

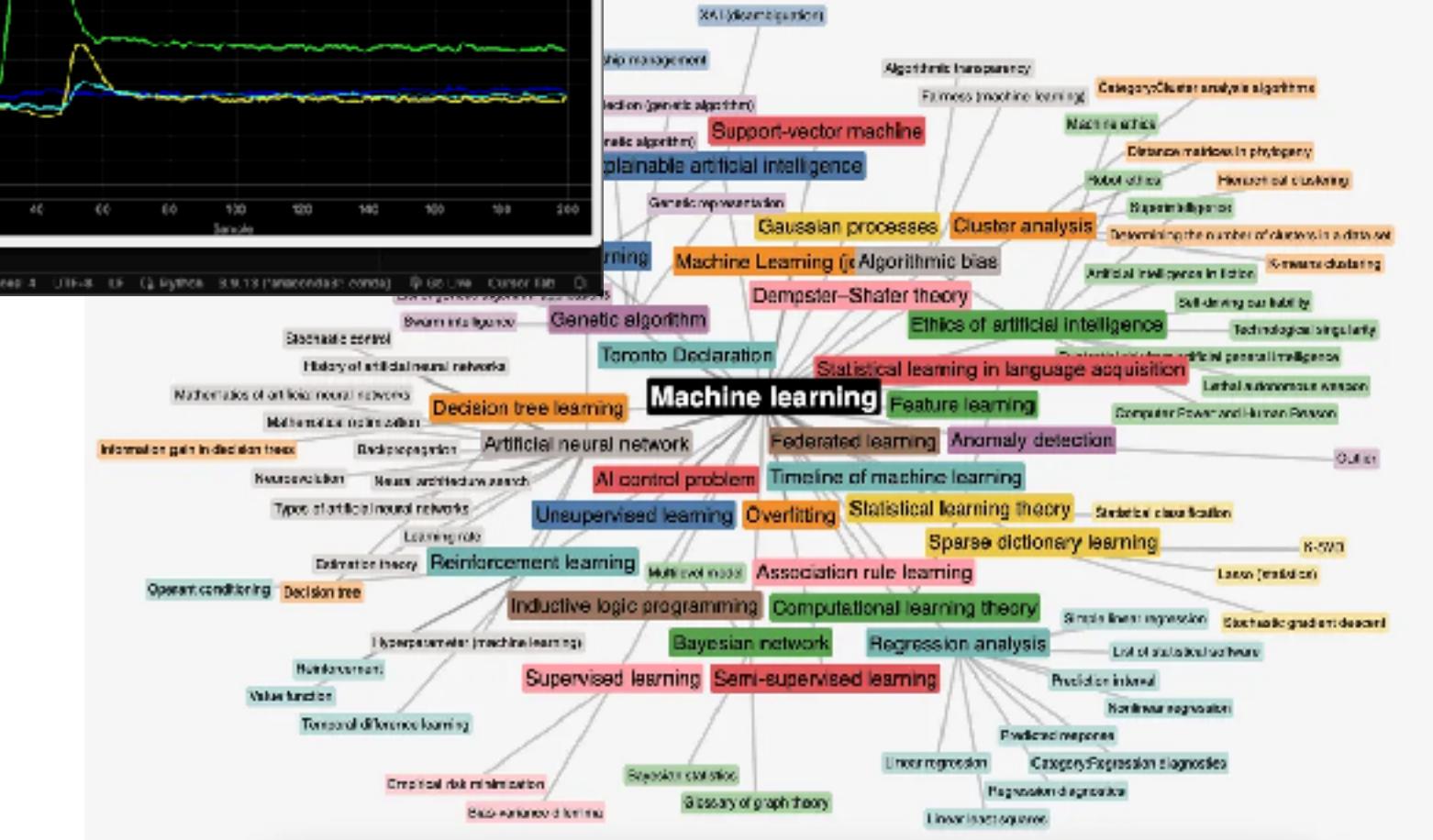
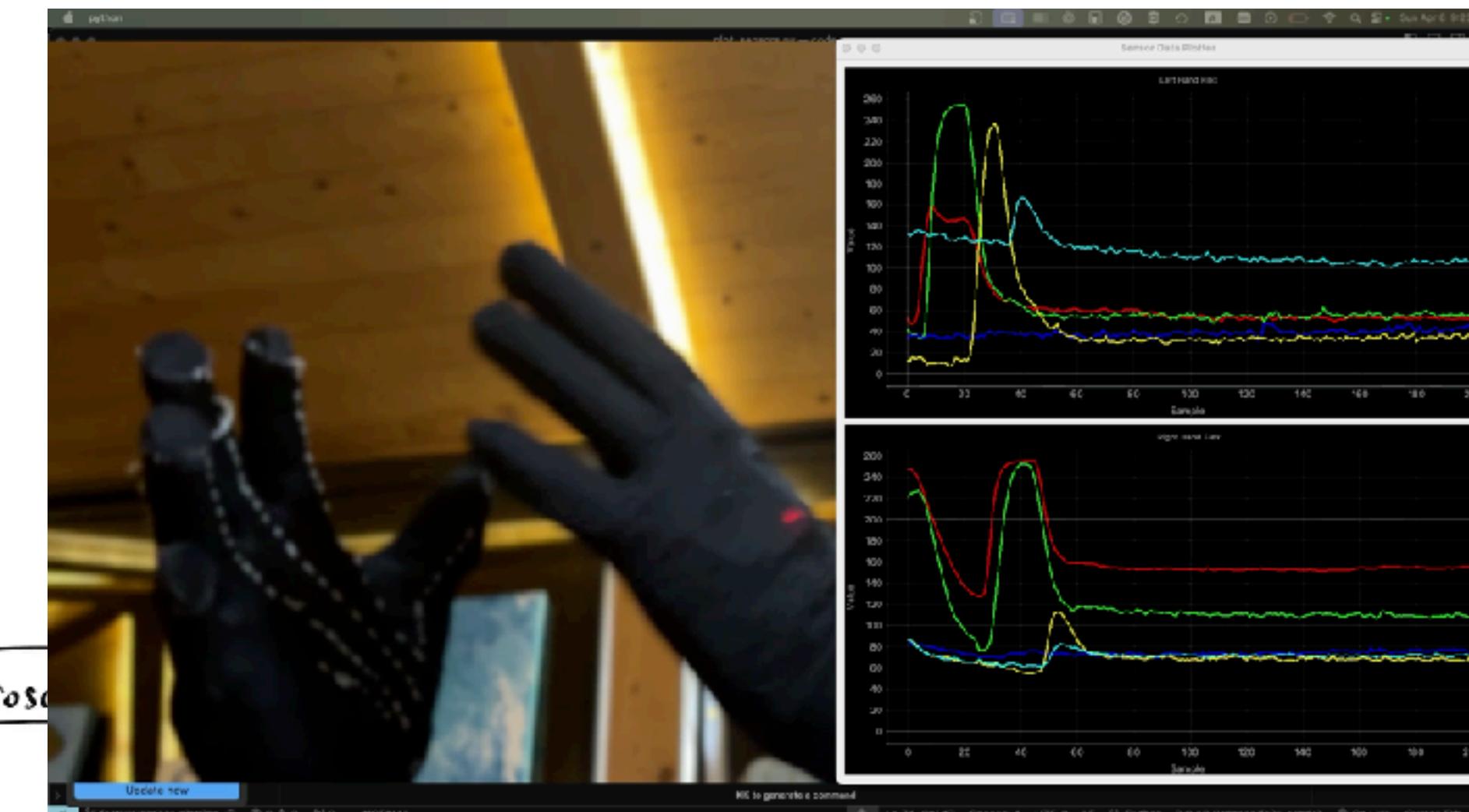
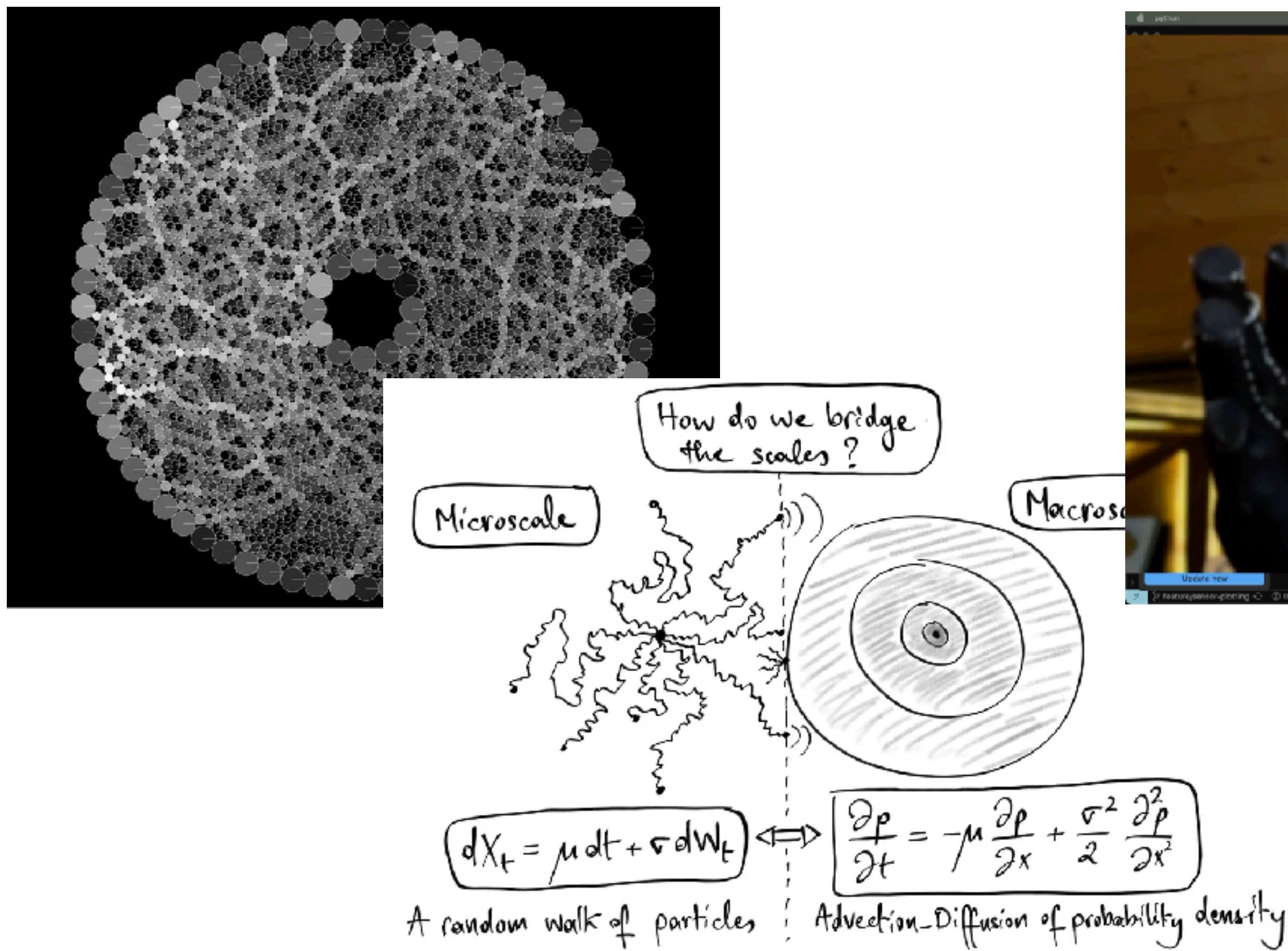
Data-Driven Modeling in Science and Engineering

EECE 798K | MECH 798M

Instructor: Joseph Bakarji

A bit about me

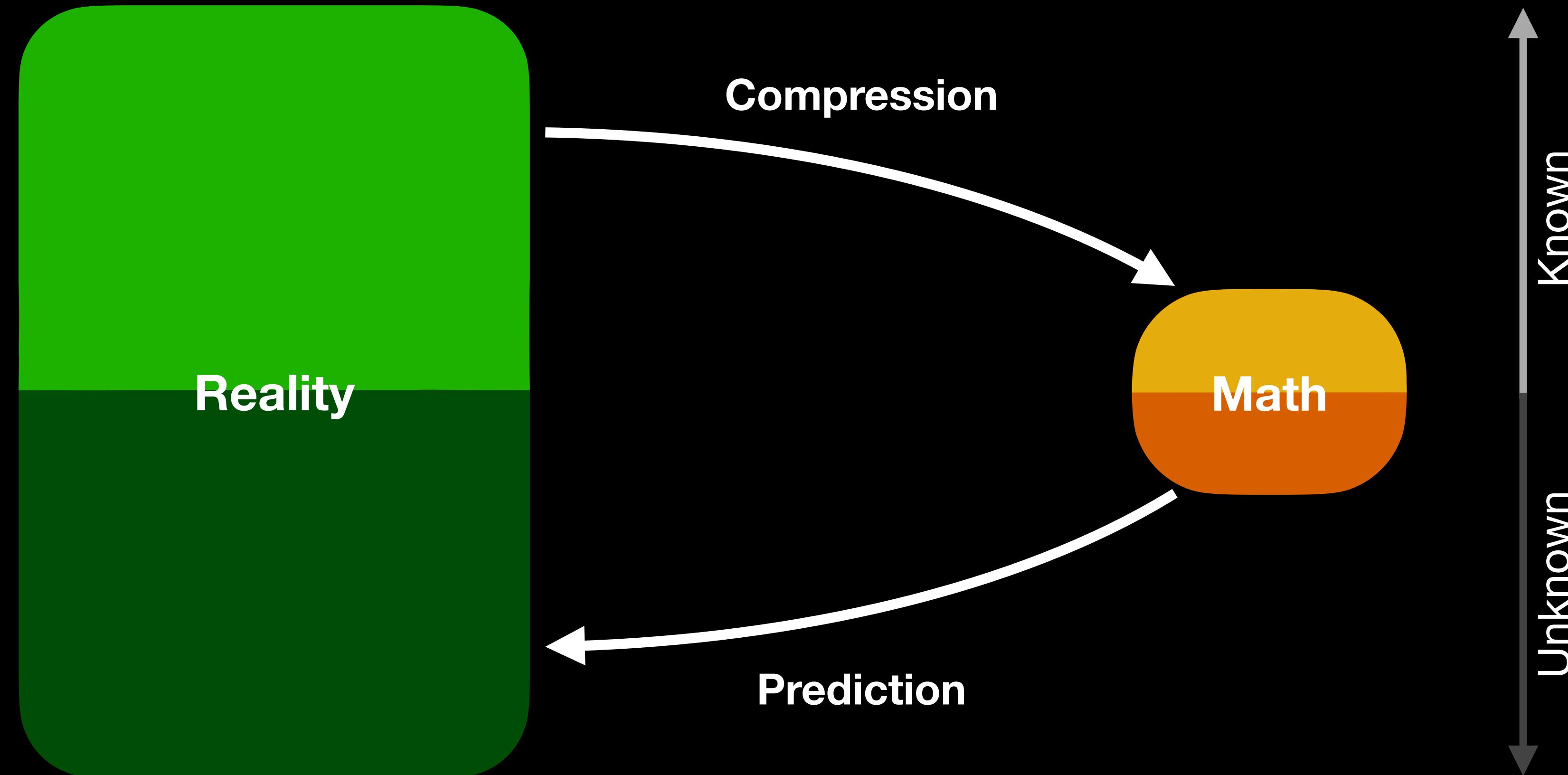
- Name and contact: **Joseph Bakarji** (jb50@aub.edu.lb) - www.josephbakarji.com
 - My research is at the intersection of machine learning, scientific modeling and music.



What is Science?

How can it be that mathematics [...]
is so admirably adapted to the objects of reality?

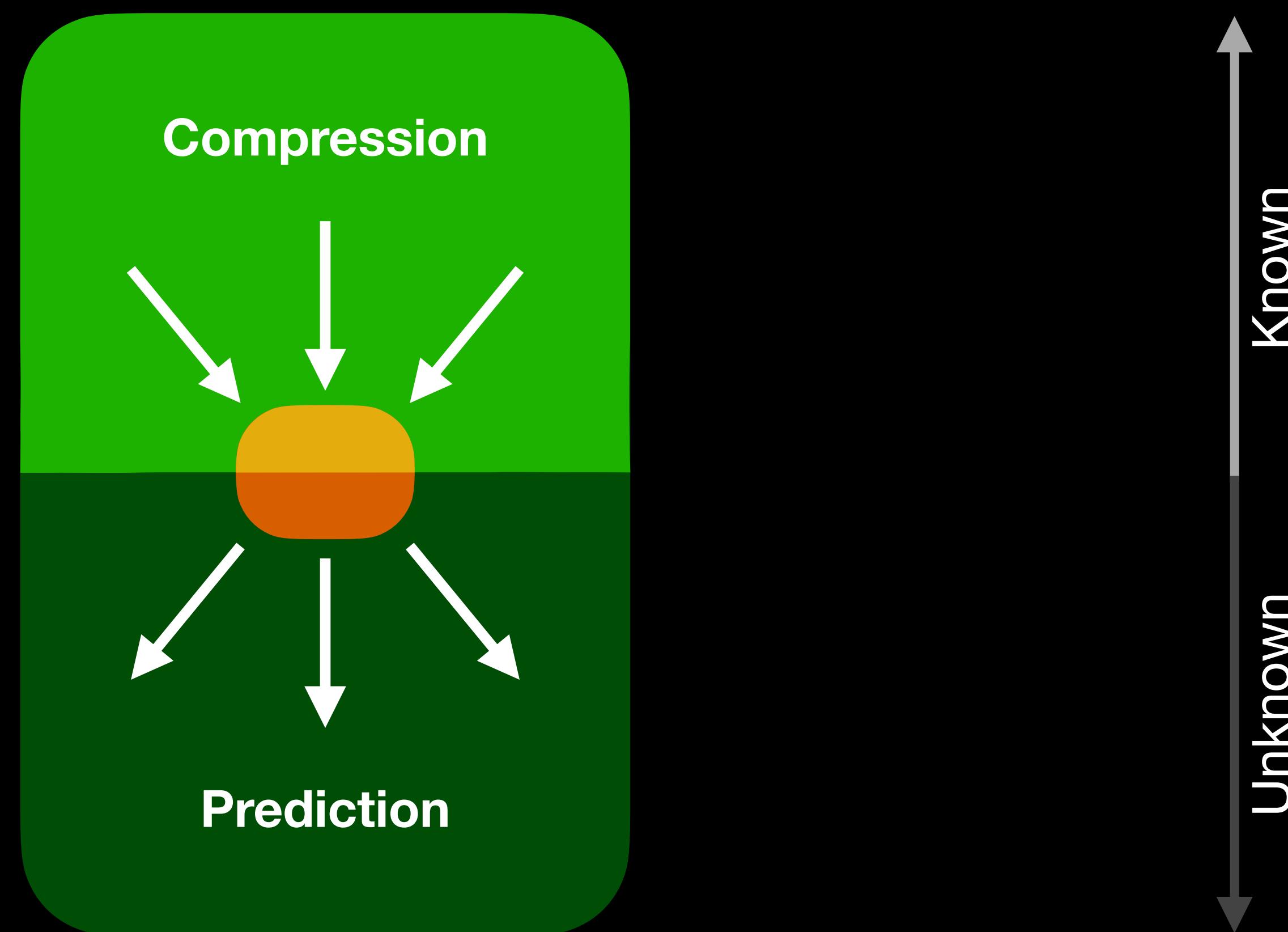
Albert Einstein



What is Science?

How can it be that mathematics [...]
is so admirably adapted to the objects of reality?

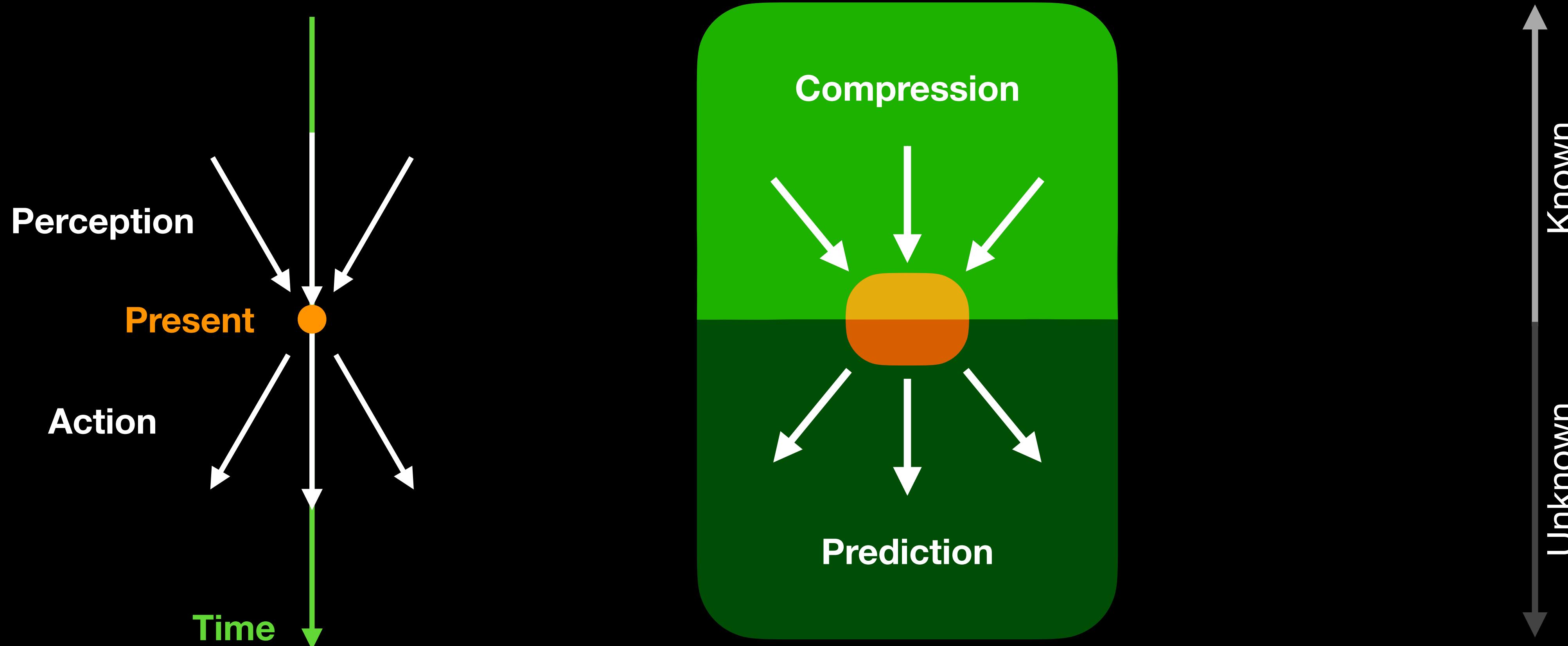
Albert Einstein



What is Intelligence?

The limits of my language mean the limits of my world.

