



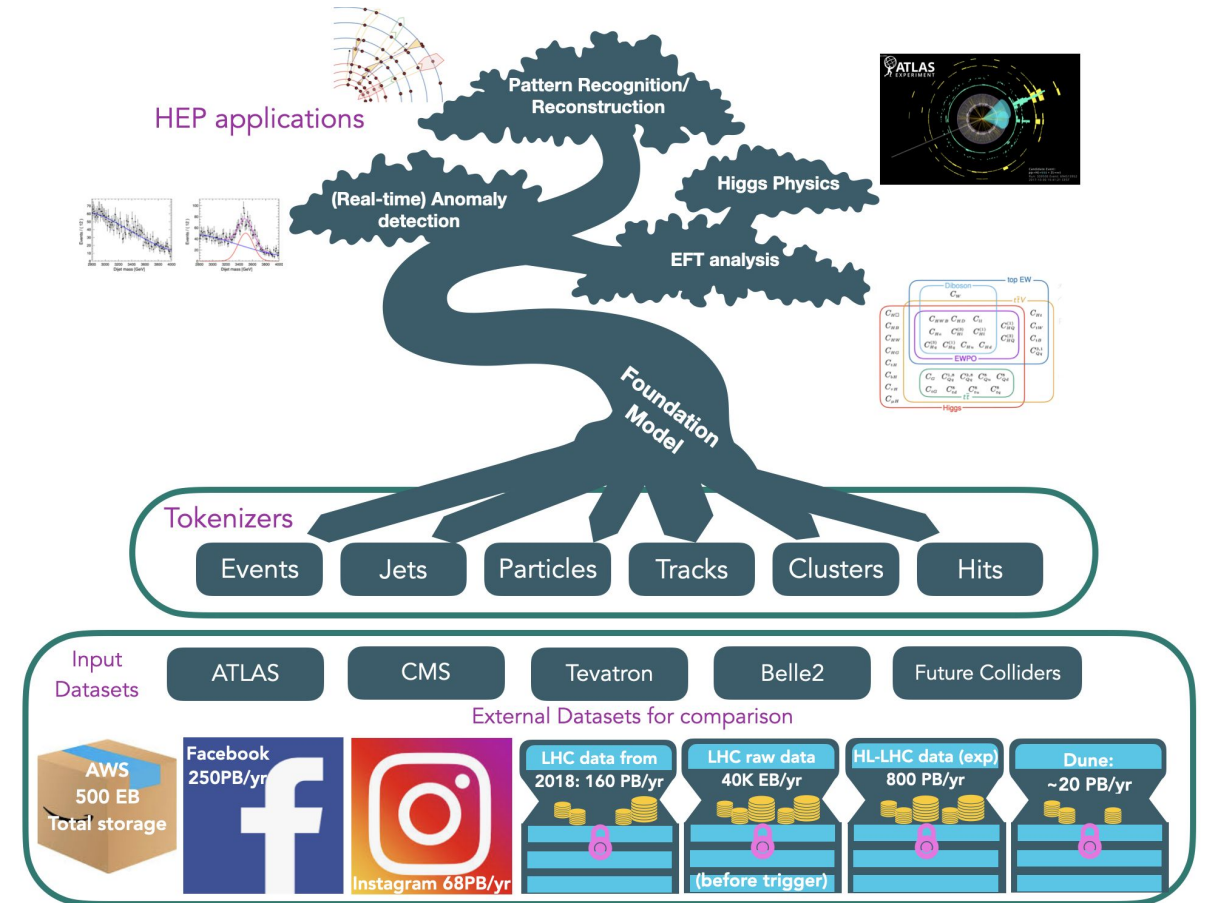
Treasure status

News

- *Github organization:*
 - <https://github.com/orgs/Treasure-AmSC>
 - *Paolo/Michael/Kevin/Walter/Viviana are owners so can make any change*
- *Google drive:*
 - <https://drive.google.com/drive/folders/1LMSFQ4kqSuUlrNC1r3ZdTq-9bnn90fY8?usp=sharing>
 - *You can add stuff there*
- *April 27th in person 3-day workshop at BNL: <https://indico.global/event/16443/>*
 - *Agenda is close to final. Take a look*
 - *Main Objectives:*
 - *What to add to the ATLAS/CMS Open Data ⇒ Beojan is organizing this session, will be great to get in advance what is needed*
 - *Try to get everyone to use the same code ⇒ lots of working time*
 - *Publication plan*

