



FAIR Universe

Fair Universe

Unbiased Data Benchmark Ecosystem for Physics HiggsML Uncertainty Challenge

Wahid Bhimji,
NERSC, Berkeley Lab

Paolo Calafiura, Ragansu Chakkappai, Yuan-Tang Chou, Sascha Diefenbacher, Steven Farrell, Aishik Ghosh, Isabelle Guyon, Chris Harris, Shih-Chieh Hsu, Elham Khoda, Benjamin Nachman, Benjamin Thorne, Peter Nugent, Mathis Reymond, David Rousseau, Ihsan Ullah, Daniel Whiteson

<https://fair-universe.lbl.gov/>



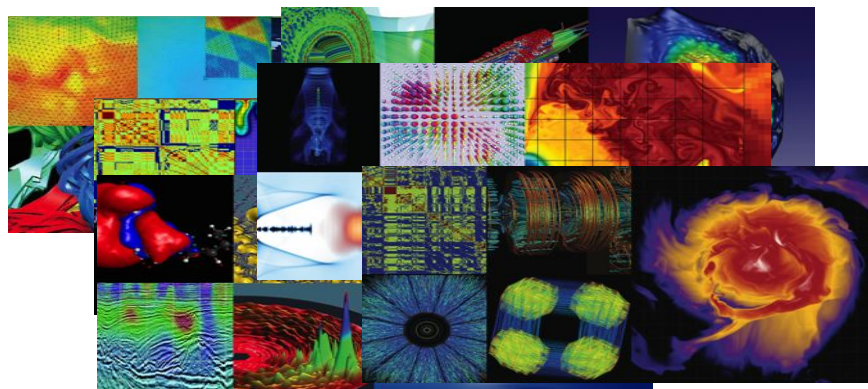
Fair Universe: Unbiased Data Benchmark Ecosystem for Physics

Project aims to:

- Provide a **large-compute-scale AI ecosystem for hosting challenges and benchmarks.**
- **Organize a challenge series** initially focused on **measuring and minimizing the effects of systematic uncertainties** in HEP (particle physics and cosmology).

Large-compute-scale AI ecosystem for
hosting challenges and benchmarks

NERSC: Mission HPC for the Dept. of Energy Office of Science



Large compute systems

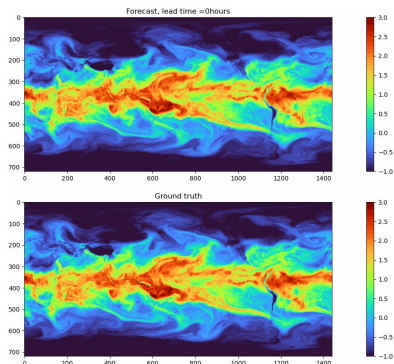
- E.g. Perlmutter: ~7k A100 GPUs
- Also high-capacity/ fast filesystems,
1 Tbit/s WAN and flexible services
- E.g. SPIN: Rancher/K8s platform for user-defined services

Broad science user base

- > 10,000 users,
- > 1000 projects
- Across all DoE Science e.g. HEP; NP; Climate; Fusion Chemistry; Materials; Genomics; etc ...

NERSC-AI Ecosystem:

- **Deploy** optimized hardware and software (working with vendors)
 - Improve performance, e.g through benchmarking ([MLPerf HPC](#))
- **Apply** cutting edge AI for science: e.g. “NESAP” program with postdocs
- **Empower** through e.g. over 20 DL@Scale tutorials, 1000s of total participants: ([SC23](#))
- **Many AI for science highlights** not covered here. e.g.



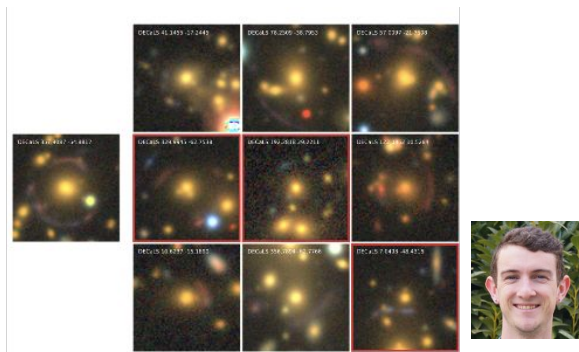
FourCastNet- [Pathak et al. 2022](#)

Collab with Nvidia, Caltech

First DL model with skill of numerical weather prediction (NWP)

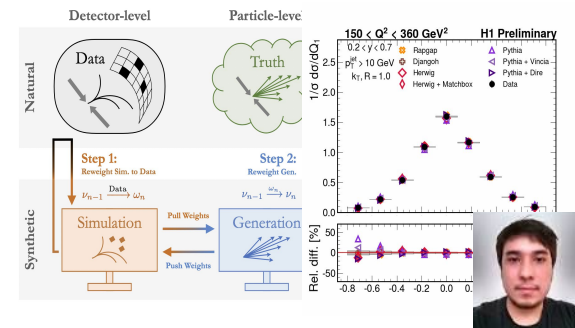
Train up to 1000s of GPUs

Forecasts 1000s times faster than NWP



Self-supervised sky surveys [Stein et. al. \(2021\)](#)

