Andrey Ustyuzhanin

Machine Intelligence

as a new kind of science

2021

MLHEP 2021

HIGHER SCHOOL OF ECONOMICS
NATIONAL RESEARCH UNIVERSITY

LAMBDA • HSE

Yandex

SCHOOL OF DATA ANALYSIS

EPFL

SIT
Schaffhausen Institute of Technology

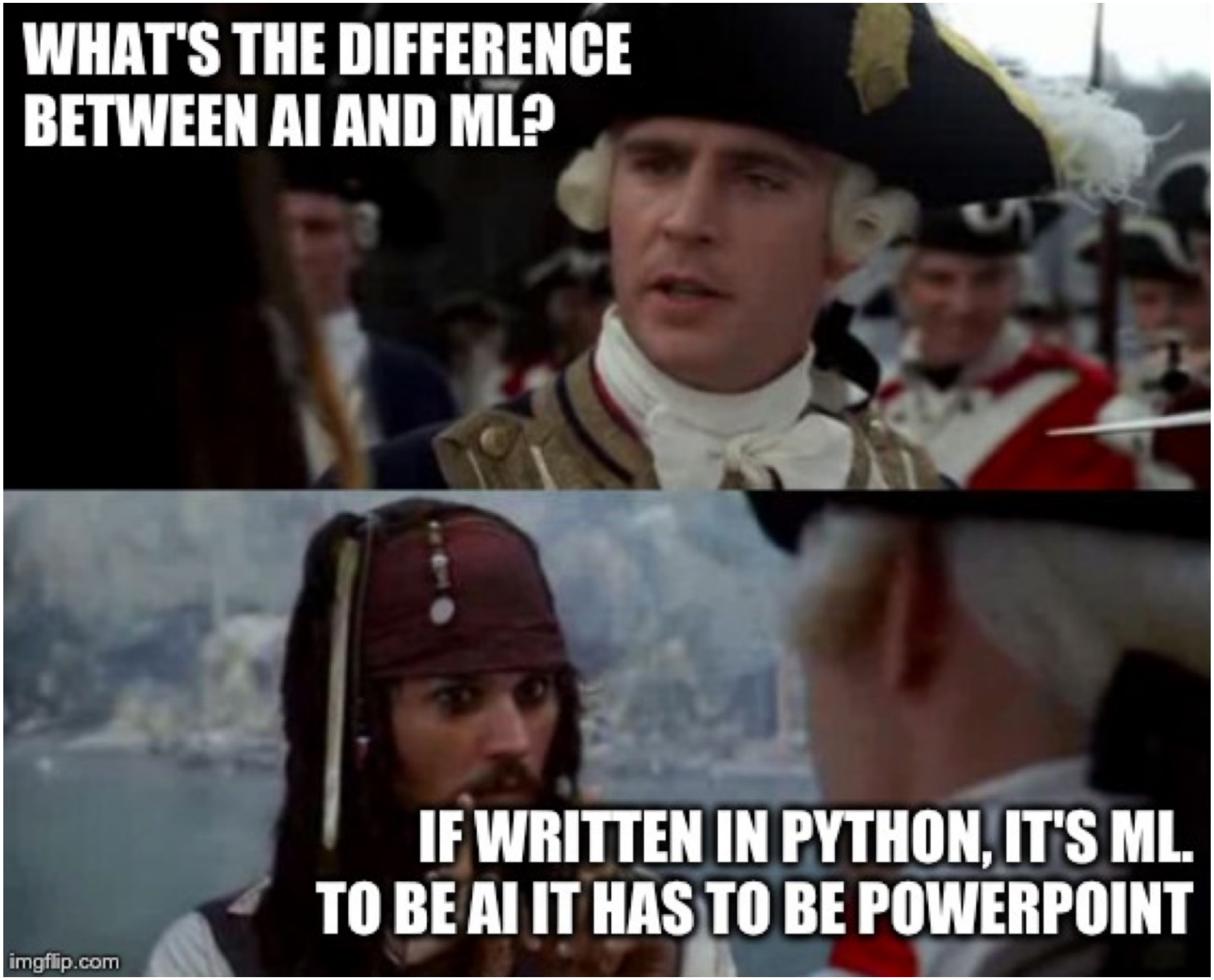
A few words about your lecturer

- Head of Laboratory of methods for Big Data Analysis at HSE University, AI/ML expert @SIT, researcher @MISiS
  - Applications of Machine Learning to natural science challenges
- Head of LHCb Yandex School of Data Analysis (YSDA) team
- Co-organizer of data-driven Kaggle competitions for Physics and IDAO challenges
- Lecturer and organizer: 6 MLHEP summer schools, ML courses at ICL, Clermont-Ferrand, URL Barcelona, Coursera

Machine Intelligence Magic

Tesler’s theorem:
“AI is whatever hasn’t been done yet”

Larry Tesler


WHAT'S THE DIFFERENCE BETWEEN AI AND ML?
IF WRITTEN IN PYTHON, IT'S ML. TO BE AI IT HAS TO BE POWERPOINT
imgflip.com

Remarkable examples of ML technologies

- Human-level playing in computer games (Go, StarCraft, Dota 2) and winning world's Go champion Lee Sedol by Google AlphaGo;
- Understanding and generation of human-readable and understandable texts;
- Recognition and generation of images indistinguishable from photos by the naked eye;
- Simulation of complicated physical processes;
- Controlling complicated real-time systems like quantum qubits;
- Controlling autonomous vehicles in populated regions;
- and many others.

This X does not exist

This Person Does Not Exist

The site that started it all, with the name that says it all. Created using a style-based generative adversarial network (StyleGAN), this website had the tech community buzzing with excitement and intrigue and inspired many more sites.

Created by Phillip Wang.

This Cat Does Not Exist

These purr-fect GAN-made cats will freshen your feeline-gs and make you wish you could reach through your screen and cuddle them. Once in a while the cats have visual deformities due to imperfections in the model – beware, they can cause nightmares.

Created by Ryan Hoover.

This Rental Does Not Exist

Why bother trying to look for the perfect home when you can create one instead? Just find a listing you like, buy some land, build it, and then enjoy the rest of your life.

Created by Christopher Schmidt.

https://thisxdoesnotexist.com/

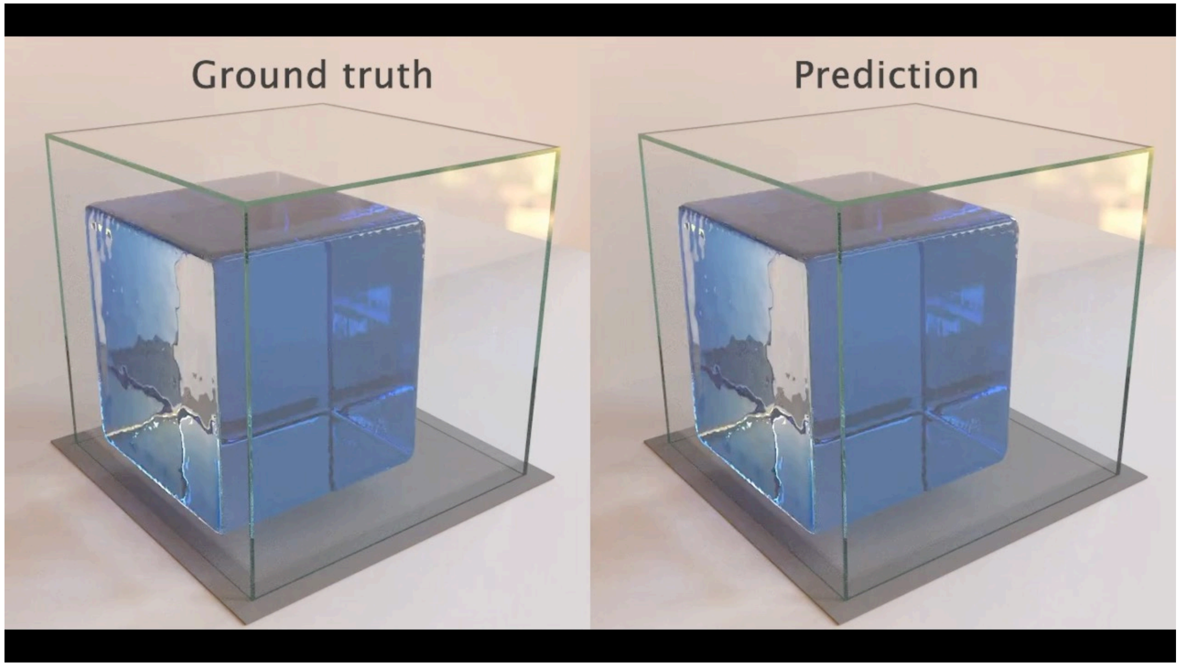
City street view simulation

This time lap shows the original scene (left), segmentation map (bottom right) and neural-network produced scene (right) by NVIDIA.

https://www.youtube.com/watch?v=ayPqjPekn7g

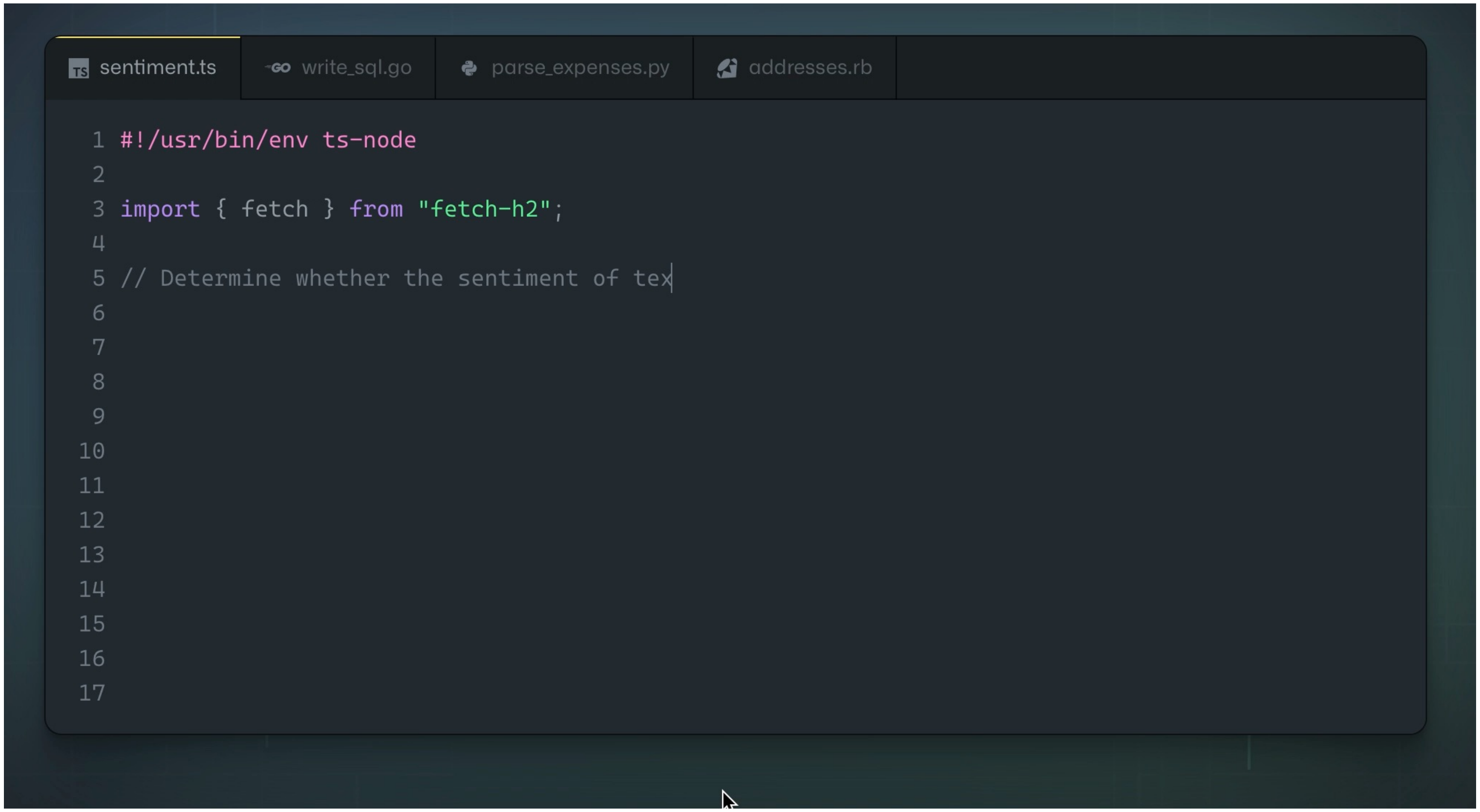
https://arxiv.org/abs/1808.06601

Fluid dynamics computation



- The time lap shows properly simulated water volume evolution (left) and simulated evolution by the trained neural network (right) with "Graph Network-based Simulators» (Alvaro Sanchez-Gonzalez et al.)

