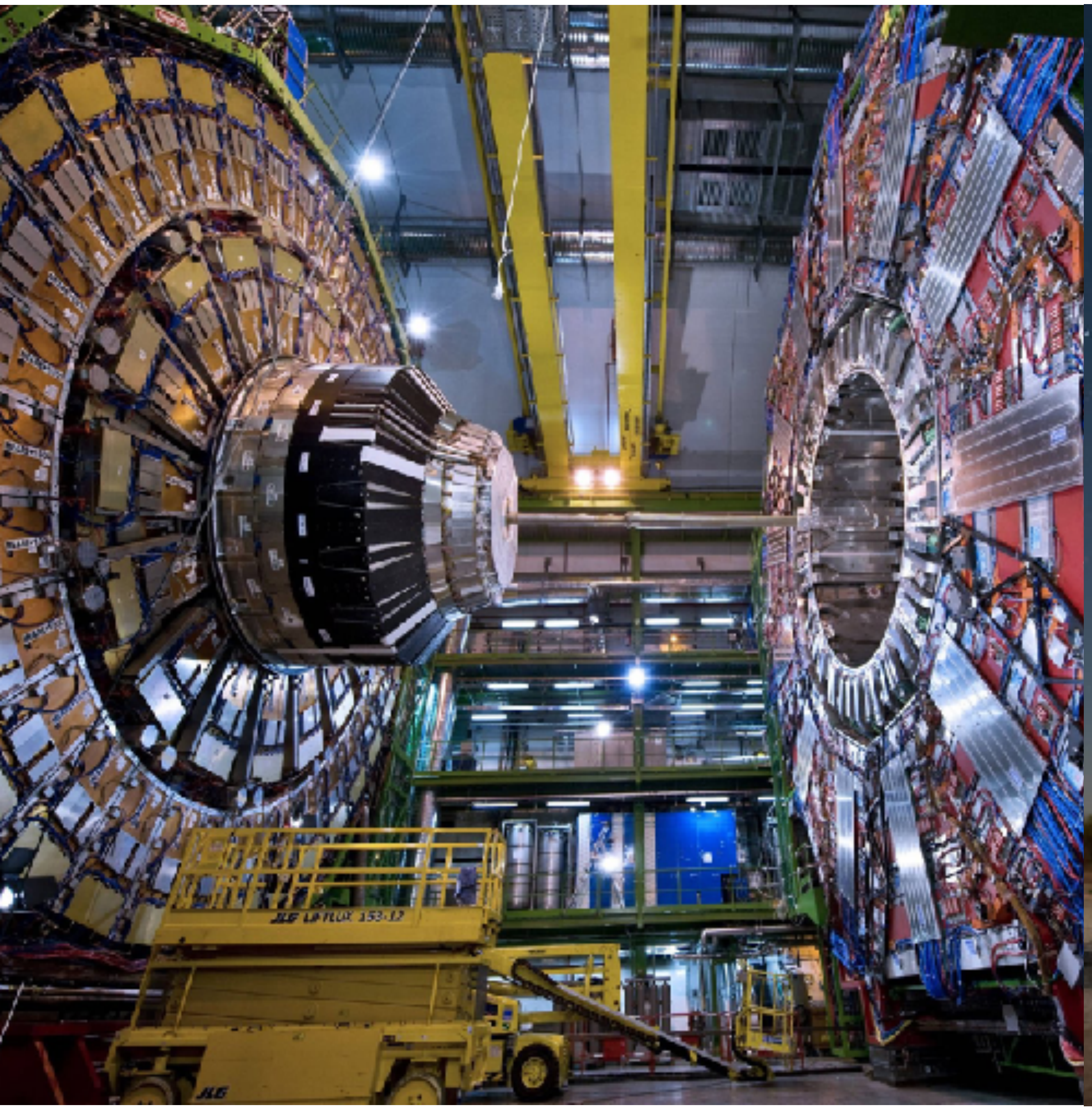


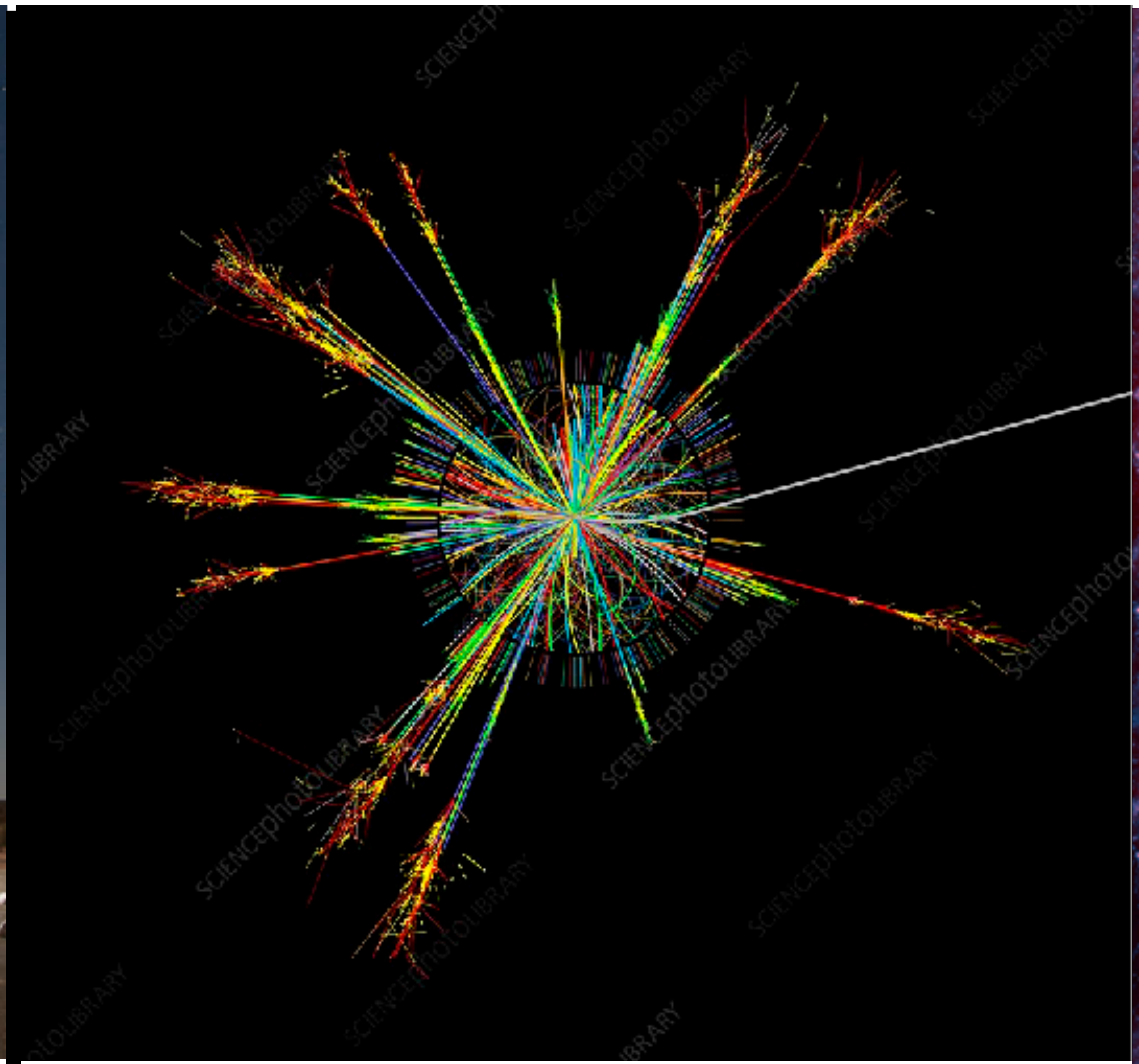
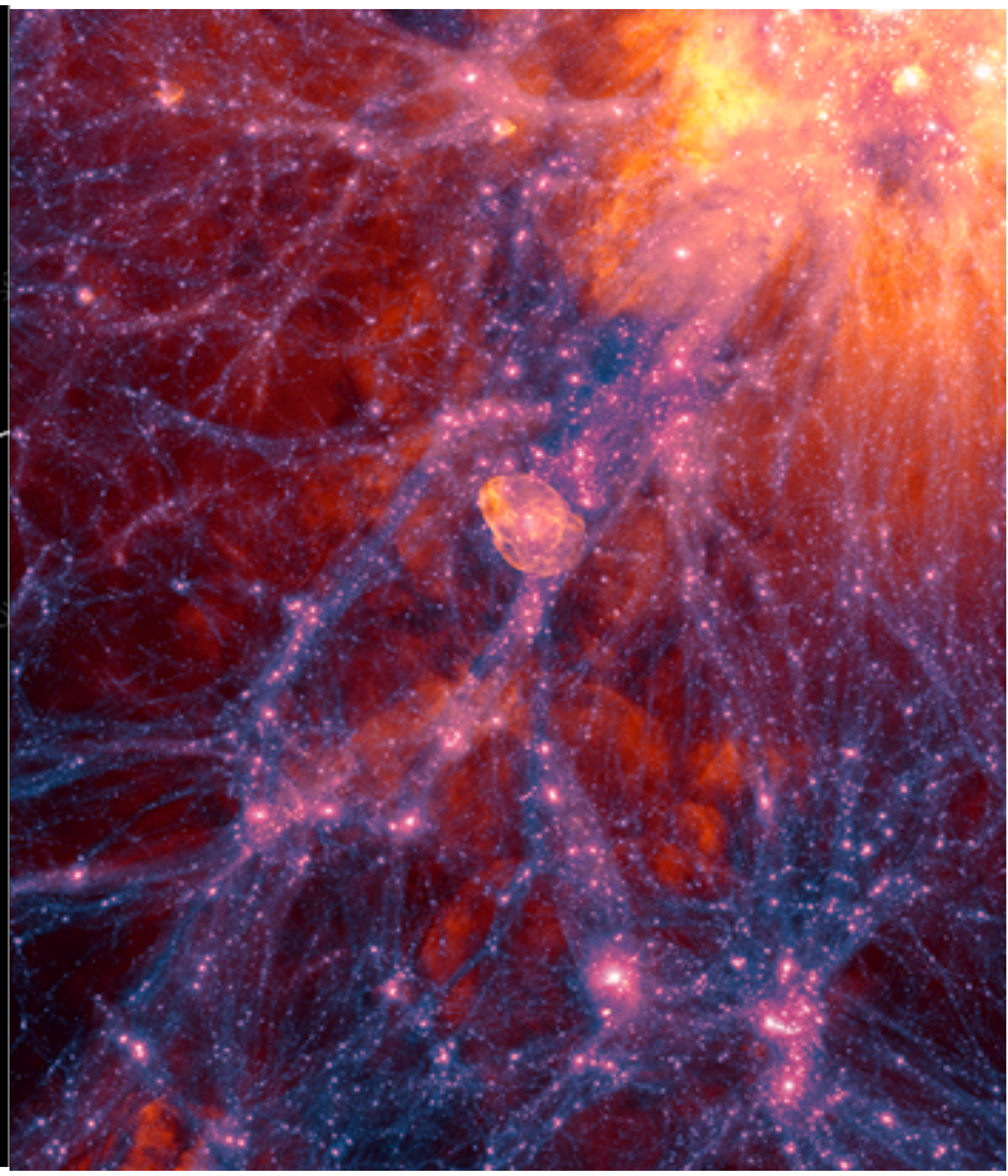
better

The key to find algorithms that surpass human performance

Fundamental Physics = AI Utopia

We are well equipped to capitalize on massive AI progress



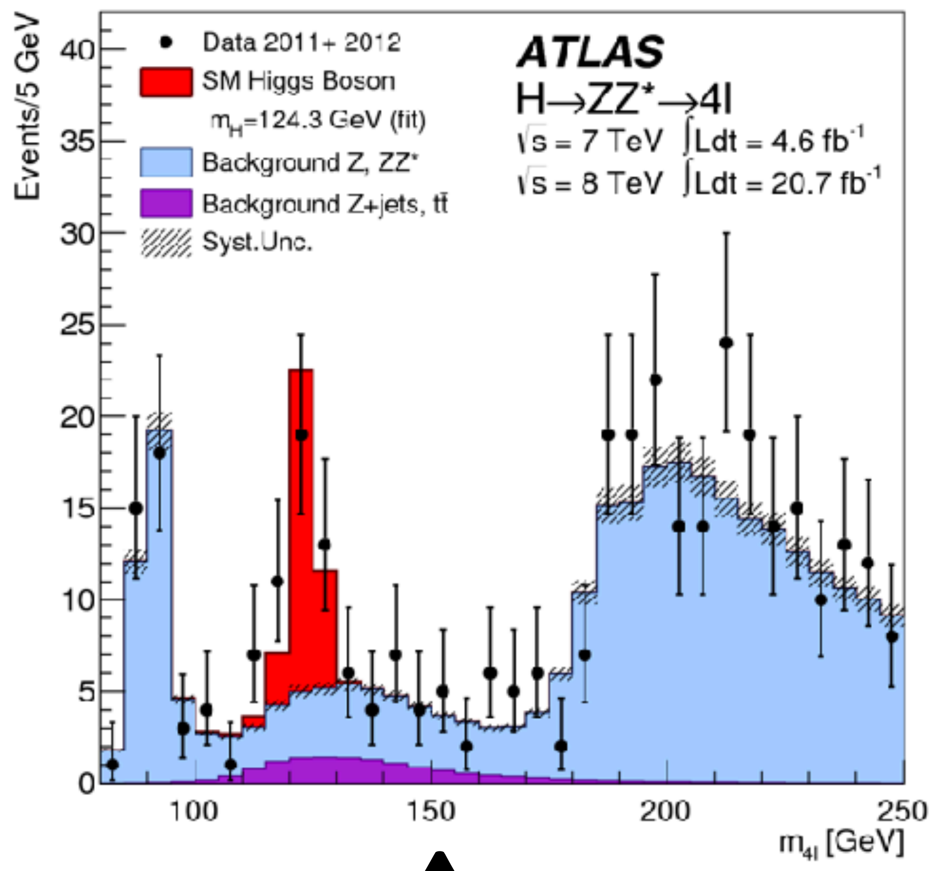



Exabytes of Experimental Data from Large-Scale Experiments
→ *much more than used to train ChatGPT*

High Quality Simulators allow us to explore new hypothetical models of the universe
→ *perfect training data for AI*

AI and HEP

$$\mathcal{L} = -\frac{1}{4} F_{\mu\nu} F^{\mu\nu} + i \bar{\psi} \not{D} \psi + h.c. + \bar{\psi}_i y_{ij} \psi_j \phi + h.c. + |D_\mu \phi|^2 - V(\phi)$$

