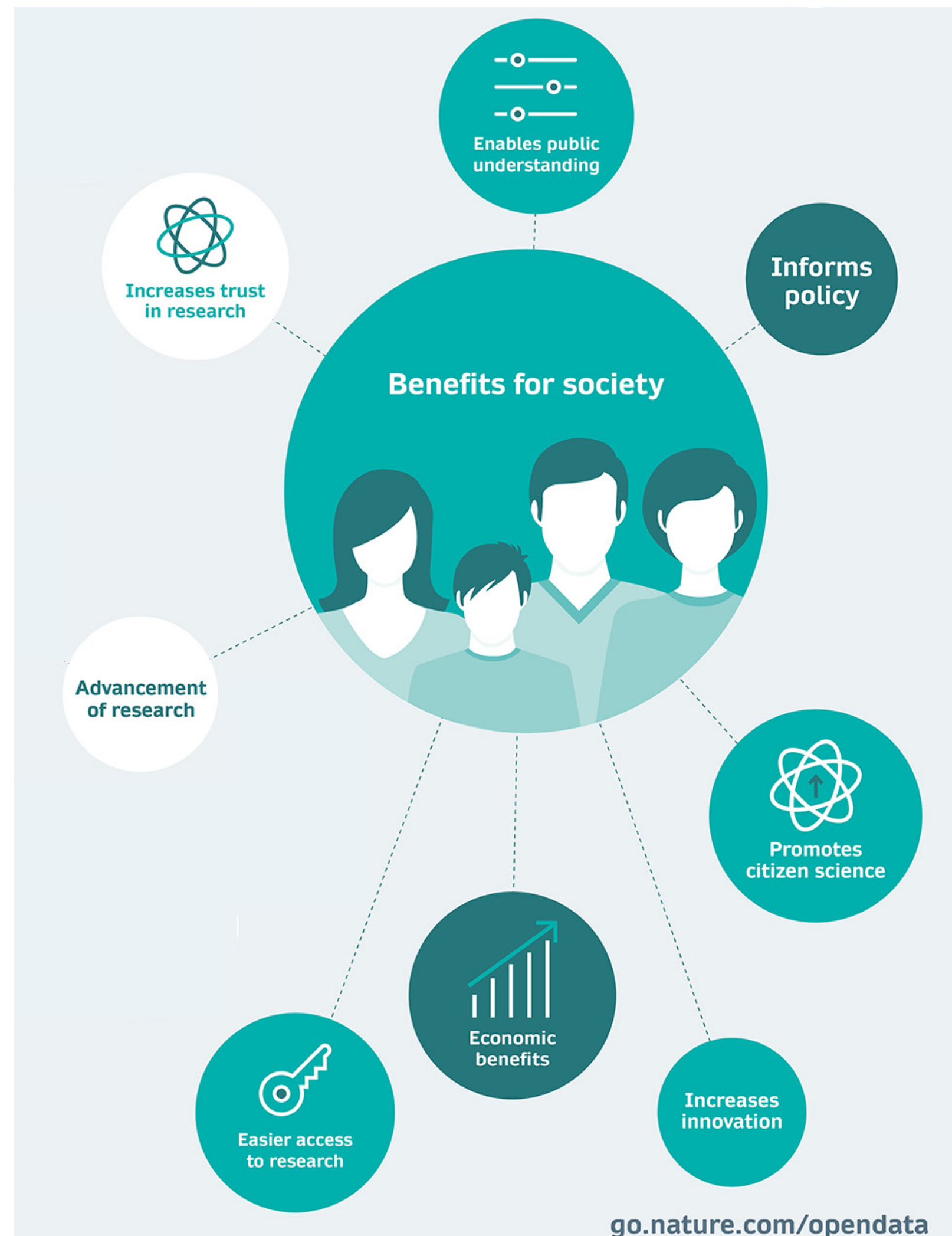


Meirin Oan Evans, on behalf of the ATLAS Education & Outreach group

# Teaching Machine Learning with ATLAS Open Data


## Open Data: a crucial part of science



- ▶ Enables public understanding
- ▶ Promotes citizen science
- ▶ Economic growth
- ▶ Easier access to research
- ▶ Increases trust in scientists



# Open Data: a crucial part of education

- ▶ Data are an essential part of creating science knowledge, and so they should be when students acquire knowledge
  - ▶ Make learning realistic
  - ▶ Make learning active
  - ▶ Make learning hands-on
  - ▶ Make research transparent
  - ▶ Make research trustworthy
- 
- An illustration showing the silhouettes of five students looking at a large digital screen. The screen displays a network of circular icons representing various scientific fields: a clock, a person, a tablet, a wrench, a telescope, a globe, a laptop, a magnifying glass, a flask, a microscope, a pencil, an atom, and a book. The icons are interconnected by lines, symbolizing the interconnected nature of scientific knowledge.





