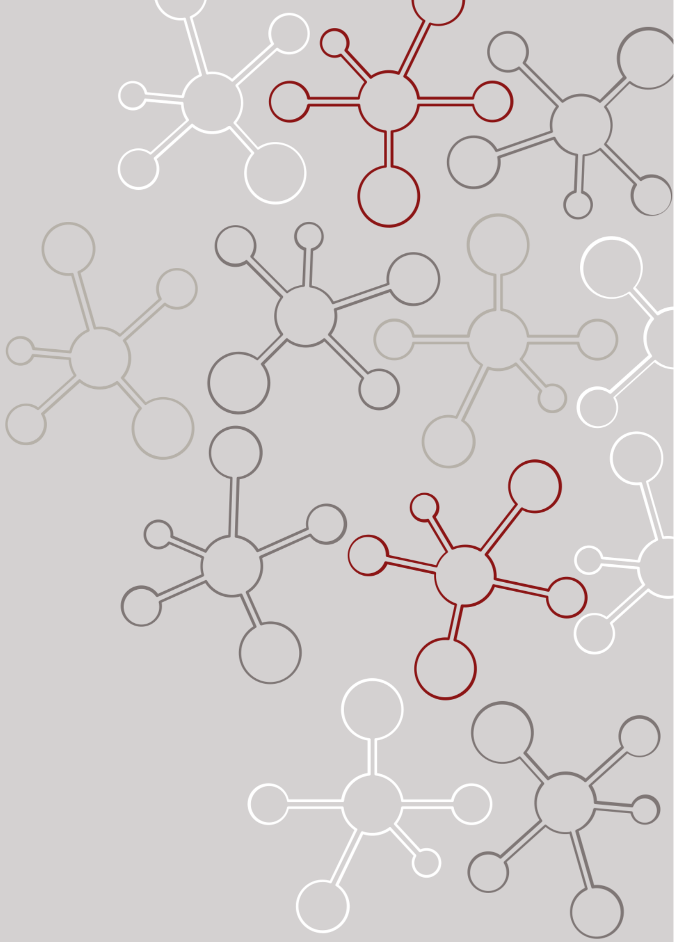
New Frontiers in AI for Fundamental Physics



Michael Kagan, SLAC

1st TREASURE Workshop at BNL
April 28, 2026

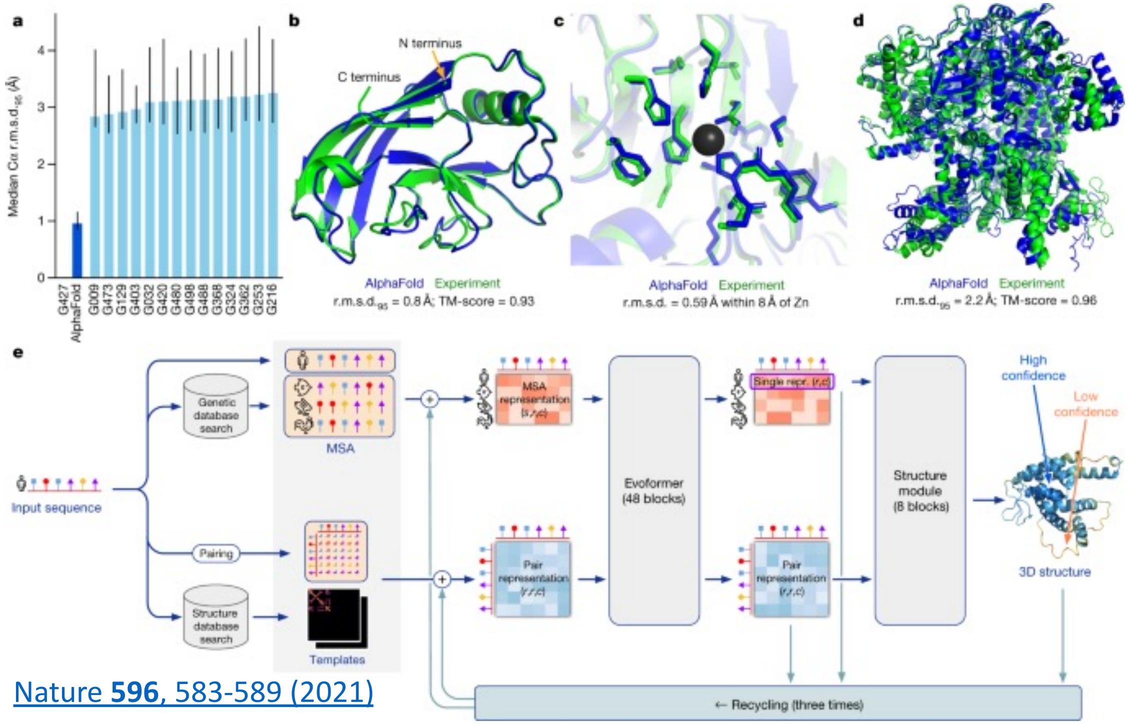
SLAC NATIONAL ACCELERATOR LABORATORY

Stanford University | U.S. DEPARTMENT OF ENERGY

AI is everywhere in science

Energy Department Launches 'Genesis Mission' to Transform American Science and Innovation Through the AI Computing Revolution

RAISE sets the stage for Europe's next chapter in AI-accelerated discovery



Nature **596**, 583-589 (2021)

Nobel Prize in Chemistry 2024

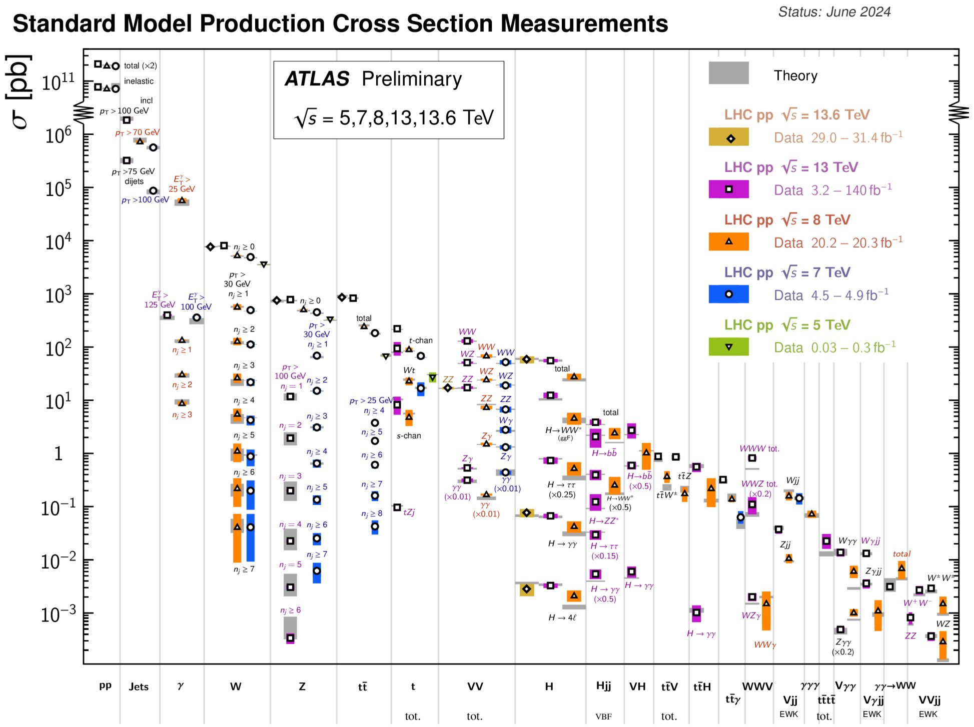
© Nobel Prize Outreach. Photo: Clément Morin
Demis Hassabis
Prize share: 1/4

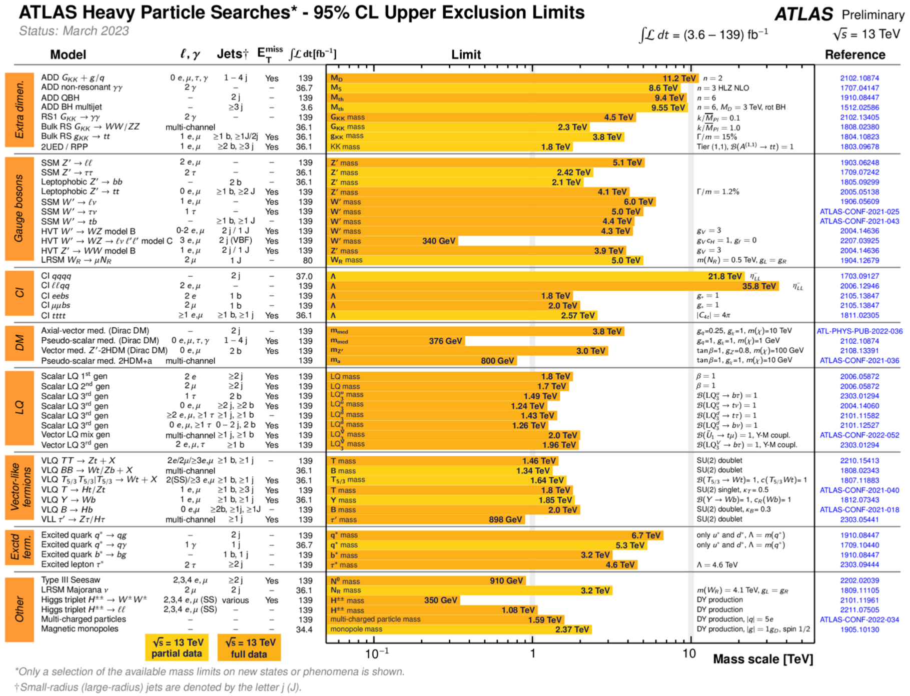
© Nobel Prize Outreach. Photo: Clément Morin
John Jumper
Prize share: 1/4

AI in science is everywhere… how might fundamental physics change?

>15 years of amazing results at the LHC

Yet a typical measurement or search analysis takes 2-3 years…





AI in science is everywhere… how might fundamental physics change?

>15 years of amazing results at the LHC

Yet a typical measurement or search analysis takes 2-3 years…

How would x10 – x100 speedup in time-to-result change the field?
i.e. analysis only takes a day, or a week?

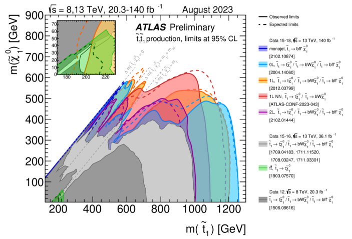
- Less time on mechanics of data analysis
- More time on questions, hypotheses, interpretations
- *Revolutionize the way we do science*

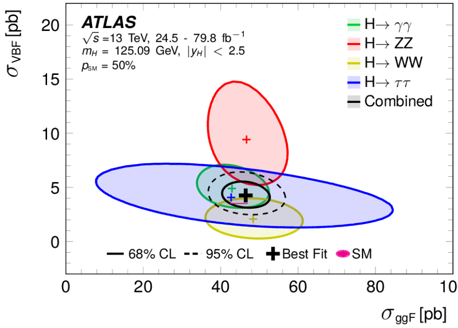
What might that look like?

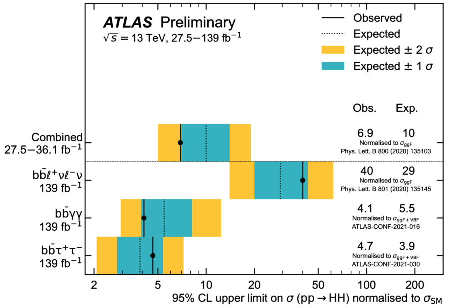
Theory

Simulation

Data Analysis







⋮

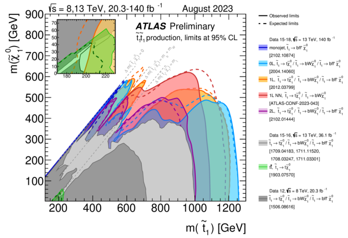
What might that look like?

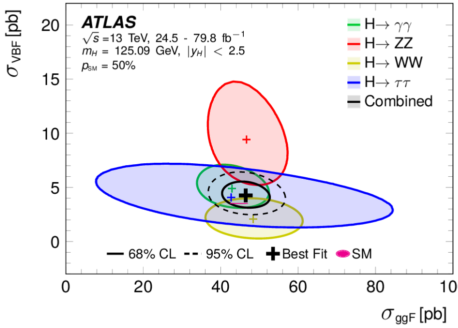
Theory

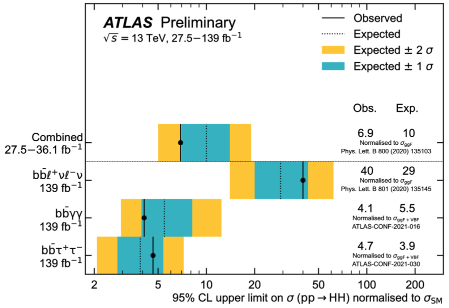
Simulation

Data Analysis

Per-Theory-Point (hypothesis)
Fine-tuning and optimization

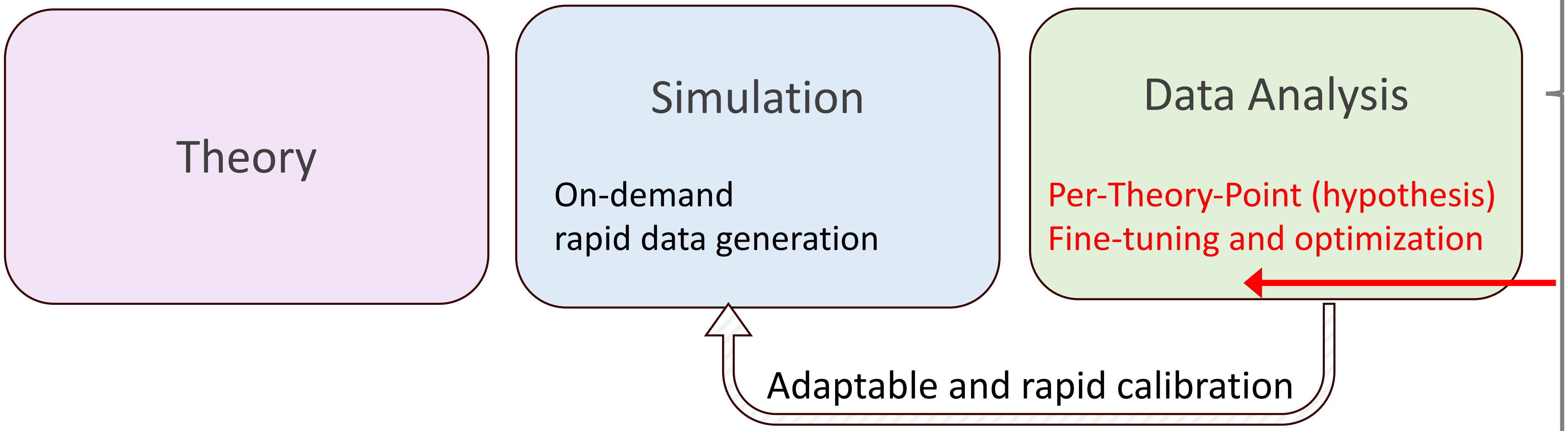






⋮

What might that look like?



Theory

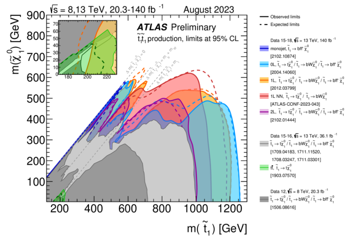
Simulation

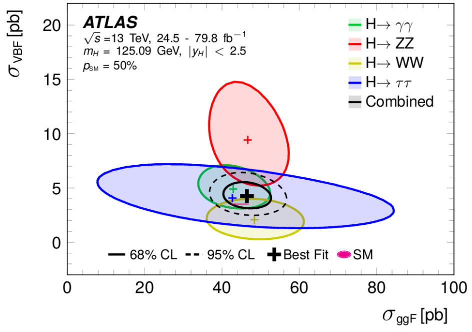
On-demand
rapid data generation

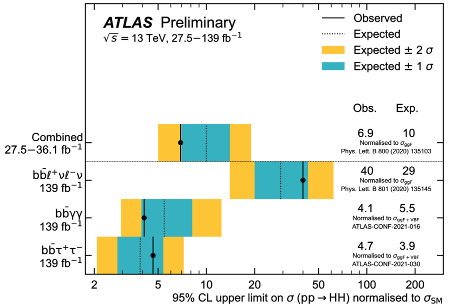
Data Analysis

Per-Theory-Point (hypothesis)
Fine-tuning and optimization

Adaptable and rapid calibration

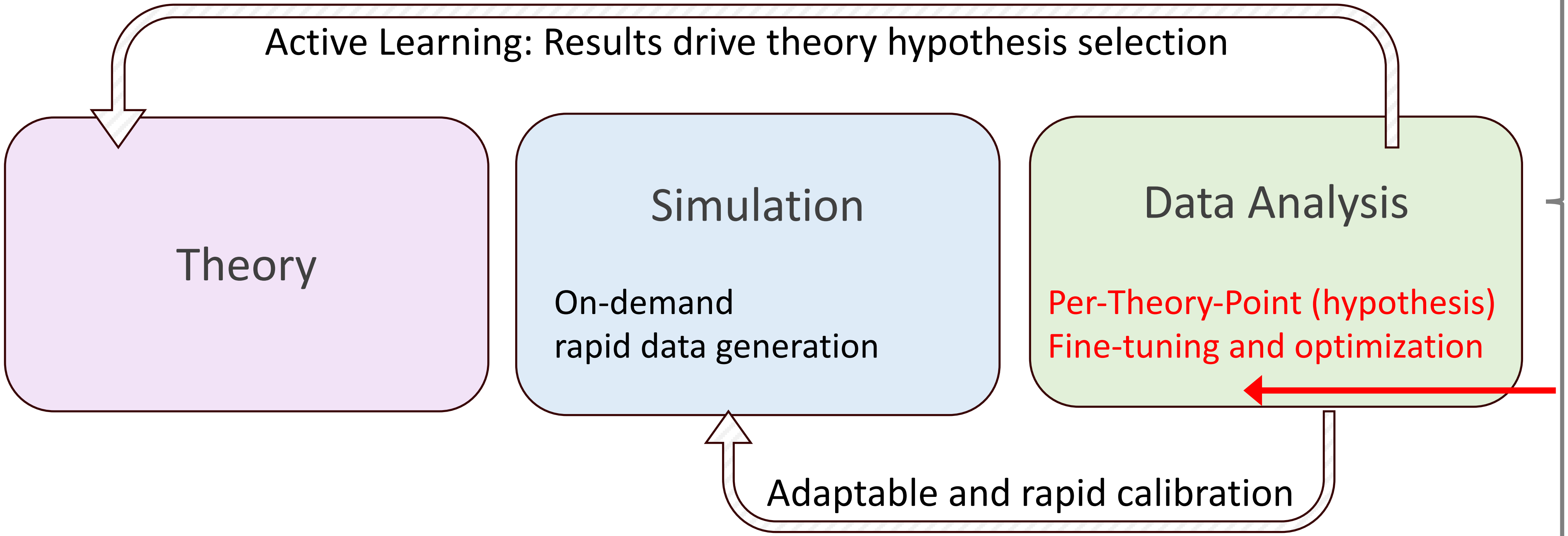






⋮

What might that look like?



