Who am I?

I'm **Mohammed Touami,** from **Algeria.**

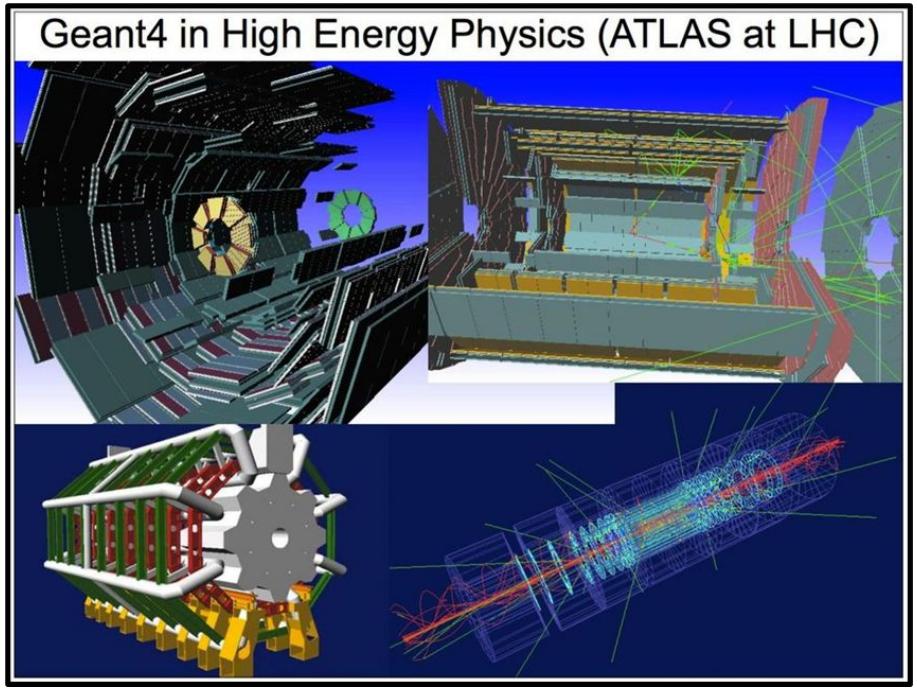
5th Year **Masters** student in **Computer Science**.

I do **Machine Learning** and **Data Science**

Project: **HybridSim**

A **hybrid** particle physics **simulation** framework, a middle ground between **Monte Carlo** and **Generative Models**

Supervised by: **Dalila Salamani, Anna Zaborowska, Piyush Raikwar, Peter McKeown**