Github Copilot



Writing code by natural-text task description. Based on OpenAI Codex model (https://copilot.github.com/)

"Any sufficiently advanced technology is indistinguishable from magic."
Arthur Clarke

Machine Intelligence (MI)

Interdisciplinary technological area that embraces ML and bridges the gap between ML and domain specifics by iterative procedure:

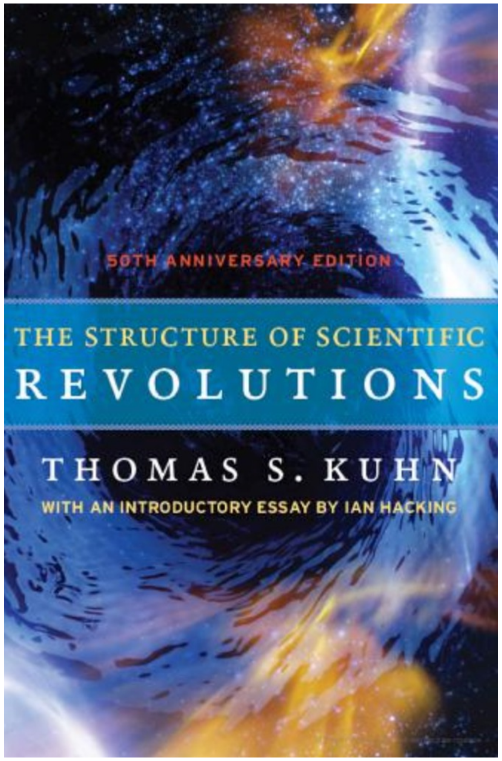
- Data collection from experiment or software simulation
- Prior choice and Hypothesis formulation
- Algorithm family selection from ML world (Decision Tree, Convolutional Neural Networks, Flows, etc.)
- Training of the algorithm using the data collected
- Validation of the trained algorithm
- Production deployment

Abridged history of science

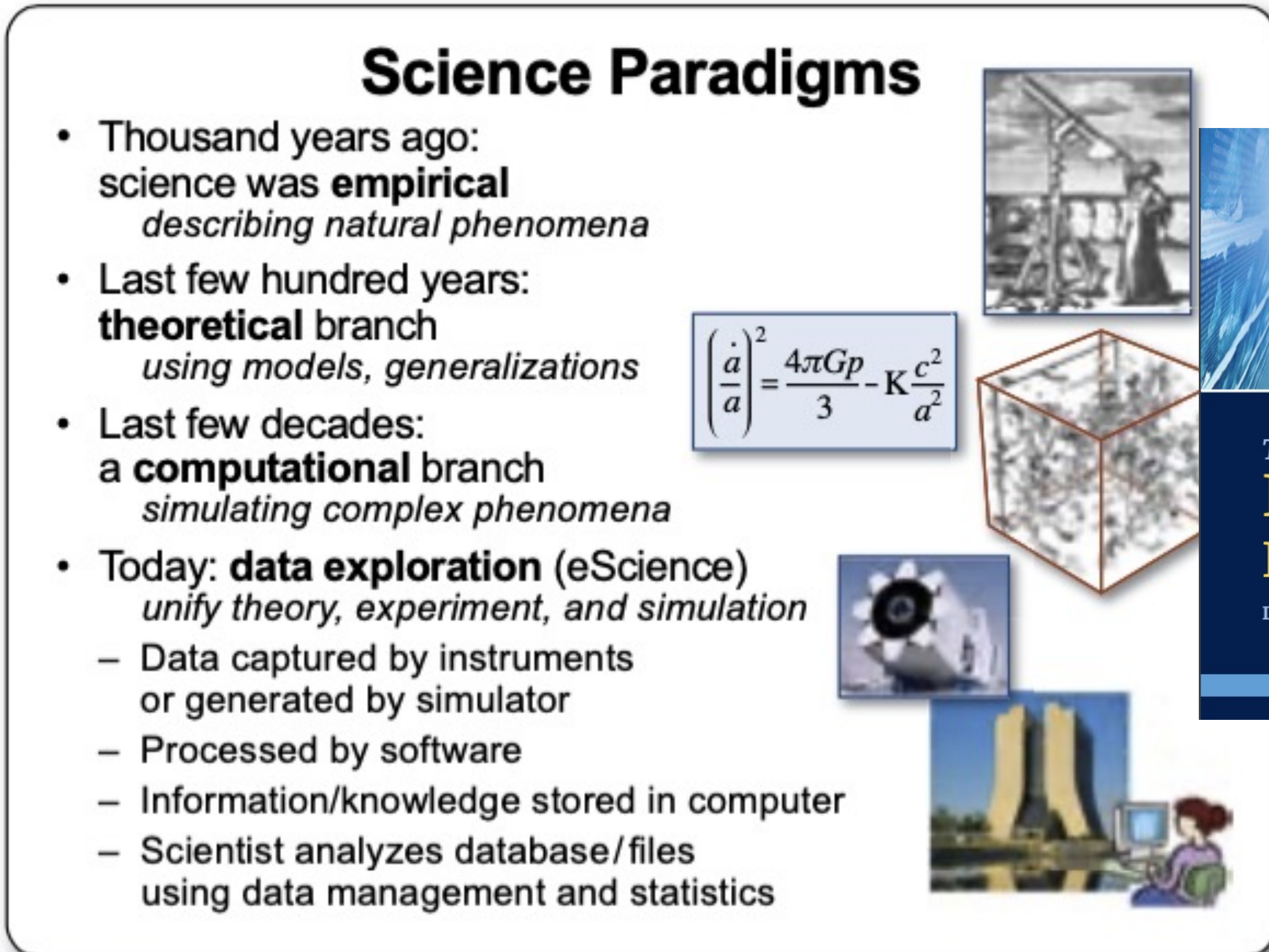
Scientific Paradigm by Thomas Kuhn

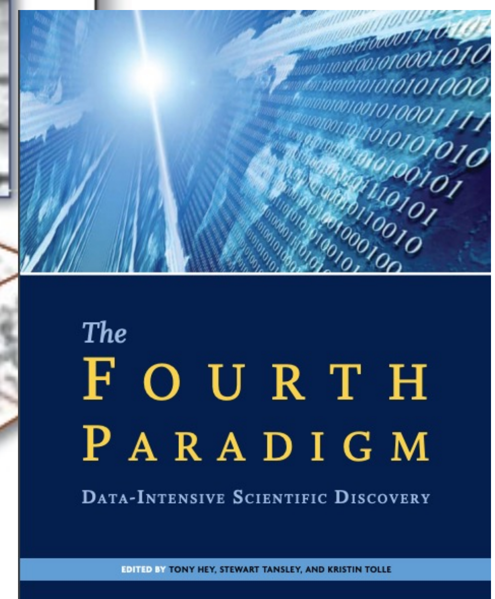
Universally recognized scientific achievements that, for a time, provide model problems and solutions for a community of practitioner:

1. ***what*** is to be **observed** and scrutinized;
2. What are the ***questions*** that are **probed** for answers and *how* these questions are to be **structured**;
3. ***what* predictions** made by the primary theory within the discipline;
4. *how* the **results** of scientific investigations should be **interpreted**;
5. *how* an **experiment** is to be conducted, and *what* equipment is available.



Jim Gray’s vision of science, 2009



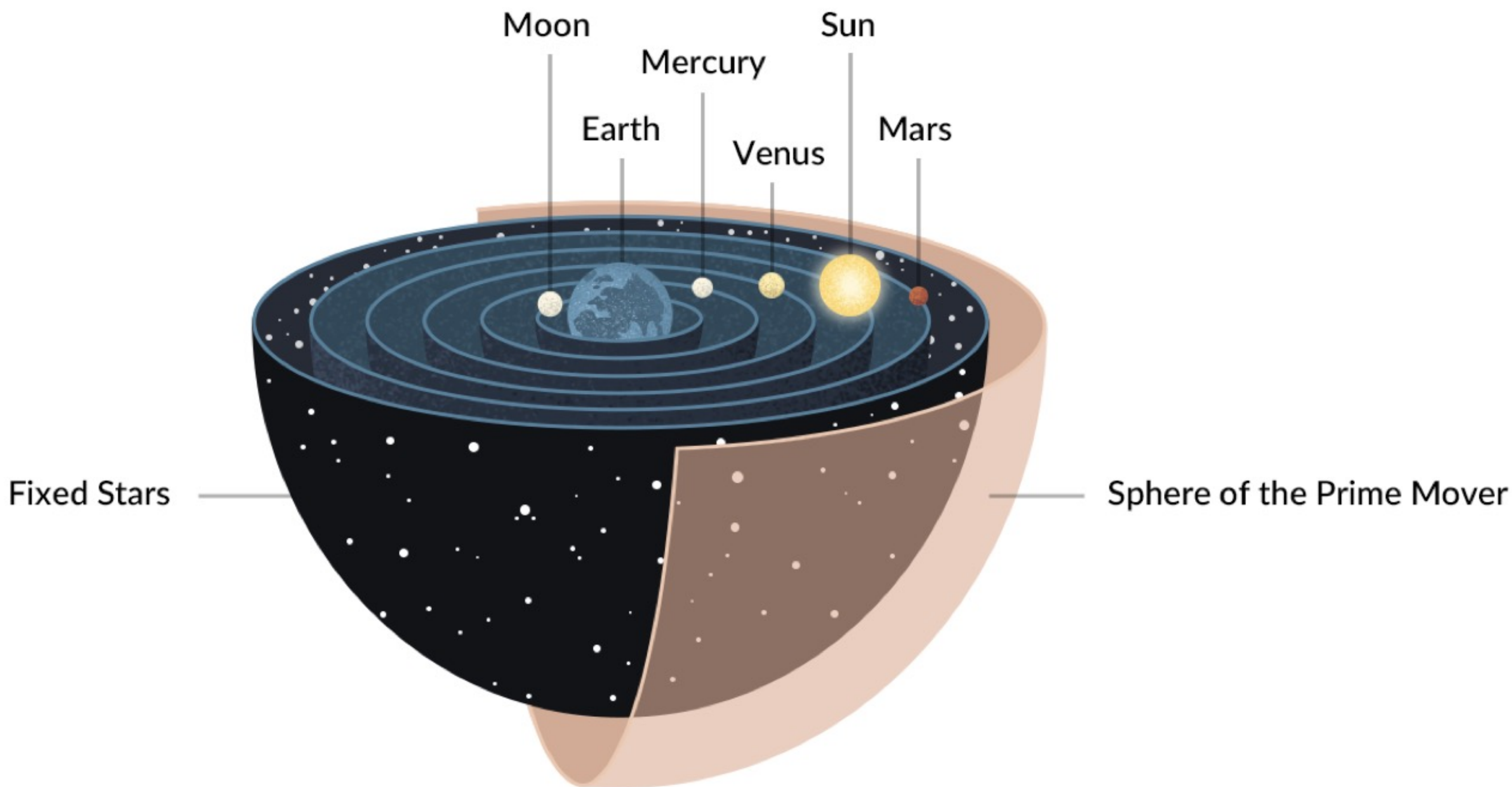


link

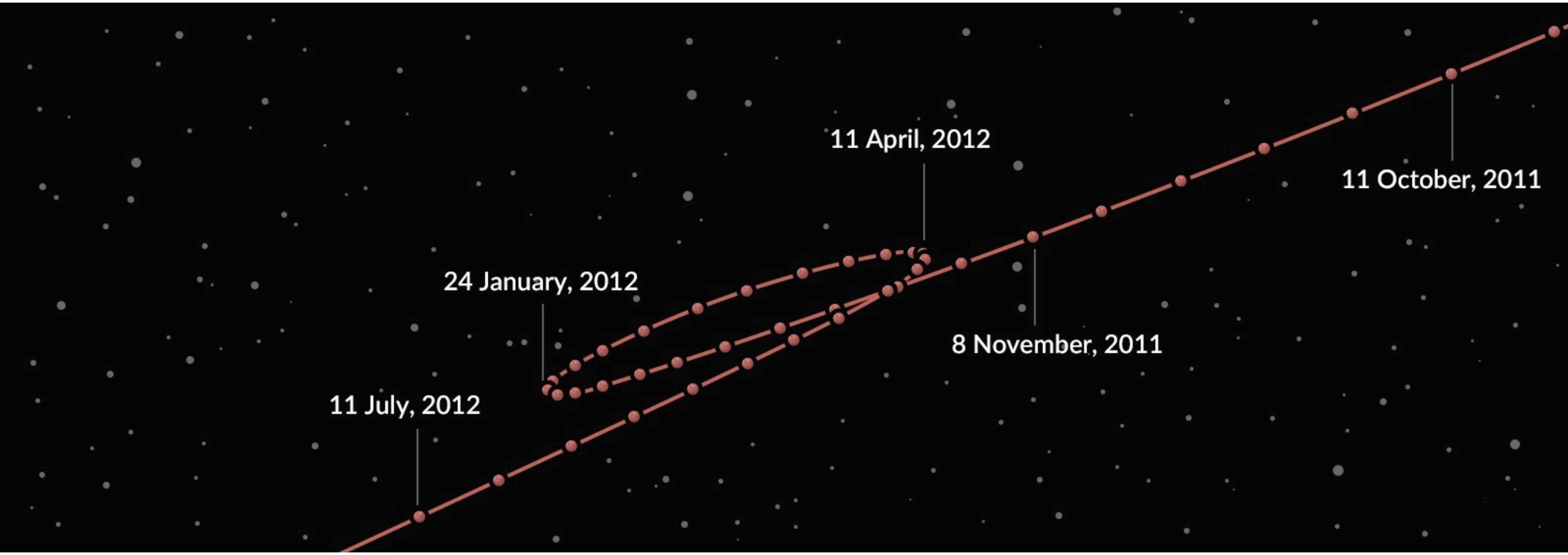
Empirical Science Questions

- How can we navigate using stars?
- Does the sun rotate around the Earth or vice versa?
- Which body does fall faster?
- What are the causes of solar eclipse?
- Can we estimate time of the next eclipse?
- How to describe motion of the moon and the planets?
- Is Earth flat?