Active Learning: Results drive theory hypothesis selection

Theory

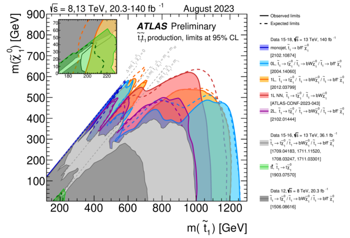
Simulation

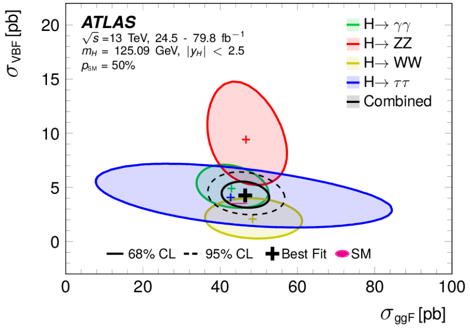
On-demand
rapid data generation

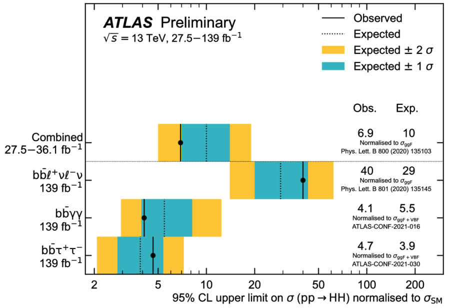
Data Analysis

Per-Theory-Point (hypothesis)
Fine-tuning and optimization

Adaptable and rapid calibration



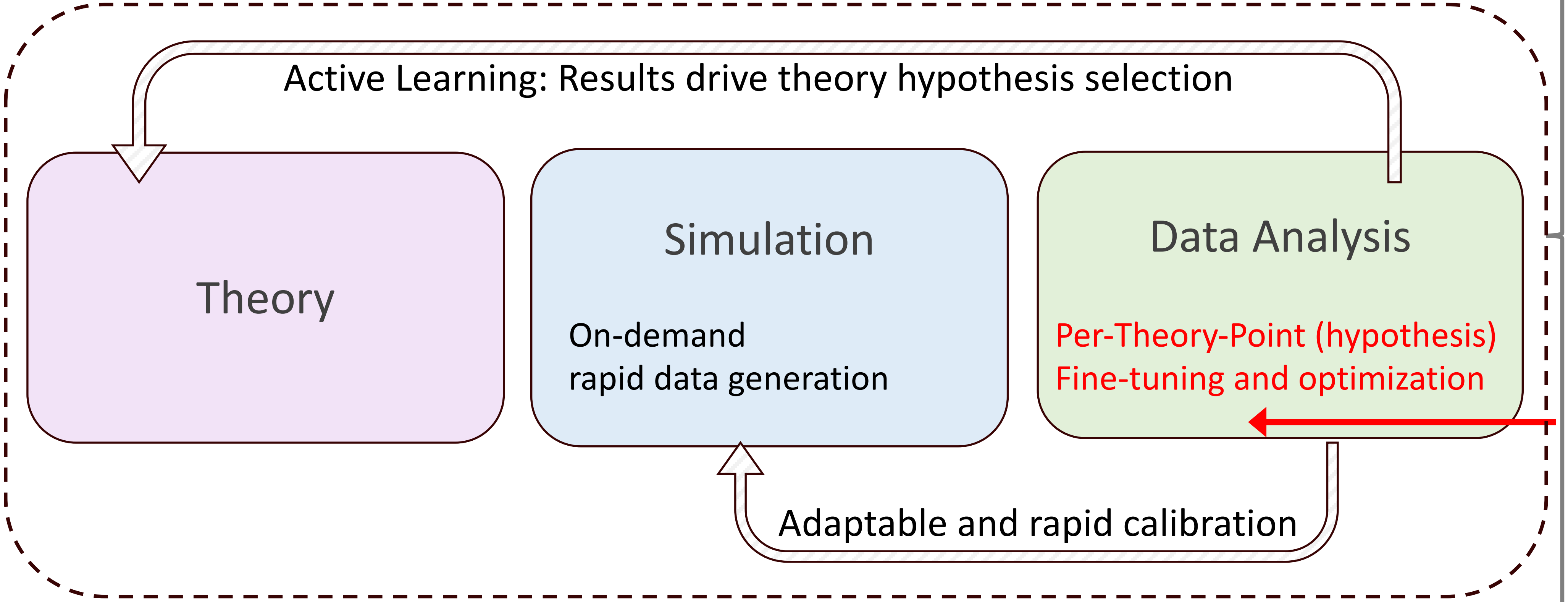




⋮

What might that look like?

Automated workflows → Agentic science



Active Learning: Results drive theory hypothesis selection

Theory

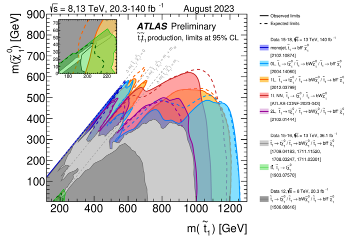
Simulation

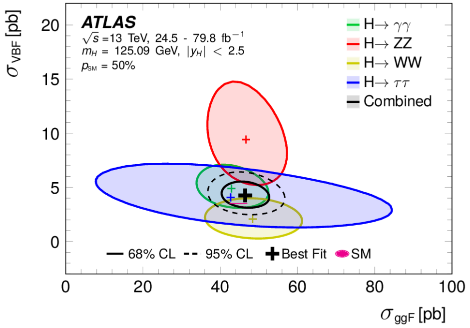
On-demand
rapid data generation

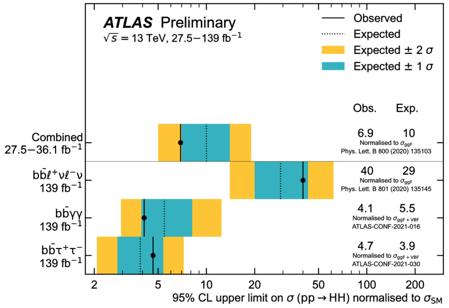
Data Analysis

Per-Theory-Point (hypothesis)
Fine-tuning and optimization

Adaptable and rapid calibration







⋮

Systems to do rapid, optimal, and automated searches through lots of data and theory space

How might we get there?

Tools that adapt to our data analysis goals → ML is the infrastructure

Optimizable data analysis → Analysis-as-fine-tuning

Rapid and on-demand data simulation

Adaptable calibration

Each is an exciting research frontier!
Not enough time to discuss all!

Active learning to drive hypothesis selection and parameter scans

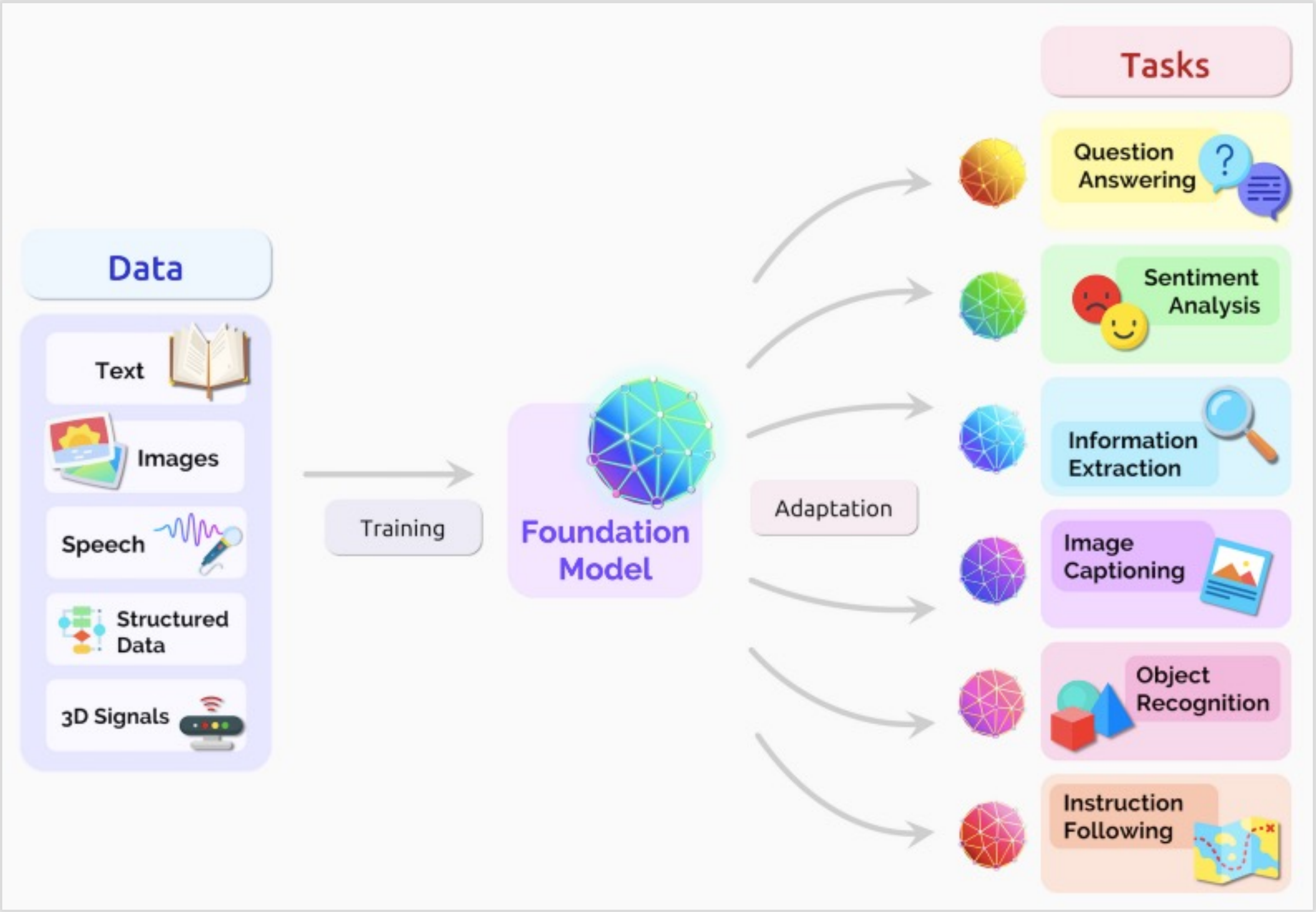
Automating processes through workflows, active learning and agents

What are the optimizable and fine-tune-able data analysis tools?

Foundation models

They are the ML infrastructure, the reusable, adaptable and generalized tools that propel these workflows

What is a foundation model?

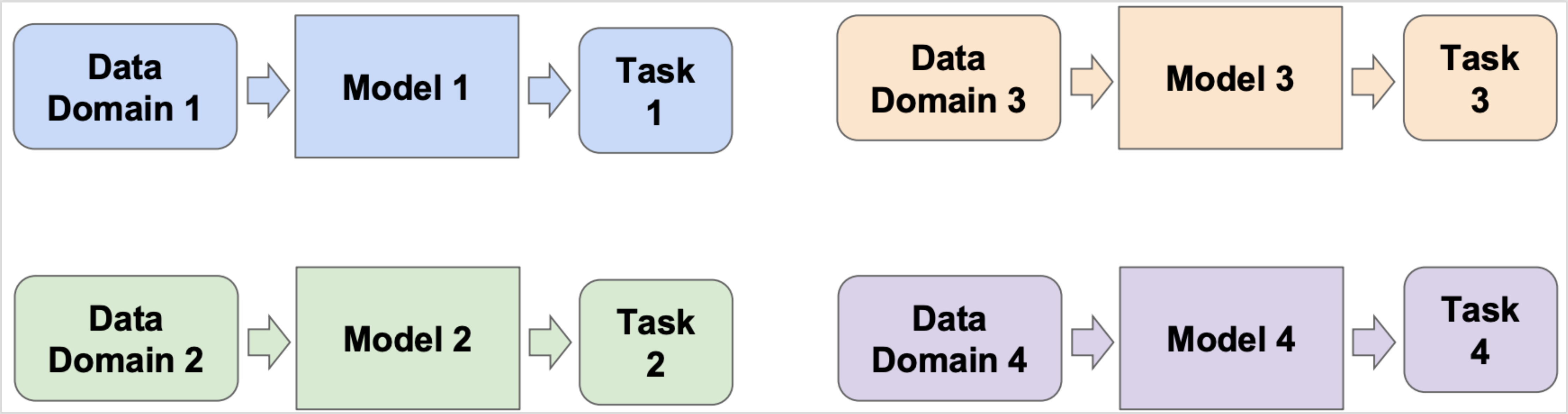


Prompt: `Several giant wooly mammoths approach treading through a snowy meadow […]` OpenAI Sora

Image credit: 2108.07258

How we typically think about ML models

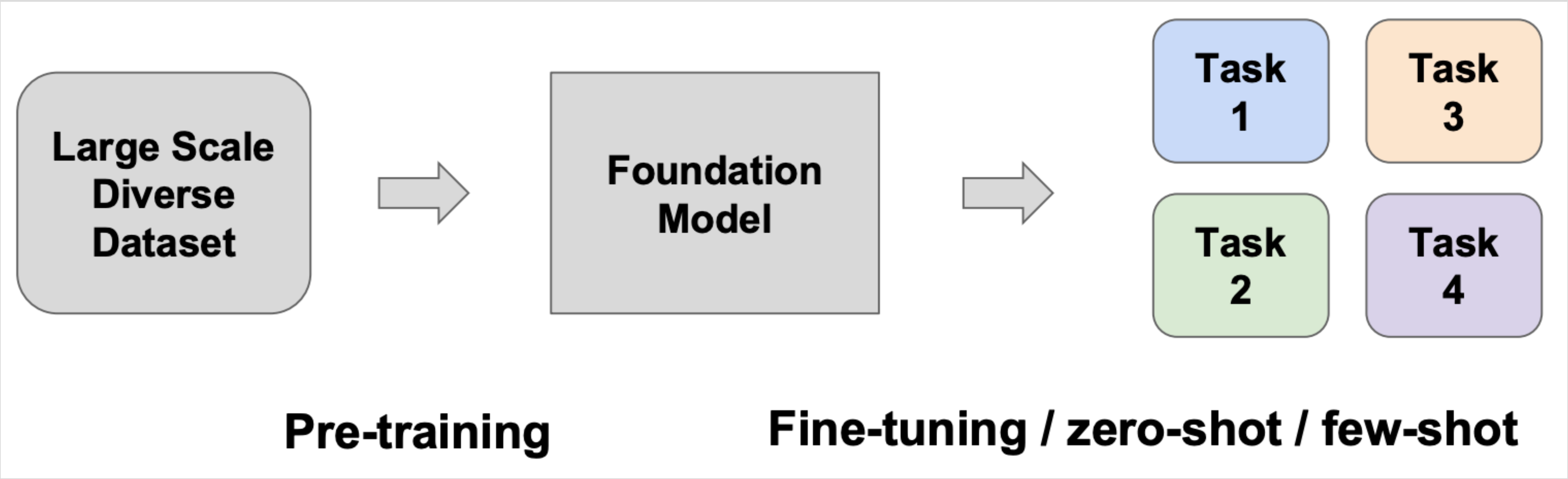
Train a specialized model to solve each task (HEP right now)



Bespoke solutions

Foundation models take a different approach

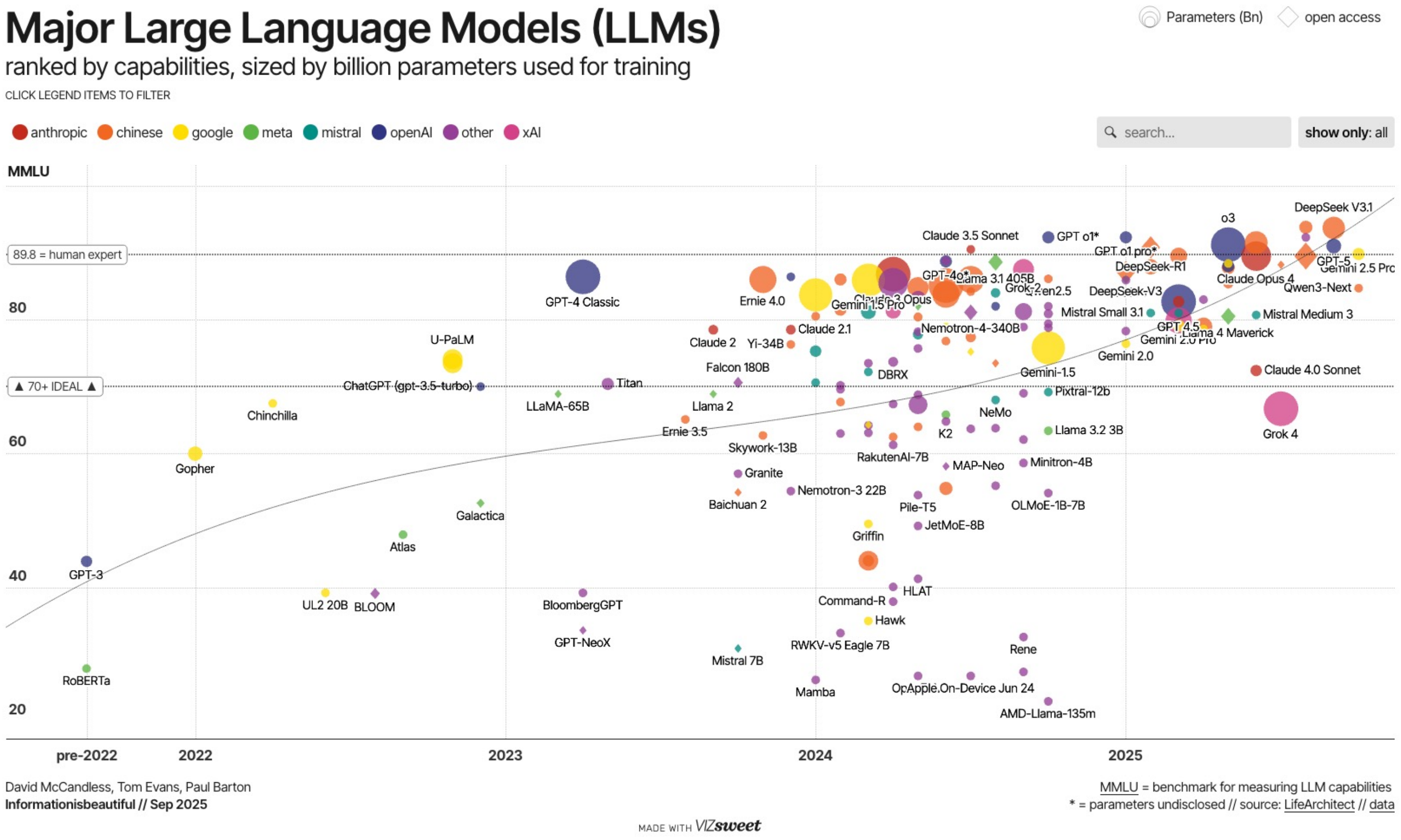
Pre-train one model that acts as the ***foundation*** for many different tasks



Pre-training learn data ***representation*** that is useful / transfers to many tasks.