# Hi from ATLAS 🖐️



- ▶ ~5500 members
- ▶ Physics, Detector, Upgrade, Trigger, Computing, Data prep, Education, Outreach...



## ATLAS Open Data

- ▶ 8 & 13 TeV research data for science education worldwide
- ▶ Along with tools, software, documentation for usability
- ▶ Designed in collaboration with students and teachers

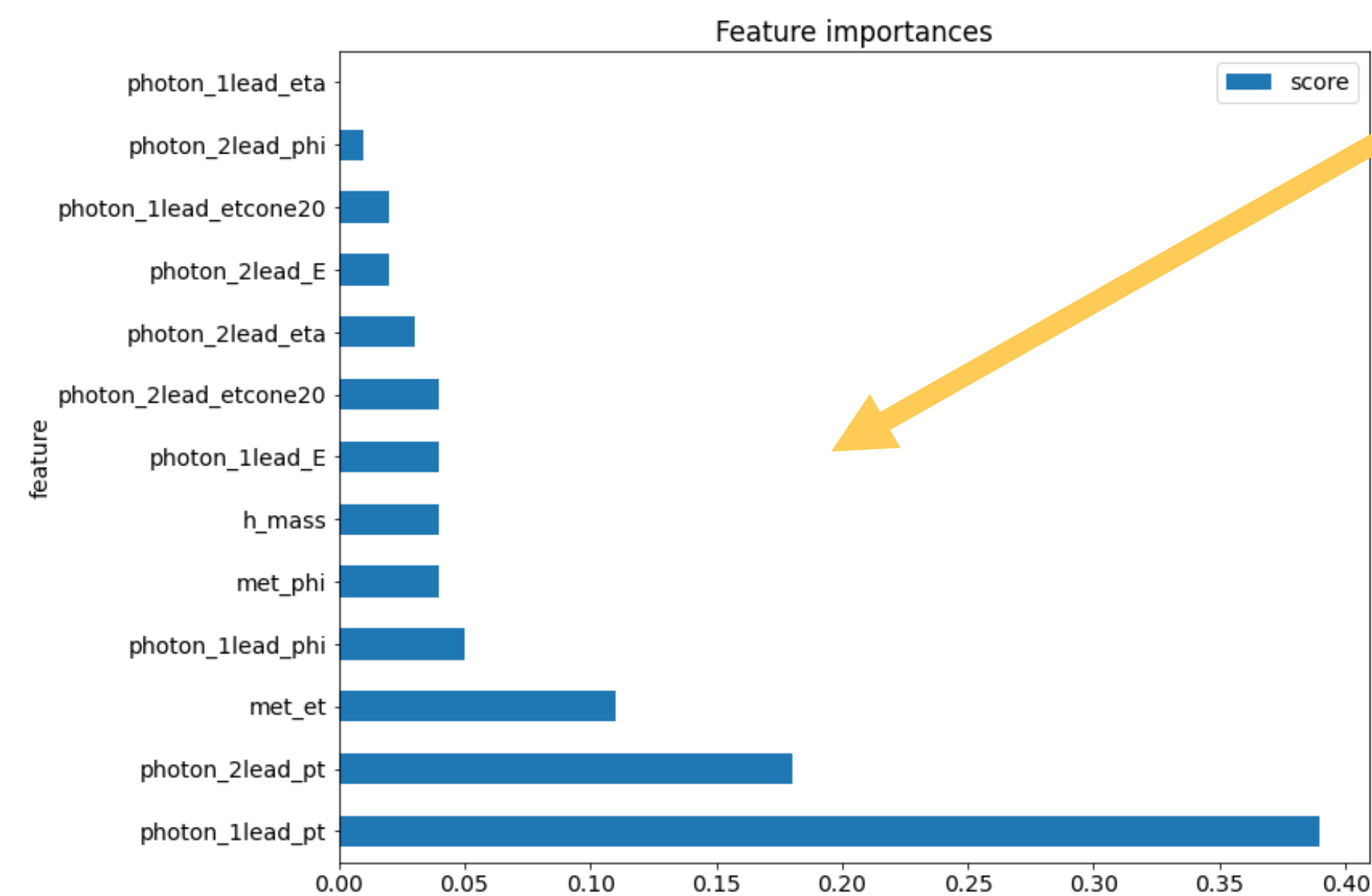




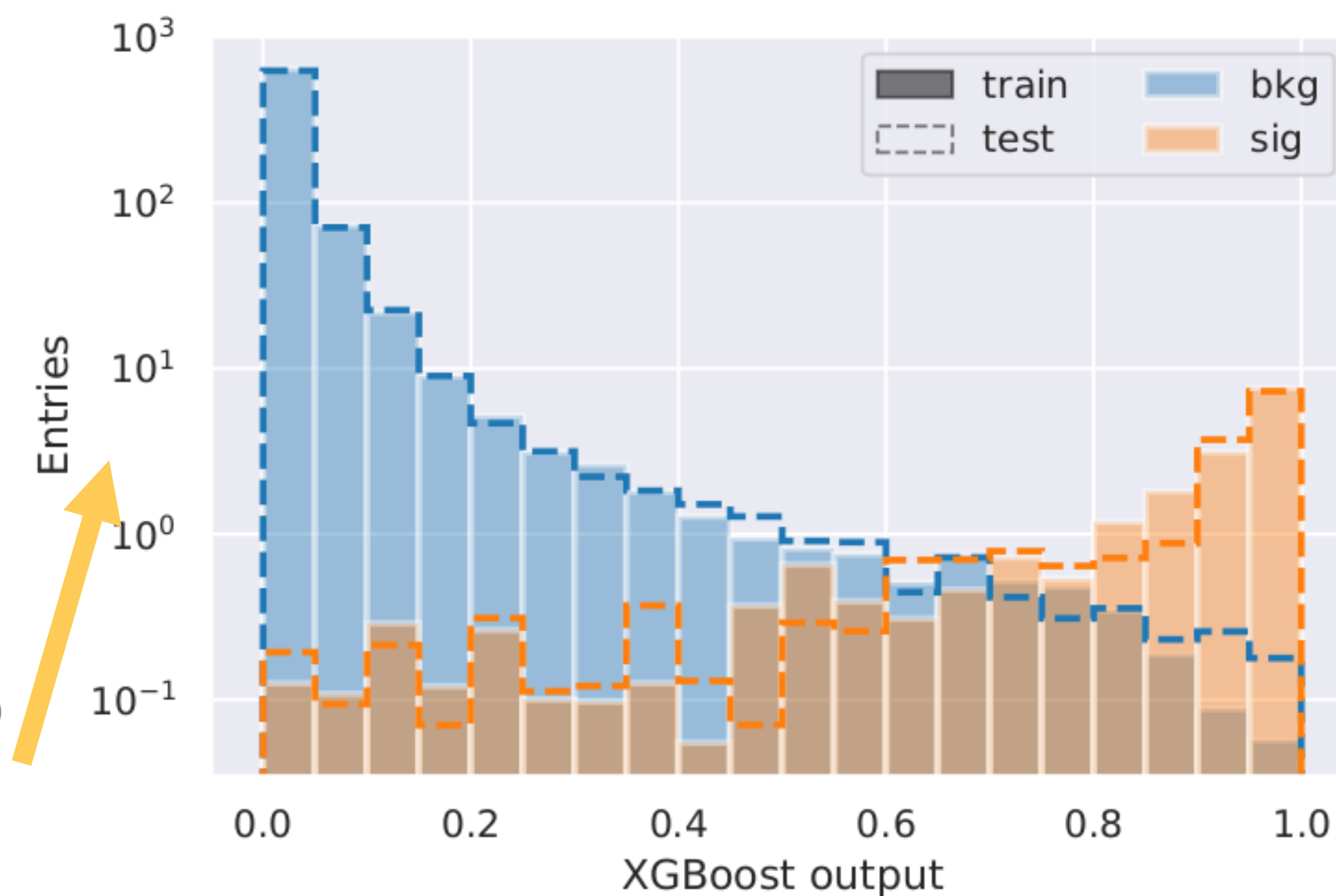


## Master level projects

- [Research-Based Particle Physics course @ University of Oslo](#)