Ludwig Wittgenstein



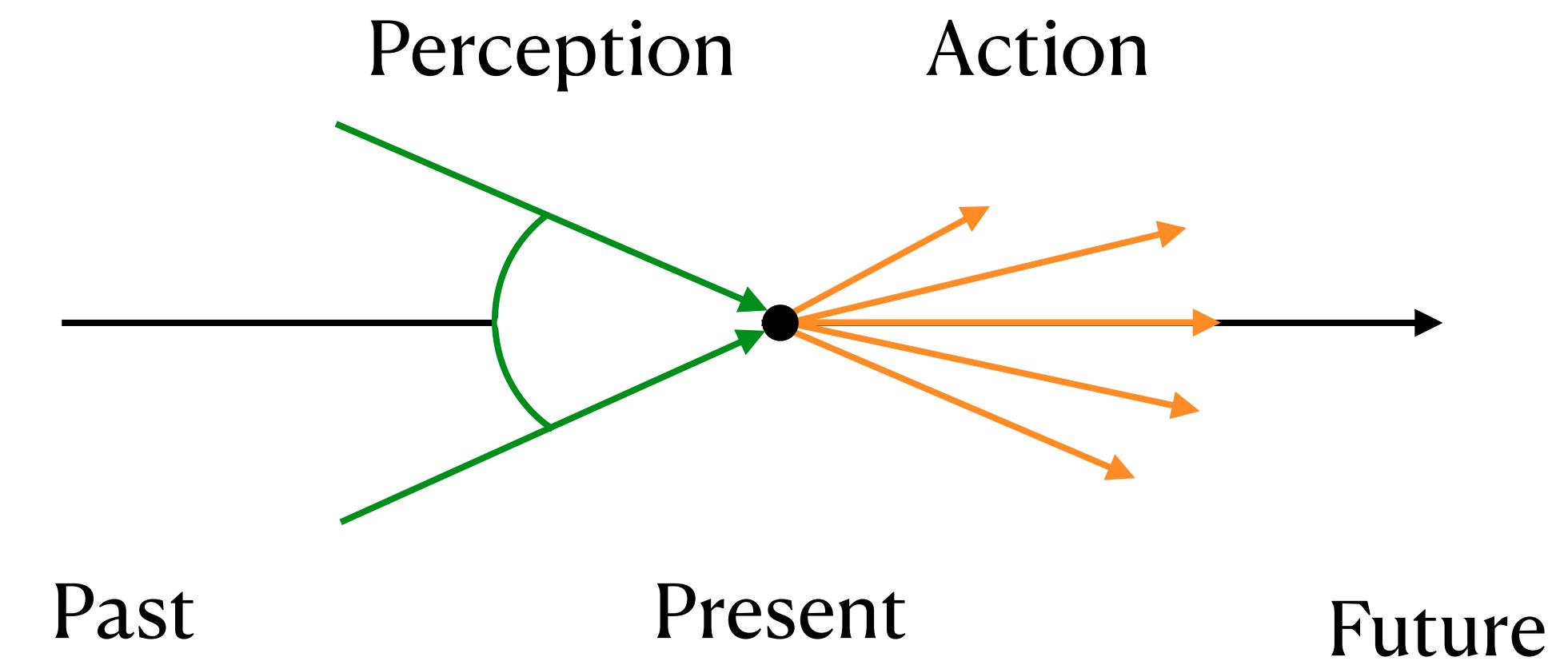
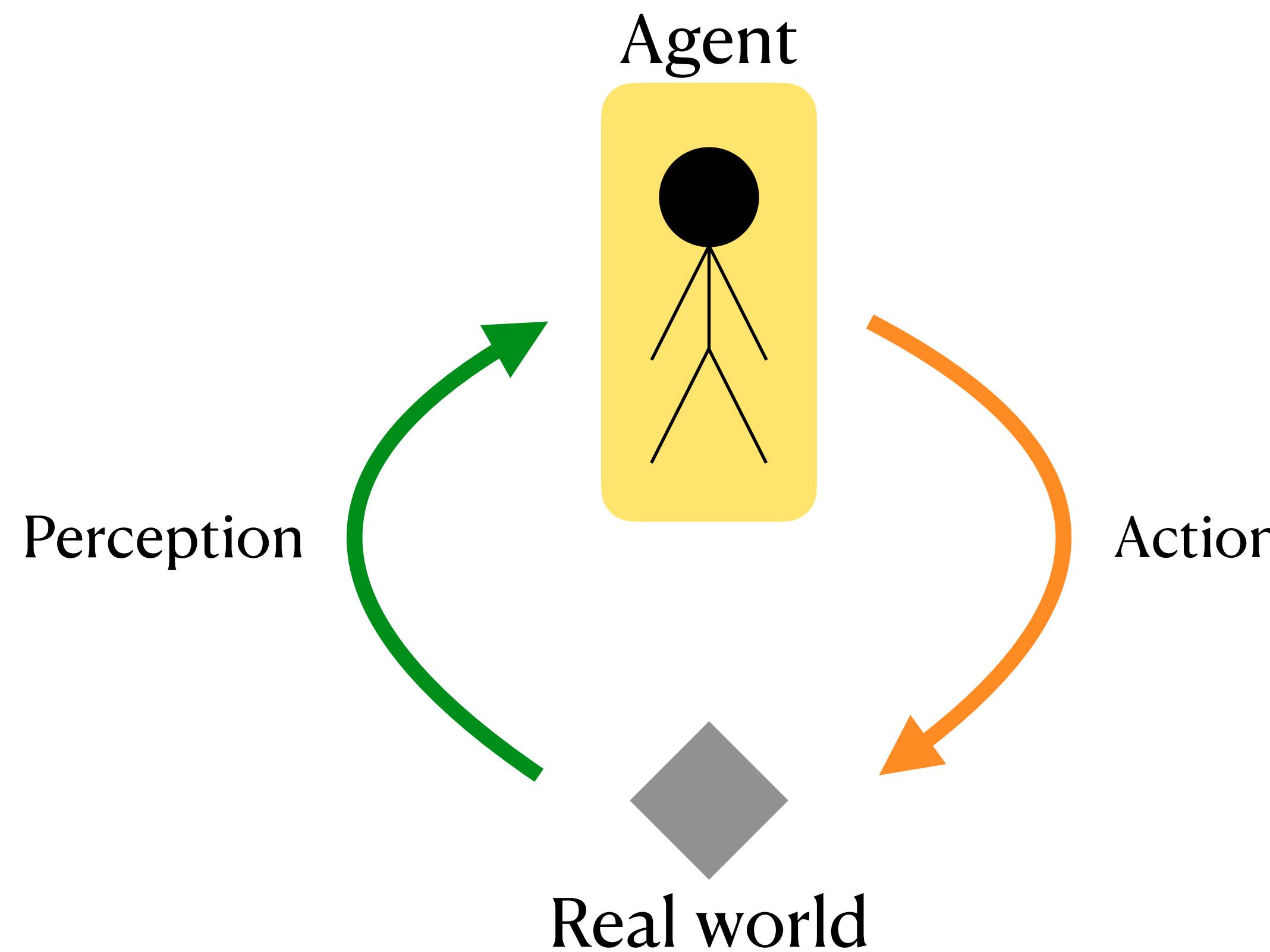
How does Intelligence create Science ?

**What is it that we *do*
when we do science?**

**Can we automate the process
of scientific discovery?**

**Can we build
scientist machines?**

What is Intelligence?



World Models

Can agents learn inside of their own dreams?

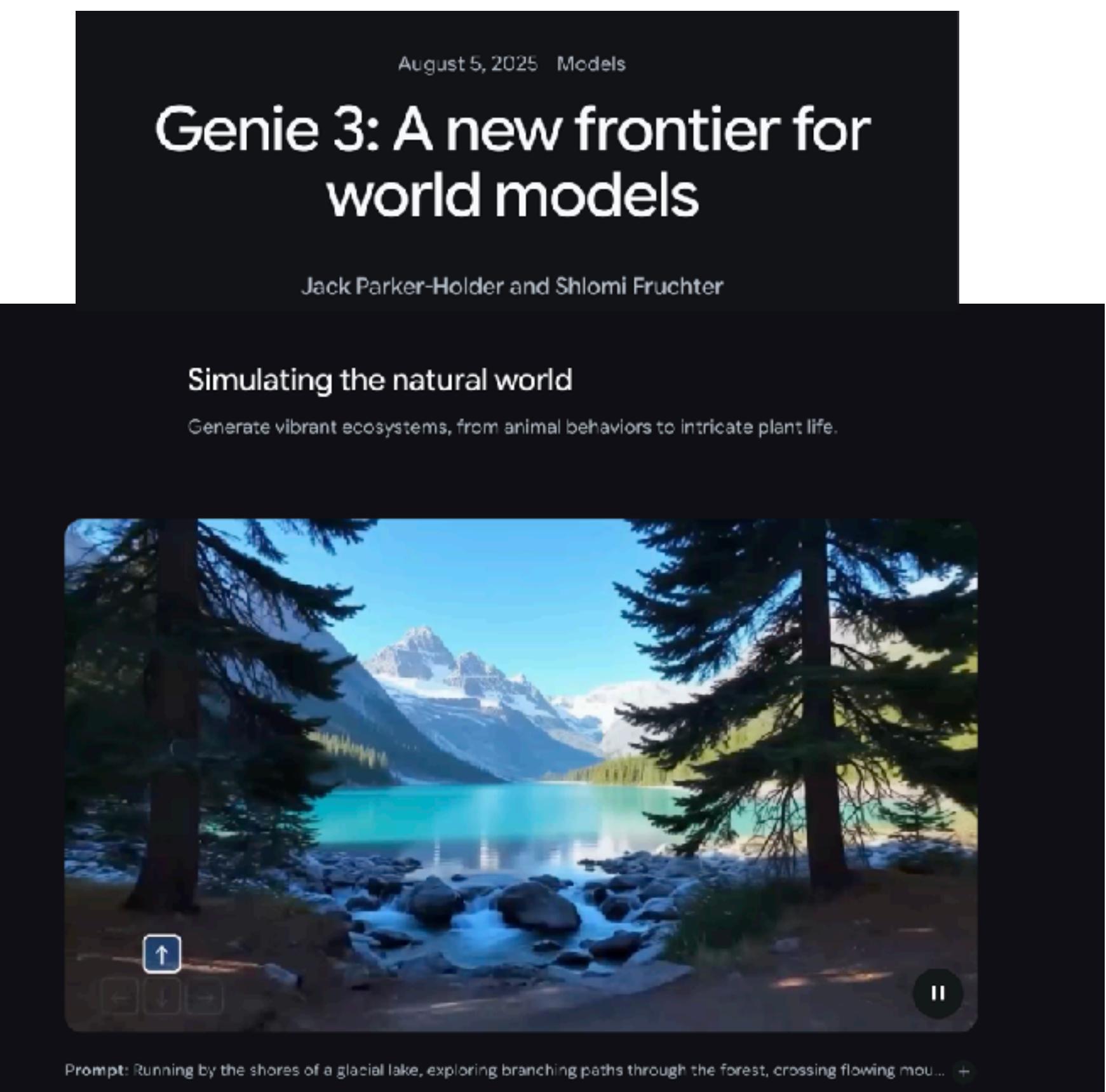
DAVID HA JÜRGEN SCHMIDHUBER
Google Brain NNAISENSE
Tokyo, Japan Swiss AI Lab, IDSIA (USI & SUPSI)

March 27
2018

NIPS 2018
Paper

YouTube
Talk

Download
[PDF](#)



<https://deepmind.google/blog/genie-3-a-new-frontier-for-world-models/>

From Words to Worlds: Spatial Intelligence is AI's Next Frontier



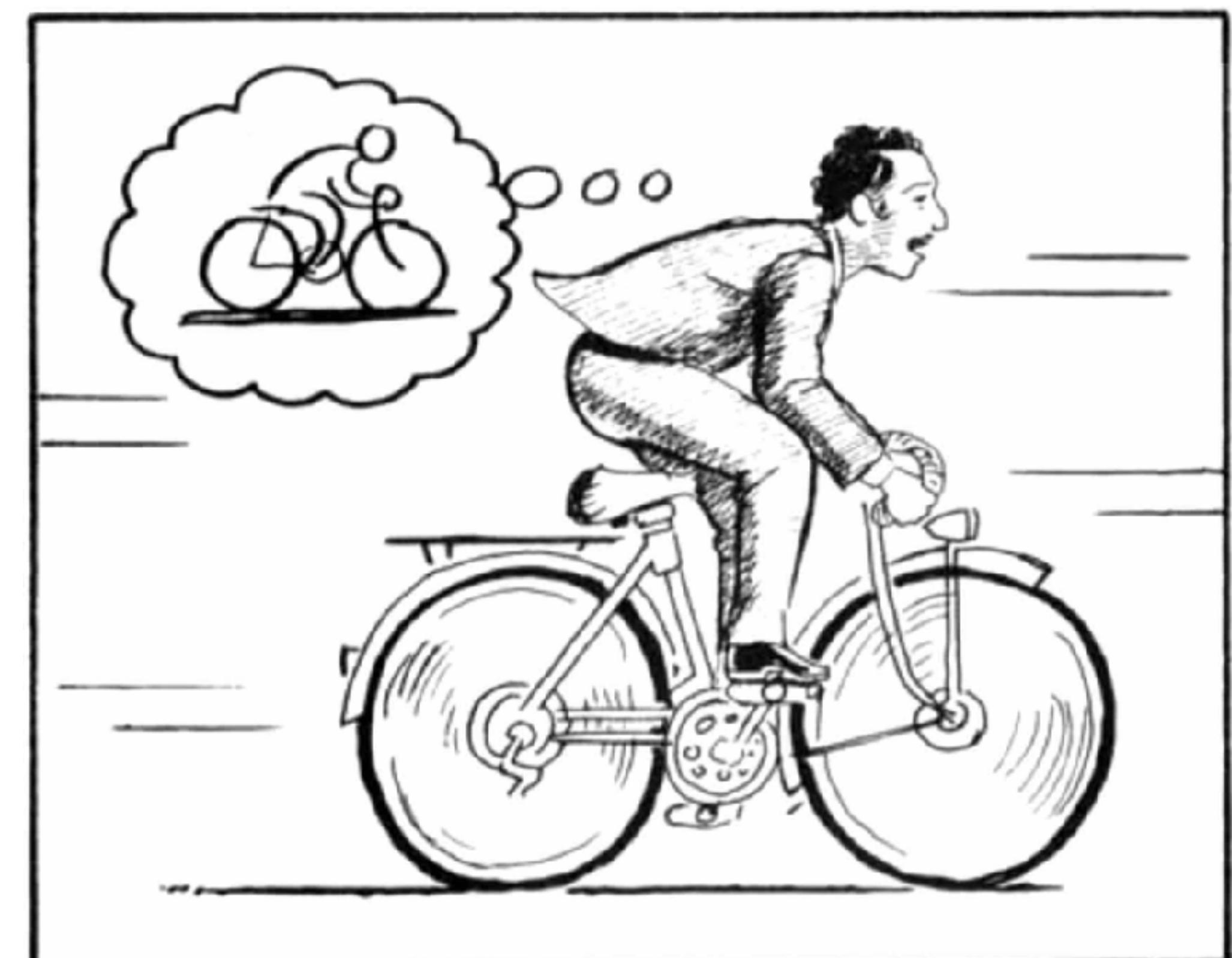
FEI-FEI LI
NOV 10, 2025

1,907 303

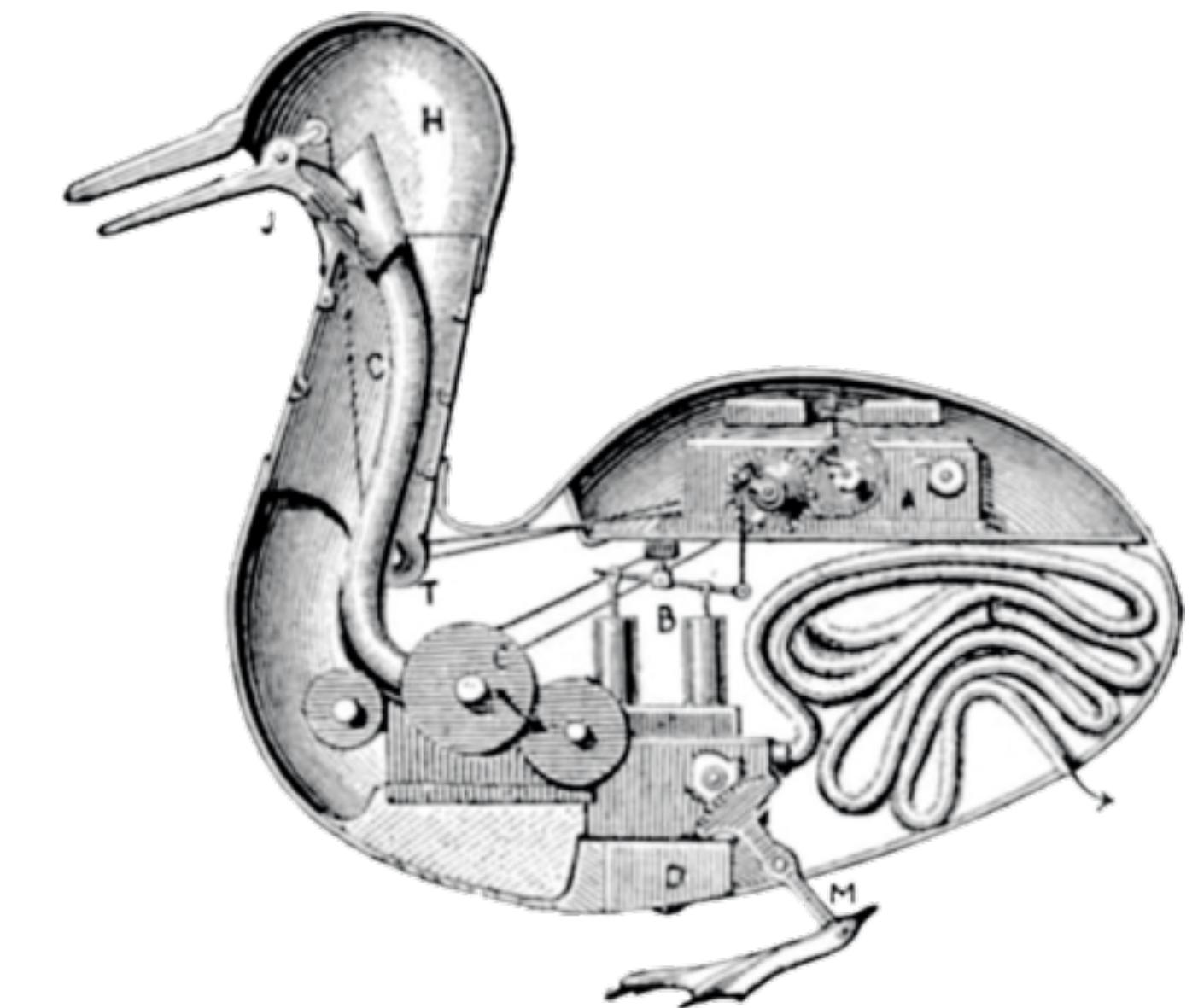
Share

...

<https://worldmodels.github.io/>



Intelligent mechanical machines

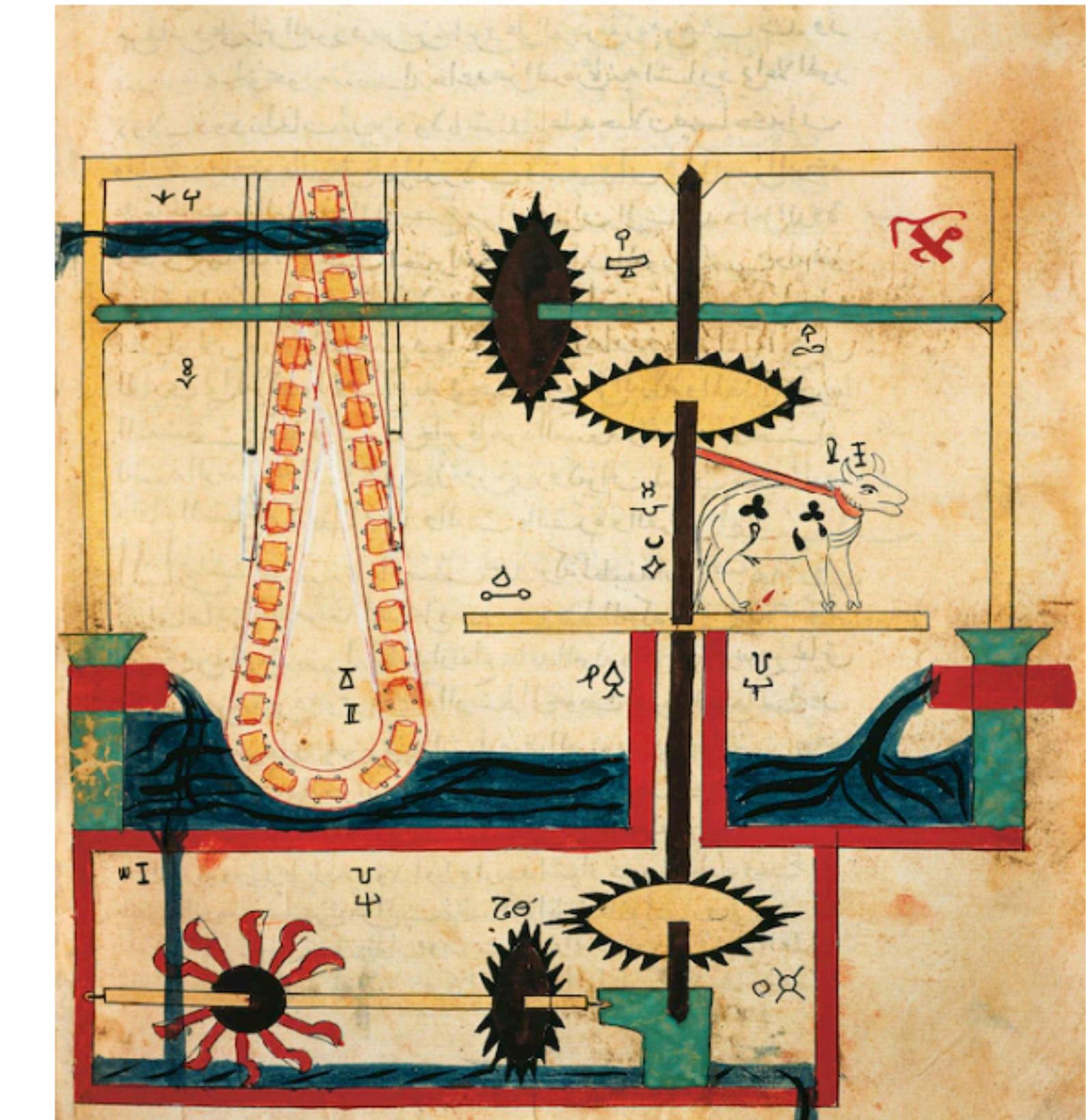
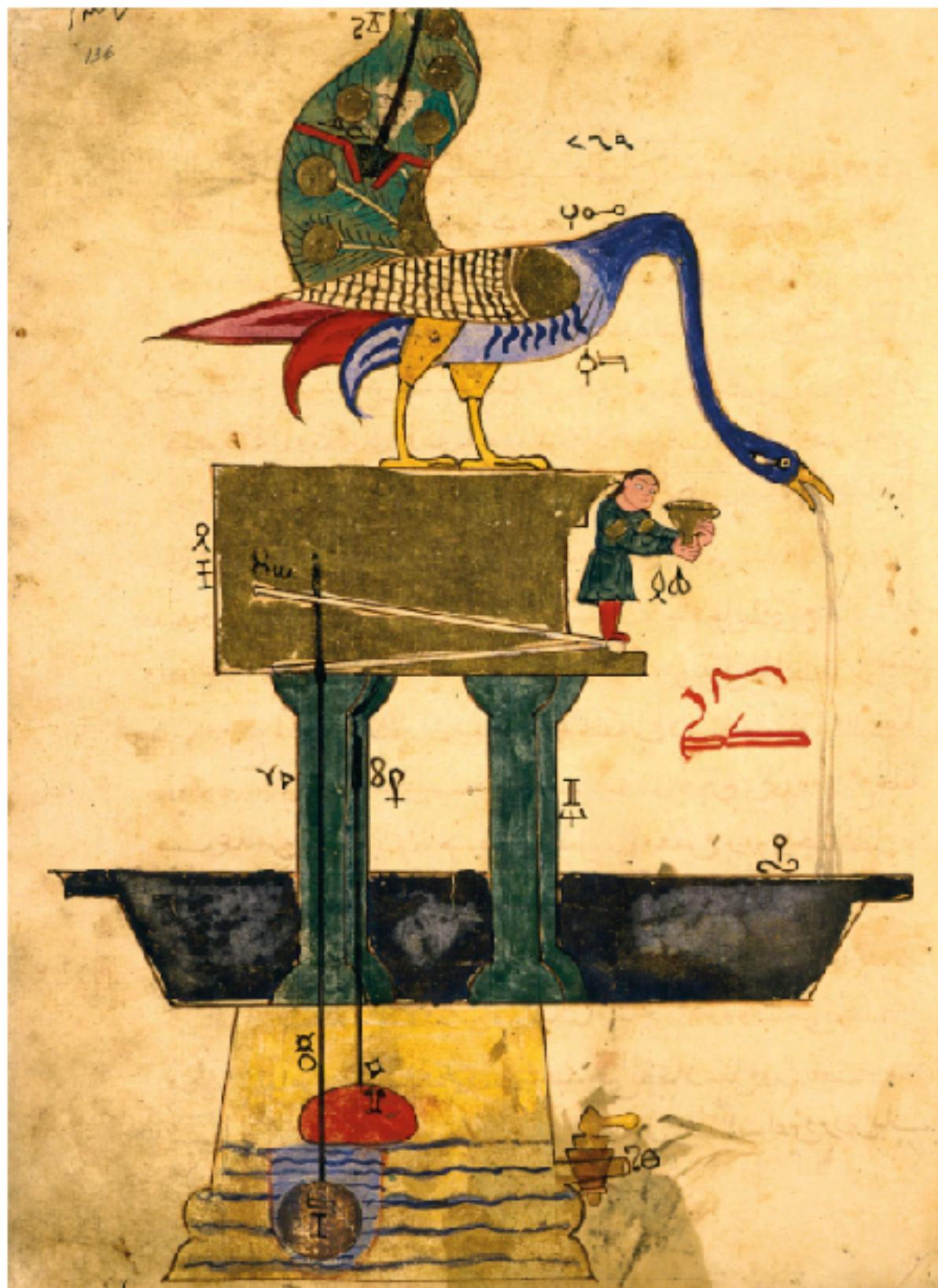


