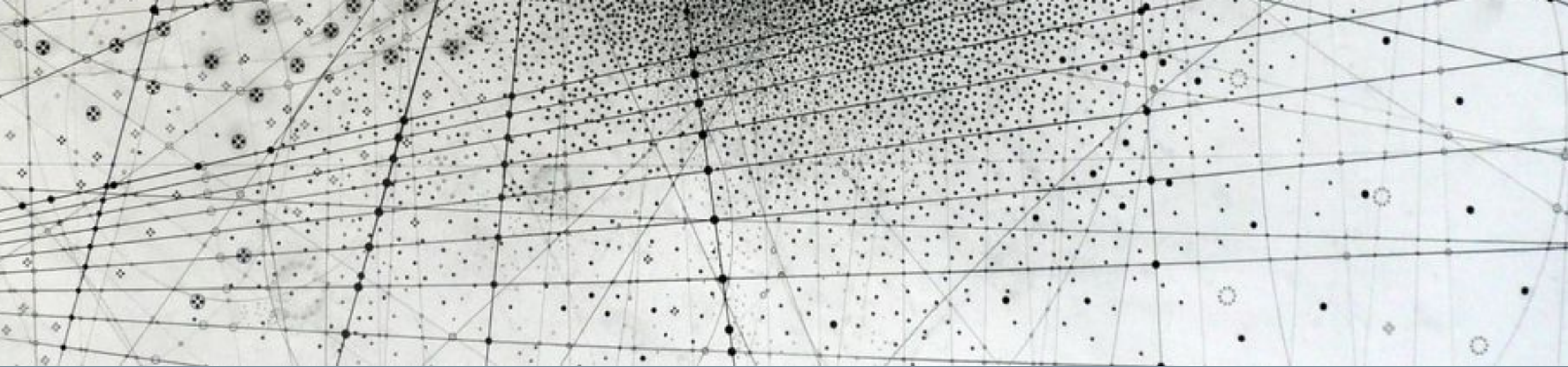


Research Methods in Machine Learning & Physics

Prof. Mariel Pettee

ATLAS EXPERIMENT Polymathic WISCONSIN UNIVERSITY OF WISCONSIN–MADISON

Course inspiration

- This is the course I wish I could have had as a graduate student!
- My graduate coursework was a solid theoretical foundation essential to being in a community of physicists, but I got very little practical training that I could apply in my area of research.
- Physicists have long been entwined in the field of Machine Learning, and we bring with us specialized datasets, research challenges, and statistical requirements that a general course in ML wouldn’t cover
- **Not** a survey course – instead, a sandbox to develop your research toolkit

Course overview

- Graduate-level seminar
- **Hands-on, project-based practicum**, i.e. a series of realistic simulations of how to apply machine learning methods for research-grade physics data in the context of a collaboration
- In-class sessions are largely dedicated to collaborative problem-solving and active discussion
- Emphasizes collaboration & scientific communication skills (written & oral) alongside the computational skills that you'll hone
- Reflects my strong bias toward **active learning** as a teaching strategy

Course logistics

Class takes place in-person on **Mondays & Wednesdays, 4:00pm - 5:15pm.**

We will meet in **Sterling 1333.**

Canvas site: https://canvas.wisc.edu/courses/480728

A portion of each class session will be devoted to an interactive presentation/lecture, and the remainder of each class will feature in-person collaborative work on problem sets and projects.

(I intend to put all my lecture slides on Canvas & Github. This is my first time working with Canvas as an instructor – please bear with me. If something looks wrong, let me know!)

Prerequisites

The most important prerequisites are a familiarity with Python and a solid foundation in modern physics principles.

I will assume you are at least familiar with some, but not all, of the following software tools/languages: Git, Jupyter, LaTeX, Numpy, Matplotlib, Pandas, etc.

Prior experience training machine learning models is not expected or required, but any familiarity with some ideas from machine learning will of course be helpful.

Learning objectives

By the end of this course, you should feel very comfortable:

- Using contemporary data science approaches to solve open-ended research problems in physics. These include:

    - Sketching out new research ideas with **Python** & **Jupyter** notebooks
    - Incorporating version control into your daily workflow with **Git**, and presenting well-organized and documented code
    - Using data science libraries like **numpy, scipy, pandas, & matplotlib**
    - Writing scientific papers in **LaTeX** using journal templates
    - Training a couple of common types of neural networks using **Pytorch** and tracking runs using tools like **Wandb**

Learning objectives

By the end of this course, you should feel very comfortable:

- Presenting your work in formal and informal settings:
  - **Writing:** blog posts & LaTeX-formatted scientific papers
  - **Oral:** in-person collaboration, formal talks with slides, impromptu Q&A

- Sharing feedback with peers:
  - Asking questions during talks
  - Providing useful comments during peer review
  - Doing code review

Community guidelines

- In this course, I expect you to treat your fellow students as research colleagues.

- Classroom time should feel friendly, respectful, fun, and lively. Your peers will really benefit from your attendance and participation.

- If you need to miss a class, arrive late, or leave early, I'd appreciate it if you could let me know in advance so I can make any adjustments to group assignments.

- You may miss up to 2 class sessions, no questions asked, without affecting your participation grade.

Topics covered

The projects (and their accompanying problem sets) will engage directly with real data or highly realistic simulations from a selection of major contemporary experiments. The first 3 projects will involve data analysis using ML in the areas of **high-energy particle physics, gravitational waves, and astrophysics**.

The final project can engage with a research question of your choice, or I will have a few suggested project ideas you can choose from. If you would like to choose your own research topic for the final project, please come meet with me during my office hours to discuss the potential scope.

