PARCOURS IA – 2019/2020

Fondements de l'Intelligence Artificielle et du Machine Learning

SESSION 1 – INTRODUCTION

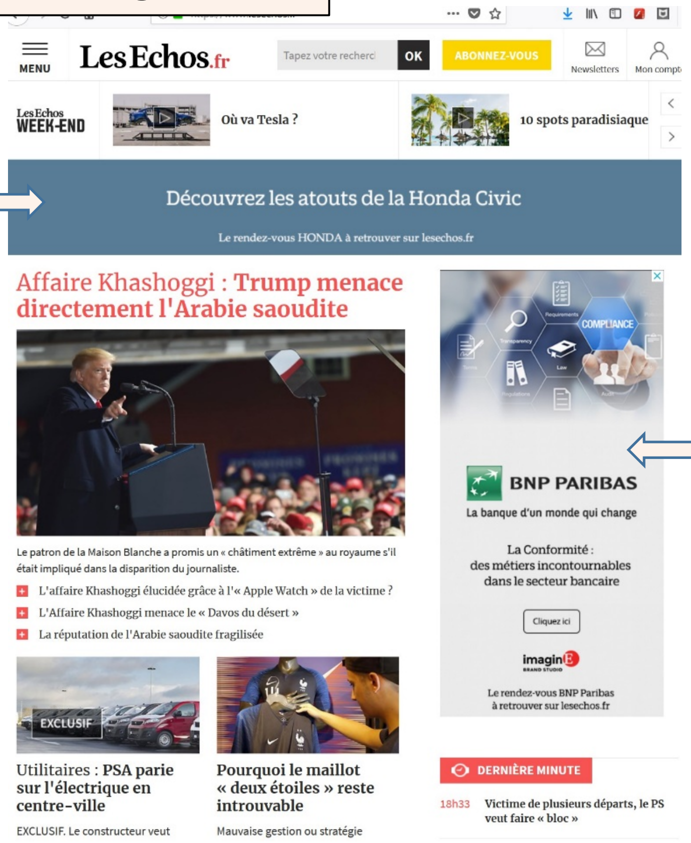
NICOLAS VAYATIS

école normale supérieure paris—saclay

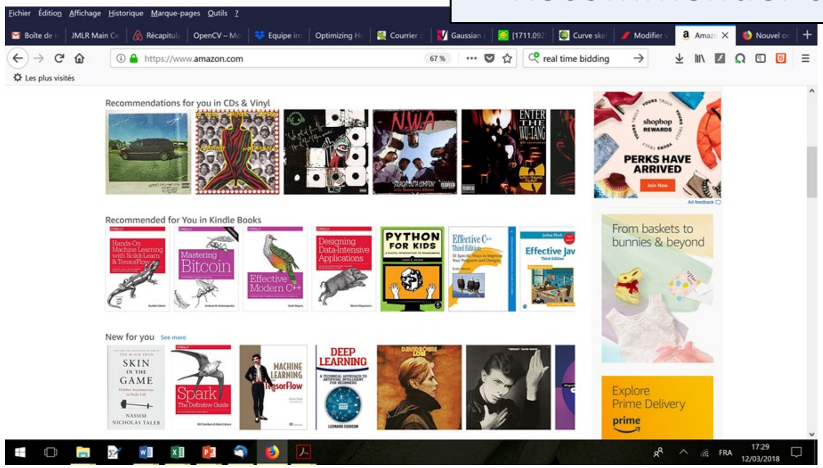
How to start a talk about AI?

Commercial success of AI

Advertising



Recommender systems



Facebook for the blind

A fact: AI may outperform Humans

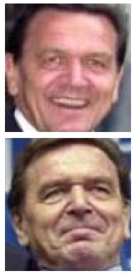
Board games

1996 : Gary Kasparov vs. Deep Blue (IBM)

2016 : Ke Jie vs. AlphaGo (DeepMind)

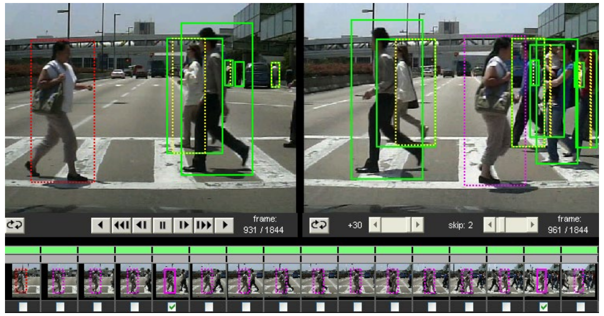
Supervised learning algorithms outperform human performance in many pattern recognition tasks

- LeCun et al. (1989): Handwritten zip code digit recognition
  - → USPS database; about 10,000 digits
  - → 10 categories; 7000 training data (16x16 gray level images)



- Lu and Tang (2015): Face recognition
  - → Life Faces in the Wild (LFW) data set
  - → 5749 public figures; 13,233 uncontrolled face images
  - → Training on 40,000 pairs of images (matched/mismatched)

- Zhang et al. (2017): Pedestrian recognition
  - →Caltech pedestrian data set
  - → 10hours video at 30Hz;10^6 frames
  - → 10% contain pedestrians; 2300 unique pedestrians
  - → Some trouble with partial occlusions…



Questions raised

- What are the drivers of success for AI components and their current limitations?
- Why successful AI applications are typically related to images and text data ?
- Is cross-validated prediction performance the only criterion to adopt AI-driven technologies?
- How does AI rely on heavy (and energy-consuming) computational resources?
- Will there still be a Human in the loop in ten years?