The Digesting Duck
Jacques Vaucanson (1739)

18th century automata

Early programmable “robots” for practical applications

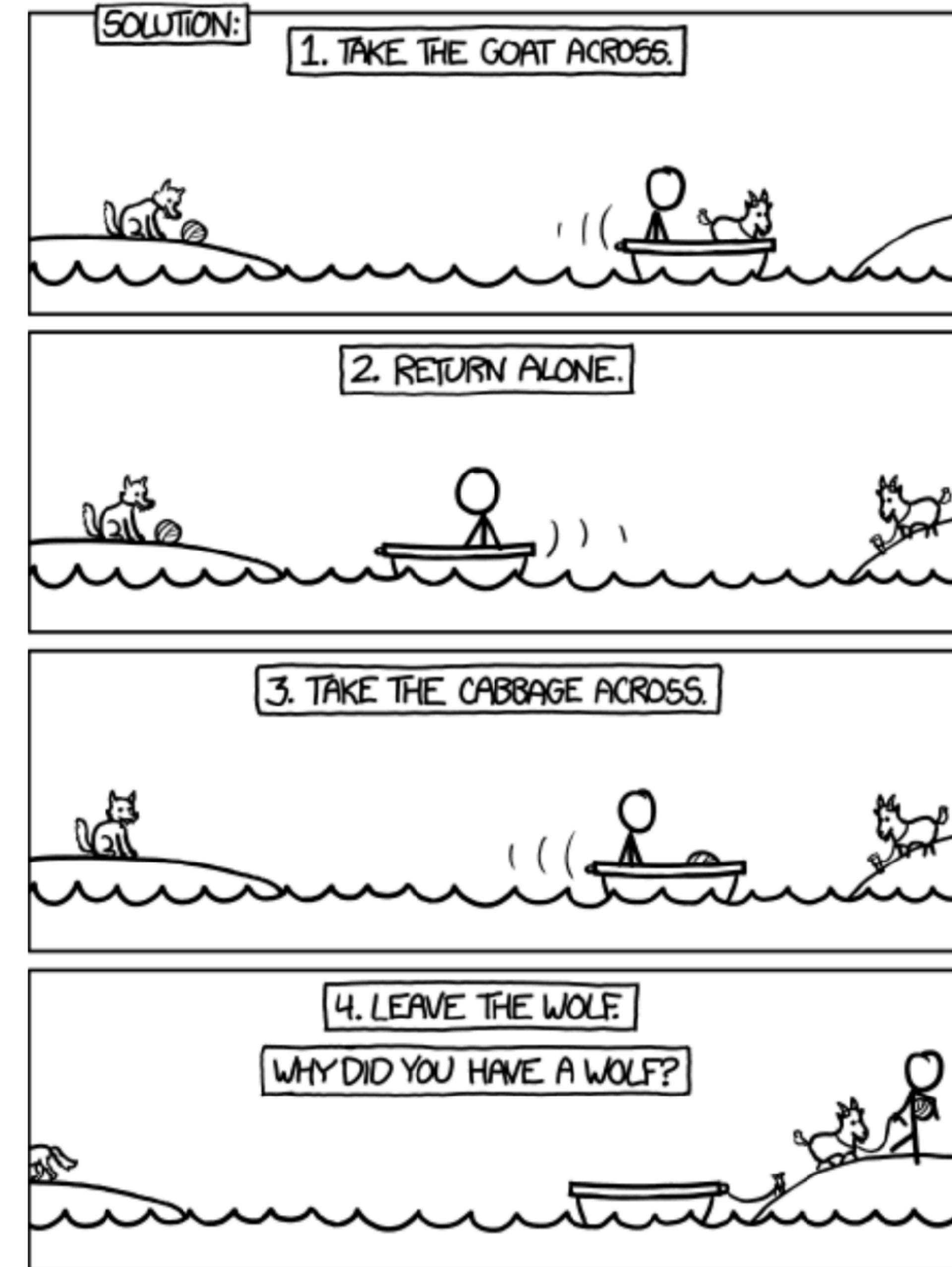
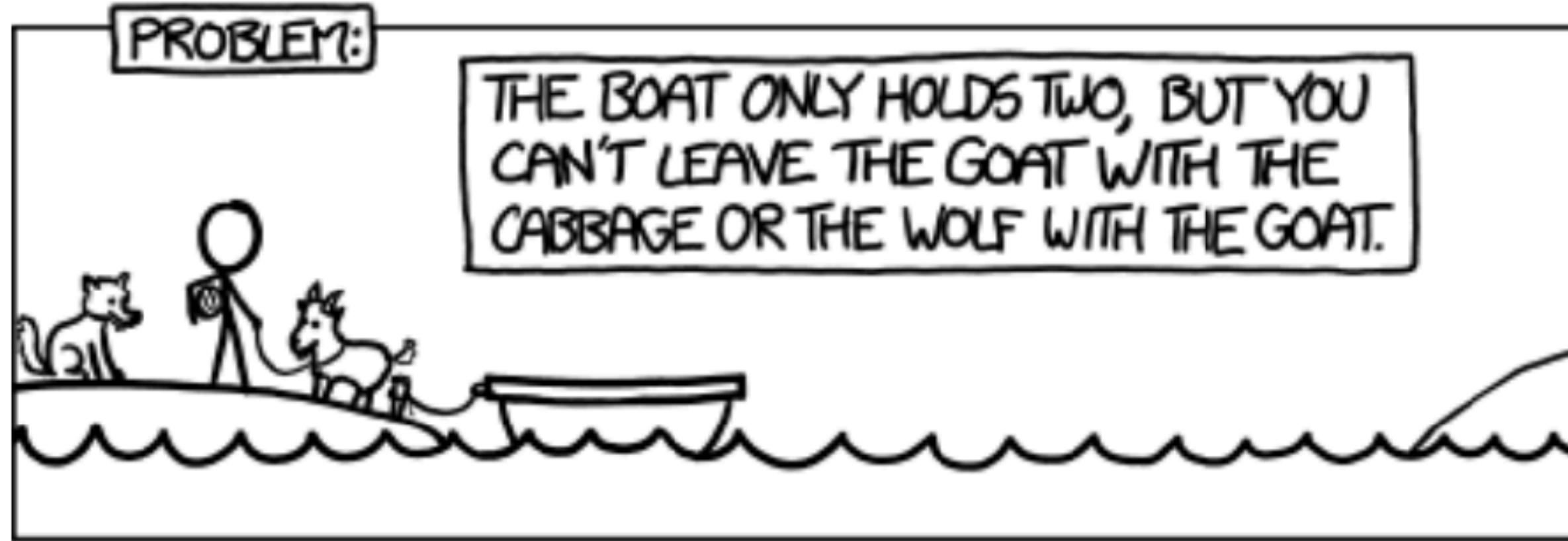
بديع الزمان أَبُ الْعِزِّ إِبْنُ إِسْمَاعِيلِ إِبْنُ الرِّزَازِ الْجَزَرِي



1206

Artificial Intelligence 1950's-

Solving puzzles

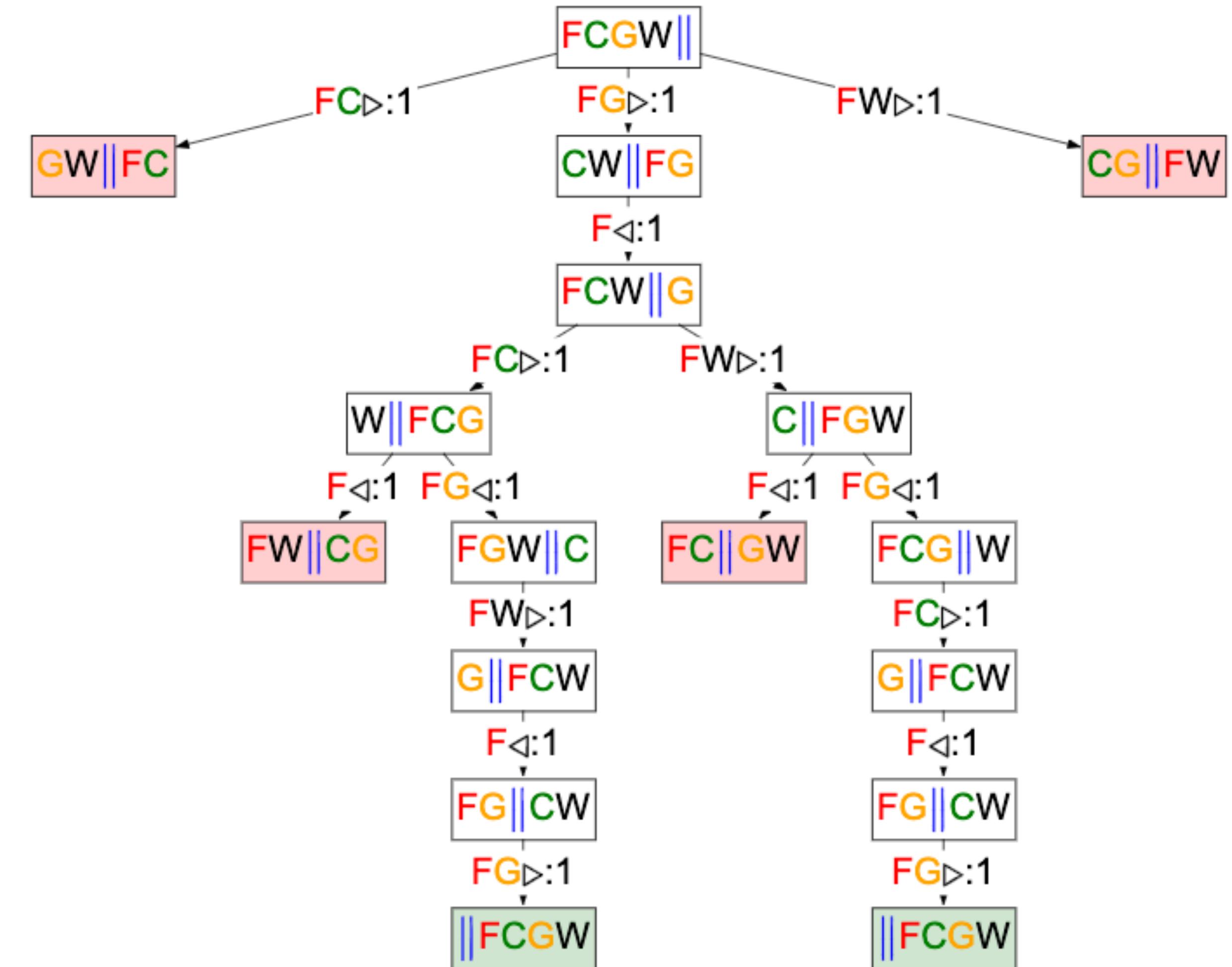


Find optimal search strategies



Farmer Cabbage Goat Wolf

$F\triangleright$	$F\triangleleft$
$FC\triangleright$	$FC\triangleleft$
$FG\triangleright$	$FG\triangleleft$
$FW\triangleright$	$FW\triangleleft$



Overwhelming optimism

The spirit is willing but the flesh is weak.



(Russian)



The vodka is good but the meat is rotten.

1950's

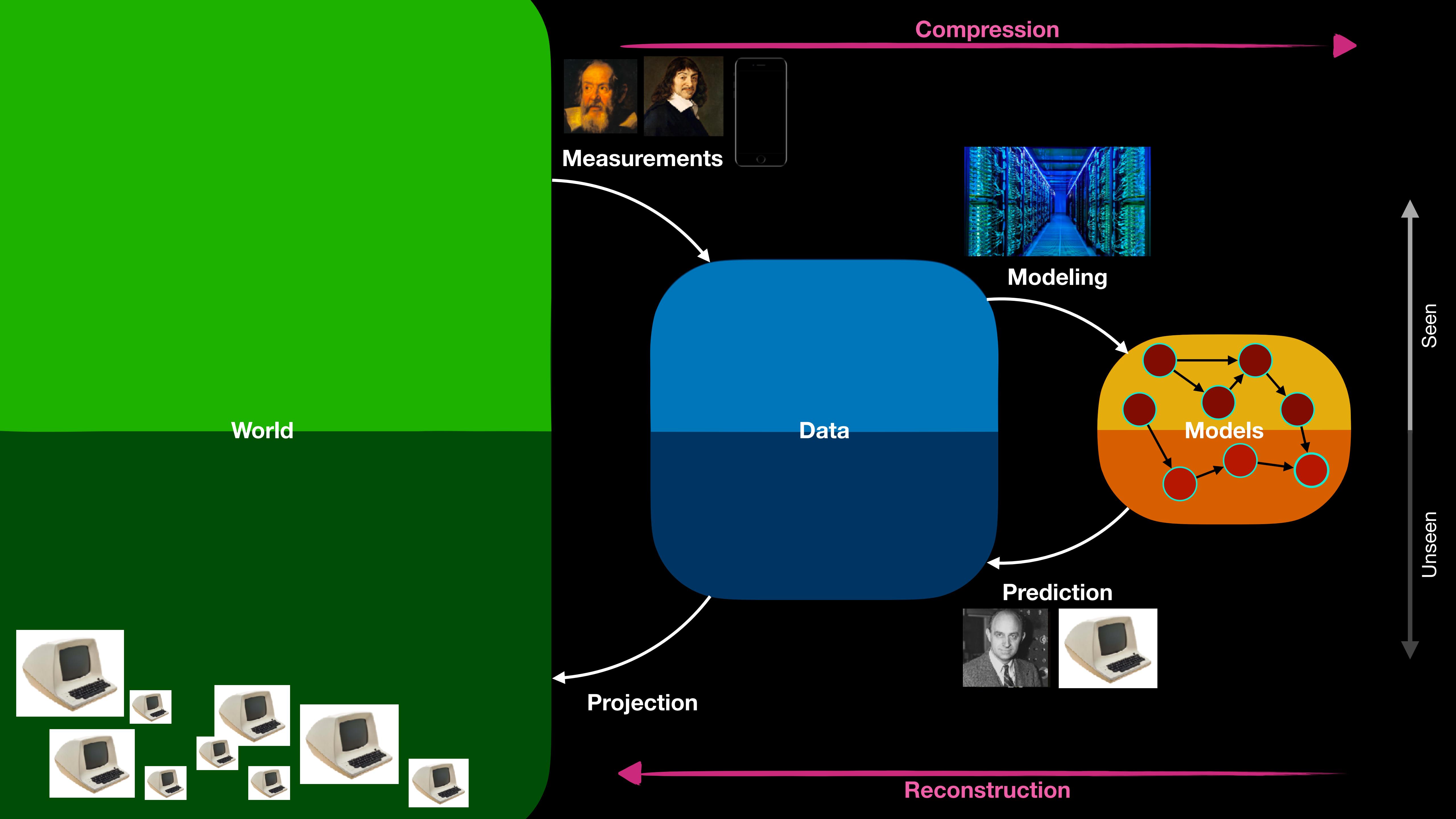
Rebirth of Artificial Intelligence

More real-world data

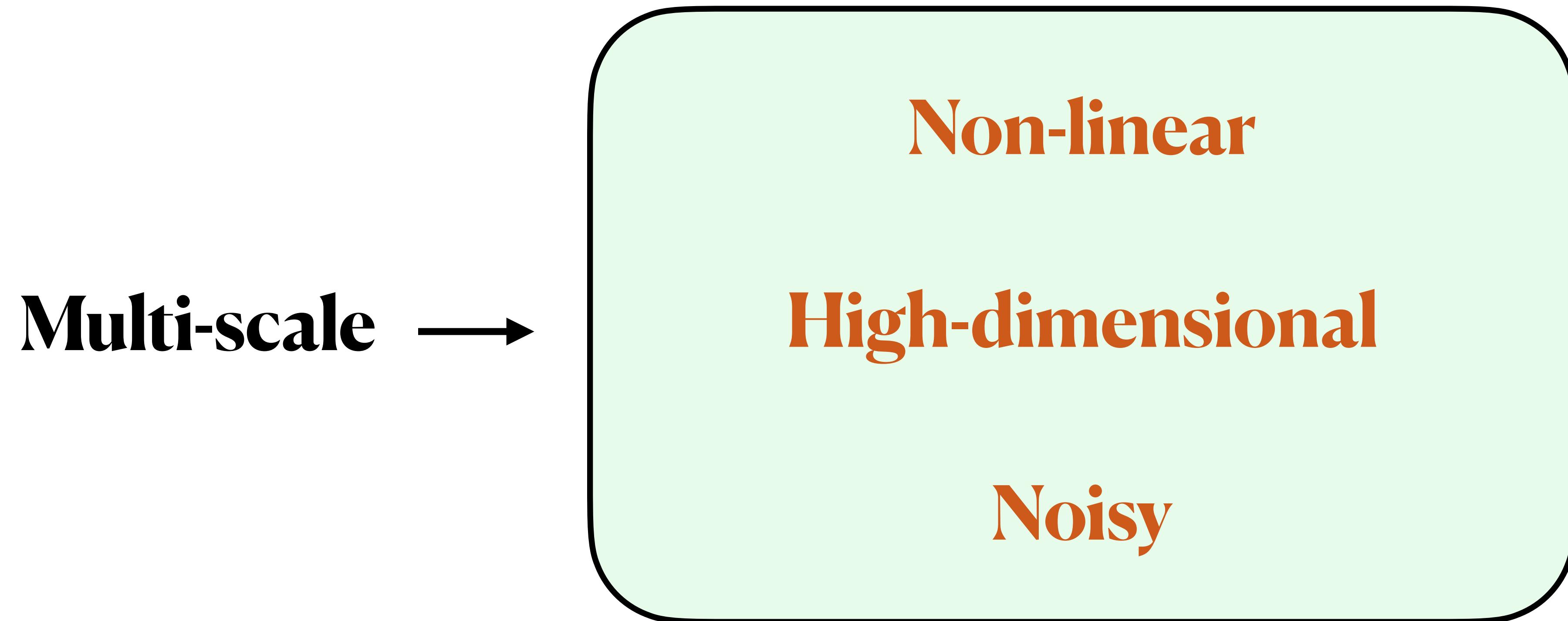
Faster computers

Better algorithms

How does the world work?



The world is:



Probability

Course plan

Discovering **Non-linear Functions from Data**

Supervised Learning. Parametric/non-parameteric functions etc.

Linear regression and classification

Learning theory: variance-bias trade-off, regularization etc.

Deep learning

....

Course plan

Reducing **High-dimensionality** of observations

Singular value decomposition

Reduced-order modeling: proper orthogonal decomposition

Spectral methods: Fourier transform

Auto-encoders

.....

Course plan

Machine Learning for Physical Modeling

Discovering/solving equations from data

System Identification

Symbolic Regression

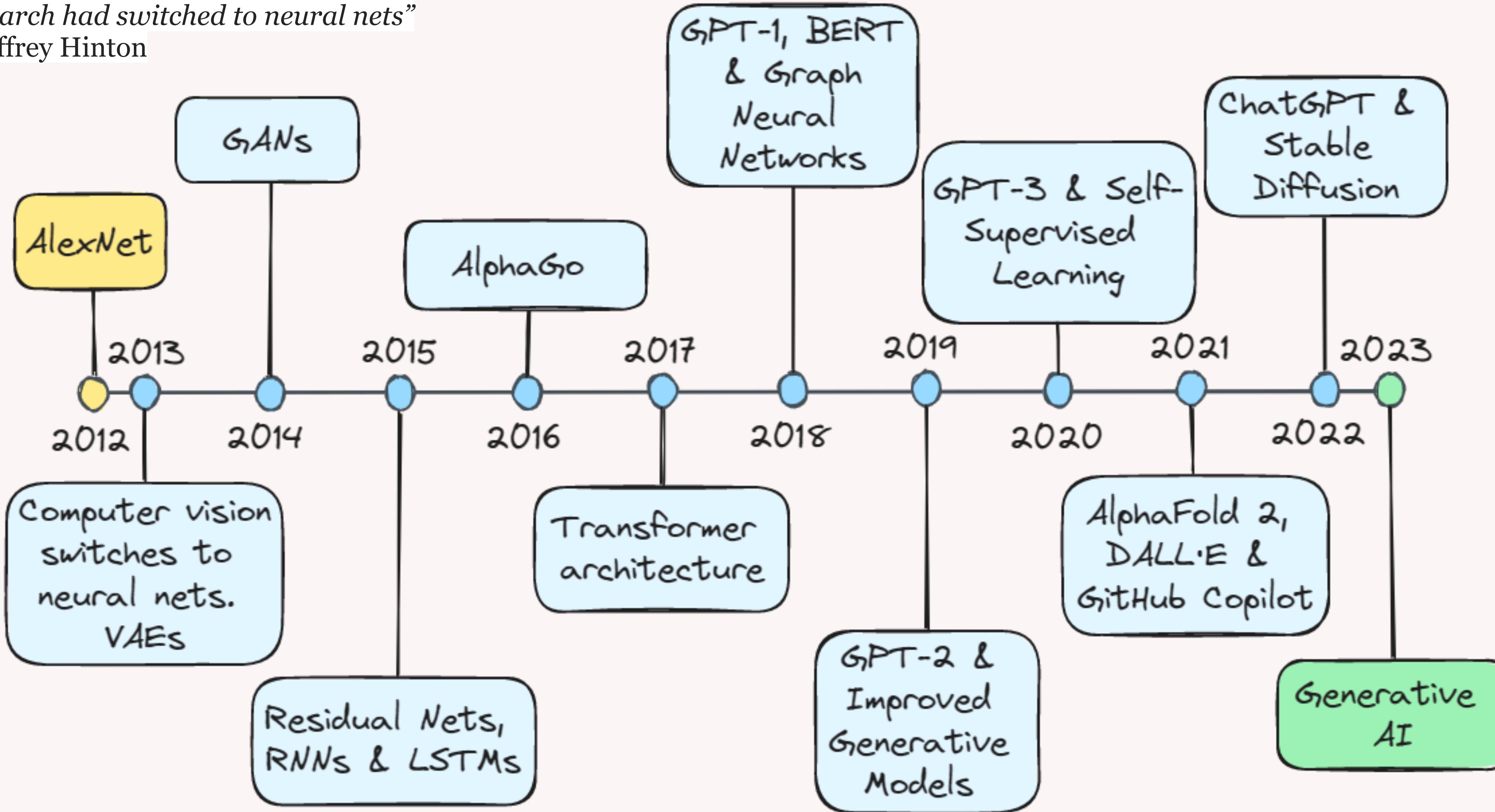
Sparse Identification of Differential Equations

Physics-Informed Neural Networks

Fourier Neural Operators

Dynamic Mode Decomposition

“By 2013, pretty much all the computer vision research had switched to neural nets”
Geoffrey Hinton



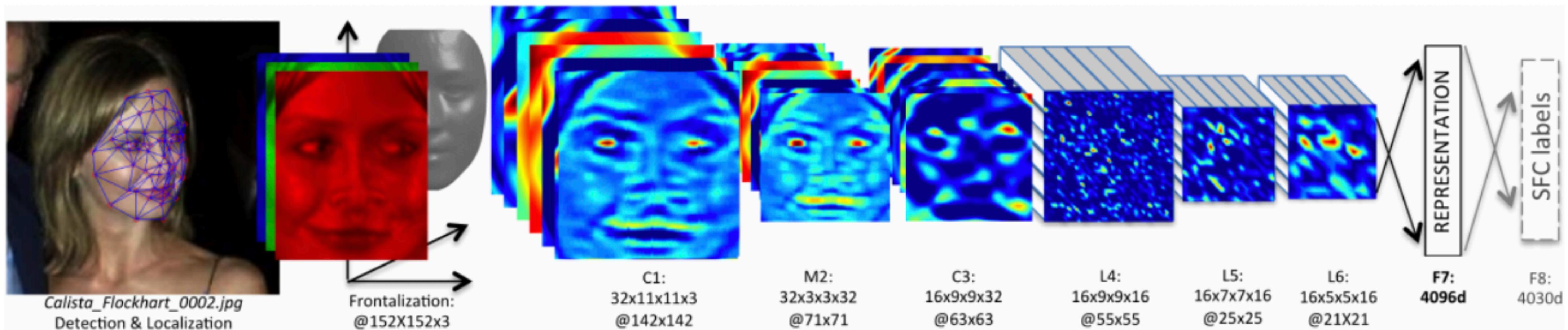
Autonomous Driving



Google announces Waymo in 2010

Image recognition

- Unlocking your phone
- Recognizing people in your photos



- Google Brain uses deep learning to recognize cats in images (2012)
- Facebook recognizes faces with 97% accuracy (2014)