Note that this class will *not* be a comprehensive overview of all major contemporary machine learning methods – instead, **we will cover a smaller selection of methods at a deeper level**.

Topics covered

ML-related topics may include:

- Deep learning & neural networks
- Training dynamics
- Parameter inference
- Anomaly detection
- Graph Neural Networks
- Convolutional Neural Networks
- Likelihood-free inference
- Unfolding / deconvolution
- Transformers

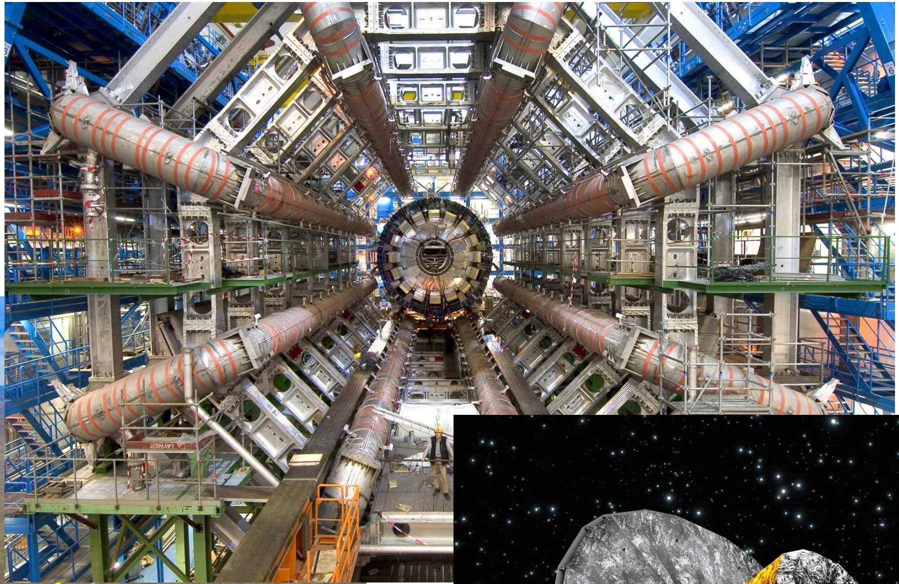
Course materials

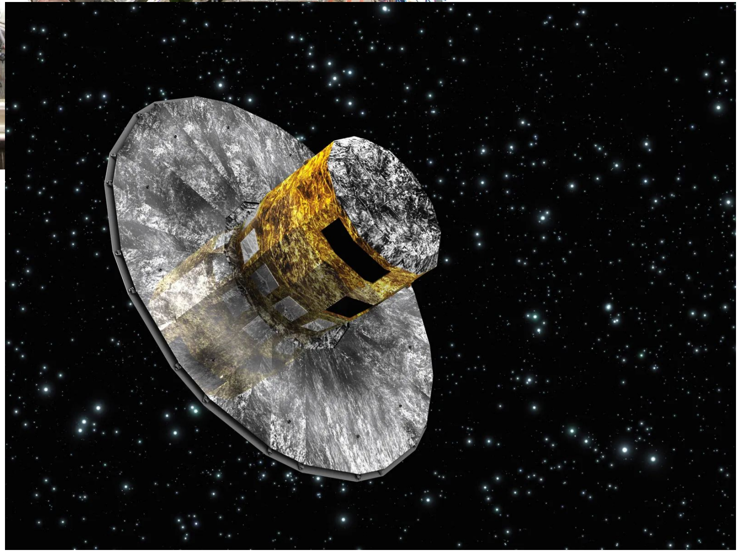
There is no required textbook. Resources for learning Python, machine learning, and more are plentiful online – you are encouraged to find the sources that help you learn best, whether they are video tutorials, textbooks, etc.

Take note of good references you find, and **as a class, we will compile our favorites into a shared list.** Here are some of my recommendations:

- Understanding Deep Learning by Simon J.D. Prince
- Deep Learning by Ian Goodfellow, Yoshua Bengio, and Aaron Courville
- Deep Learning for Physics Research by Martin Erdmann, Jonas Glombitza, Gregor Kasieczka, and Uwe Klemradt
- Neural Networks, a video playlist by 3Blue1Brown

Experiments





Grading

My biggest goal is for you to learn the critical practical skills you'll need to do strong research.

I designed this grading scheme to support your learning best, not to add undue stress – please come see me during office hours if you have any questions or concerns.

- **Participation:** 5%
  - Up to 2 missed classes (no questions asked) and in-class engagement will earn full credit here
- **Problem sets** (4 total): 28%
  - These will be graded on Canvas as quizzes
- **Projects** (4 total): 56%
  - These will be partially graded using a peer review system
- **Final written & oral presentation:** 11%

LLM Policy

This course emphasizes problem-solving and communication skills over specific knowledge recall. Because we are engaging with real research-grade datasets and problems, **you will encounter many points when you'll feel stuck and need to ask for help**. I encourage you to **turn to your peers for this as your first impulse** as much as possible for these reasons, and more:

1. Teaching material in an off-the-cuff manner is absolutely the best way to learn it.

2. Research is collaborative by nature, and communicating frequently with peers / learning to effectively ask for help is a critical skill to practice now. One of the biggest mistakes I see young researchers make is not asking for help soon enough or showing vulnerability.