Aristotle’s model

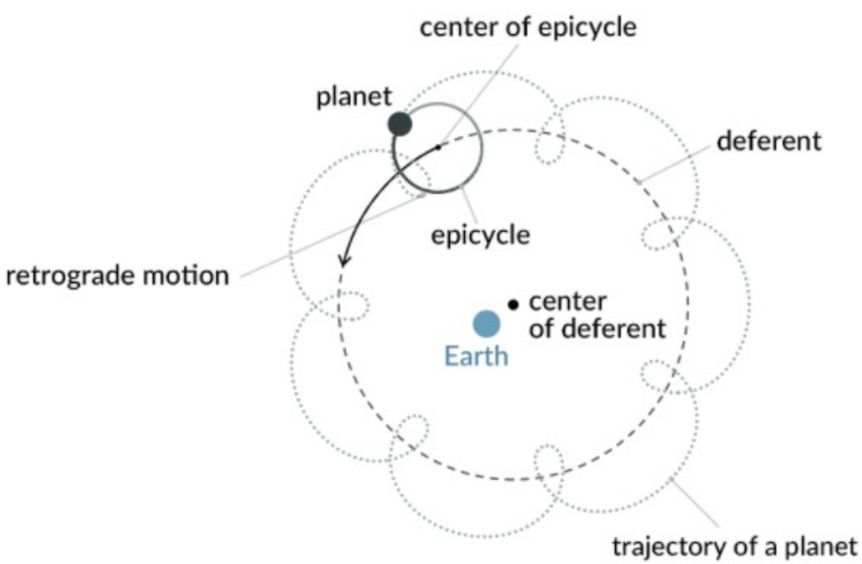


Retrograde planet motion as observed from Earth



https://en.pelican.study/static/bundles/demonstrations/helio/index.html

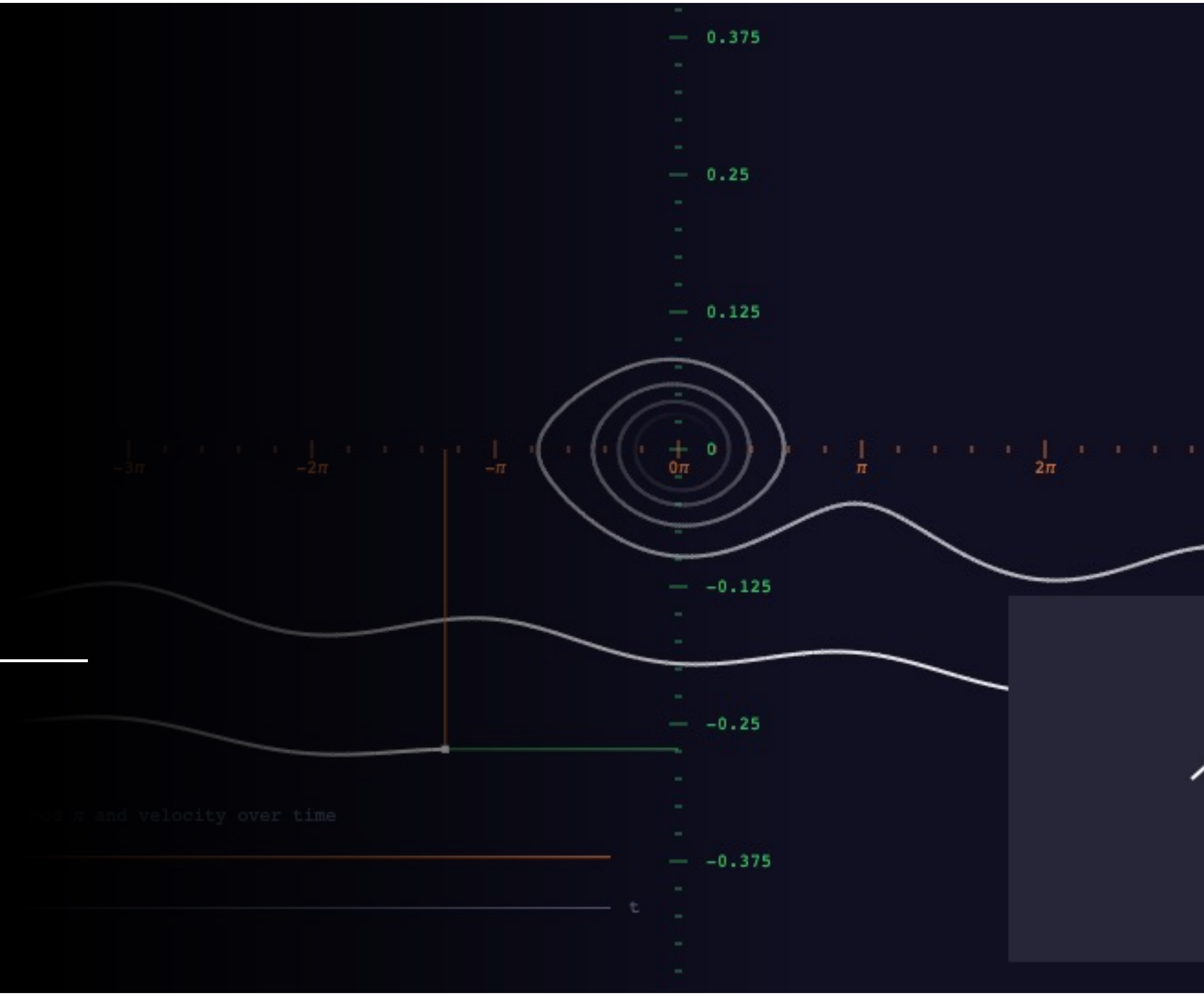
Ptolemy's model (100 AD)



Widget link

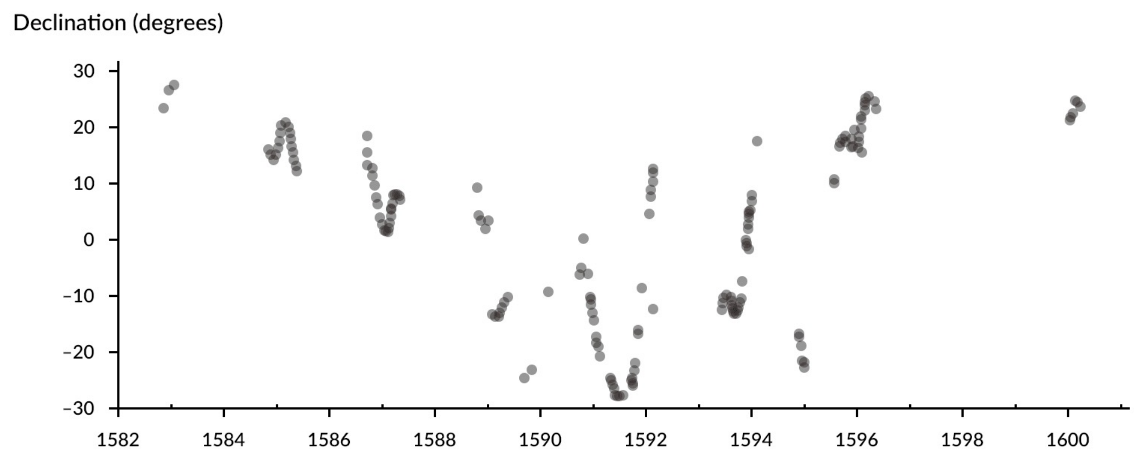
- Solves retrograde problem
- Also explains motion of Mercury more accurately
- However, takes 2x more parameters

Theoretical branches of science



From Greeks to the Enlightenment Age

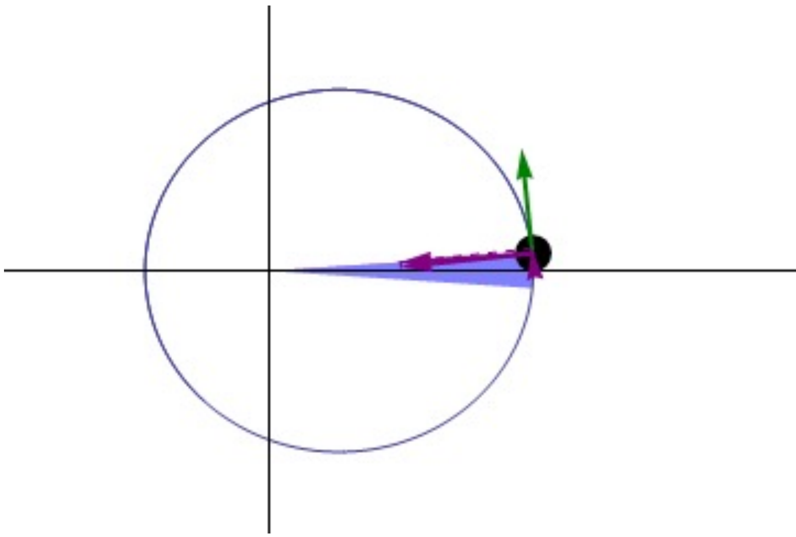
- Medicine
- Astronomy
  - Galileo Galilei, first telescope
  - Nicolaus Copernicus, heliocentric system
  - Tycho Brahe, collected da
  - Johannes Kepler, derived laws of planetary motion from Tycho's data
- Newton's laws



http://www.pafko.com/tycho/observe.html

Kepler’s laws

- The first law states that the orbit of every planet is an ellipse with the Sun at one of the two foci.
- The second law describes the quantitative dependency of the speed of a planet over the position: a line joining a planet and the Sun sweeps out equal areas during equal intervals of time.
- The third law states that the square of the orbital period of a planet is directly proportional to the cube of the semi-major axis of its orbit.



Newton’s laws

- Laws of motion
- Law of gravity

$$F = G\frac{m_1 m_2}{r^2}$$

- pretty compact and easy to grasp;
- pretty hard to get from raw numbers;
- captures the qualitative and quantitative dependencies between state variables and observables;
- can explain the dependencies between the observable effect and some input variables;
- allows to derive other laws;
- gives a way to estimate the future state of the system under observation if the past state of the system is known, or in other words, one can plan an experiment to test this model;

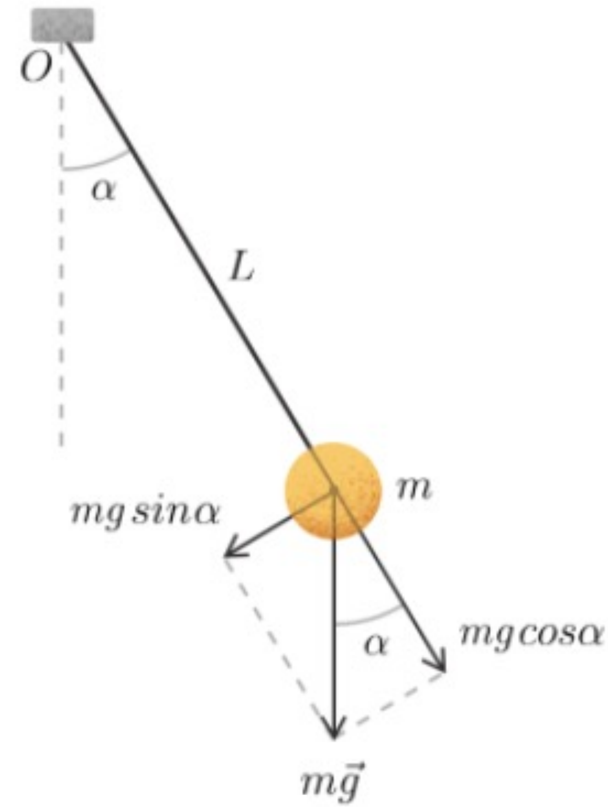
Theoretical branch: differential equations

- Developed by Newton, Leibniz
- At every moment of time t, we can express the dependency of angular acceleration ε

$$\varepsilon = \frac{d^2\alpha}{dt^2} = \frac{M}{I}$$

$$\frac{d^2\alpha}{dt^2} = \frac{mgL\sin\alpha}{mL^2} = -\frac{g\sin\alpha}{L}, \Rightarrow \frac{d^2\alpha}{dt^2} + \frac{g}{L}\sin\alpha = 0$$

$$\frac{d^2\alpha}{dt^2} + \frac{g}{L}\alpha = 0 \text{ or } \frac{d^2\alpha}{dt^2} + \omega^2\alpha = 0, \text{ where } \quad \omega = \sqrt{\frac{g}{L}}$$

