WHAT IF WE COULD USE LLM-LIKE MODELS TO DO ANOMALY DETECTION?




Ambre Visive
ambre.visive@cern.ch
Nikhef
Universiteit van Amsterdam

LLM ARE GOOD AT LEARNING DATA DISTRIBUTION, RIGHT?

Ambre Visive

ambre.visive@cern.ch

Nikhef

Universiteit van Amsterdam

EuCAIFCon 2025

LLM ARE GOOD AT LEARNING DATA DISTRIBUTION, RIGHT?

So why not making them learn the particle physics data distribution

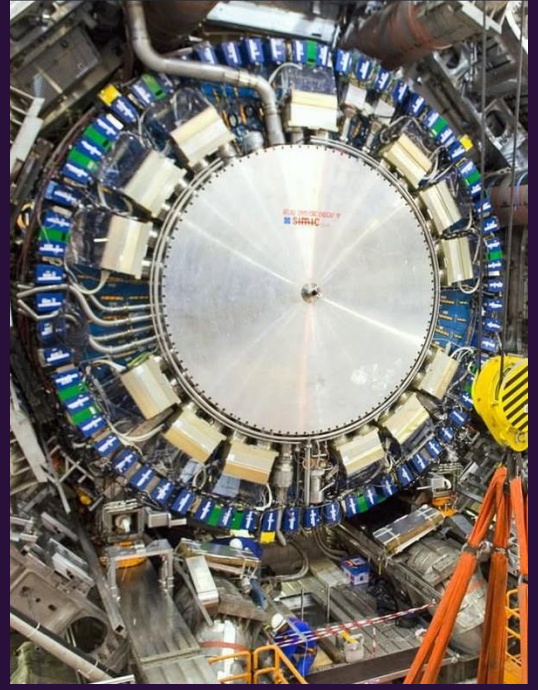
Ambre Visive

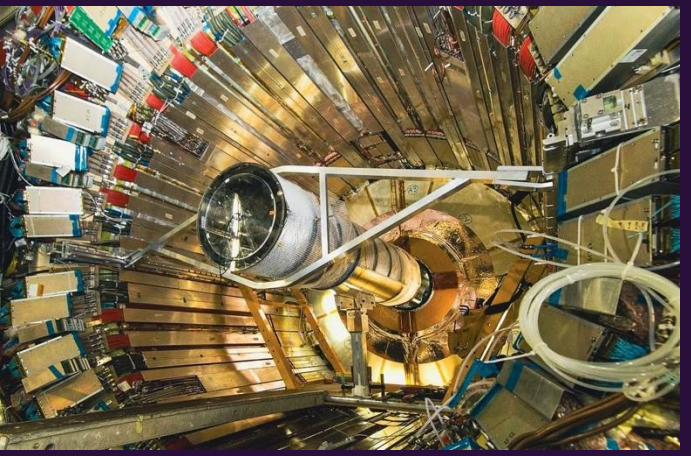
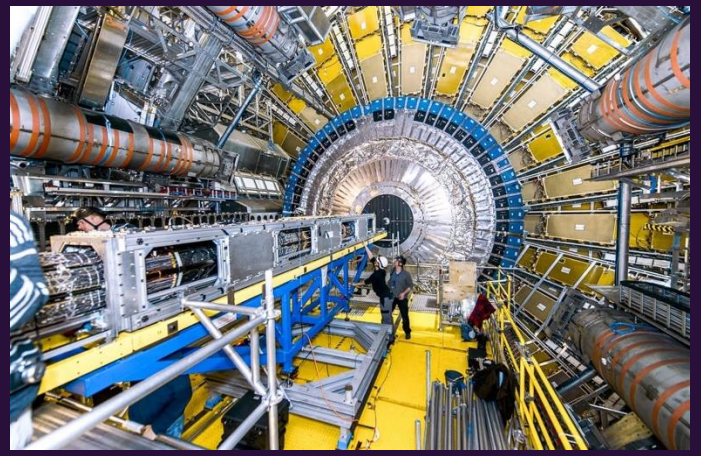
ambre.visive@cern.ch

