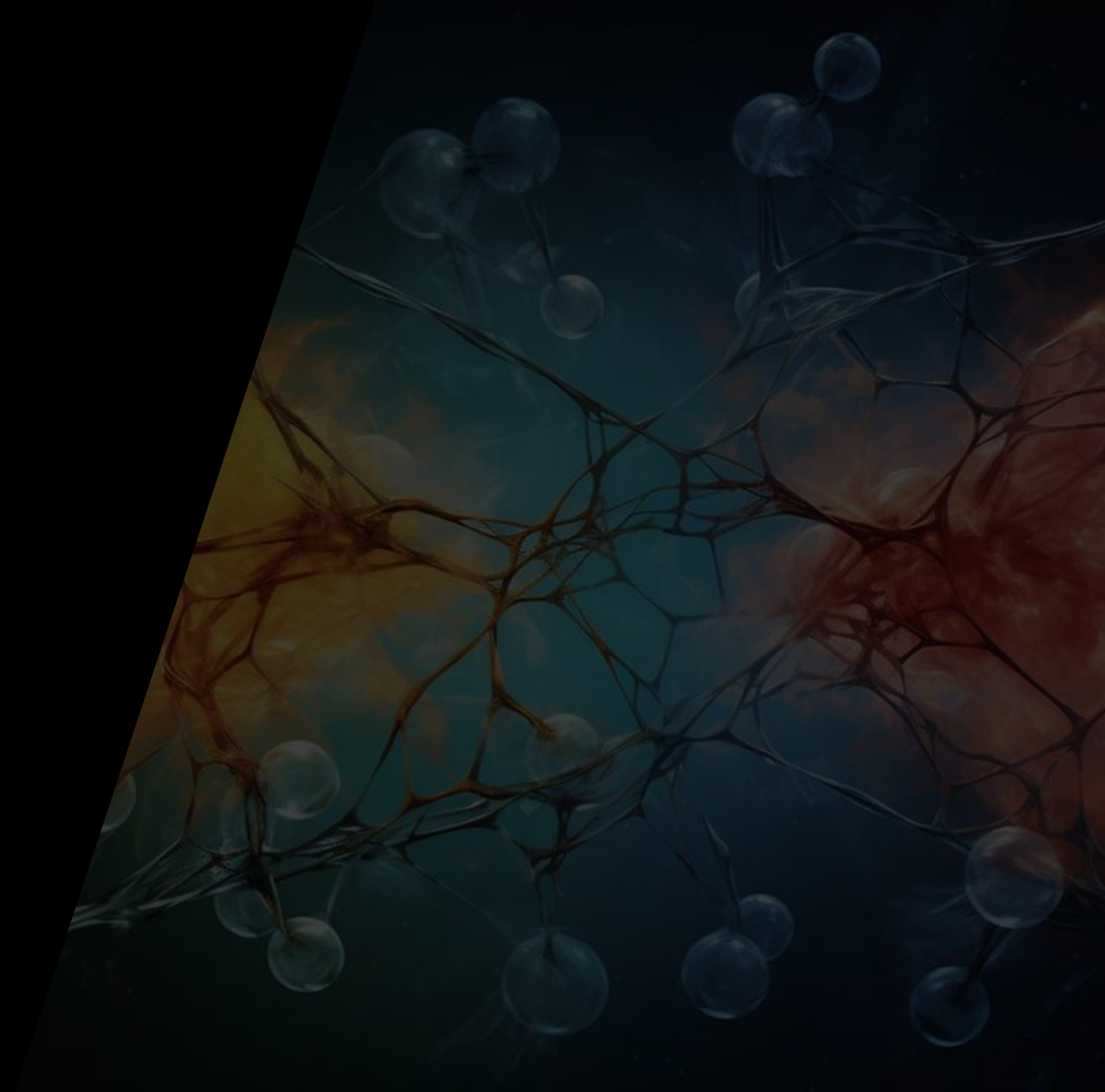
Projects



Contacts:

tony.bonnaire@ens.fr

sebastien.pierre@phys.ens.fr

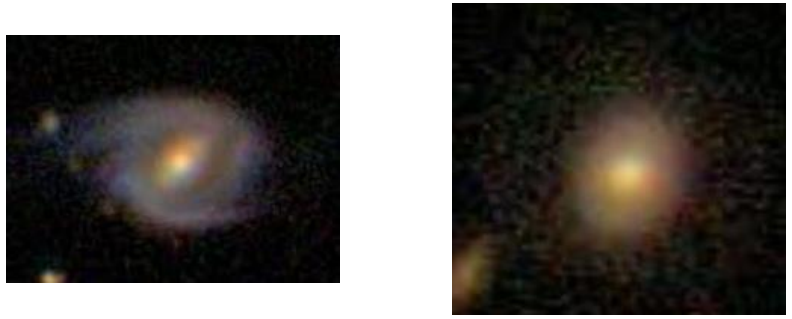
Difficulty: 1/3

I. Galaxy classification with computer vision algorithms

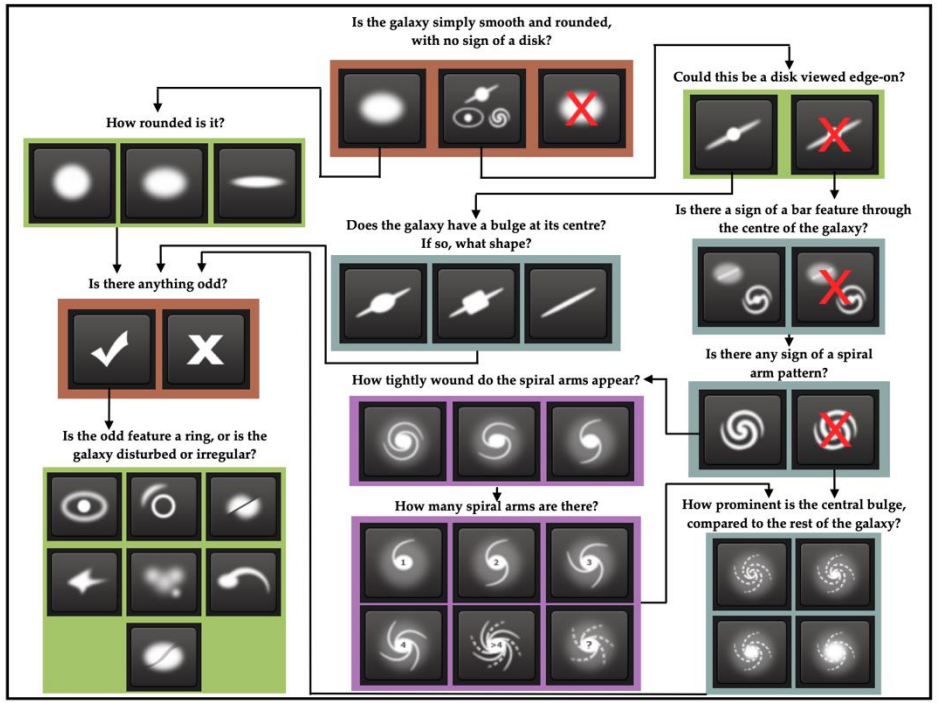
Keywords: Astrophysics, supervised learning, classification, 2D images

What you'll do:

- Preprocess the image dataset (424× 424 pixels),
- Identify and **implement** several machine-learning based solutions for predicting classifying the galaxies (examples: **CNNs, Residual Nets, Vision transformers**)
- **Discuss** possible **improvements** and implement them,
- If time, different directions are possible: how to handle uncertainties, improve the models, look for more complex datasets, etc.