Progress in the last 10 years

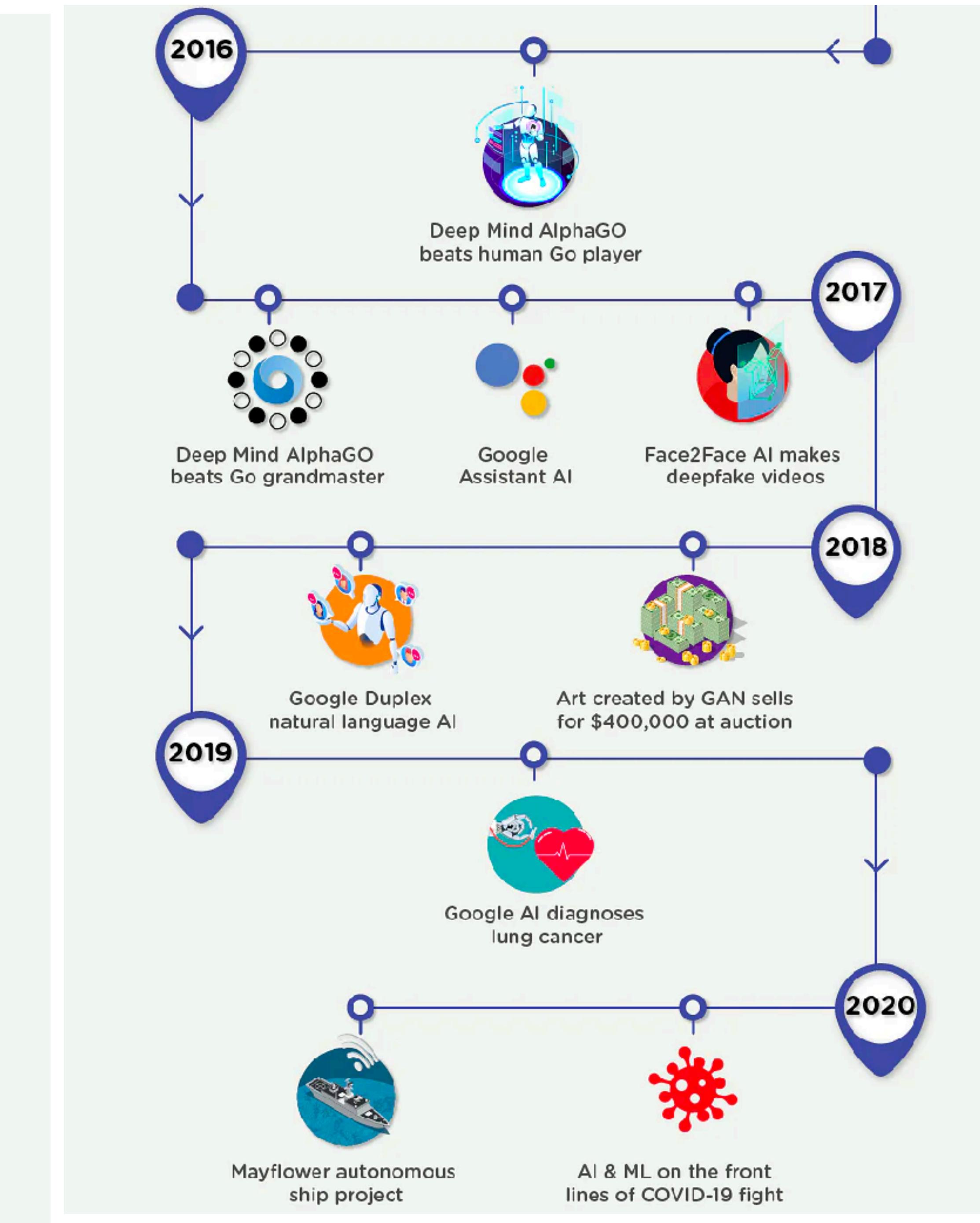
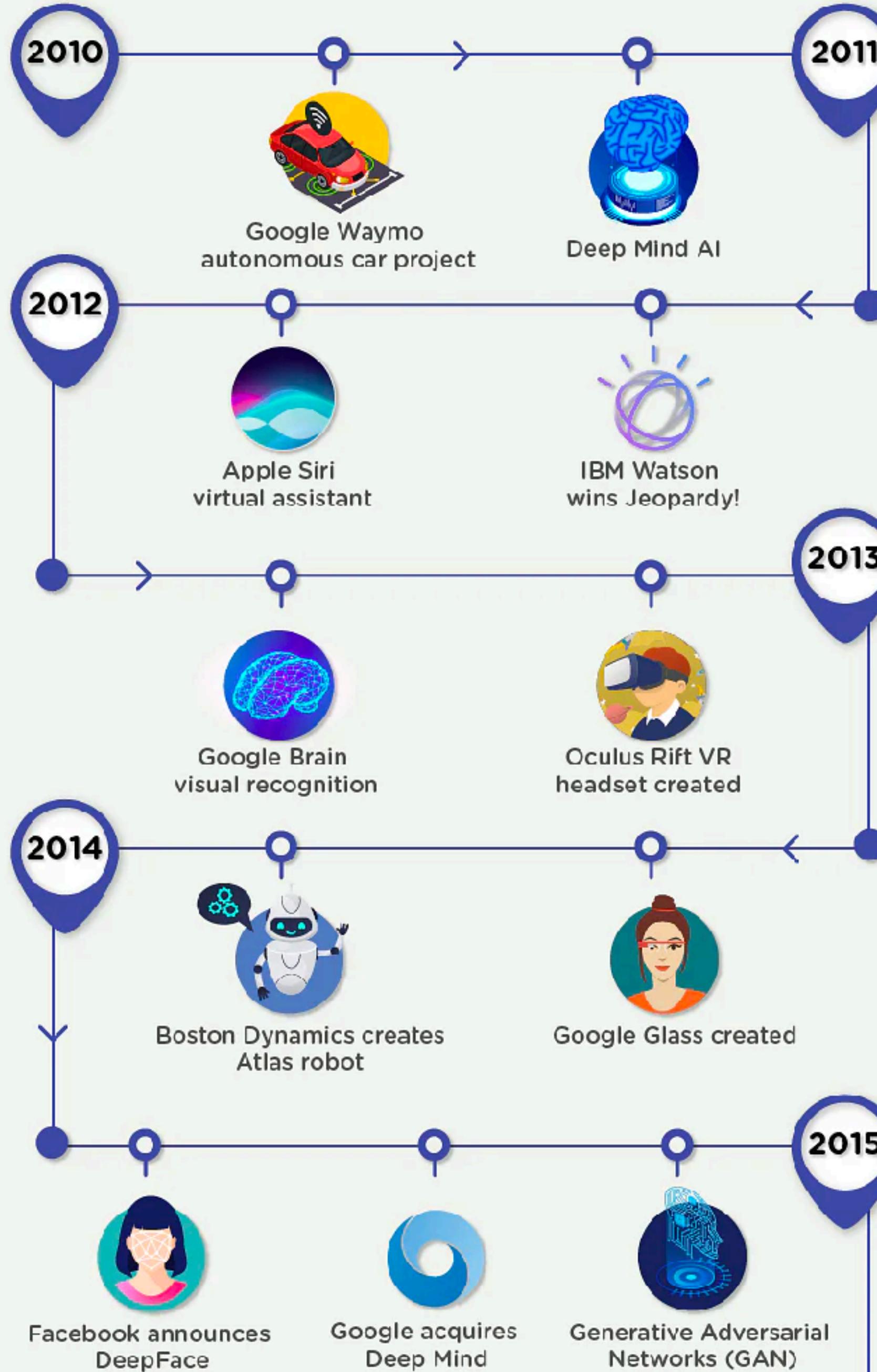
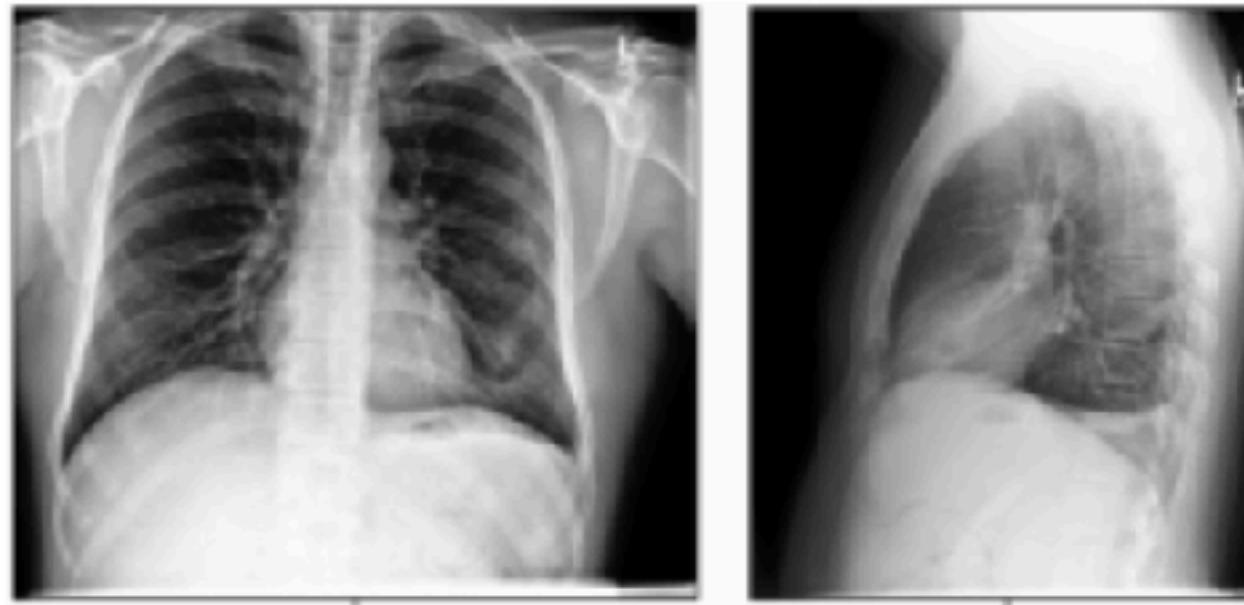


Image recognition

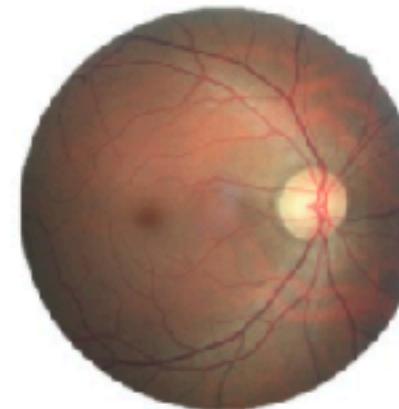
Chest radiology



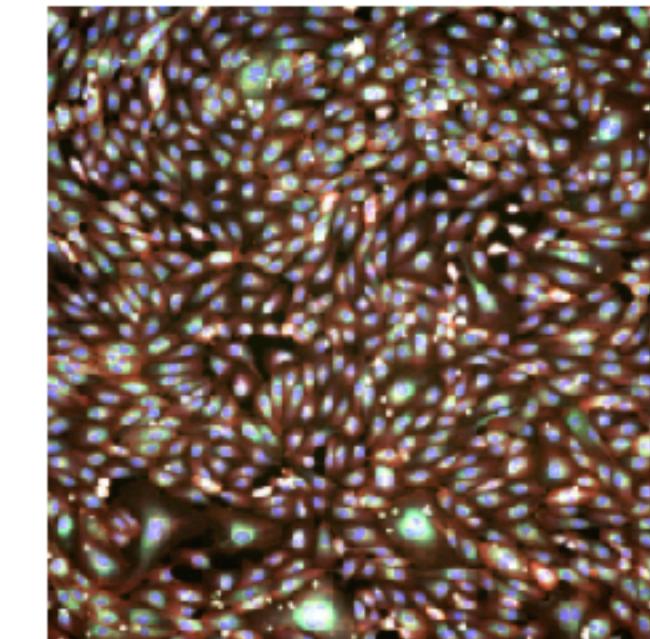
2019

**Google AI outperforms
radiologists at lung cancer
detection**

Diabetic retinopathy



Drug screening for COVID-19



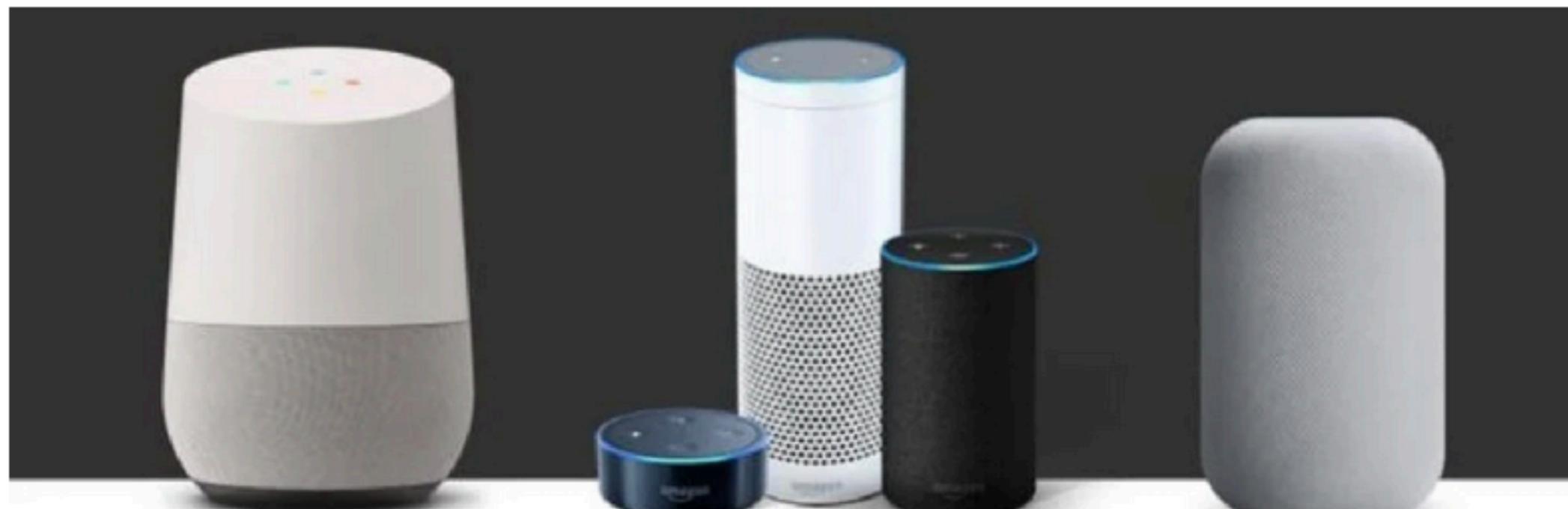


DEEPFAKE MORGAN FREEMAN

This is not Morgan Freeman - A look behind the Deepfake Singularity - <https://youtube.com/watch?v=F4G6GNFzoO8>

Natural Language Processing

Deep learning surpasses traditional approaches
in speech recognition in 2010



2016

2014

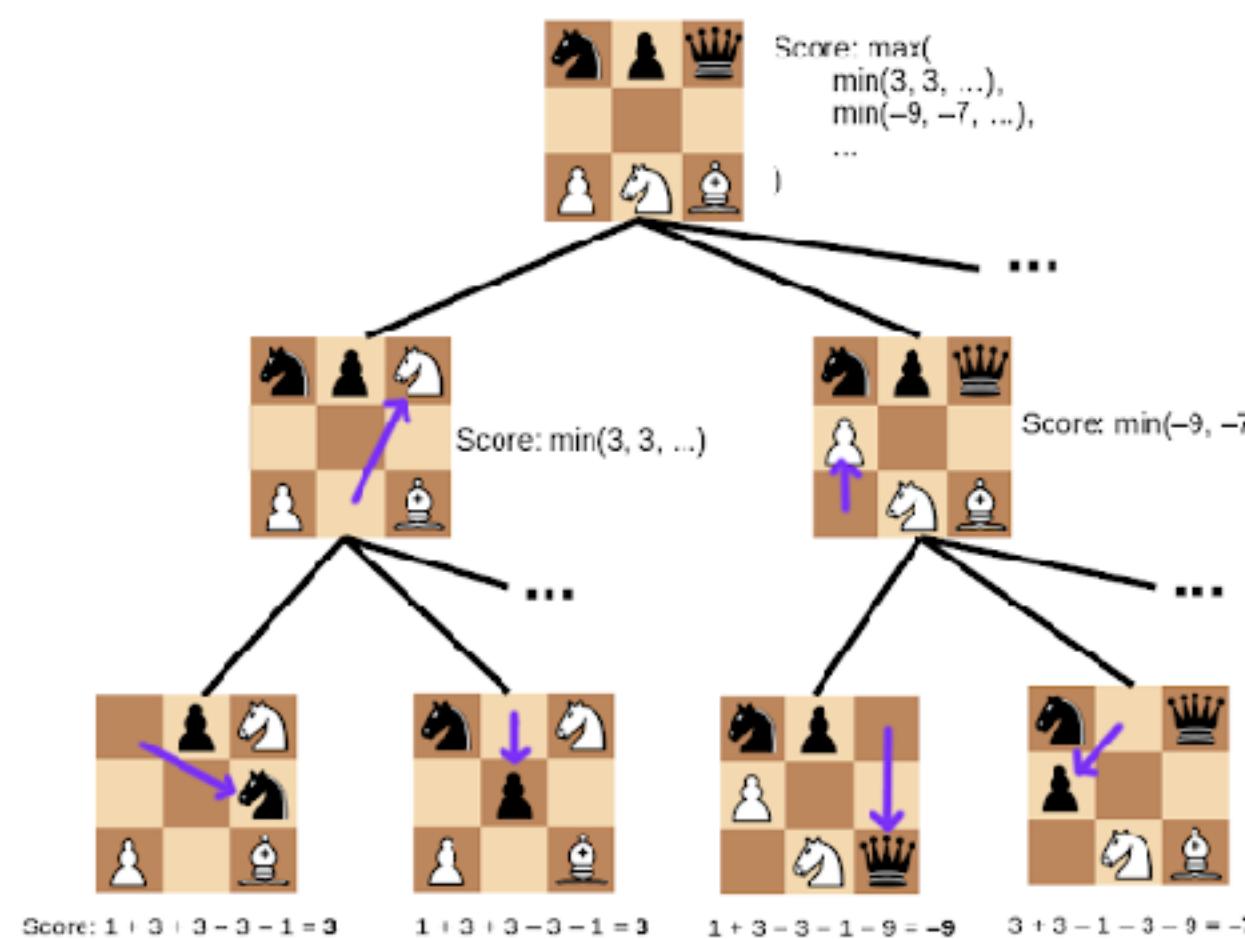
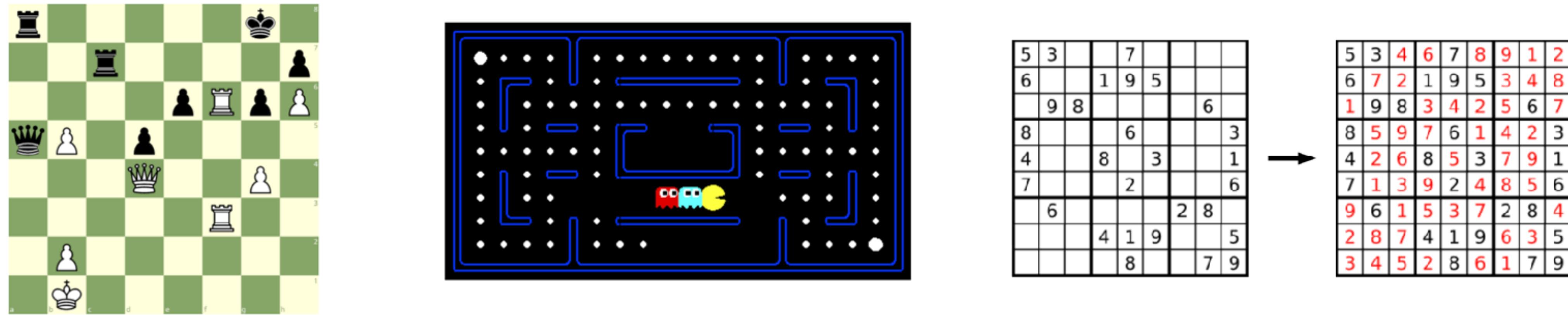
2011

IBM-Watson Defeats Humans in "Jeopardy!"

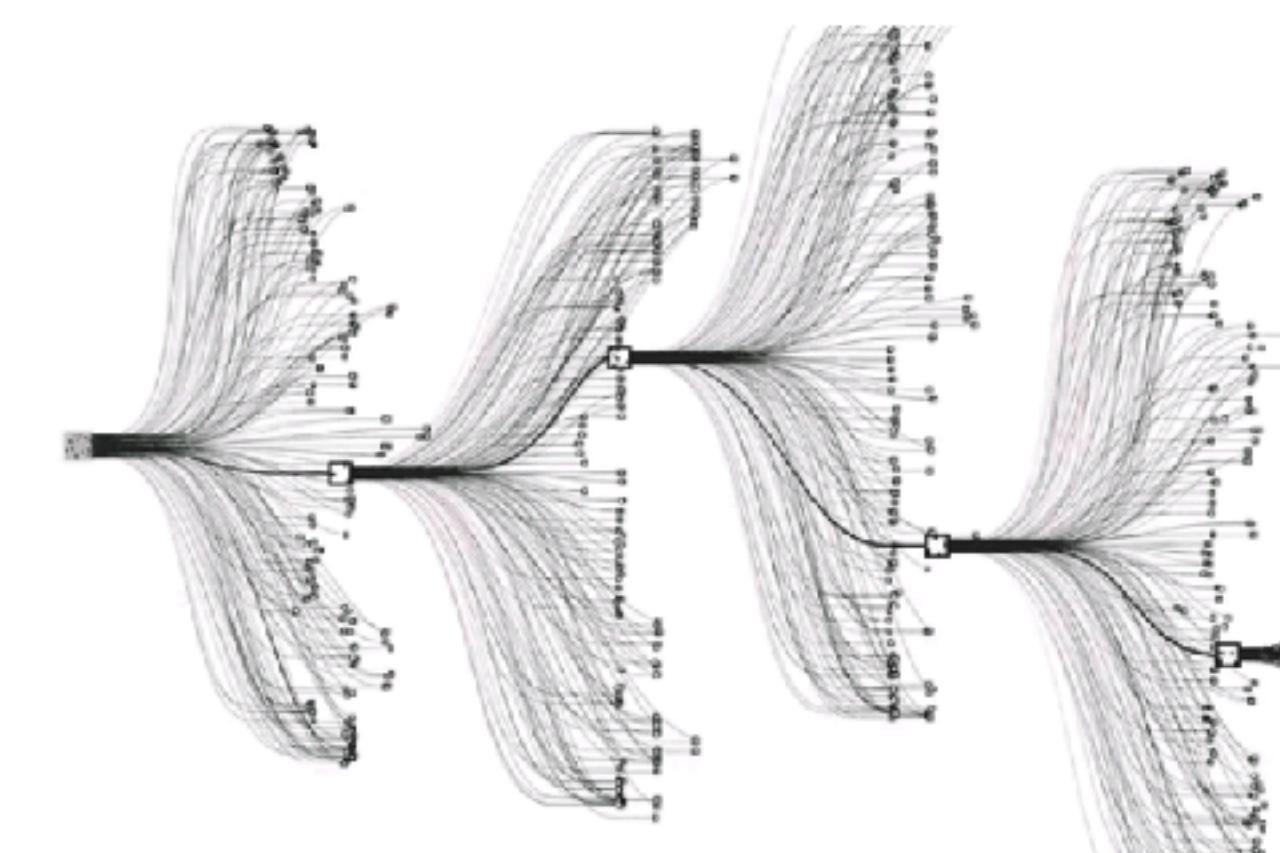


2011

Solving puzzles



10^{120} possibilities



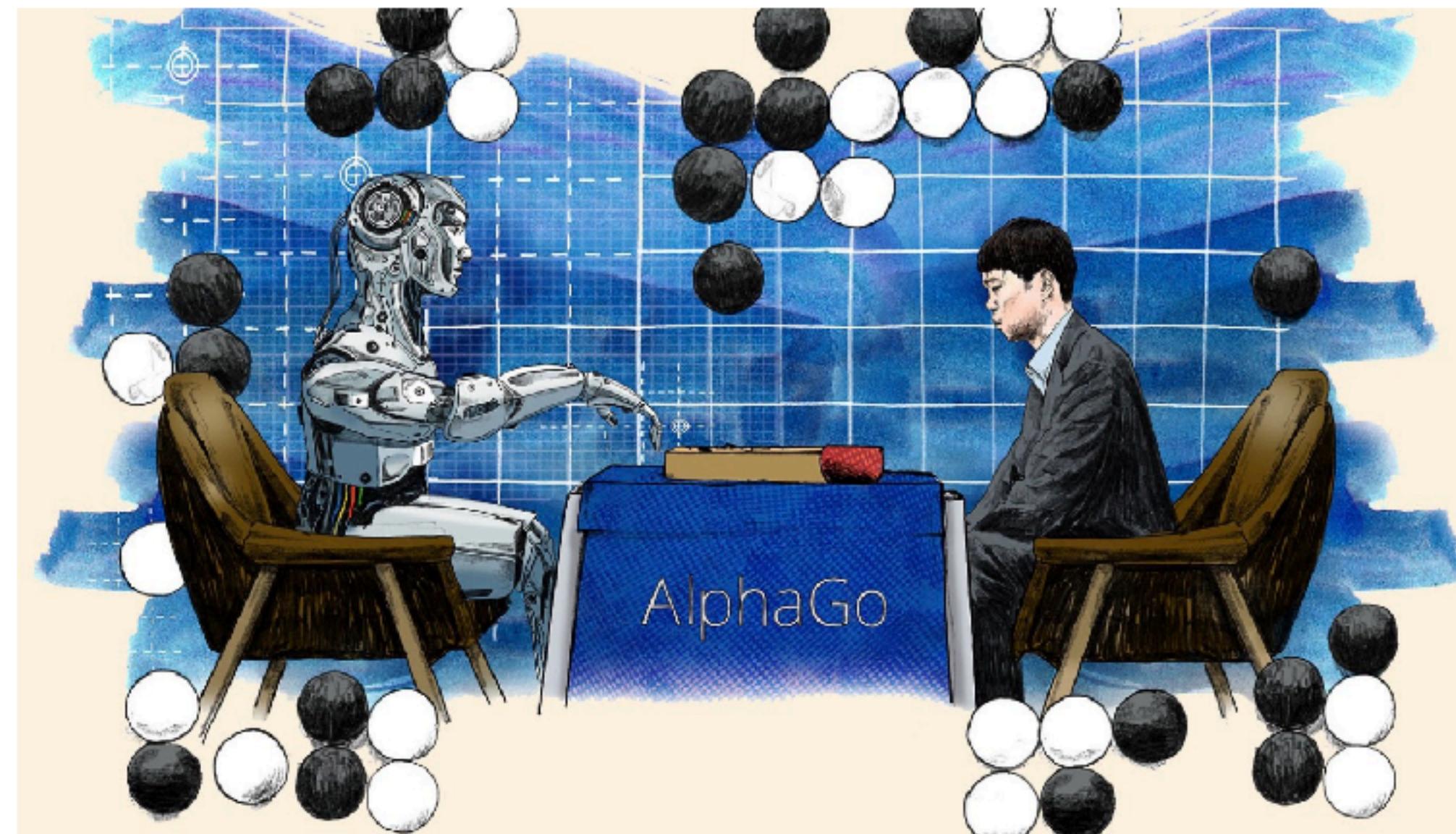
Games

IBM creates a checker playing program



1959

AlphaGo beats
world's best Go player



2016

IBM Deep Blue
beats Kasparov



1997

Trading

coursera Explore ▾ What do you want to learn? 