3. You already have time set aside for exactly this purpose in our class sessions!

LLM Policy

- That said, LLMs can also help you in your research, if you know how to use them well and are aware of their pitfalls.
- If you want to use them, I encourage you to find out how to effectively incorporate LLMs into your research practice – we will discuss these tools & strategies during one of our class sessions.
- Note that the majority of your grade in this course will come from the strength of your projects and your oral & written communication about your work, as judged by myself as well as your peers.
- As in real research, all tools are available to you in this course, but code and writing that reads as obviously LLM-generated will not inspire confidence that your work is rigorous and trustworthy. You must be able to explain and justify your choices/methods.

Connecting with me

- **Pronouncing my name:** Feel free to call me Mariel ("mary-ELL") or Professor Pettee ("puh-TEE")
- **Office hours/tea time: Tuesdays from 3pm-4pm** (to be confirmed), Chamberlin 4275. I will have tea available. You are encouraged to come by and say hello, chat, and/or ask me questions.
- **Contact:** mpettee@wisc.edu

About me

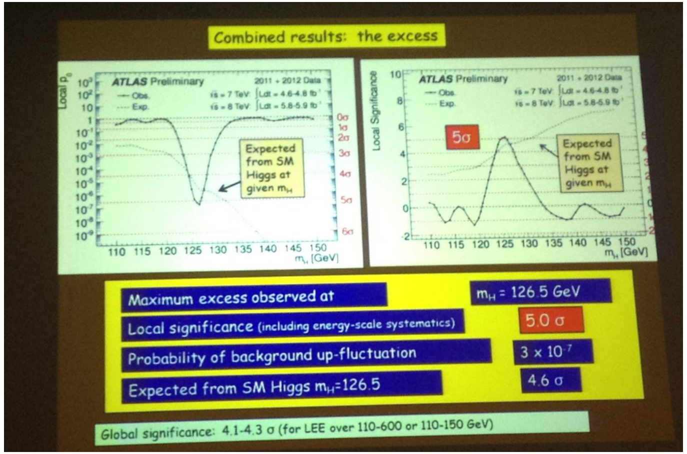
https://marielpettee.com/

Prof. Mariel Pettee

New assistant professor in the Physics Department at UW-Madison & RISE AI affiliate

I have been involved with the ATLAS Experiment at CERN since 2012:

ATLAS EXPERIMENT



https://marielpettee.com/

Prof. Mariel Pettee

New assistant professor in the Physics Department at UW-Madison & RISE AI affiliate

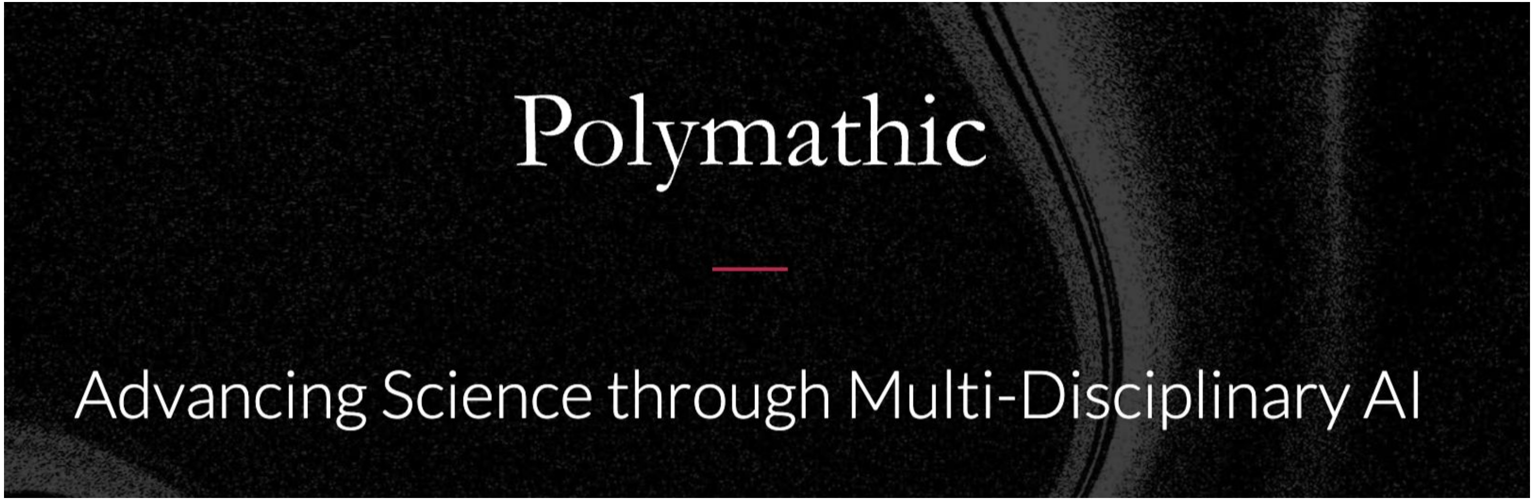
I've lived in many places throughout my academic career...

https://marielpettee.com/

Prof. Mariel Pettee

New assistant professor in the Physics Department at UW-Madison & RISE AI affiliate

For the past 2 years, I've been a member of the Polymathic AI collaboration:



https://marielpettee.com/

Prof. Mariel Pettee

New assistant professor in the Physics Department at UW-Madison & RISE AI affiliate

For the past 2 years, I've been a member of the Polymathic AI collaboration:

OUR MISSION

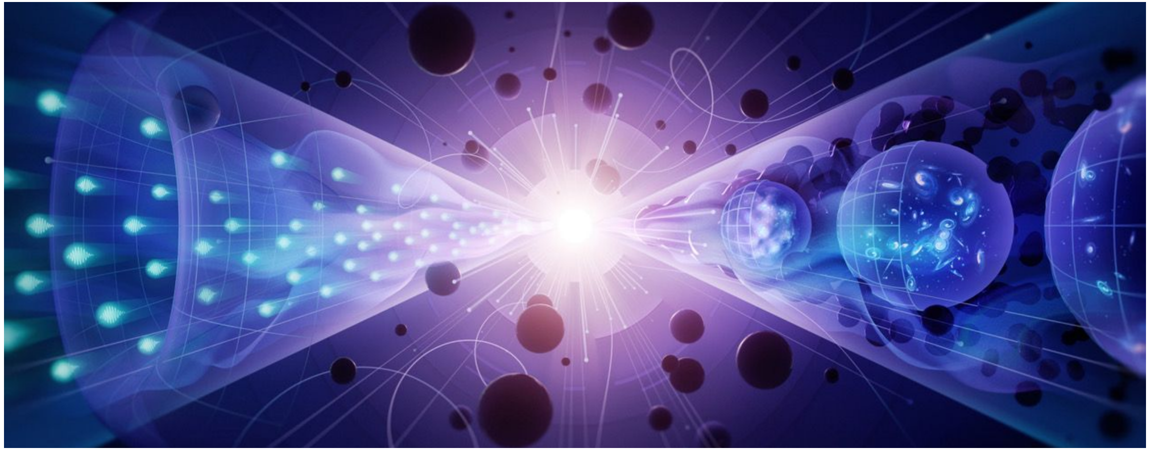
To usher in a new class of machine learning for scientific data, building models that can leverage shared concepts across disciplines. We aim to develop, train, and release such foundation models for use by researchers worldwide.

https://marielpettee.com/

Prof. Mariel Pettee

New assistant professor in the Physics Department at UW-Madison & RISE AI affiliate

My research explores how to reveal subtle patterns in large-scale data and leverage shared knowledge across datasets, instruments, and even disciplines. To do this, I build custom AI models adapted for physics analysis.

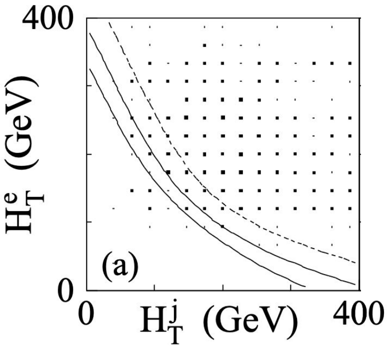


Brief & incomplete history of ML in Physics

Machine Learning & (Particle) Physics: New architectures

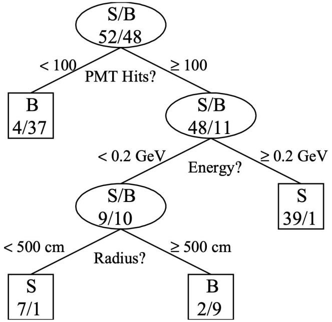
1990 **2000** **2010** **2020** **2030**