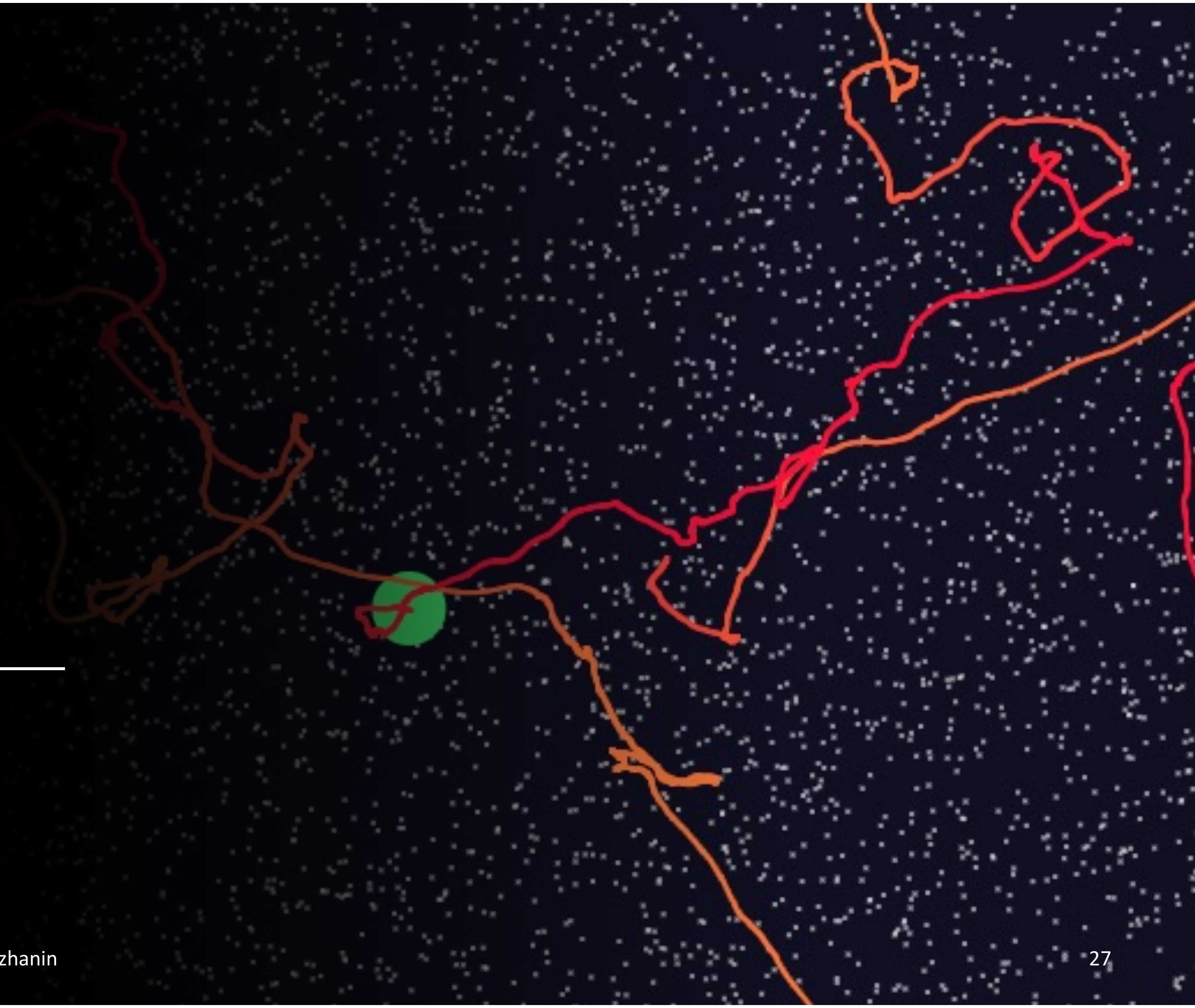


$$T = \frac{2\pi}{\omega} = 2\pi\sqrt{\frac{L}{g}}$$

Scientific method

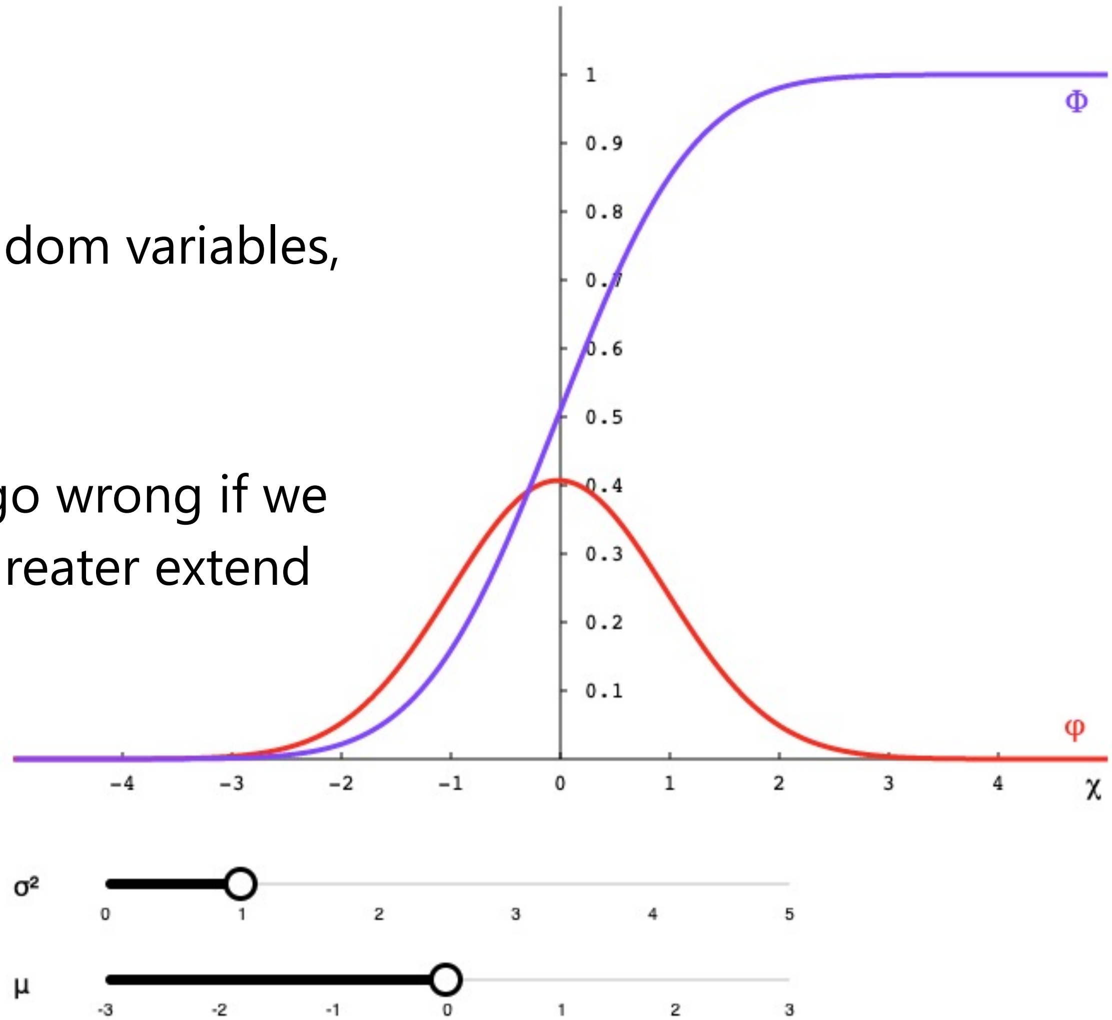
- Define a question
- Gather information and resources (observe)
- Form an explanatory hypothesis
- Test the hypothesis by performing an experiment and collecting data in a reproducible manner
- Analyze the data
- Interpret the data and draw conclusions that serve as a starting point for new hypothesis
- Publish results
- Retest (frequently done by other scientists)

Theoretical sciences and stochasticity



Account for the incomplete world knowledge

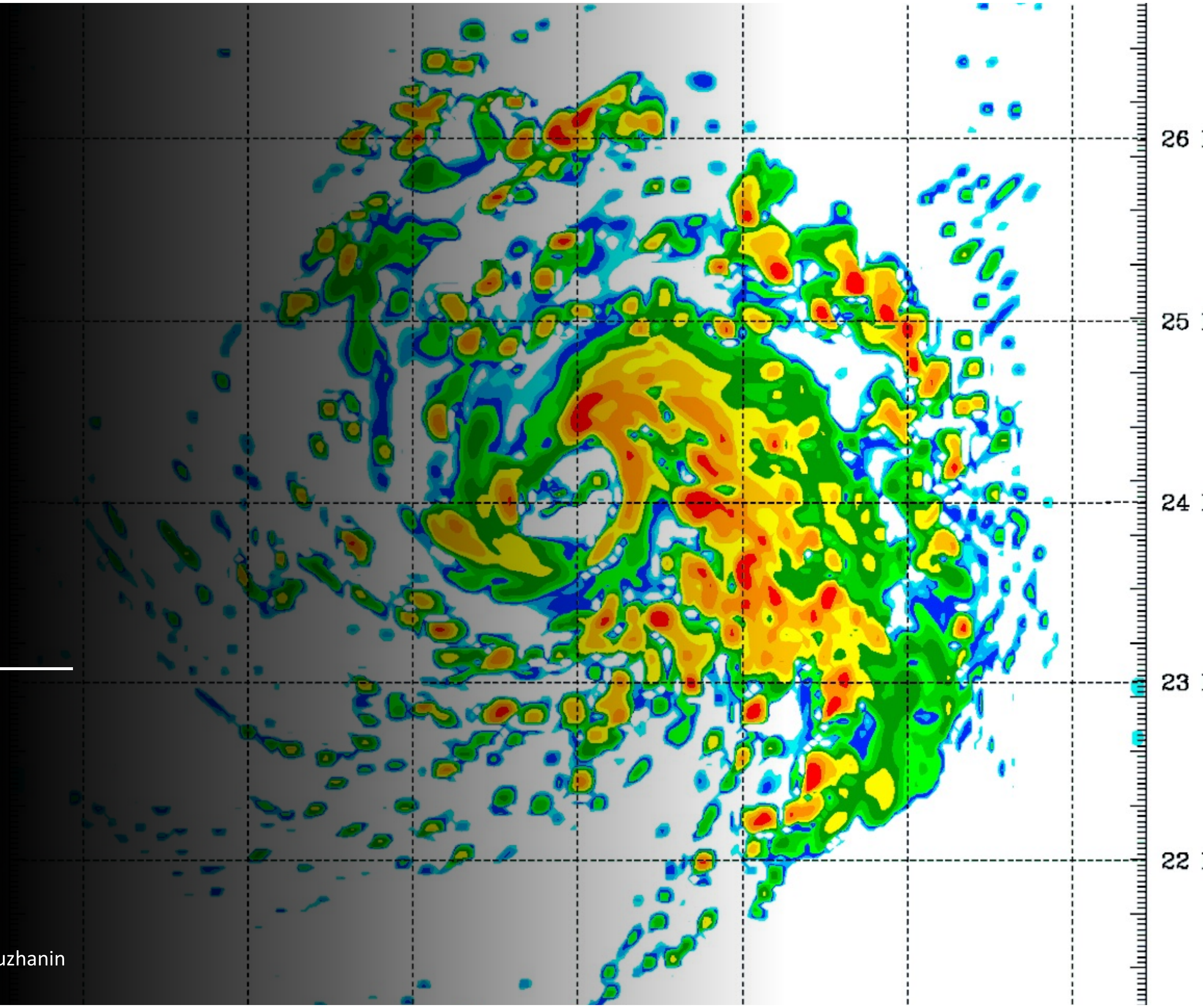
- Probability axiomatics
- Regular variables -> Random variables, distribution families
- Explicit noise term
- Awareness of what can go wrong if we apply reasoning to the greater extend
  - Selection bias
  - Sampling variability
  - Model bias
- Statistical inference



Statistical inference

- Two main types of experimental measurement:
  - Tests of a theory/model: **hypothesis testing**
  - Measurement of a quantity: **parameter estimation**
- For **parameter estimation** we usually have some data (a set of measurements) from which we want to obtain:
  - The best estimate of the true parameter; **"the measured value"**
  - The best estimate of how well we have measured the parameter; **"the uncertainty"**
- The goal of inference is to be able to make a statement about latent variable, and ideally to be able to characterize any uncertainty you have about that statement.

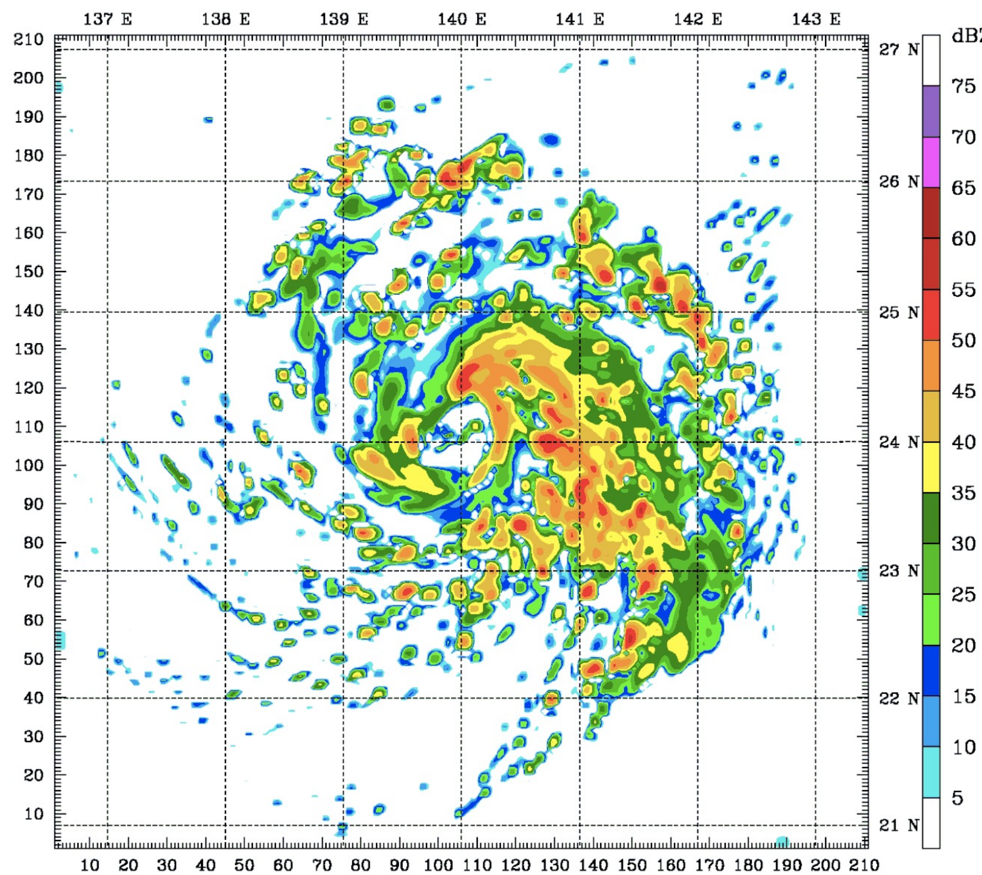
Computer Simulation



Andrey Ustyuzhanin

Computational branch: computer simulation

- Computes the evolution of mathematical models using machines;
- Especially useful when closed-form solution is not available
  - weather forecasting, earth simulator, flight simulator, molecular protein folding, and so on.
- Requires special math methods;
- Blooms with computing power availability.



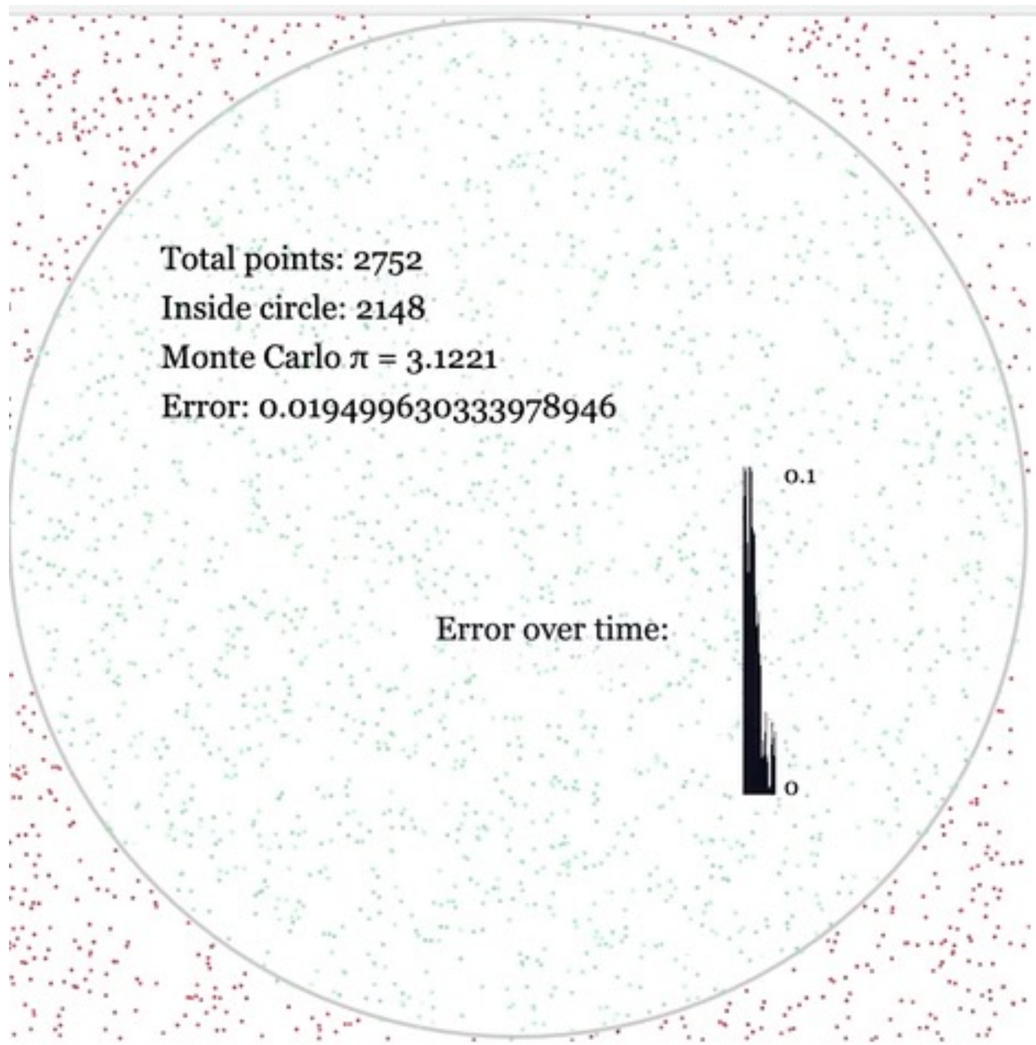
Monte Carlo Method

- Instead of accurately computing all the outcome probabilities, one can combine the randomness from different sources to replicate the overall system dynamics
- E.g., estimate the π constant:

```
import numpy as np

def pi_MC(n):
 assert n > 0, "argument should be positive"
 x = np.random.rand(n)
 y = np.random.rand(n)
 n_c = np.count_nonzero(x**2 + y**2 <= 1)
 return 4 * n_c / n

print pi_MC(1000)
```



