

**ETH** zürich

**EPFL**

PAUL SCHERRER INSTITUT

**PSI**

 **Empa**

Materials Science and Technology



**eawag**  
aquatic research 000

 **SDSC**

# LESSONS LEARNED IN DATA-DRIVEN SCIENCE

[www.datascience.ch](http://www.datascience.ch)  
@SDSCdatascience

# The Swiss Data Science Center



## Accelerate the adoption of data science and AI in Switzerland

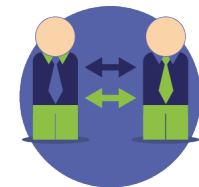
- An initiative from the ETH Domain, started in 2017
- Offices in Zurich and Lausanne
- Academic and industry collaborations



Academic  
projects

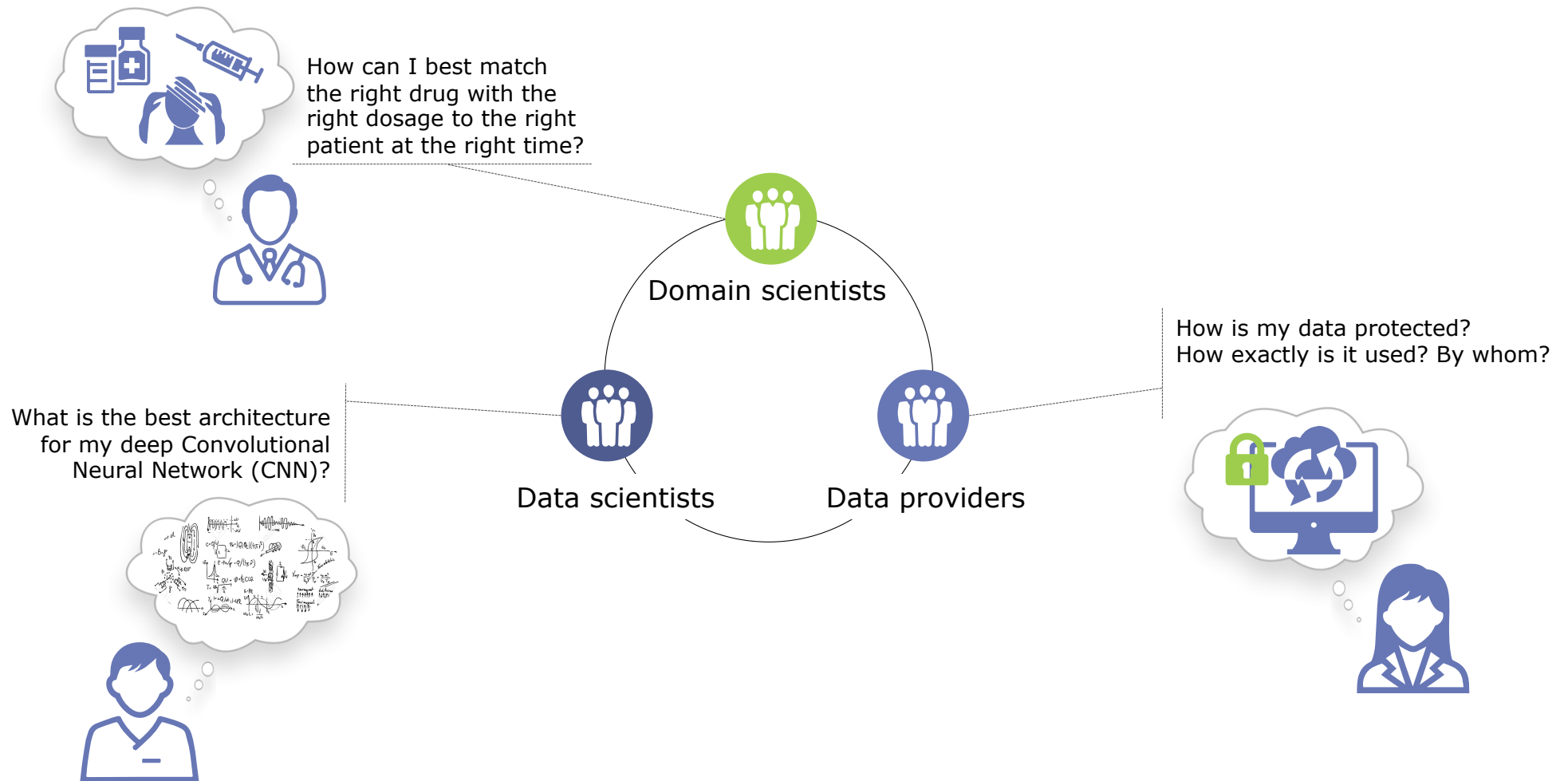


Industry  
collaborations



RENKU  
platform

# Closing the gaps in the data science journey



# Data is everywhere



# A fantastic source of data



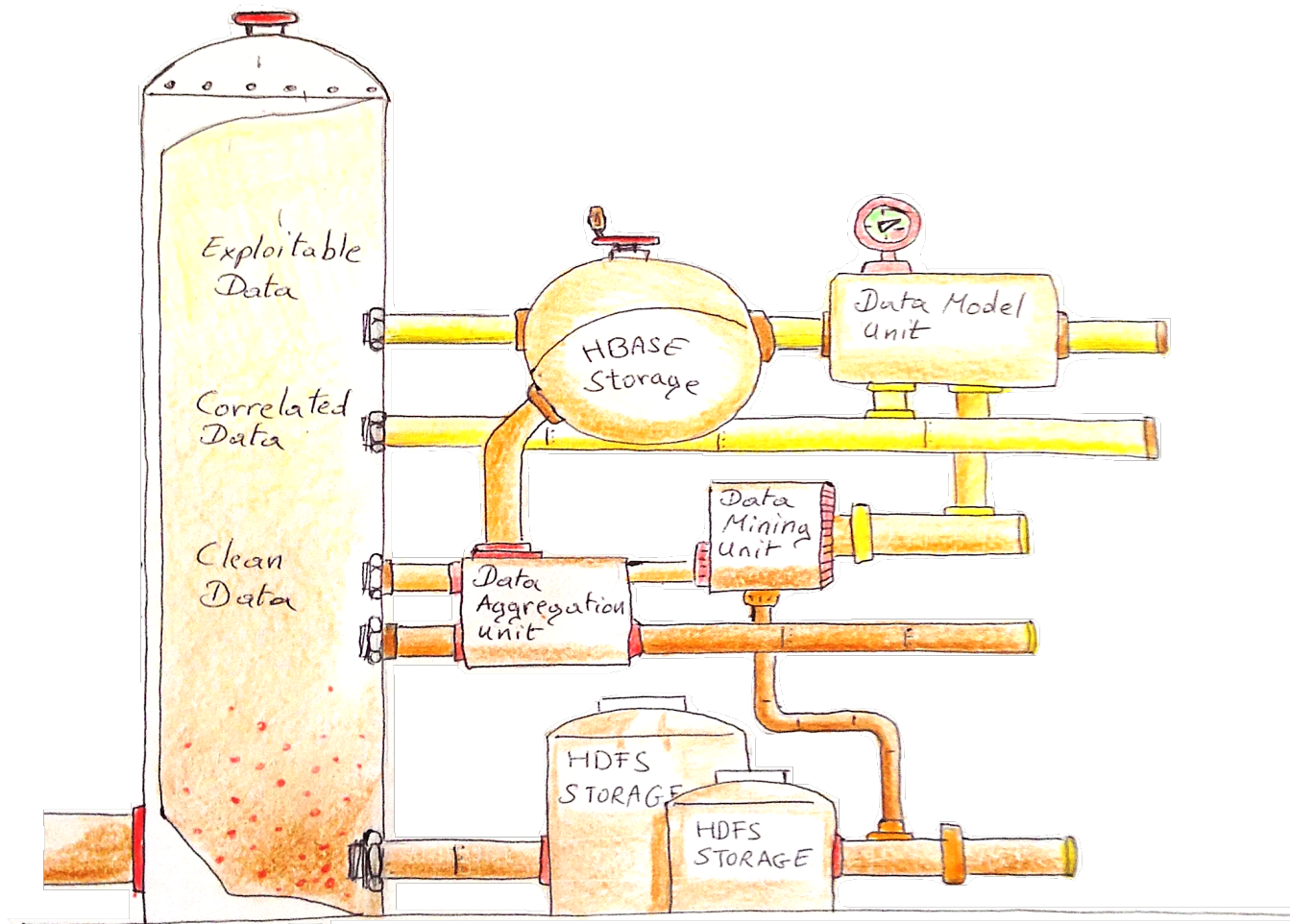
## AI needs better data, not just more data



"YEP... GOT MY CELLPHONE, MY PAGER, MY INTERNET LINK, MY WIRELESS FAX, AND THANKS TO THIS NIFTY SATELLITE NAVIGATION SYSTEM, I KNOW PRECISELY WHERE I AM AT ALL TIMES!"