Foundation-like model trained at scale - used for downstream tasks
E.g. uncovered thousands of undiscovered strong-lenses



Unfolding particle physics

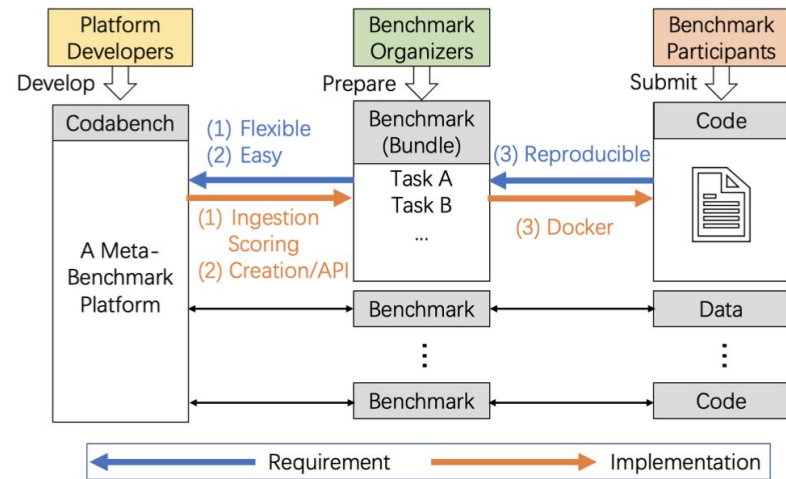
H1 Collaboration ([...] [Mikuni et. al.](#)):

Reanalyzing previous experiment data to unfold many quantities
Utilized Perlmutter for 1000s of bootstrapping and UQ runs

Codabench/“Fair Universe” Platform

Based on <https://www.codabench.org/>

- Codabench - open source platform for AI benchmarks and challenges
 - Originally (CodaLab) Microsoft/Stanford now a Paris-Saclay/[LISN](https://www.lisn.fr/) led community
 - > 500 challenges since 2013
 - Allows code submission as well as results e.g. for evaluation timing or reproducibility
 - Also data-centric AI “inverted competitions”
 - Organizers can define scoring functions
 - Queues for evaluation can run on diverse compute resources
 - Platform itself can be deployed on different compute resources



The screenshot shows the Codabench website interface. At the top, there is a search bar and navigation links for **Benchmarks**, **Resources**, **Queue Management**, and **whiting**. The main content area includes an **Announcement** section with a welcome message and a **Popular Benchmarks** section. The **Featured Benchmarks** section highlights several challenges:

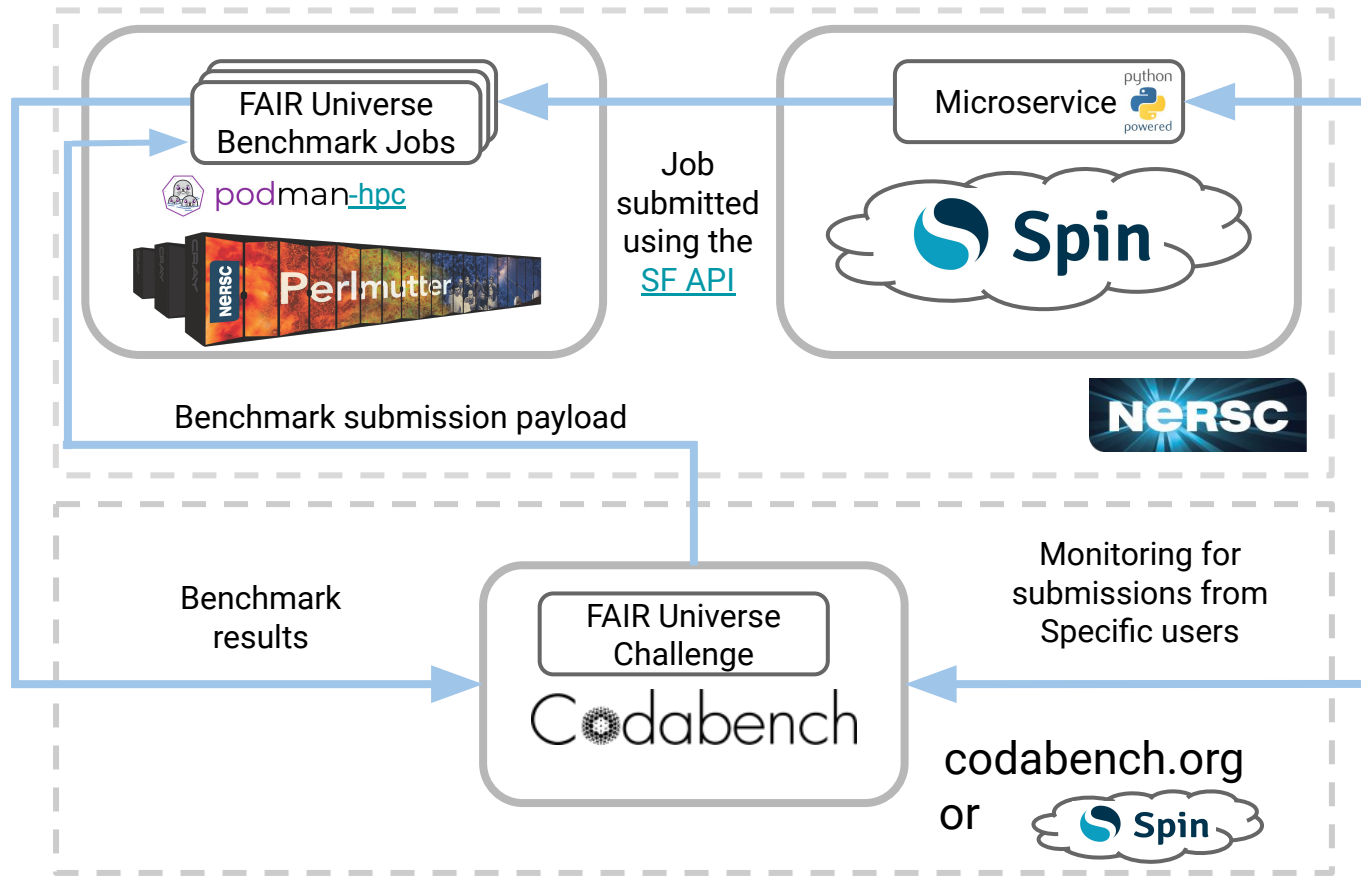
- SenEval-2024, Task 6: Multigenerator, Multidomain, and Multilingual Blank-Box Machine-Generated Text Detection** (November 27, 2023, 311 Participants)
- AutoML Cup Phase 2** (July 15, 2023, 20 Participants)
- WSDM@24 - Conversational Multi-Doc QA Challenge** (November 30, 2023, 292 Participants)
- ROBOVOX SP CUP** (The Robovox challenge field speaker verification) (Organized by: anasim)
- Capture The Flag Mech Interp Challenges** (Find hidden information from Transformer models trained on top tasks) (Organized by: aligocales)
- BioLaySumm 2024 - Shared Task on the Lay** (January 15, 2024, 3 Participants)