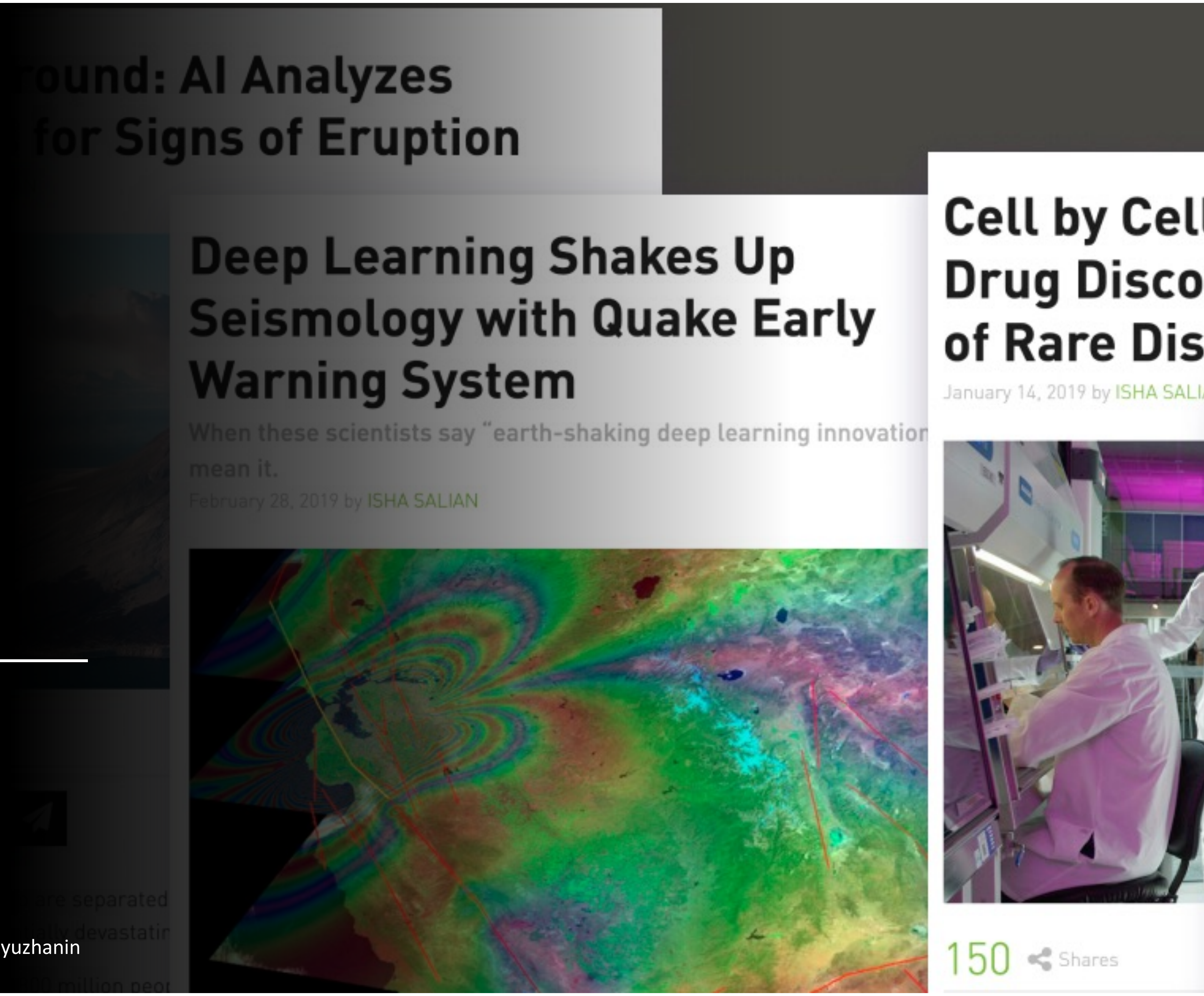
https://en.pelican.study/static/bundles/demonstrations/pi/index.html

Forward and Inverse problems

- Forward: from given initial system parameters, get the observable state
- Inverse: from the observable state, get hidden parameters
  - No single solution
  - No straightforward way to compute
  - But if one can approximate evolution of a system by some differentiable surrogate, it might profit from methods of Machine Learning
  - Systems for probabilistic programming: Stan, PyMC3, pyro, Tensorflow Probability (ex Edward) or pyprob.

Data-driven science

Andrey Ustyuzhanin



Main boosting factors

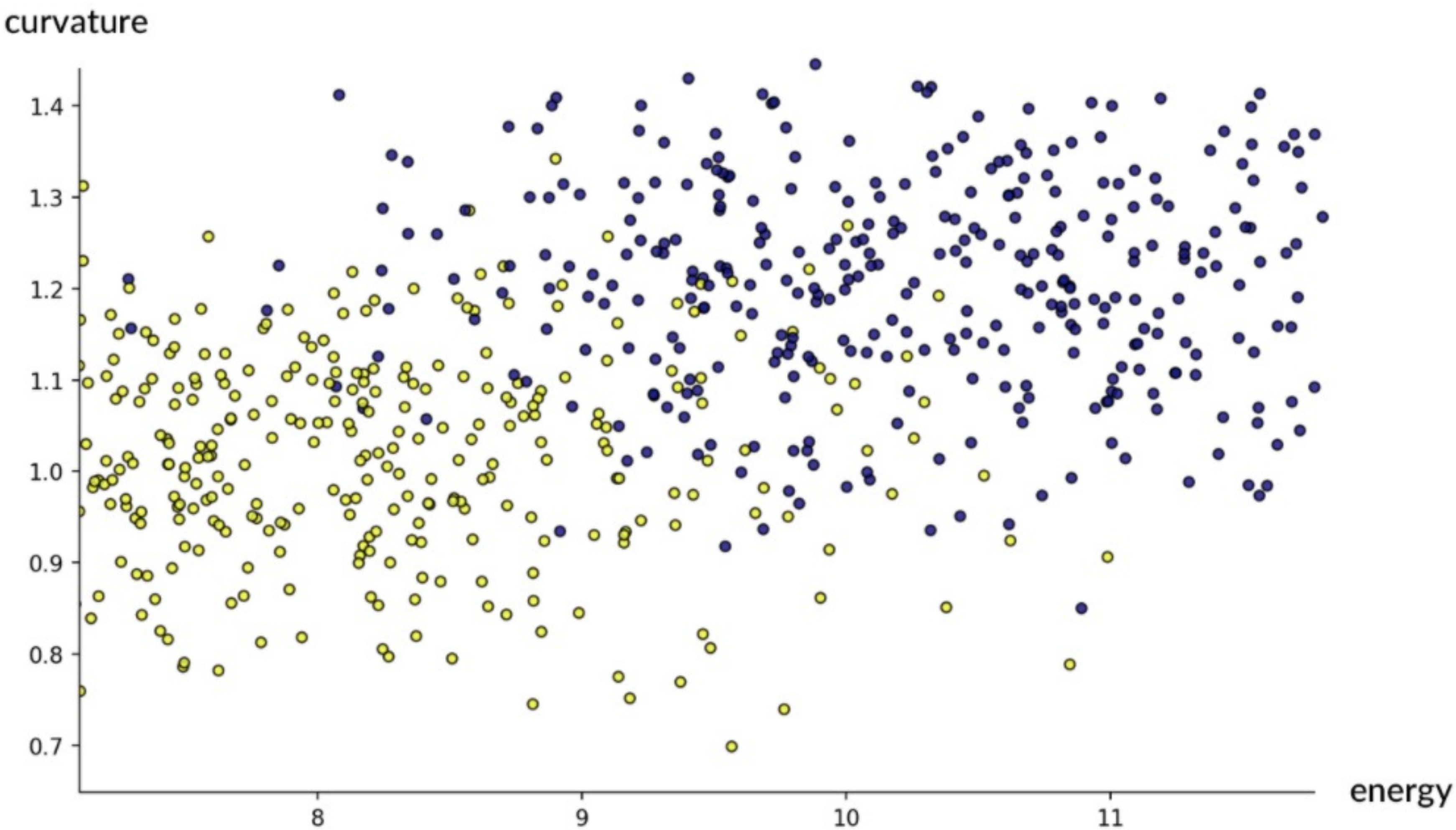
- Data deluge
  - Experiments
  - Industry
  - Simulation
- Computational power
  - Moore’s law
- Sophisticated (meta—level) algorithms

A Dataset for particle classification

- Data sample
  - Features
  - Labels
- Features
  - Type: **float**, integer, …
  - Distributions
- Labels, given from outside
  - **Boolean**, integers, real vectors, …

	Energy	Curvature	Class label
0	10.122077	1.226283	0.0
1	7.760199	1.062012	1.0
2	10.989290	0.906222	1.0
3	8.759292	1.096391	1.0
4	10.778759	1.182548	0.0

Two types of particles: muons and electrons



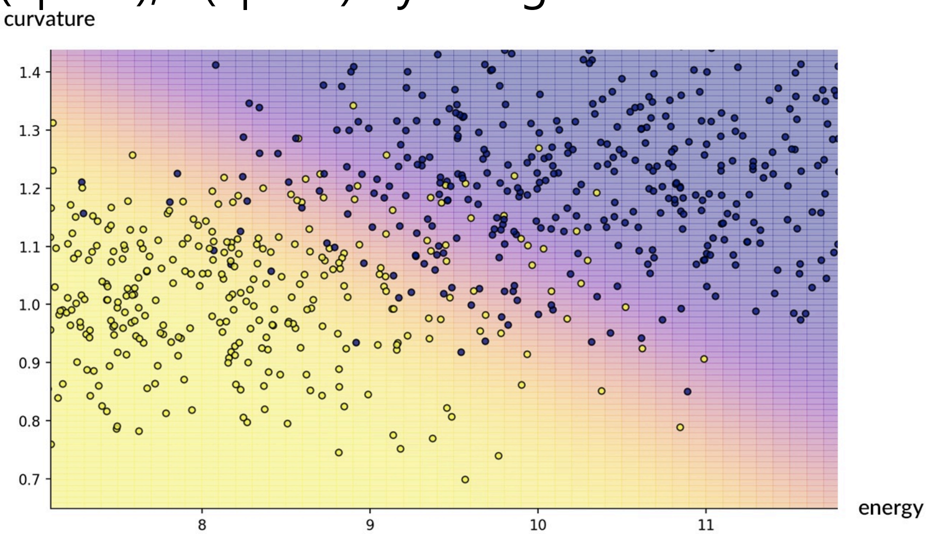
Naïve Bayes classifier

- A Naive Bayes Classifier uses the Bayes' Theorem and connects posterior probability of particle to belong to class C (either 0 or 1) given feature vector F (P(ClF)) with
  - likelihood of feature vector F given class C (P(FlC)). It comes from some simple (*naïve*) assumptions, e.g., features are independent from each other and can be represented by simple N-dimensional Gaussian distribution,
  - prior probability of particle to belong to the class (P(C)) and
  - probability of specific vector feature P(F)

$$P(C|F) = \frac{P(F|C)P(C)}{P(F)}$$

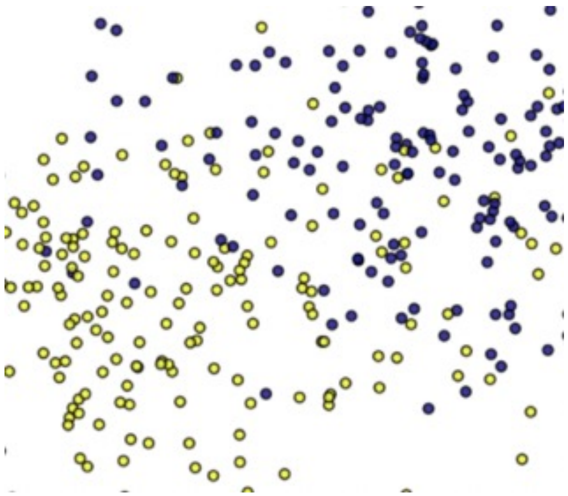
Training an NBC

- Get data sample with features and labels, split it into training and testing;
- Chose P(F|C) distribution type, described by some parameters for each class;
  - E.g., 2D Gaussian distribution with 4 parameters: μ1, μ2, σ1, σ2
- Find the parameter values that would best fit the distribution into the training sample: build P(F|C=0), P(F|C=1) by using *maximum likelihood approach*

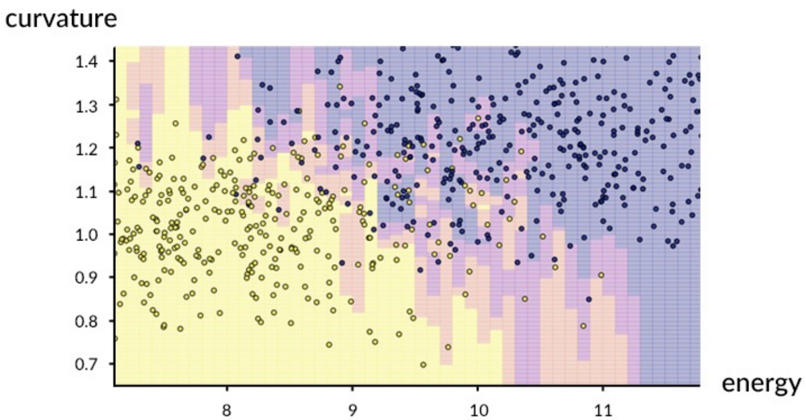


K-Nearest neighbors

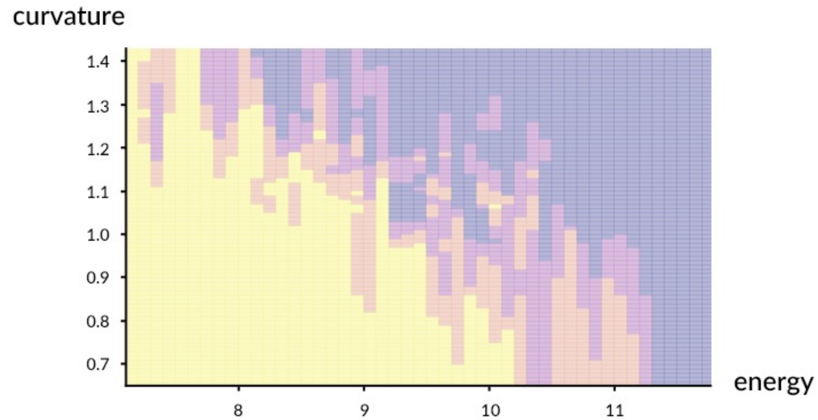
- K-nearest neighbors (or k−NN for short) is a classification algorithm that decides by
  - For every given feature vector F representing unknown particle
  - k−NN looks at the K nearest neighbors from the training sample and
  - counts the fraction of electrons and muons among the neighbors.