Bespoke solutions → **Reusable Infrastructure**

The Power of Scale: Large Models, Data, Compute



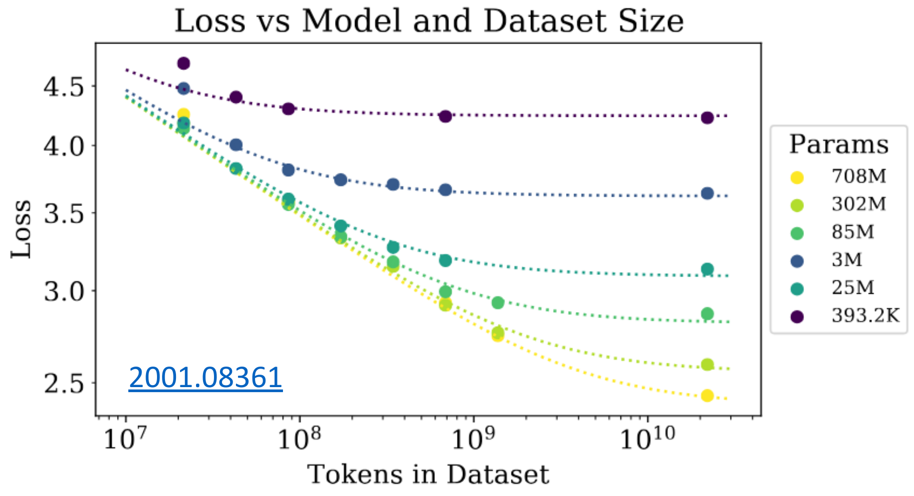
Major Large Language Models (LLMs)

ranked by capabilities, sized by billion parameters used for training

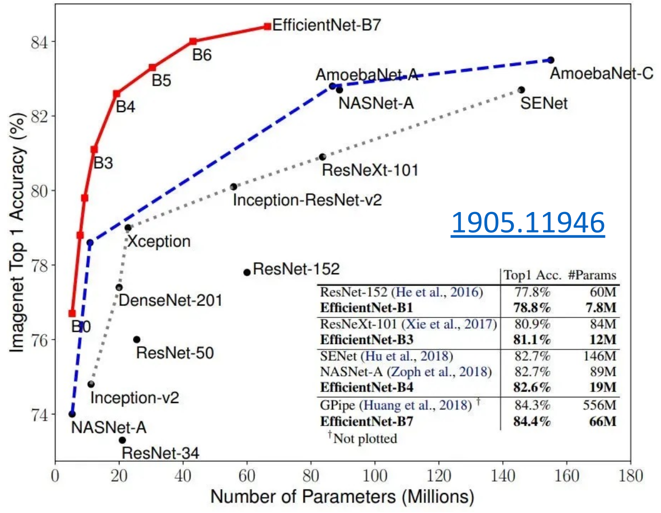
CLICK LEGEND ITEMS TO FILTER

David McCandless, Tom Evans, Paul Barton
Informationisbeautiful // Sep 2025

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



2001.08361



1905.11946

	Top1 Acc.	#Params
ResNet-152 (He et al., 2016)	77.8%	60M
EfficientNet-B1	**78.8%**	**7.8M**
ResNeXt-101 (Xie et al., 2017)	80.9%	84M
EfficientNet-B3	**81.1%**	**12M**
SENet (Hu et al., 2018)	82.7%	146M
NASNet-A (Zoph et al., 2018)	82.7%	89M
EfficientNet-B4	**82.6%**	**19M**
GPipe (Huang et al., 2018) †	84.3%	556M
EfficientNet-B7	**84.4%**	**66M**

†Not plotted

Image credit: D. McCandless, T. Evans, P. Barton

The Power of Scale: Large Models, Data, Compute

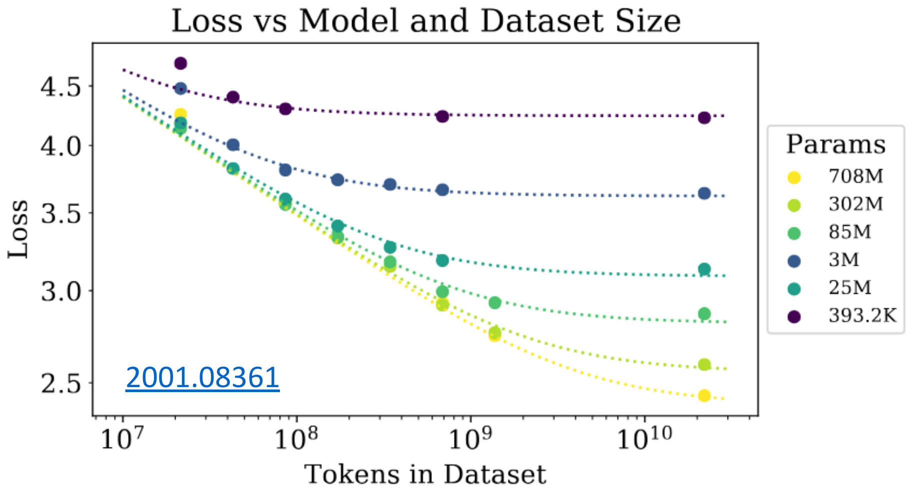
Large-scale training requires

A lot of (labelled?) data

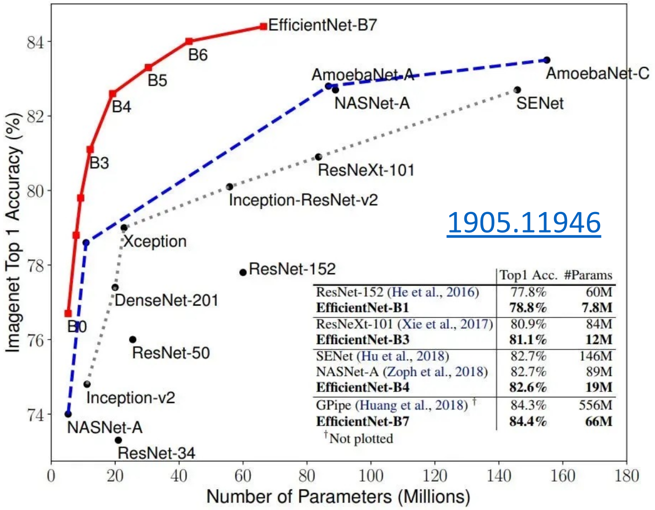
Is there a way to train neural networks without the need for huge labeled datasets?

Self-supervised learning

Self-supervision is data-type specific, adapted or new methods needed for HEP



2001.08361



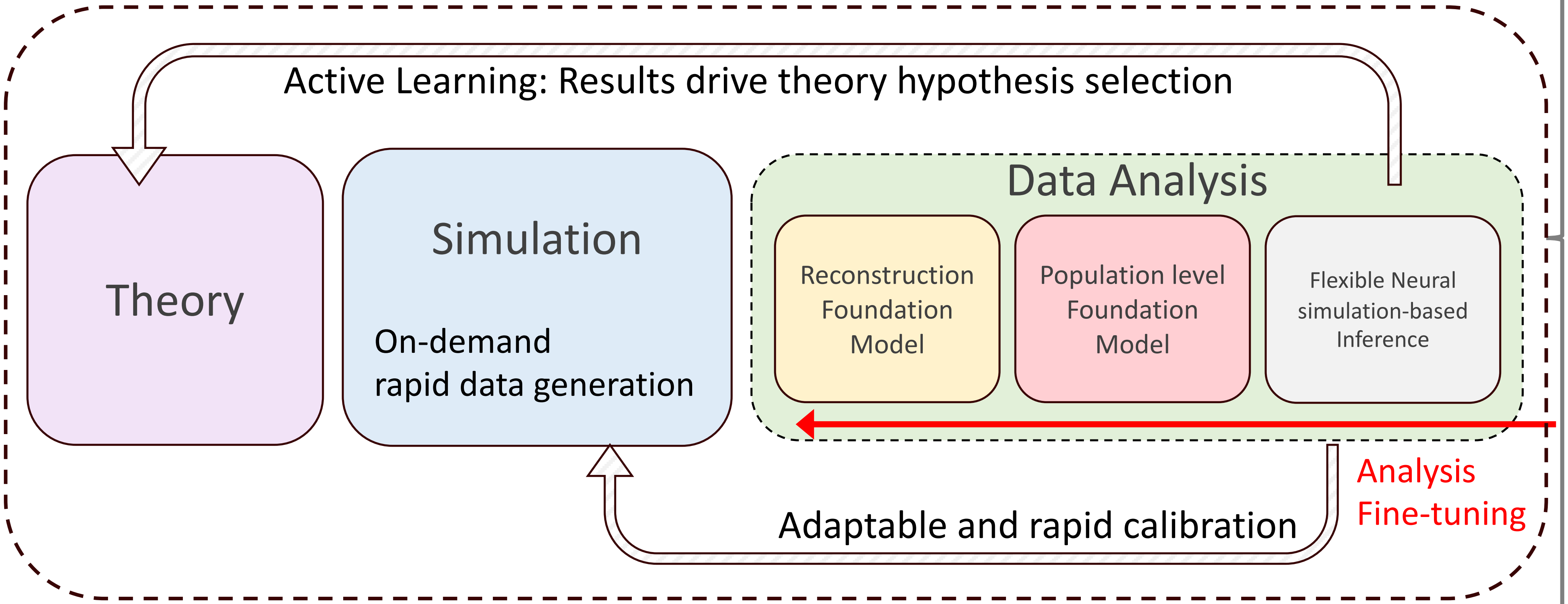
1905.11946

	Top1 Acc.	#Params
ResNet-152 (He et al., 2016)	77.8%	60M
EfficientNet-B1	**78.8%**	**7.8M**
ResNeXt-101 (Xie et al., 2017)	80.9%	84M
EfficientNet-B3	**81.1%**	**12M**
SENet (Hu et al., 2018)	82.7%	146M
NASNet-A (Zoph et al., 2018)	82.7%	89M
EfficientNet-B4	**82.6%**	**19M**
GPipe (Huang et al., 2018) †	84.3%	556M
EfficientNet-B7	**84.4%**	**66M**

†Not plotted

What might that look like?

Automated workflows → Agentic science



Active Learning: Results drive theory hypothesis selection

Theory

Simulation

On-demand
rapid data generation

Data Analysis

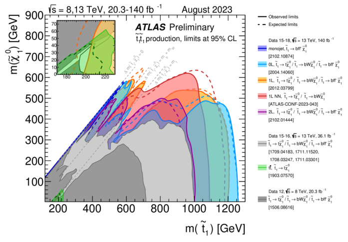
Reconstruction Foundation Model

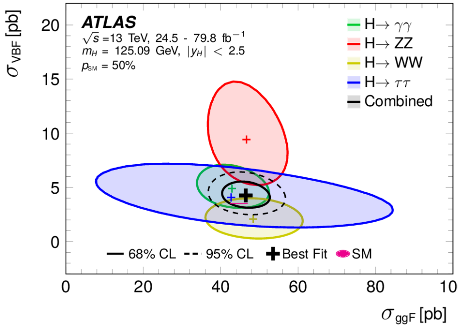
Population level Foundation Model

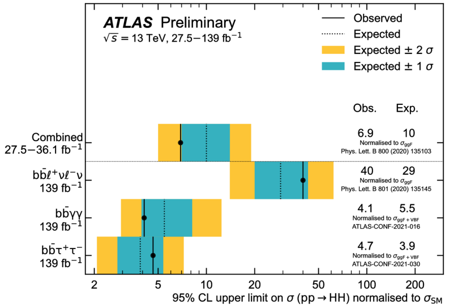
Flexible Neural simulation-based Inference

Analysis Fine-tuning

Adaptable and rapid calibration







⋮

Systems to do rapid, optimal, and automated searches through lots of data and theory space

What do we need for our foundation models?

Training procedures, like self-supervision, that enable generalized representation learning

Multi-modality, to process information form different detectors or even different experiments

Scale, to take advantage of the large data, model, and compute limits that benefits learning

Adaptability, to put these models to use for a wide array of tasks

Pre-training with Self-Supervision

Masked Language Modeling

Strategy: Mask some elements of input sequence, then predict missing tokens

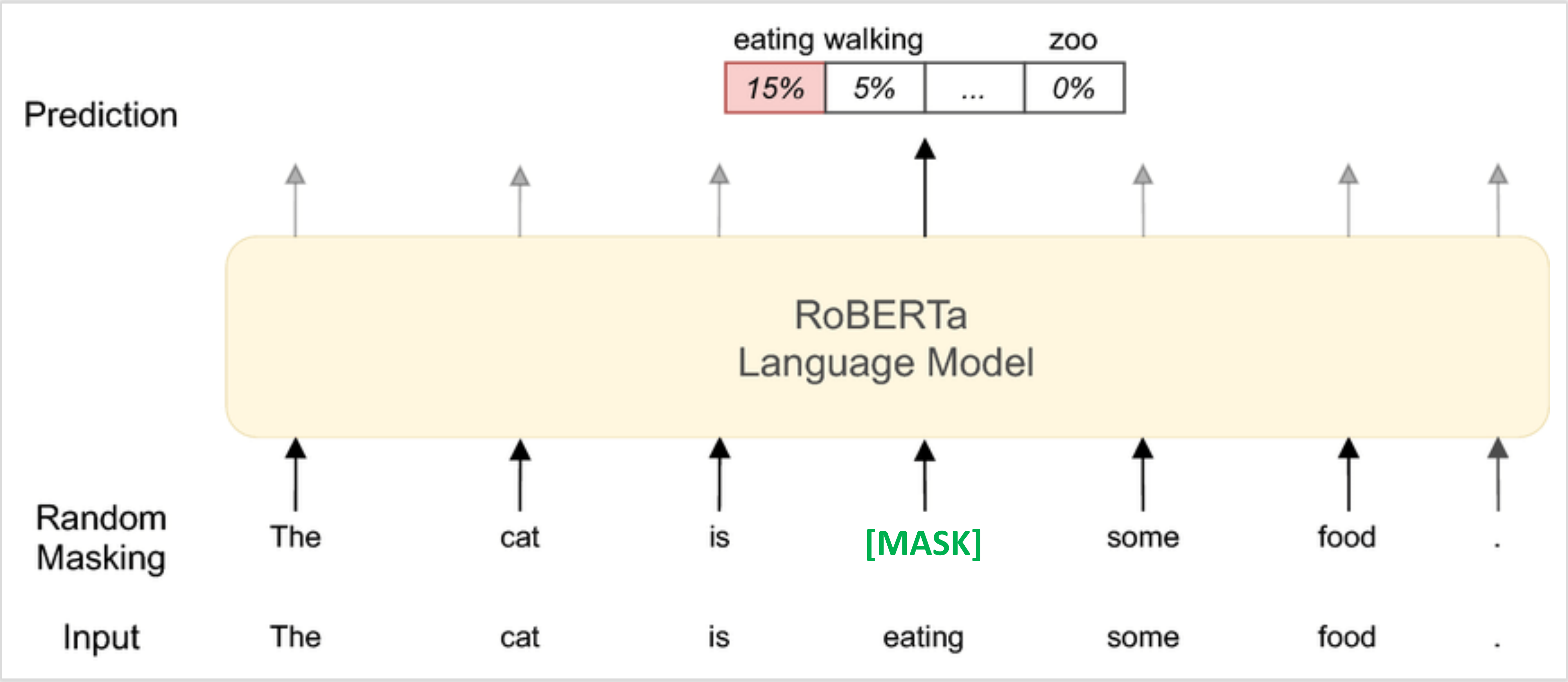
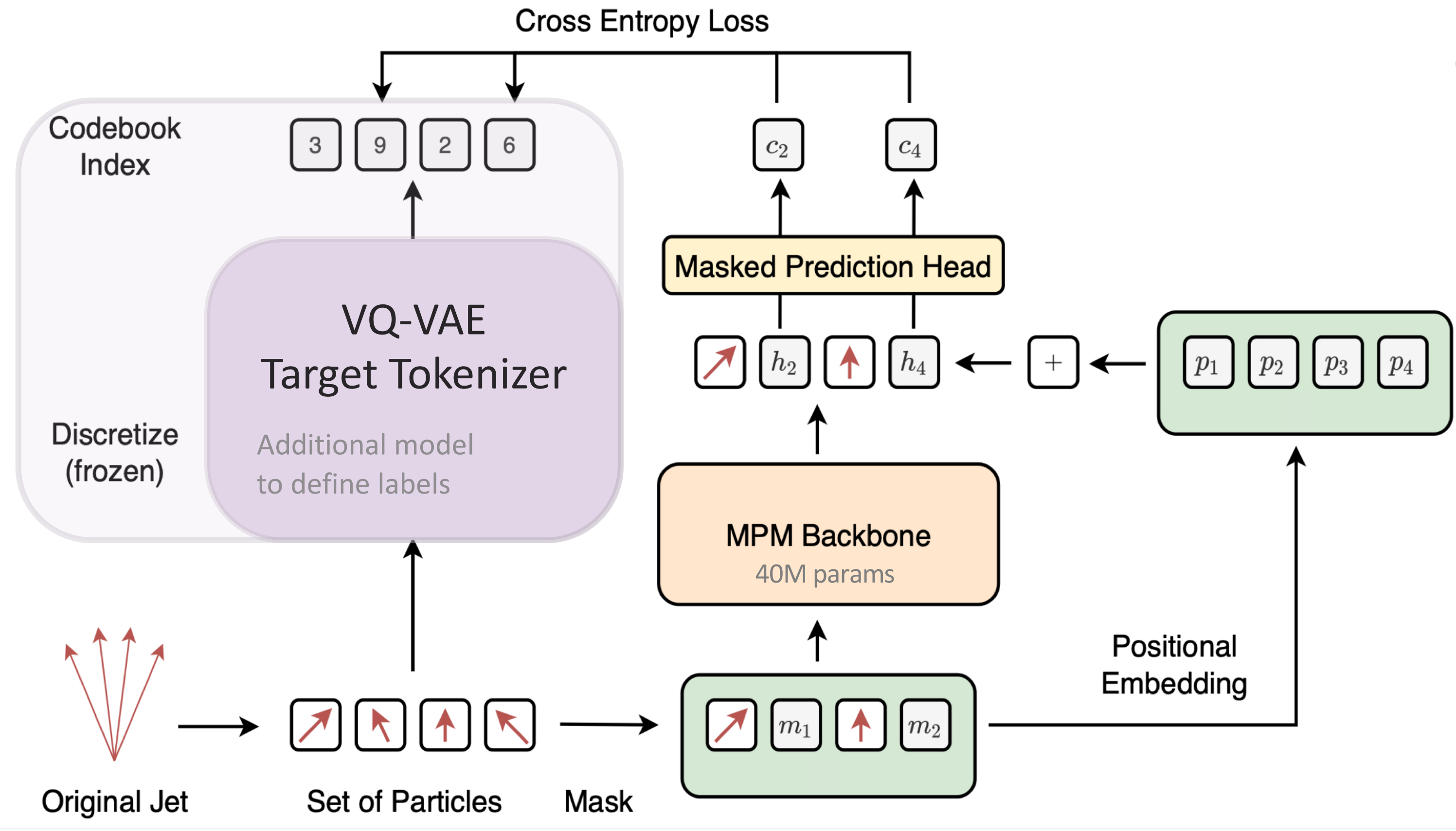