Example of train images



The decision tree that volunteers filled for each galaxies

References: *Kaggle 2013* challenge, Willet+*2013*, *Polymathic AI*

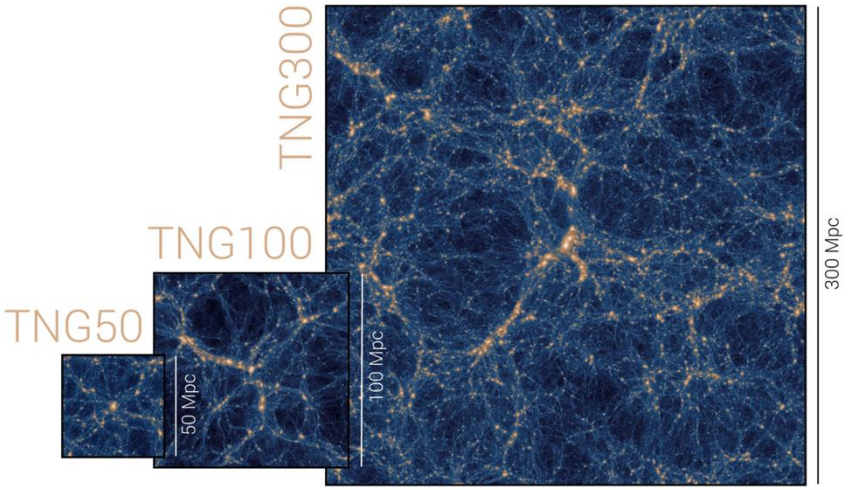
Difficulty: 2/3

II. Generating non-Gaussian data by Diffusion

Keywords: Unsupervised learning, generation, 2D images.

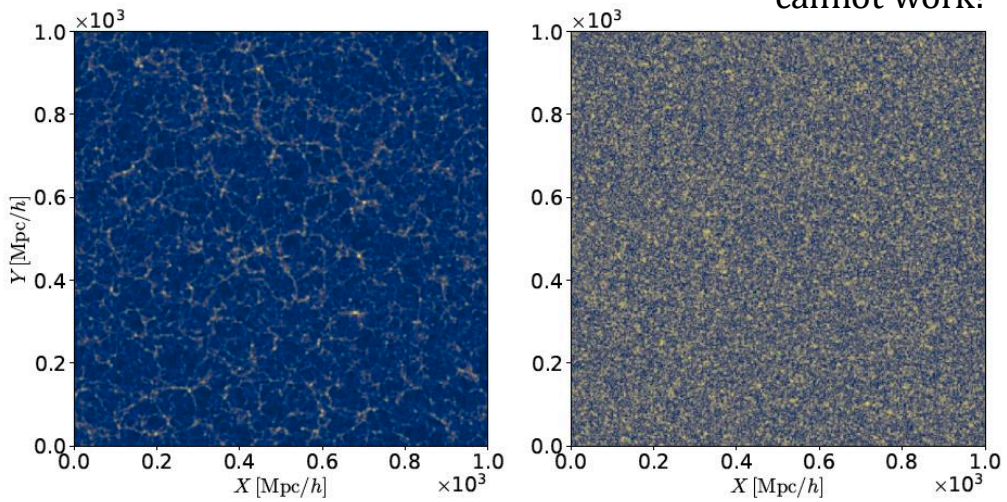
What you'll do:

- Understand the problematic of **unsupervised learning** for generation,
- Understand the **maths of Diffusion Models** and implement a neural-network-based solution for generated samples of 2D fields (cosmology or ising/phi4 fields in statistical physics),
- Discuss **shortcomings**, **evaluation metrics**, possible improvements, and alternative architectures.
- Possible directions: theoretical understanding of diffusion, conditional generation, or applications to real problems.



Several months of runtime!