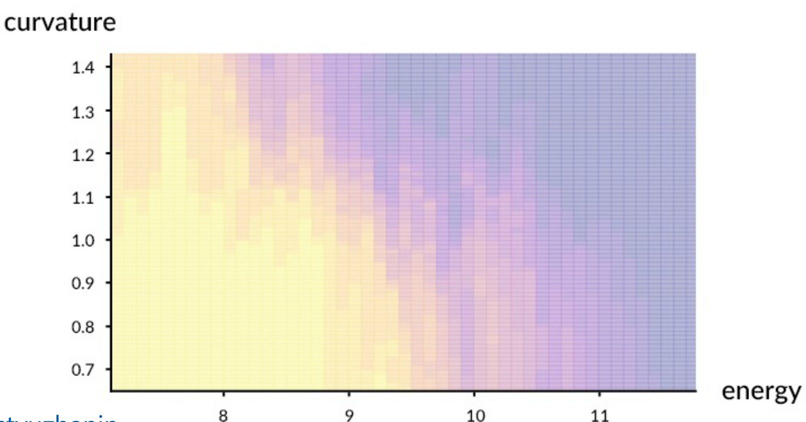
K-Nearest Neighbors (K = 3)



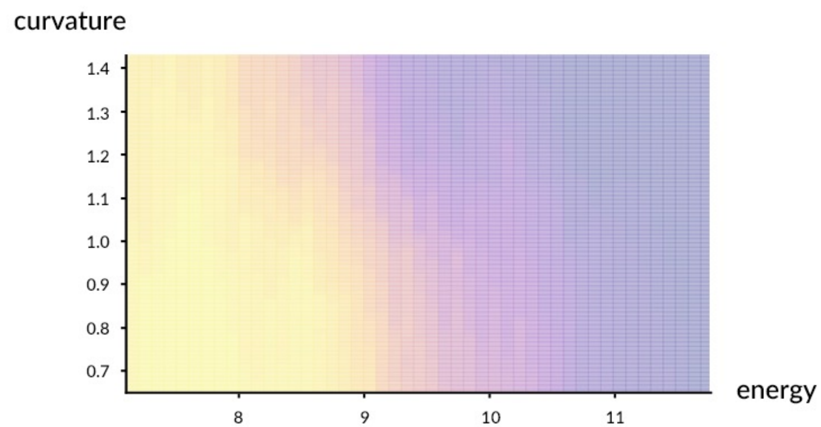
K-Nearest Neighbors (K = 3)



K-Nearest Neighbors (K = 15)



K-Nearest Neighbors (K = 50)



Taxonomies of machine learning algorithms

- By the information needed to make decision:
  - Supervised - uses labelled sample for training;
  - Unsupervised - exploits distance between different objects and doesn't need labels;
  - Reinforced learning - the ground truth becomes available to the algorithm during interaction with environment. Examples include playing computer games, controlling robots or more complicated mechanisms.
- By the type of decision:
  - Classification - attributing an object to one of given classes;
  - Regression - estimation of real number (or vector) from given features;
  - Policy search - building a control policy, e.g. strategy of moves for a game or actuator control;
  - Segmentation - selection of a region belonging to specific class object inside input data;
  - Generative - learn how to generate objects of certain kind, e.g. images of human faces or cars or even painting style.

Taxonomies of machine learning algorithms (2)

By the type of input objects (features) an algorithm can deal with:

- tabular representation;
- 2D or 3D image;
- text;
- time series;
- graphical data.

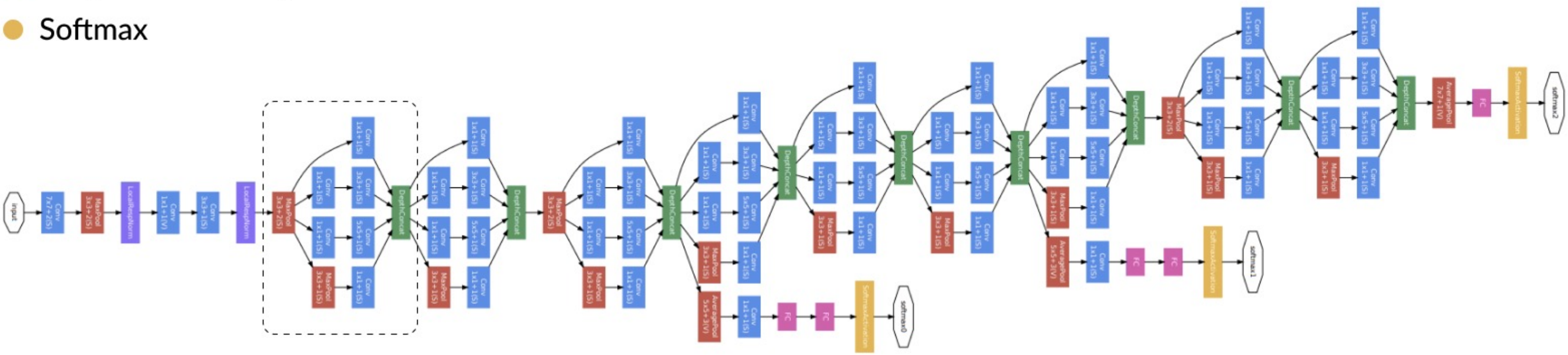
For example, Naive Bayes algorithm is supervised classification algorithm that easily deals with tabular data. Interestingly to note that the algorithm builds a generative model $P(F|C)$ behind the scenes for estimation of likelihood for different classes.

Mache Learning research

- Relies on the developed background,
- Works best in cooperation with domain science expertise,
- Meta-reasoning,
- Embeds learning patterns into a trainable algorithm
  - Convolutional neural network
  - Langevin Gradient Descent
  - Generative Adversarial Network
  - Neural [Ordinary] Differential Equations
  - Monte Carlo Tree Search
  - …and many others

Deep Learning

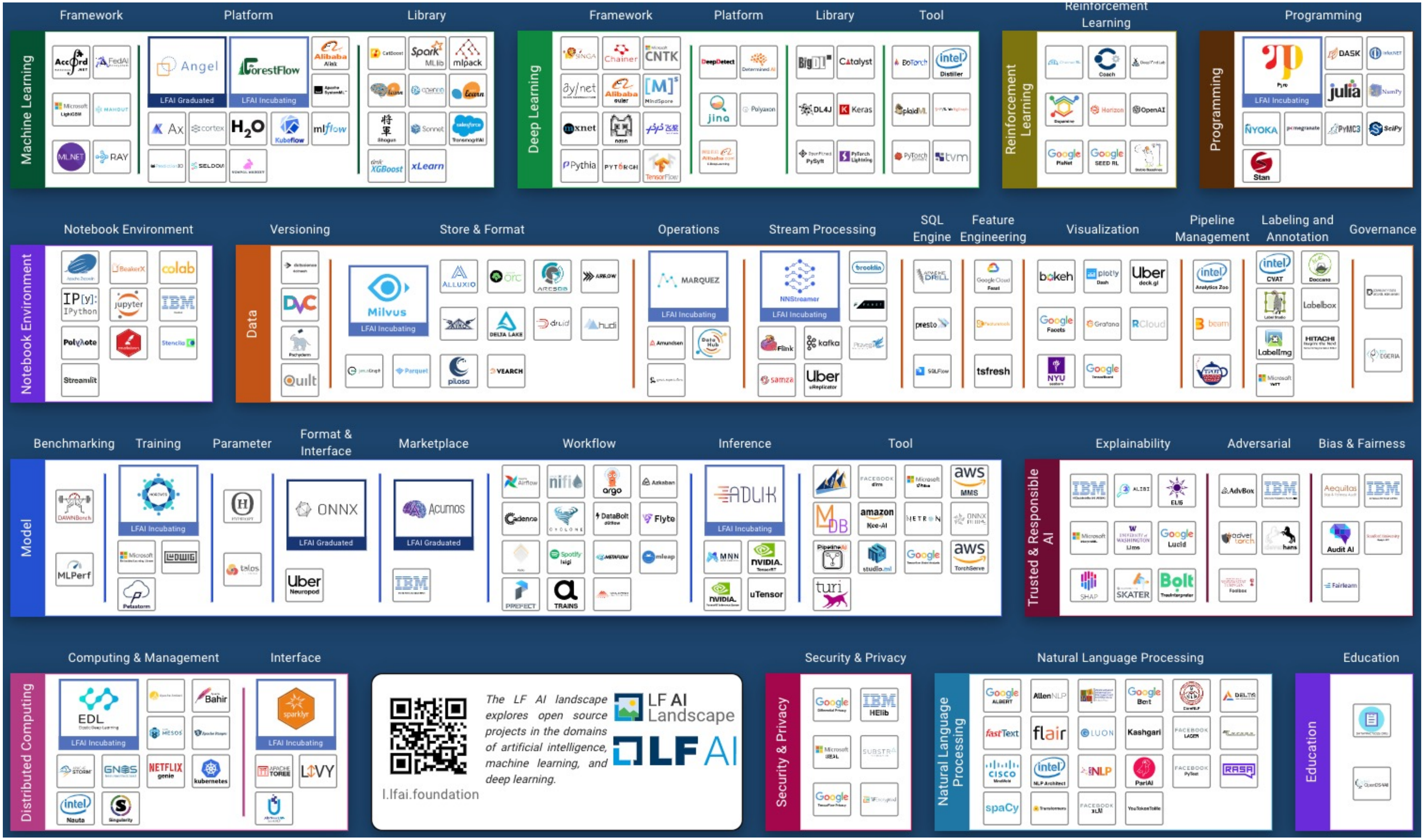
- Convolution
- Max pooling
- Channel concatenation
- Channel-wise normalization
- Fully-connected layer
- Softmax



Input　　Inception cell　　Auxiliary loss　　Auxiliary loss　　Final output

InceptionNet, 5 million parameters

Link

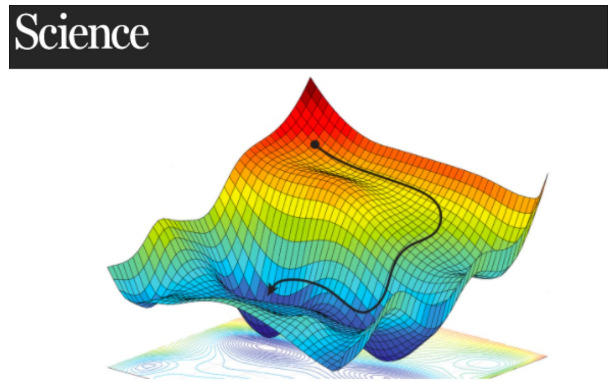


Machine Learning
Framework
Platform
Library
Deep Learning
Framework
Platform
Library
Tool
Reinforcement Learning
Programming
Notebook Environment
Data
Versioning
Store & Format
Operations
Stream Processing
SQL Engine
Feature Engineering
Visualization
Pipeline Management
Labeling and Annotation
Governance
Model
Benchmarking
Training
Parameter
Format & Interface
Marketplace
Workflow
Inference
Tool
Trusted & Responsible AI
Explainability
Adversarial
Bias & Fairness
Distributed Computing
Computing & Management
Interface
Security & Privacy
Natural Language Processing
Education
The LF AI landscape explores open source projects in the domains of artificial intelligence, machine learning, and deep learning.
LF AI Landscape
LF AI
l.lfai.foundation

Missing piece: Theory of Deep Learning

Deep Learning: Alchemy or Science?

https://www.ias.edu/ideas/videos-deep-learning-alchemy



Gradient descent relies on trial and error to optimize an algorithm, aiming for minima in a 3D landscape. ALEXANDER AMINI, DANIELA RUS, MASSACHUSETTS INSTITUTE OF TECHNOLOGY, ADAPTED BY M. ATAROD/SCIENCE