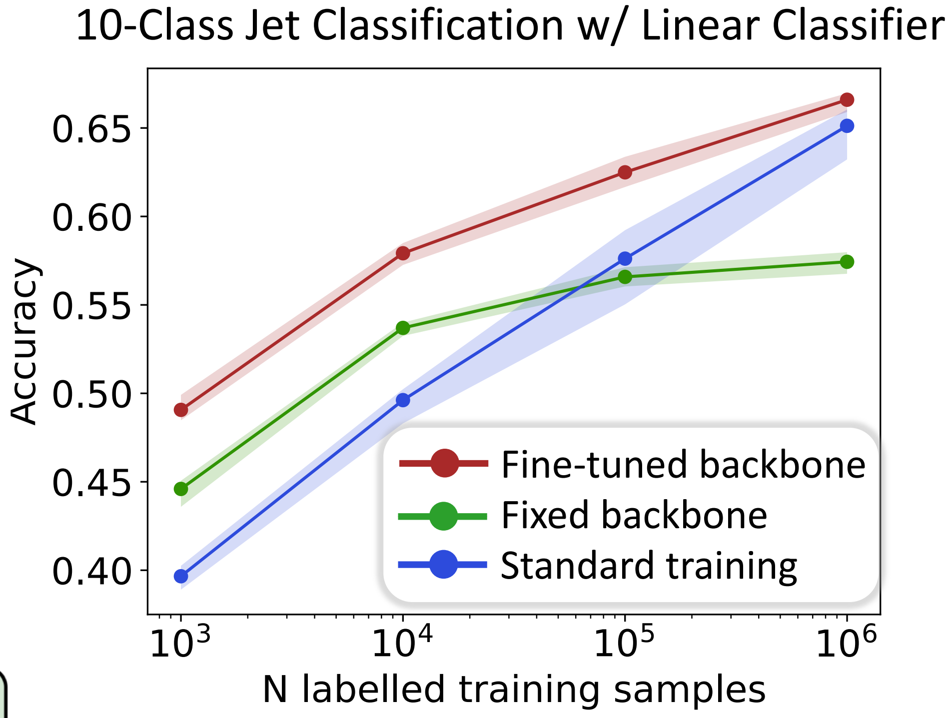
Image credit: 2104.01642

Masked Particle Modeling

Mask individual particles, predict properties

Amenable to large-scale data-based training





Pre-training gives better performance and better data efficiency on downstream tasks

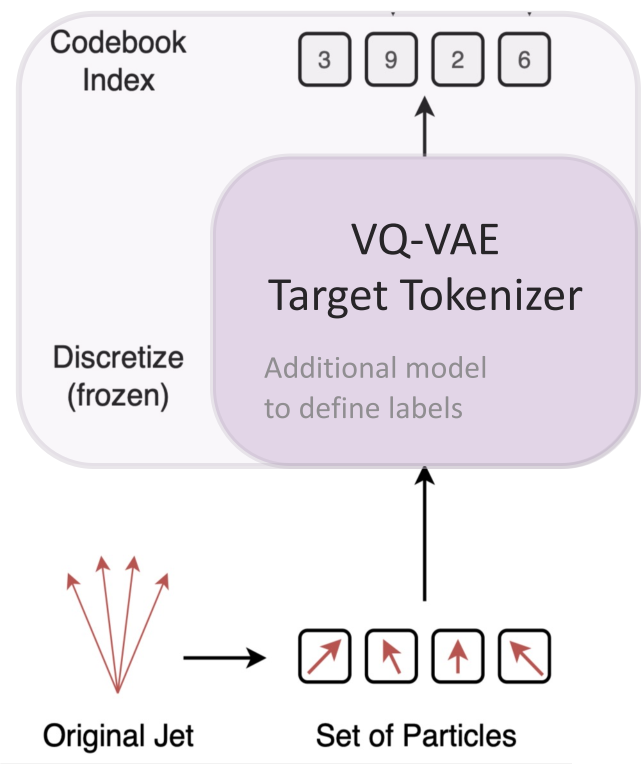
MLST **5** 035074 (2024),
MLST **6** 025075 (2025)

Tokenization

Tokenizing does “in-context” discretization and compression of data space

- Can facilitates and improve target predictions

- Poor tokenization can lead to resolution loss



Codebook Index

3 9 2 6

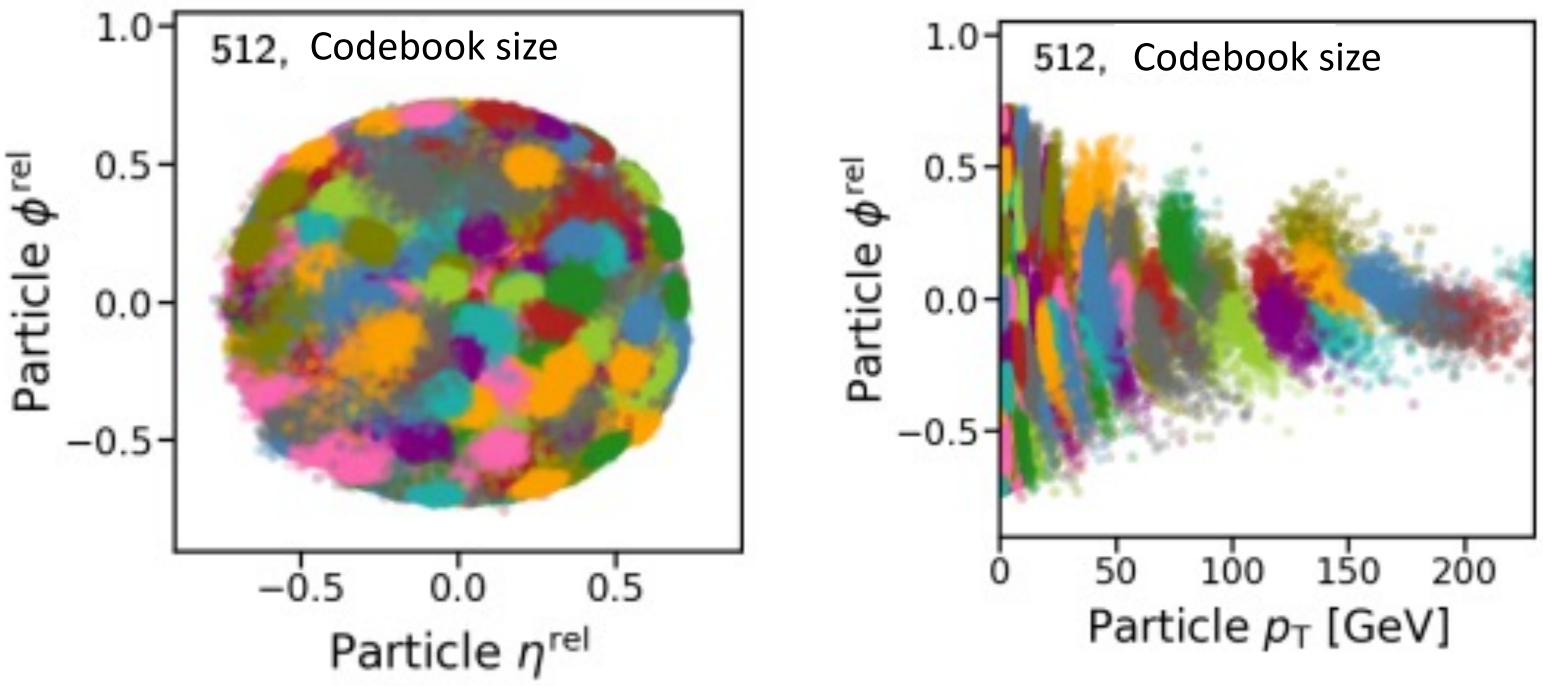
VQ-VAE Target Tokenizer

Additional model to define labels

Discretize (frozen)

Original Jet

Set of Particles

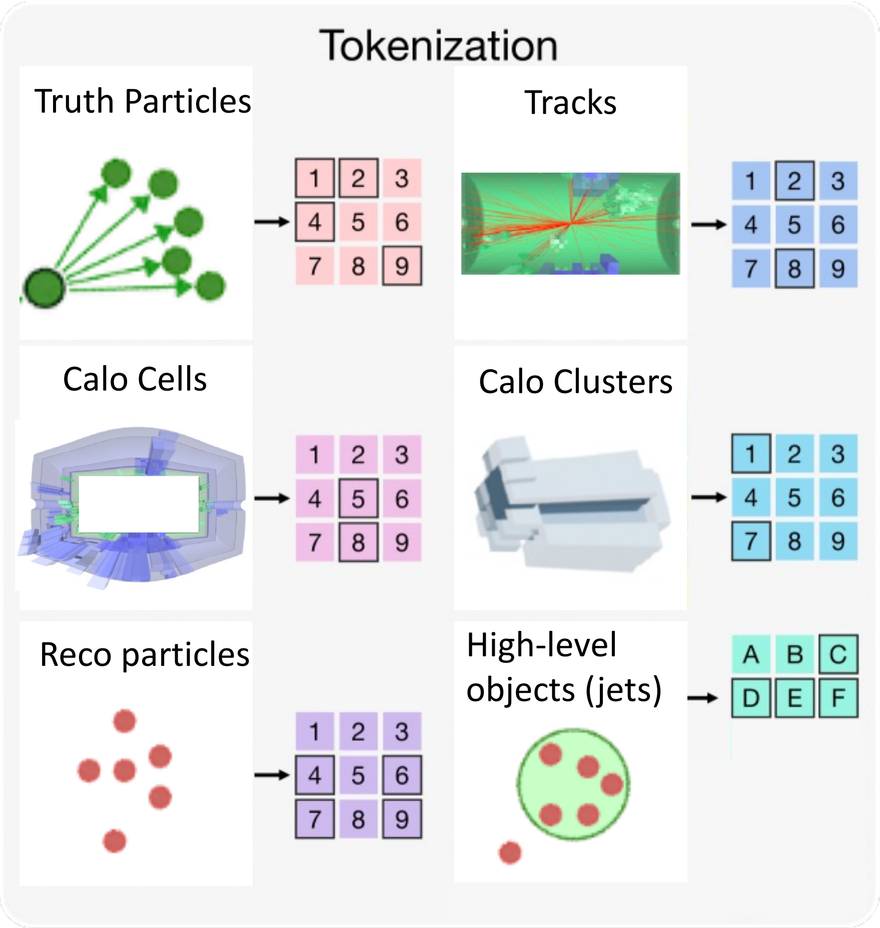


MLST **5** 035074 (2024),
MLST **6** 025075 (2025)

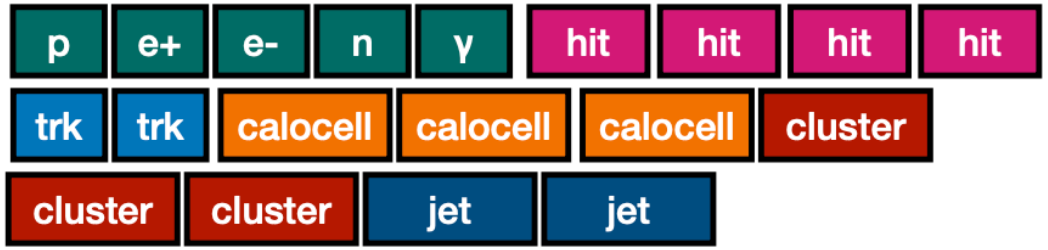
Image credit: 2403.05618

Multi-Modality

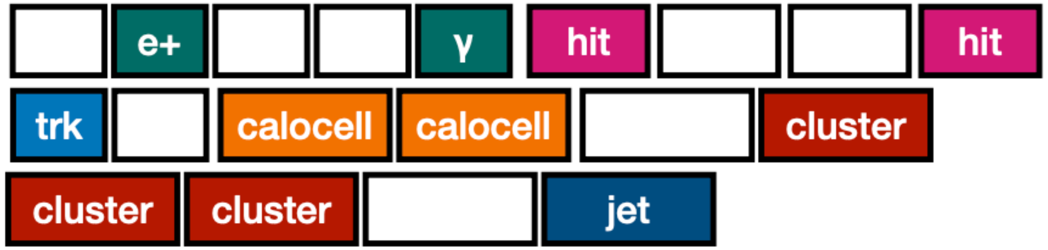
Towards Big Multi-Modal Modals for Particle Physics



Like masked language modeling but on multi-modal physics data


p e+ e- n γ hit hit hit hit
trk trk calocell calocell calocell cluster
cluster cluster jet jet

“fill in the blank” on this “text”

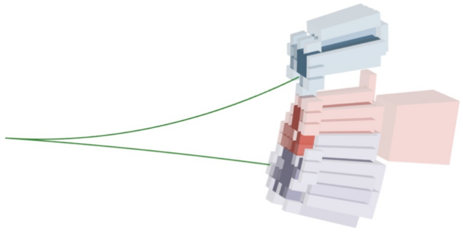

e+ γ hit hit
trk calocell calocell cluster
cluster cluster jet

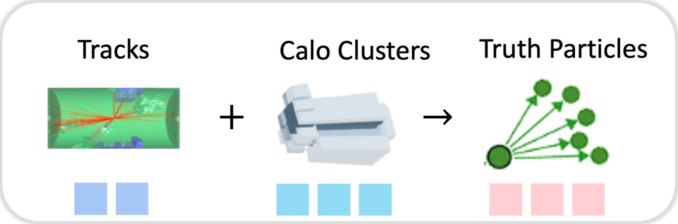
Work in Progress:
N. Kakati, B. Hashemi, E. Gross, L. Heinrich, M. K.

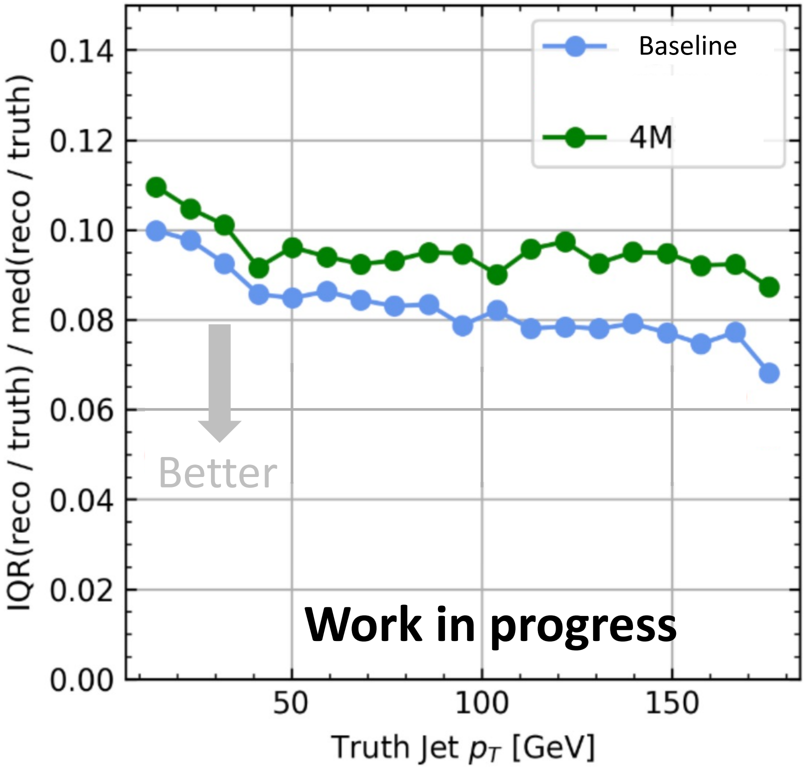
Prior work: Leigh, Klein, … , MK, … , 2401.13537
Leigh, Klein, … , MK, … , 2409.12589

Towards Big Multi-Modal Modals for Particle Physics

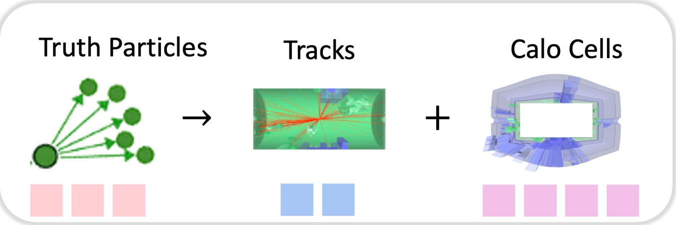
Pattern Recognition

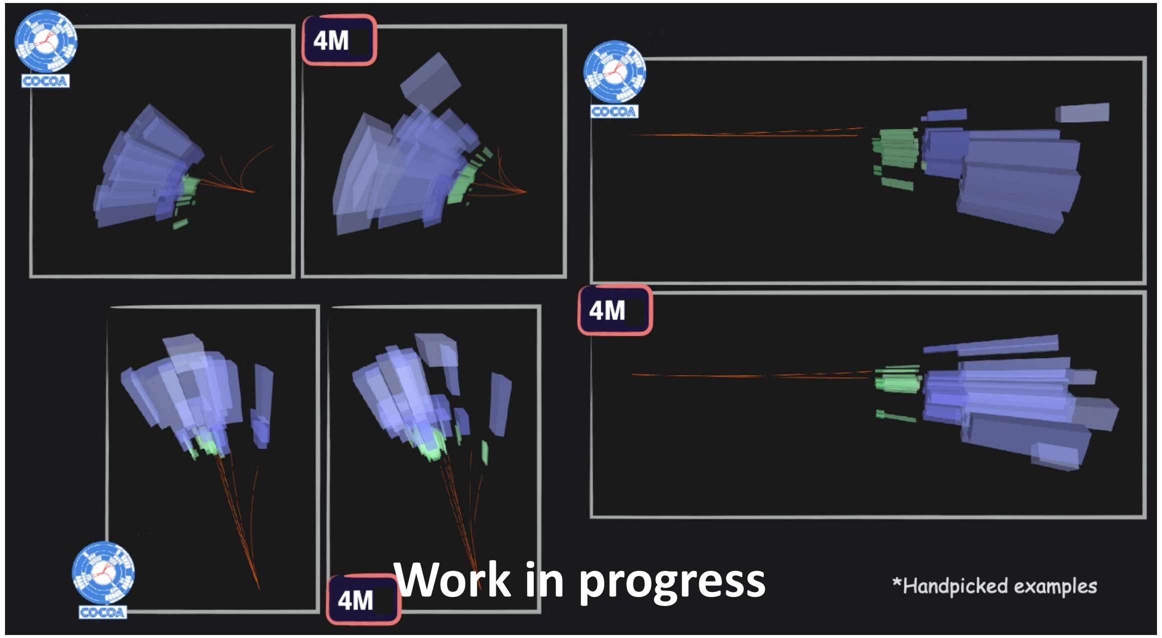






Generative Modeling



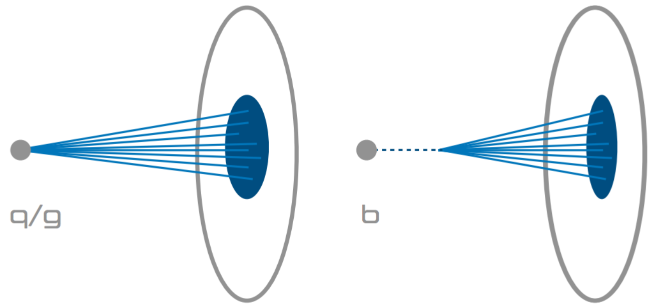


Work in Progress:
N. Kakati, B. Hashemi, E. Gross, L. Heinrich, M. K.