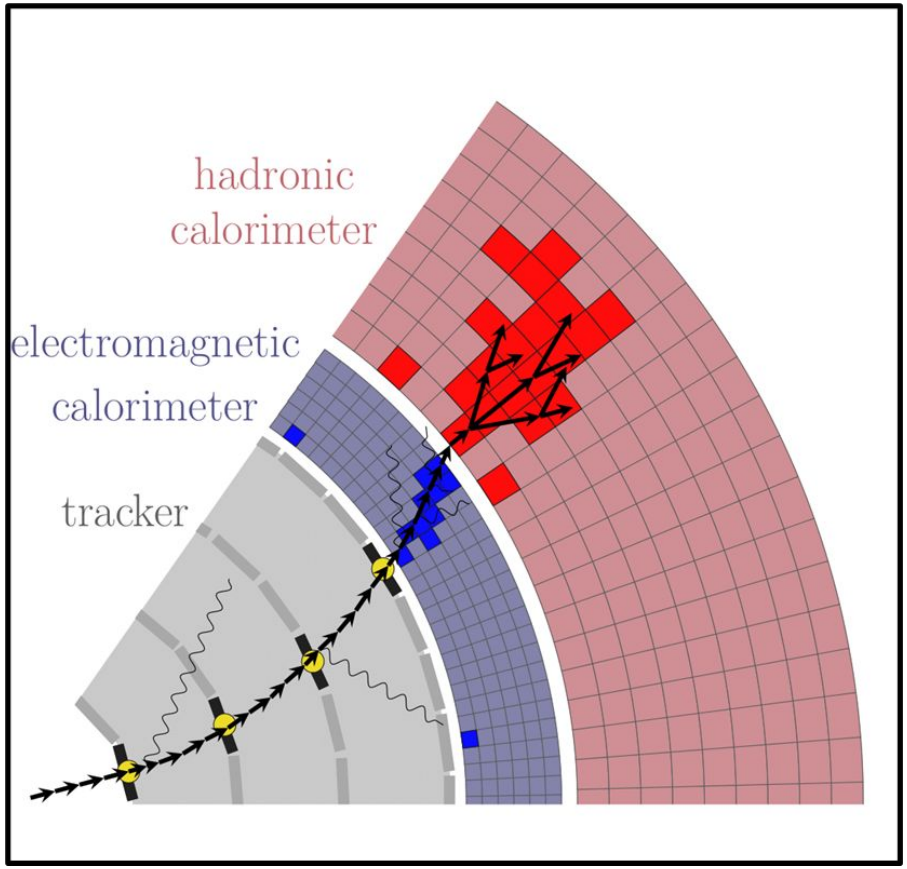


Geant4:

- Simulation toolkit Coded in C++.
- Simulates particle interaction with matter using physics informed probability (Monte Carlo).
- These simulations are called **Full Simulations**.

Monte Carlo:

- Data is collected from real experiments.
- Monte Carlo algorithms are used to calculate the probabilities.
- These simulations are considered to be the baseline (truth) for simulated experiments, as it encodes highest fidelity physics knowledge validated with real experiments.



Particle Cascades:

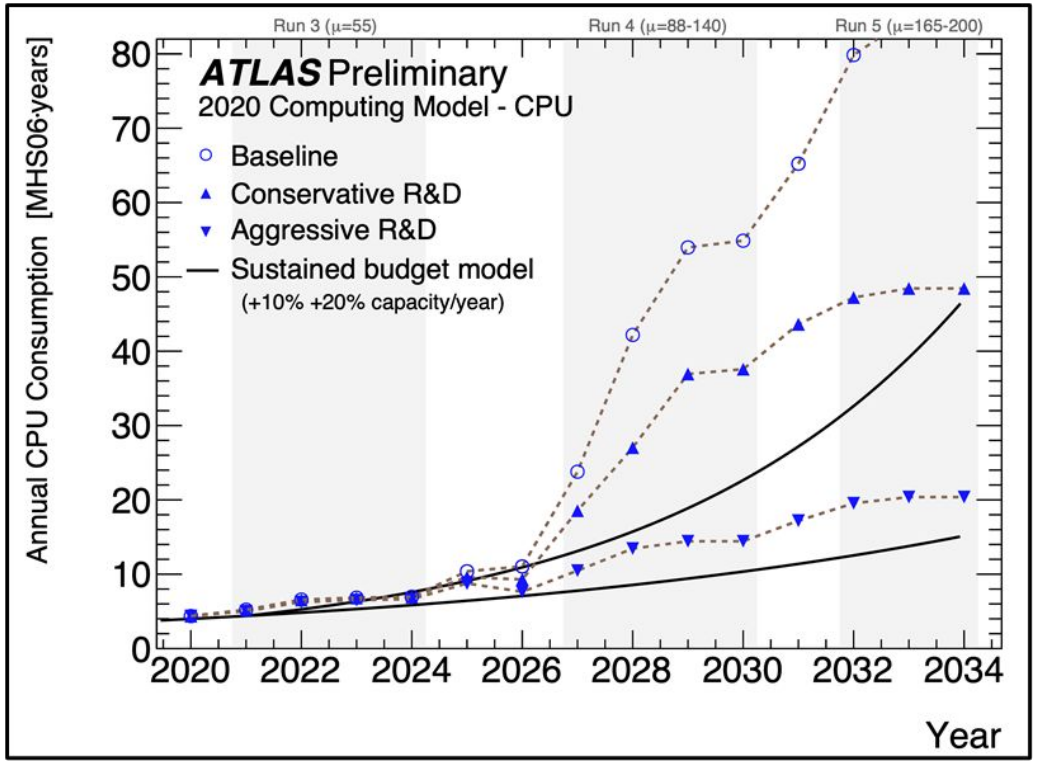
- Particles entering calorimeters break down into showers.
- The shower of sub-particles created by one particle is called Cascade.
- In simulation, cascade generation isn't deterministic but is rather based on probabilities (stochastic).

