Contribution ID: **47** Type: **Contributed talk**

# Reinforcement learning for automatic data quality monitoring in HEP experiments

*Friday 2 February 2024 11:10 (20 minutes)*

The usage of modern ML techniques to automate the search for anomalies in collider physics is a very active and prolific field. Typical cases are the search for signatures of physics beyond the Standard Model and the identification of problems in the detector systems that would lead to bad-quality data, unusable for physics data analysis. We are interested in the second type of task, which can also be referred to as data-quality monitoring. In large experimental collaborations, this kind of anomaly detection usually relies on large pools of rotating shifters, taken from within the members of the collaboration. Great benefits can be gained by the partial automation of those tasks, in terms of both an increased efficiency for the collection of good data and a reduction in the need for associated person power.

Besides the usual challenges in the detection of anomalies with ML, additional difficulties arise in situations where the nominal experimental conditions are rapidly changing, for example during the period of commissioning of a new detector. In such a case, the algorithms need to be continuously retrained in an efficient manner. Additionally, if the optimisation goal doesn't only look at the data-collection efficiency but includes human factors (for example, trying to reduce the need for redundant shifter actions), the definition of an adequate loss is not trivial. To face these extra challenges, we propose the application of Reinforcement Learning techniques with human feedback to the task of data-quality monitoring.

In this contribution, we describe a simplified simulated setup designed to study the automation of data-quality monitoring in two regimes, "online"and "offline". The "online"one deals with the problem of spotting problems in a detector while data is being collected, aiming at a prompt fixing that will increase the future data-collection efficiency. The "offline"one focuses on the problem of classifying data that has already been collected as usable or unusable. We present the progress on the application of RL algorithms in those regimes, discuss the performance achieved and identify future lines of work.

## Would you like to be considered for an oral presentation?

Yes

**Authors:** GARCIA PARDINAS, Julian (CERN); DEL PIANTA, Lorenzo; JANISCH, Maximilian; SERRA, Nicola (University of Zurich (CH)); JULLIAN PARRA, Olivia (CERN); KLAVER, Suzanne (Nikhef National institute for subatomic physics (NL)); LEHÉRICY, Thomas

**Presenter:** JULLIAN PARRA, Olivia (CERN)

**Session Classification:** Contributed Talks

**Track Classification:** 2 ML for analysis : event classification, statistical analysis and inference, including anomaly detection