Prior work: Leigh, Klein, … , MK, … , 2401.13537
Leigh, Klein, … , MK, … , 2409.12589

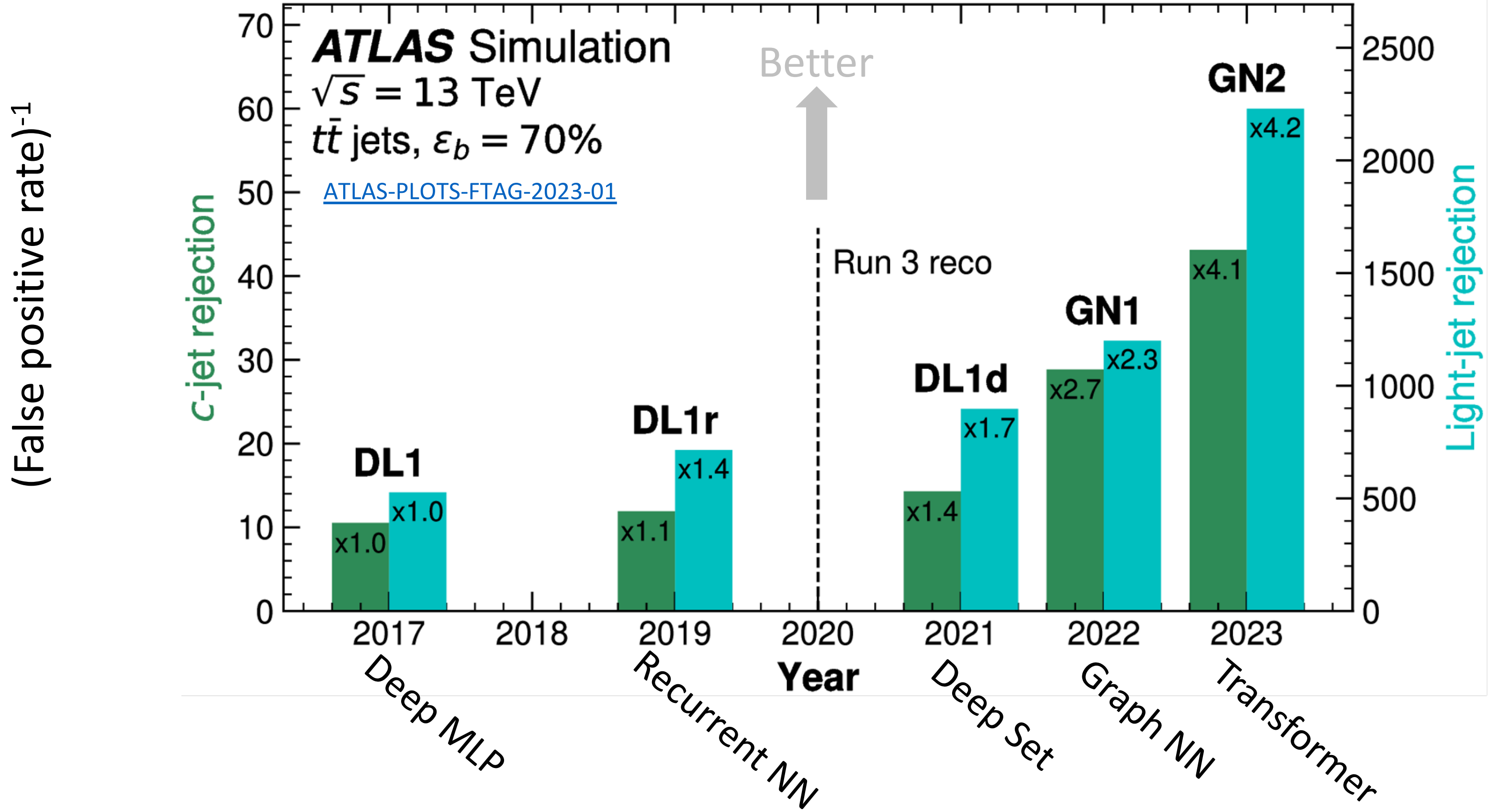
Scale

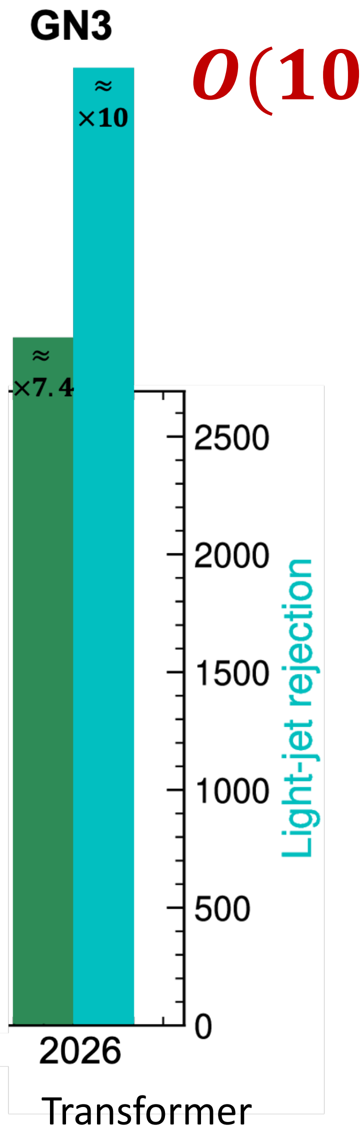
Building bigger models



$O(1M)$ param

$O(10M)$ param



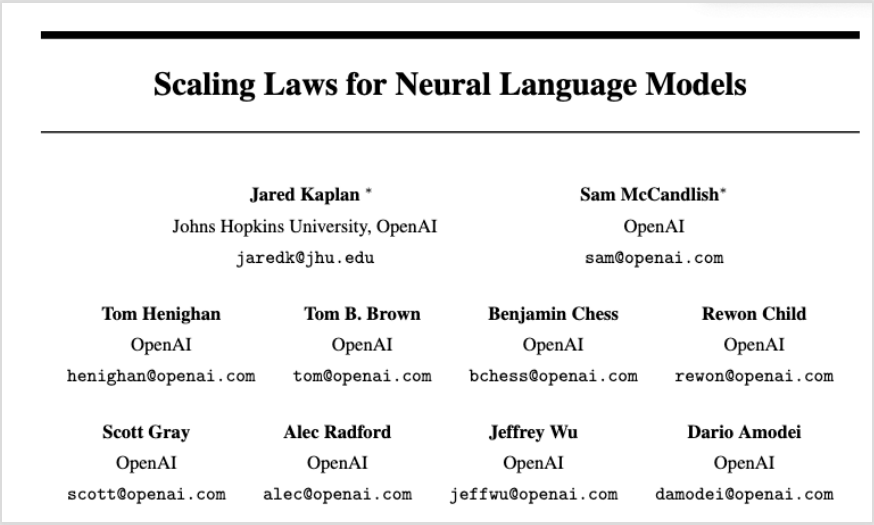


Transformer

ATL-PHYS-PUB-2026-001

Neural Scaling Laws

Performance improvement is predictable in the limit of large data and models



Scaling Laws for Neural Language Models

Jared Kaplan *
Johns Hopkins University, OpenAI
jaredk@jhu.edu

Sam McCandlish*
OpenAI
sam@openai.com

Tom Henighan
OpenAI
henighan@openai.com

Tom B. Brown
OpenAI
tom@openai.com

Benjamin Chess
OpenAI
bchess@openai.com

Rewon Child
OpenAI
rewon@openai.com

Scott Gray
OpenAI
scott@openai.com

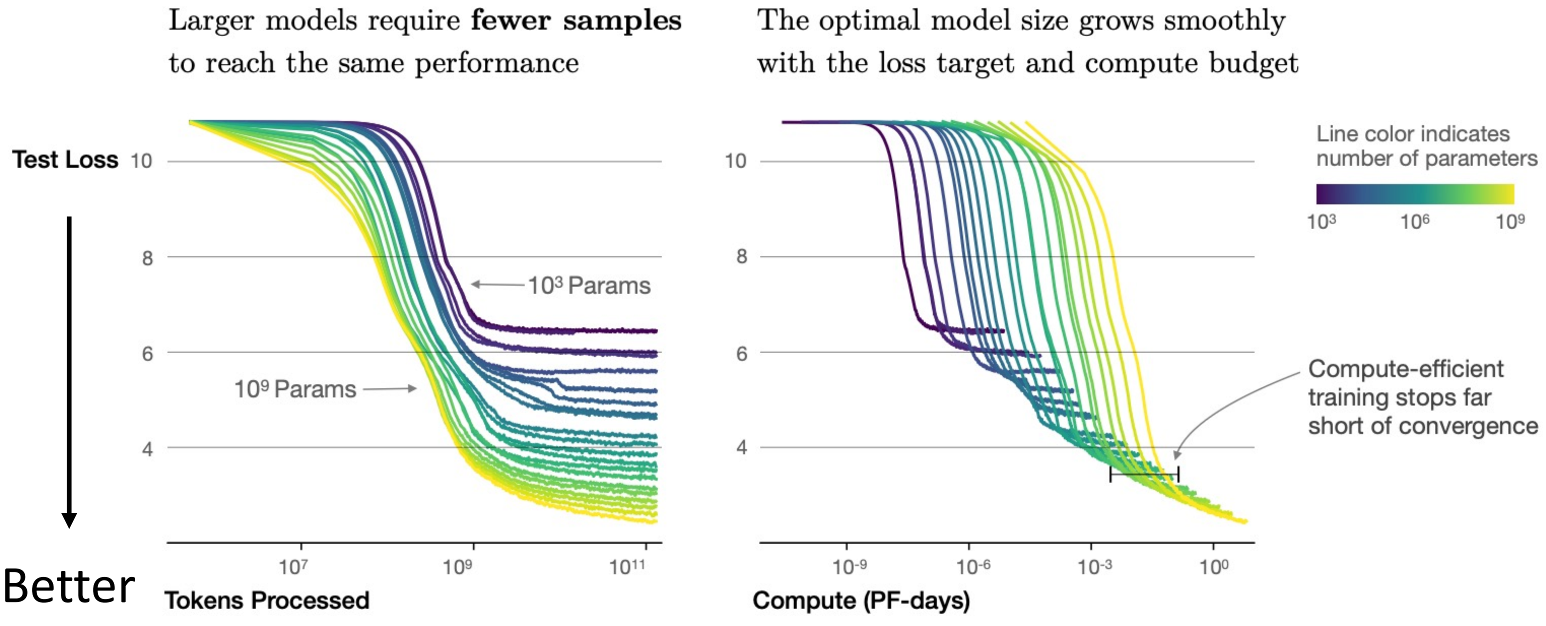
Alec Radford
OpenAI
alec@openai.com

Jeffrey Wu
OpenAI
jeffwu@openai.com

Dario Amodei
OpenAI
damodei@openai.com

Result: Power laws

$$\hat{L}(N,D) \triangleq E + \frac{A}{N^{\alpha}} + \frac{B}{D^{\beta}}$$



Larger models require **fewer samples** to reach the same performance

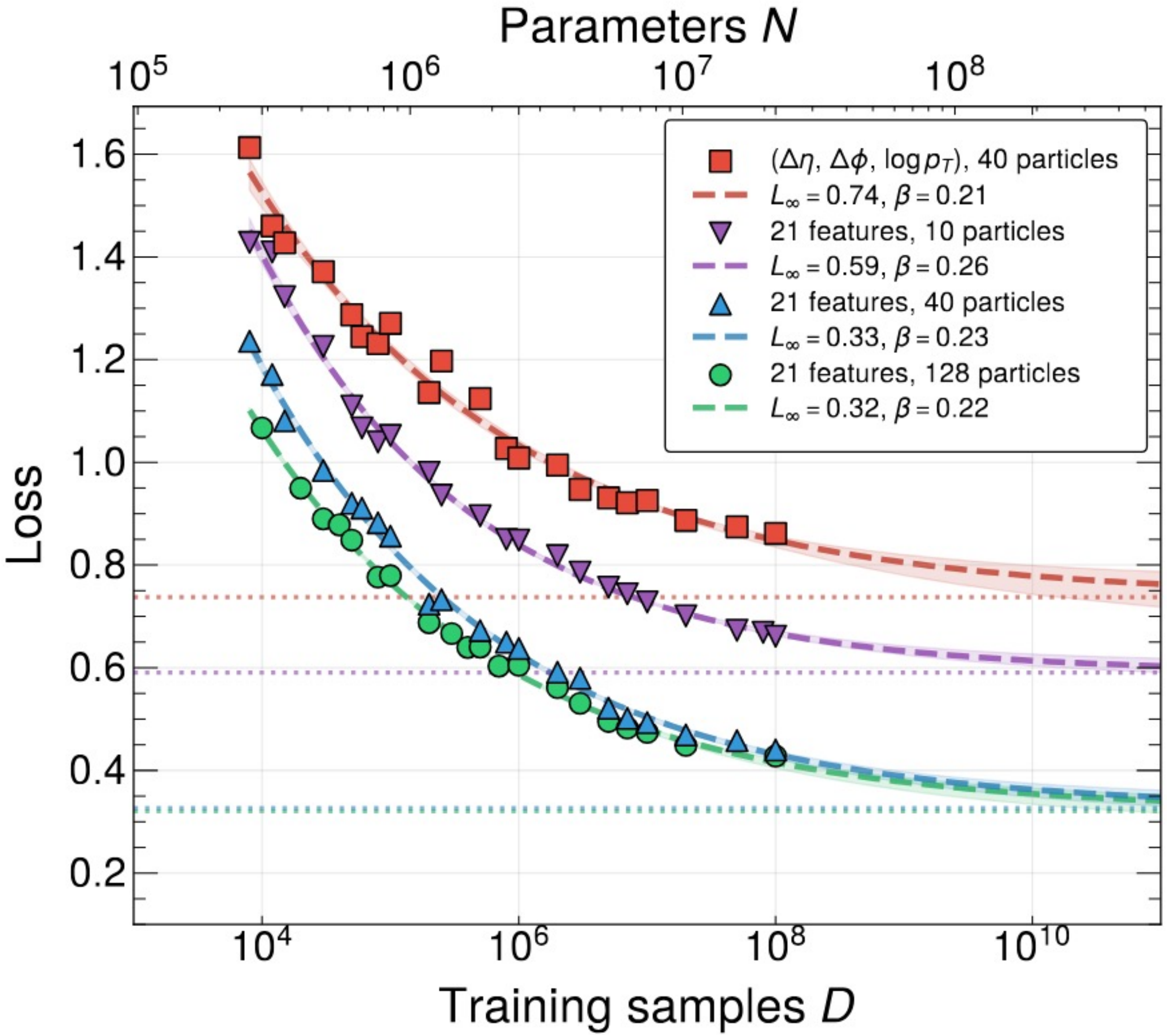
The optimal model size grows smoothly with the loss target and compute budget