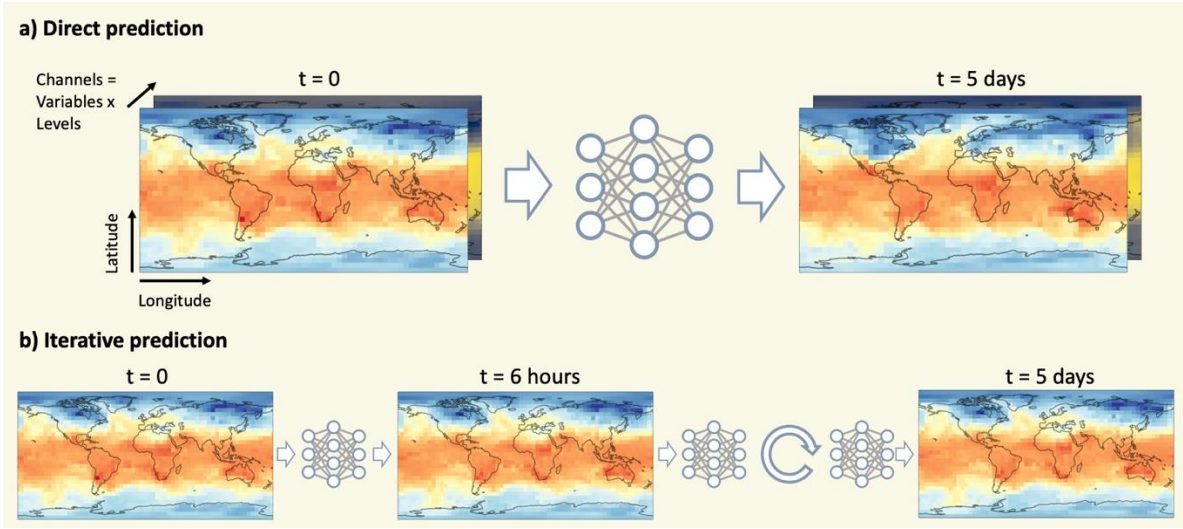
Naïve reconstruction cannot work!



References: Ho+21, Villaescusa Navarro+21, Mudur+23.

Difficulty: 2/3

III. AI-based weather forecast



Keywords: Supervised learning, 2D images.

What you'll do:

- Workout **spatio-temporal image data**
- Implement and/or use several **advanced vision models** to predict temperatures within few days' windows like **CNNs, Residual Networks and vision transformers**
- Compare and discuss several architectures.
- Possible directions: prediction of more than just temperature maps: precipitation, geopotential, or even rare and extreme events like cyclones or hurricane based on simulated datasets.

References: WeatherBench dataset paper, Rasp+2021

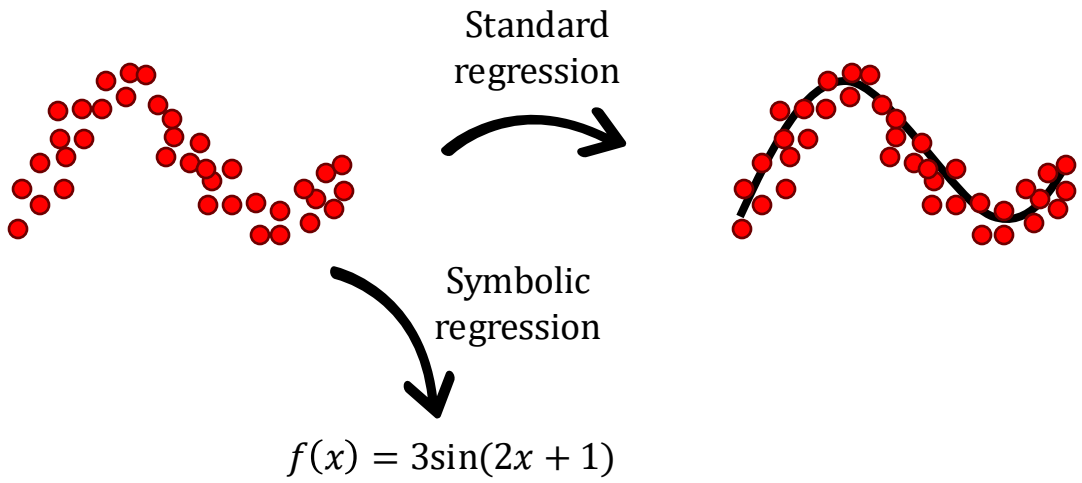
Difficulty: 3/3

IV. (Deep) Symbolic regression for new physics

Keywords: Mathematical physics, Supervised (or unsupervised) learning, symbolic regression.