BY LOWE FOR THE SUN-SENTINEL, FLO

# From raw data to unbiased information



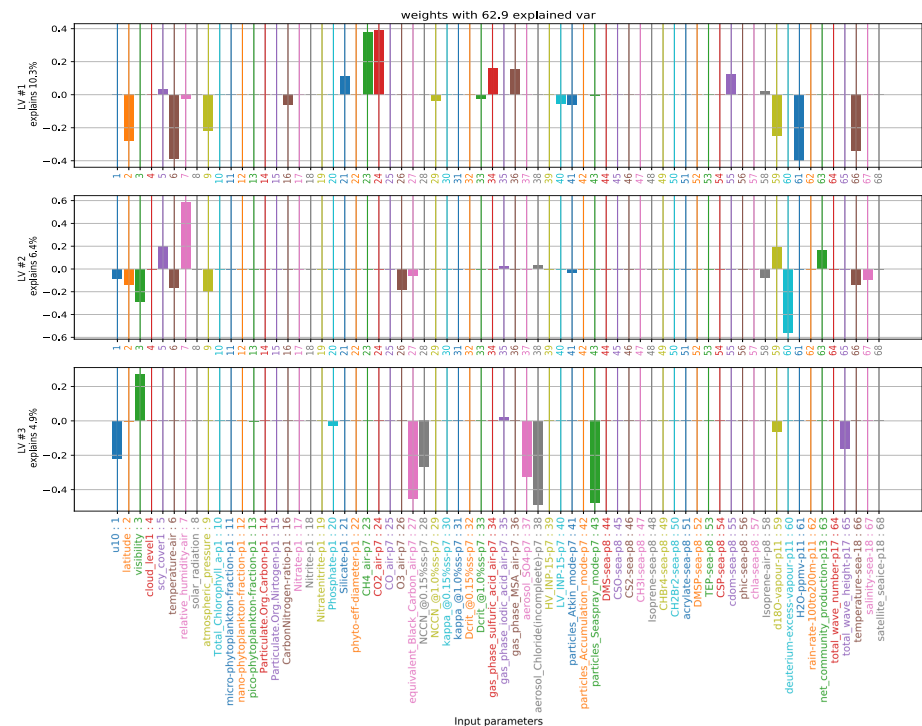
# Antarctic Circumnavigation Expedition

Expedition boat with 22 teams from South Africa to Australia to Chile in 90 days

**Context:** Foster collaboration between teams of scientists, breaking data silos

**Initial problem:** Model relationships between ocean / wave parameters and aerosols

EPFL ETH zürich






# The Data-driven Science journey

Data + Algorithms → Knowledge → Benefits

*Big Data /  
Data lake*

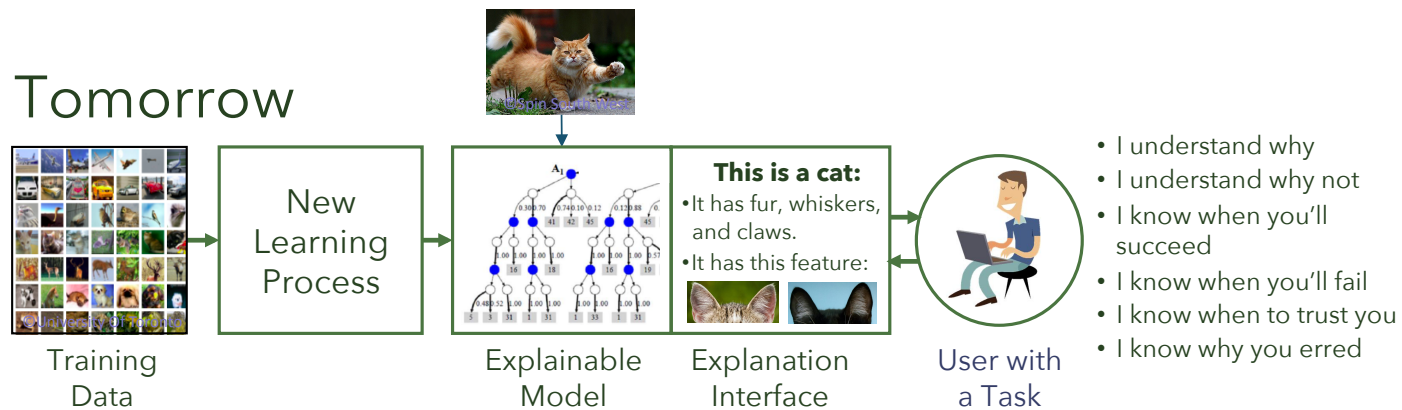
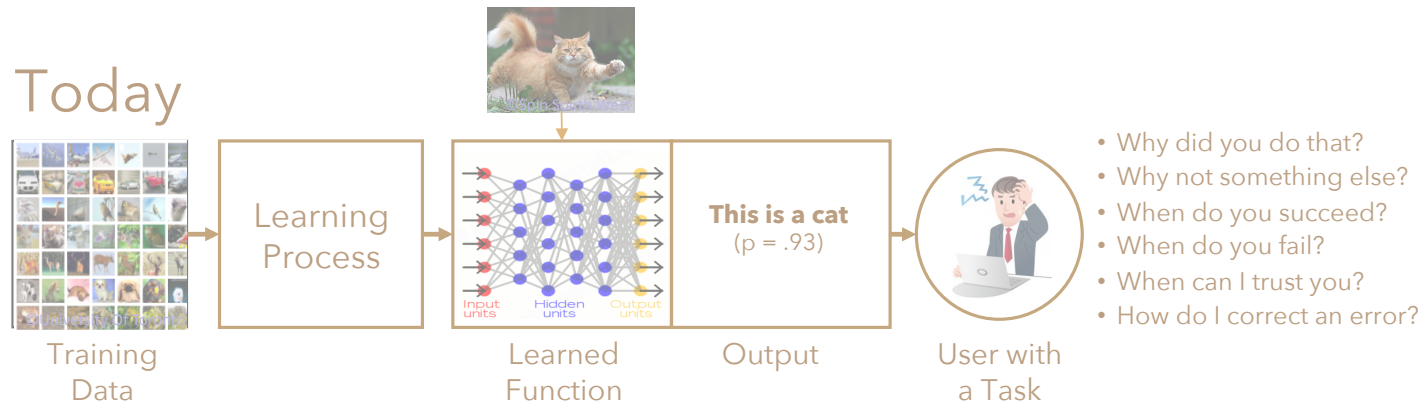
*Machine Learning /  
“Dumb” A.I.*