More data

More compute

2001.08361

Scaling Laws in Particle Physics

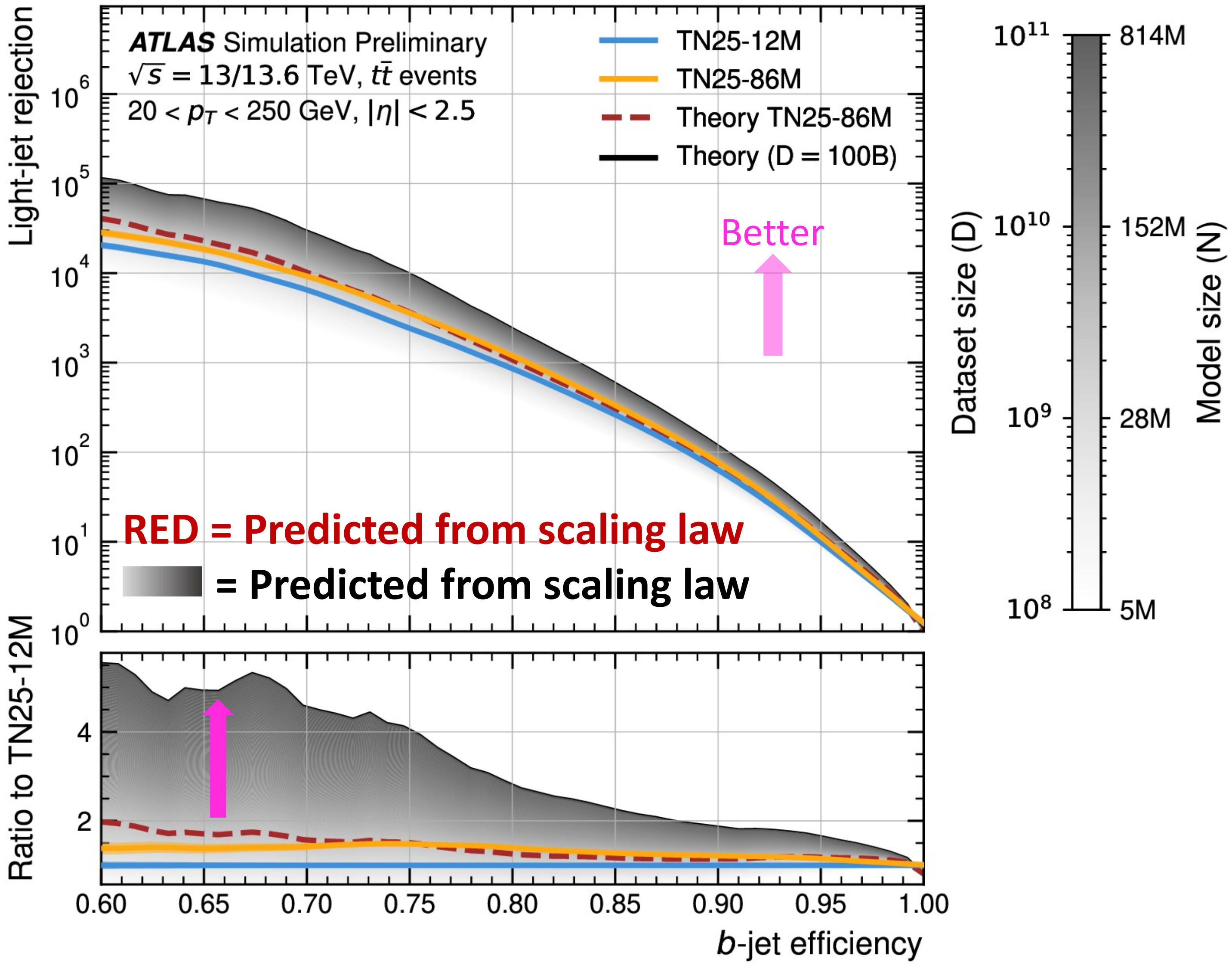


ATL-SOFT-PUB-2026-002

Vigl, Hartman, **MK**, Heinrich, 2602.15781,

Scaling Laws in Particle Physics

(False positive rate)$^{-1}$



Better

RED = Predicted from scaling law

▨ **= Predicted from scaling law**

True positive rate

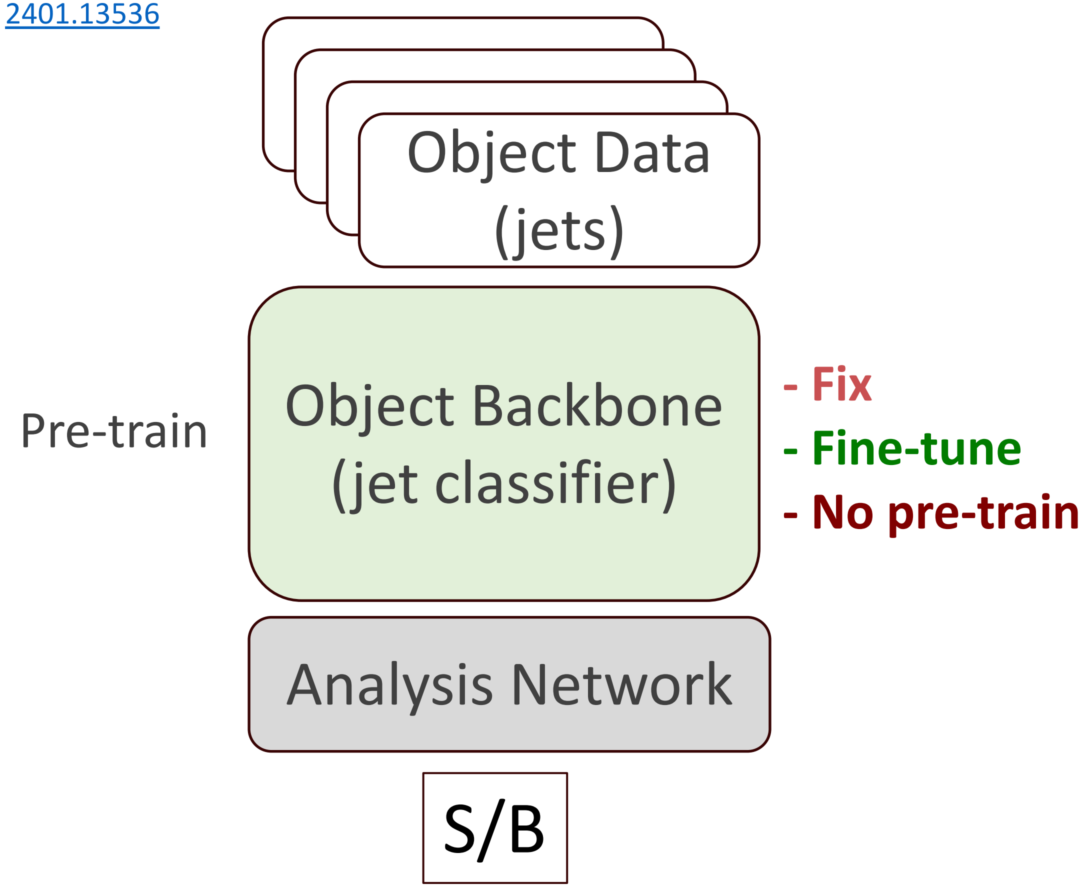
ATL-SOFT-PUB-2026-002

Vigl, Hartman, **MK**, Heinrich, [2602.15781](2602.15781),

Adaptability

Does Pre-training help in Measurements?

2401.13536



Object Data (jets)

Pre-train

Object Backbone (jet classifier)

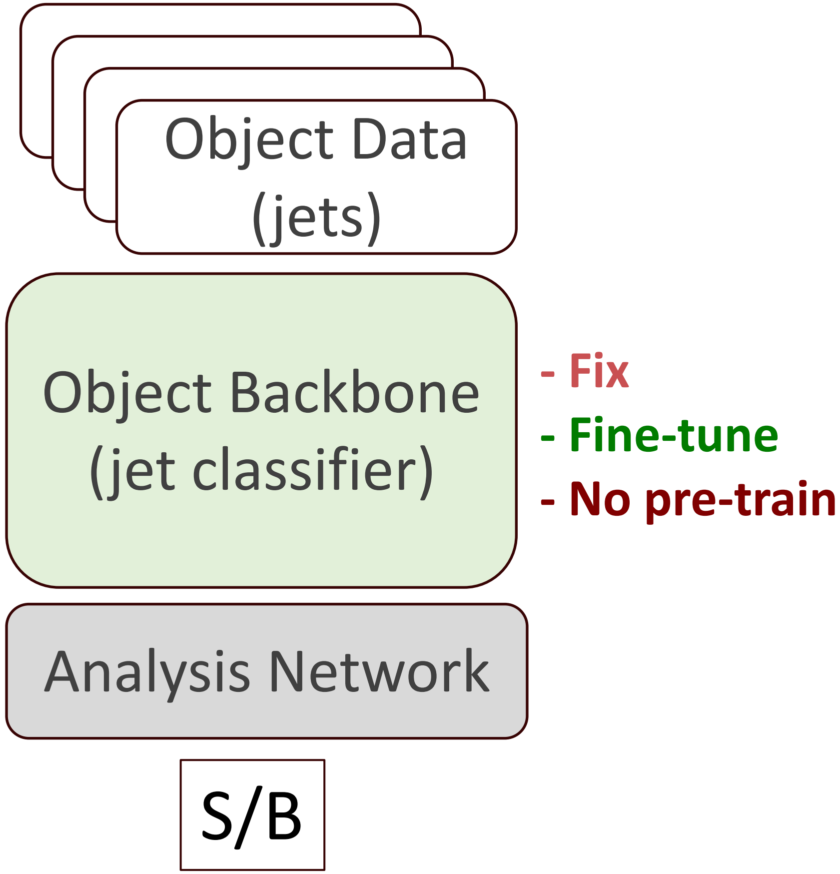
- **Fix**
- **Fine-tune**
- **No pre-train**

Analysis Network

S/B

Does Pre-training help in Measurements?

2401.13536



Object Data (jets)

Pre-train

Object Backbone (jet classifier)

- Fix
- Fine-tune
- No pre-train

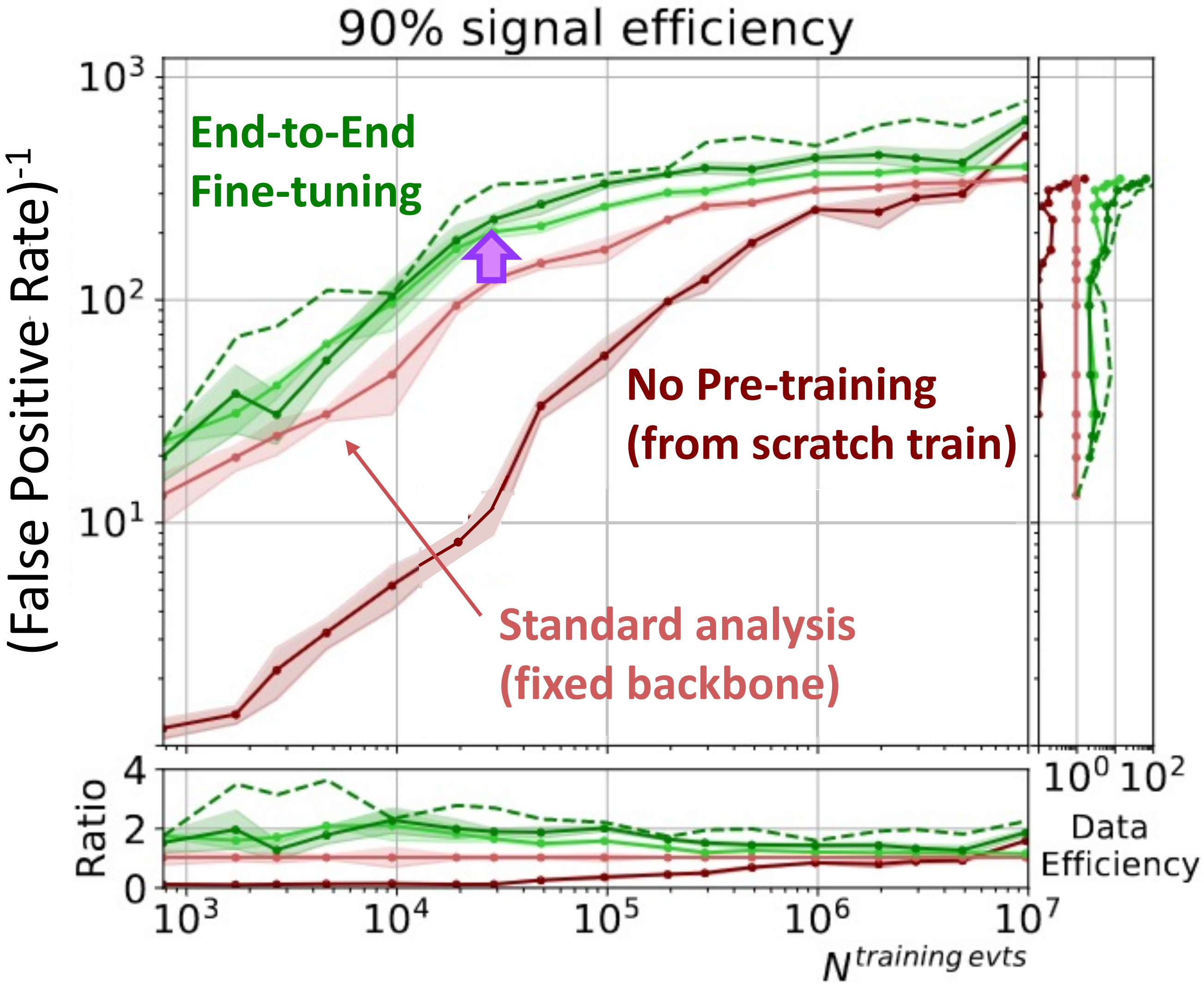
Analysis Network

S/B

Up to:

2-4X better Signal / Background

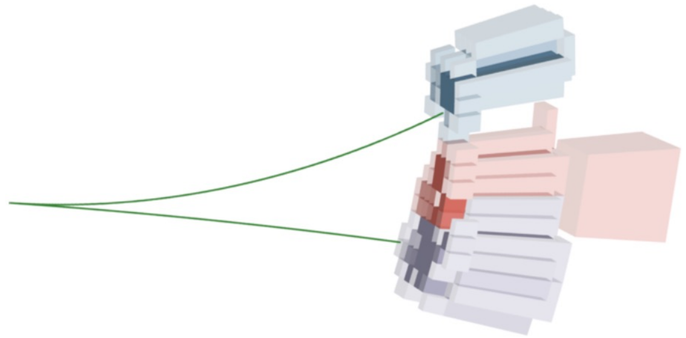
10X more data efficient

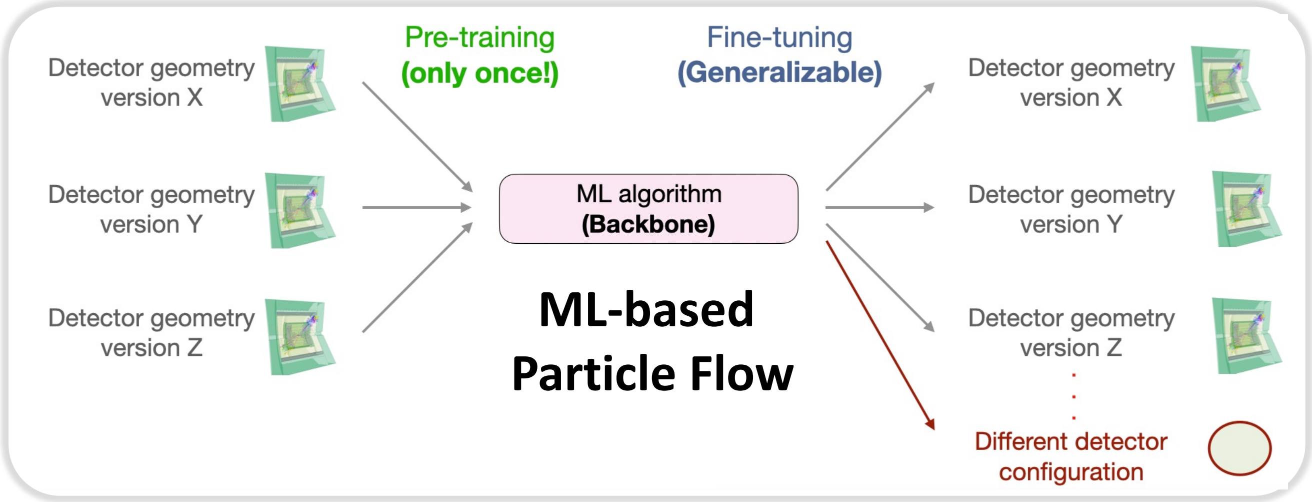


Double Higgs HH→4b vs Background Events

Adaptable Reconstruction for Detector Optimization

Data reconstruction with ML-based Particle Flow





Detector geometry version X

Detector geometry version Y

Detector geometry version Z

Pre-training **(only once!)**

Fine-tuning **(Generalizable)**

ML algorithm **(Backbone)**

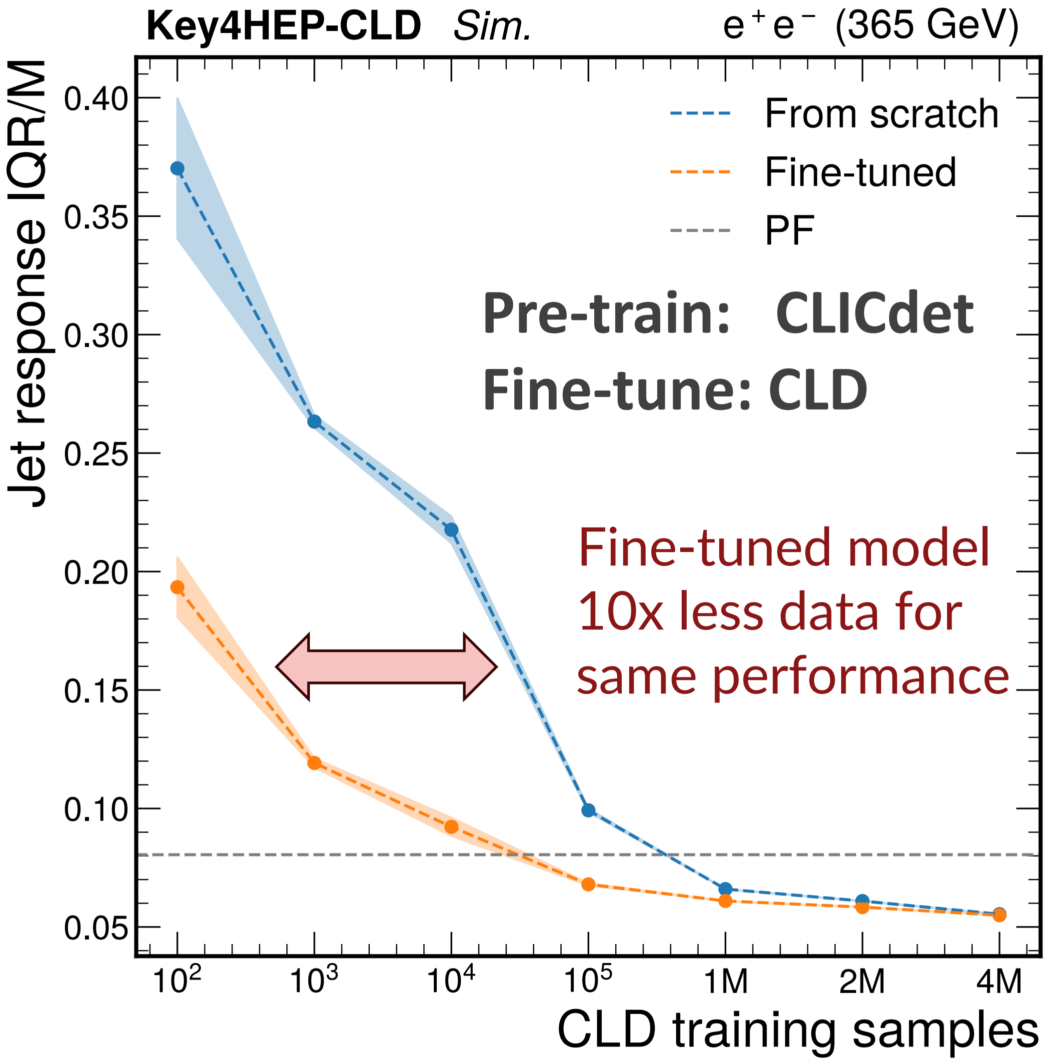
ML-based Particle Flow

Detector geometry version X

Detector geometry version Y

Detector geometry version Z

Different detector configuration



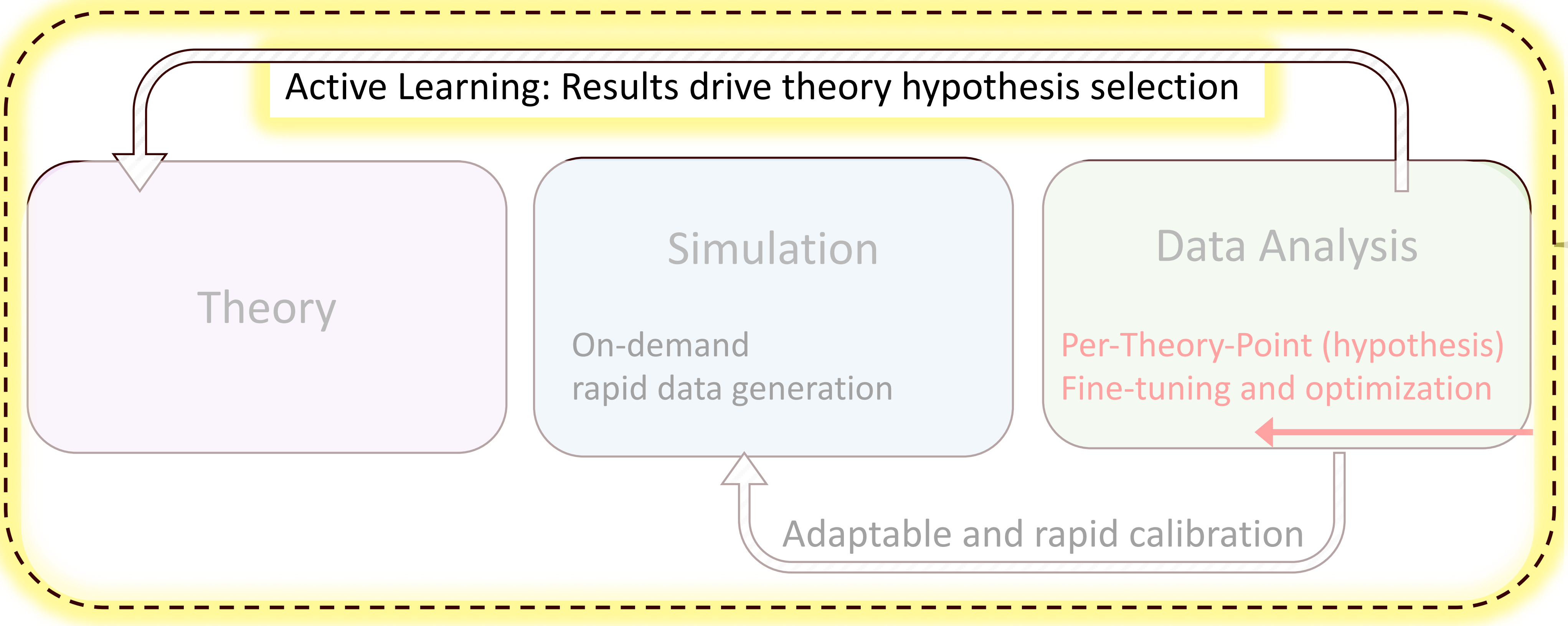
Pre-train: CLICdet
Fine-tune: CLD

Fine-tuned model 10x less data for same performance

PRD 111, 092015 (2025)

What about the outer parts of the loop?