At the bottom, a summary of platform statistics is displayed:

- 515** TOTAL COMPETITIONS
- 73** PUBLIC COMPETITIONS
- 3444** USERS
- 3418** COMPETITION PARTICIPANTS
- 26604** SUBMISSIONS

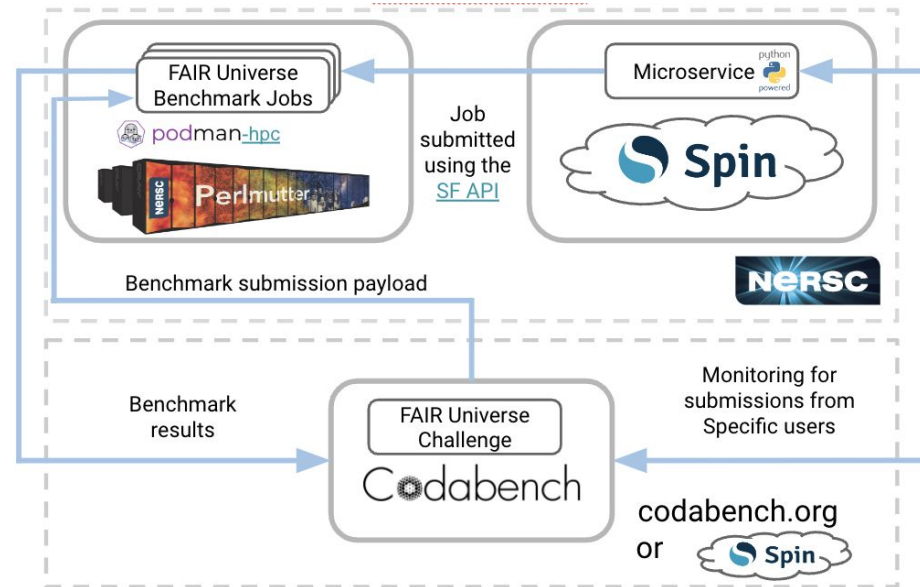
“Fair Universe” brings Codabench to HPC at NERSC!

