ML for hadronization (MLHAD) and tuning (Mrenna)*

*work supported by NSF grants through associated position at UCinc

Theme: ML to reweight good models to be better models

SubTheme: Models as differentiable as possible using Reweighting

Rejection Sampling with Autodifferentiation - Case study: Fitting a Hadronization Model (LBL + UCinc)

MC rejection sampling is paired with autodifferentiation to reweight generated samples to data in the fitting process.

Describing Hadronization via Histories and Observables for Monte-Carlo Event Reweighting (UCinc)

HOMER: train a classifier on simulation and data (physical observable), infer single fragmentation weights, and calculate the weight for the full hadronization chain (generator level non-observables).

Towards a data-driven model of hadronization using normalizing flows (UCinc)

Hadronization model based on invertible NNs. Reproduces a simple Lund string model for meson hadronization.

MAGIC: new training method for normalizing flows that improves the agreement between simulated and experimental distributions of high- level (macroscopic) observables by adjusting single-emission (microscopic) dynamics

Close to being able to train models on DATA instead of SIM

Parametric weighting of MC tools (including GEANT) will be critical!

LLM for answering Pythia questions (Mrenna)

80% of issues raised in git have already been addressed somewhere

15% need our limited attention

5% need our precious dedicated debugging (subtle, time-consuming)

Exploring NotebookLM (and now ollama)

Pythia-specific resources for training
- website
- manual
- git issues
- example programs

Little work – reasonable payoff – awesome AI generated podcast

Bridging the Gap between Simulation and Data

Build ML models based on observed discrepancies between Pythia and data at the LHC that fill in the physics behind these gaps

(1) Corrections to an existing physics model to describe data using reweighting
(2) Symbolic relationships between theory features to construct better physics models.

(1) Use weakly supervised ML to automatically identify collider data (Rivet) that is poorly described by theory predictions and has no obvious physics explanation

(2) Develop an ML-based regression model to describe the gap between theory and data

(3) Construct metamodels, using Simulation-based inference (SBI) for example, to determine correlations between the gap and the latent variables of the generator. These metamodels will be used to supplement the standard generator predictions and provide candidate physical explanations and descriptions of the data.