ML algorithms to distinguish Higgs production mechanisms



Multivariate classification to separate EW SUSY from SM

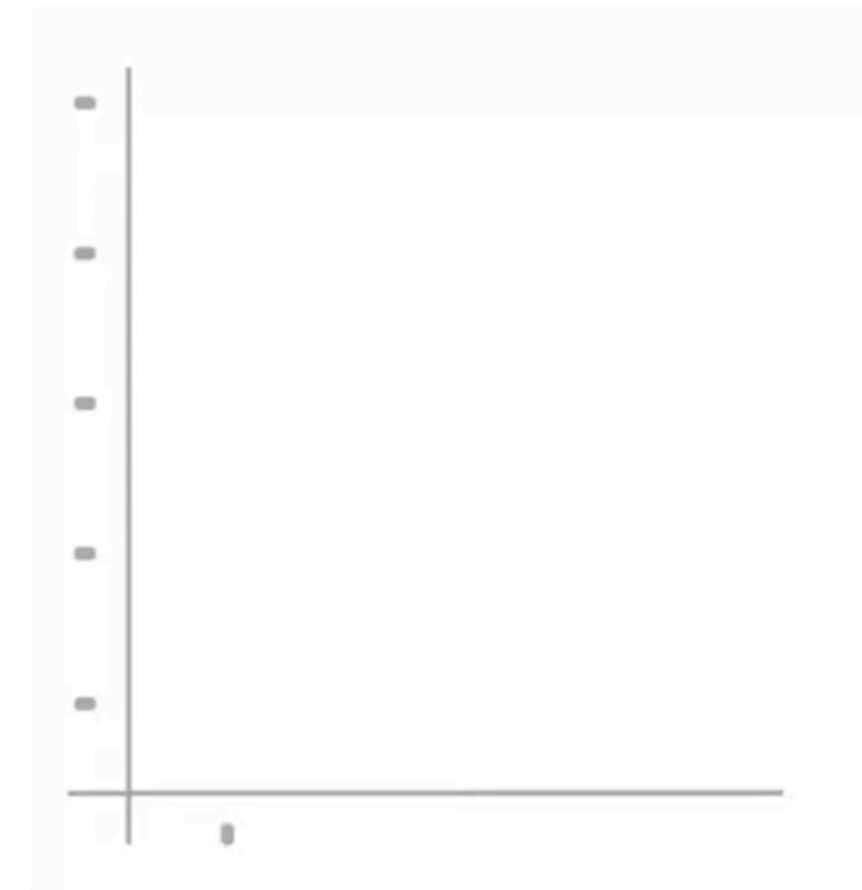
### Example notebooks

- ▶ ML algorithm only just runs out of the box
- ▶ Up to the student to tweak and improve

The screenshot shows the GitHub interface for the repository `atlas-outreach-data-tools / notebooks-collection-opendata`. The repository has 3 stars and 6 forks. The `Code` tab is selected, showing the file structure. The current view is for the `master` branch, specifically the path `notebooks-collection-opendata / 13-TeV-examples / uproot_python /`. A commit by `meevans1` is highlighted, titled "Change master branch uproot README to point to no-root binder", with commit hash `ff8bc6c` and dated "on 9 Jul". Below the commit, two files are listed: `HZZ_BDT.ipynb` and `HZZ_NN.ipynb`, both with the description "add links within uproot notebooks" and dated "3 months ago".



## Video tutorial



### Machine Learning with Neural Networks

ATLAS 13 TeV Open Data Tutorial

*Meirin Oan Evans, University of Sussex*

- ▶ To be published on [ATLAS YouTube channel](#)
- ▶ Used in 1<sup>st</sup> ever online ATLAS Software Tutorial

# ATLAS US/CA computing bootcamp

## Schedule

	Setup	Download files required for the lesson
00:00	1. <a href="#">Introduction</a>	What is machine learning? What role does machine learning have in particle physics? What should I do if I want to get good at machine learning?
00:00	2. <a href="#">Mathematical Foundations</a>	What is the common terminology in machine learning?
00:00	3. <a href="#">Resources</a>	Where should I go if I want to get better at python? What are the machine learning libraries in python? Where should I go if I want to get better at machine learning?
00:10	4. <a href="#">Data Discussion and Preprocessing</a>	What dataset is being used How must we organize our data such that it can be used in the machine learning libraries?
00:25	5. <a href="#">Model Training</a>	How does one train machine learning models in python?
00:45	6. <a href="#">Model Comparison</a>	How do you use the sci-kit learn and tensorflow packages for machine learning?
01:05	7. <a href="#">Neural Networks</a>	What is a neural network? How can I visualize a neural network?
01:15	Finish	

► [Lesson webpage](#)

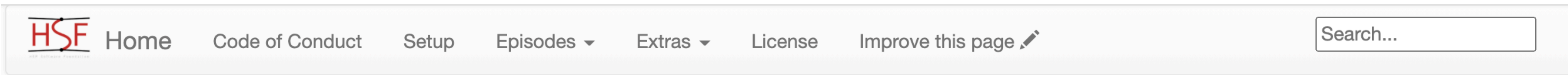
► Developed by [Luke Polson](#)