Genesis call for proposal

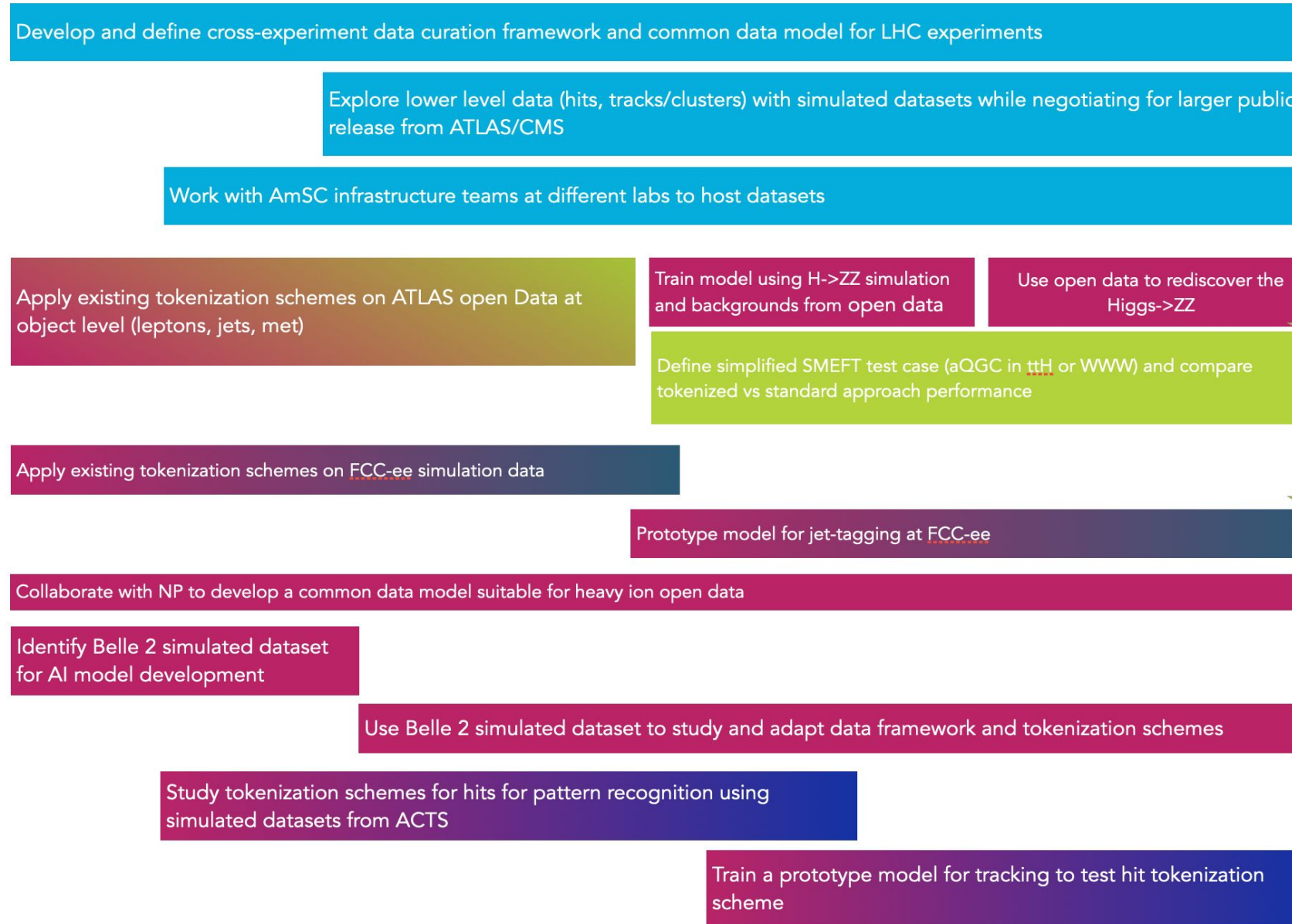
- Treasure is not applying for a Phase 2 in this round
- Plan to participate in different phase 1 and bring everyone back together for a Phase 2 in December



FY26



Legend



Work by all

BNL specific

ANL specific

SLAC specific

LBL specific

FNAL specific



FY26



Legend



BNL Plan for Higgs Physics in ATLAS

Task 1

$H \rightarrow ZZ \rightarrow 4\ell$

Train classifier to separate Higgs signal from background. Direct comparison: standard classifier vs foundation model.