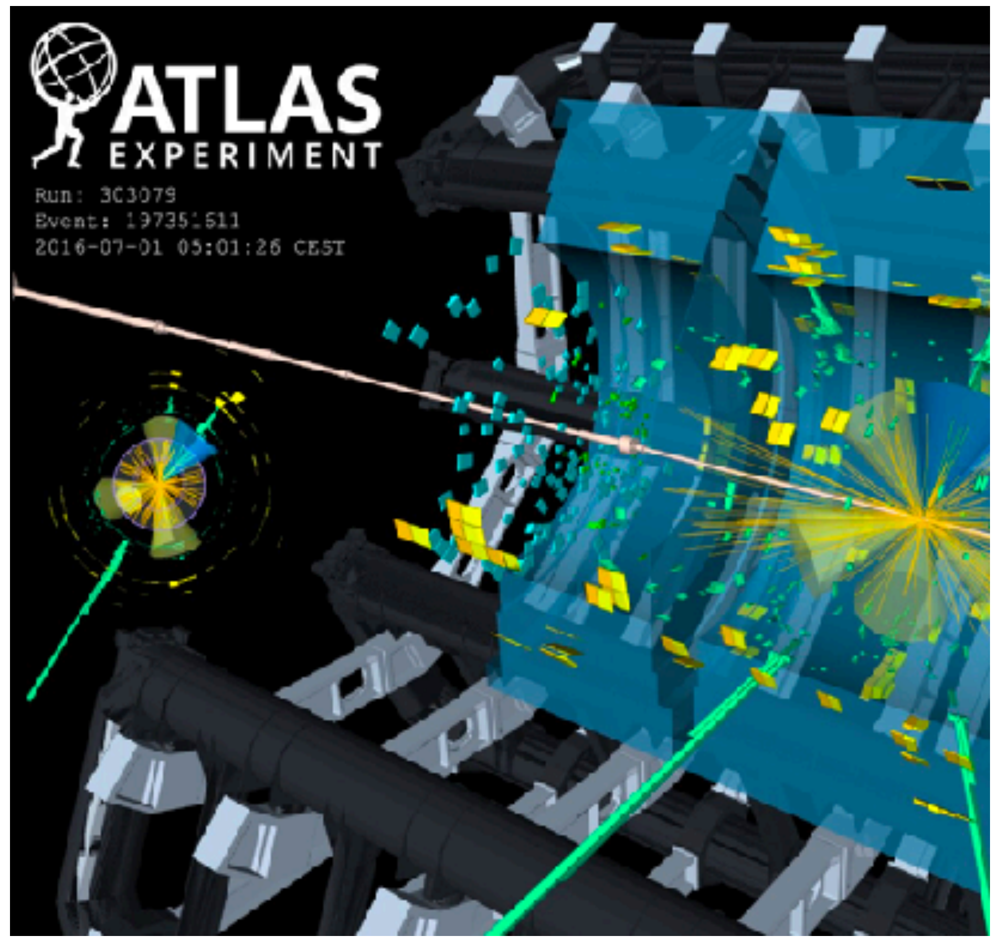


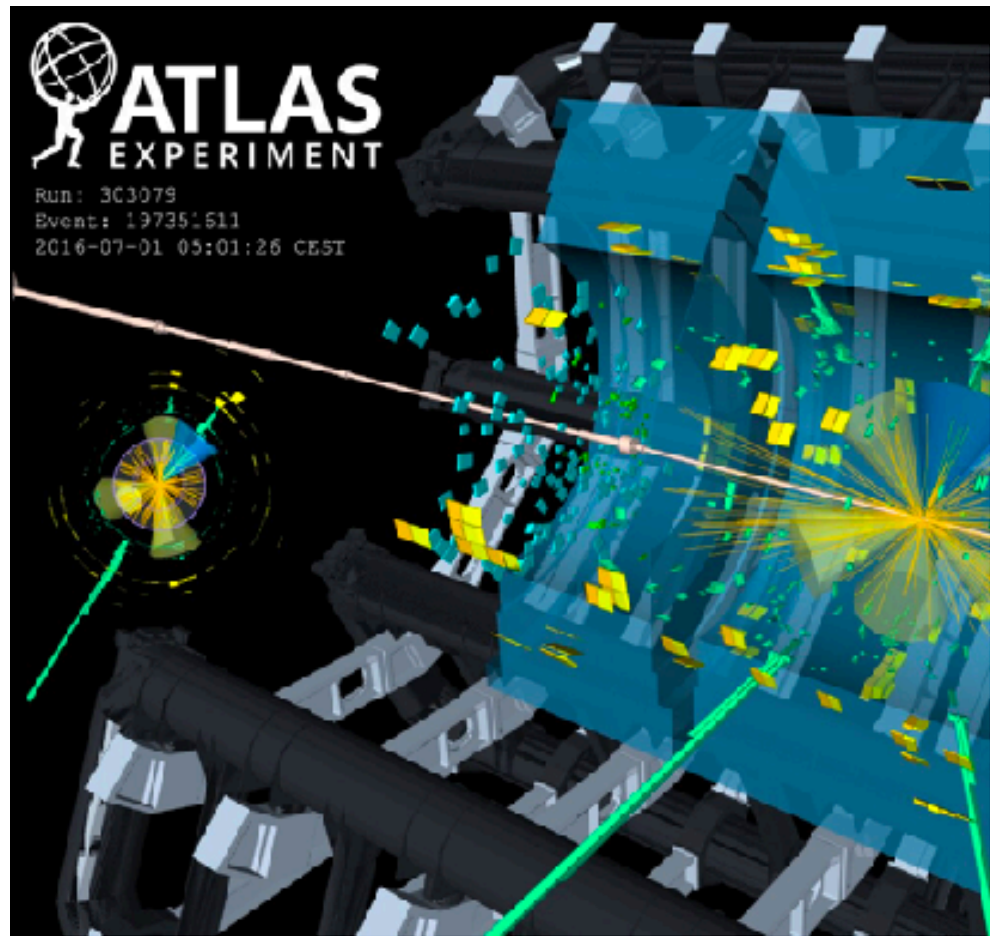
High-Level Concept
(here: Physics Theory)

Stochastic Prediction
$(x, \theta) \sim p(x\,|\,\theta)p(\theta)$

High-Dimensional Inference
$\{(x_1, \theta_1,), (x_2, \theta_2), \ldots\} \to p(\theta\,|\,x)$

Low-Level Data
(here: Experiment Sensors)





AI and HEP

street style photo of a woman selling pho at a Vietnamese street market, sunset, shot on fujifilm

High-Level Concept
(here: Language)

This is a picture of Barack Obama.
His foot is positioned on the right side of the scale.
The scale will show a higher weight.

Stochastic Generation
$(x, \theta) \sim p(x\,|\,\theta)p(\theta)$

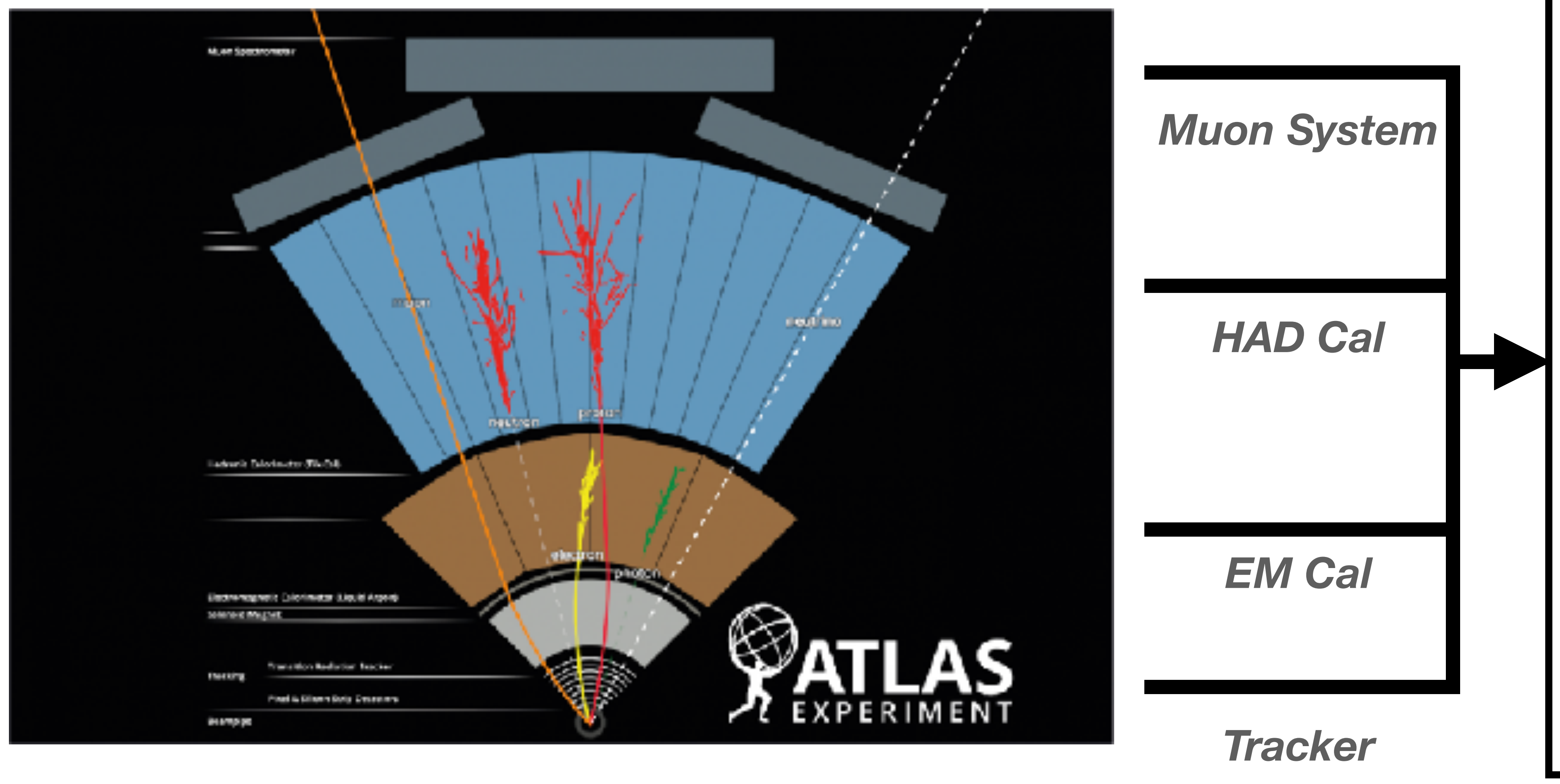
Sample-based Inference
$\{(x_1, \theta_1,), (x_2, \theta_2), \ldots\} \rightarrow p(\theta\,|\,x)$

Low-Level Data
(here: Pixels)

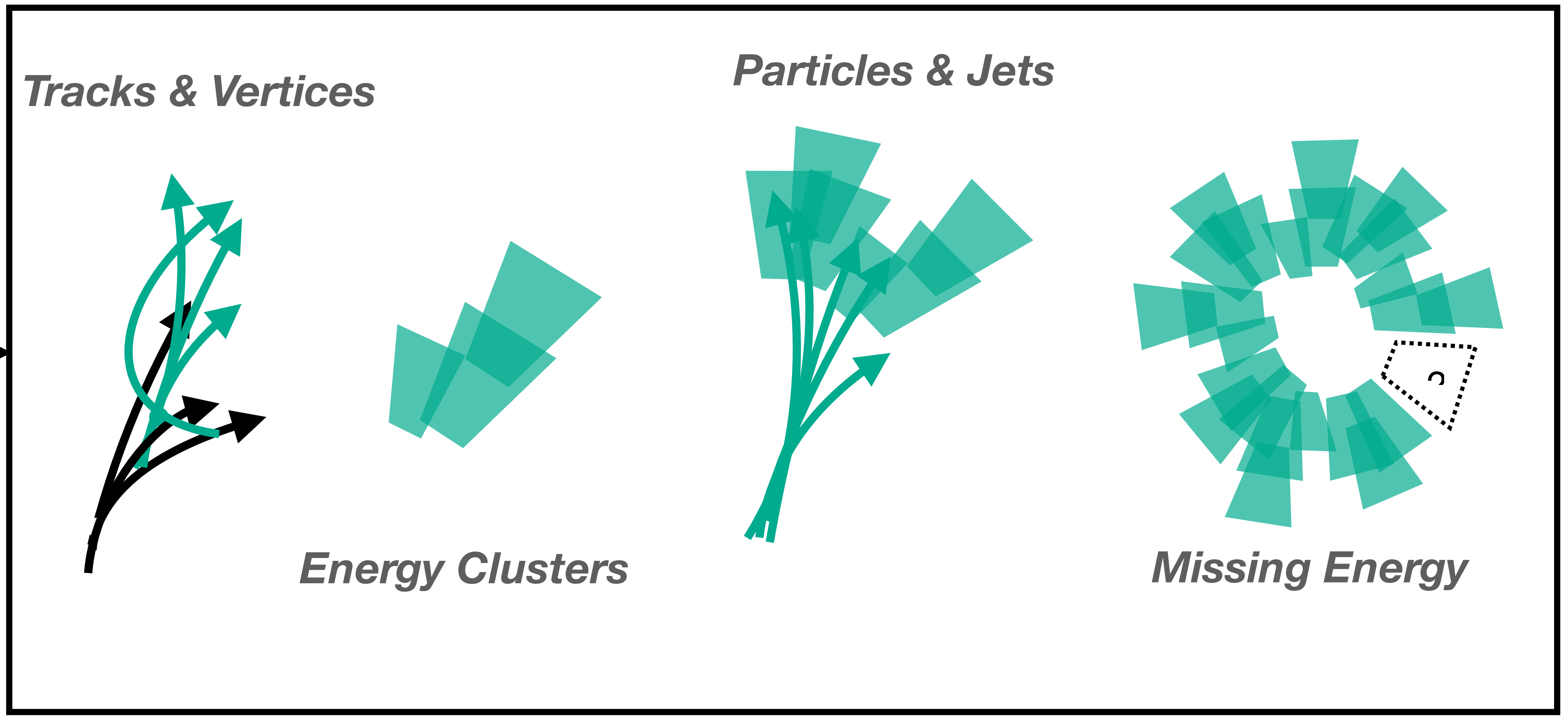
How should we analyze data

The **science we can extract from data** depends crucially on **what language we use** to describe it

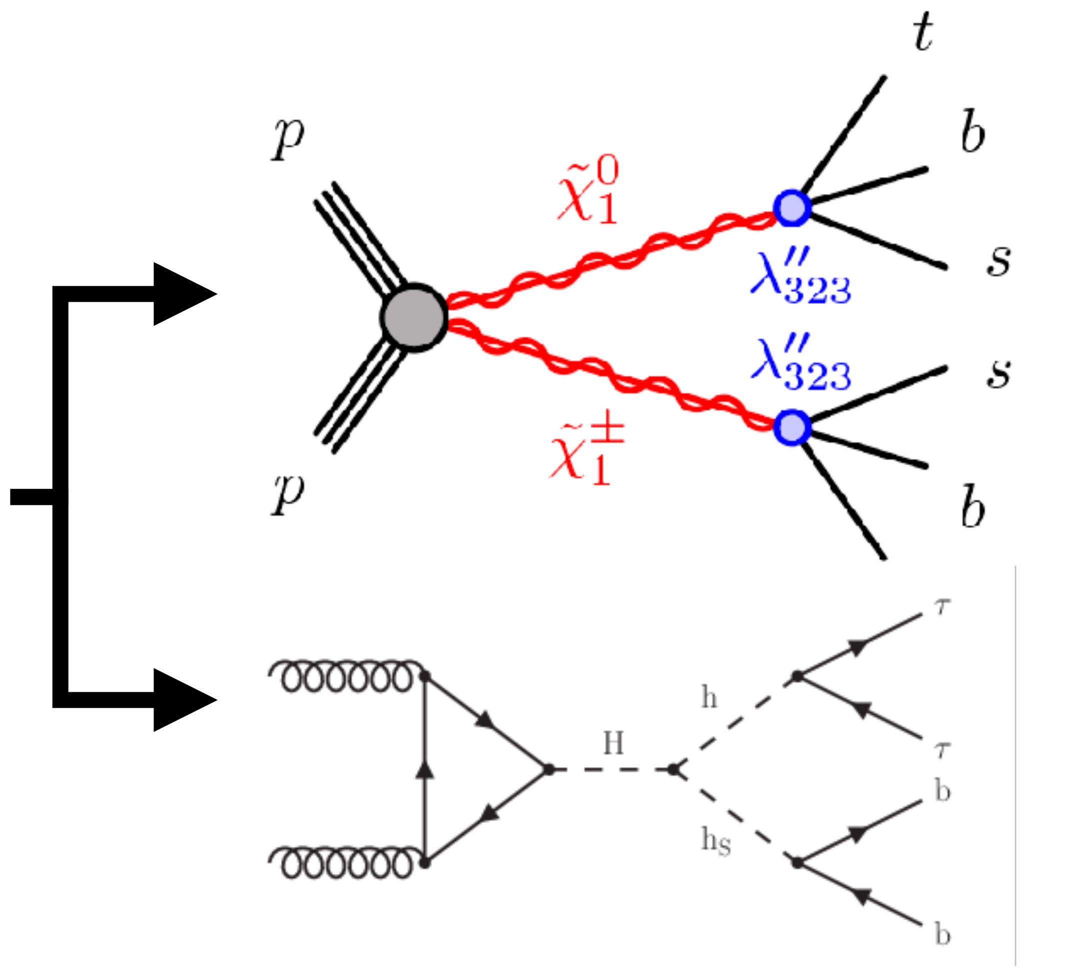
Raw Data → **Reconstruction** → Interpretation → **Analysis** → Result



100 M sensor readouts



Our Vocabulary: O(1k-10k) physics objects

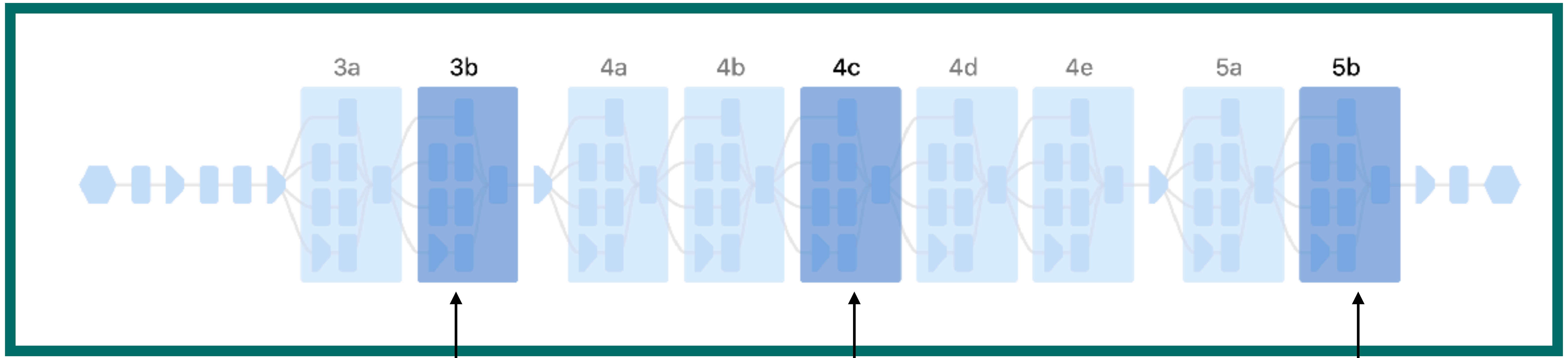


O(1) Analysis Observable

Are we sure that we - humans - can find the best vocabulary?