Fair Universe Platform: Codabench/NERSC integration



Fair Universe Platform: Codabench/NERSC integration

- Benchmark submissions pulled to workers running on Perlmutter:
 - Use [podman\(-hpc\)](#) container runtime: secure and scalable
 - Enable parallelism/scale for
 - **Intensive methods** - use multiple A100 GPUs for training or evaluation
 - **Many participants** - through running many parallel workers
 - **Many evaluations** - e.g for Uncertainty Quantification
- Workers submitted as needed by microservice on SPIN service platform
 - NERSC's "[SF API](#)" for job submission
 - Monitor/filter submissions



- Also deploy instances of Codabench platform itself within SPIN
 - Customization and future OIDC integration with NERSC authorization

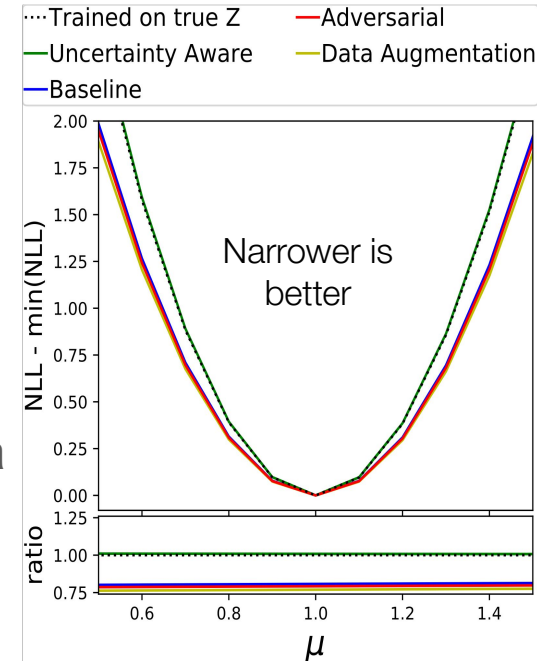
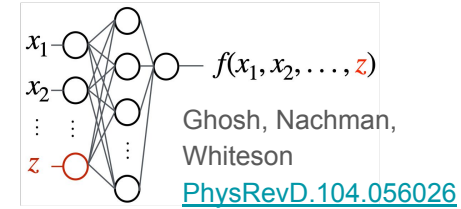
Organize a challenge series focused on measuring and minimizing the effects of systematic uncertainties in HEP

Bias and uncertainty in ML in HEP

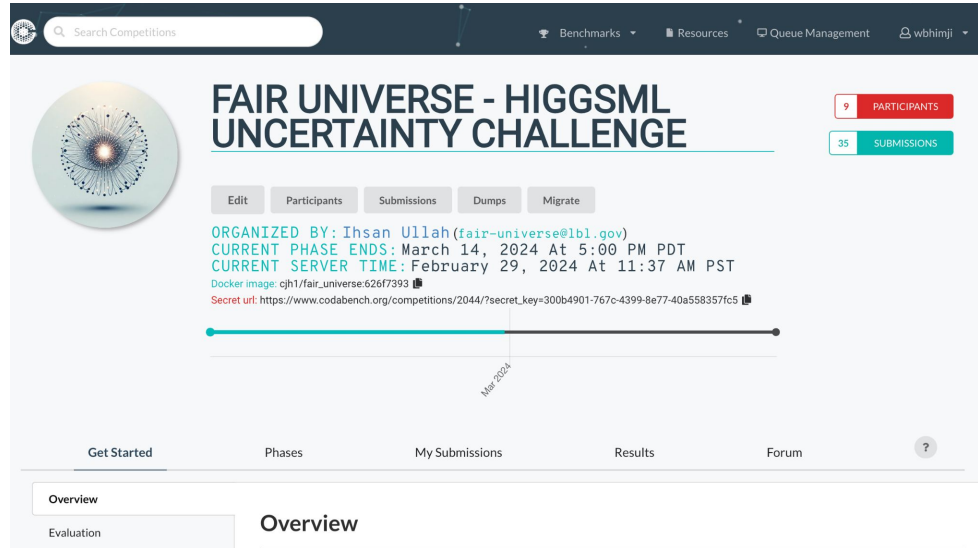
- ML trained on simulation with estimated systematics (“Z”)
 - Then applied in data with different state Z’
 - Commonly dealt with by shifting Z and measuring impact

Non-optimal so many other techniques proposed - e.g.:

- Decorrelation - augmentation; adversarial training
- “Uncertainty-aware” Ghosh, Nachman, Whiteson [PhysRevD.104.056026](https://arxiv.org/abs/1806.04743)
 - Parameterize classifier using Z
- Other novel approaches e.g. (not comprehensive)
 - Inferno: [arxiv:1806.04743](https://arxiv.org/abs/1806.04743);
 - Full profile-likelihood: e.g. [arxiv:2203.13079](https://arxiv.org/abs/2203.13079)
- Mainly demonstrated on single systematic with limited data
 - Can be hard to scale, e.g. retraining and profiling expensive
- **Need for larger datasets and novel metrics to compare uncertainty quantification for ambitious approaches**



