$whoami

SANJIBAN SENGUPTA

SFT Group Meeting│ 13 January 2025

Intro

- **25 years old, from Kolkata, India**

Intro

- **25 years old, from Kolkata, India**
- **Education**
  - Bachelor of Technology, *2022*
    (Major: Computer Engineering)
    IIIT Bhubaneswar

Intro

- **25 years old, from Kolkata, India**
- **Education**
  - Bachelor of Technology, *2022*
    (Major: Computer Engineering)
    IIIT Bhubaneswar
- **Previously**
  - Engineering at Voltron Data, Cisco, and several startups working on machine learning and data engineering.
  - Contributing to open-source: ROOT, Apache Arrow, Substrait, Velox, etc.

Intro

- **25 years old, from Kolkata, India**
- **Education**
  - Bachelor of Technology, *2022*
    (Major: Computer Engineering)
    IIIT Bhubaneswar
- **Previously @CERN**
  - Google Summer of Code Student - 2021
  - Summer Student - 2022
  - Mentoring in GSoC for CERN-HSF - 2022, 23, 24

Intro

- **25 years old, from Kolkata, India**
- **Education**
  - Bachelor of Technology, *2022*
    (Major: Computer Engineering)
    IIIT Bhubaneswar
- **Previously @CERN**
  - Google Summer of Code Student - 2021
  - Summer Student - 2022
  - Mentoring in GSoC for CERN-HSF - 2022, 23, 24
- **Hobbies**
  - Theatre/stage play, Poetry, Yoga

What I will be working on?

- **Doctoral Student**
  - at the **University of Manchester**
  - Thesis on **Optimizing inference interfaces for learning models deployed in Next-Gen triggers**
  - **University Supervisors:** Dr. Caterina Doglioni, Dr. Elena Gramellini
  - **CERN Supervisor:** Dr. Lorenzo Moneta

- **ML4EP**
  - Service and support to the experiment on common ML issues
  - **NGT/1.7**
    - Common Software Developments for Heterogeneous Architectures

What I will be working on?

ML4EP

- **Enhancing SOFIE**
  - SOFIE (System for Optimized Fast Inference code Emit) - developed at CERN
  - Memory optimization, extending operator support

- **NGT/1.7**
  - Efficient Accelerator Interfaces to ML Inference
    - Eliminating intermediate data transformations with direct inference -> minimizing execution latency
    - Portability of inference architectures on GPUs
    - Continuous benchmarking suite

Motivation

- Feels great to be back!
- Researching AI+Physics at the forefront
- Learn and contribute

Thanks!

Let's connect!

sanjiban.sengupta@cern.ch   32/S-B15   linkedin.com/in/sanjiban-sengupta/