Task 2

$H \rightarrow \gamma\gamma$ or $H \rightarrow bb$

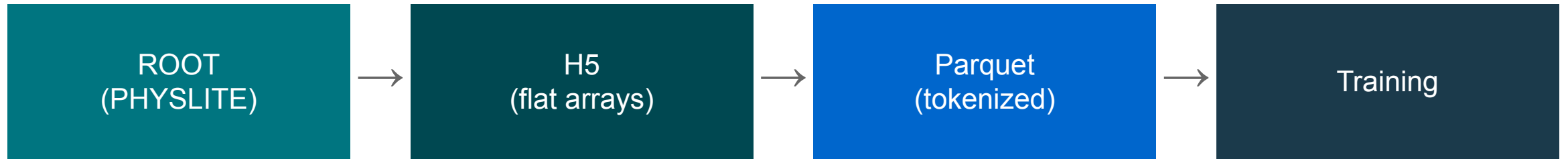
Apply same pre-trained FM with new classification head. Demonstrate transfer learning across Higgs decay channels.

Task 3

Anomaly Detection

Use FM representations for unsupervised anomaly search and test if we would have discovered the Higgs .

Data Pipeline



Physics objects extracted (PCDF-aligned):

Taus

Electrons

Muons

Photons

Jets

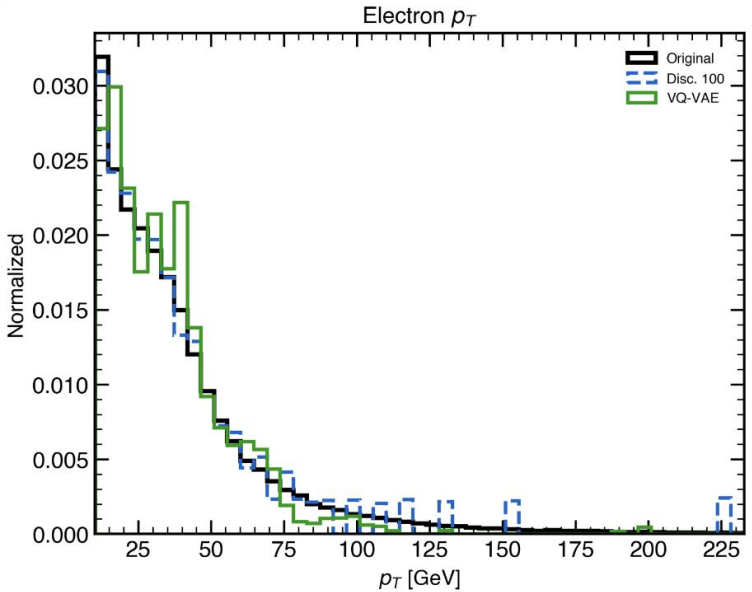
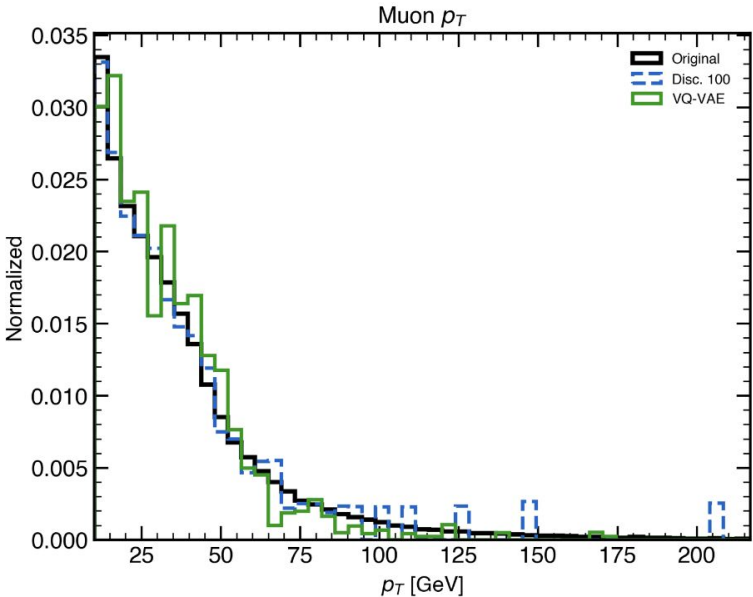
Tracks

MET

3 features

met, phi, sumet

Tokenization Comparison



Electrons

Muons

Photons

Jets

Tracks

MET

Continuous Raw features \rightarrow linear projection into transformer.
Input: H5 direct
No information loss.

Discrete Quantile binning into N bins per feature.
Input: Parquet
Simple and fast, but lossy at bin boundaries.

VQ-VAE Learned codebook per object type (6 separate VQ-VAEs).
Input: Parquet
Inspired by OmniJet-alpha (arXiv:2412.10504).