Artificial NNs



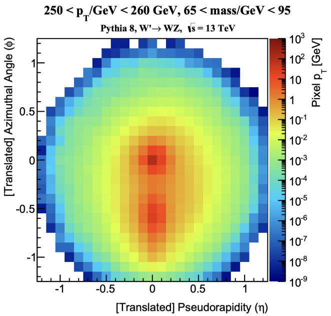
arXiv:hep-ex/9707033

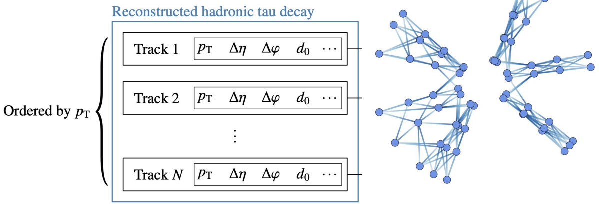
BDTs



arxiv:physics/0408124v2

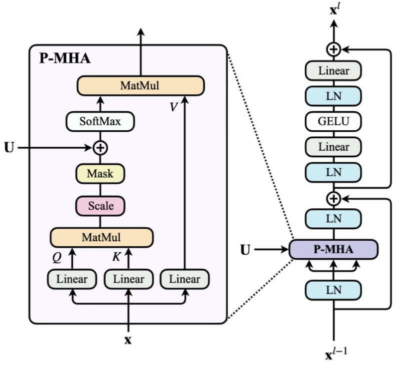
Neural networks





arXiv:1511.05190, arXiv:2007.13681

Transformers

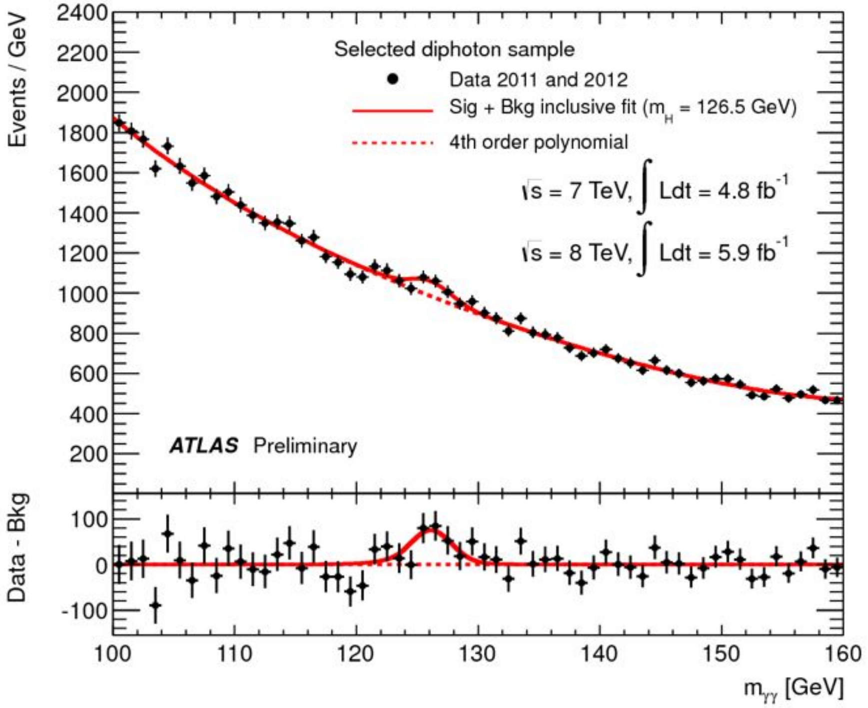


(b) Particle Attention Block

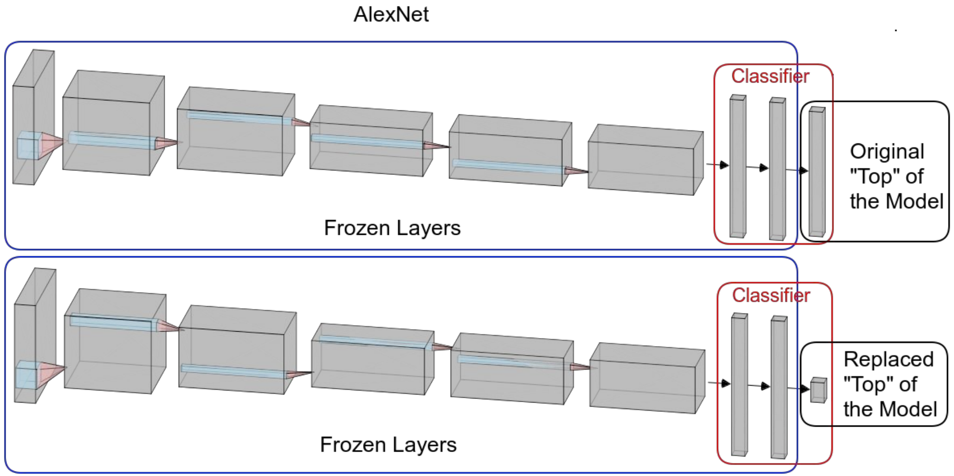
arXiv:2202.03772

2012

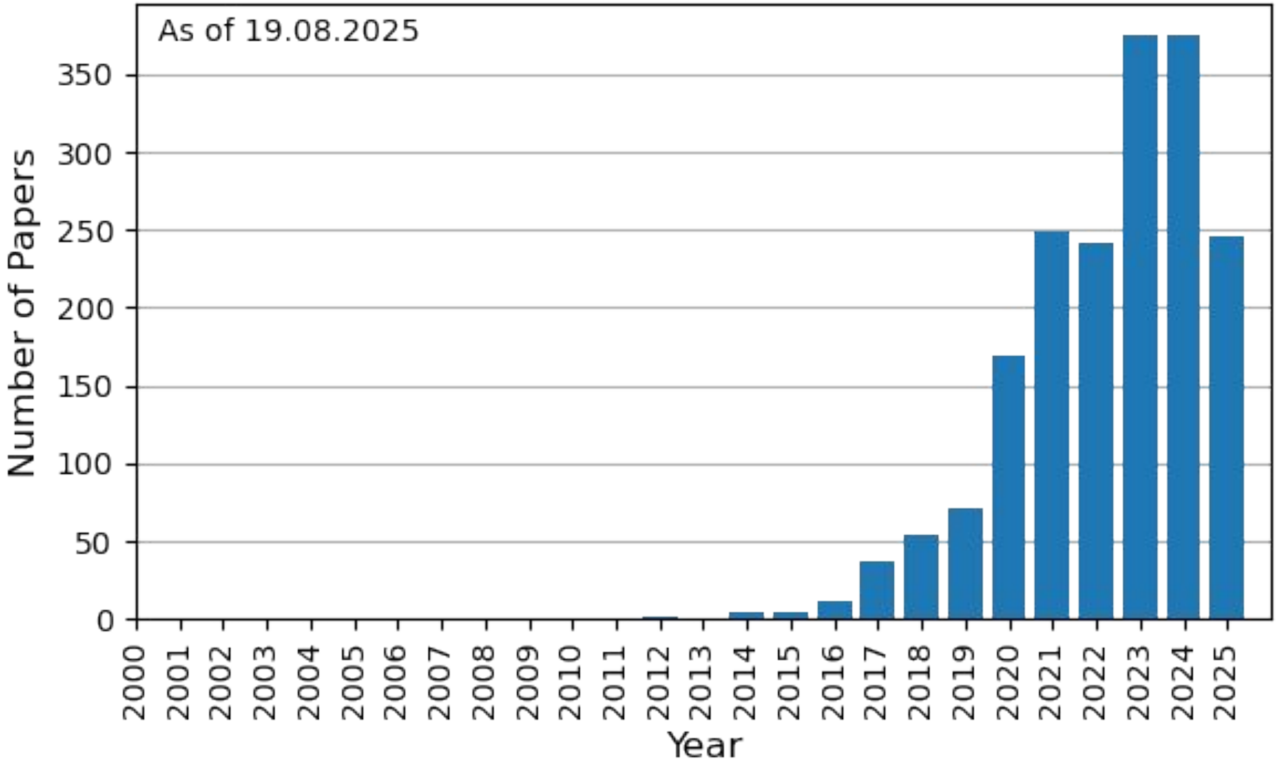
The Higgs boson



