IOP Institute of Physics

2026

Open Data for Machine Learning

Key Challenges

Overview of four core challenges in using open physics data for machine learning education.

Shaadil Shah Mandarry - Doctoral Researcher on ATLAS Experiment

14 May 2026

UNIVERSITY OF SUSSEX
BRIGHTON

- Biggest Barriers?
- Simplify Datasets?
- Best Data Formats?
- Accessibility Focus

Challenge 1: Open vs Usable Data

Key Challenge

How do we make open physics data genuinely accessible for machine learning education?

UNIVERSITY OF SUSSEX BRIGHTON

IOP Workshop

Challenge 2: Sustainable ML Workflows & Infrastructure

Key Challenge

How do we build sustainable and reproducible ML environments for open physics data education?

Sustainable ML education requires reproducible software environments, accessible computing resources, and long-term support for tutorials, datasets, and benchmark workflows.

Reproducible ML Environments

How do we maintain ML tutorials as software frameworks, dependencies, and ML libraries rapidly evolve?

Computing & Accessibility

Should ML educational workflows rely on local computing, cloud platforms, or shared GPU infrastructure?

UNIVERSITY OF SUSSEX BRIGHTON

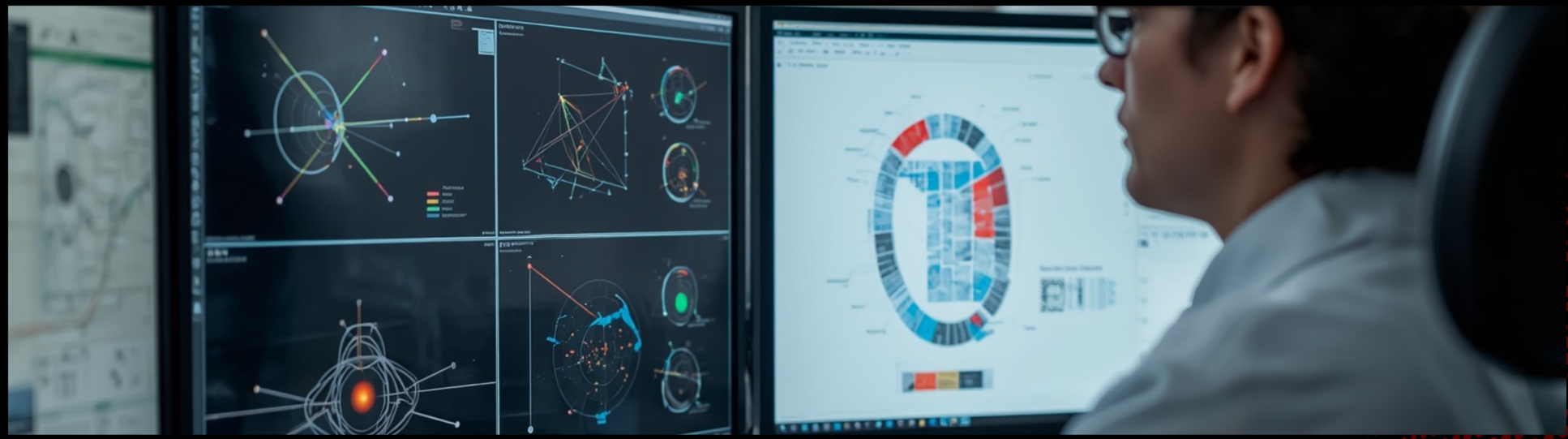
IOP Workshop

Challenge 3: ML-Ready Data vs Experimental Complexity

Key Challenge

How much experimental complexity should be exposed in ML educational datasets?

How do we balance ML accessibility with realistic detector effects, uncertainties, and reconstruction challenges?



UNIVERSITY OF SUSSEX BRIGHTON

IOP Workshop

Challenge 4: Future Skills & Education

Key Challenge

What machine learning and data-analysis skills should students gain from working with open physics datasets?

ML in Curricula?

Should ML become a standard component in physics education curricula?

ML Learning Tasks

Which ML tasks best translate open physics data into meaningful educational experiences?

Career Preparation

How can ML workflows using open data prepare students for both research and industry?

Essential ML Skills

Which practical ML and data-analysis skills should students develop through open-data workflows?

UNIVERSITY OF SUSSEX BRIGHTON

IOP Workshop

Group Discussion Activity

Open vs Usable Data

What are the biggest barriers students face today?

Should experiments release simplified ML-ready datasets?

Which data formats are most suitable for education?

Sustainable ML Workflows and Infrastructure

Should experiments or universities maintain ML benchmark workflows?

How do we keep ML tutorials reproducible as frameworks evolve?

Should open educational data be treated as long-term scientific infrastructure?

ML-Ready Data vs Experimental Complexity

Should students begin with simplified ML-ready datasets or realistic detector-level workflows?

How do we balance ML accessibility with realistic experimental complexity?

Can simplified datasets still preserve meaningful physics insight?

Future Skills and Education

Should ML become a standard part of physics curricula?

Which ML tasks best support learning using open physics data?

How can ML workflows using open data prepare students for research and industry?