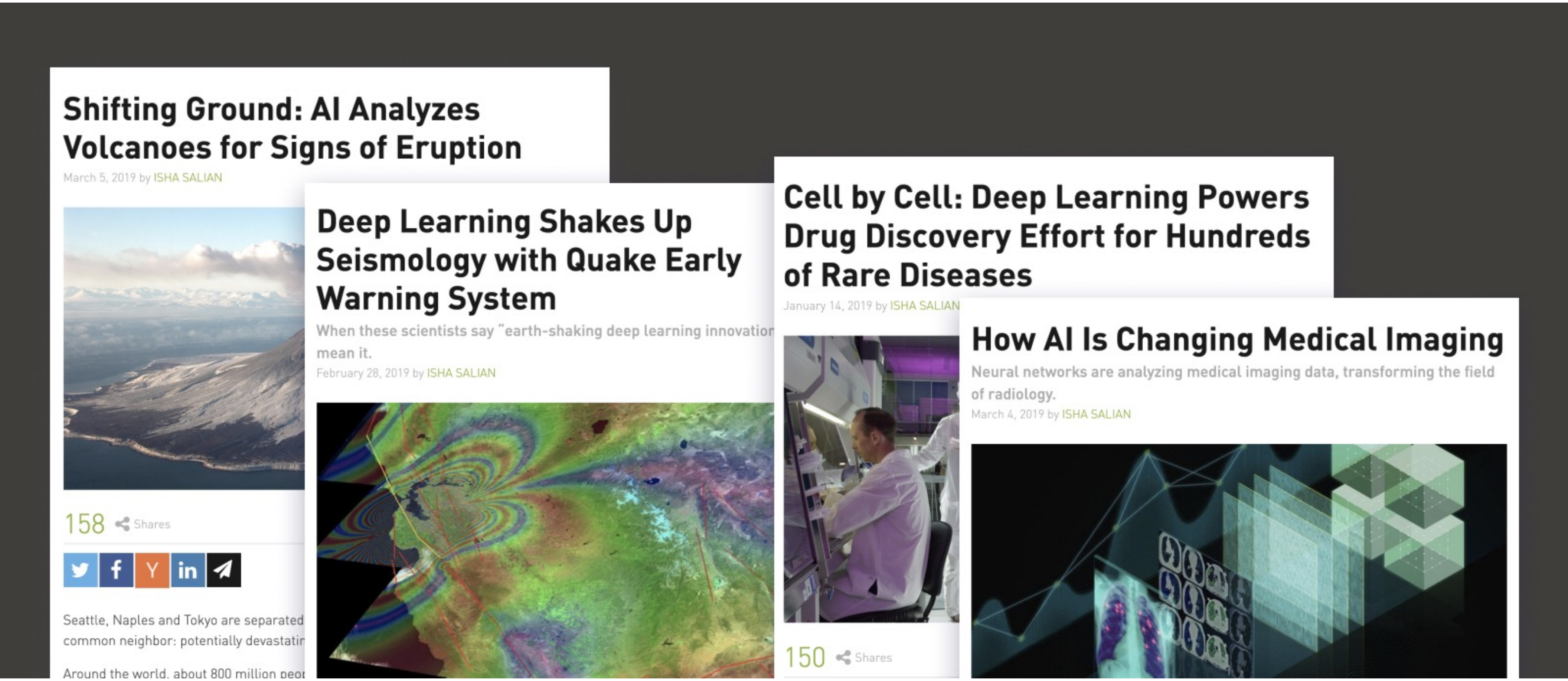
AI researchers allege that machine learning is alchemy

https://www.sciencemag.org/news/2018/05/ai-researchers-allege-machine-learning-alchemy

"People gravitate around cargo-cult practices, relying on folklore and magic spells". Francoise Collet

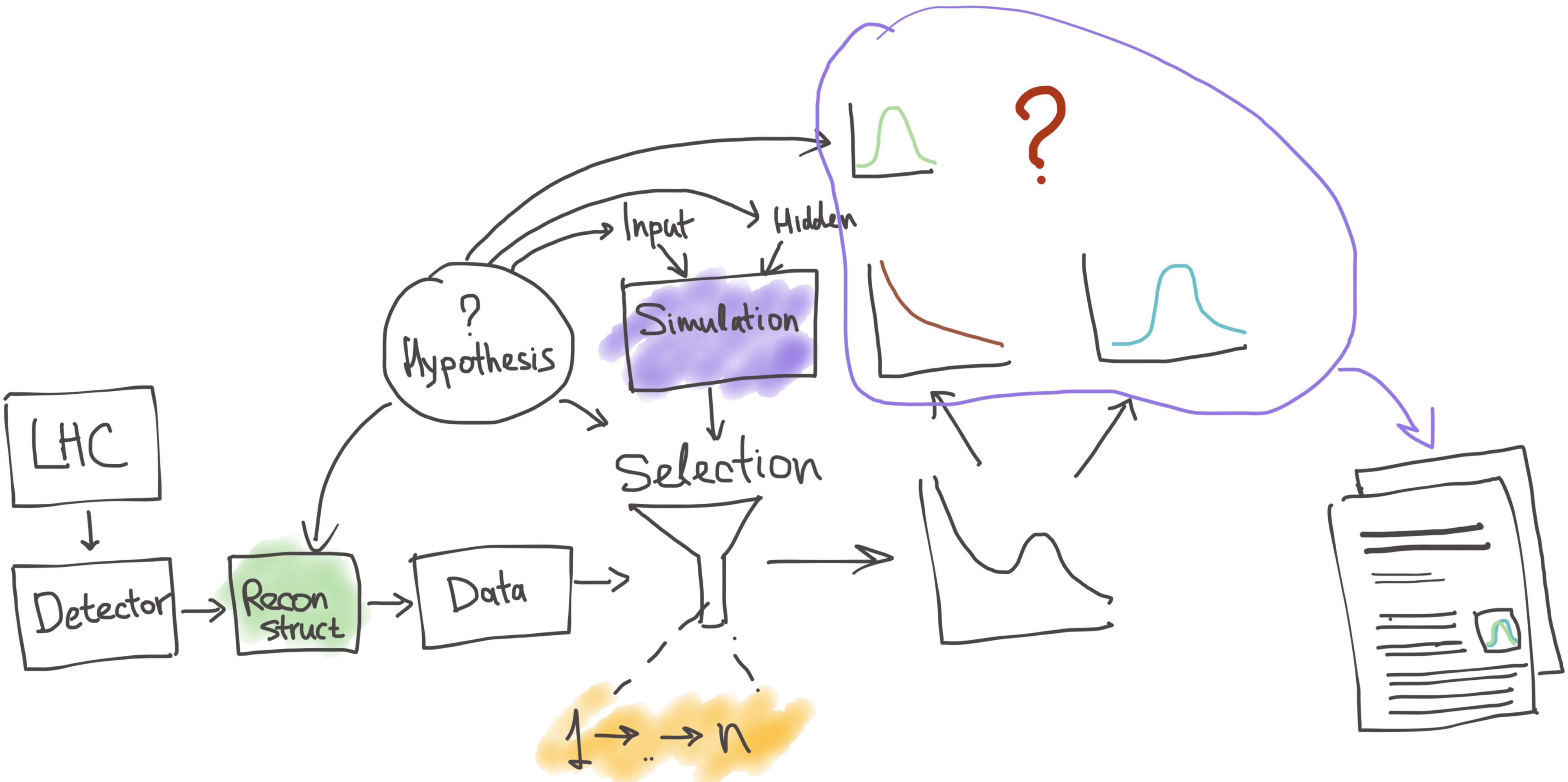
There is no solid theory that would predict properties of a given architecture for a given dataset, which makes applied deep learning pretty much like engineering discipline with heuristics and plenty of trial-and-error activities.

Deep Learning Applications



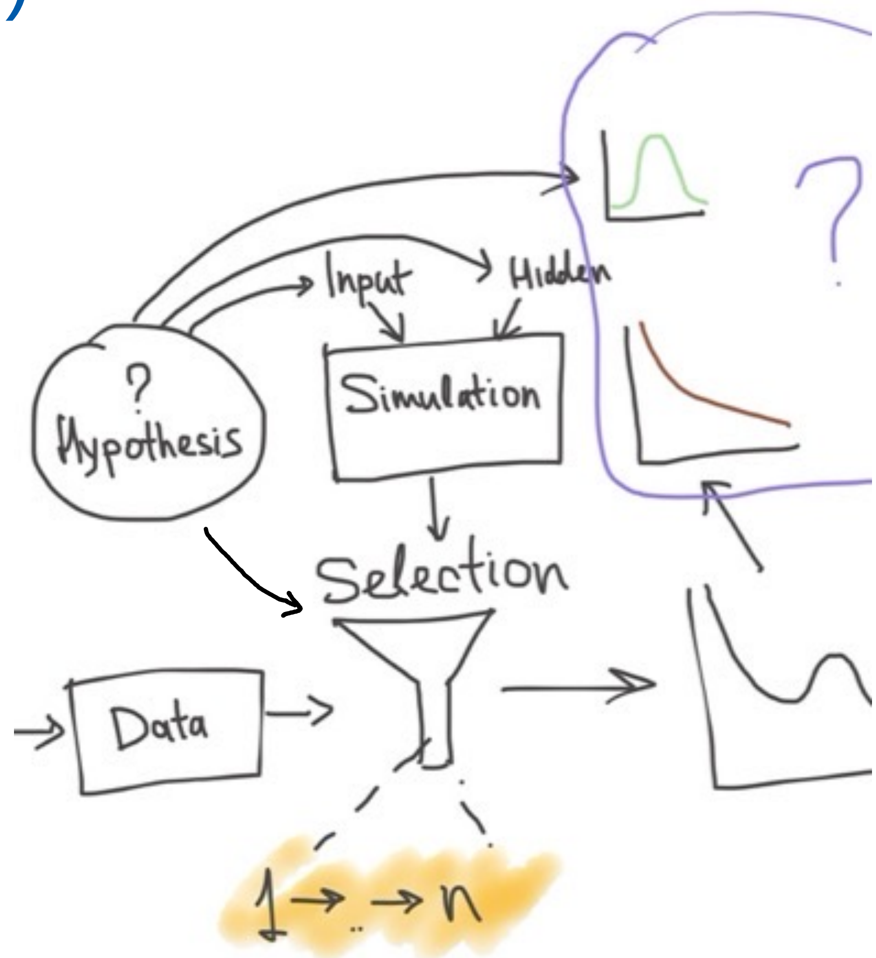

Large Hadron Collider
CMS
p
p
LHCb
ATLAS
ALICE

Generic (simplified) discovery plan



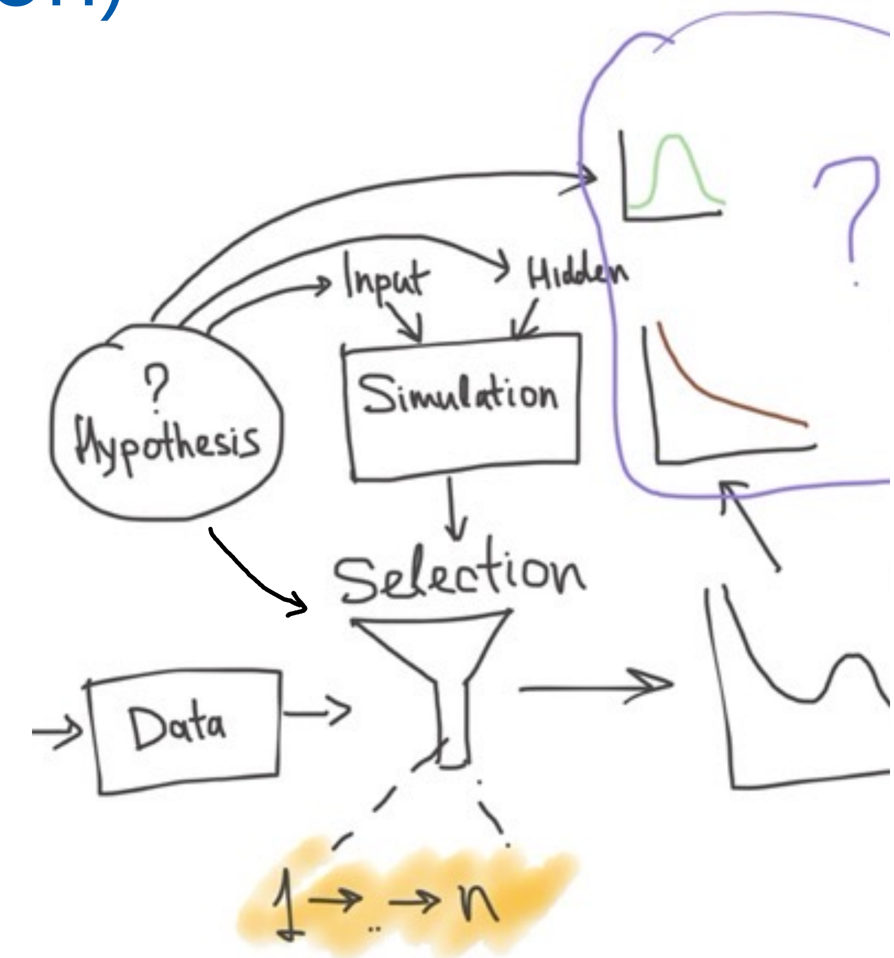
MI challenges – I (selection)

- Optimize selection given data, hypothesis to **maximize sensitivity**
  - Triggers, particle or jets identification, etc.
- Optimize selection given data, hypothesis to **maximize sensitivity and minimize model-induced bias**
- Optimize selection given data, null hypothesis to **maximize unexpected (unexplainable) signal yield**



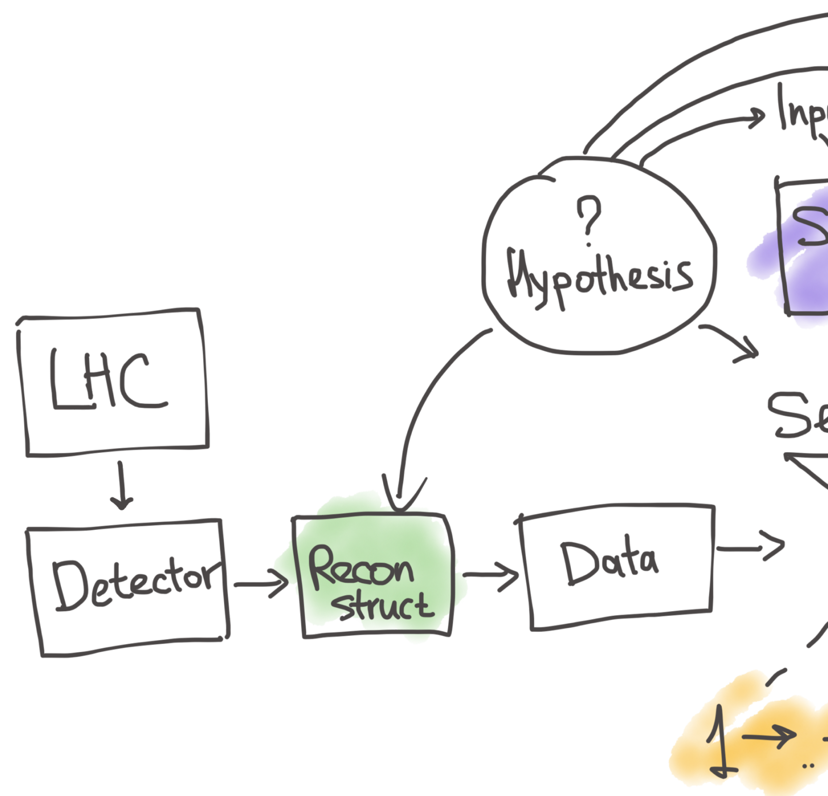
MI challenges – II (simulation)