AlexNet

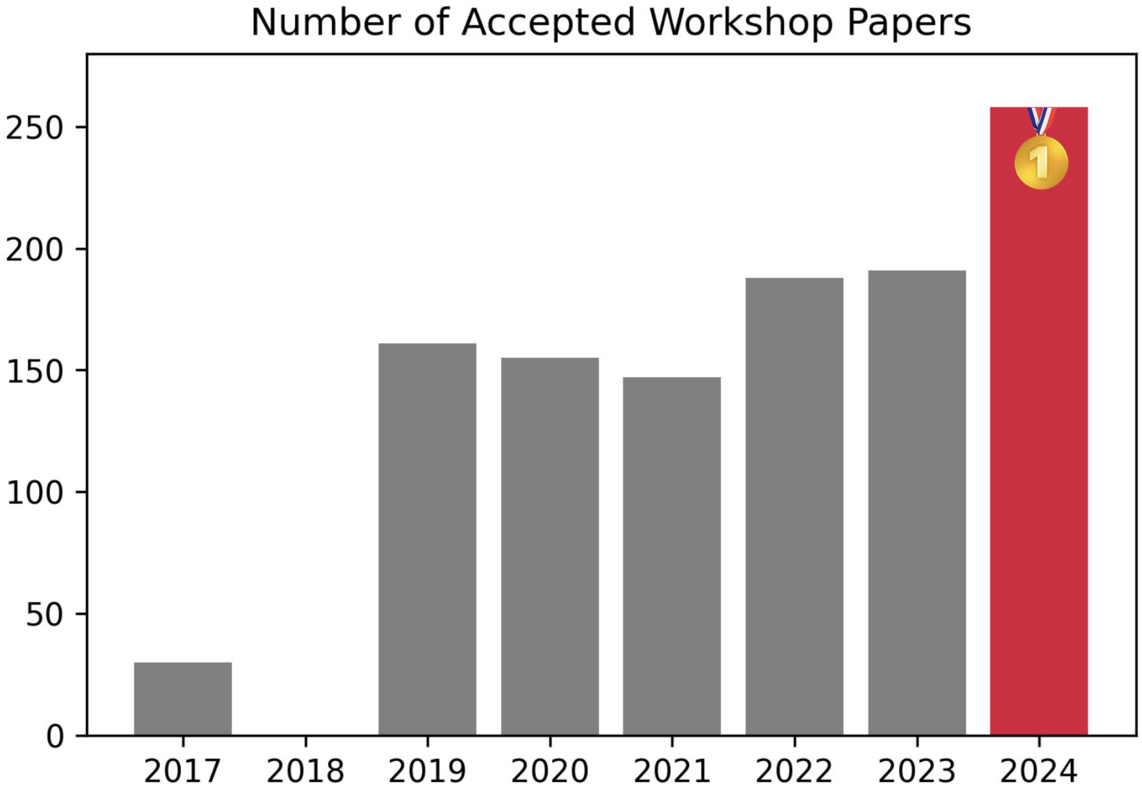


HEP-ML Papers by Year



https://iml-wg.github.io/HEPML-LivingReview/

NeurIPS ML for the Physical Sciences Workshop



+40% increase in submissions in 2025!

https://ml4physicalsciences.github.io/

2024

The Nobel Prize in Physics (2024)

Ill. Niklas Elmehed © Nobel Prize Outreach

John J. Hopfield

Prize share: 1/2

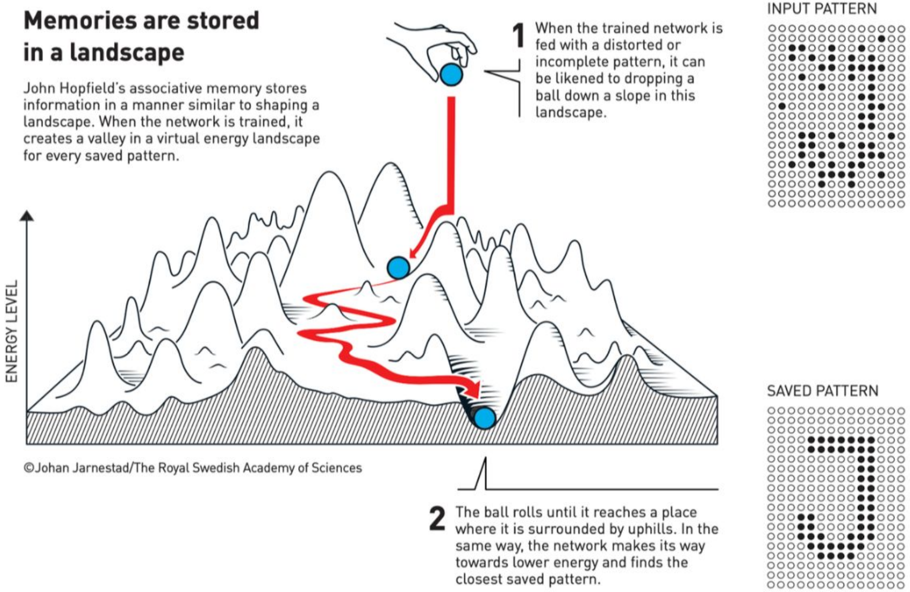
Ill. Niklas Elmehed © Nobel Prize Outreach

