CERN
openlab

Foundation Model

Renato Cardoso, Sofia Vallecorsa

Work realized in collaboration with IBM

Foundation Models

- A model trained on broad data and adaptable to a range of different downstream tasks, zero-shot, few-shot learning.
- Foundation Models concepts:
  - self/semi-supervised learning + transfer learning but at scale:
    - Billions of parameters and gigabytes of data
    - Large and diverse datasets → powerful representations
- Examples:
  - BERT (340M params.), GPT-2, GPT-3 (175B params.) – Generative language models
  - CLIP – Language-Image pre-training
  - DALL-E, DALL-E 2, Imagen – Text to Image models
  - GATO – Sequence to sequence model

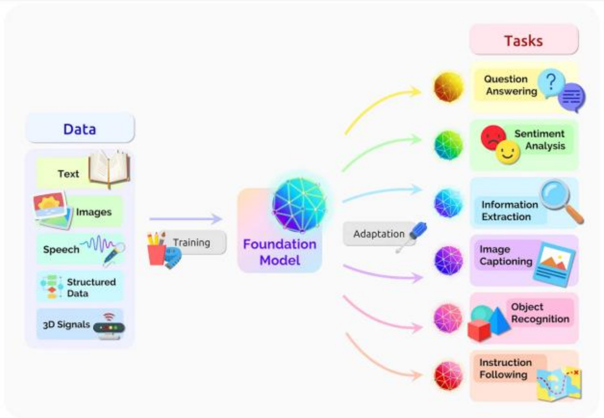


Image obtained from:
On the Opportunities and Risks of Foundation Models

- Stanford CRFM (2021) : On the Opportunities and Risks of Foundation Models [arxiv.2108.07258]

Foundation Models

Why use Foundation Models:

- ML is computational expensive
    - Train once. Then, adapt to new detector geometries, quickly.
- Transformers as building block in foundation models:
    - A generalized architecture without any inductive bias
    - Model long-range dependencies (Attention mechanism)
    - Permutation invariant
    - [arXiv:1706.03762]

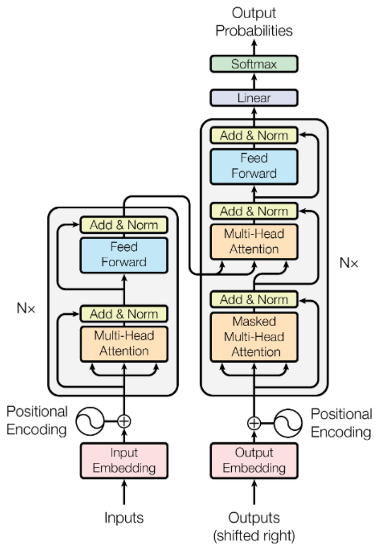


Figure 1: The Transformer - model architecture.

Our Objective:

- Foundation model trained on MC data to perform different physics related tasks
    - Simulations - one lengthy training, then fast adaptation to different detector geometries
    - Reconstruction - one base model adaptable to different tasks (particle identification, regression on phys. variables, etc.)
- Understand how foundation model concept apply to our use case:
    - Understand the minimal scale of the model for reaching meaningful results (No need to reach BERT / GPT-3 scale)

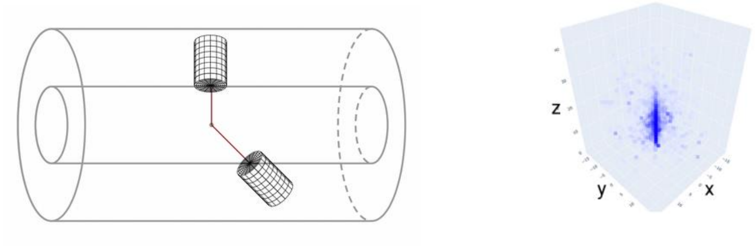
Work done

Dataset: High Granularity Electromagnetic Calorimeter Shower Images

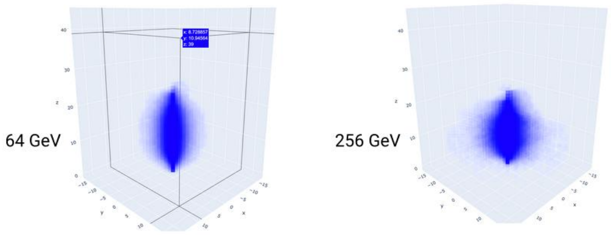
Our first task Foundation model for fast and accurate calorimetry simulation

Single dataset training multiple model architectures:

- Vision Transformer (ViT) based architecture [arXiv:2010.11929]
  - Masked Model
- VAE-like learning model with transformers
- Graph neural network
- VQ-VAE model [arXiv:1711.00937]
- DDPM model [arXiv:2006.11239]
- Other tests:
  - Preprocessing
  - Sinkhorn Loss
  - Regression Loss
  - Etc.



Dataset



Results Obtained from ViT based architecture model

Infrastructure

Why do we need computational infrastructure for this project:

- Models with a high number of parameters
  - High parallelizable but take time to train
- Multiple test being realized simultaneously
  - Multiple people working in the same project
  - Optimization of a single model takes a lot of time with minimal resources
- Memory requirements
  - Big models not only take time to train they need GPUs with a high amount of memory

CERN openlab