Fair Universe: HiggsML Uncertainty Challenge

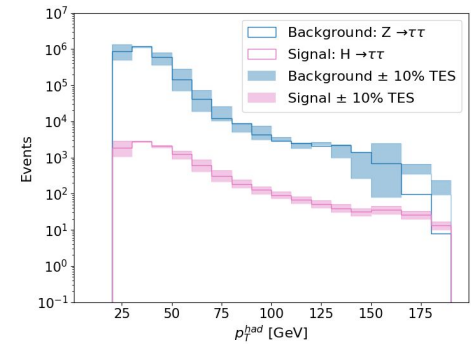
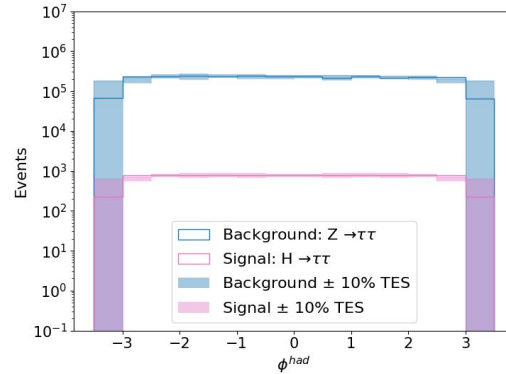
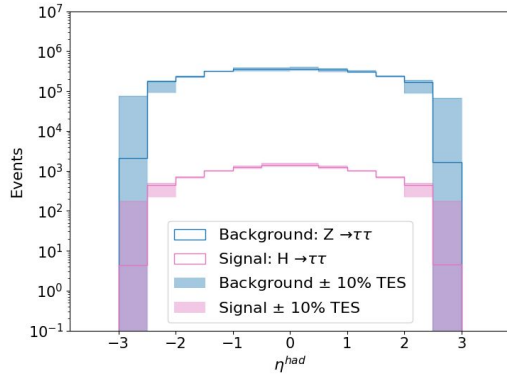


The screenshot shows the competition page for 'FAIR UNIVERSE - HIGGSML UNCERTAINTY CHALLENGE'. The page includes a search bar, navigation links for Benchmarks, Resources, and Queue Management, and a user profile for 'wbhimji'. The challenge title is prominently displayed, along with statistics: 9 PARTICIPANTS and 35 SUBMISSIONS. Below the title are buttons for 'Edit', 'Participants', 'Submissions', 'Dumps', and 'Migrate'. The organizer is listed as 'Ihsan Ullah (fair-universe@lbl.gov)'. Key dates are provided: 'CURRENT PHASE ENDS: March 14, 2024 At 5:00 PM PDT' and 'CURRENT SERVER TIME: February 29, 2024 At 11:37 AM PST'. A Docker image and a secret URL are also shown. A timeline indicates the current phase ends in March 2024. At the bottom, there are navigation tabs for 'Get Started', 'Phases', 'My Submissions', 'Results', and 'Forum', with 'Overview' selected.

- Extension of [HiggsML Kaggle challenge](http://go.lbl.gov/fair-universe-higgsml-spring24) ($H \rightarrow \tau\tau$) from 2014
- HiggsML Uncertainty Challenge Pilot - **launched today** → <http://go.lbl.gov/fair-universe-higgsml-spring24>
 - Evaluate methods and metrics and gain feedback
- Full HiggsML Uncertainty Challenge ~May-Oct 2024
 - Submitted as [NeurIPS competition](#) - results presented at NeurIPS in December

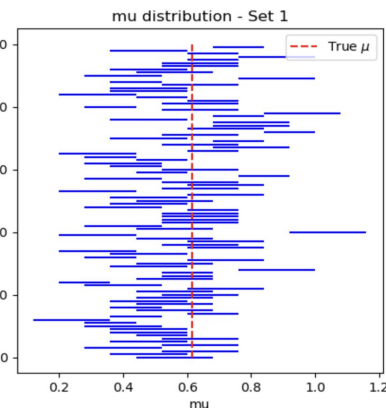
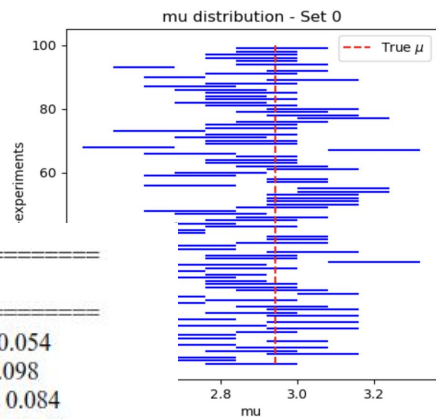
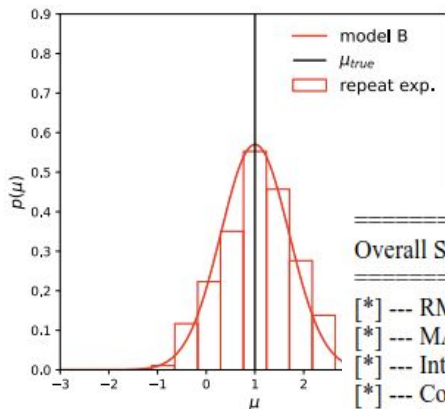
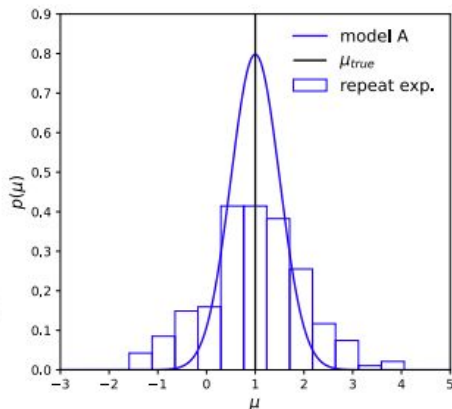
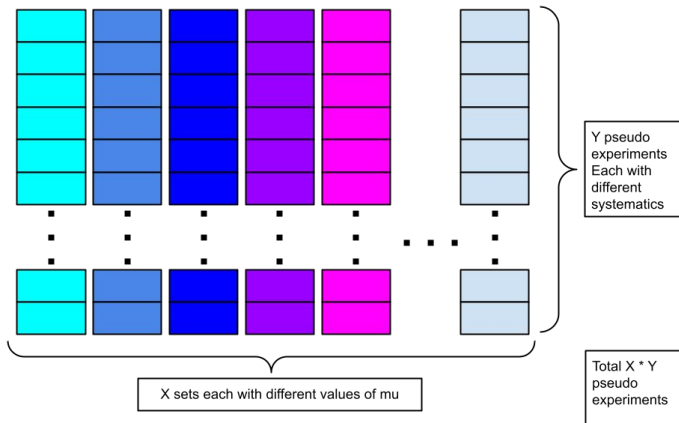
HiggsML Uncertainty Challenge - Datasets