Very first try, trained on data and MC (HZZ and ZZ continuum)

Task : $H \rightarrow ZZ \rightarrow 4\ell$

Rediscover the Higgs boson in the 4-lepton channel using ATLAS 2015-2016 open data.

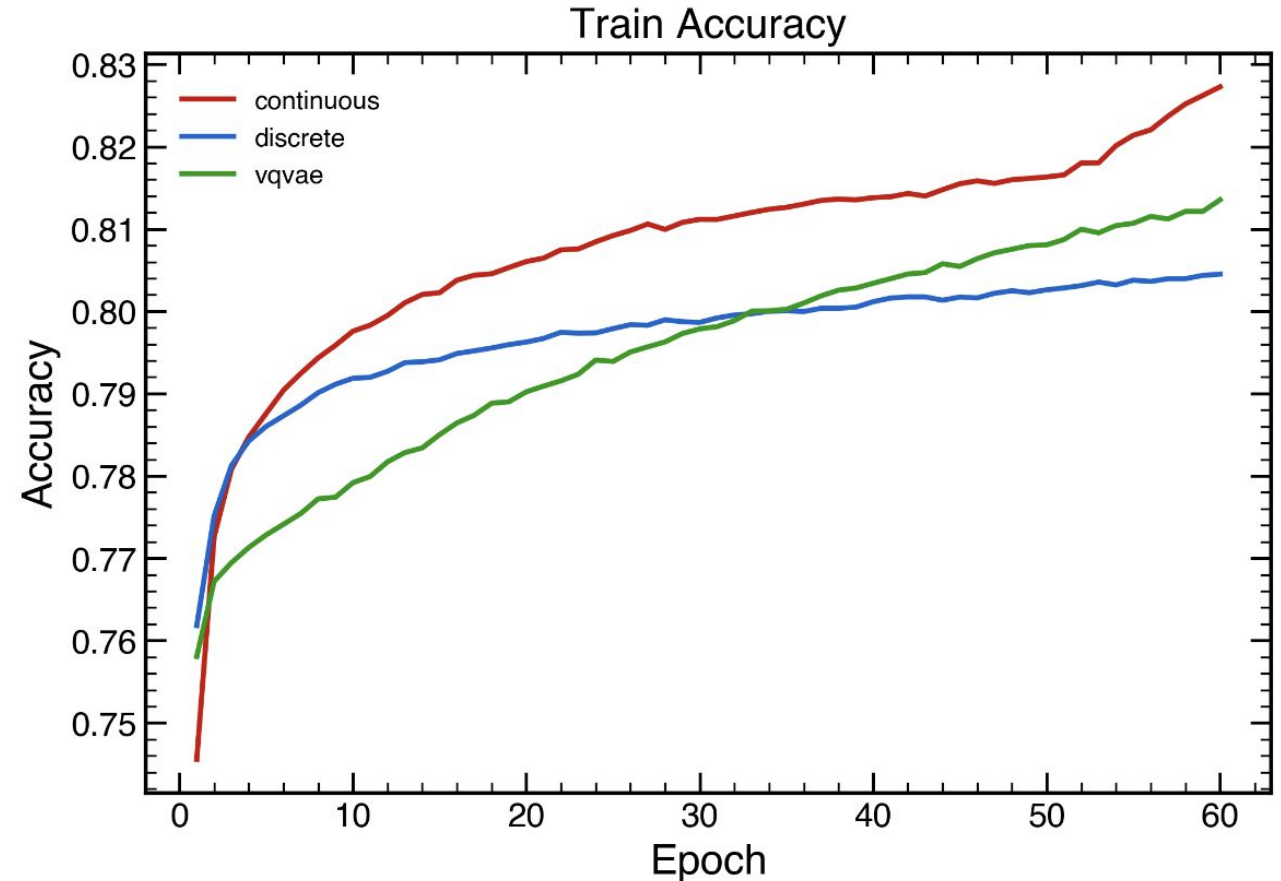
Direct Classifier

Train transformer from scratch on labelled signal vs background

Signal: $H \rightarrow ZZ \rightarrow 4\ell$ MC

Background: ZZ continuum, Z+jets, etc.

Baseline performance



BNL status

- *ATLAS* → *Punit Sharma*
 - *working on Tracking with ML and using ACTS simulated datasets to see if it can be included with FM4NPP*
- *Belle 2* → *Riccardo*
- *FCC-ee Andrea*
 - *working on tokenizing a simulated jet dataset*

