

NNN25

International Workshop on Next Generation
Nucleon Decay and Neutrino Detectors

October 1-3, 2025



Contribution ID: 98

Type: **Plenary Talk**

Vision Foundation Models and Big, High-resolution Neutrino Detectors

Wednesday, October 1, 2025 12:30 PM (25 minutes)

Foundation Models (FMs), such as Large Language Models (LLMs), have dramatically transformed many aspects of modern workflows. FMs are machine learning (ML) models designed to learn flexible and general representations of data, which can then be adapted to solve a wide variety of tasks. Successful pre-training yields a model with a broad range of applications, and fine-tuning further optimizes for specific tasks. A major challenge, especially relevant to physics, is building FMs for sensory datasets (i.e. 1D waveforms, 2D images, and 3D point clouds). Learning useful representations directly from these more complex and varied sensor data is an active research area.

High Energy Physics (HEP), and neutrino physics offer unique opportunities to push the research boundaries. Massive, high-resolution sensor datasets benefit from precise detector knowledge, making them excellent test beds for developing robust, general-purpose AI models. Current AI models in neutrino experiments are typically task-specific and reliance on simulations can lead to differences between model performance on simulated vs. real data. FMs can be trained directly on real data, producing versatile models that support multiple applications by sharing learned features, addressing both efficiency and consistency challenges.

In this talk, I will present the development of the first sensor-level FMs for simulated 3D point cloud data from a liquid argon time projection chamber (LArTPC) detector. We demonstrate that the representations learned by these models can effectively support multiple downstream applications. Our fine-tuned models match or surpass the performance of state-of-the-art AI models in several benchmarks. We are releasing our trained model, associated software tools, and a large dataset of 1M simulated 3D scenes for public use and reproducibility. I will discuss plans to foster sensor-level FM research in high energy physics as part of broader U.S. initiatives to develop AI-ready scientific datasets, with the goal of benefitting the wider research community.

Submitter Email

stephen.sekula@snolab.ca

Submitter Name

Stephen Sekula

Submitter Institution

SNOLAB, Queen's University, and Laurentian University

Primary author: TERAQ, Kazuhiro (SLAC National Accelerator Laboratory)

Presenter: TERAQ, Kazuhiro (SLAC National Accelerator Laboratory)

Session Classification: Plenary Talks

Track Classification: Plenary Talk: Invited Talk