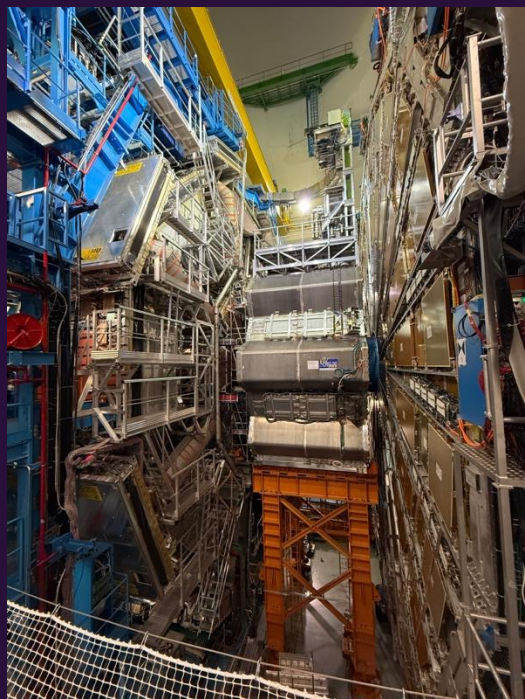
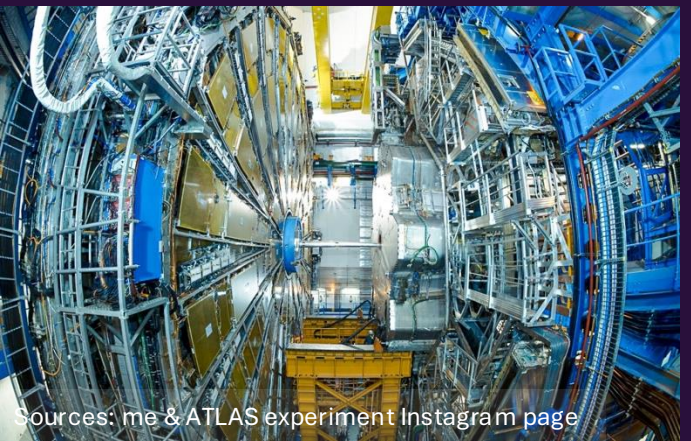
Nikhef