Deep Learning shows: Not always

We know that **AI with access to the Raw Data** can **learn its own vocabulary** and outperform human-designed algorithms



Labrador!

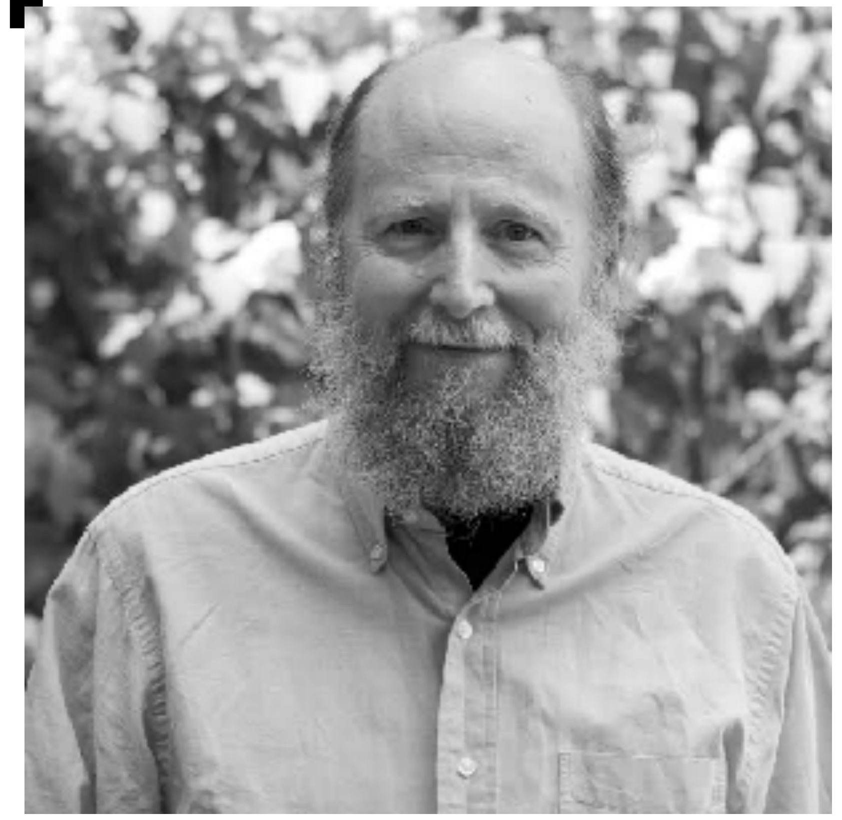
Low-level (Edges)

Medium-level (Textures)

High-Level (Animals)

To detect dogs in images, nobody writes a flappy-ear algorithm by hand!

The Bitter Lesson



The Bitter Lesson

Rich Sutton

March 13, 2019

The biggest lesson that can be read from 70 years of AI research is that general methods that leverage computation are ultimately the most effective, and by a large margin. The ultimate reason for this is Moore's law, or rather its generalization of continued exponentially falling cost per unit of computation. Most AI research has been conducted as if the computation available to the agent were constant (in which case leveraging human knowledge would be one of the only ways to improve performance) but, over a slightly longer time than a typical research project, massively more computation inevitably becomes available. Seeking an improvement that makes a difference in the shorter term, researchers seek to leverage their human knowledge of the domain, but the only thing that matters in the long run is the leveraging of computation. These two need not run counter to each other, but in practice they tend to. Time spent on one is time not spent on the other. There are psychological commitments to investment in one approach or the other. And the human-knowledge approach tends to complicate methods in ways that make them less suited to taking

There were many examples of AI researchers' belated learning of this bitter

ominent.

hampion, Kasparov, in 1997, were based on massive, deep search. At the

of computer-c

hen a simple

sed chess res

search may have won this time, but it was not a general strategy, and a

wanted methods based on human input to win and were disappointed

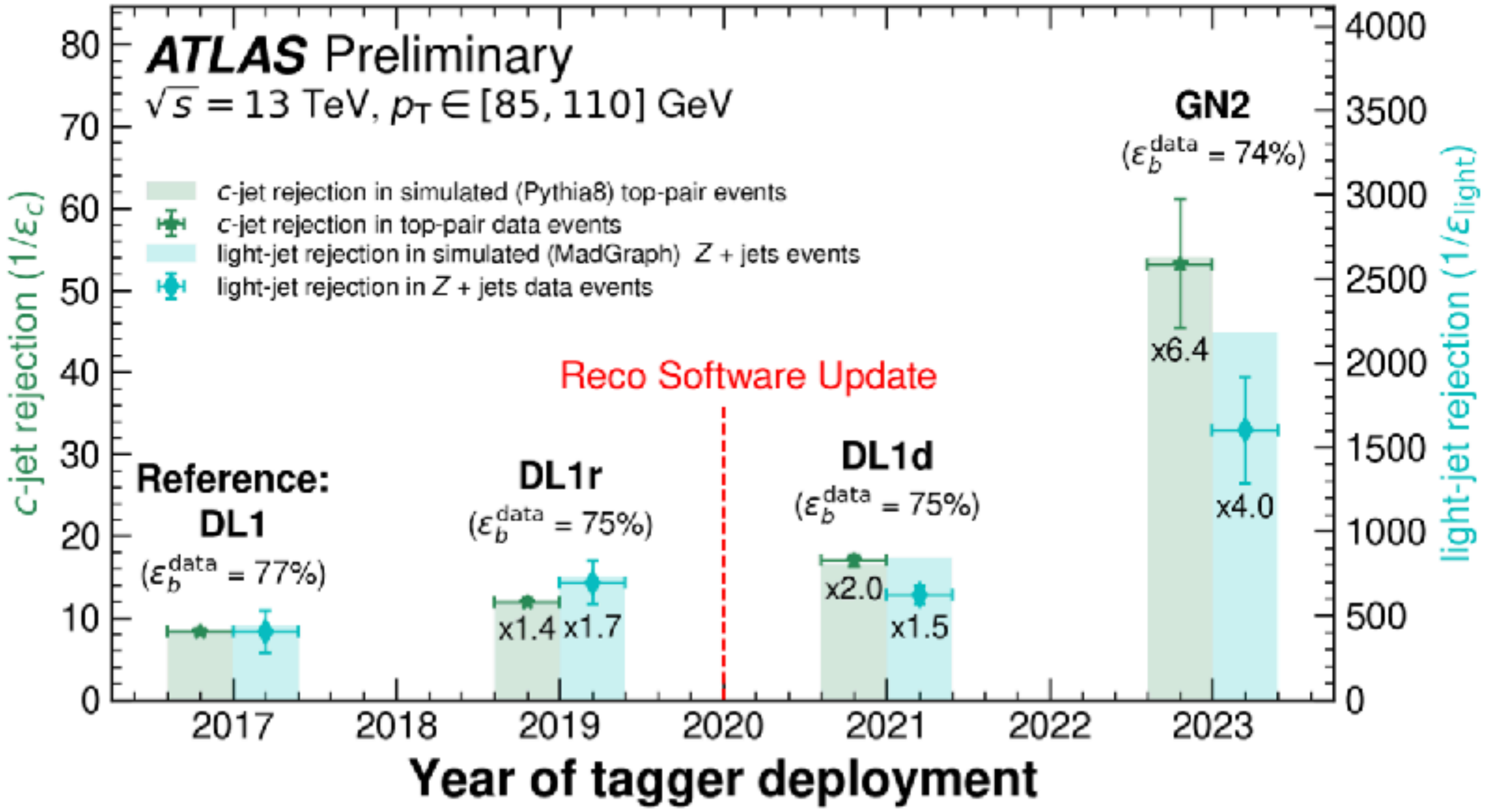
Researchers seek to leverage their human knowledge […], but the only thing that matters in the long run is the leveraging of computation

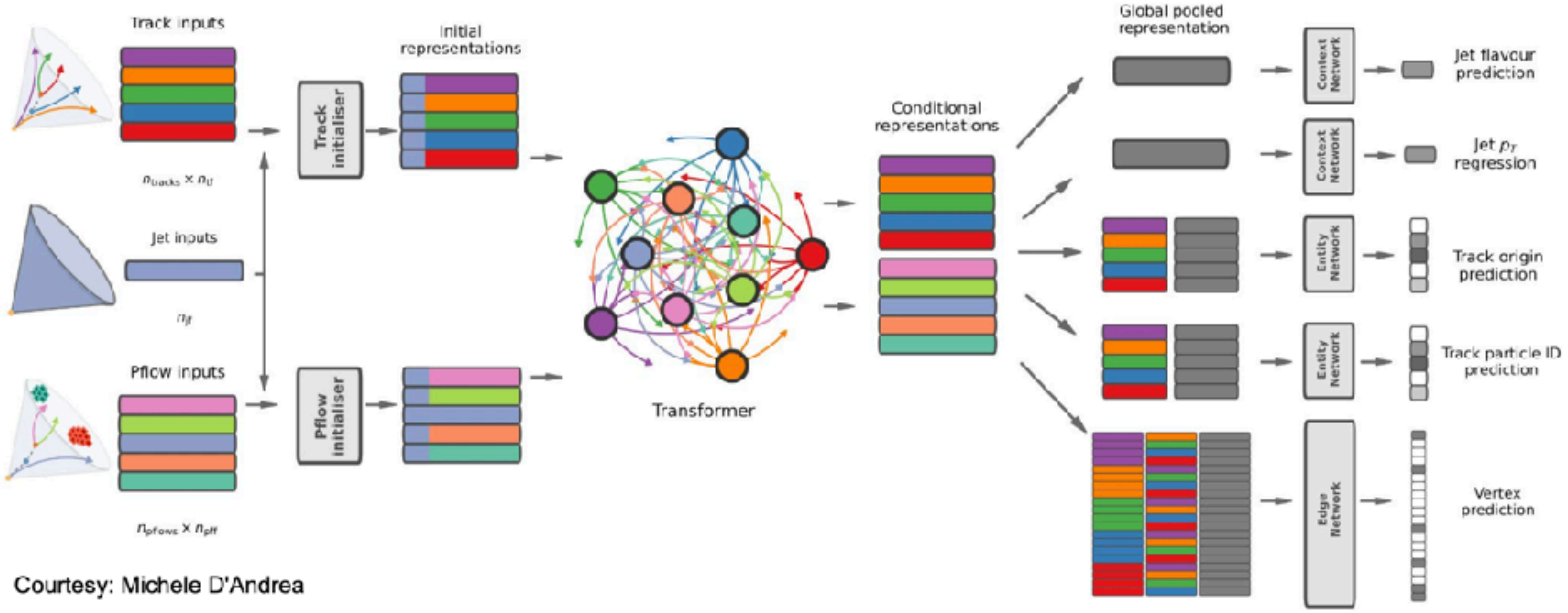
… many examples of AI researchers' belated learning of this bitter lesson

How much can we gain if we embrace it?

It's now happening in HEP. Improvements e.g. in Flavor Tagging are equivalent to **years(!) of LHC data**

→ "hands-off" approach, let the network see low-level data



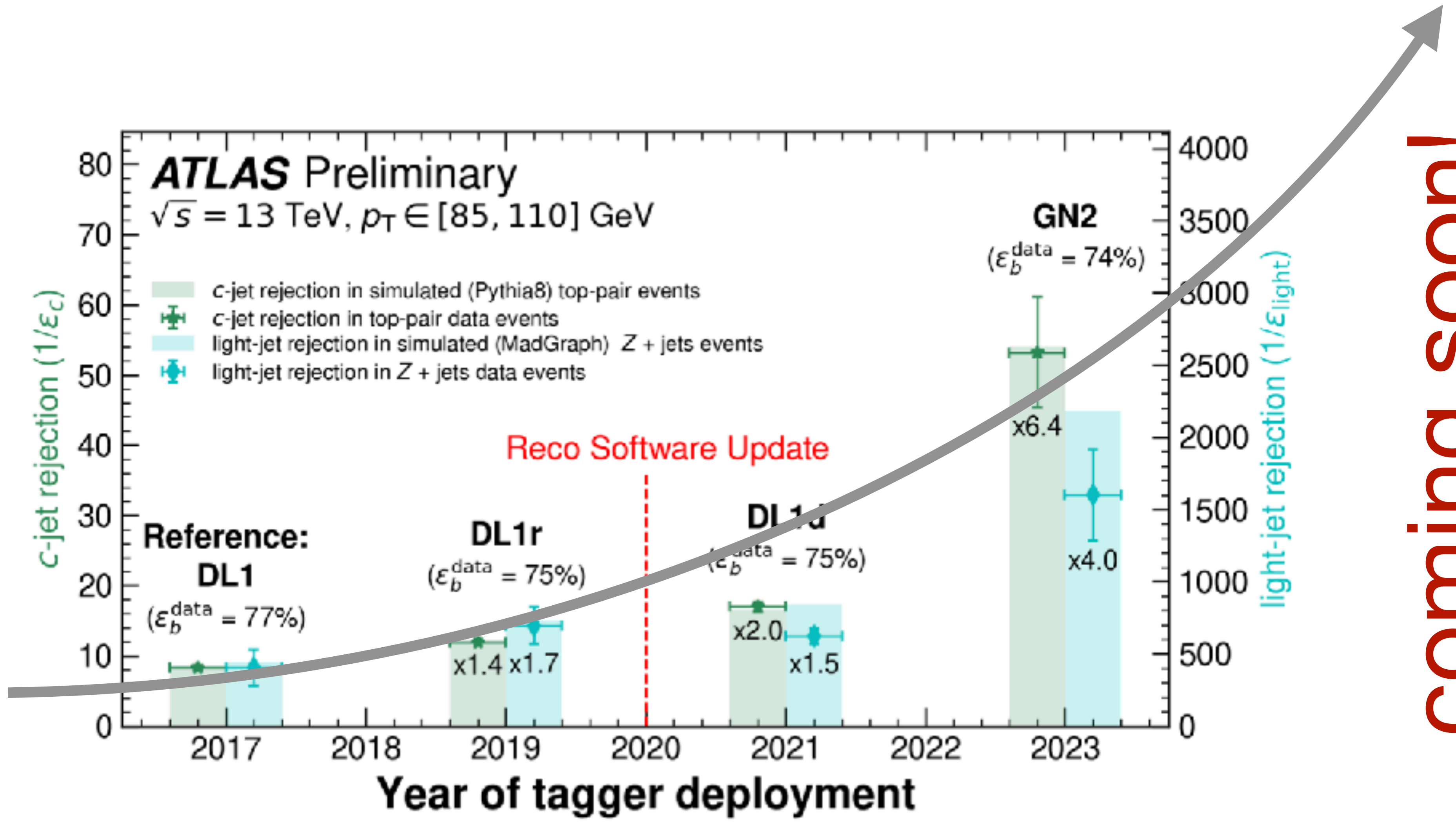


A O(100 M$) Impact already, can we even more?

