Contribution ID: **12**                                                                 Type: **not specified**

# Fairness

*Friday, 8 November 2024 14:10 (20 minutes)*

In social sciences, fairness in Machine Learning (ML) comprises the attempt to correct or eliminate algorithmic bias of gender, ethnicity, or sexual orientation from ML models. Many high-energy physics (HEP) analyses that search for a resonant decay of a particle employ mass-decorrelated event classifiers, as the particle mass is often used to perform the final signal extraction fit. These classifiers are designed to maintain fairness with respect to the mass, which is accomplished primarily by retaining mass-correlated information during training.

Our studies present a first proof-of-concept for systematically applying, testing and comparing fairness methods for ML-based event classifiers in HEP analyses. We explore techniques that mitigate mass correlation during and after training. Through simulations and a case studies, we demonstrate the effectiveness of these methods in maintaining fairness while preserving the classifier performance.

**Primary author:** DE VRIES, Karel

**Presenter:** DE VRIES, Karel

**Session Classification:** Parallel Sessions (II)