NATIONAL CENTRE FOR NUCLEAR RESEARCH

Abstract

New simulation software and machine learning technologies in the LHCb experiment to evaluate physics performance of Run 3

Michał MAZUREK

The main goal of the work presented in this thesis was to explore new software and machine learning-based technologies to improve the performance of the LHCb data processing applications and, in particular, the simulation framework of LHCb, GAUSS.

GAUSSINO is the new core simulation framework that was created by extracting all the experiment-independent functionalities of GAUSS. In this work, the GAUSSINO framework was moved from the advanced prototype stage to the production-ready framework, which can act as an ideal test bed for all the new simulation and detector developments in a standalone mode, as well as a robust core simulation framework for HEP experiments. GAUSS-ON-GAUSSINO is the new version of the LHCb simulation framework, based on GAUSSINO. In addition to ensuring the smooth transition to the new simulation framework for Run 3 and beyond, the work included integrating a new interface to fast simulations, adding support for new detector description toolkits (DD4HEP and EXTERNALDETECTOR), as well as new visualization tools and web-based documentation.

In addition to the improvements in the software of the simulation framework, new machine learning-based technologies were explored and integrated into the simulation framework. In particular, infrastructure for running Generative AI (GenAI) models for calorimeter fast simulations in GAUSSINO was integrated. Moreover, the performance of the first, production-ready CALOML+VAE model trained on the LHCb electromagnetic calorimeter data was evaluated. An exploration of the additional use of the ML models in object detection algorithms for cluster energy reconstruction in the LHCb electromagnetic calorimeter was also presented.

Finally, validation of the new simulation framework and machine learning-based fast simulation techniques was presented using a few relevant LHCb decay channels in the last chapter of this thesis. Validation was done with respect to samples produced with GAUSS framework when using the well-tested GEANT4 toolkit.