► ~1hr:15min  
tutorial for new  
PhD students

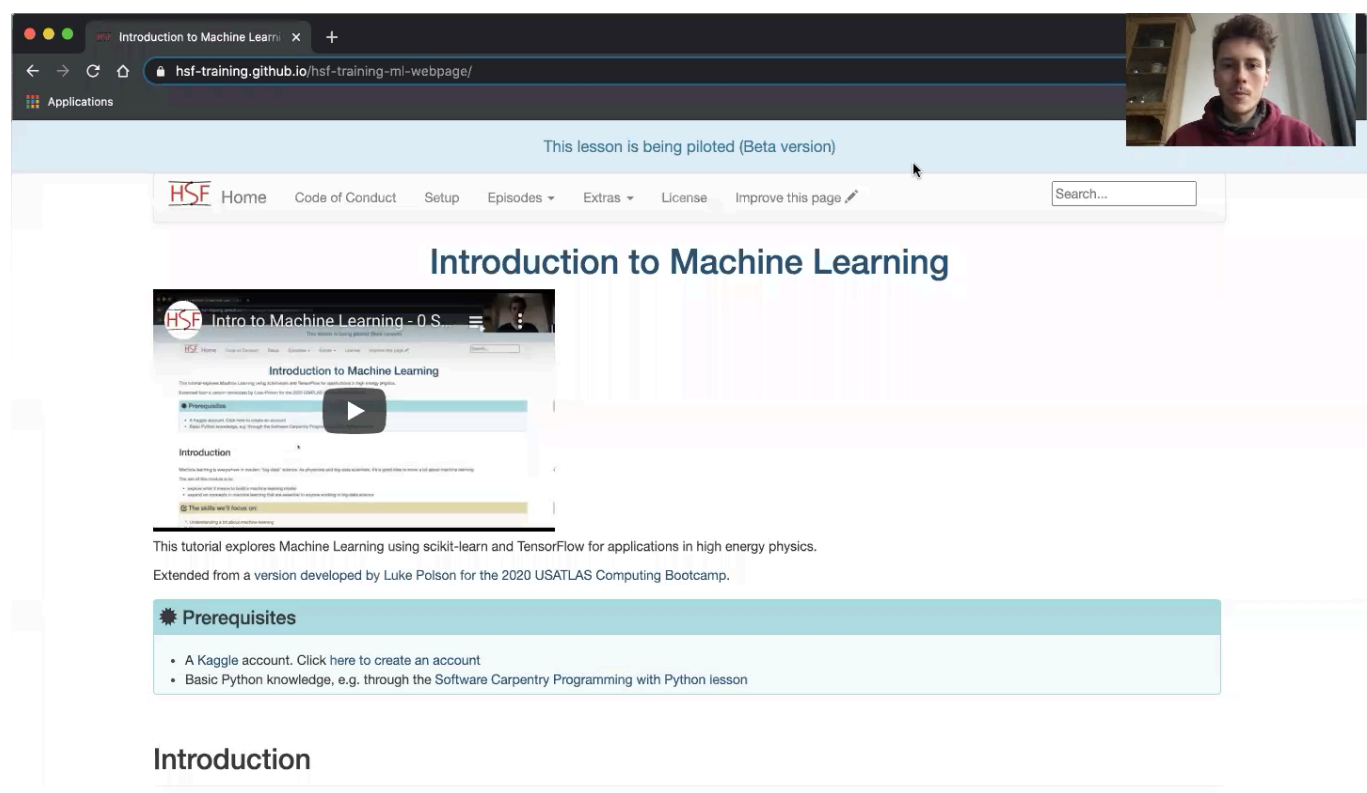
► In [Software Carpentry](#) style

► A success!

# HEP Software Foundation training




## Introduction to Machine Learning



This tutorial explores Machine Learning using scikit-learn and TensorFlow for applications in high energy physics.

Extended from a version developed by Luke Polson for the 2020 USATLAS Computing Bootcamp.

 **Prerequisites**

- A Kaggle account. [Click here to create an account](#)
- Basic Python knowledge, e.g. through the [Software Carpentry Programming with Python lesson](#)



# HEP ML tutorial

- ▶ Convert files to [hdf5](#) using [uproot](#)
- ▶ Notebooks teach [XGBoost](#), [sklearn](#), [LightGBM](#)

dhrou / HEPMLtutorials

Watch

1

Star

0

Fork

0

<> Code

Issues

Pull requests

Actions

Projects

Wiki

Security

Insights

master

1 branch

0 tags

Go to file

Add file

Code

dhrou first version

a5a1ba3 on 9 Jun 13 commits

HEPML_HandsOn_BDT.ipynb	add weight discussion	5 months ago
HEPML_HandsOn_NN.ipynb	clean version	5 months ago
PreProcessingMerge.ipynb	major new version	5 months ago
PreProcessingUR.ipynb	major new version	5 months ago
README.md	first commit	11 months ago
ST4PNT_ML_EI.ipynb	small fix	4 months ago
higgsml_syst.py	first version	4 months ago

About

ML introductory tutorials for the (high energy) physics oriented minds

Readme

Releases

No releases published

Packages

No packages published

In progress

## Future challenges

- ▶ How to ensure our tools and resources are **accessible** without guidance from physicists?
- ▶ How to incorporate our tools and resources into more university (and maybe pre-university?!) teaching?
- ▶ Can we spread our tools and resources into **wider use**? e.g. on platforms that teach machine learning, Kaggle...
- ▶ How can we teach more than just particle physics – skills in computing, analysis, data science, machine learning...?