Main Problems

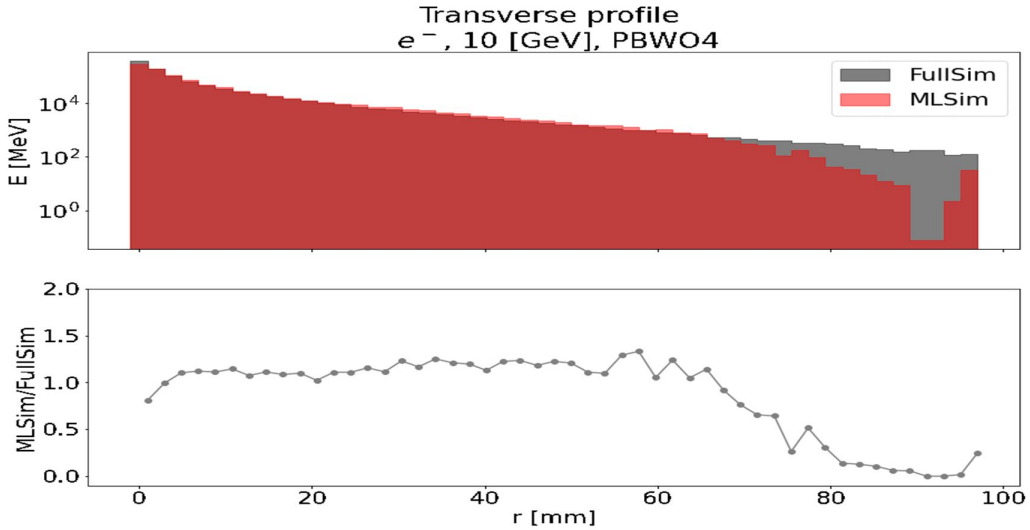
The problem

CPU Limitations

- CERN plan to upgrade current LHC to **HL-LHC** (More **Data**).
- Future **CPU usage** is set to **surpass** current resources.
- **Hardware** upgrades are **limited** by Moore's law.
- Focus on **Software** upgrade is therefore a **must**.



The problem



Previous Works

- Some work on Geant4 uses Generative Machine Learning models (Variational Autoencoders, and Generative Adversarial Networks).
- Previous works see a clear drop off in **accuracy** for large shower sizes, as **Sparsity** is a limitation of voxelised data representation.

Project Objectives

Master Thesis

Remote Contribution

OpenLab

Automate Data Generation

Create optimization algorithms to find optimal (or close to) parameters for Geant4 simulations given a set of constraints.

Find Optimal Data representation

Switch from Voxelized (3D Pixel) cylindrical data structure, to Point Clouds (cylindrical coordinates) based data, or explore other representations.

Find Optimal GenAI model & Hybridize

Train and evaluate generative AI model for low energy showers, integrate into Geant4 and test.

Targets

- Baseline Optimization Algorithm for Data Generation.
- Data representation that resolves sparsity and large tensor issue.
- Generative AI model(s) that is performative, and generalizable on different energies, incident angles, material types and geometries.

The aim of the larger project is to create a machine learning pipeline for Geant4 that would simulate any detector accurately to replace Monte Carlo.

Thank You

Questions ?