# Machine Learning 101

$f$  :   $\rightarrow$  Cat



# Explainable AI – What Are We Trying To Do?



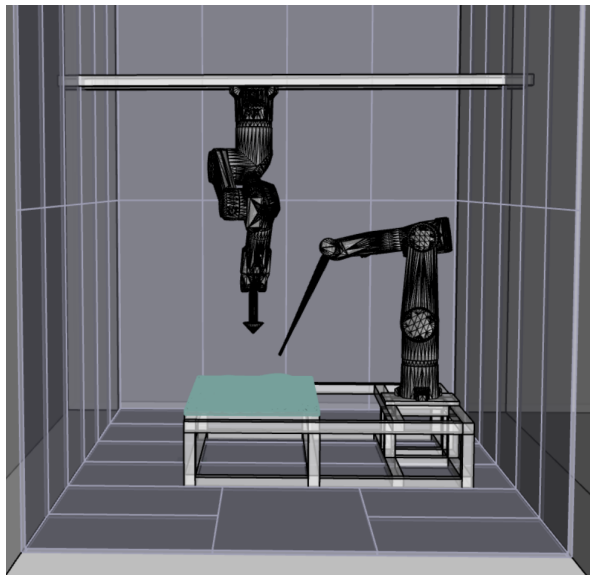
# Data-Driven Acoustical Design

Joint research collaboration with the Architecture group at ETH Zurich

**Problem:** Modeling sound propagation and diffusion in everyday rooms

**Initial results:** Estimation of impulsive