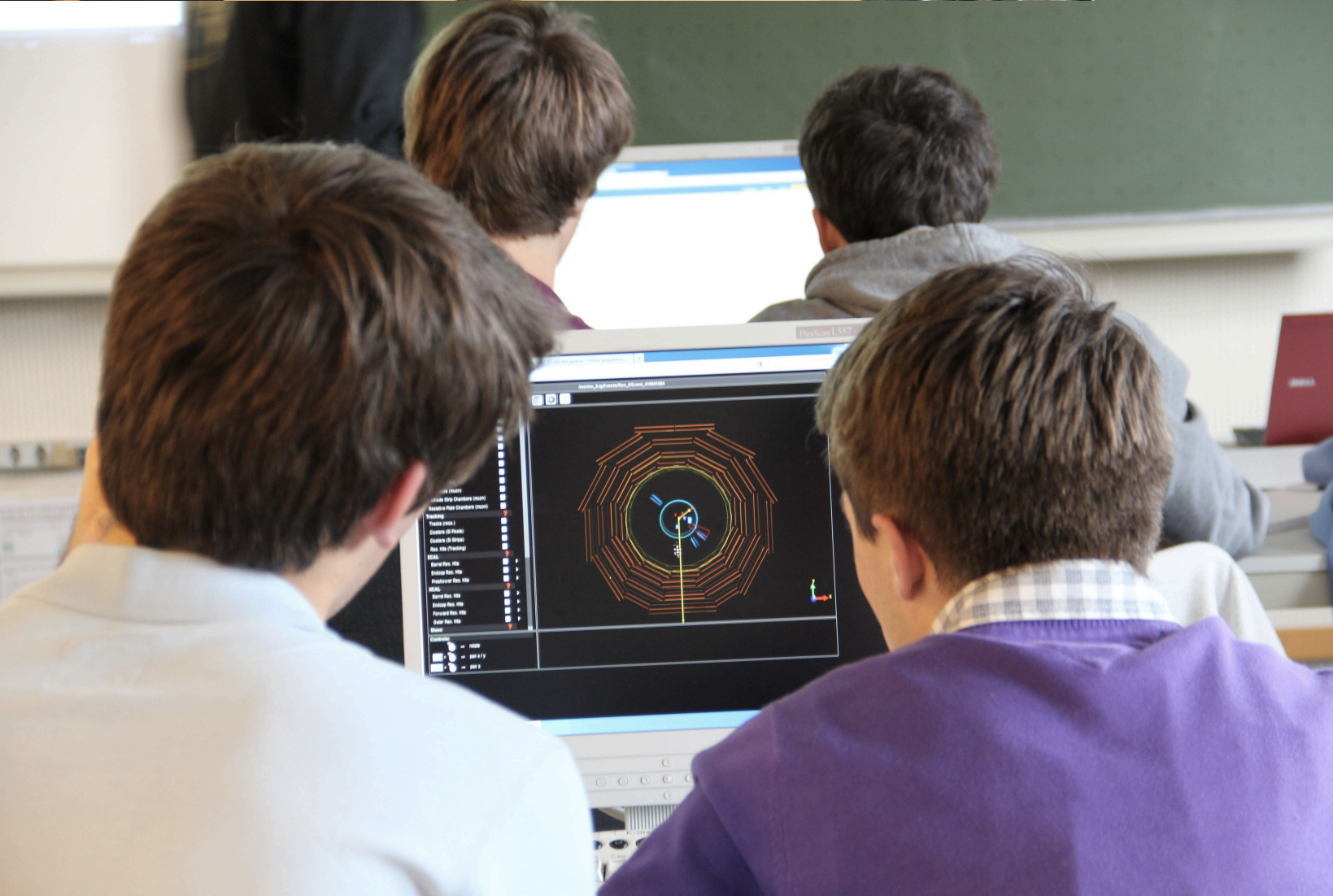
### For the future

- ▶ Lots of exciting ML teaching using ATLAS Open Data:
  - [Example notebooks](#)
  - ~30 min video tutorial
  - [~1hr:15min online tutorial](#)
  - [~3hr:30min online tutorial](#)
- ▶ Encourage friends/enemies to use/develop these resources
- ▶ Use these resources to teach ML yourself!