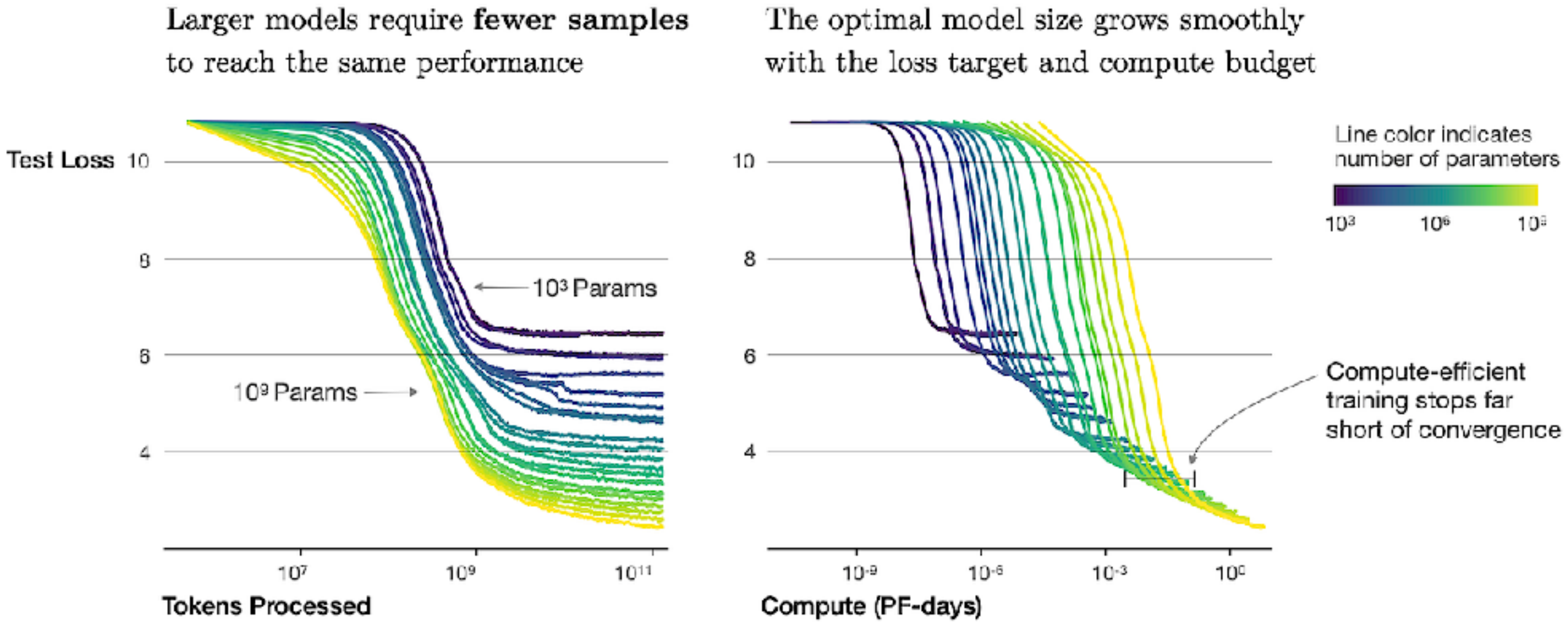
How much can we gain if we embrace it?

Recent Insight: we are nowhere close to the detector limit

→ we can predict performance of scaling up (“Scaling Laws”)



coming soon!



State of the Art Flavor-Tagging ⟷ **State of the Art Language Models**

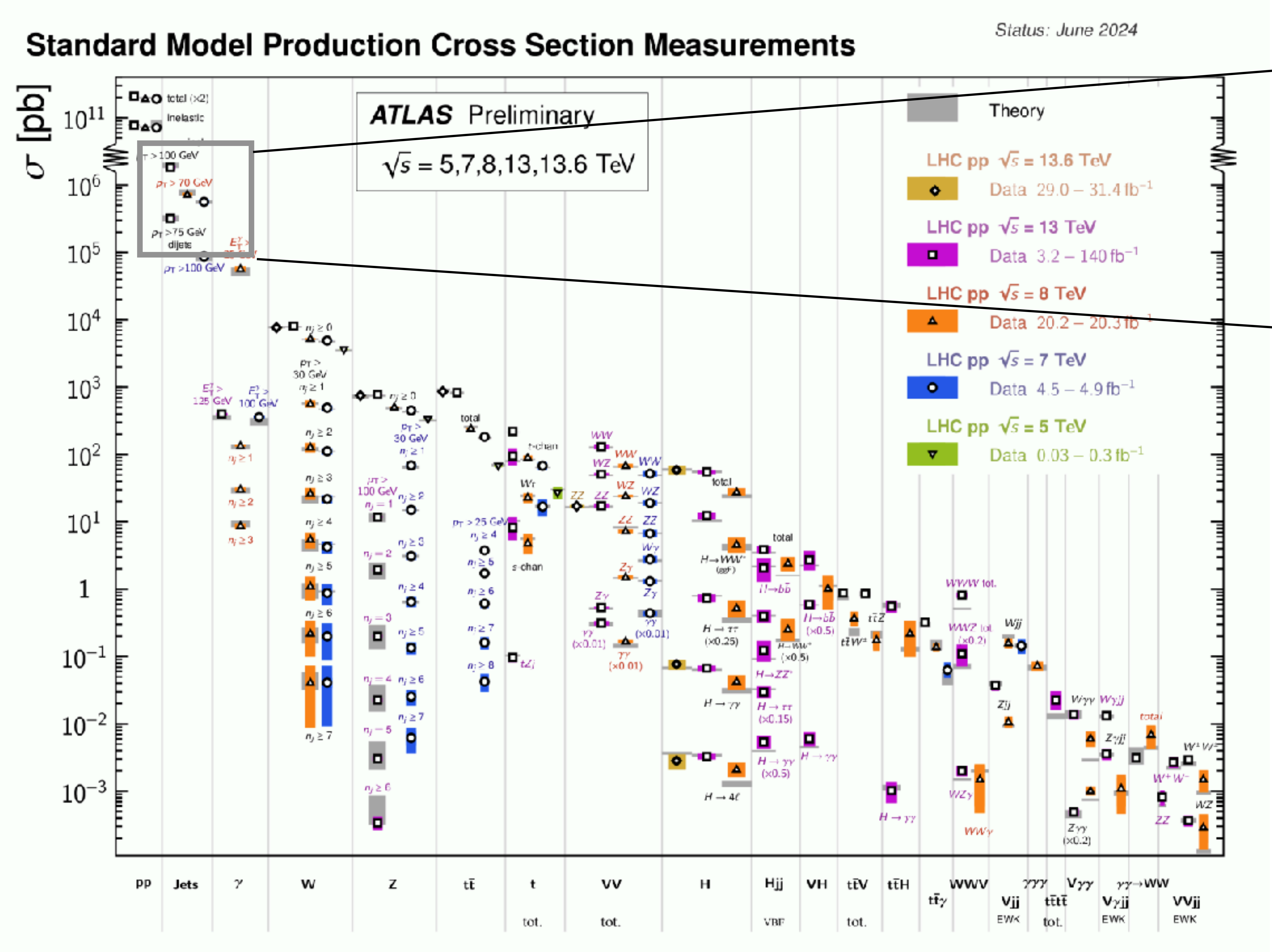
tiny(!) 10M parameter model
trained on 300M jets

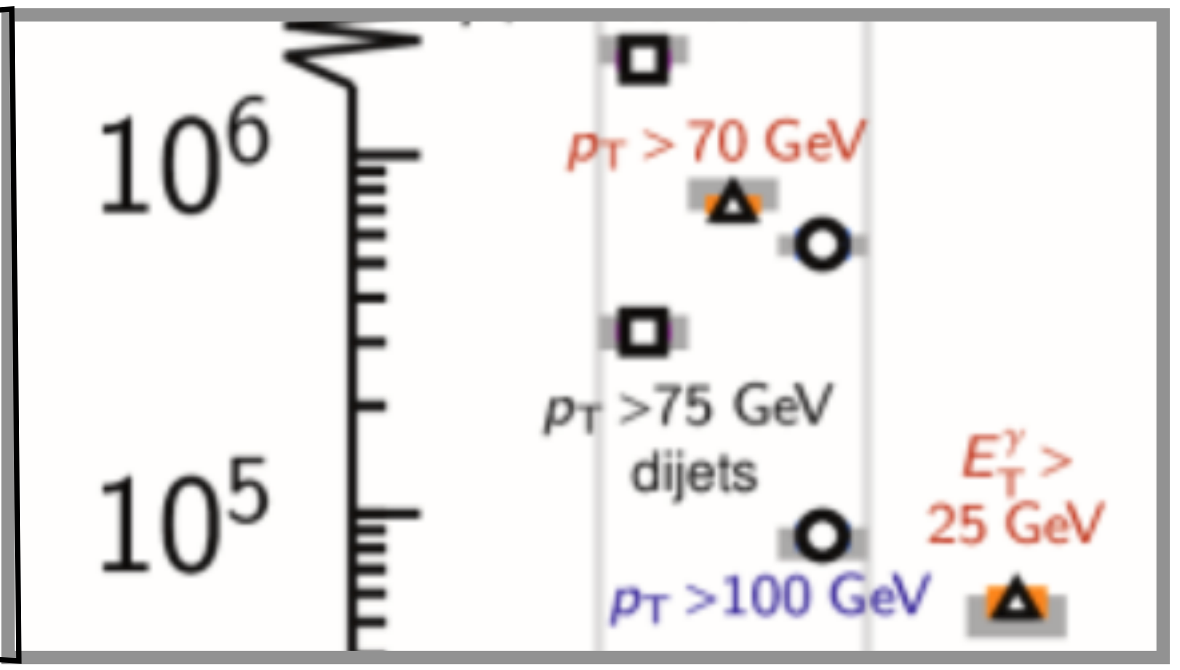
10,000x more data / parameters

100 Billion(!) parameter model
trained on trillions(!) of tokens

[ATLAS FTAG Group]

Too Few Jets at a Hadron Collider?





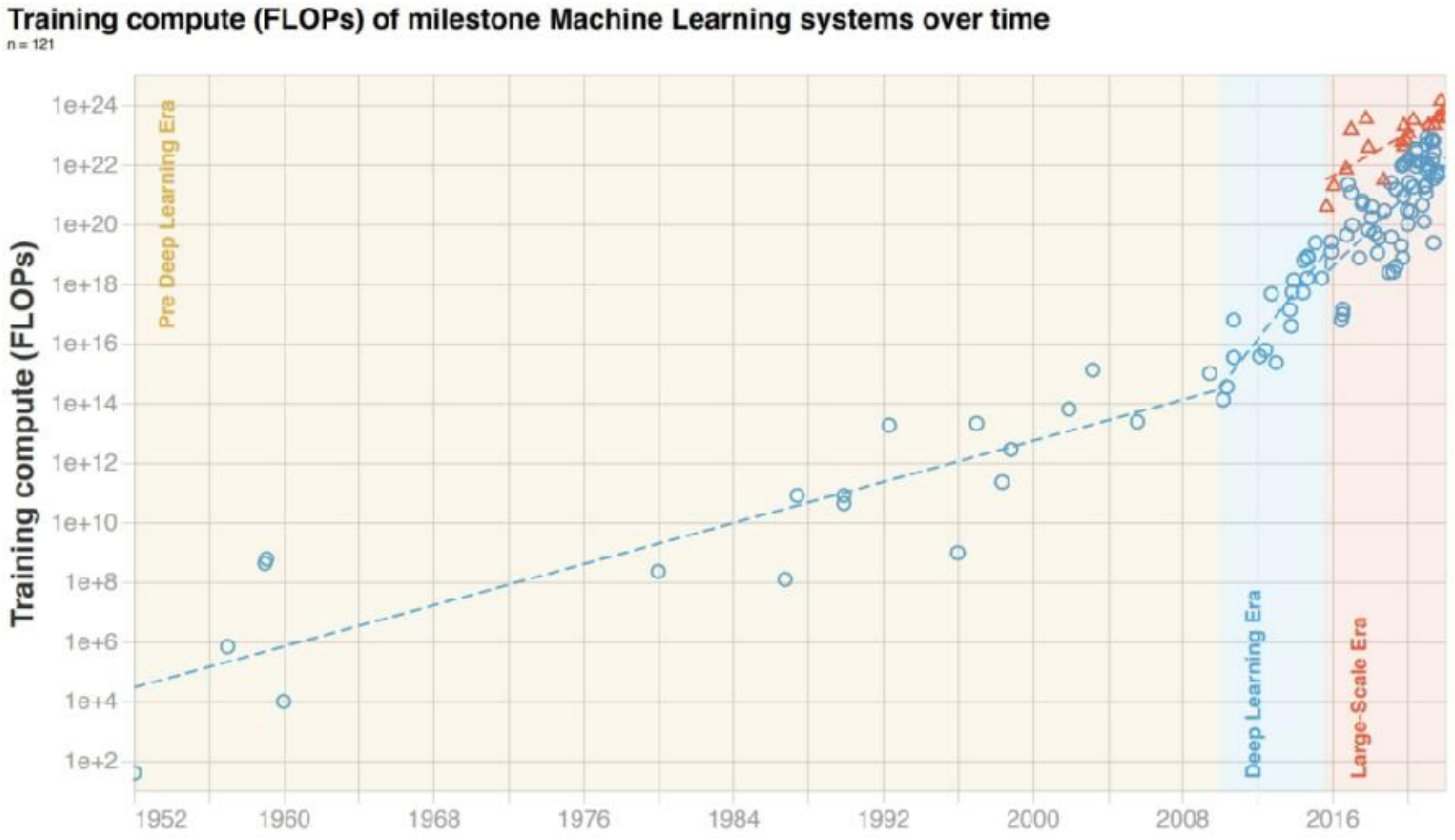
1 ab^{-1} of data gives us “only” a O(trillion) jets

For modern AI this is now the **required** scale

→ Fast Simulation?!