- Much larger datasets than original HiggsML challenge ($12 \cdot 10^6$ training events)
- [Delphes](#) detector simulation (see “Data” tab on competition page for details)
- Apply parameterized systematics (Nuisance Parameters):
 - In current pilot: Tau Energy Scale - shift hadronic Tau Pt (and correlated MET)
 - Further planned for full competition: Jet Energy Scale (and correlated MET impact); Additional randomised Soft MET; Background normalisation; W-boson background normalisation
 - Also ongoing work to add systematics into Delphes generation for comparison



HiggsML Uncertainty Challenge - Evaluation

- Form multiple pseudo-experiment test sets:
 - different signal strengths (μ) and systematics
 - Current pilot - 4μ and 50 pseudo-experiments
- Task: predict uncertainty interval $[\mu_{16}, \mu_{84}]$
 - E.g. 68% quantile of likelihood or assume 1σ
- Score: balance accuracy and precision with matching the true uncertainty and accuracy



Overall Score

- [*] --- RMSE: 0.054
- [*] --- MAE: 0.098
- [*] --- Interval: 0.084
- [*] --- Coverage: 0.46
- [*] --- Quantiles score: -1.651

Uncertainty Quantification Metric

- **Interval width (w)** averaged over N test sets
- **Coverage (c)**: fraction of time μ is contained
- Combined using a **coverage function f(x)**:

$$x \geq 0.68 - 2\sigma_{68} \text{ and } x \leq 0.68 + 2\sigma_{68} : 1.$$

$$x < 0.68 - 2\sigma_{68} : 1 + \left| \frac{x - (0.68 - 2\sigma_{68})}{\sigma_{68}} \right|^4$$

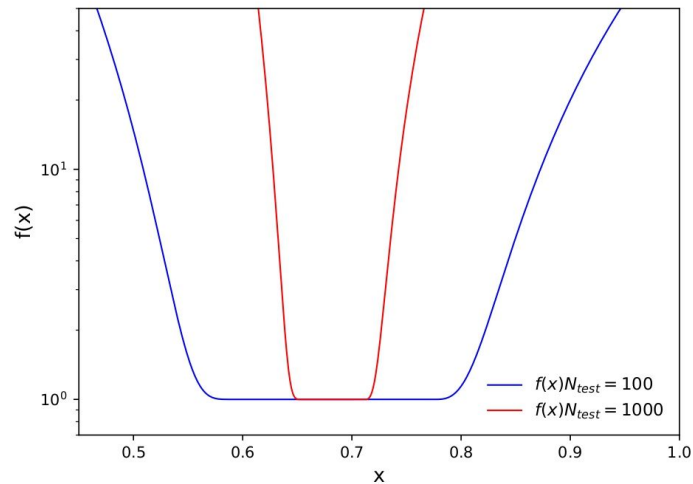
$$x > 0.68 + 2\sigma_{68} : 1 + \left| \frac{x - (0.68 + 2\sigma_{68})}{\sigma_{68}} \right|^3$$

$$\text{with } \sigma_{68} = \frac{\sqrt{(1-0.68)0.68N}}{N}$$

- N dependence for equivalent ideal coverage
- Penalizes undercoverage more
- **Final score (s)** designed to avoid large values or gaming

$$w = \frac{1}{N} \sum_{i=0}^N |\mu_{84,i} - \mu_{16,i}|.$$

$$c = \frac{1}{N} \sum_{i=0}^N \mathbf{1} \text{ if } (\mu_{true,i} \in [\mu_{84,i} - \mu_{16,i}])$$