How far can AI go?

(Super)creativity

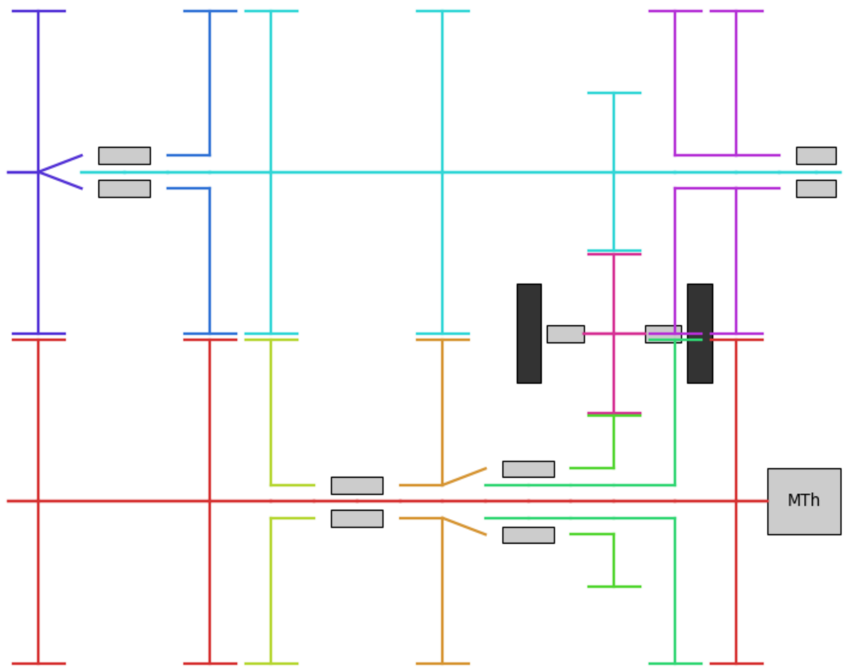
(Super)creativity

The next Rembrandt

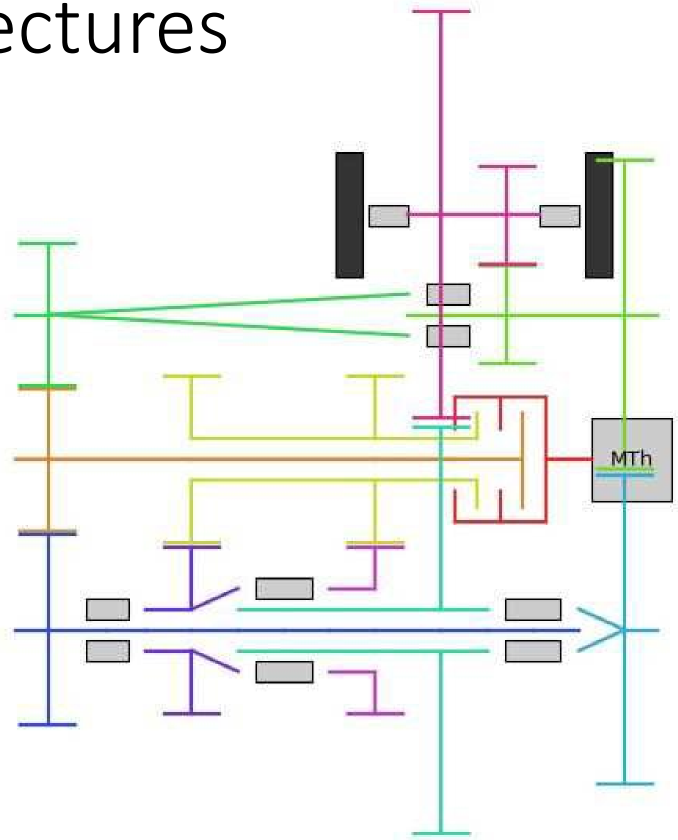
AICAN exhibition, 2018

An example of system design (with HPC)

Exploring the space of gearbox architectures



2D model of a six gearbox ratios of a manual transmission



2D model of a five gearbox ratios of a dual-clutch transmission

From S. Masfaraud (2016) - Méthodes numériques pour la recherche et la conception d’architectures optimales de boîtes de vitesses, Thèse ENS Paris-Saclay.