The Problem with Deep Learning

end2end DL has a cost: both **computational** and **conceptual**



vs

vs

Throws away everything we know

every task learned from scratch. ***expensive & slow!***

Solving for a single task is brittle

→ prone to ***non-sensical shortcuts***

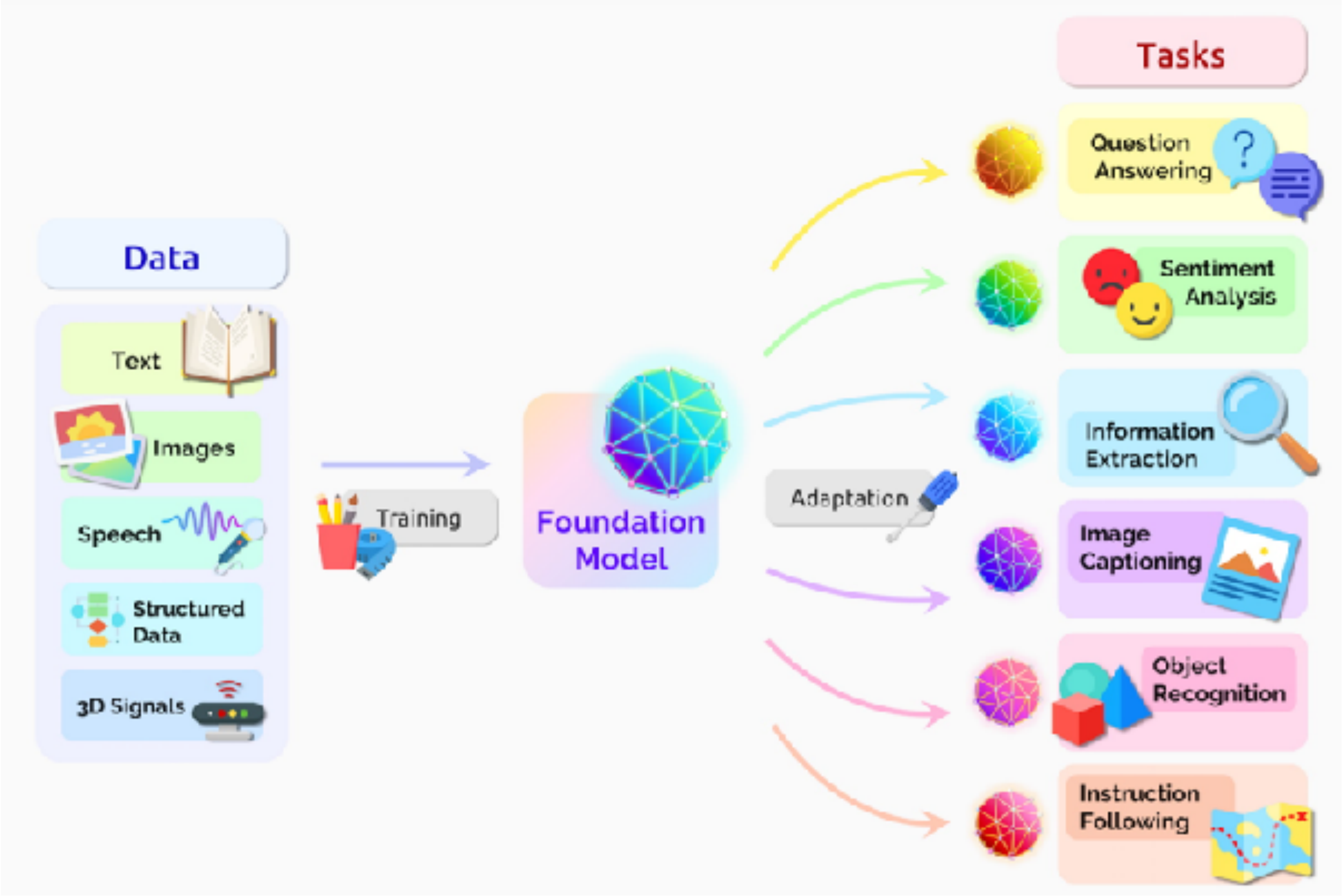
Foundation Models

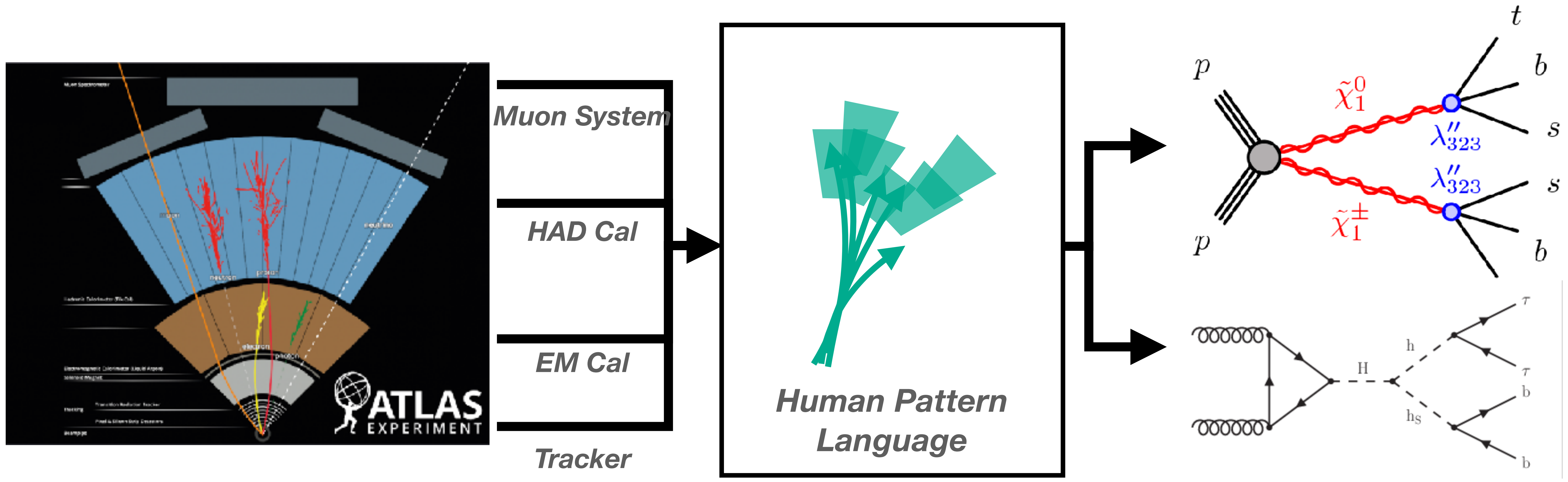
Train AI to learn its **own, multi-purpose, pattern vocabulary**

→ **Goal:** make it useful for not one, but many tasks

Raw Data → **Foundation Model** → Embedding → **Task-Network** → Result

Raw Data → **Reconstruction** → Interpretation → **Analysis** → Result





Actually very similar to the human two-step strategy we know so well

→ but combine with precision and autonomy of AI

Foundation Models

Train AI to learn its **own, multi-purpose, pattern vocabulary**

→ **Goal:** make it useful for not one, but many tasks

1 ***Pretraining***

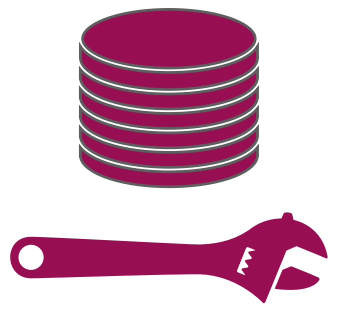
Raw Data → Foundation Model → Interpretation → Puzzles → Puzzle Solution

Large-Scale Datasets
(General Purpose)

2 ***Downstream Use***

Raw Data → Foundation Model → Interpretation → Application Model → Result

Small-scale Data
(Application Specific)



Mirror the human two-step strategy that we know so well

→ but combine with precision and autonomy of AI

How do you build Knowledge?

Idea: find a way to **synthesize many mini-tasks** that require (hopefully) developing a general “understanding” of the domain

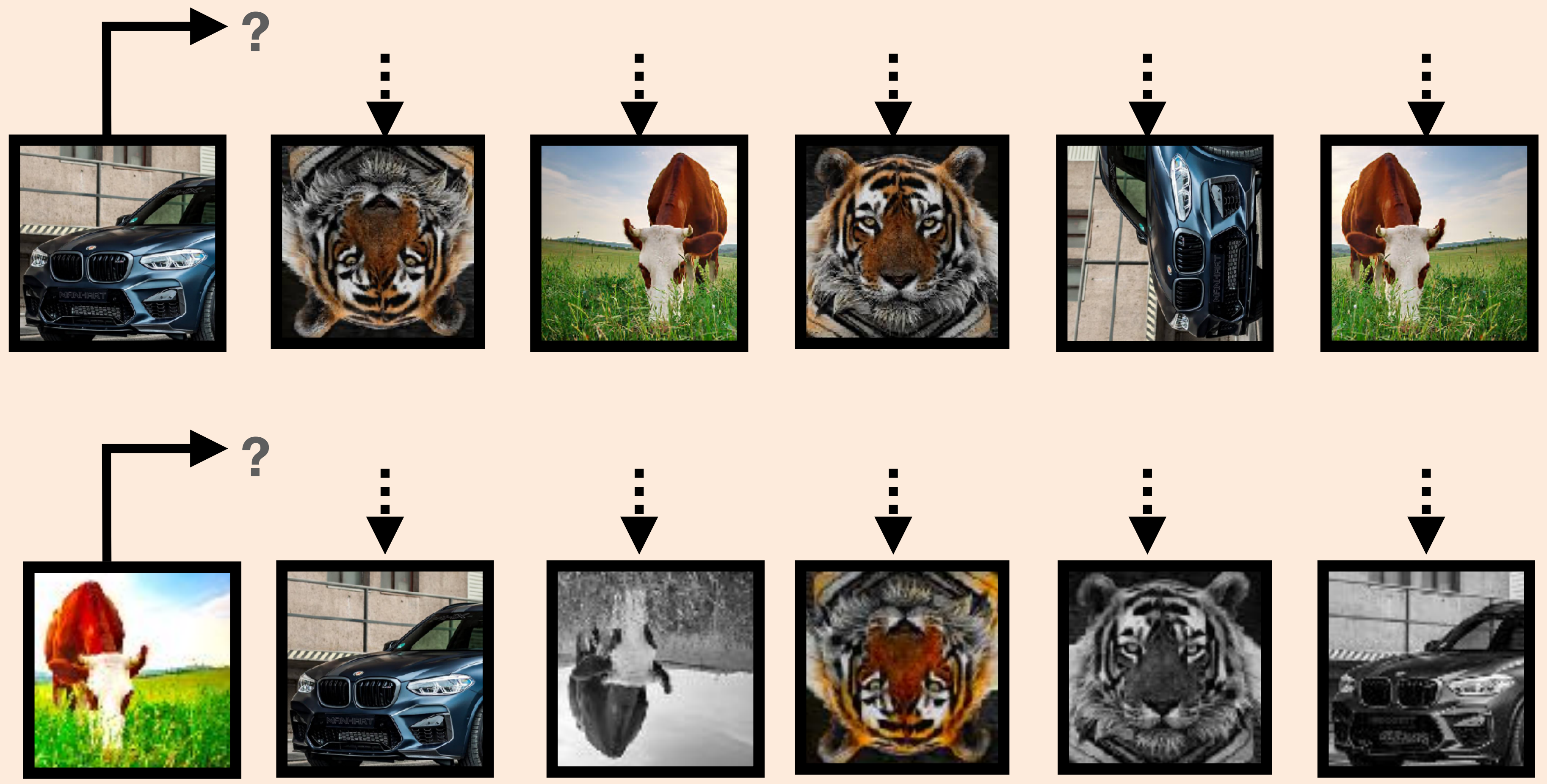
Data

Bormio is the annual meeting of nuclear physicists in Italy

Puzzles

[?] is the annual meeting of nuclear physicists in [?]

Bormio is an annual [?] of [?] physicists in Italy



Fill in the Blank

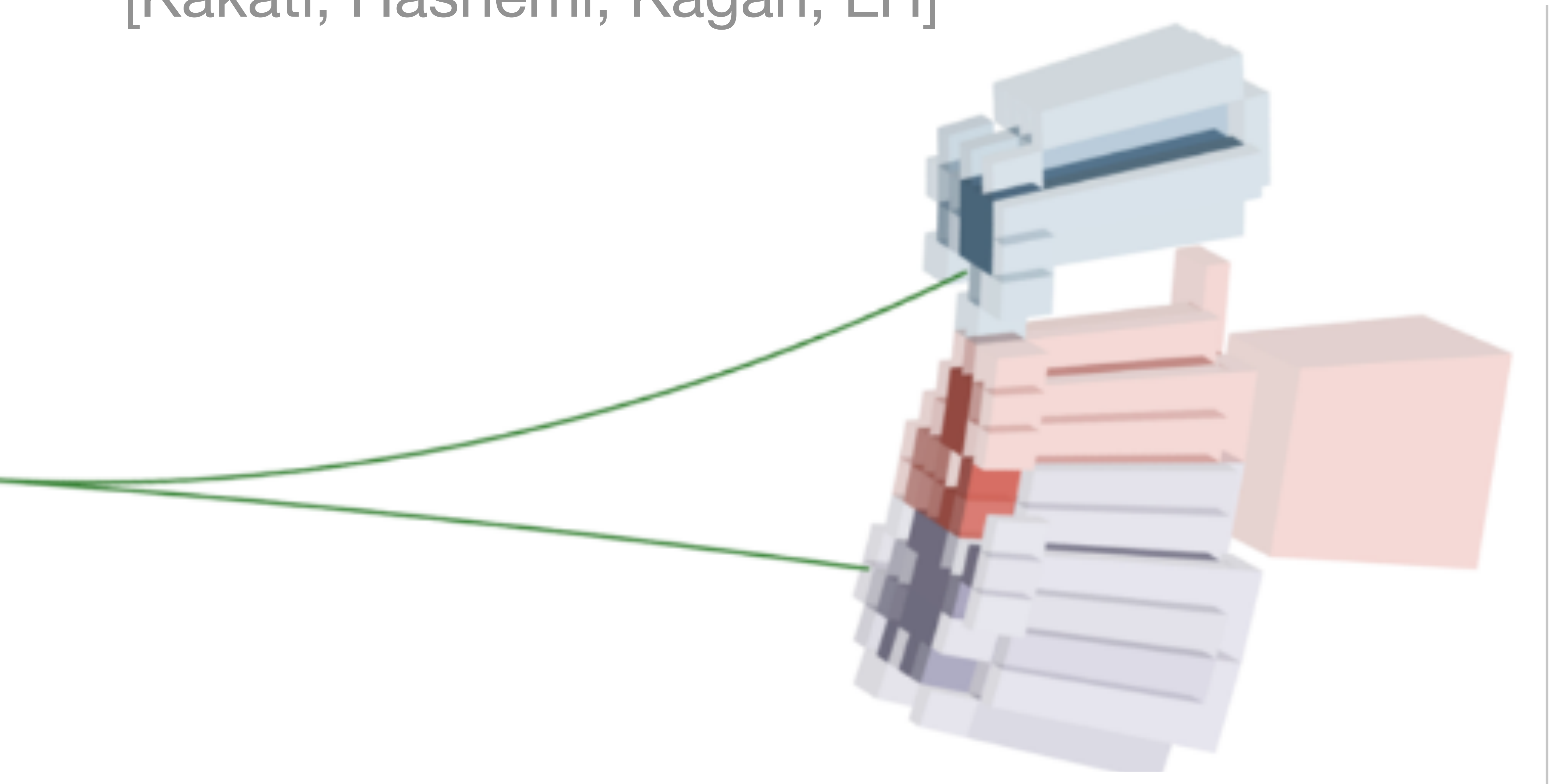
Association Games

Foundation Models in HEP

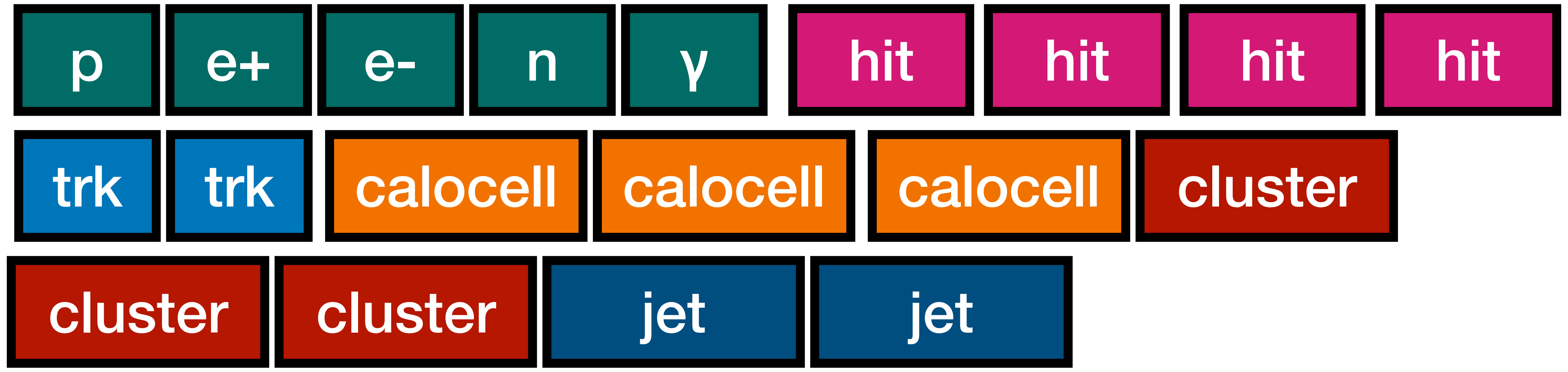
We can do the same “fill in the blank” idea in particle physics

Idea: our data is like a “narrative”, we can treat it like text

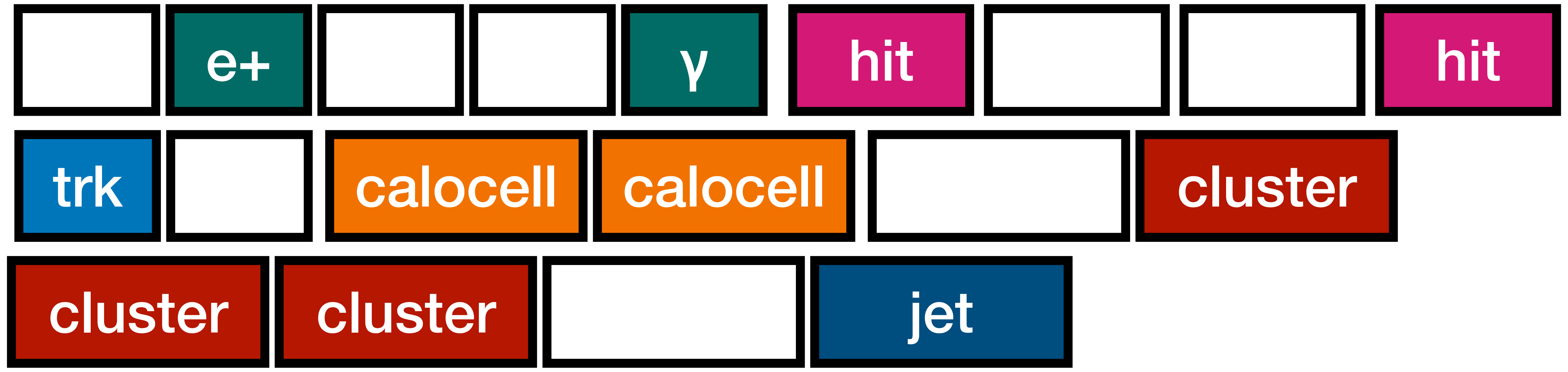
[Kakati, Hashemi, Kagan, LH]