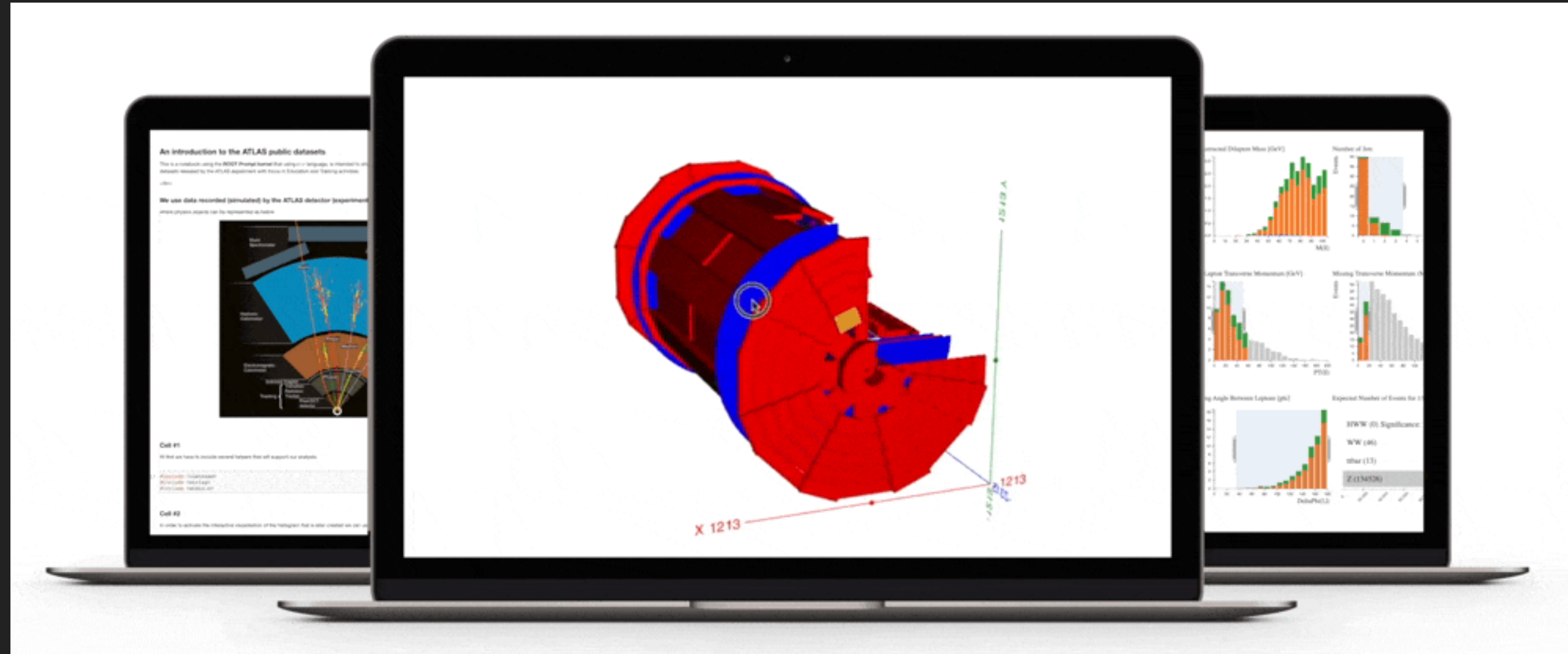
Thanks!



\*to everybody teaching machine learning

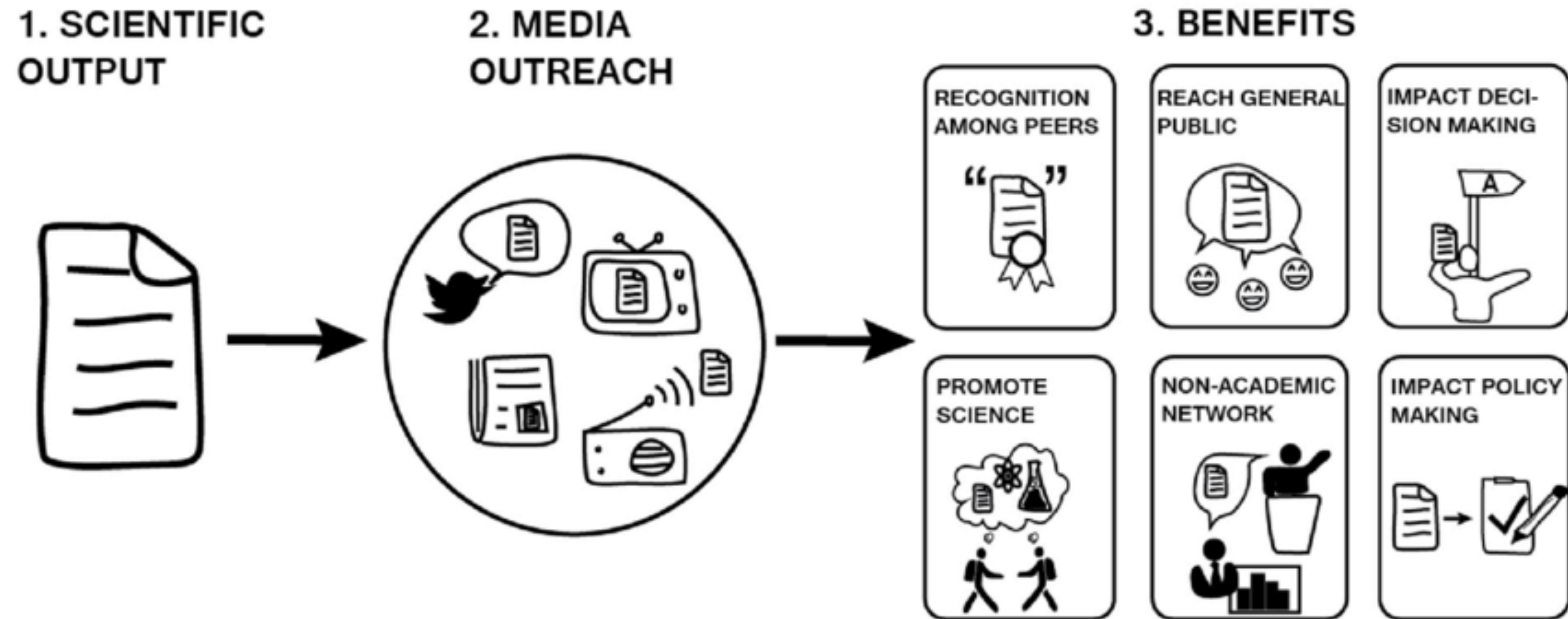
\*\*to all students exploring machine learning