Browse > Data Science > Machine Learning

Machine Learning for Trading Specialization

Start Your Career in Machine Learning for Trading. Learn the machine learning techniques used in quantitative trading.

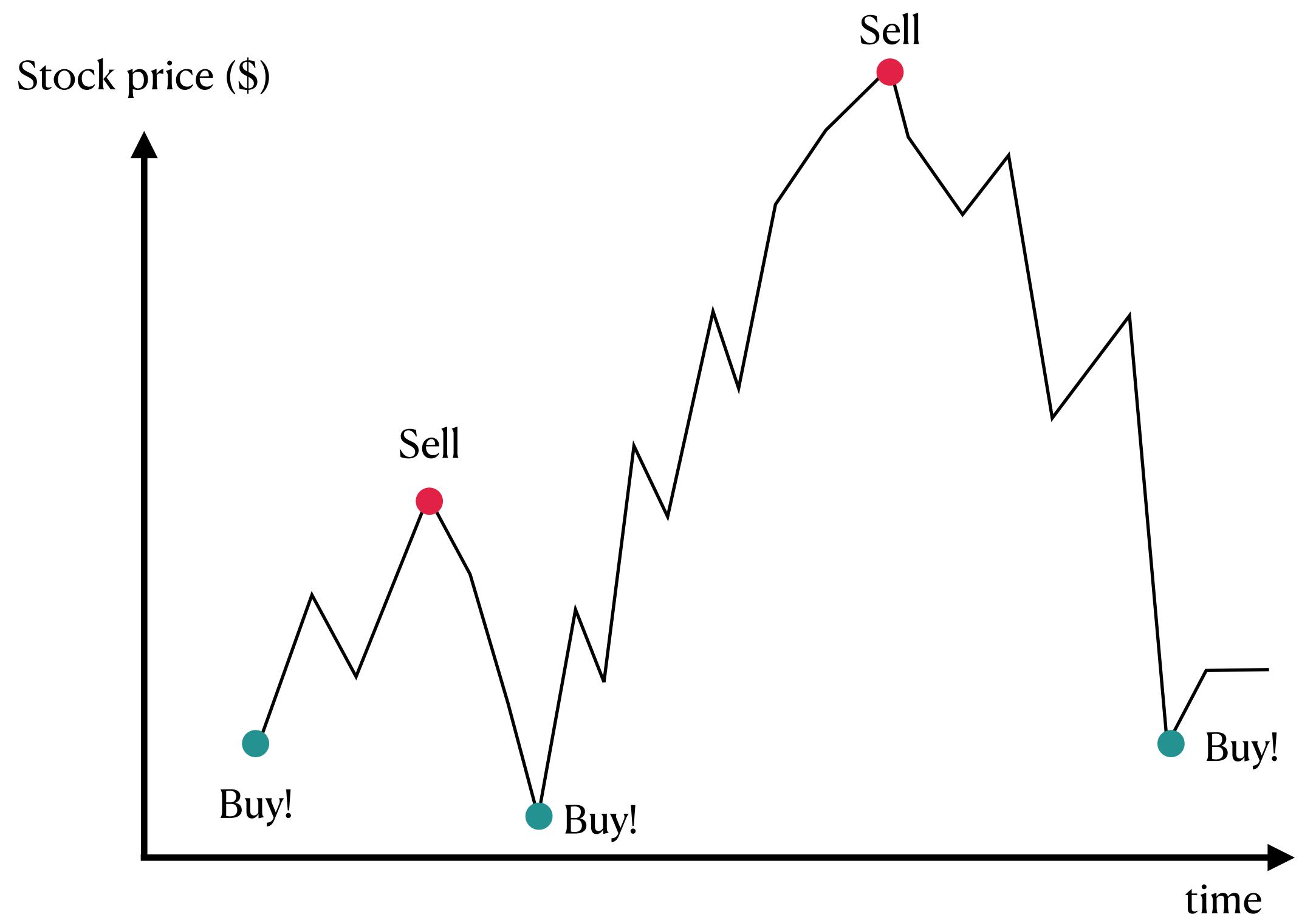
★★★★★ 3.9 795 ratings

 Jack Farmer [+1 more instructor](#)

Enroll for Free Starts Aug 9

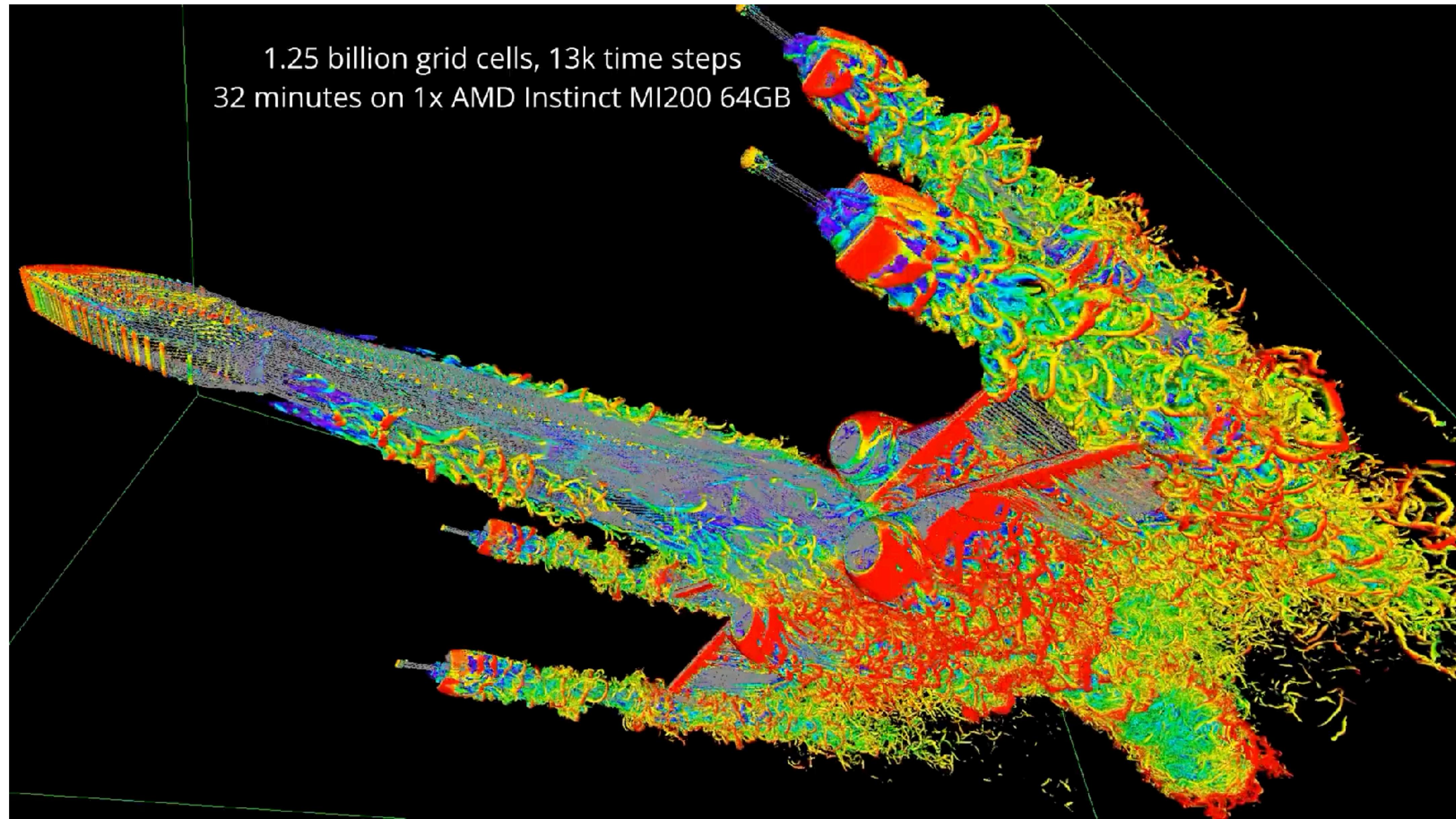
Financial aid available

14,817 already enrolled



**Can we automate
the process of scientific discovery?**

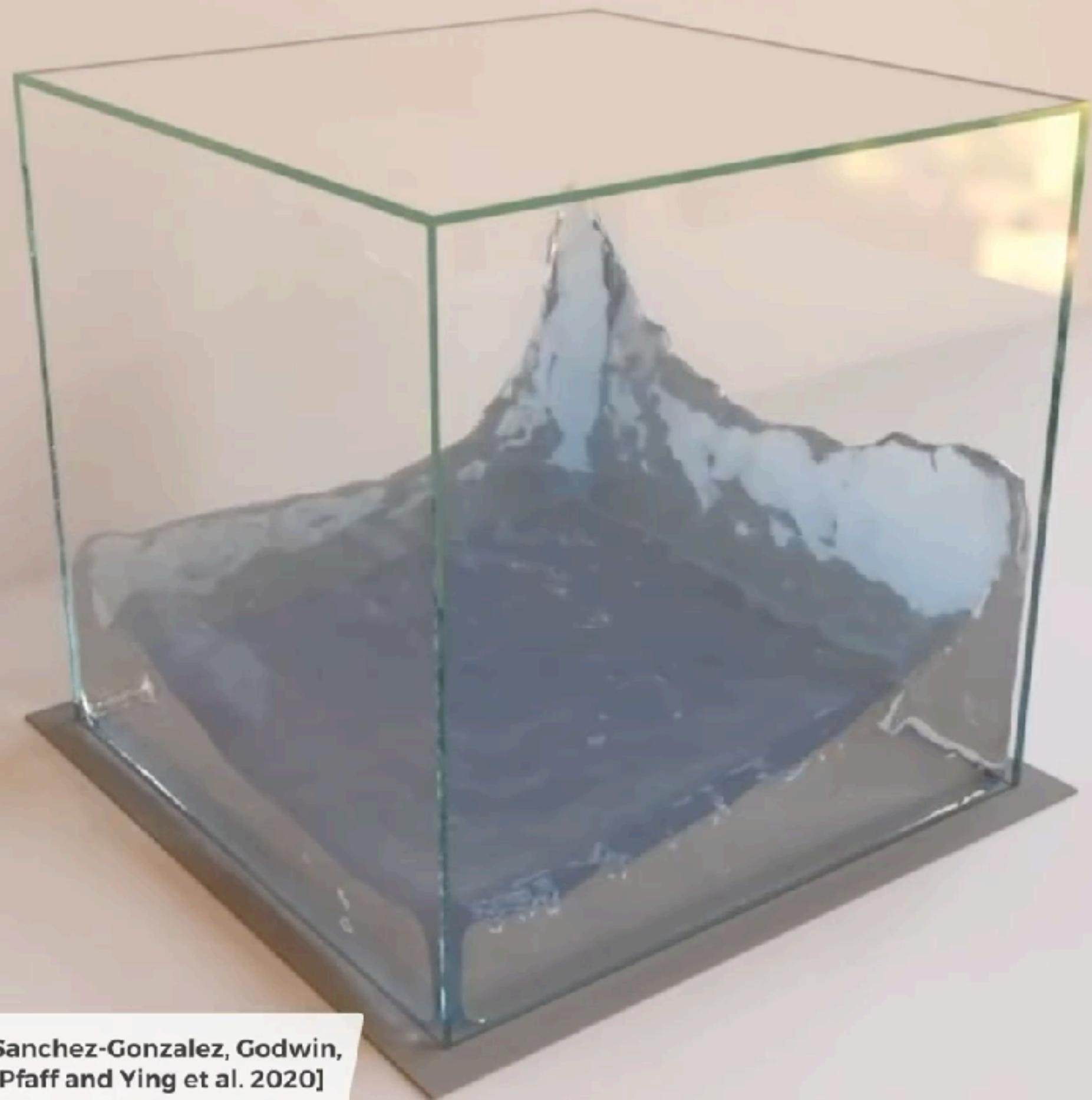
Fluid dynamics



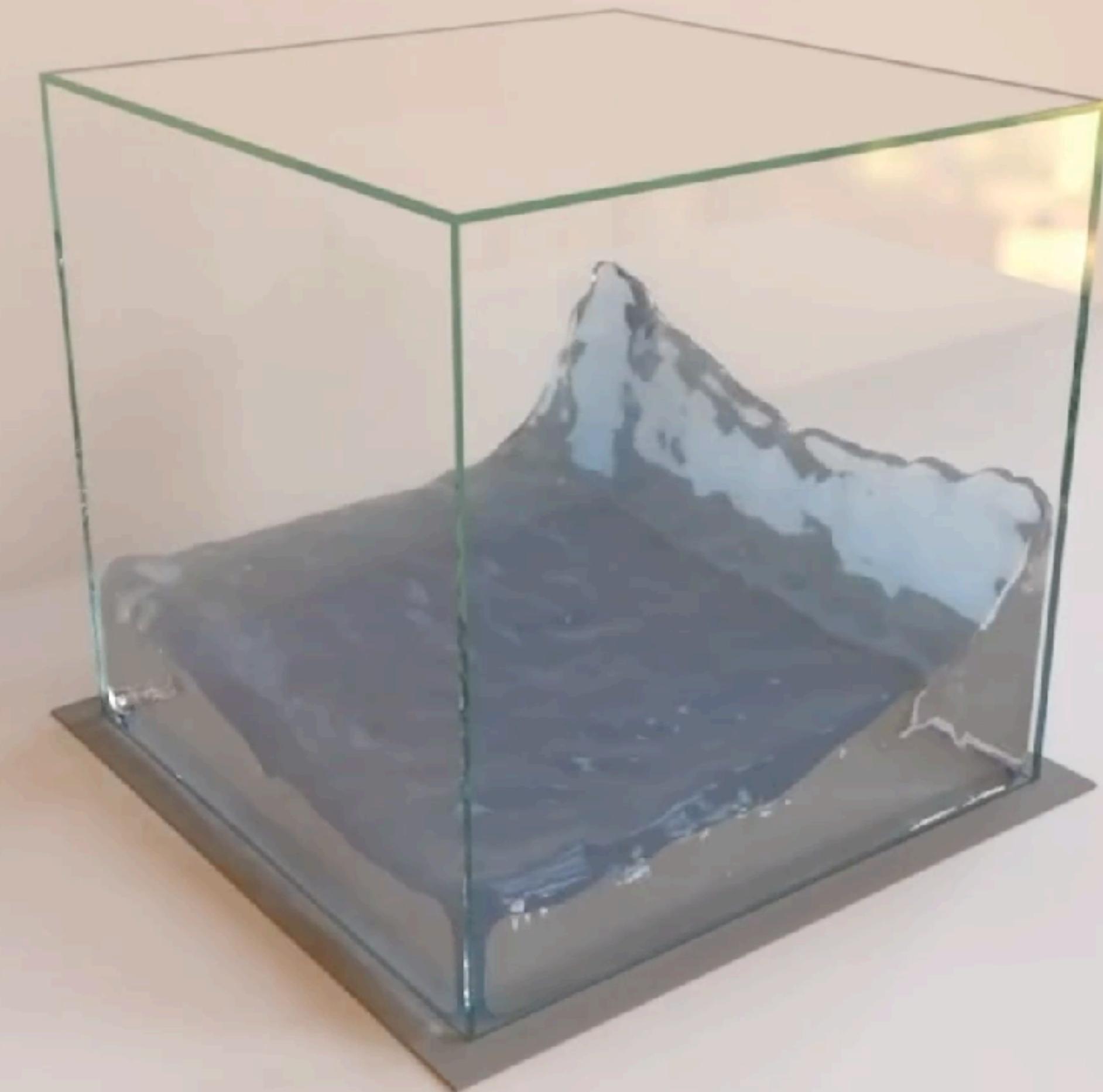
Fluid dynamics

Just to make sure: Ground truth = simulation, Prediction = AI

Ground truth



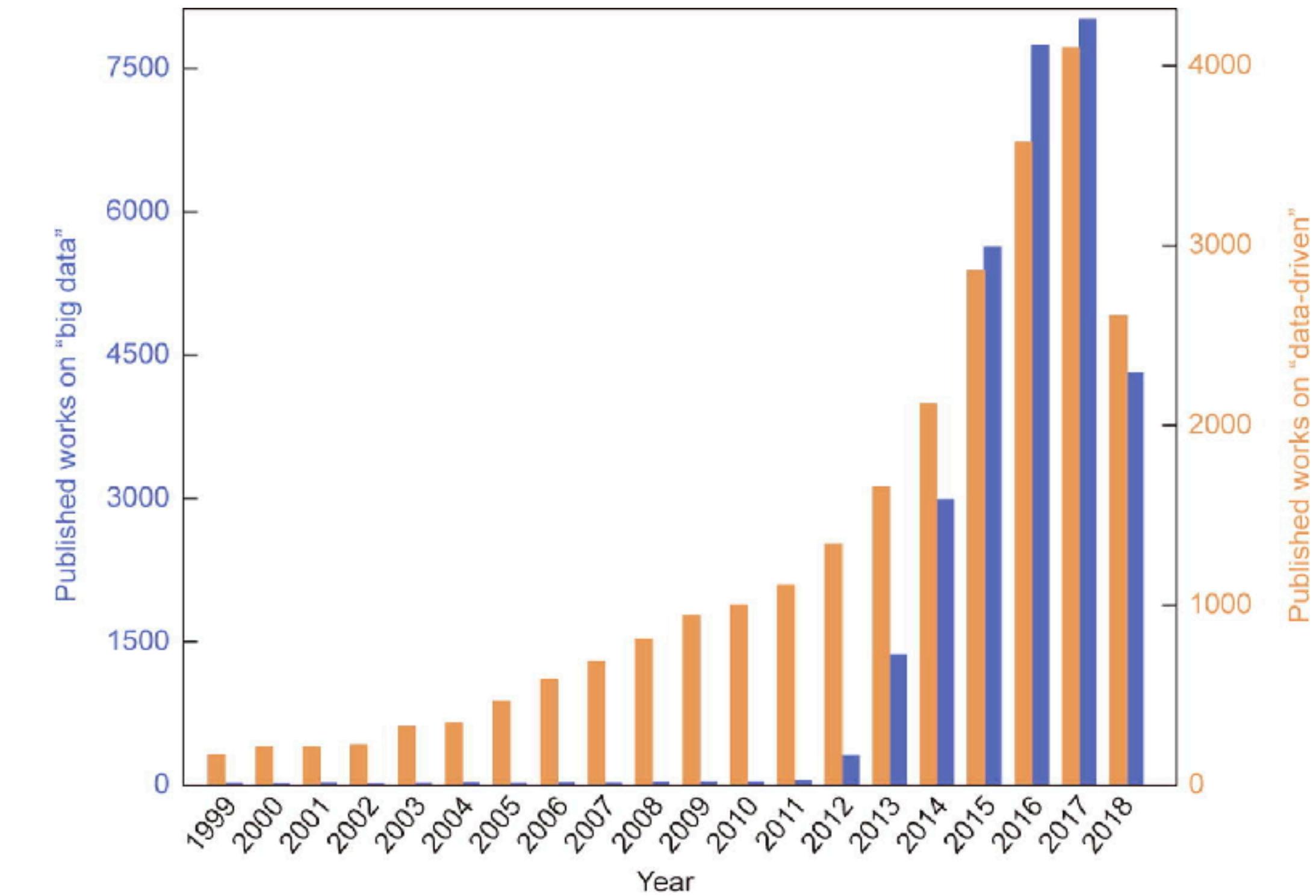
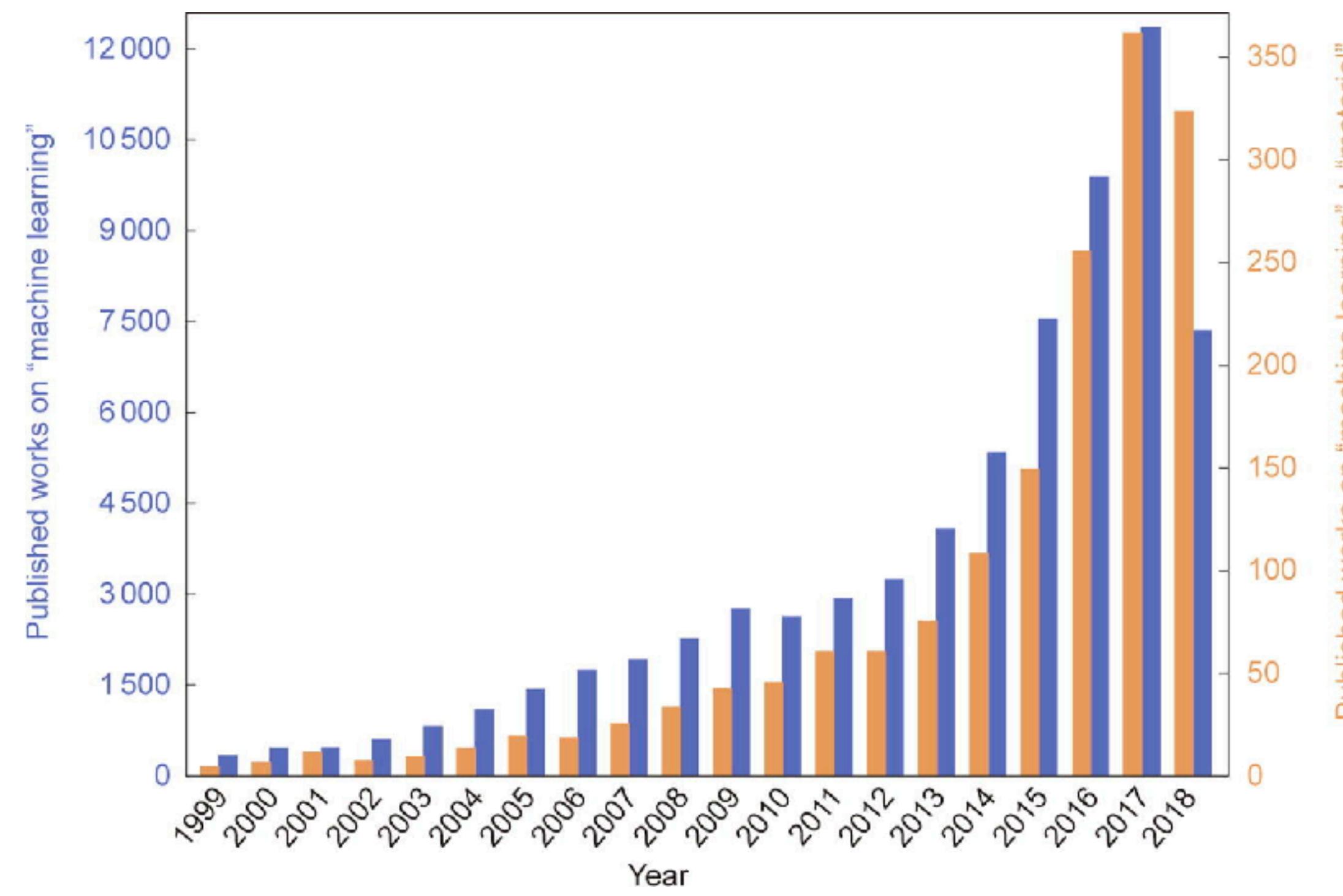
Prediction