***4M:** Massively multi-modal masked modelling*

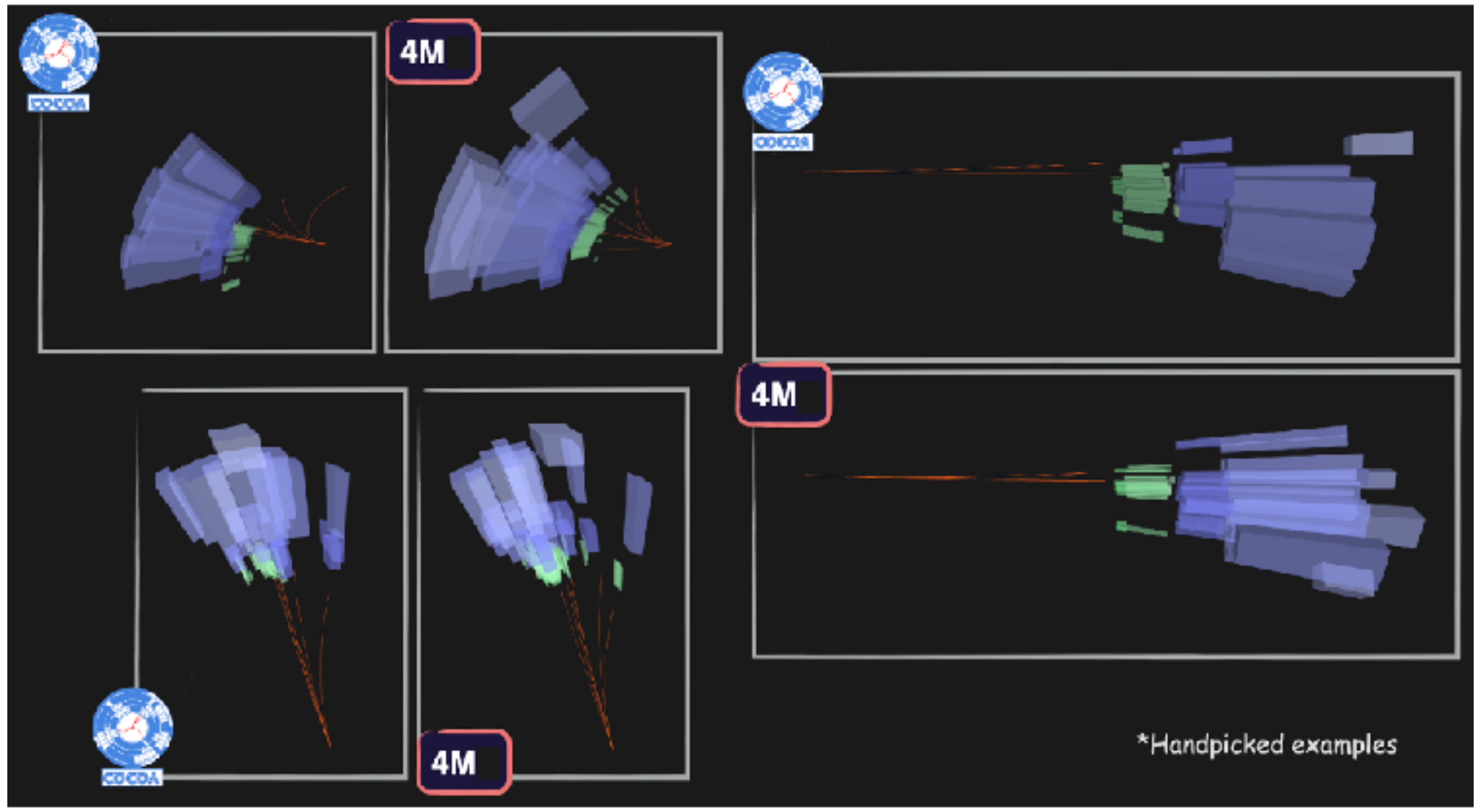


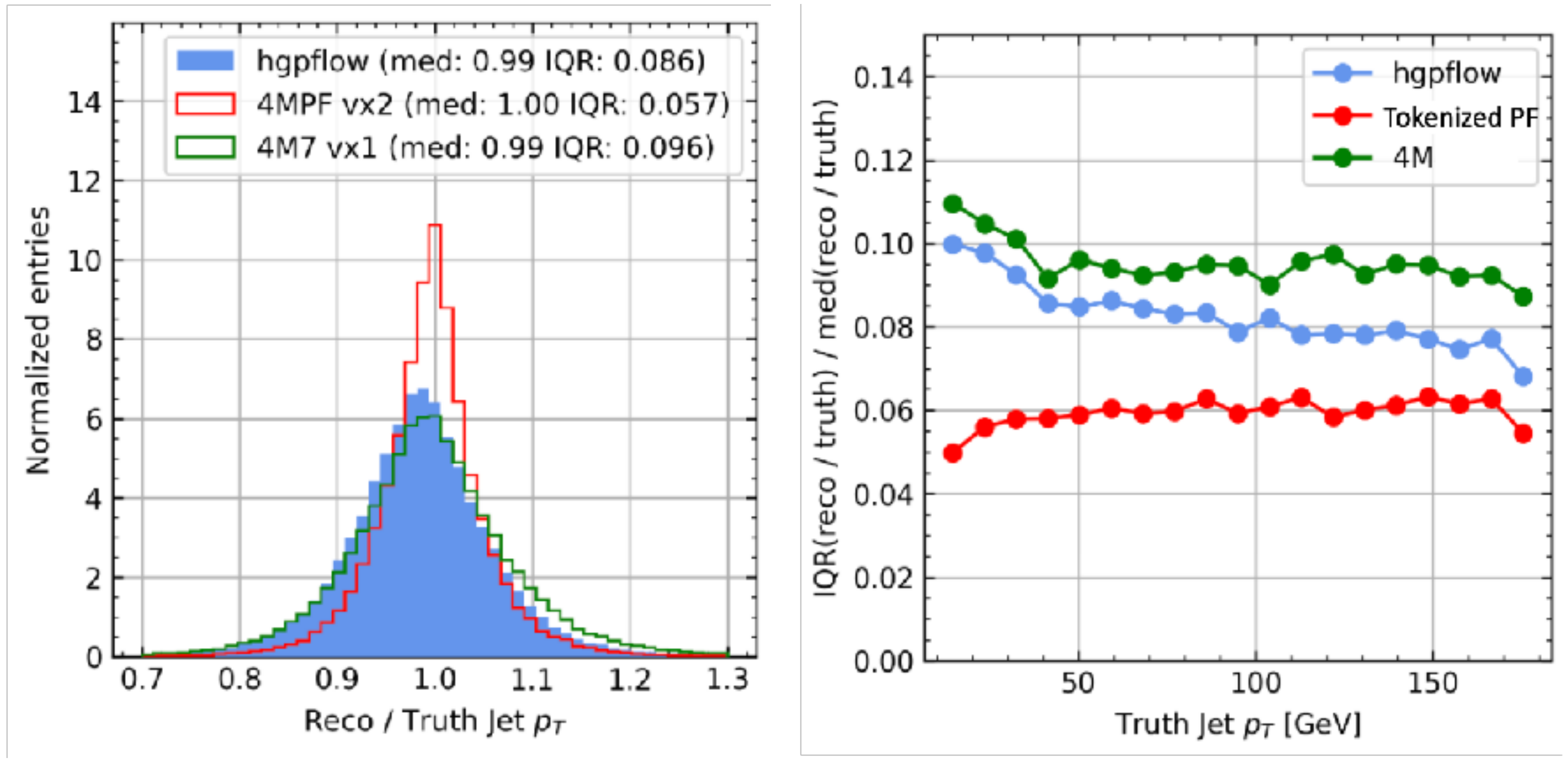
→ “fill in the blank” on this “text”



Foundation Models in HEP

When trained, we can do many things **out of the box**
→ one big general model vs. many small dedicated ones





p | e+ | e- | n | γ

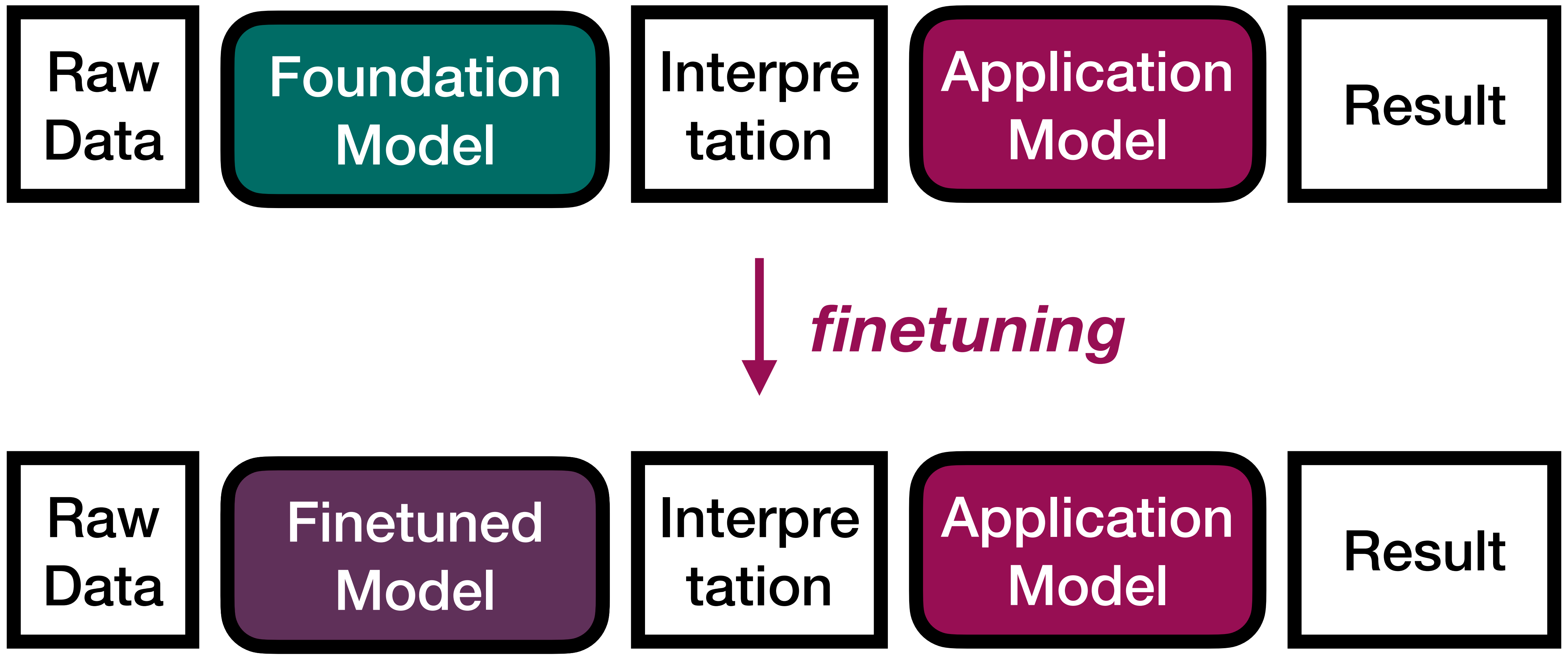
Detector Simulation
(given particles generate the rest)

trk | trk
calocell | calocell | calocell | cluster

Particle Reconstruction
(given tracks / calo infer particles)

Foundation Models in HEP

The best vocabulary to use **depends on the question** you ask.
→ Enables **Rapid Adaptation / Reoptimization** to a new task

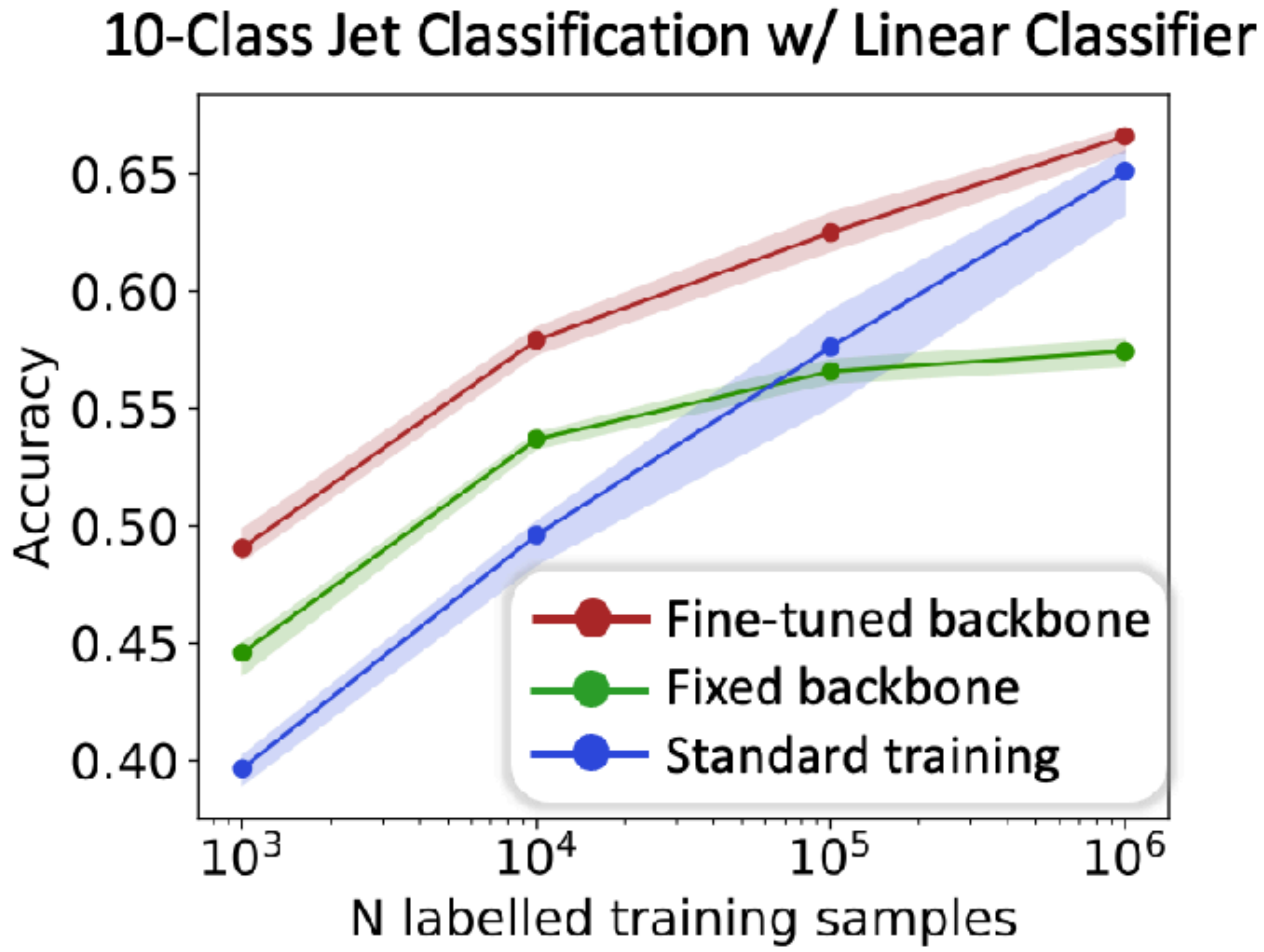


Possible thanks to differentiable (optimizable) nature of neural networks
→ key advantage over fixed human "Foundation Model"