Geoffrey Hinton

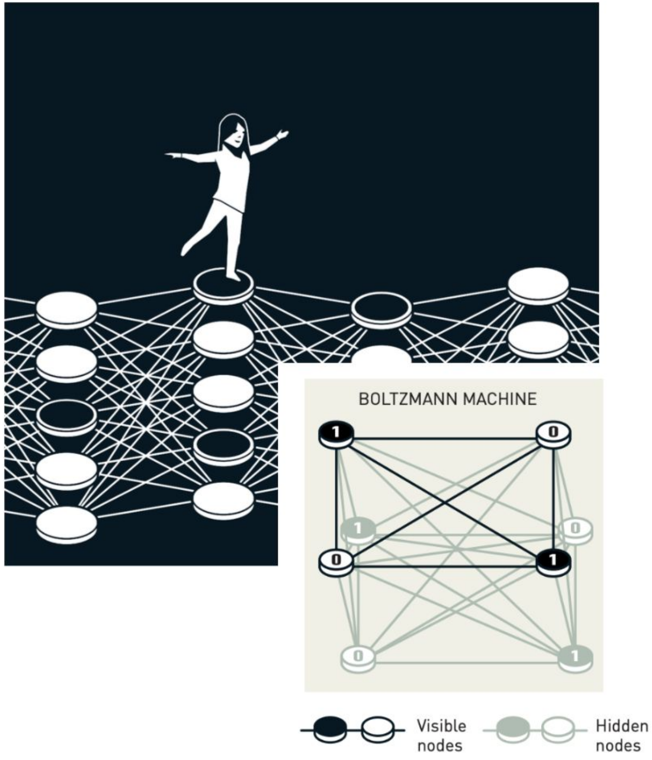
Prize share: 1/2

"for foundational discoveries and inventions that enable machine learning with artificial neural networks"