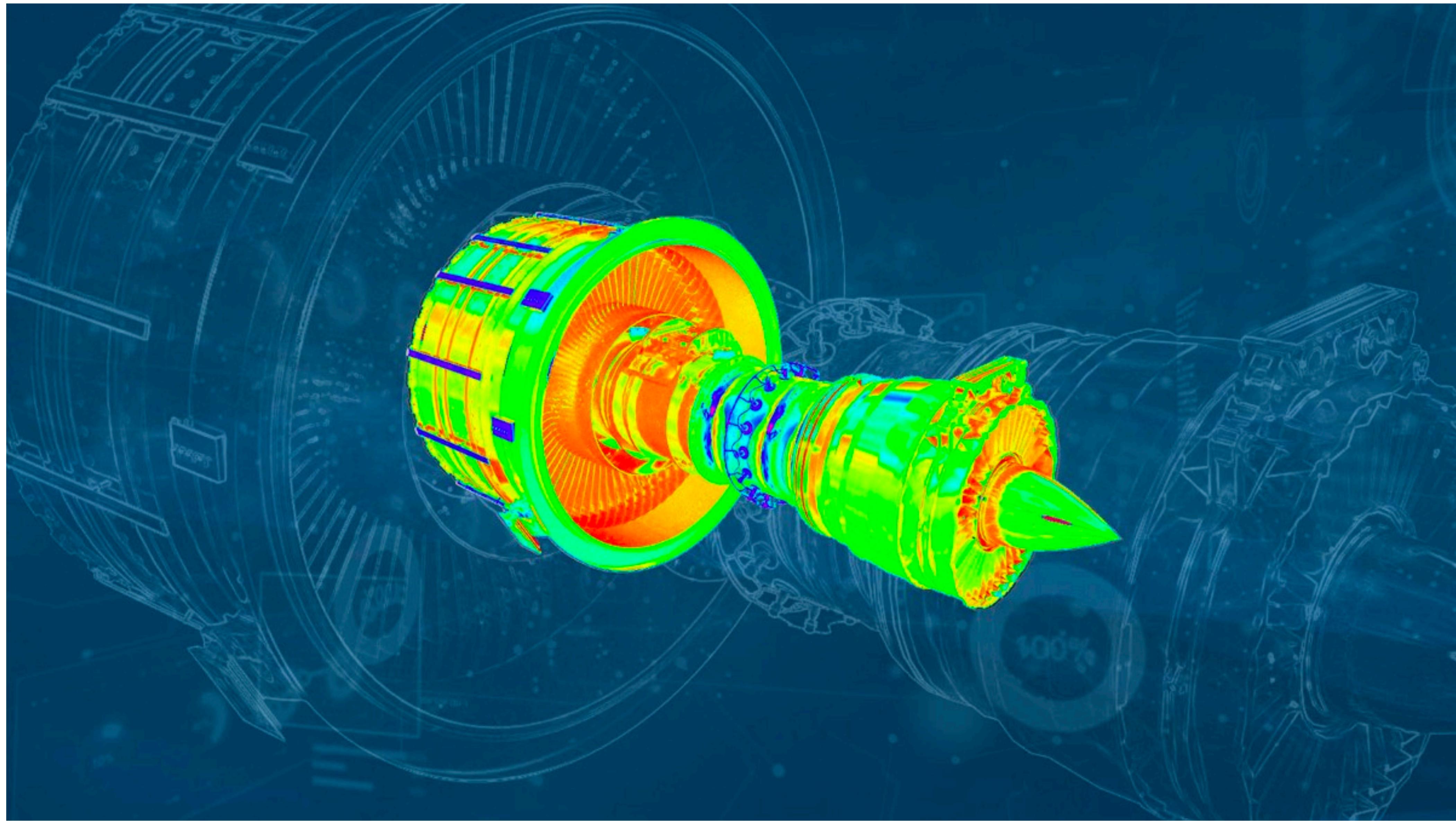
[Sanchez-Gonzalez, Godwin,
Pfaff and Ying et al. 2020]

Source

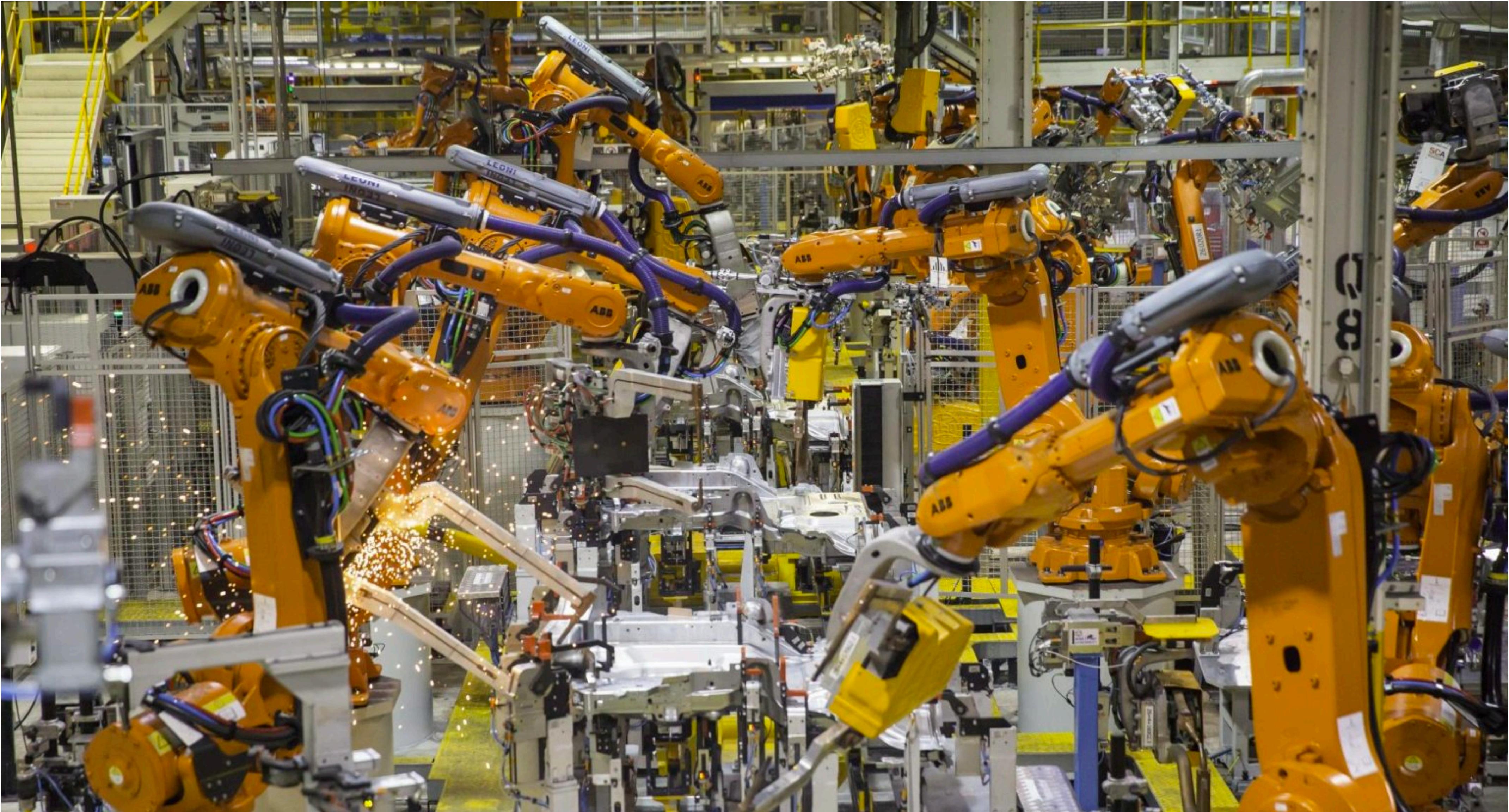
Machine learning in materials research



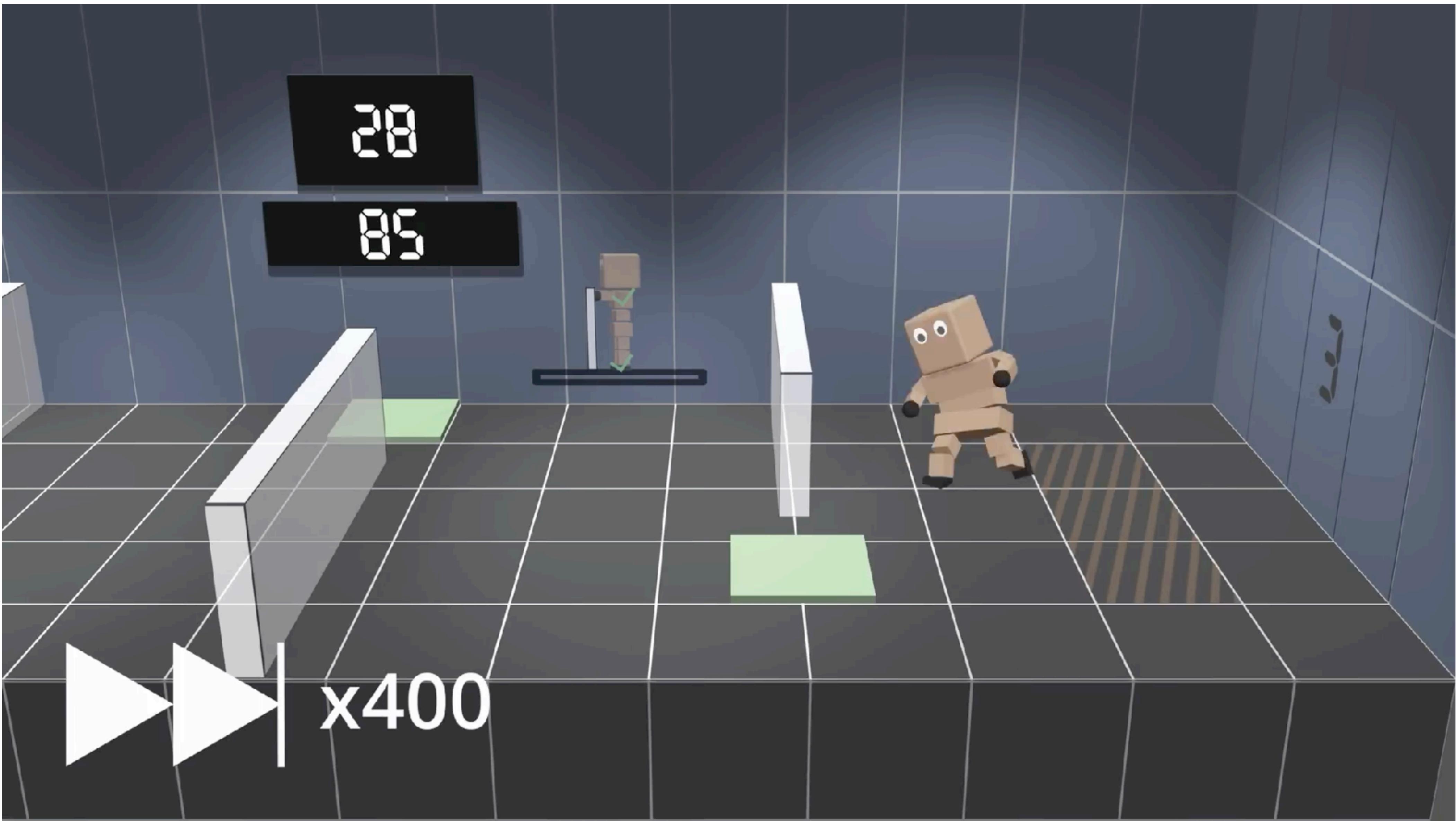
Machine learning in materials research



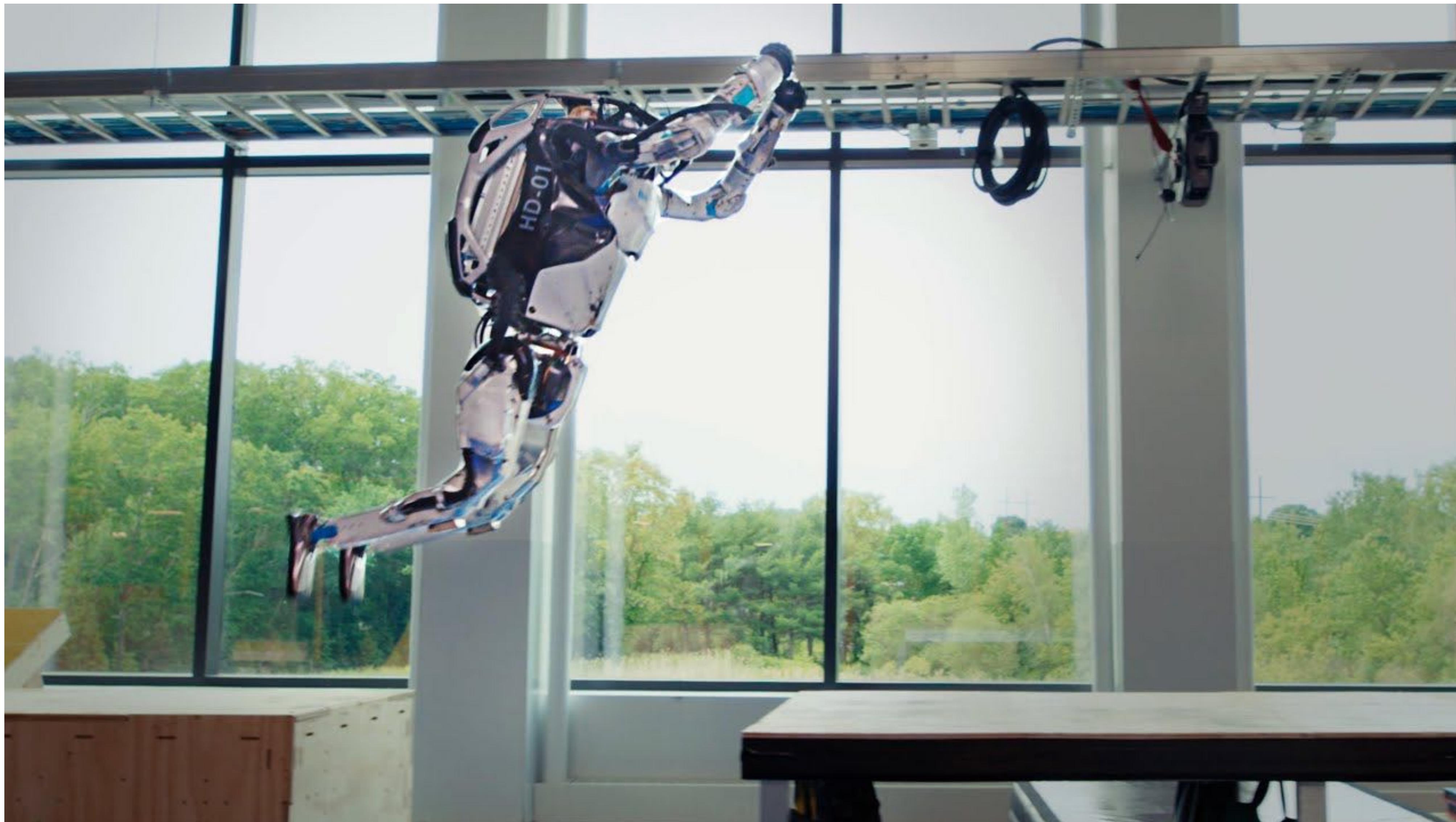
Robotics and Manufacturing



Robotics and games



Robotics and games



Boston Dynamics - <https://youtube.com/watch?v=tF4DML7FIWk>

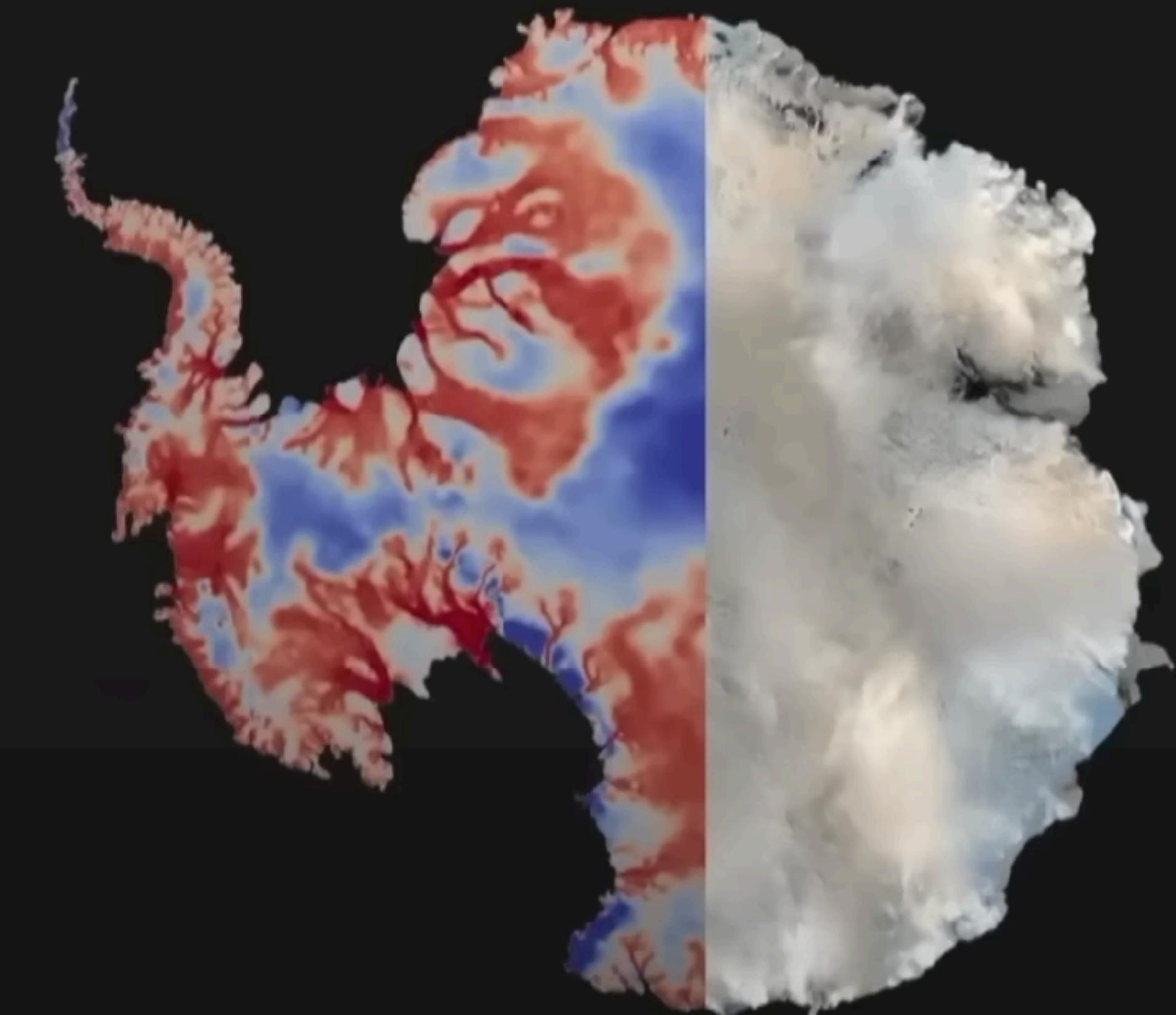
Self-driving cars



<https://youtube.com/watch?v=tlThdr3O5Qo>

Digital twin of an Ice Sheet

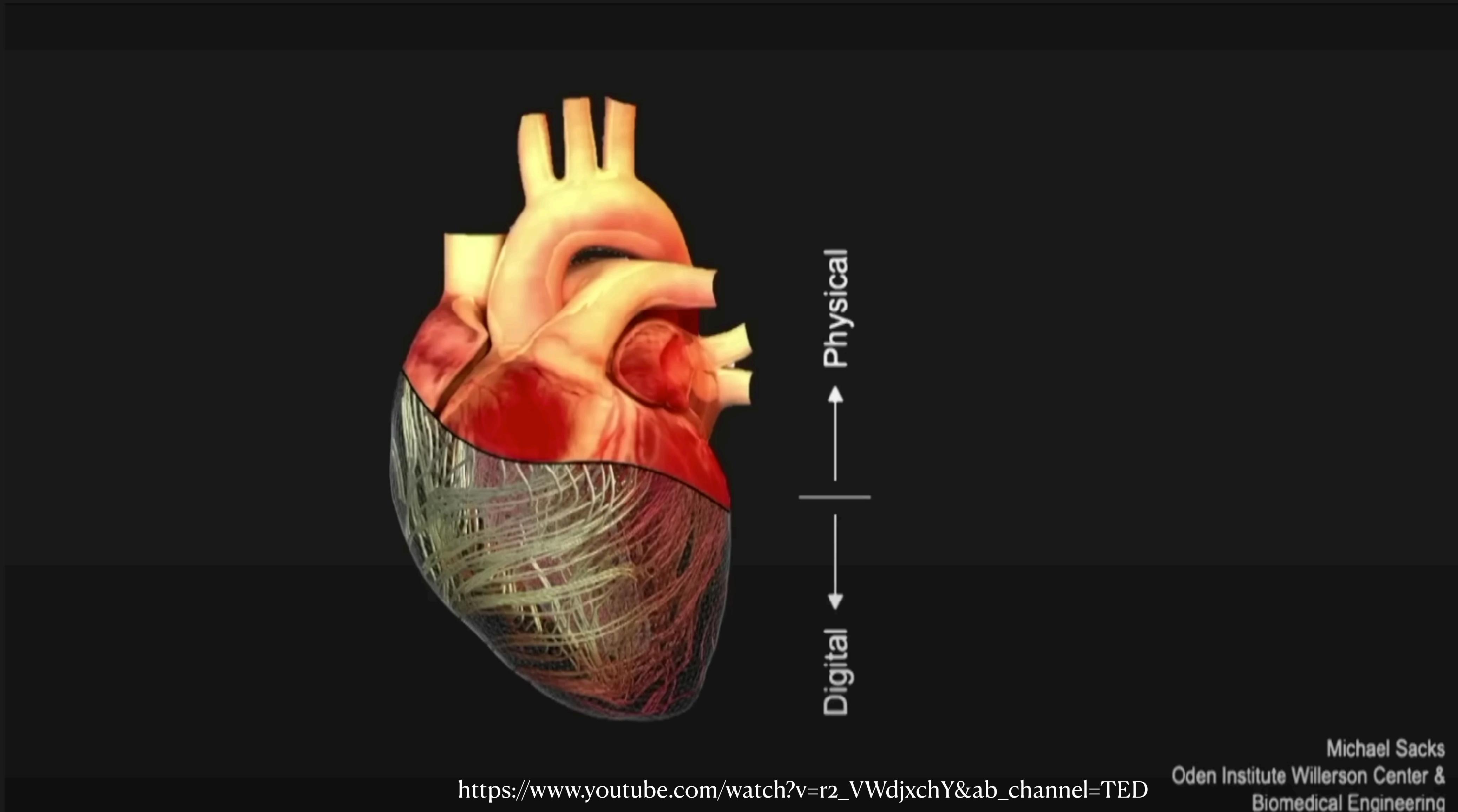
Digital ← → Physical



https://www.youtube.com/watch?v=r2_VWdjxchY&ab_channel=TED

Omar Ghattas
Oden Institute &
Jackson School of Geosciences

Digital twin of a heart?



https://www.youtube.com/watch?v=r2_VWdjxchY&ab_channel=TED

Michael Sacks
Oden Institute Willerson Center &
Biomedical Engineering

Biomedical Engineering

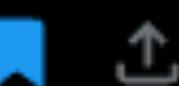
 **Mindaugas Galvosas, MD** @MGalvosas · Jan 18 ...