Foundation Models in HEP

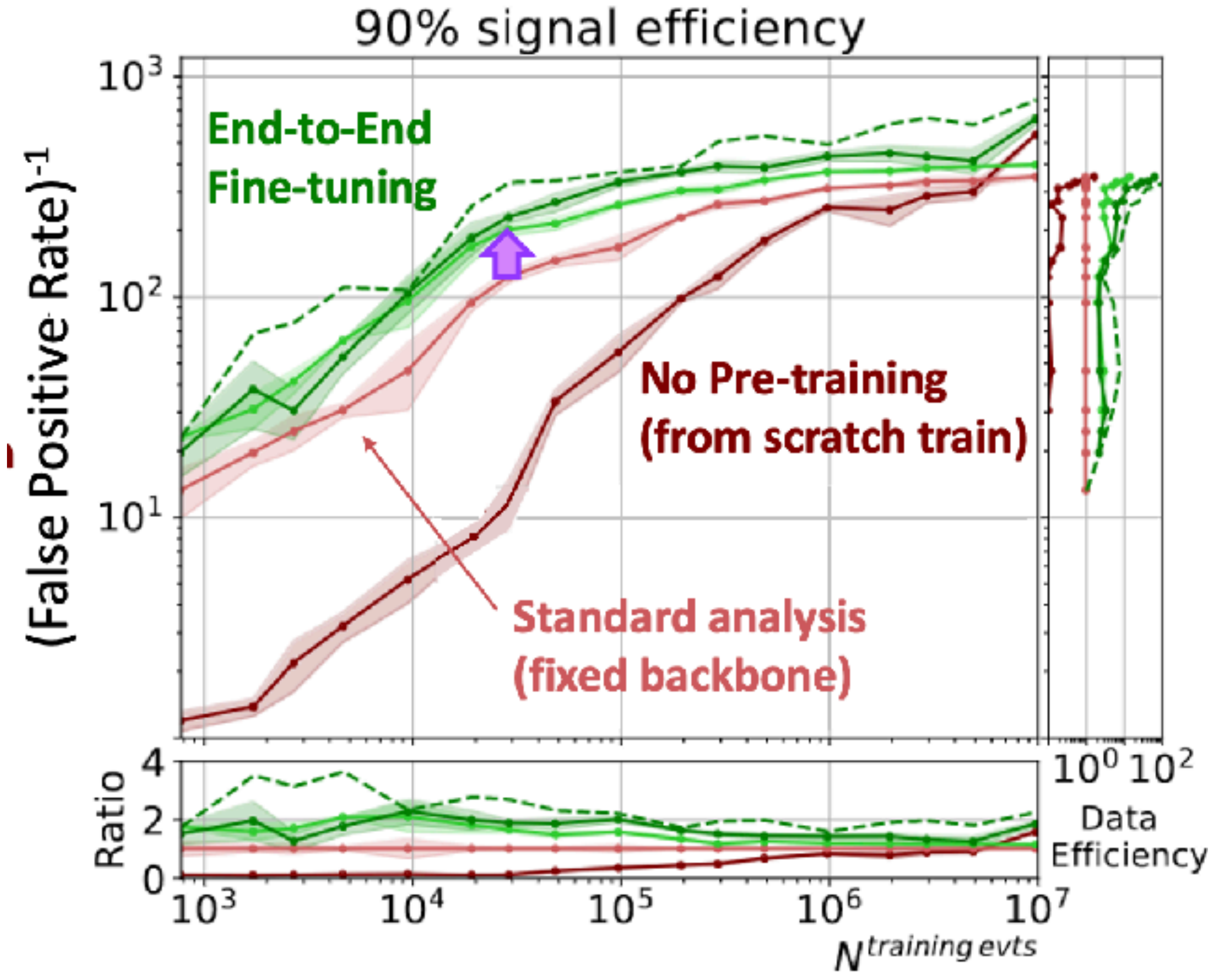
The best vocabulary to use **depends on the question** you ask.
→ Enables **Rapid Adaptation / Reoptimization** to a new task

[Klein, ..., Golling, Kagan, LH]



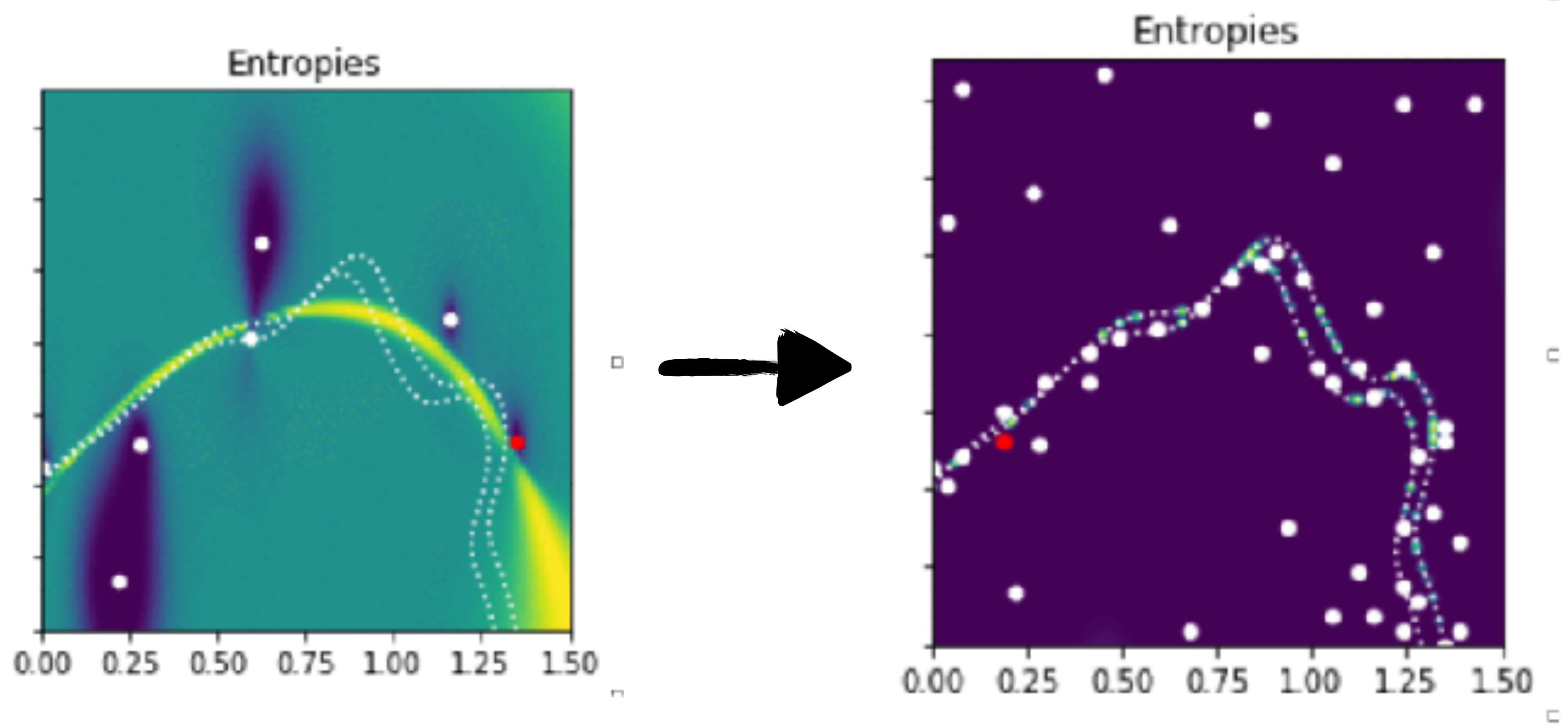
Jet Tagging

[Vigl, Hartman, LH]

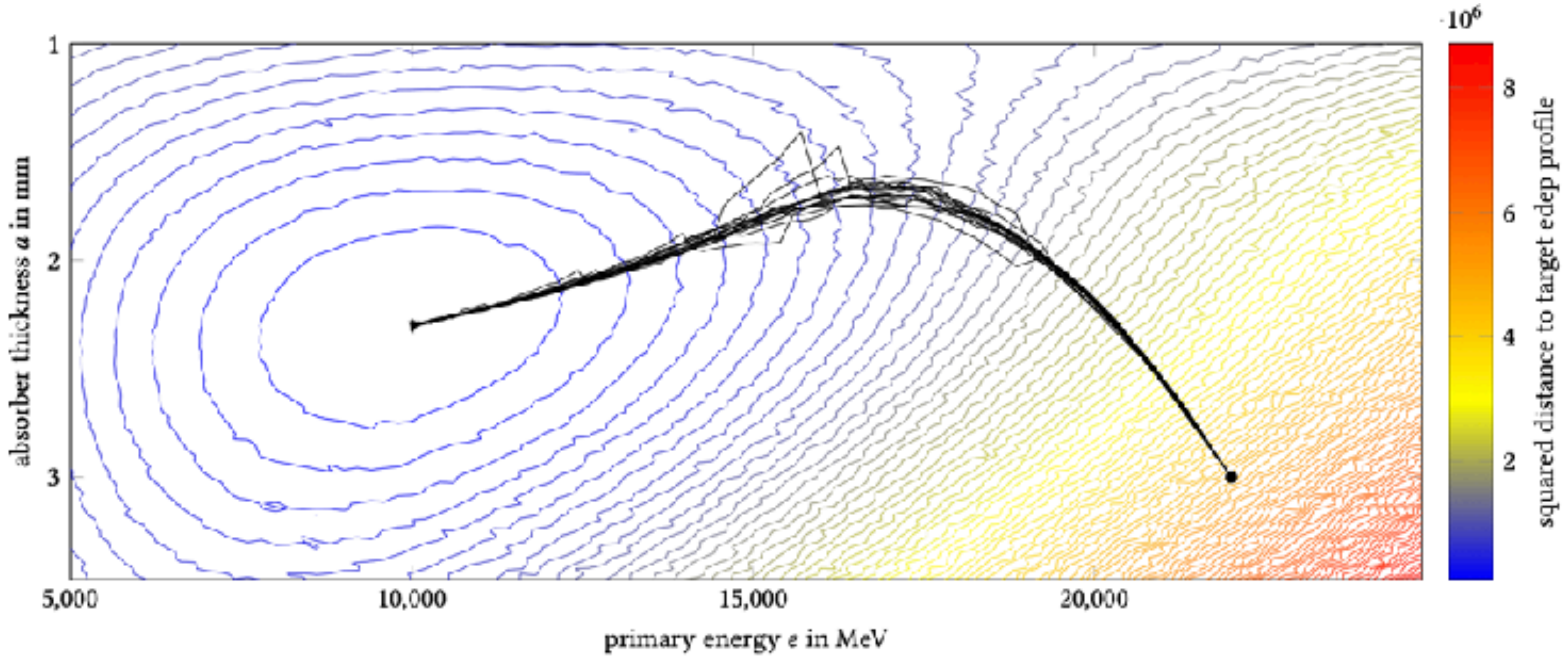


HH BSM Scenario

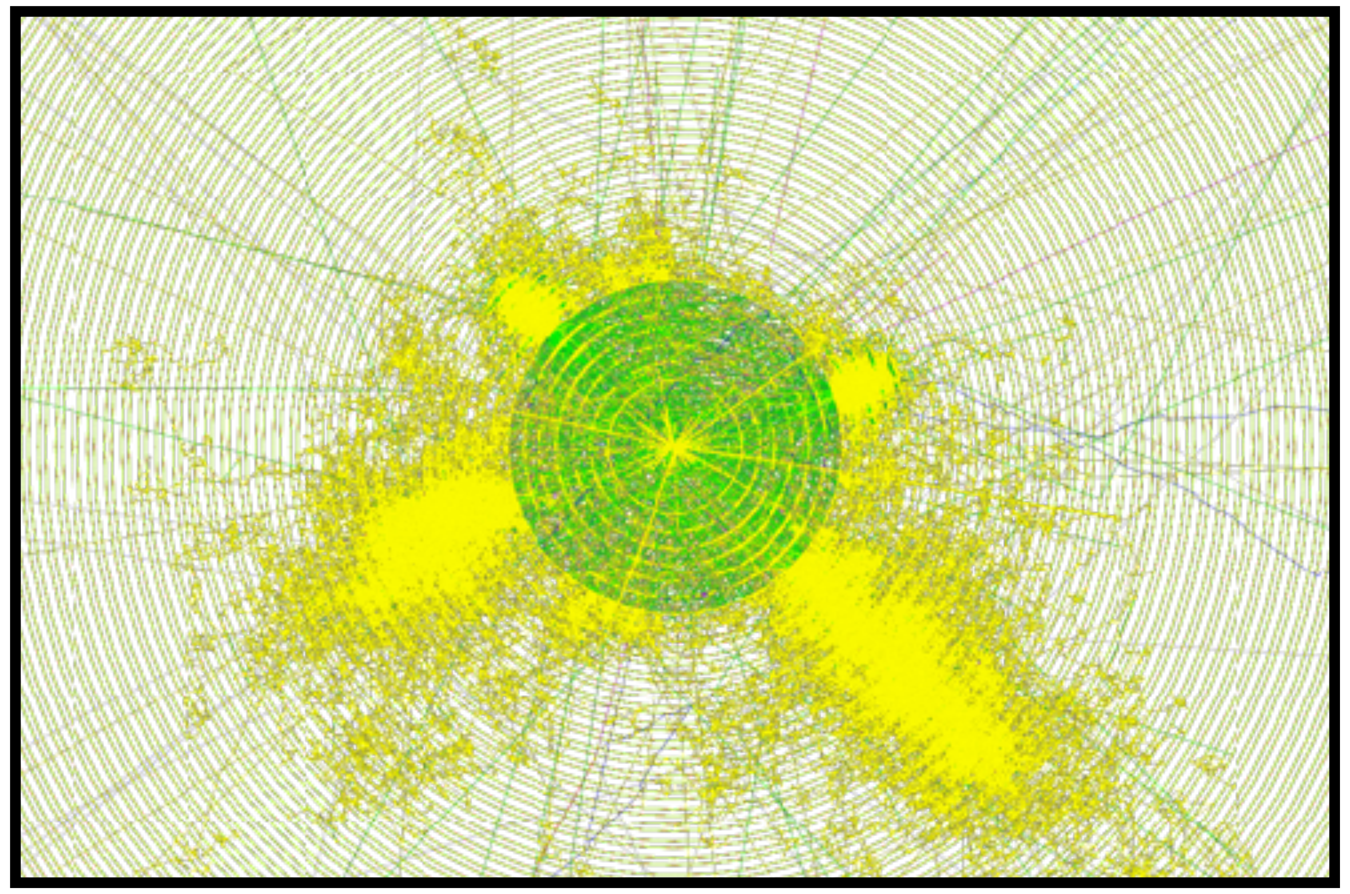
Exciting Topics I'd love to Have Time for



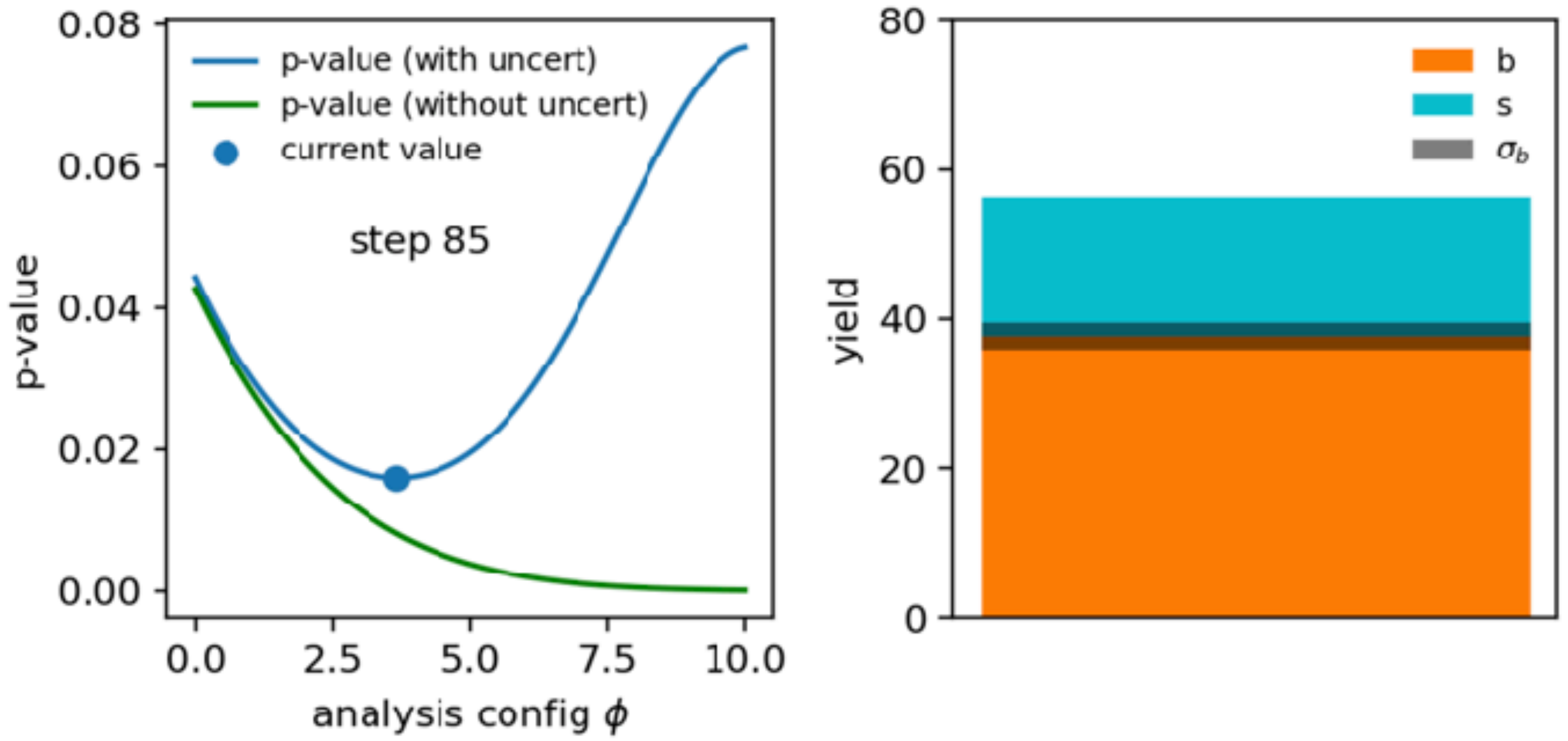
AI-driven Theory Exploration



AI-driven Detector Design



AI-enhanced general-purpose simulation



Uncertainty-aware AI

“We are at the cusp of something exhilarating and terrifying”