Automated workflows → Agentic science



Active Learning: Results drive theory hypothesis selection

Theory

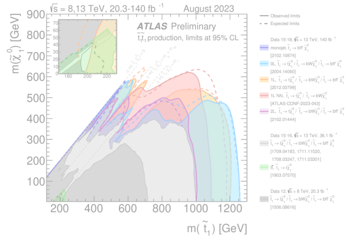
Simulation

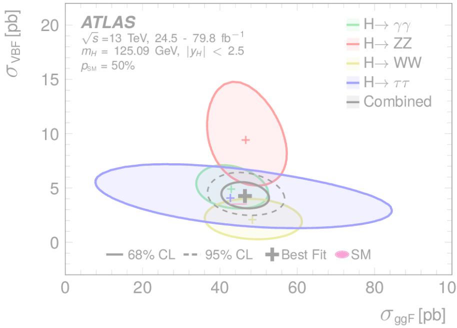
On-demand
rapid data generation

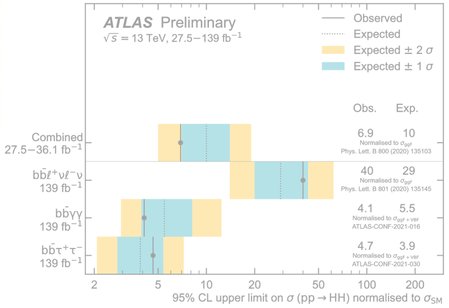
Data Analysis

Per-Theory-Point (hypothesis)
Fine-tuning and optimization

Adaptable and rapid calibration



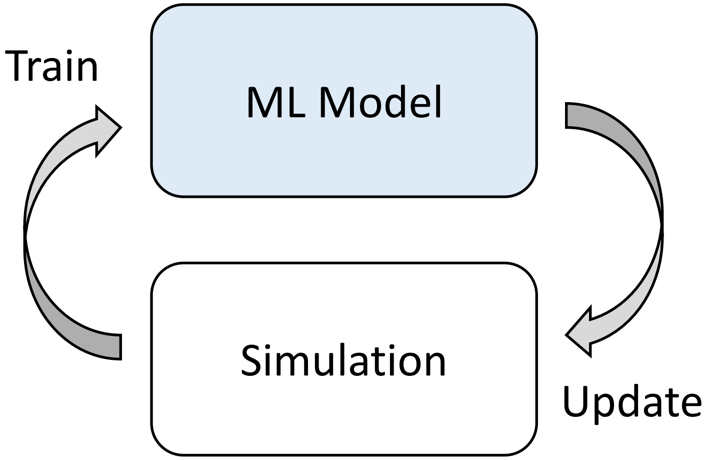




⋮

Systems to do rapid, optimal, and automated searches through lots of data and theory space

Active Learning– Exploring Theory Space

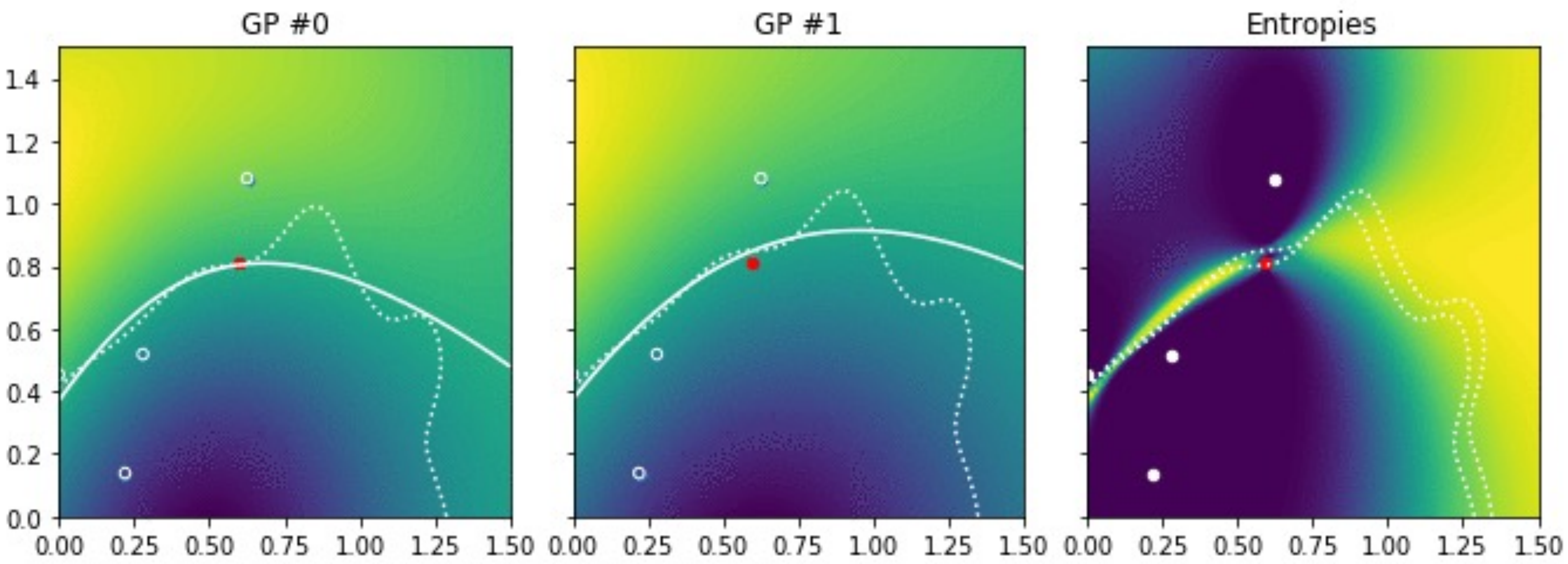


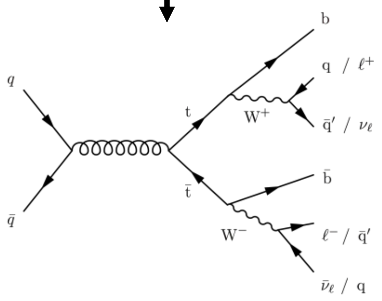
Train

ML Model

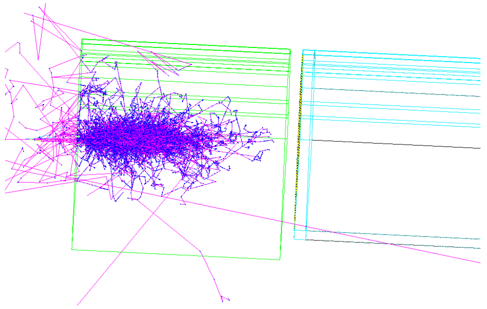
Simulation

Update

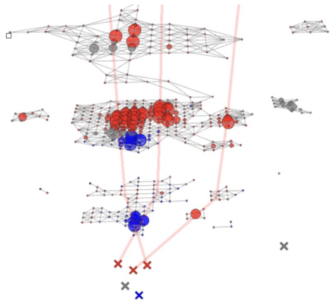




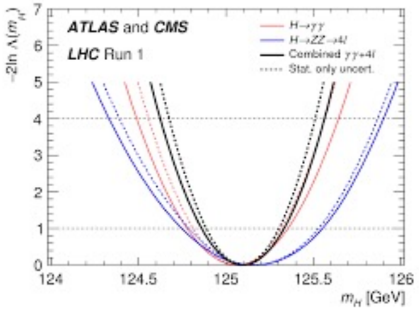
Theory



Simulation



Data Analysis Pipeline

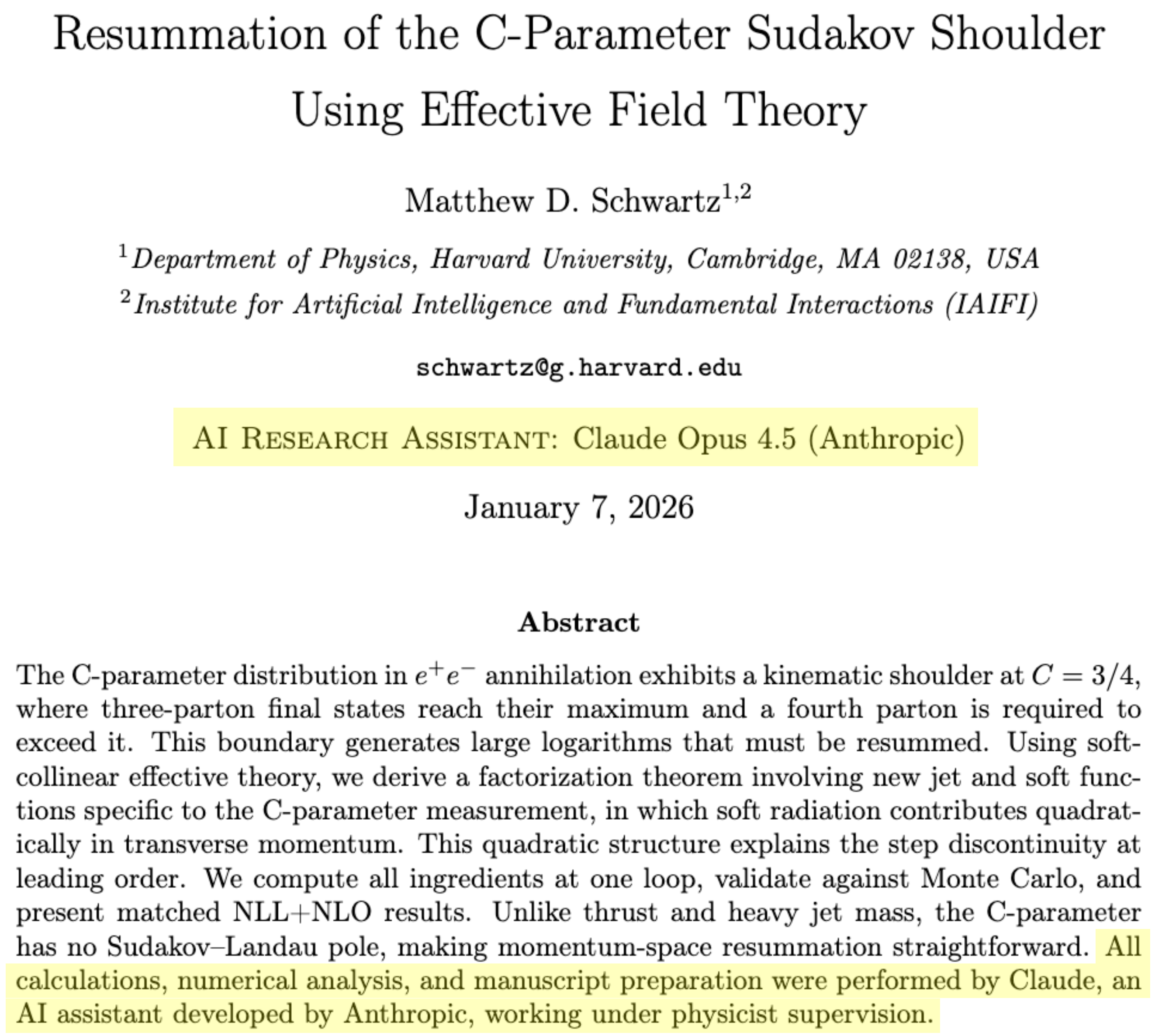


Statistical Analysis

ML to suggest new Theory

ATL-PHYS-PUB-2022-045

Towards Agentic Science in HEP



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.

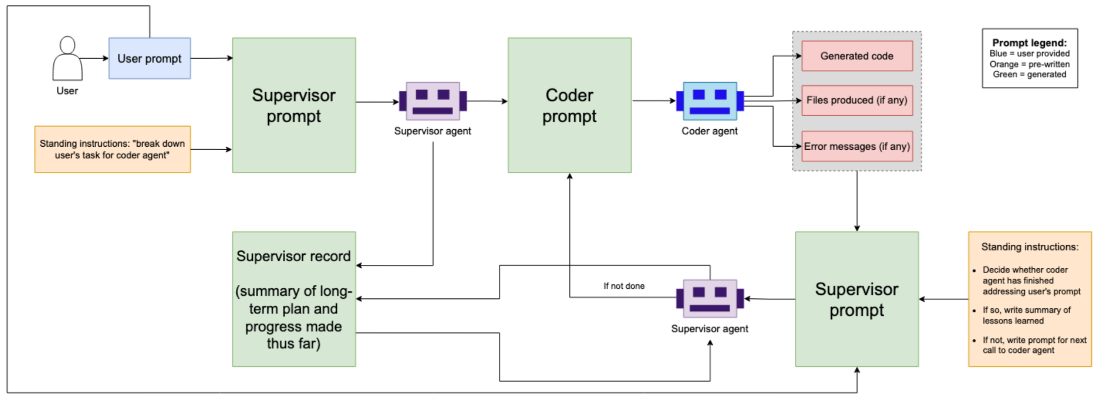
2601.02484

Towards Agentic Science in HEP

First steps using LLM agents to plan, code, and evaluate LHC data analysis

Automating High Energy Physics Data Analysis with LLM-Powered Agents

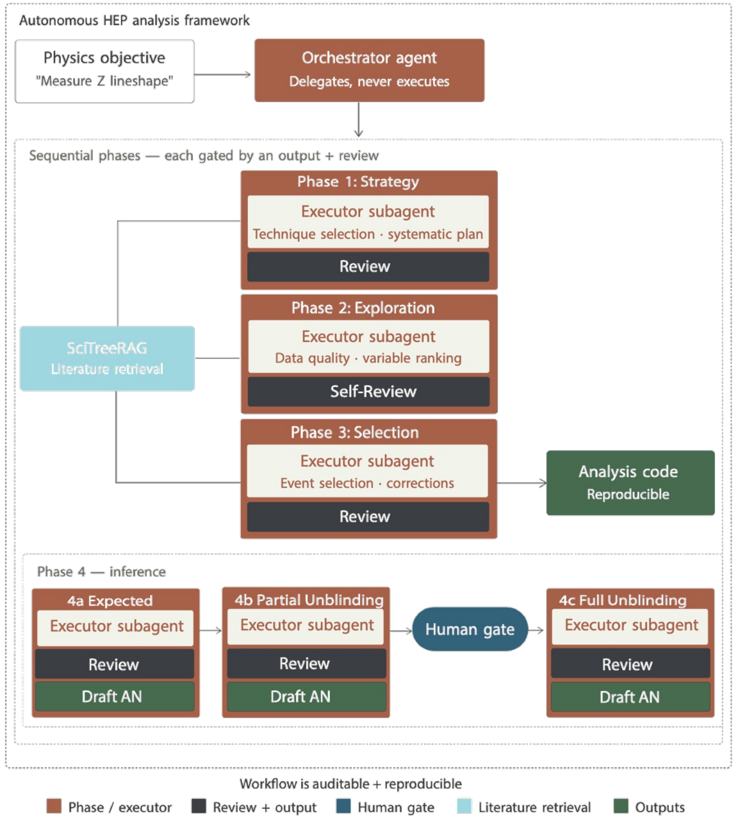
Eli Gendreau-Distler,[1,2,*] Joshua Ho,[1,2,†] Dongwon Kim,[1,2,‡] Luc
Tomas Le Pottier,[1,2,§] Haichen Wang,[1,2,¶] and Chengxi Yang[1,2,**]
[1]*Department of Physics, University of California, Berkeley, Berkeley, CA 94720, USA*
[2]*Physics Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA*



2512.07785

AI Agents Can Already Autonomously Perform Experimental High Energy Physics

Eric A. Moreno[*†1,2], Samuel Bright-Thonney[*‡1,2], Andrzej Novak[*§1,2], Dolores Garcia[♯3],
and Philip Harris[♭1,2]



2603.20179

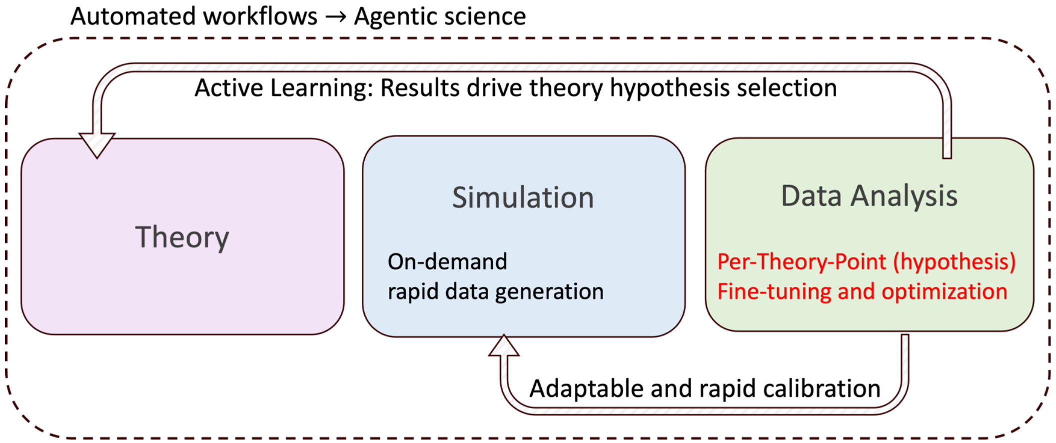
Summary

More rapid, optimal, automated data analysis
→ Broader and faster searches through data space, theory space, & design space

Foundation models are the infrastructure
→ Reusable, adaptable, and generalized ML tools that solve a variety of tasks