Universiteit van Amsterdam

EuCAIFCon 2025

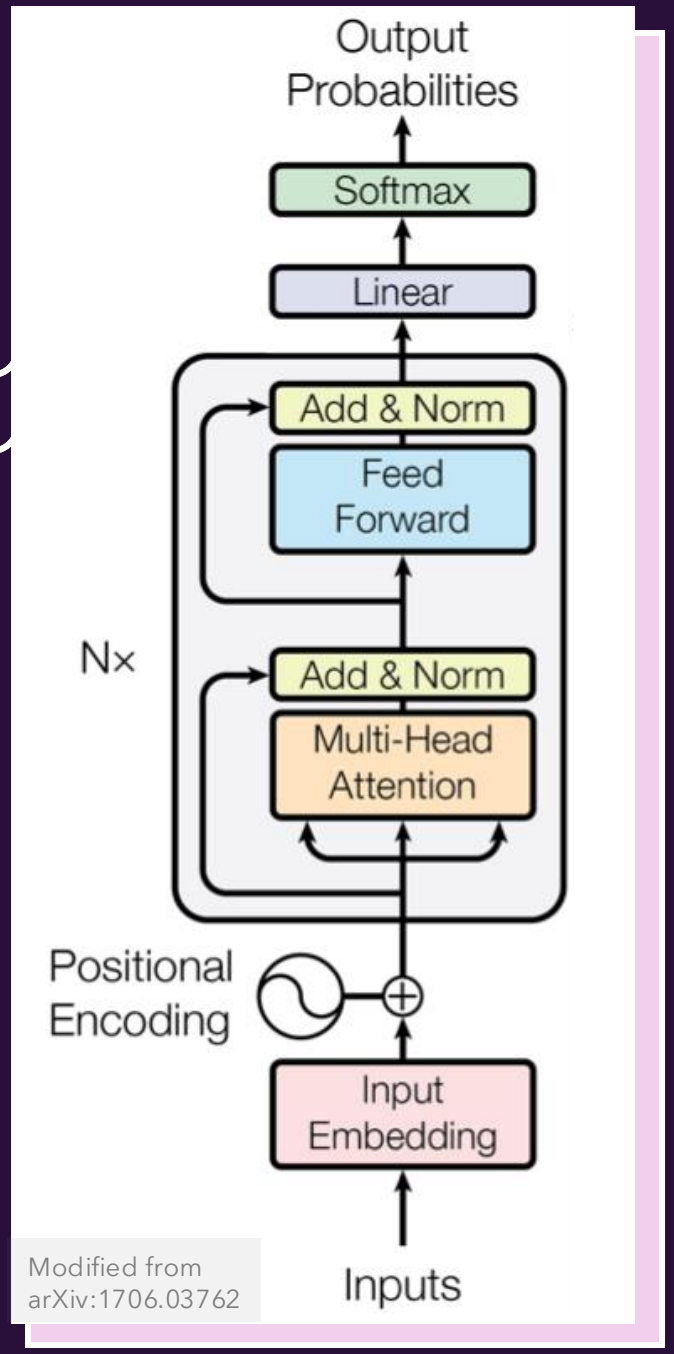





Sources: me & ATLAS experiment Instagram page

IDEA:
Build the encoder part of a simple LLM-like model

Use it to separate background from signal and more generally for anomaly detection



Learn distribution of background data

WHAT EXACTLY ARE WE DOING?

- Train our LLM-like model to reconstruct a masked particle (from an event) -like BERT

WHAT EXACTLY ARE WE DOING?

- Train our LLM-like model to reconstruct a masked particle (from an event) -like BERT

- We train only on background events

WHAT EXACTLY ARE WE DOING?

- Train our LLM-like model to reconstruct a masked particle (from an event) -like BERT

- We train only on background events

- Model is able to reconstruct the masked particle of background events and it should struggle to reconstruct the masked particle of other events

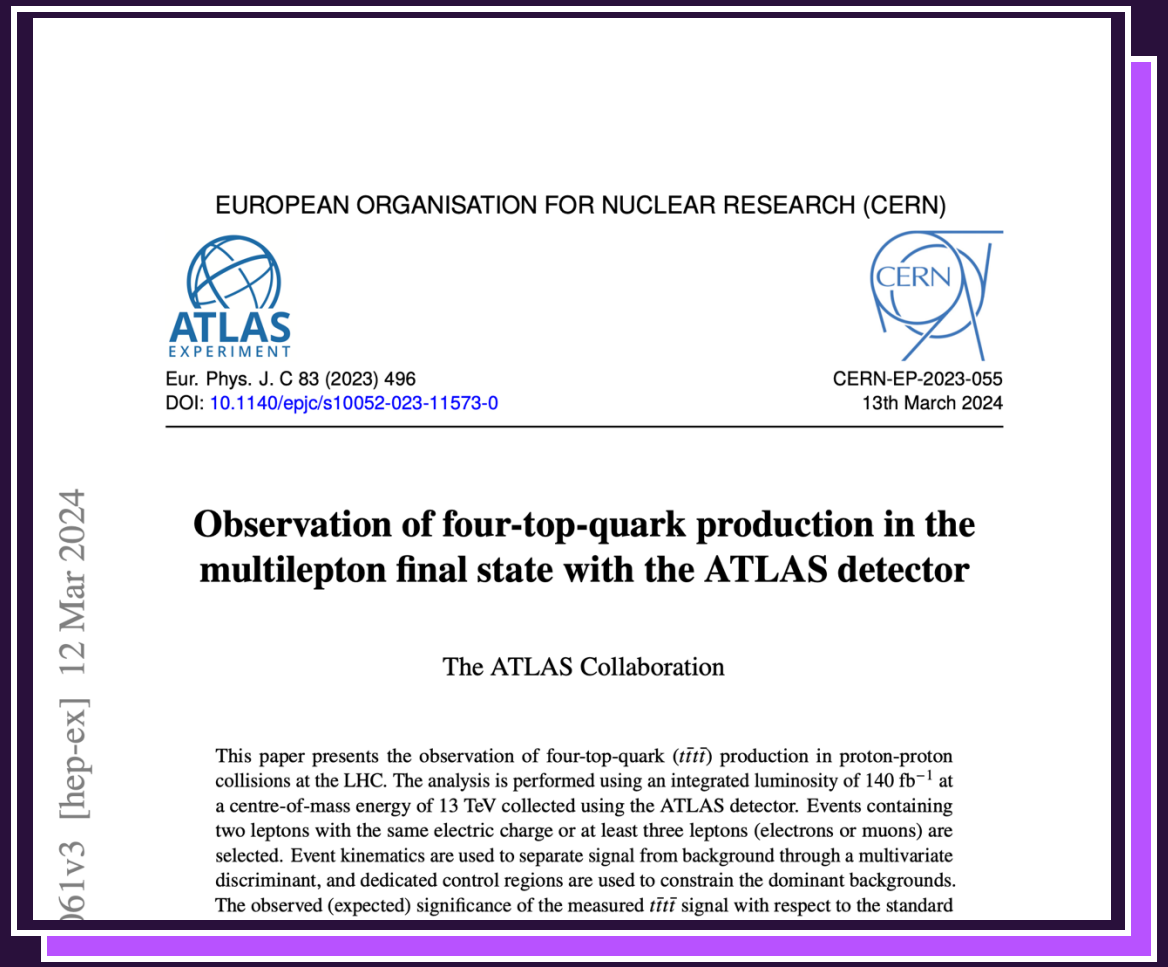
WHAT EXACTLY ARE WE DOING?

