One (important) of many
downstream task

WG1: Foundation models
(for scientific discovery)

Tobias Golling, University of Geneva

Lukas Heinrich, Technical University of Munich

Keen interest in FMs

20+ attendants

Very interactive

Thanks everyone !

Mission / goals

Facilitate research on large-scale foundation models (FMs) for fundamental physics

Provide infrastructure, resources, data and models, connect researchers, define problems & metrics

The Vision

Data sets / benchmarks (MNIST) / challenges (?):
- Jets
- Events
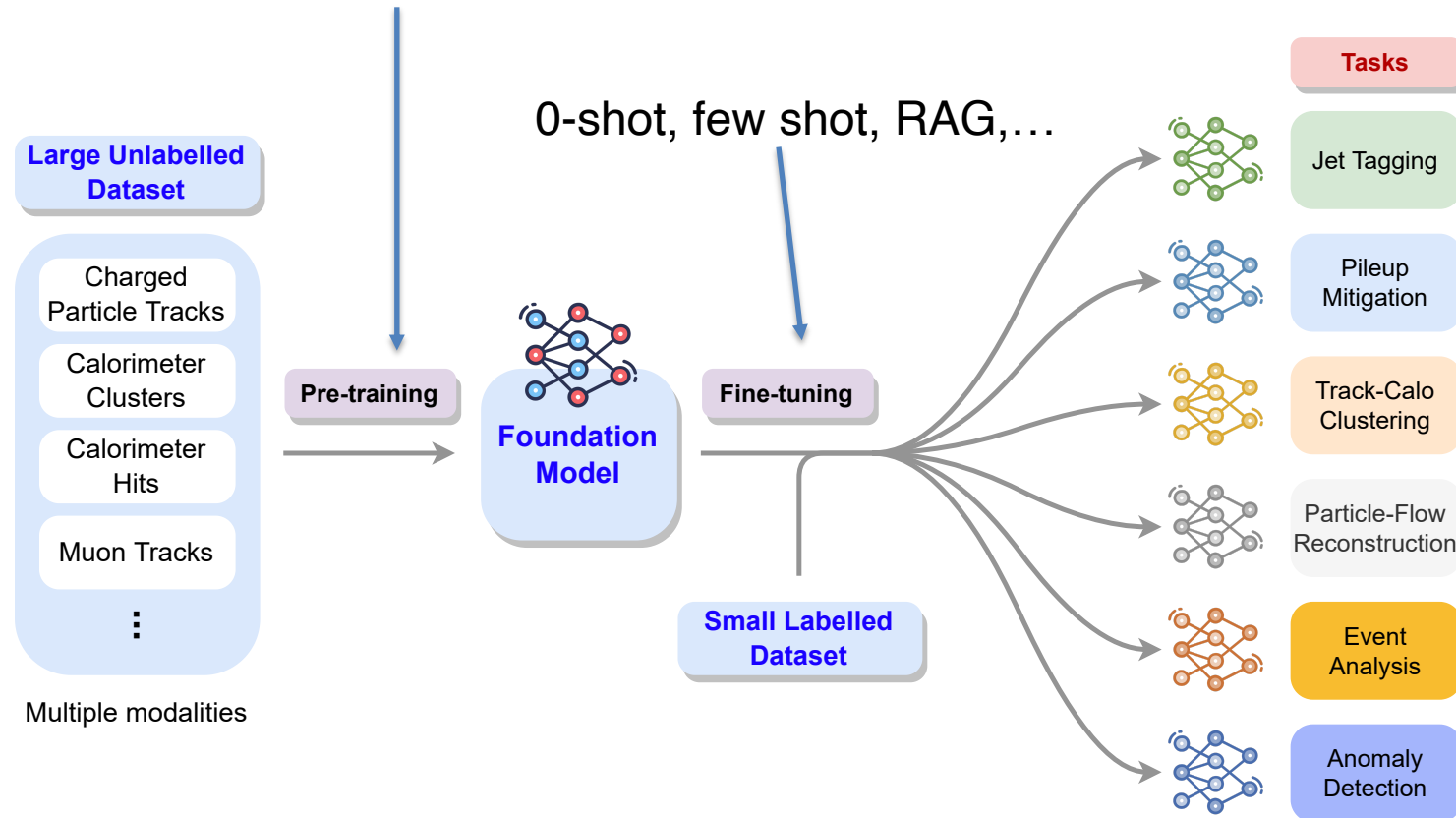- GW?
- Astro?
- Nuclear?
- Theory?

Evaluation metrics:
- Go beyond downstream tasks?
- Continued monitoring…

SSL, tokenization, transformers (& beyond?), (flexible) inductive bias?...

More to do:
- More features
- Scale up 100M →10B…
- Multi-modal & -scale
- Common latent space (align?)
- Language modality?
- Adopt industry models to physics
- FM → agents of experts
- Interpretability
- Explore & exploit (→ experiment)

Set example for AI infrastructure:
- Platform: Huggingface?
- Common software stack
- Common workflow management
- …

0-shot, few shot, RAG,…

Large Unlabelled Dataset

Charged Particle Tracks

Calorimeter Clusters

Calorimeter Hits

Muon Tracks

⋮

Multiple modalities

Pre-training

Foundation Model

Fine-tuning

Small Labelled Dataset

Tasks

Jet Tagging

Pileup Mitigation

Track-Calo Clustering

Particle-Flow Reconstruction

Event Analysis

Anomaly Detection

Demonstrator: end-to-end prototype pipeline, modular, well-defined deliverables (individual papers!) & match with interests, add realism,…

4

Next steps

- Please sign up: https://bit.ly/eucaifcon24-wg1 & **provide input** (FTE, compute,…)
- Email list
- Slack ($ by EuCAIF?)
- Periodic meetings
- Propose **in-person meeting at CERN** in coming months

- Think big, aim high !
- **Bottom-up**: what interests **YOU**, **YOU** shape effort, community consensus
- Collective > individuals
- WG1 provides structure

- Understand compute needs
- Funding opportunities…

Why foundation models?

What Matt said:

2. What makes a question interesting?

- It connects to nature
- You can make progress on it
- Someone else thought it was interesting

From Savannah: **Can/should we automate science?**

Great panel yesterday – mission statement to come!

Let's talk about FMs, write white paper, make the case !

Stroopwafels & Hagelslag

Common shared portable model, automation, efficient science, end-to-end, differentiable, democratize,…