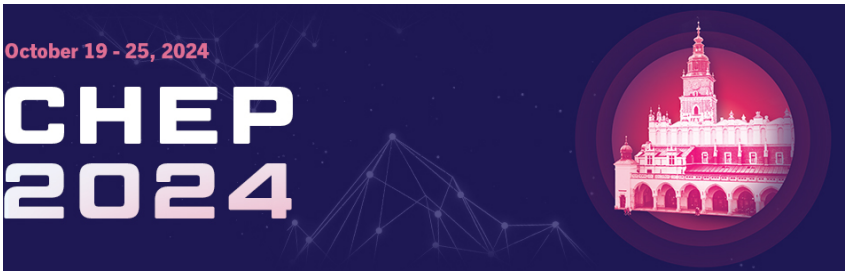

Contribution ID: **90** Type: **Poster**

Machine Learning Inference in Athena with ONNXRuntime

Machine Learning (ML)-based algorithms play increasingly important roles in almost all aspects of data processing in the ATLAS experiment at CERN. Diverse ML models are used in detector simulation, event reconstruction, and data analysis. They are being deployed in the ATLAS software framework, Athena. Our primary approach to perform ML inference in Athena is to use ONNXRuntime. ONNXRuntime is a cross-platform ML model acceleration library, with a flexible interface to integrate hardware-specific libraries. In this talk, we will describe the ONNXRuntime interface in Athena and the impact of advanced ONNXRuntime settings on various ML models and workflows at ATLAS.

Primary authors: KRASZNAHORKAY, Attila (CERN); STANISLAUS, Beojan (Lawrence Berkeley National Lab. (US)); Dr LEGGETT, Charles (Lawrence Berkeley National Lab (US)); ELMSHEUSER, Johannes (Brookhaven National Laboratory (US)); ESSEIVA, Julien (Lawrence Berkeley National Lab. (US)); CALAFIURA, Paolo (Lawrence Berkeley National Lab. (US)); TSULAIA, Vakho (Lawrence Berkeley National Lab. (US)); JU, Xiangyang (Lawrence Berkeley National Lab. (US))

Presenter: JU, Xiangyang (Lawrence Berkeley National Lab. (US))

Session Classification: Poster session

Track Classification: Track 3 - Offline Computing