What you'll do:

- Understand the problematic of **symbolic regression**,
- Discuss its applications in physics,
- Learn how to handle **text data,**
- Implement a **transformer-based solution** for symbolic regression,
- **Discuss** possible **extensions** and state-of-the-art methods.



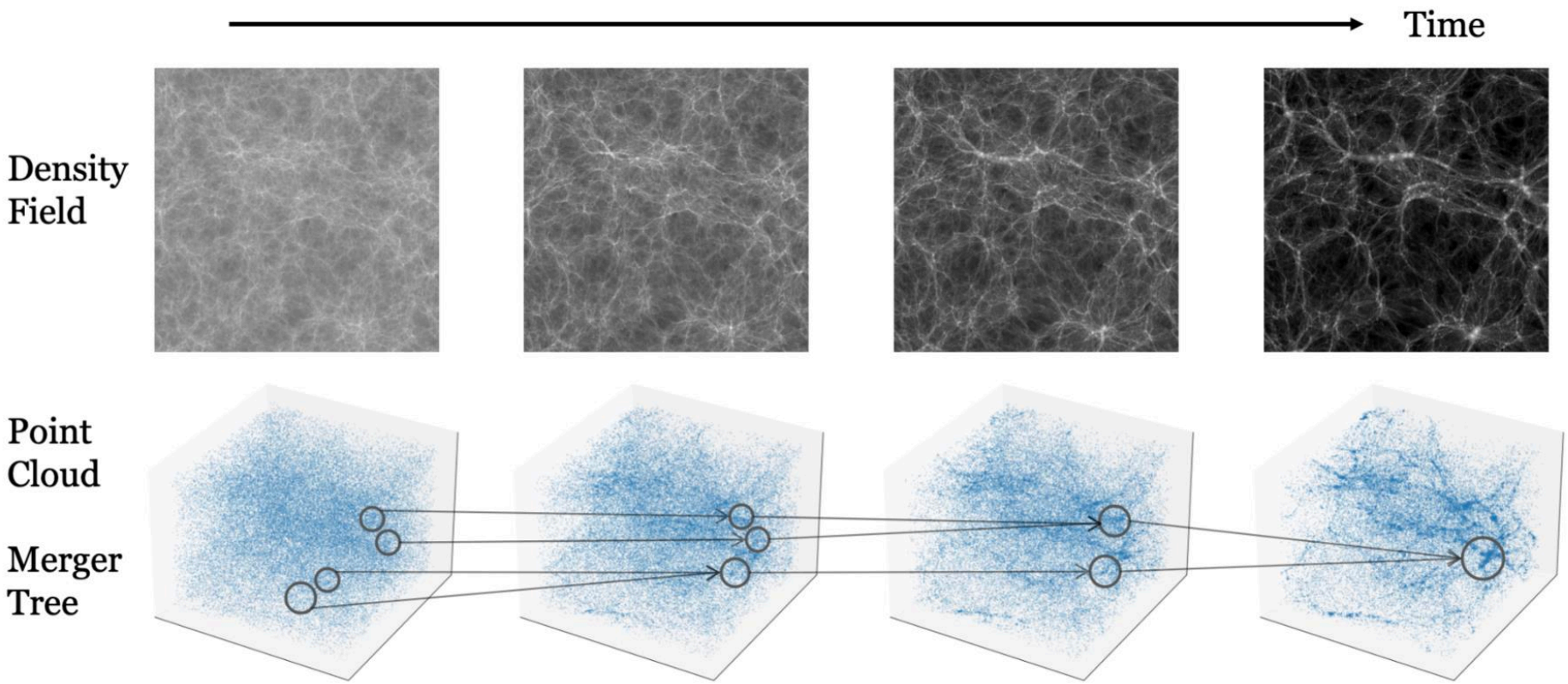
Applications in all fields of physics, from quantum physics to cosmology