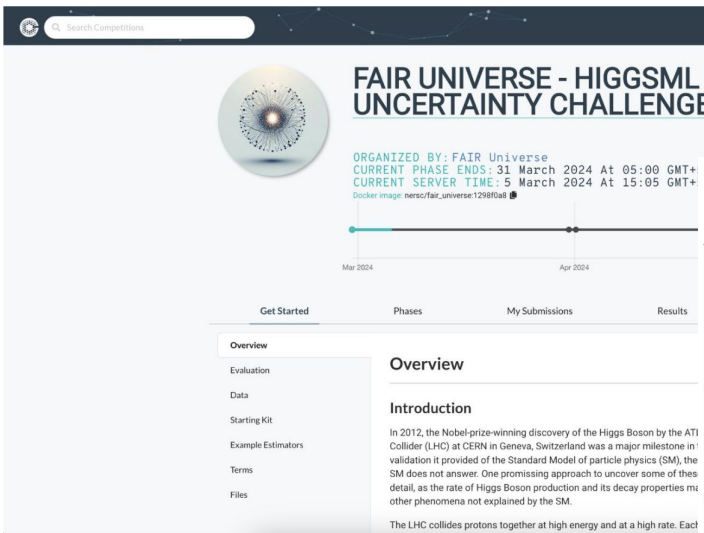


$$s = -\ln((w + \epsilon)f(c))$$

Enter the HiggsML Uncertainty Challenge Pilot!

<http://go.lbl.gov/fair-universe-higgsml-spring24>

- See more detailed: [walkthrough slides](#); and “starting kit”



FAIR UNIVERSE - HIGGSML UNCERTAINTY CHALLENGE

ORGANIZED BY: FAIR Universe
CURRENT PHASE ENDS: 31 March 2024 At 05:00 GMT+5
CURRENT SERVER TIME: 5 March 2024 At 15:05 GMT+5
Docker image: nerscfair_universe:1298f0a8

0 PARTICIPANTS
39 SUBMISSIONS

Get Started Phases My Submissions Results

Overview
Evaluation
Data
Starting Kit
Example Estimators
Terms
Files

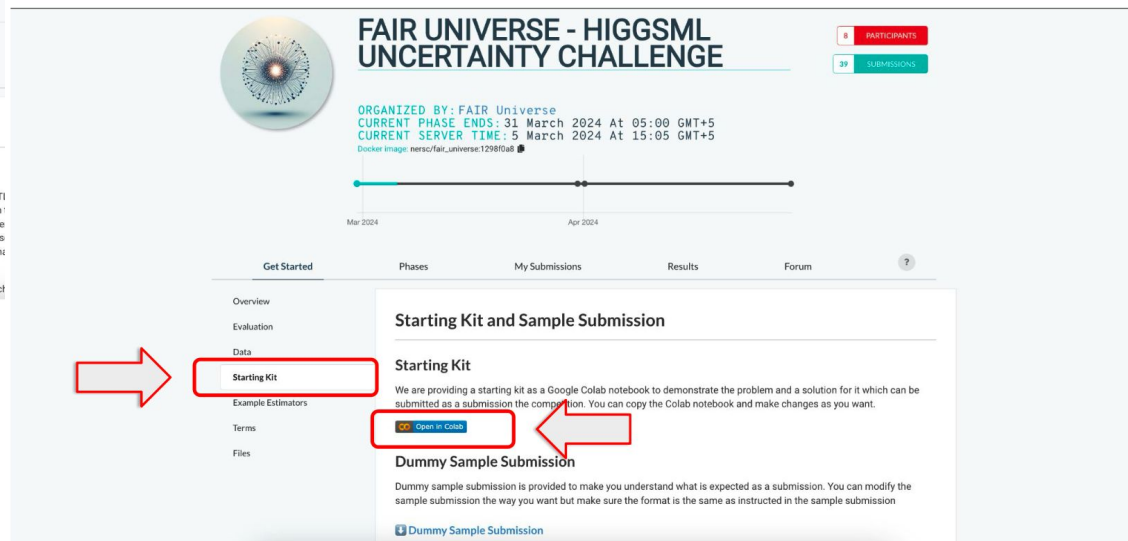
Overview

Introduction

In 2012, the Nobel-prize-winning discovery of the Higgs Boson by the ATLAS and CMS experiments at CERN in Geneva, Switzerland was a major milestone in the validation it provided of the Standard Model of particle physics (SM), the SM does not answer. One promising approach to uncover some of these details, as the rate of Higgs Boson production and its decay properties in other phenomena not explained by the SM.

The LHC collides protons together at high energy and at a high rate. Each

6. Check out the starting kit



FAIR UNIVERSE - HIGGSML UNCERTAINTY CHALLENGE

ORGANIZED BY: FAIR Universe
CURRENT PHASE ENDS: 31 March 2024 At 05:00 GMT+5
CURRENT SERVER TIME: 5 March 2024 At 15:05 GMT+5
Docker image: nerscfair_universe:1298f0a8

0 PARTICIPANTS
39 SUBMISSIONS

Get Started Phases My Submissions Results Forum

Overview
Evaluation
Data
Starting Kit
Example Estimators
Terms
Files

Starting Kit and Sample Submission

Starting Kit

We are providing a starting kit as a Google Colab notebook to demonstrate the problem and a solution for it which can be submitted as a submission the competition. You can copy the Colab notebook and make changes as you want.