response from different walls

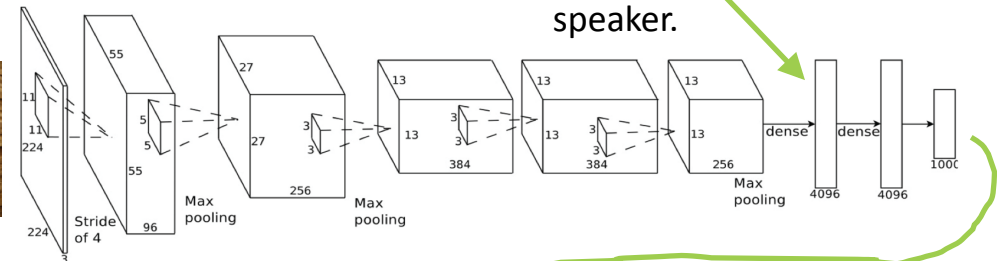
**ETH** zürich



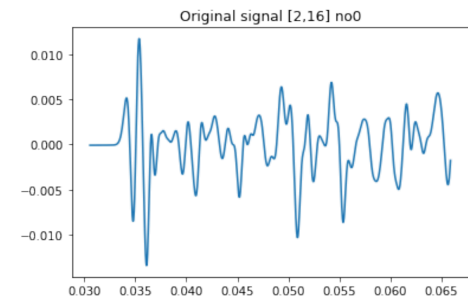
Wall tensor



100×60×10



From the wall, we predict the signal that would be received on each microphone



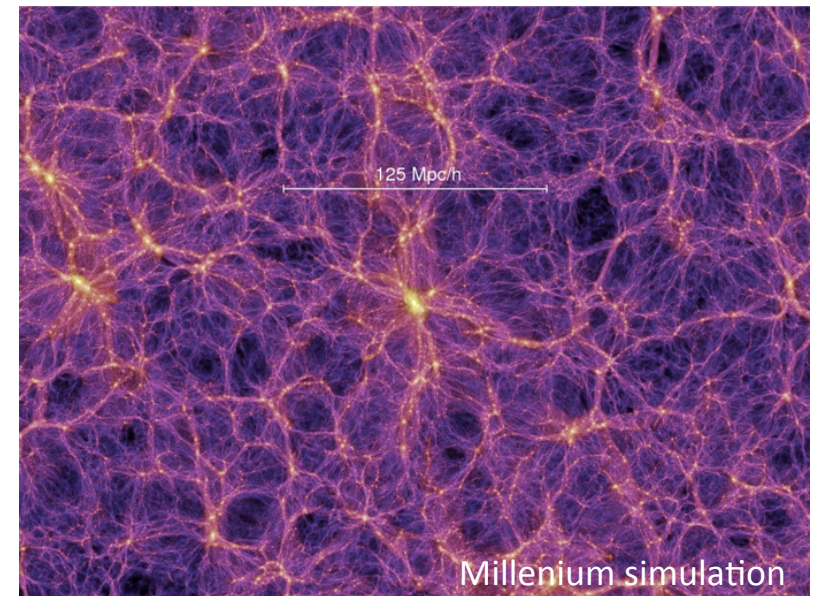
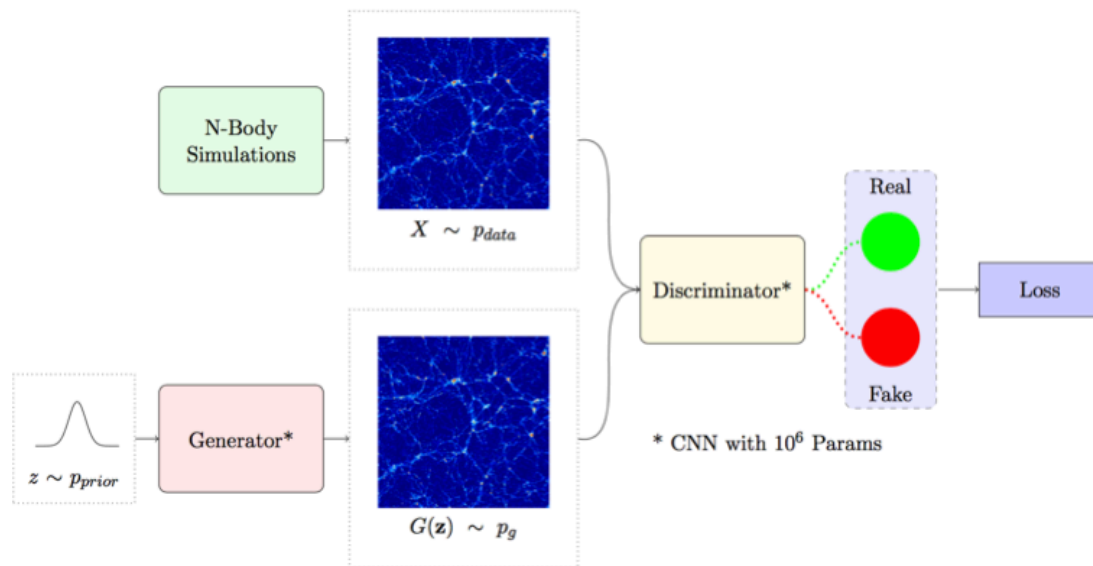
# Deep Learning for Observational Cosmology