Ideas of data: Physics expressions or Feynman diagrams in quantum physics

References: Kimmieny+21, Alnqaydan+23

Difficulty: 1/3

V. Inferring cosmological parameters from galaxy positions

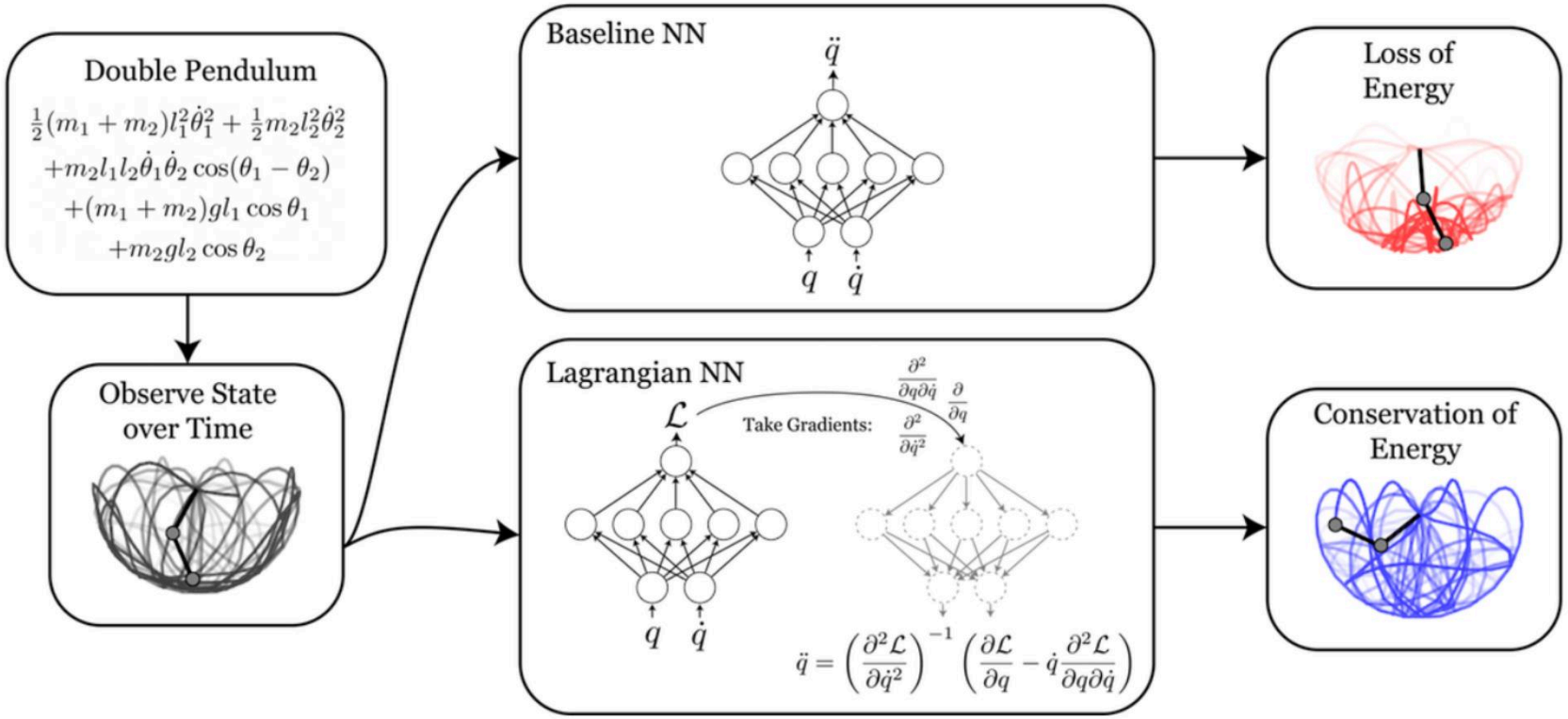


- Goal : infer Ω_m, σ_8 from galaxy point cloud

Reference: https://cosmobench.streamlit.app/, https://arxiv.org/pdf/2507.03707