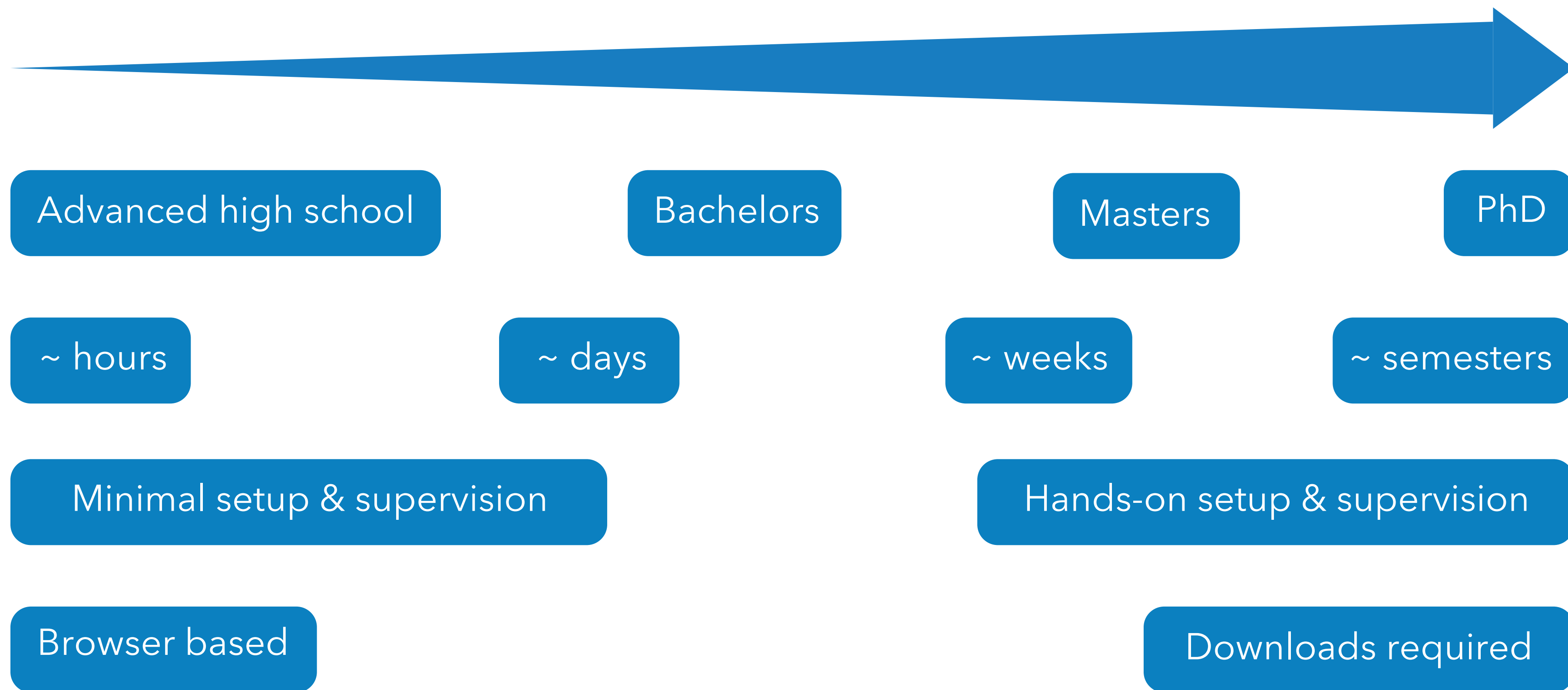


Backup

## Open Data: a crucial part of research



## Reaching a broad audience





## Data Intensive, AI, ML summer school

- ▶ [Optimise Signal/Background Ratio Using Stats Techniques](#)
- ▶ GitHub ["How to rediscover the Higgs!"](#)

The screenshot shows the GitHub interface for the repository 'meevans1 / How-to-rediscover-the-Higgs'. The repository has 0 watches, 11 stars, and 3 forks. The 'Code' tab is selected, showing a file tree for the 'notebooks' directory. A commit by 'meevans1' is highlighted, showing updates to several Jupyter Notebook files.

File	Commit Message	Time
..		
BDT.ipynb	Change tuple path inside BDT notebook	15 months ago
NN-keras.ipynb	Updated input file path in ML notebooks	14 months ago
NN.ipynb	Updated input file path in ML notebooks	14 months ago
Random-Forest.ipynb	Updated input file path in ML notebooks	14 months ago
SVM.ipynb	Updated input file path in ML notebooks	14 months ago