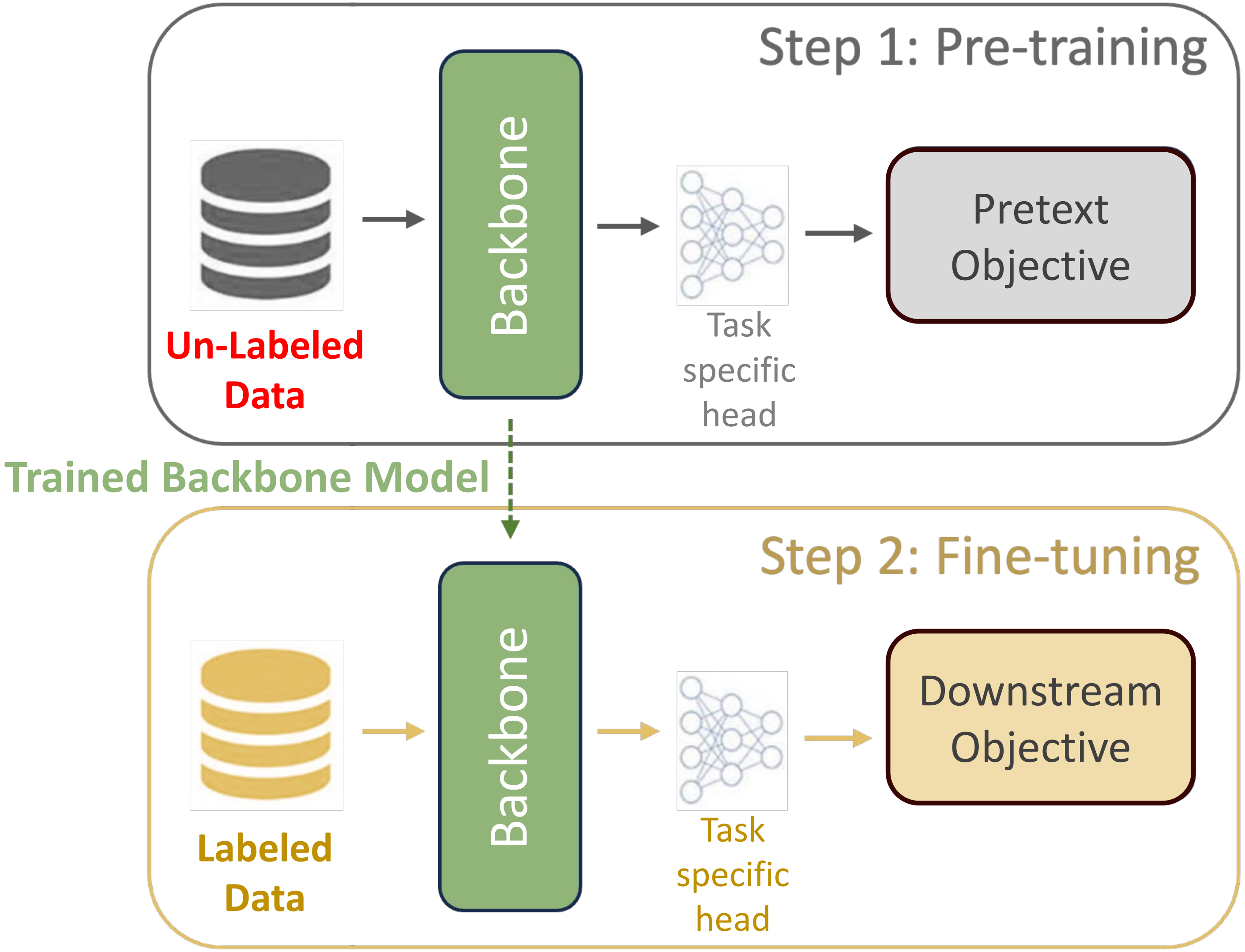
Exciting developments underway
→ Key progress on new training methods, multi-modality, scaling, adaptability

Agentic science is quickly developing
→ Exploring the automation frontier



Backup

Self-Supervised Learning and Fine-Tuning



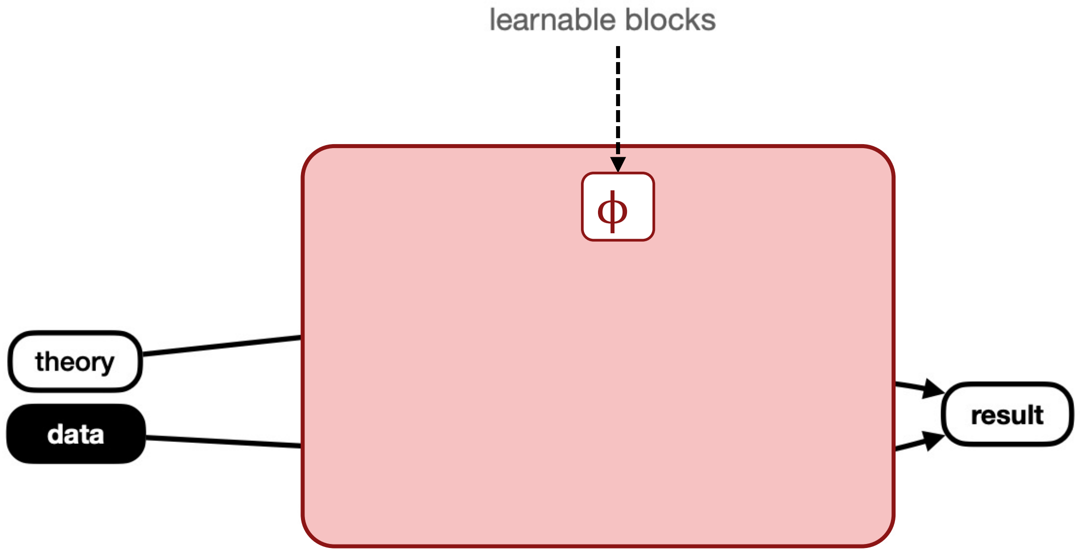
Pretext Task

Downstream Task

Replace everything with one BIG neural Network?

I don’t believe this will work…

At least not with any “reasonable” model size, data size, compute



Recent work: 2508.19190 (backup slide)

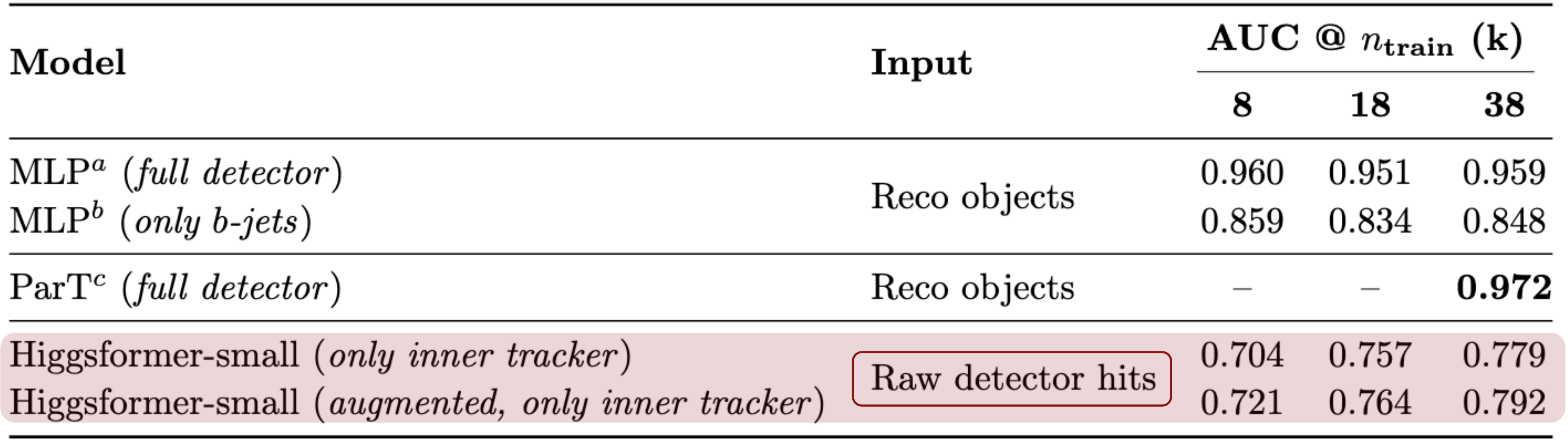
Very large gap between using only raw data versus using some physics knowledge

Hits ⟶ Higgs: Yes? No?

Replace everything with one BIG neural Network?

Still enormous gap between using only hits
and using some physics knowledge

Causation vs correlation
How to use knowledge without over-constraining systems?

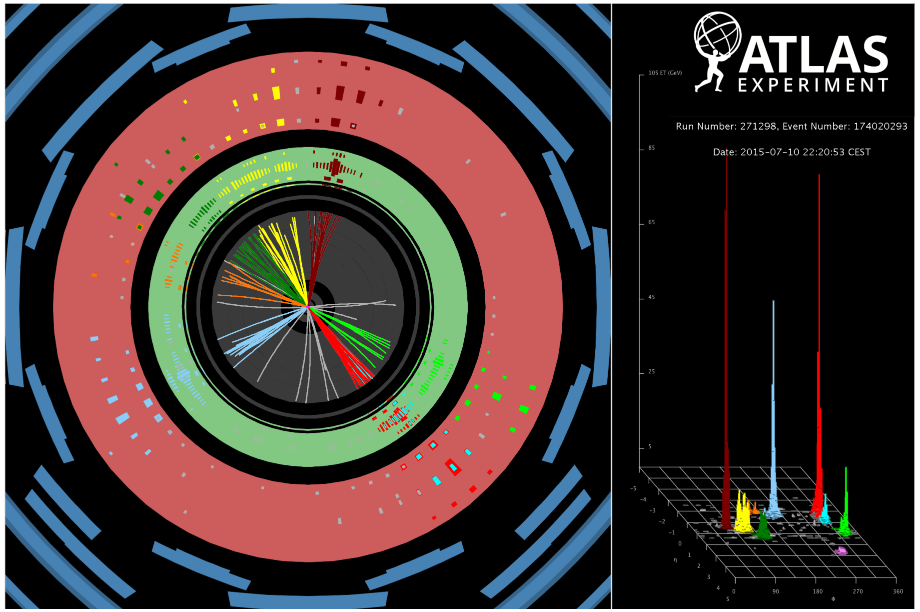


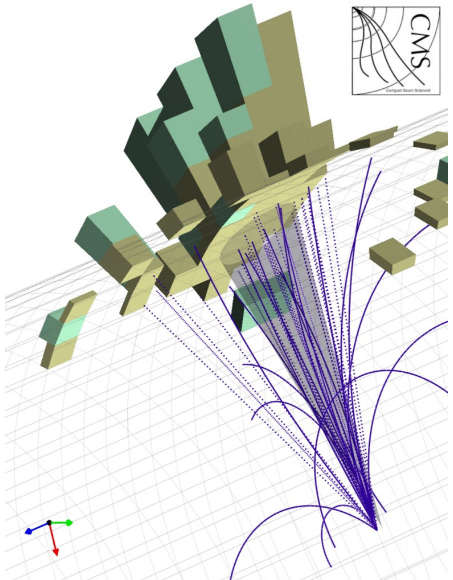
Model	**Input**	**AUC @ n_{train} (k)**		
		8	**18**	**38**
MLP^a (*full detector*)	Reco objects	0.960	0.951	0.959
MLP^b (*only b-jets*)		0.859	0.834	0.848
ParT^c (*full detector*)	Reco objects	–	–	**0.972**
Higgsformer-small (*only inner tracker*)	Raw detector hits	0.704	0.757	0.779
Higgsformer-small (*augmented, only inner tracker*)		0.721	0.764	0.792

Transformers classifiers operating only on hits

2508.19190

A note about data: Jets





Jet = Unordered set of particles

Each particles has a list of features:

Particle = {momentum, direction, position, … }

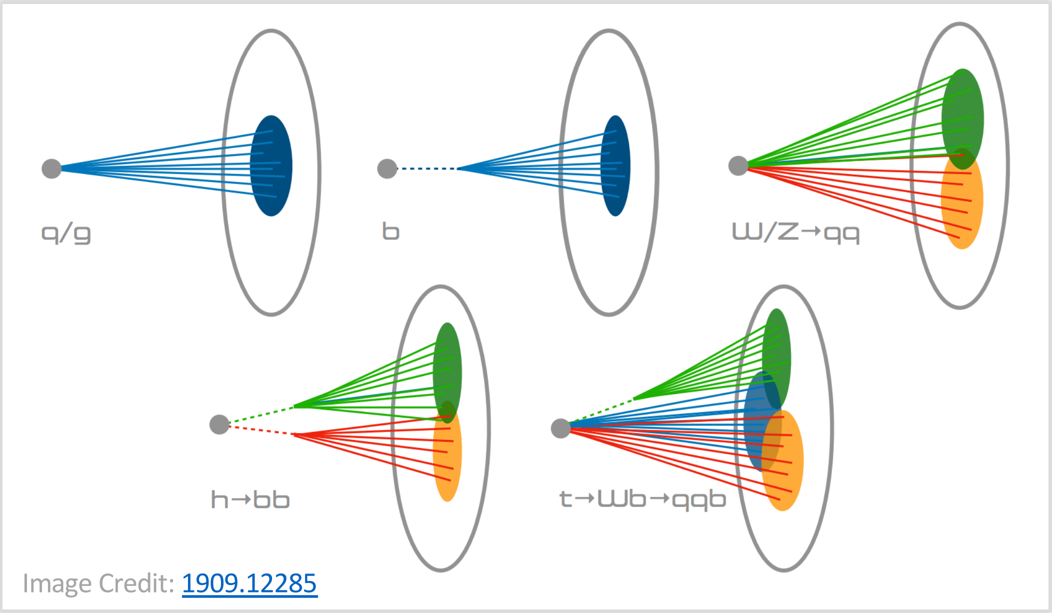
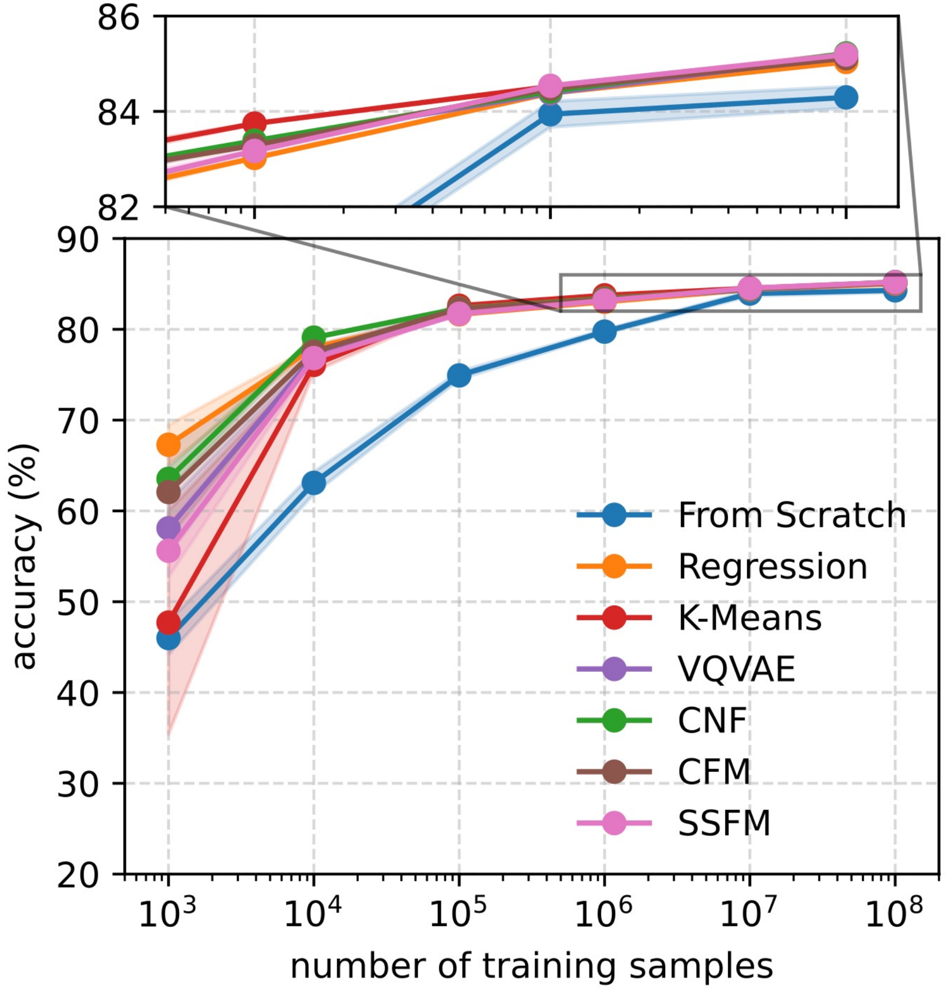


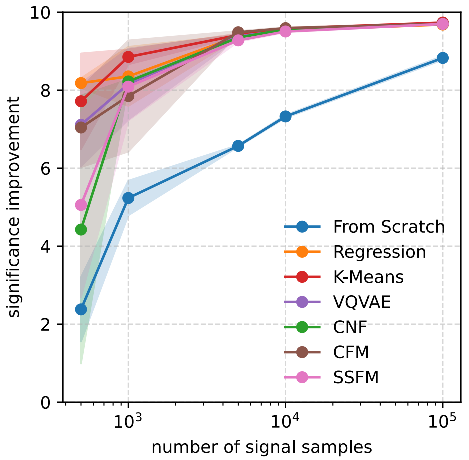
Image Credit: 1909.12285

Transfer to many tasks

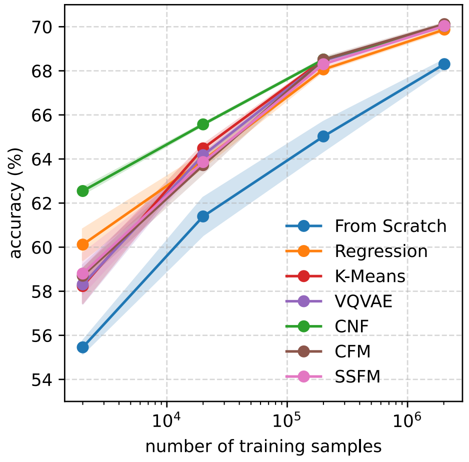
10-Class Jet Classification with Transformer Head



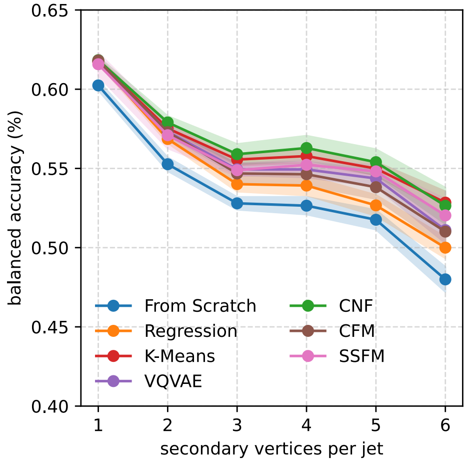
Weakly Supervised Classification



Out-of Distribution Classification on different data set (b-tagging)



Segmentation (vertex finding)



Element-wise (track) classification