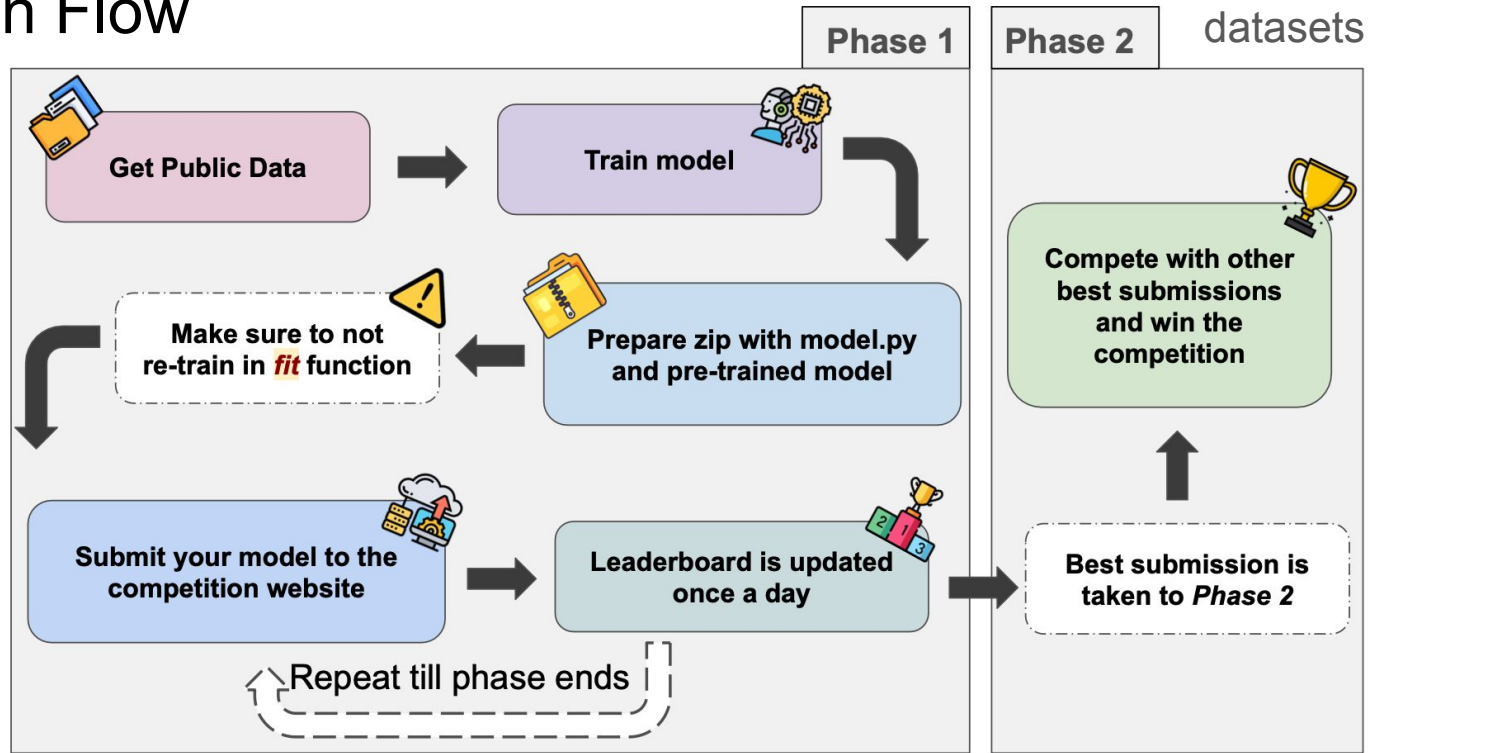
[Open in Colab](#)

Dummy Sample Submission

Dummy sample submission is provided to make you understand what is expected as a submission. You can modify the sample submission the way you want but make sure the format is the same as instructed in the sample submission

[Dummy Sample Submission](#)

Competition Flow



- Submissions are run, and the leaderboard updated, at least once a day
- Participants code is run on the platform, in parallel on many test sets, so there can be differences to when tested locally - contact us on [slack](#) with any issues

Conclusions

- We have built a flexible platform for hosting challenges and benchmarks - extending Codabench, backed by HPC at NERSC
- Launching a series of challenges for uncertainty aware methods for HEP
 - You can enter the HiggsML Uncertainty Challenge Pilot now!
 - <http://go.lbl.gov/fair-universe-higgsml-spring24>
- Open to feedback on metrics/datasets to encourage advanced approaches and ensure challenge is interesting

Help and feedback: [#higgsml-uncertainty-challenge-spring-24](#) channel on the [Fair Universe Slack workspace](#)

Ongoing information Google Group: [Fair-Universe-Announcements](#)

Collaborations, questions, comments: fair-universe@lbl.gov

General collaboration on NERSC-AI: wbhimji@lbl.gov