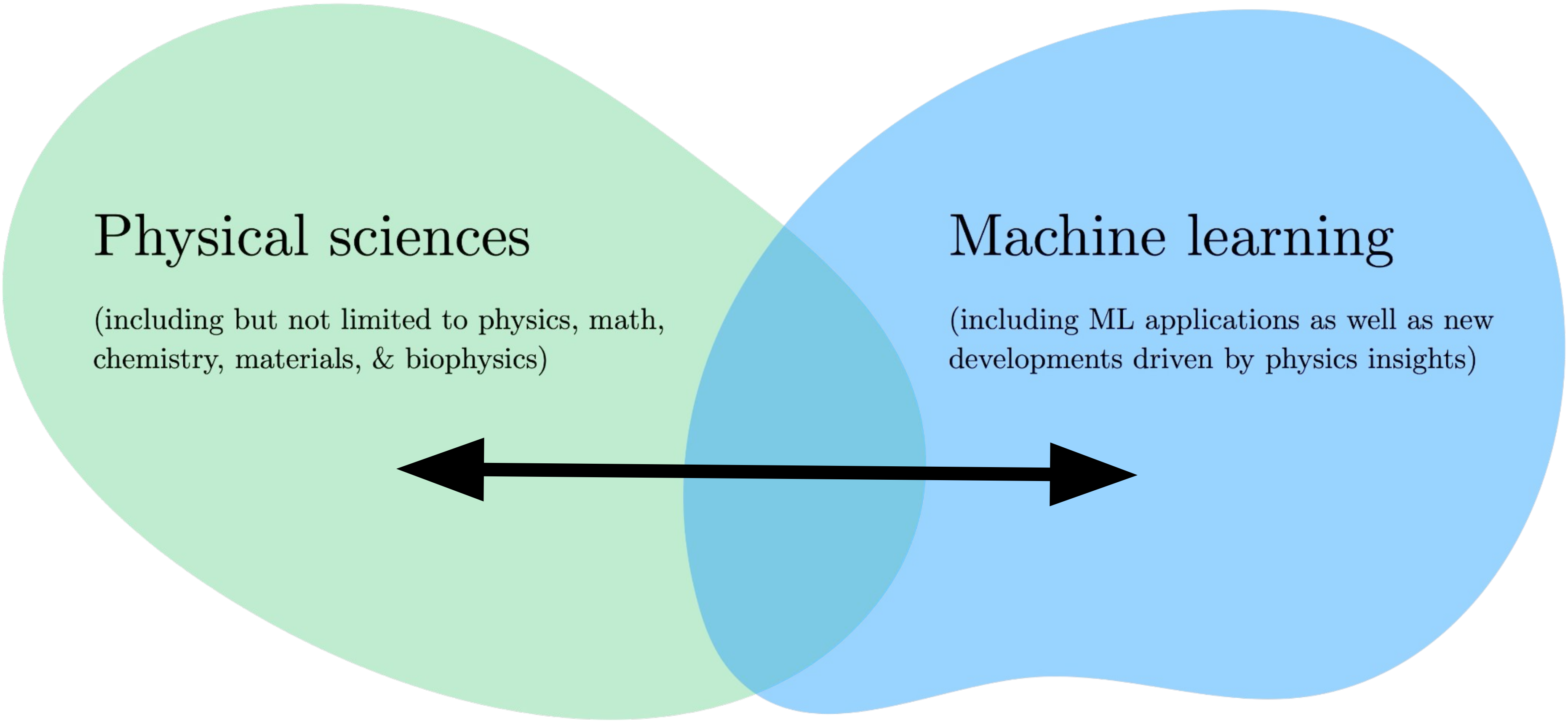
2024

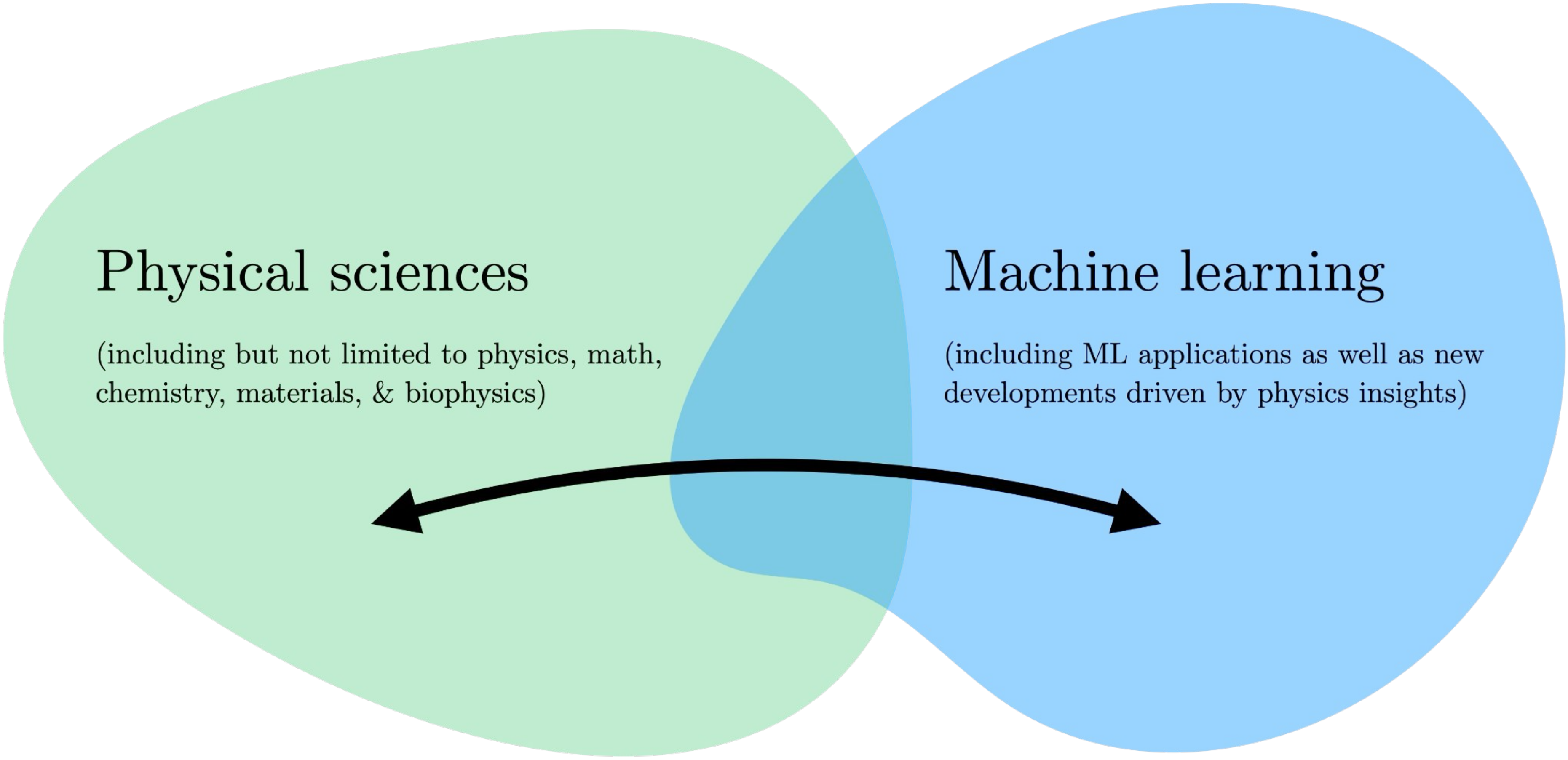


Memories are stored in a landscape
John Hopfield's associative memory stores information in a manner similar to shaping a landscape. When the network is trained, it creates a valley in a virtual energy landscape for every saved pattern.
1 When the trained network is fed with a distorted or incomplete pattern, it can be likened to dropping a ball down a slope in this landscape.
ENERGY LEVEL
©Johan Jarnestad/The Royal Swedish Academy of Sciences
2 The ball rolls until it reaches a place where it is surrounded by uphills. In the same way, the network makes its way towards lower energy and finds the closest saved pattern.
INPUT PATTERN
SAVED PATTERN



BOLTZMANN MACHINE
1
0
1
0
0
1
0
1
Visible nodes
Hidden nodes


Physical sciences
(including but not limited to physics, math, chemistry, materials, & biophysics)
Machine learning
(including ML applications as well as new developments driven by physics insights)


Physical sciences
(including but not limited to physics, math, chemistry, materials, & biophysics)
Machine learning
(including ML applications as well as new developments driven by physics insights)

Setting up your compute environment

Setting up your compute environment

- This course is based heavily on Python and Jupyter notebooks
    - The notebooks should be able to run on your personal laptop or on Google Colab
- I will also ask you to create a Github repository to share your problem sets and projects
- Finally, you will also be asked to do some written and oral summaries of your work. For the formal written work, you'll want to use LaTeX.
    - You can download this locally, if you prefer
    - Otherwise, I recommend using Overleaf

- ❏ Create a Jupyter notebook and run the following successfully:

```
import numpy as np
import matplotlib.pyplot as plt
import pandas as pd
```

  - **Bonus:** Make the most elegant plot you can of 3 decaying sine waves and save it as a PDF. (Also, if on Colab, try connecting to a GPU runtime.)

- ❏ Create a Git repository on Github and push your notebook to this repository.
  - **Bonus:** Add a folder structure for psets & projects and a README.md file

- ❏ Write up your “results” using the NeurIPS 2025 paper template in LaTeX. (Make up a title and abstract only.)
  - **Bonus:** Add a figure and table of contents. Add a citation & bibliography.