Contribution ID: **3** Type: **Poster + Flashtalk**

Adversarial Machine Learning for Robust Event Classification in Particle Physics

Adversarial machine learning is a collection of techniques used to study attacks on machine learning algorithms. It is commonly used in cybersecurity and still has few applications in fundamental physics. Since systematic effects in detector data, often absent in Monte Carlo simulations, challenge the performance of machine learning models in particle physics, in this work we model the detector as an "adversary", either as an "enemy" introducing realistic perturbations to expose model vulnerabilities or as a "friend" simulating systematic effects during training. This dual approach forces classifiers to learn invariant features, improving robustness and accuracy on real data. Applied to simulated detector events, our method demonstrates superior generalization compared to traditional approaches, addressing systematic uncertainties in experimental science.

AI keywords

simulation-based inference; adversarial machine learning

Primary author: BRUNDU, Davide (Istituto Nazionale di Fisica Nucleare)

Presenter: BRUNDU, Davide (Istituto Nazionale di Fisica Nucleare)

Track Classification: Inference & Uncertainty