- Train our LLM-like model to reconstruct a masked particle (from an event) -like BERT
- We train only on background events
- Model is able to reconstruct the masked particle of background events and it should struggle to reconstruct the masked particle of other events
- At inference, gives both signal and background data to get an anomaly score

WHAT EXACTLY ARE WE DOING?

- Train our LLM-like model to reconstruct a masked particle (from an event) -like BERT
- We train only on background events
- Model is able to reconstruct the masked particle of background events and it should struggle to reconstruct the masked particle of other events
- At inference, gives both signal and background data to get an anomaly score
- Use this anomaly score on real data

SO WHERE IS THE GREAT ARTICLE, RIGHT?



EUROPEAN ORGANISATION FOR NUCLEAR RESEARCH (CERN)

Eur. Phys. J. C 83 (2023) 496
DOI: 10.1140/epjc/s10052-023-11573-0

CERN-EP-2023-055
13th March 2024

Observation of four-top-quark production in the multilepton final state with the ATLAS detector

The ATLAS Collaboration

This paper presents the observation of four-top-quark ($t\bar{t}t\bar{t}$) production in proton-proton collisions at the LHC. The analysis is performed using an integrated luminosity of 140 fb^{-1} at a centre-of-mass energy of 13 TeV collected using the ATLAS detector. Events containing two leptons with the same electric charge or at least three leptons (electrons or muons) are selected. Event kinematics are used to separate signal from background through a multivariate discriminant, and dedicated control regions are used to constrain the dominant backgrounds. The observed (expected) significance of the measured $t\bar{t}t\bar{t}$ signal with respect to the standard

Super tricky part

Tokenization

Bottleneck of the model

Super tricky part

Tokenization

Bottleneck of the model

Have to try and evaluate different tokenization strategy

Tokenization strategies

- 6 different tokenization for 4-VECT
  → with/without MET, METphi included in the event
- different numbers of bins for MET, METphi
- VQ-VAE

A colleague focuses on that part.

NOW IF YOU WANT TO KNOW THE ACTUAL RESULTS, YOU KNOW WHAT YOU HAVE TO DO!

Talking to me is also a good option

Come check out my poster!
N°117

Ambre Visive
ambre.visive@cern.ch
Nikhef
Universiteit van Amsterdam