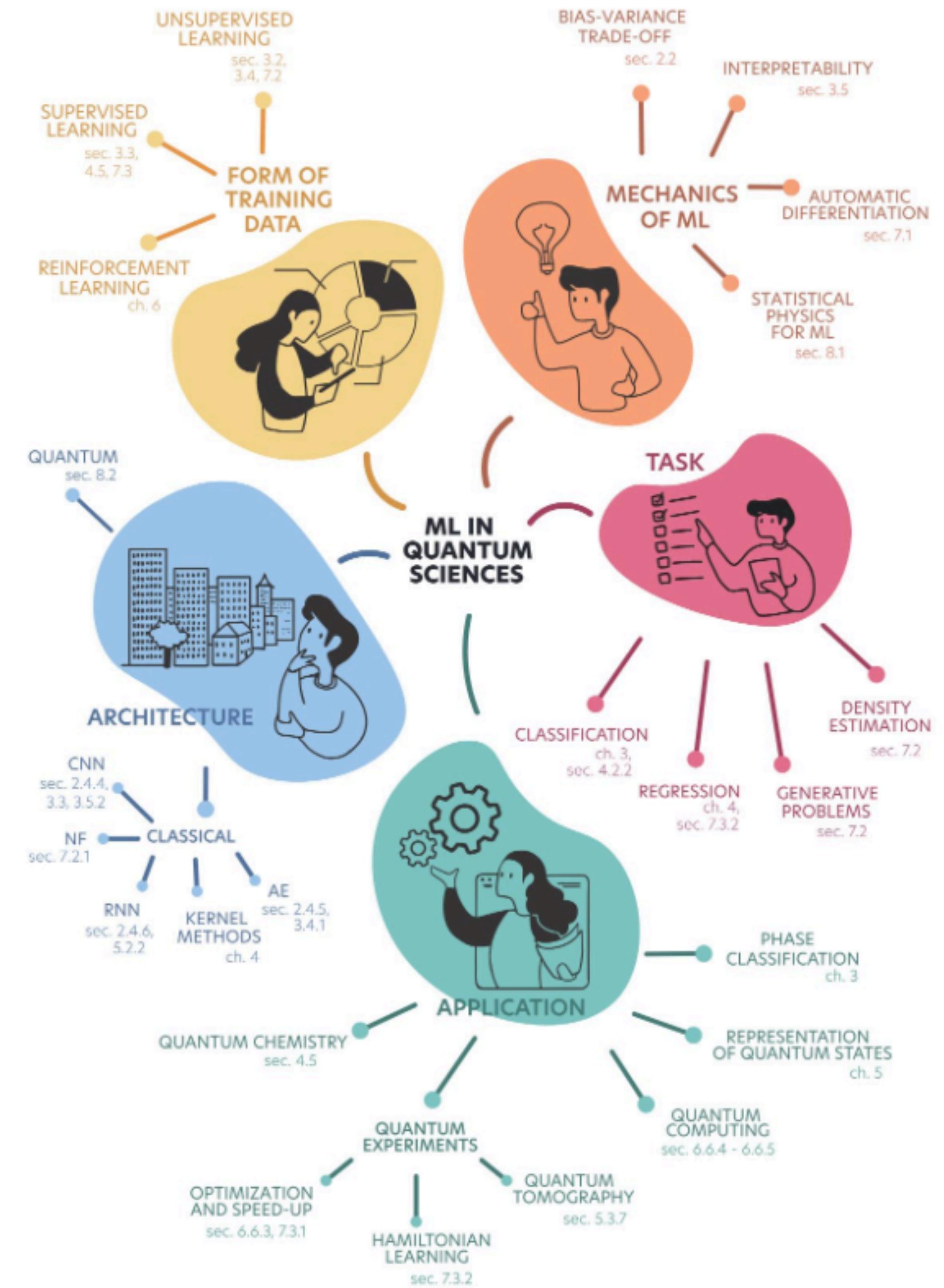
FDA just cleared the first AI device detecting all major skin cancer by DermoSensor.

Its pivotal trial showed sensitivity of 96% and a 97% chance of accurately identifying a skin lesion as benign.

Notably, the approval process began back in 2016.



Q 42 t 248 C 1K I 257K 



VI - Course Logistics

Course Objectives

- Gain experience in getting computers to **Model the World from Data**
- Master computational and data-driven thinking **with scientific data**
- Develop a better intuition for scientific modeling through examples
- Learn powerful machine learning algorithms
- Learn and develop machine learning tools **for scientific computing and discovery**

Course Details

- **Pre-requisites:**
 - MATH 218 or 219 - Linear Algebra
 - MATH 201 - Calculus
 - STAT 230 - Probability and Statistics
 - EECE 230 - Introduction to Programming
 - EECE 490 - Machine Learning – Not officially a pre-requisite but highly recommended
- **References:**
 - Stanford's **CS229**: Introduction to Machine Learning Notes
 - Course website: www.ml4science.com
 - Data-Driven Science and Engineering - by Brunton and Kutz

Course Details

- **Office Hours and Contact:**
 - Bechtel 418: Thursday 3:30-5:00pm (or by apt) - Monday: 1:00-2:00pm
 - Email: jb50@aub.edu.lb
- Lecture Schedule:
 - T-Th: 2pm-3:15 pm
 - There will be practical problem solving lab days: **bring your laptops!**
- Getting Assistance:
 - Teaching Assistants: **Khalil Elhaj, Joe Germany, Ali Tayyar, Issar Amro (?)**
 - You'll get a **Slack** invitation

Grades

- **Grades Distribution:** (detailed distribution might vary)
 - Short Quizzes (10%)
 - Assignments (25%)
 - Late Midterm Exam (25%)
 - Group project (40%)
 - Proposal (5%)
 - Final Report (30%)
- **Homework policy:** discuss the homework, help each other understand the material, **BUT** write the solution on your own. Copied assignments will be penalized.
- **Project:** find a topic, groups of **2 to 3**, put code on **GitHub**, make a **product**, final presentation.
- See course policy on website: www.ml4science.com

Generative AI Policy

- Let's start with a basic question: why are you here?
 - Acquire the confidence to **think on your own**
 - Acquire a deep understanding of useful **tools** (LLMs included!)
 - Use computational tools to **create new science/engineering**
- Given these goals, can you use LLMs in this course?
 - **Yes**, you have to use them in the future to solve some complex problems or accelerate your work. Treat them as smart friends with a lot of knowledge; and like friends they make mistakes. So, always keep a (very) healthy amount of skepticism in mind.
 - **But**, you have to be fully aware that it might make you **dumber**. For example, if it does all the thinking for you, copy paste its answers without understanding. In this case, you will make it much easier for it to replace you.
 - Your future's at stake and it's your responsibility to make sure that you're learning.

Tips for effectively using LLMs

- **Learn how to ask good questions:** before consulting an LLM, try to solve and, whenever possible, define the problem on your own. Use LLMs to empower your thinking process, not replace it.
- **Cross-check with reliable sources:** Knowing when to trust LLM content is the same as learning how to trust social media posts. It's critical thinking. It comes down to having a good model of the world (in your head), to know what is plausible, and when you should ask more questions.
- **Learn from it:** if the LLM provides a solution, take the time to understand the steps and rationale behind it. You can effectively learn anything if you ask the right questions, and you're curious enough.
- **Always aim for higher quality:** evaluating the quality of your work is as important as creating it. In the long run, this will distinguish you in the job market.

Discovering Functions From Data

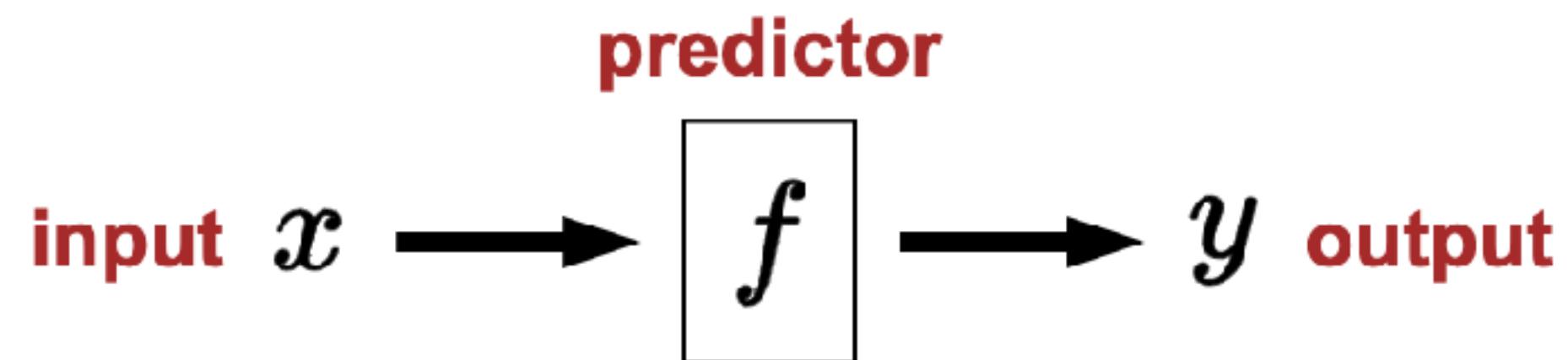
Machine Learning

$$f(x) = 2^x$$

x	0, 1, 2, 3, 4, 5
y	1, 2, 4, 8, 16, ?

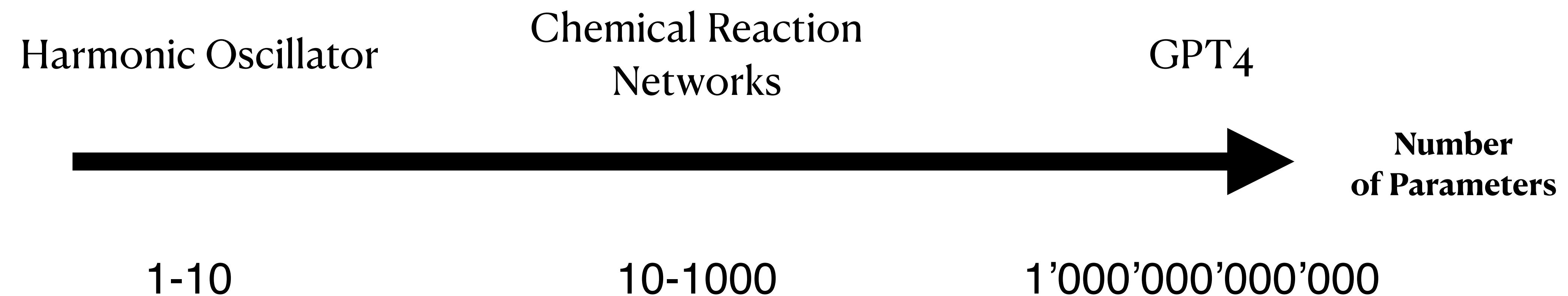
$$\begin{matrix} x & y \\ \curvearrowright & \end{matrix}$$

$$f(x) = 2x$$

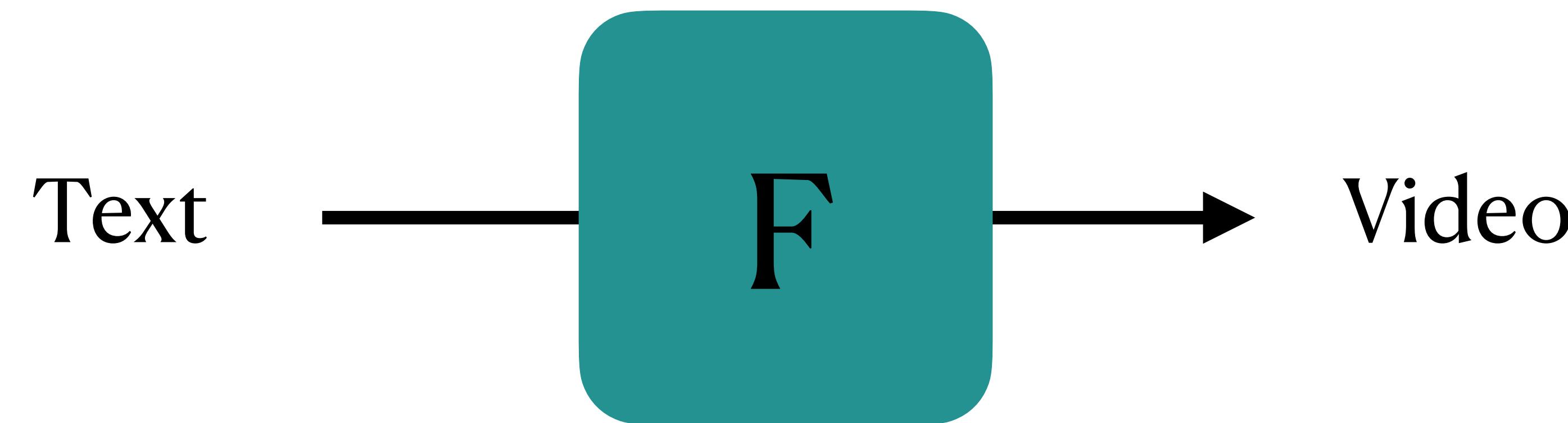


Better algorithms

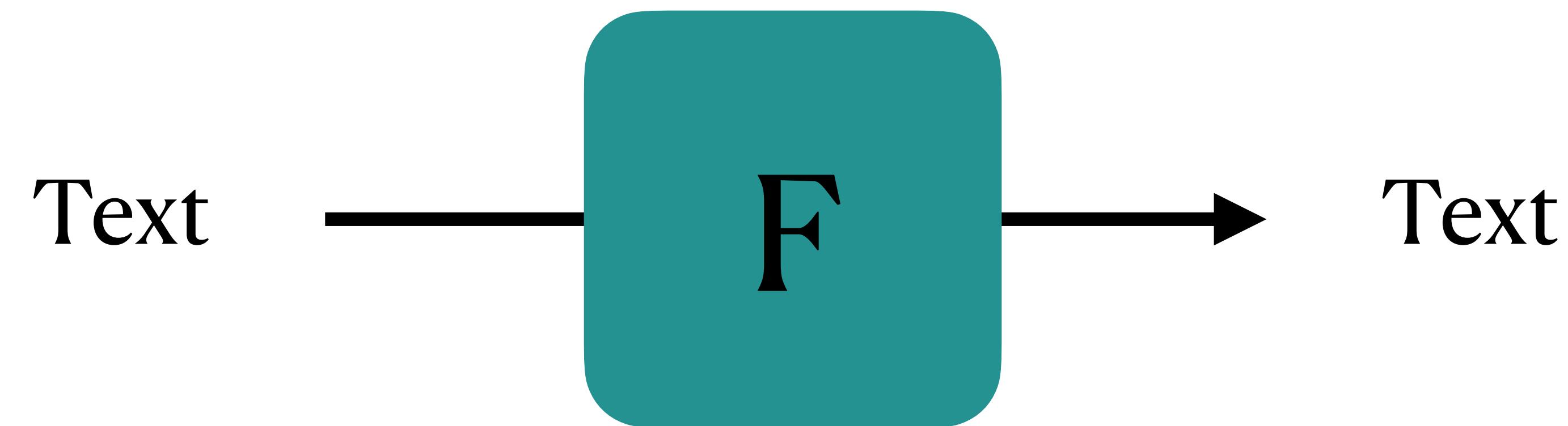
Bigger Models



Sora - Text-to-Video

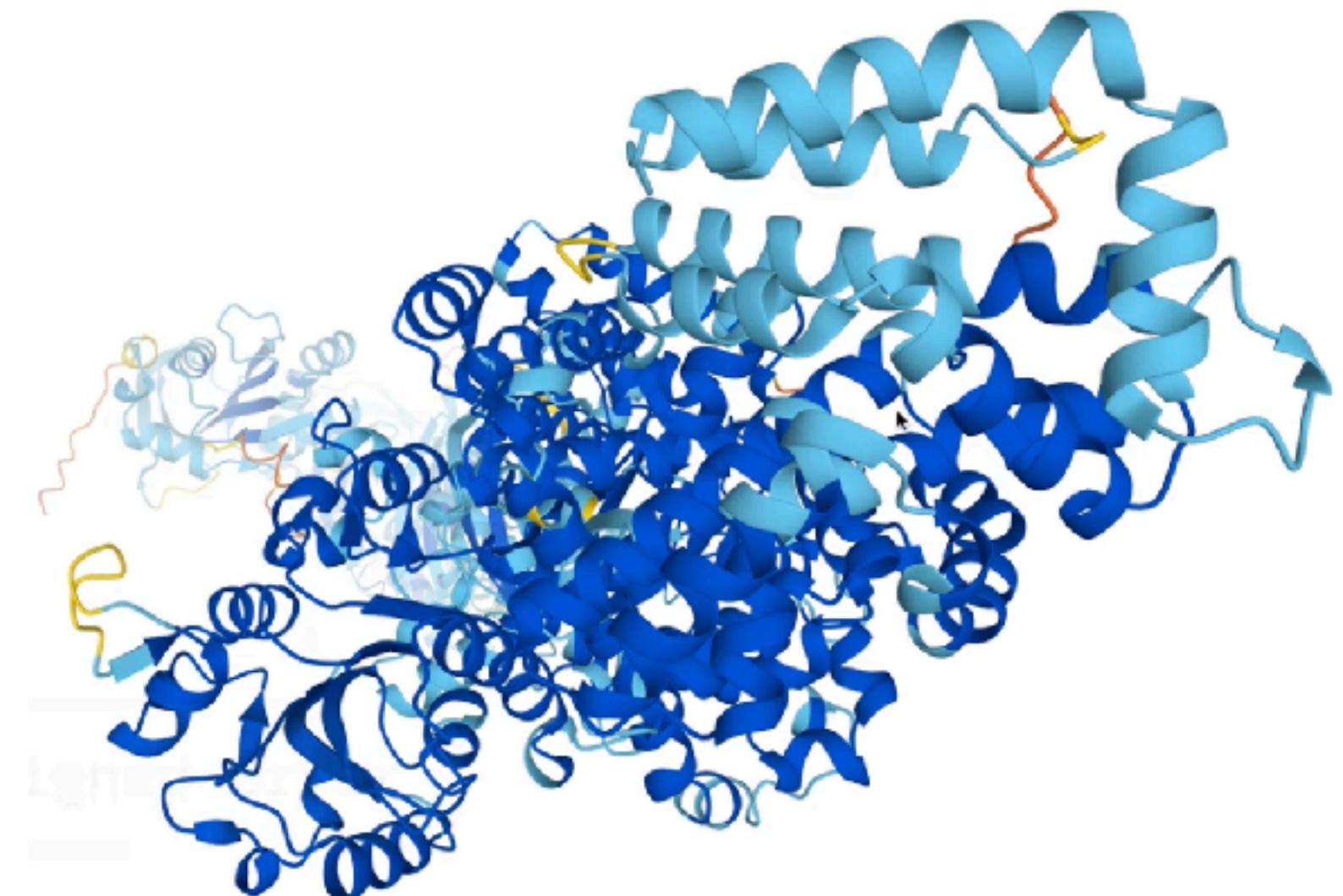
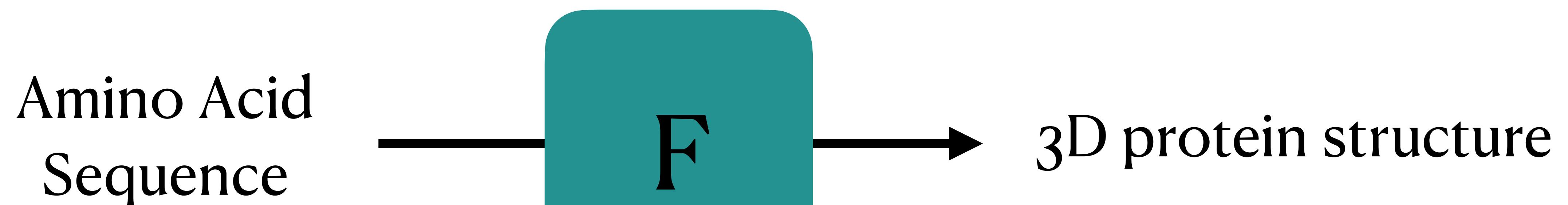


ChatGPT

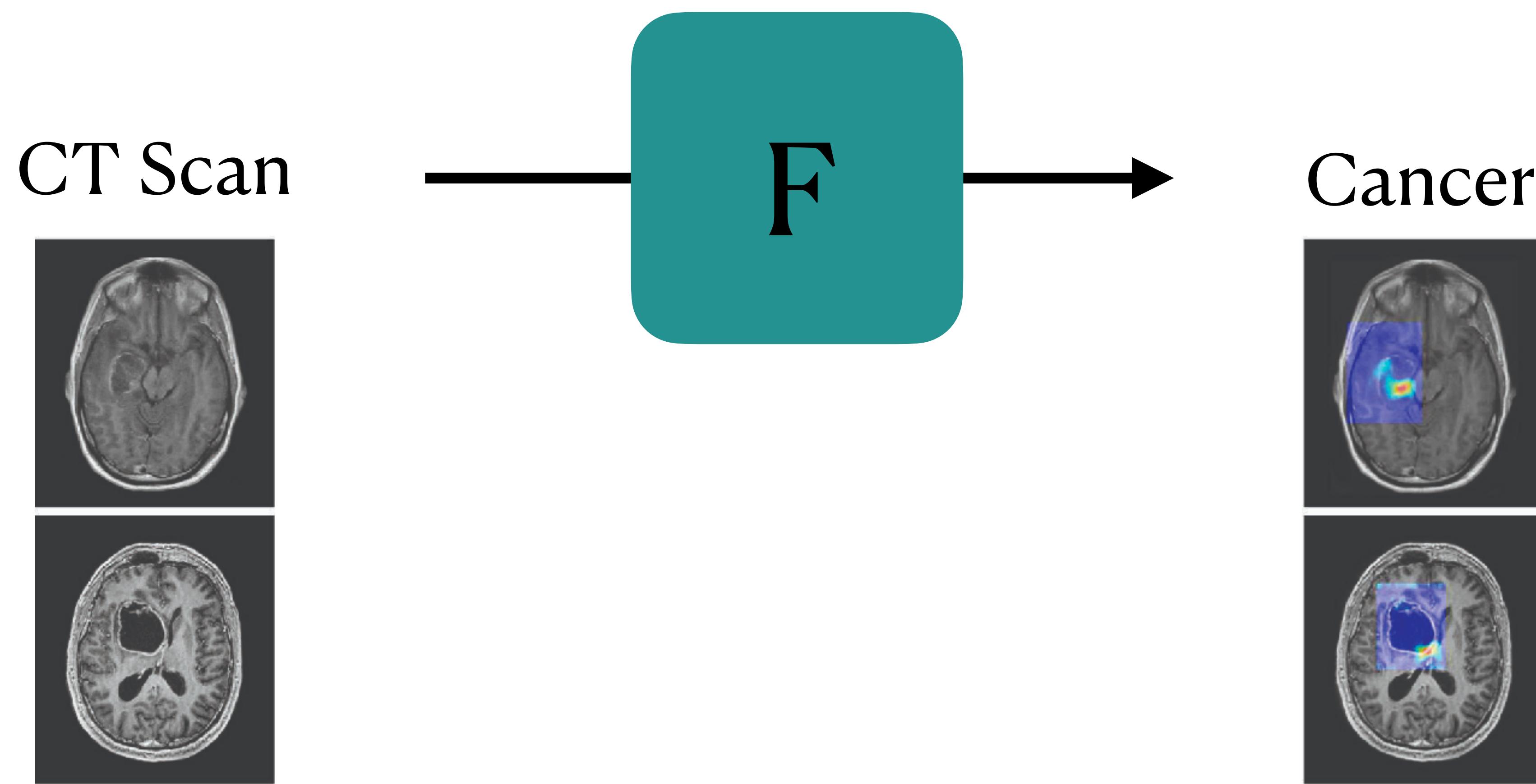


AlphaFold is an AI system developed by **DeepMind** that predicts a protein's 3D structure from its amino acid sequence. It regularly achieves accuracy competitive with experiment.