Joint research collaboration with the Cosmology Research Group at ETH Zurich

**Problem:** Observational cosmology relies on computationally expensive simulations

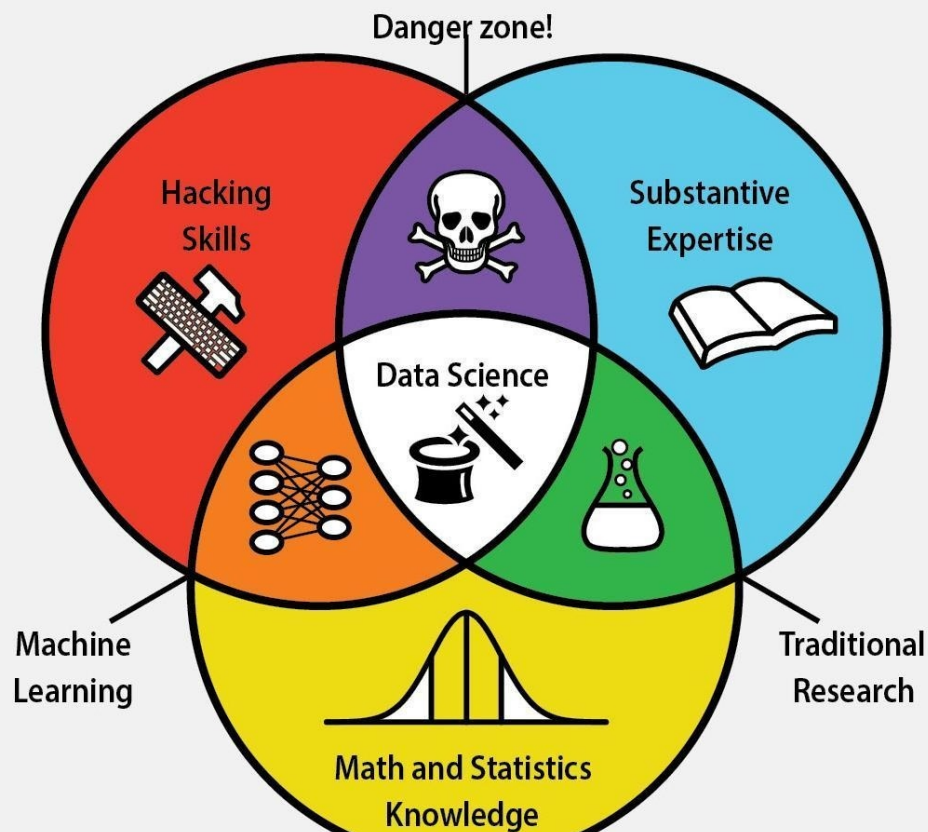
**Results:** Using a generative adversarial network (GAN), we can generate new approximate simulations for a fraction of the computational resources.

**ETH** zürich



# From AI to Data Science

## DATA SCIENCE SKILLSET



Data science, due to its interdisciplinary nature, requires an intersection of abilities: **hacking skills**, **math and statistics knowledge**, and **substantive expertise** in a field of science.



**Hacking skills** are necessary for working with massive amounts of electronic data that must be acquired, cleaned, and manipulated.



**Math and statistics knowledge** allows a data scientist to choose appropriate methods and tools in order to extract insight from data.



**Substantive expertise** in a scientific field is crucial for generating motivating questions and hypotheses and interpreting results.



**Traditional research** lies at the intersection of knowledge of math and statistics with substantive expertise in a scientific field.



**Machine learning** stems from combining hacking skills with math and statistics knowledge, but does not require scientific motivation.



**Danger zone!** Hacking skills combined with substantive scientific expertise without rigorous methods can beget incorrect analyses.

# Sharing data and knowledge, or lack thereof



credit: oxford creativity, <https://www.triz.co.uk/>



# Five FAQs in Data-Driven Research

1. How did I compute this result?
2. How does new data change this result?
3. How did you compute *your* result?
  - Can I use your data to reproduce it?
  - With your code?
  - On your infrastructure?
4. Has anyone ever used an <XYZ-algorithm> on this data? How?
5. Who is using my data? and my algorithm?
  - Why are they not citing me?!





Five Questions → Three Words

**Reproducibility**

Reusability

Collaboration



THANK YOU

[www.datascience.ch](http://www.datascience.ch)  
@SDSCdatascience