- David Bowie on the rise of the Internets (1999)

“ChatGPT doesn’t have a clue”

“The Power of Automation”

- Stafano Frixione, yesterday

Credit: Masayoshi Sukita

Resummation of the C-Parameter Sudakov Shoulder Using Effective Field Theory

Matthew D. Schwartz[1,2]

[1] *Department of Physics, Harvard University, Cambridge, MA 02138, USA*

[2] *Institute for Artificial Intelligence and Fundamental Interactions (IAIFI)*

`schwartz@g.harvard.edu`

January 7, 2026

Abstract

The C-parameter distribution in e^+e^- annihilation exhibits a kinematic shoulder at $C = 3/4$, where three-parton final states reach their maximum and a fourth parton is required to exceed it. This boundary generates large logarithms that must be resummed. Using soft-collinear effective theory, we derive a factorization theorem involving new jet and soft functions specific to the C-parameter measurement, in which soft radiation contributes quadratically in transverse momentum. This quadratic structure explains the step discontinuity at leading order. We compute all ingredients at one loop, validate against Monte Carlo, and present matched NLL+NLO results. Unlike thrust and heavy jet mass, the C-parameter has no Sudakov–Landau pole, making momentum-space resummation straightforward.

Resummation of the C-Parameter Sudakov Shoulder Using Effective Field Theory

Matthew D. Schwartz[1,2]

[1] *Department of Physics, Harvard University, Cambridge, MA 02138, USA*

[2] *Institute for Artificial Intelligence and Fundamental Interactions (IAIFI)*

schwartz@g.harvard.edu

AI RESEARCH ASSISTANT: Claude Opus 4.5 (Anthropic)

January 7, 2026

Abstract

The C-parameter distribution in e^+e^- annihilation exhibits a kinematic shoulder at $C = 3/4$, where three-parton final states reach their maximum and a fourth parton is required to exceed it. This boundary generates large logarithms that must be resummed. Using soft-collinear effective theory, we derive a factorization theorem involving new jet and soft functions specific to the C-parameter measurement, in which soft radiation contributes quadratically in transverse momentum. This quadratic structure explains the step discontinuity at leading order. We compute all ingredients at one loop, validate against Monte Carlo, and present matched NLL+NLO results. Unlike thrust and heavy jet mass, the C-parameter has no Sudakov–Landau pole, making momentum-space resummation straightforward. All calculations, numerical analysis, and manuscript preparation were performed by Claude, an AI assistant developed by Anthropic, working under physicist supervision.

Fast Simulation

We might be entering an era of unprecedented automation
→ must try if you haven't. A bit of an unsettling experience

Terence Tao
@tao@mathstodon.xyz

Recently, the application of AI tools to Erdos problems passed a milestone: an Erdos problem (#728 erdosproblems.com/728) was solved more or less autonomously by AI (after some feedback from an initial attempt), in the spirit of the problem (as reconstructed by the Erdos problem website community), with the result (to the best of our knowledge) not replicated in existing literature (although similar results proven by similar methods were located).

This is a demonstration of the genuine increase in capability of these tools in recent months, and is largely consistent with other recent demonstrations of AI using existing methods to resolve Erdos problems, although in most previous cases a solution to these problems was later located in the literature, as discussed in mathstodon.xyz/deck/@tao/11578... . This particular case was unusual in that the problem as stated by Erdos was misformulated, with a reconstruction of the problem in the intended spirit only

AI to solve open number theory problems

```
> can we compare a few options 5x5, 10x10, 15x15, 20x20

● Write(som_size_comparison.py)
  └ Wrote 113 lines to som_size_comparison.py
      """Compare SOM with different grid sizes: 5x5, 10x10, 15x15, 20x20."""
      import numpy as np
      import pandas as pd
      import matplotlib.pyplot as plt
      from sklearn.preprocessing import StandardScaler
      from minisom import MiniSom
      # Load data
      df = pd.read_csv('classified_emitters_reduced.csv')
      features = ['antibunching_ns', 'bunching_ns', 'min_g2', 'HBTEventsPerSecond', 'num_peaks',
    … +103 lines (ctrl+o to expand)

● Bash(pixi run python som_size_comparison.py) timeout: 3m 0s
  └ Training 5x5 SOM...
    Training 10x10 SOM...
    Training 15x15 SOM...
    … +2 lines (ctrl+o to expand)

● Done. som_size_comparison.png shows 4 columns (5x5, 10x10, 15x15, 20x20) with 4 rows each:

  What you'll see:
```

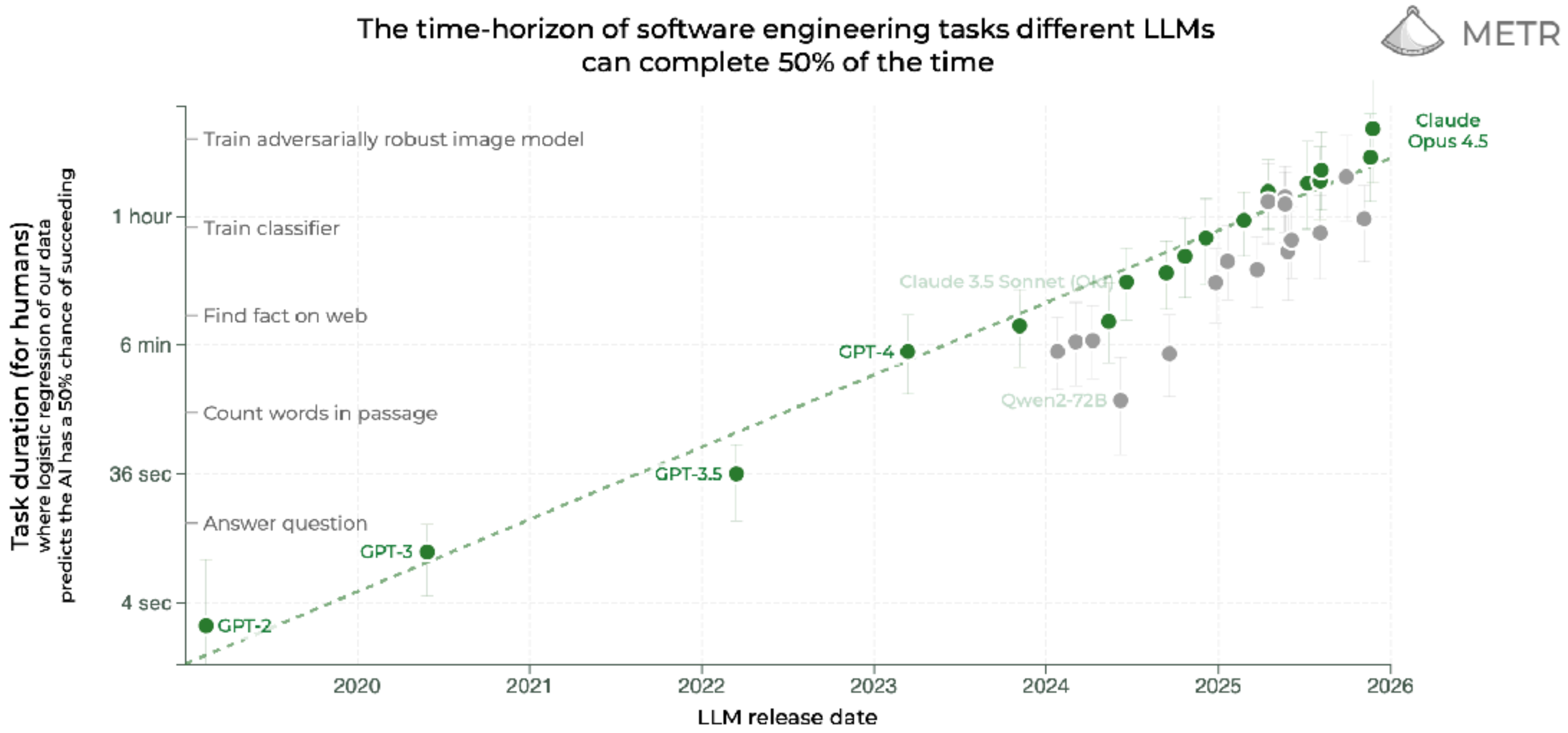
Grid Size	Neurons	Samples/Neuron (avg)	Characteristics
5x5	25	~136	Very coarse – clusters forced together, hard to see boundaries
10x10	100	~34	Moderate – main structure visible, some cluster overlap
15x15	225	~15	Good balance – clear cluster separation, visible U-matrix ridges
20x20	400	~8	Fine detail – may start overfitting, sparse hit map

Tradeoffs:
- **Too small** (5x5): Multiple clusters collapse into same neurons, lose resolution
- **Too large** (20x20): Many empty neurons, U-matrix becomes noisy, harder to interpret
- **Sweet spot** (10x10 to 15x15): Clear topology with meaningful density patterns

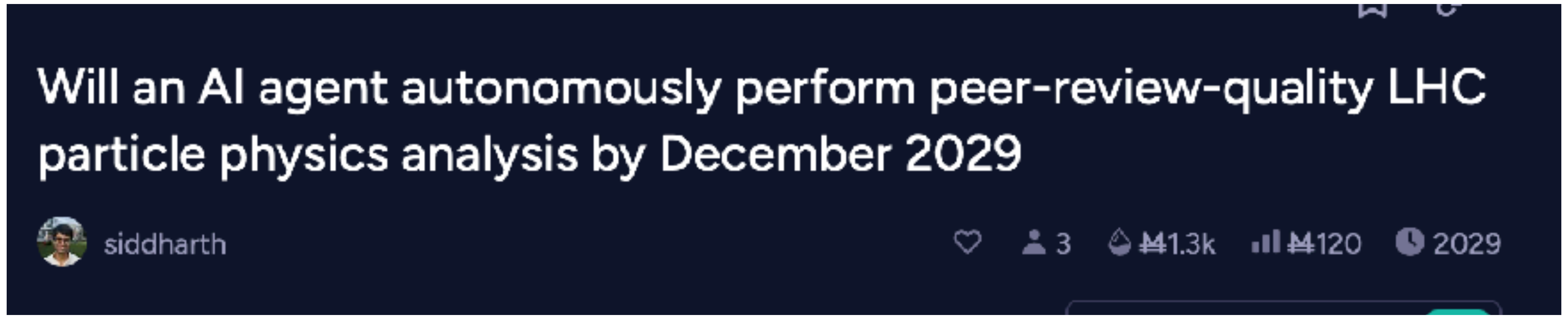
AI to analyze data in near-autonomous fashion

AI Agents

Frontier AI Labs estimate e.g. O(month) long human tasks can be done autonomously by AI in a few years (7 month doubling)



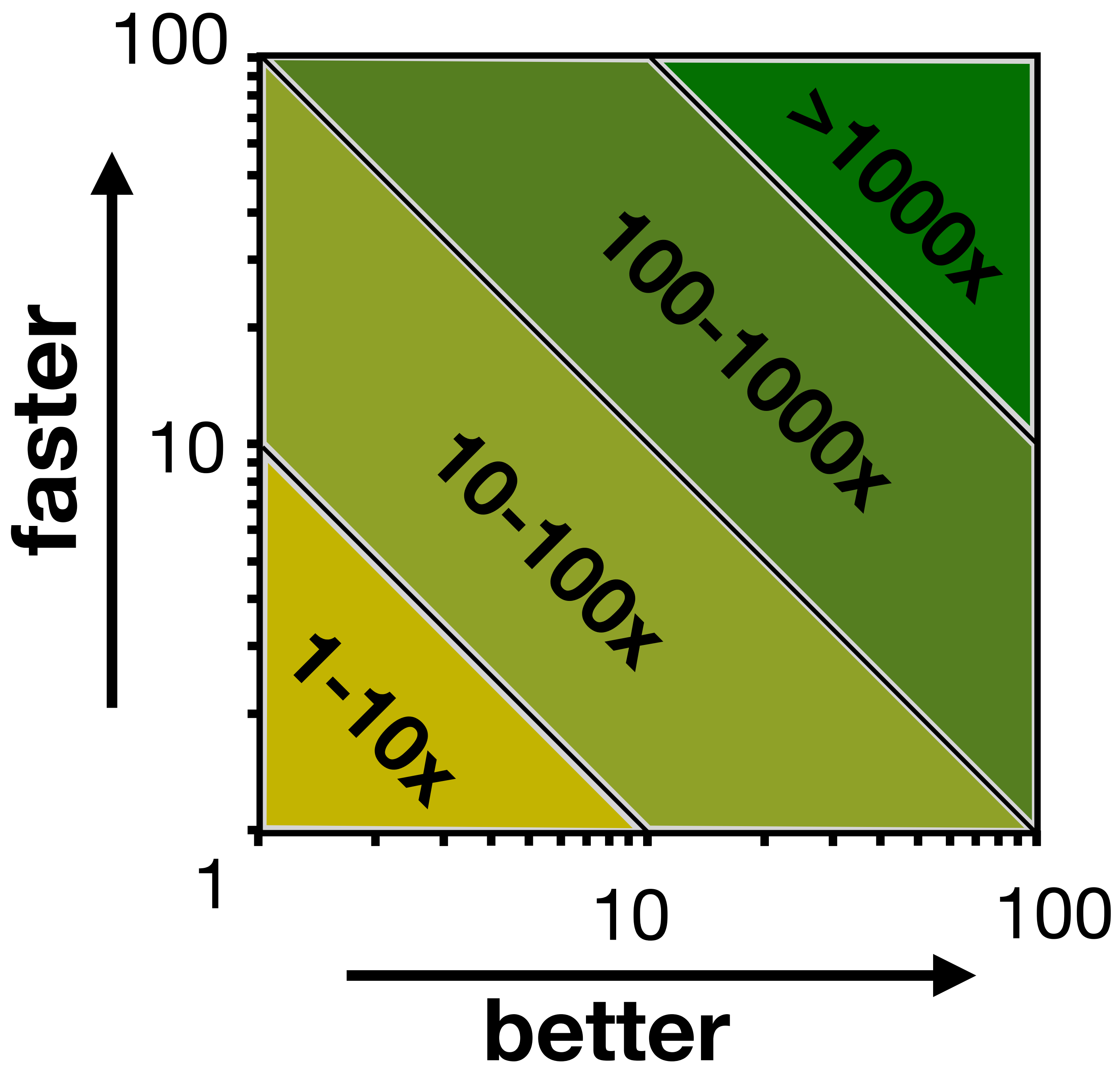
I bet against AI optimists & fairly sure I win, but not 100%



Takeaways

We know what we're doing, but it's not enough. We must find ways to massively improve and accelerate our research

→ **find the AlphaFold moment for HEP**



We found out, that we can still be orders of magnitude better at the LHC

→ requires 100-1000x in scaling models/data

AI Tooling as the infrastructure for extreme automation

→ enable a new paradigm of data/theory exploration

Acknowledgements

Group

Collab

M. Kagan

R. Osadchy

T. Golling

S. Klein

M. Leigh

N. Kakati

B. Hashemi