AlphaFold



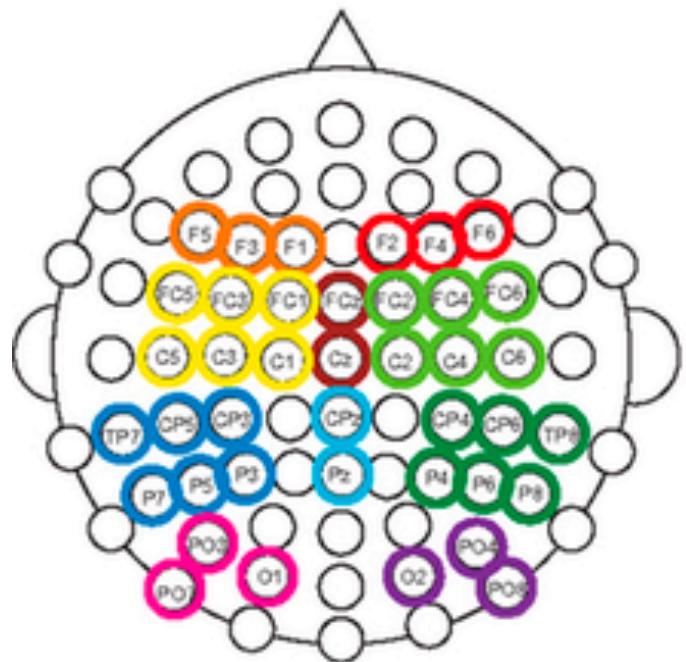
Healthcare



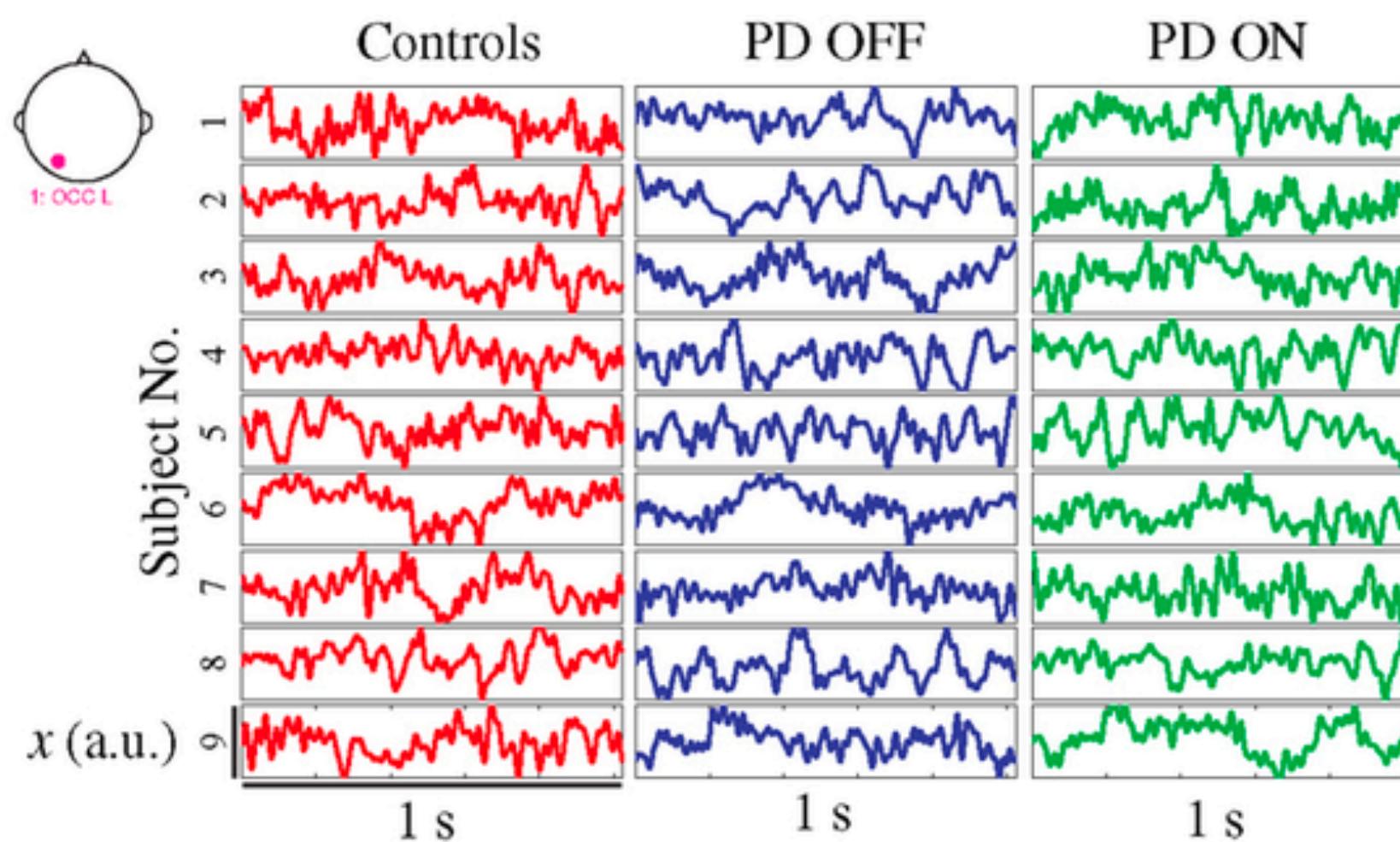
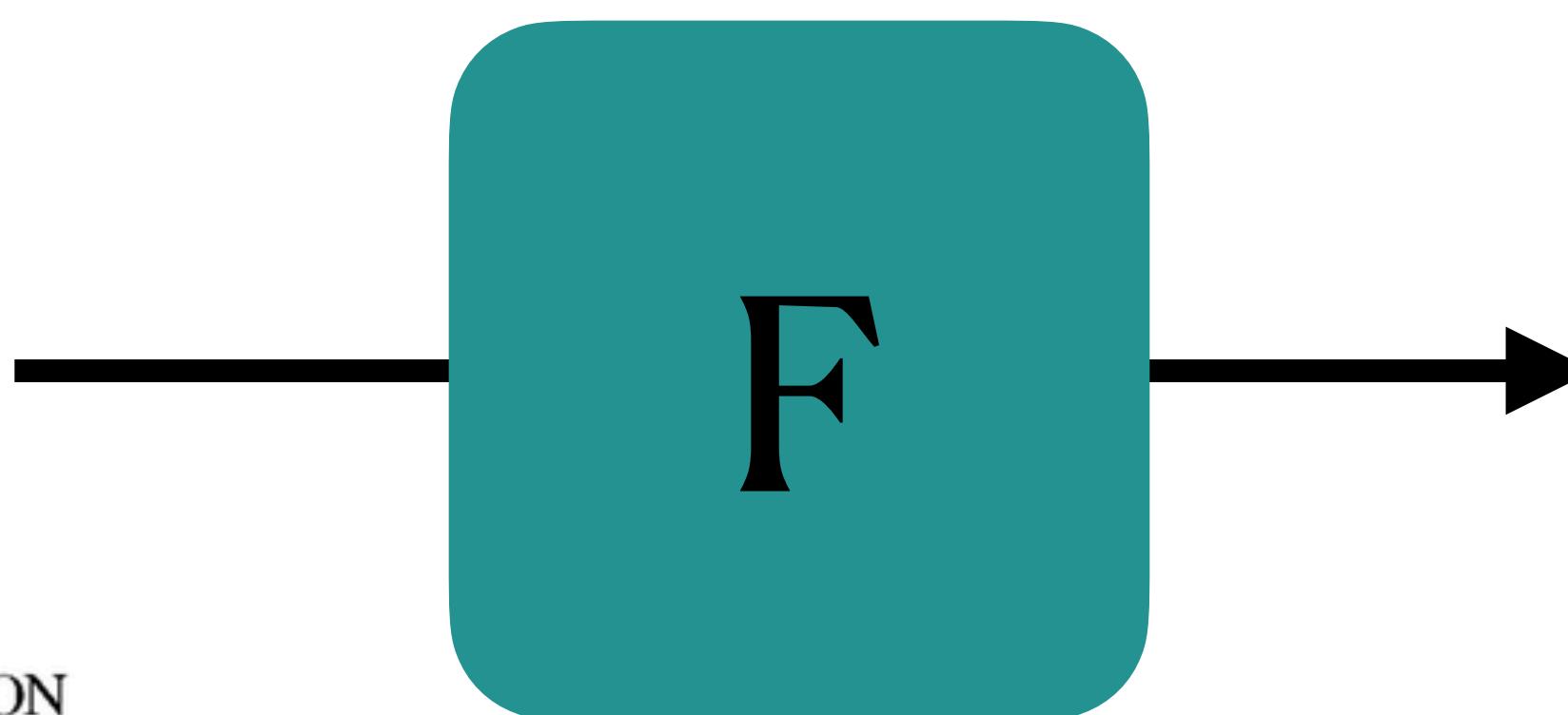
Healthcare



Neuroscience

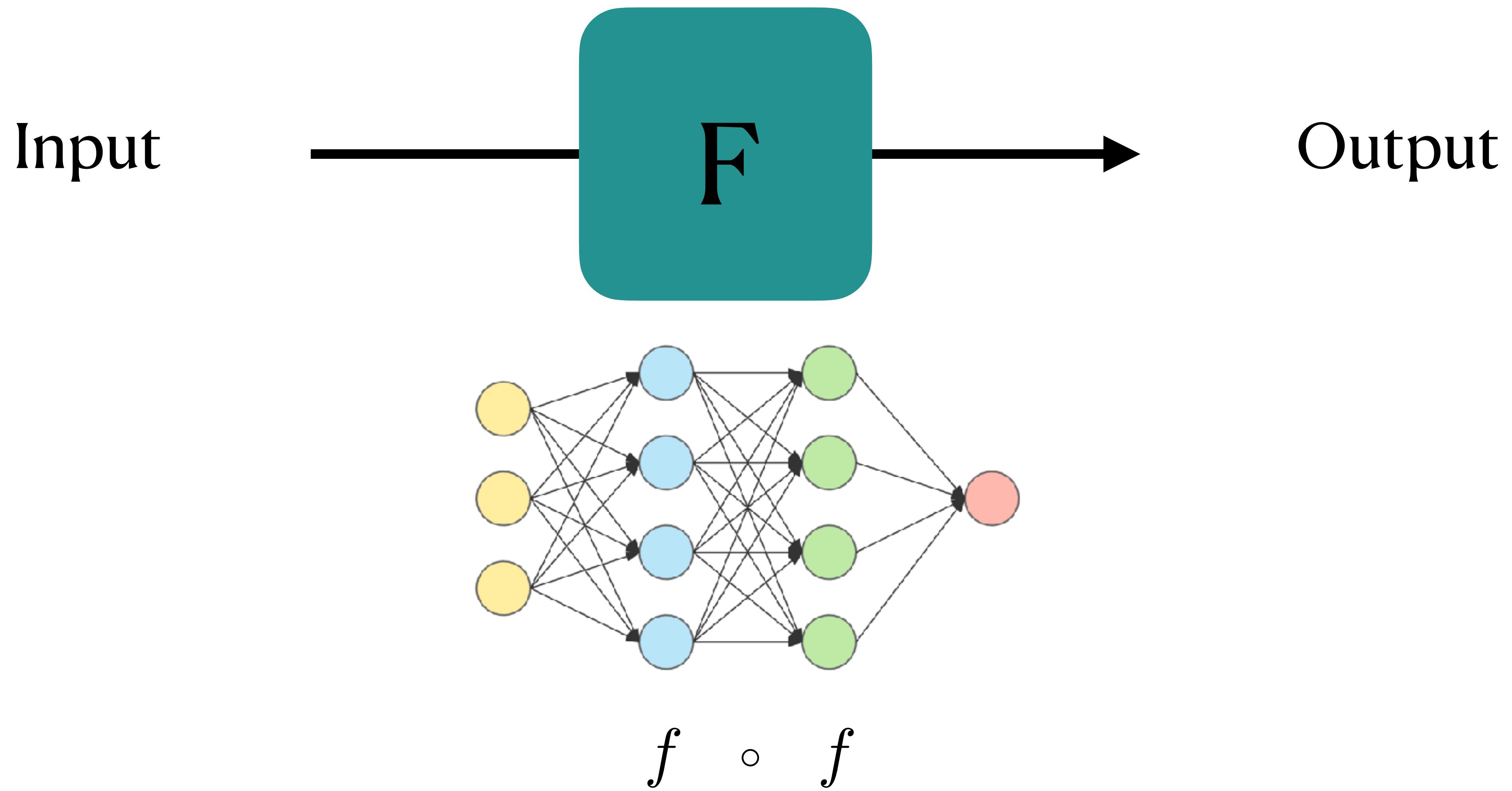


EEG



Lainscsek, Claudia, et al. "Non-linear dynamical analysis of EEG time series distinguishes patients with Parkinson's disease from healthy individuals." *Frontiers in neurology* 4 (2013): 200.

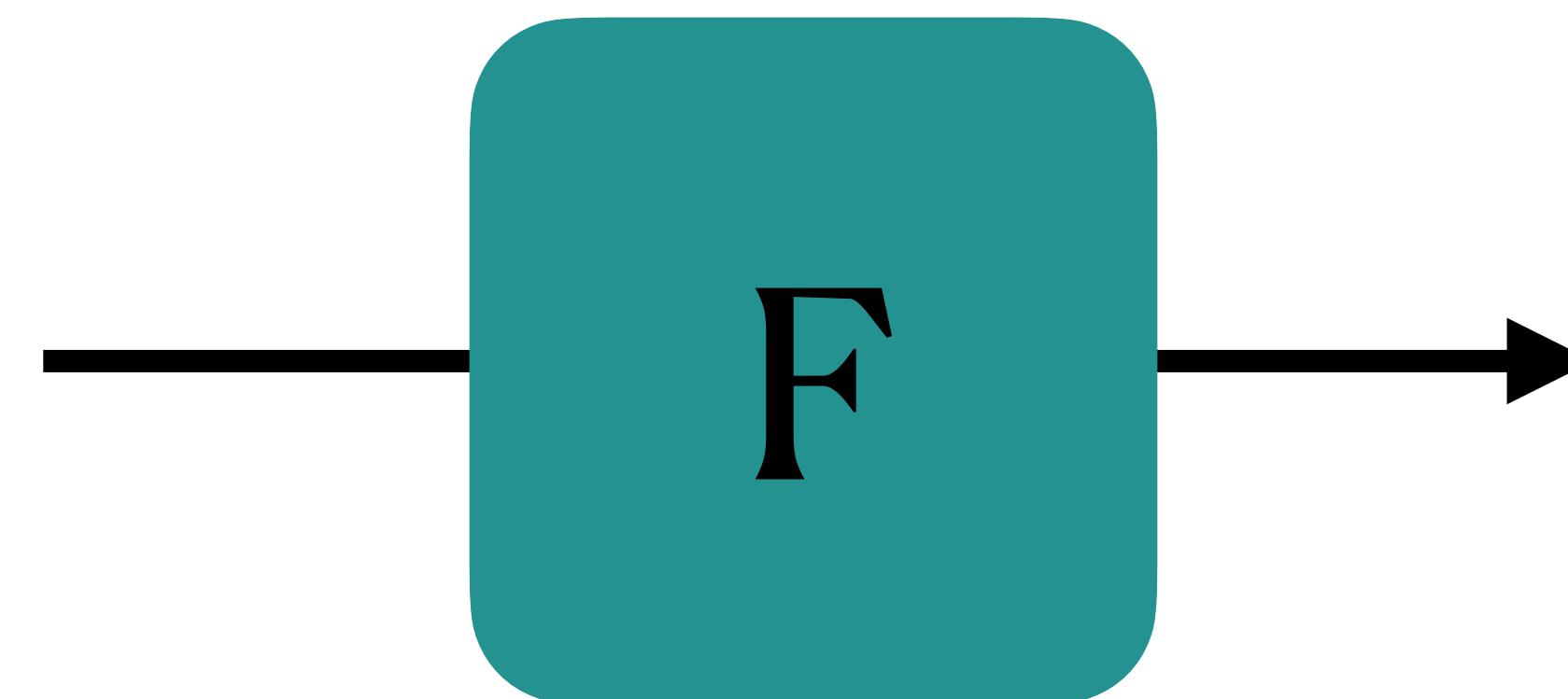
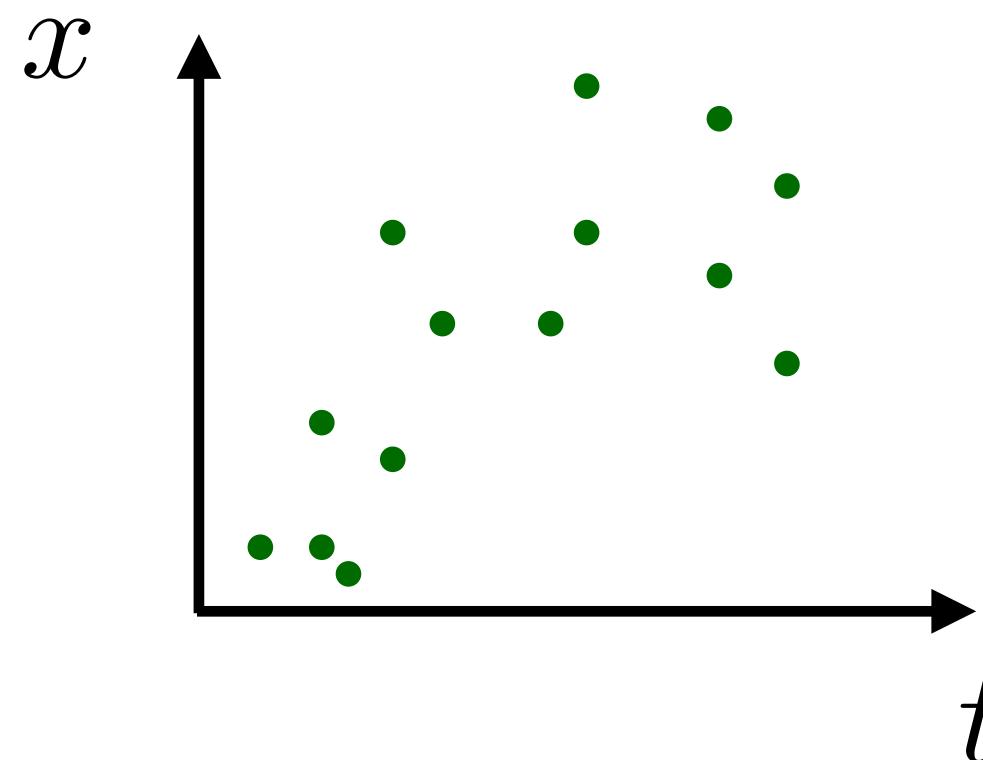
Neural Networks



Applications in Scientific Computing

Symbolic Regression

Time series



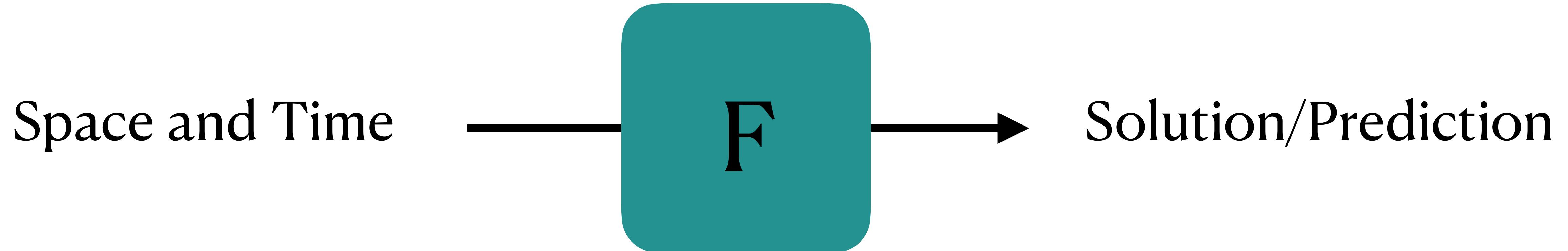
+ Physical Constraints
e.g. Conservation of Energy

???

$$\frac{dx}{dt} = f(x)$$

Applications in Scientific Computing

Physics Informed Neural Networks



+ Physical Constraints

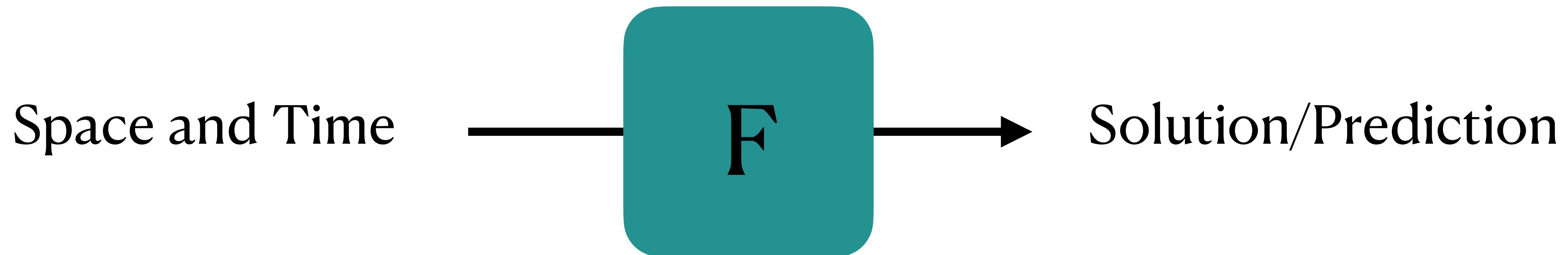
e.g. Conservation of Energy

???

A decoder-only foundation model for time-series forecasting

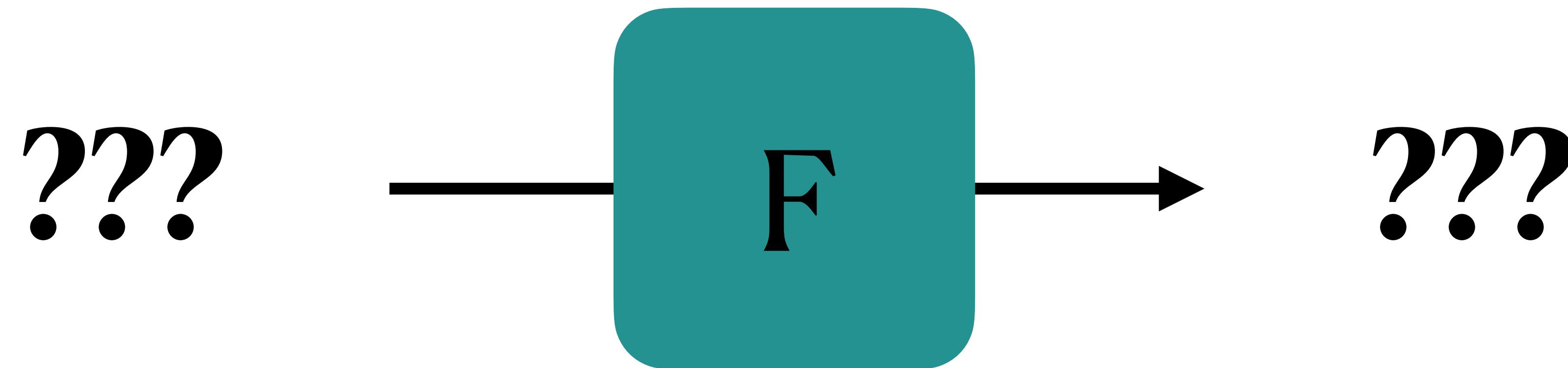
FRIDAY, FEBRUARY 02, 2024

Posted by Rajat Sen and Yichen Zhou, Google Research



Despite DL-based forecasters largely **outperforming** traditional methods and progress being made in **reducing training and inference costs**, they face challenges: most DL architectures require **long and involved training and validation cycles** before a customer can test the model on a new time-series. A foundation model for time-series forecasting, in contrast, can provide decent out-of-the-box forecasts on unseen time-series data with no additional training, enabling users to focus on refining forecasts for the actual downstream task like **retail demand planning**.

What is your input-output of interest?



+ Constraints

???