- Optimize simulation parameters given data, hypothesis to **minimize difference**
  - *What is the difference?*
- Optimize simulation surrogate given data, hypothesis to **minimize difference** and **speed**
- Make simulation invertible: Can we learn a mapping from observables to simulation input under different physics hypothesis to infer its likelihood given data, while **minimizing the complexity** and **difference between real and simulated data**



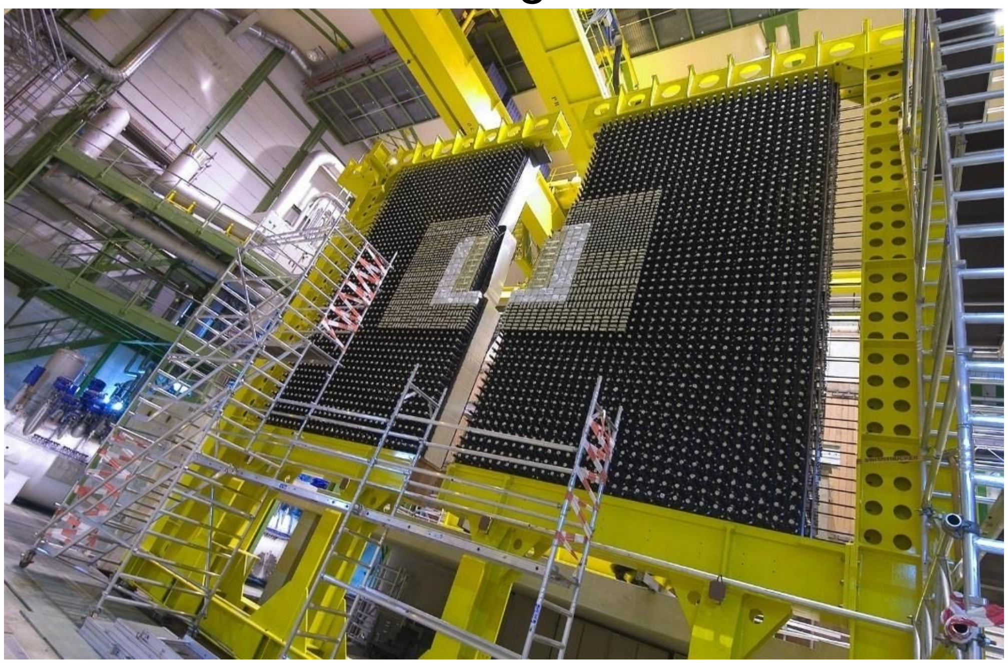
MI challenges – III+

- Optimize detector (and LHC as well?) given budget, physics laws and new physics expectations to **maximize signal yield**
  - Can it be solved recursively?
  - Implies solving challenges I and II
- Optimize set of physics laws given knowledge collected so far and agent cognitive capabilities to **minimize complexity** of the laws for the agent

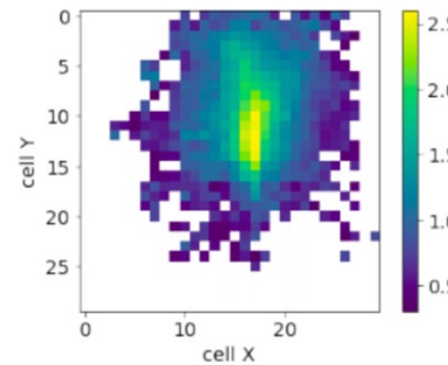


LHCb Electronic Calorimeter (ECAL) @ CERN

Current configuration



Size: 7.8x6.3x0.5 m

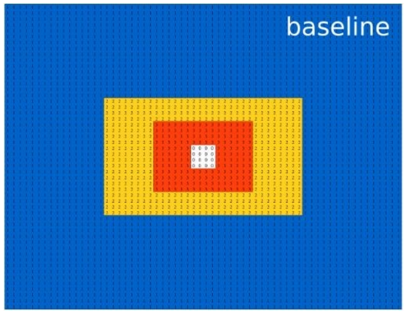


Module size 12x12 cm^2

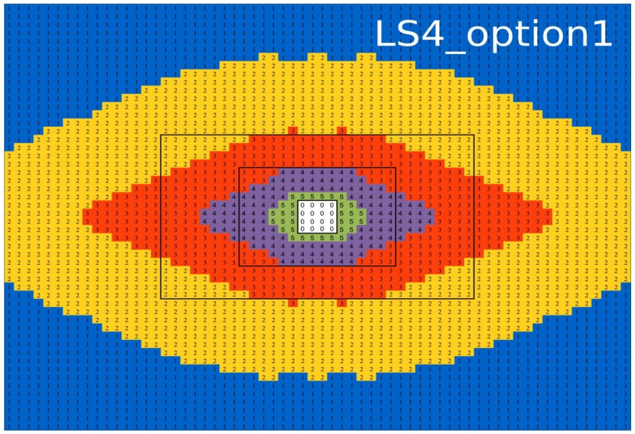
176 inner modules: 9 cells with size 4x4 cm^2
448 middle modules: 4 cells with size 6x6 cm^2
2688 outer modules: 1 cell with size 12x12 cm^2

Optimization Motivation

- What is the best configuration for given modules (fix cost) in terms of given physics metric?
- What is the best way to arrange a certain number of new modules?



Module type	# of modules
(inner): 3x3 cells (4.04×4.04 cm^2 each)	176 (1536 ch.)
(middle): 2x2 cells (6.06×6.06 cm^2 each)	448 (1792 ch.)
(outer): single cell (12.12×12.12 cm^2)	2688 (2688 ch.)

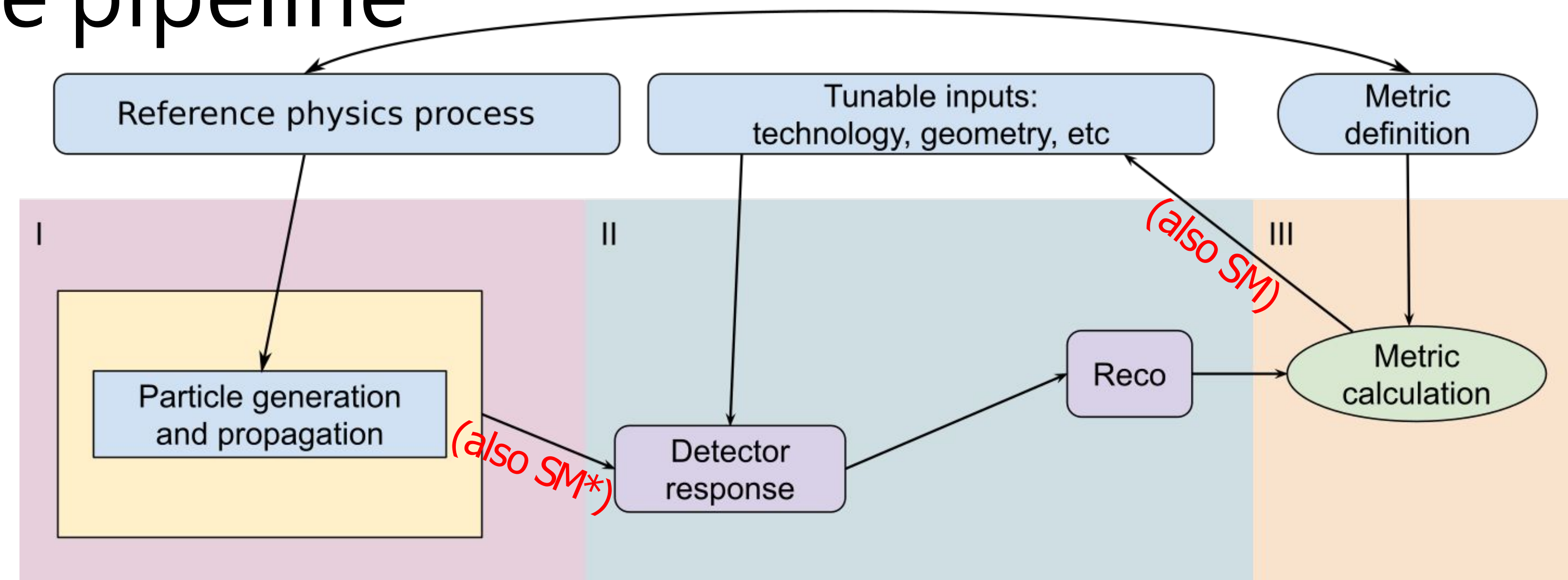


New "Spacal" modules

- 4 : cell size = 3.03×3.03 cm^2
- 5 : cell size = 1.515×1.515 cm^2

The pipeline

*SM = Surrogate Models



Optimisation cycle itself does not depend on the modules technology & arrangement, reconstruction, metric, etc.

In our realisation of ECAL simulation:

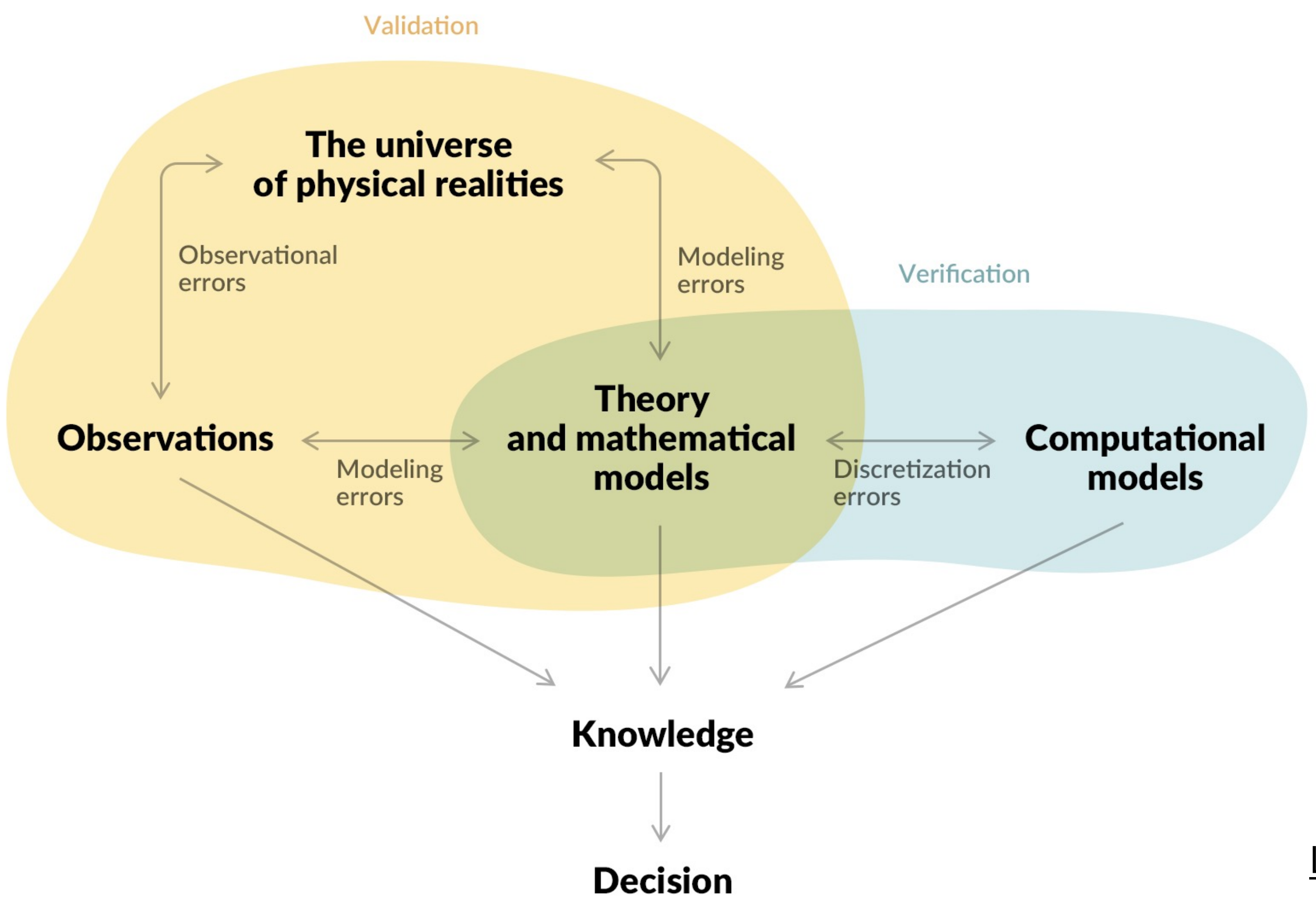
- Step I requires 250 core hours (once). Realised in Docker container.
- Steps II-III require **20 core hours per option.** Realised in vectorised python or in Docker.
  Sounds far too much... There is an option to significantly reduce the number of calls of the simulation

Modern MI Challenges in Physics

- Can MI help with computationally costly problems, like simulation or the combinatorial challenge?
- Can fast O(ns-μs) MI inference be done with FPGAs to put MI early in the trigger / data acquisition process?
- Can we make MI models robust to data change?
- Can we build more efficient machines with the help of MI?
- Can we encode physics-motivated reasoning into MI computations?
- How can we make the best use of simulation engine for inference (of latent variables) given that observation likelihood is intractable?
- Can ML/MI close the loop : Experiment → hypothesis → theory → experiment?

https://nature.com/articles/s41586-018-0361-2 https://doi.org/10.1073/pnas.1912789117

Wholistic picture



Link

Conclusion

- Data-driven approach is the new paradigm of scientific research, closely related to MI technologies
- ML transforms ways of reasoning into usable tools (computational logic, statistical inference, computational reasoning, etc.)
- MI is based on ML/DL and is grounded in all known scientific paradigms
  - Empirical, theoretical, probabilistic and computational
- ML is the unified language for solving problems for each of the paradigms, also for unifying those in a wholistic system
- Biggest challenges:
  - Define the theory of Deep Learning
  - Close the loop: Experiment -> hypothesis -> theory -> experiment

Thank you!

austyuzhanin@hse.ru

anaderiRu

hse_lambda

Andrey Ustyuzhanin

Backup

Physics-inspired approaches in ML

- Simulated Annealing
- MCMC techniques
- Gibbs sampling
- Gaussian process
- Gradient descent
- Boltzmann Machine
- Energy-based GANs

Arxiv:1903.10563