Difficulty: 2/3

VI. Physics-inspired neural networks (PINN)

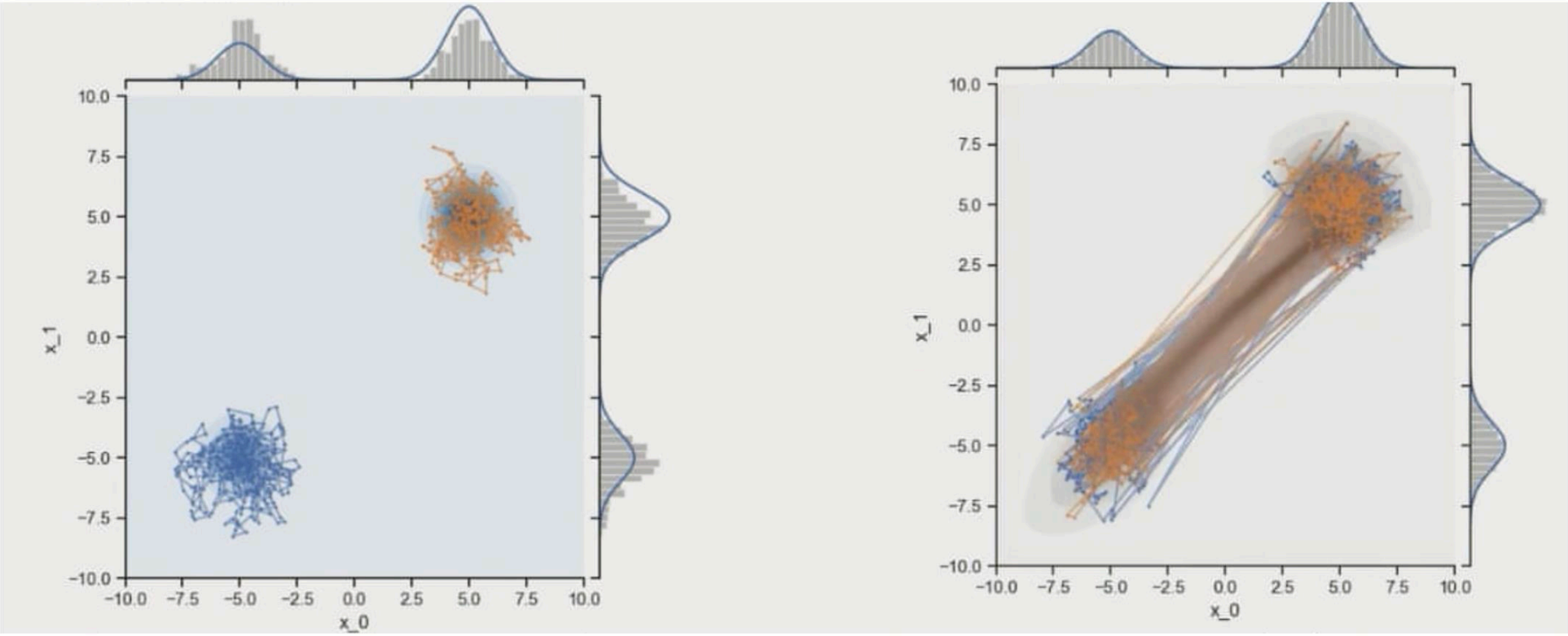


- Goal : Incorporate physical knowledge into classical NN

References: [Cranmer+2020](), [Greydanus+2019]()

Difficulty: 3/3

VII. Enhancing sampling with machine learning



- Goal : Use ML to assist sampling of physical systems

Reference: [Gabrié+2022]