What does it take to sample and screen the space of gearbox architectures

- 992 architectural schemes scanned
- 1.5×10^9 architectures generated
- 1.5×10^8 architectures tested
- 1,390 viable architectures extracted
- 13,600 CPU-hours on Intel Xeon E5-1620v2
- Further screening based on price and mass constraints
- Expert assessment to evaluate plausibility regarding to volume optimization

2D model of a six gearbox ratios of a manual transmission

- 142 architectural schemes scanned
- 2.5×10^8 architectures generated
- 2.5×10^7 architectures tested
- 320 viable architectures extracted
- 13,600 CPU-hours on Intel Xeon E5-1620v2
- Further screening based on price and mass constraints
- Expert assessment to evaluate plausibility regarding to volume optimization

2D model of a five gearbox ratios of a dual-clutch transmission

From S. Masfaraud (2016) - Méthodes numériques pour la recherche et la conception d’architectures optimales de boîtes de vitesses, Thèse ENS Paris-Saclay.

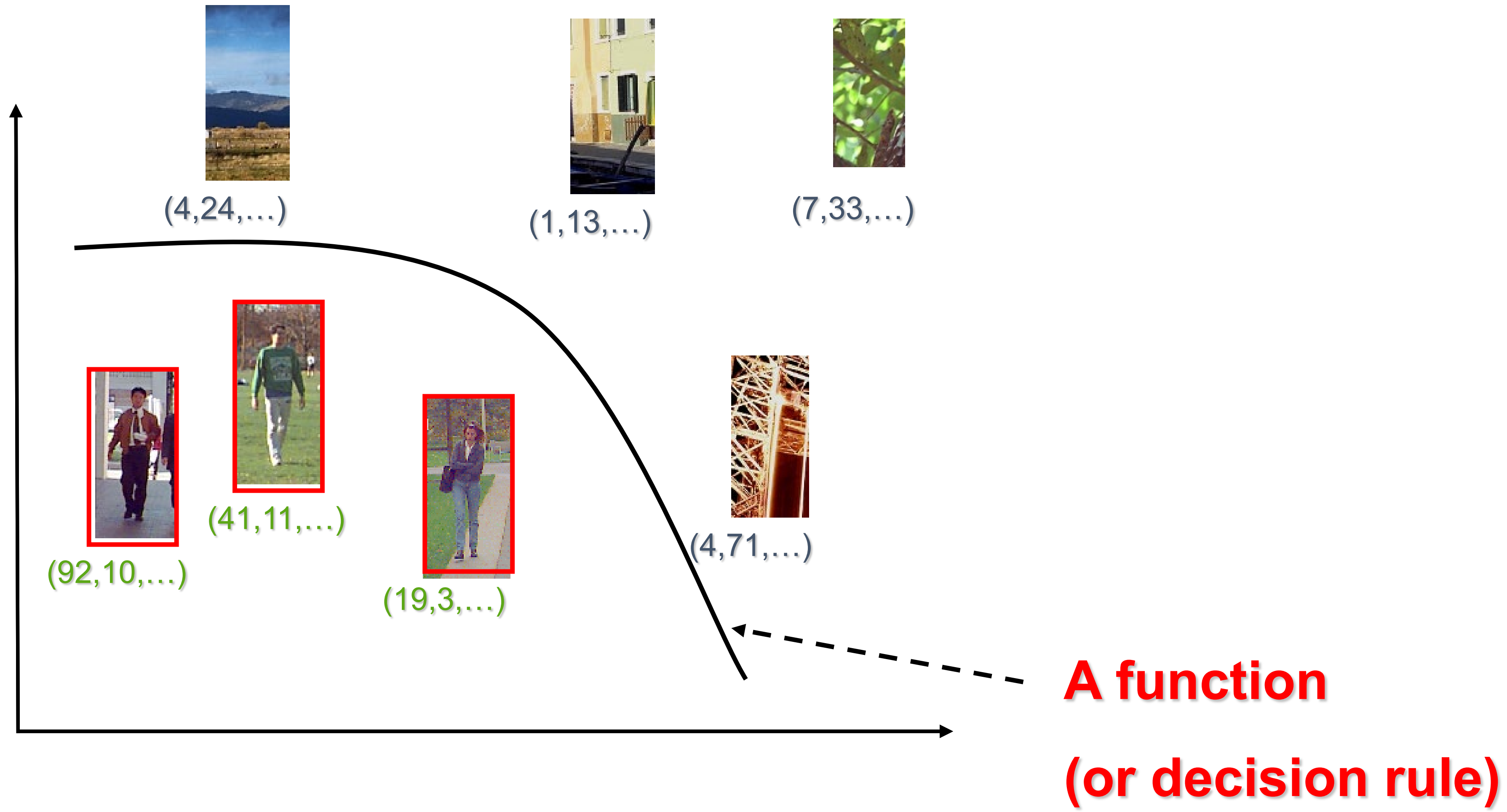
What we learned from innovative gearbox design

- There might be a DeepBlue for gearbox design
  - Not clear what is is the complexity ceiling to extend it to engine design for instance
  - Requires the potential of HPC to sample and screen architectures in order to scale up
- Need to embark field expertise together with modeling ability:
  - Gearbox engineering, mechanical systems, optimization, graph sampling…
- Contribution of machine learning?
  - Not obvious at this stage, but…
  - … it may help to better select high level design parameters and save brute force exploration time
- How to embrace such a design process disruption?
  - Mindset of the organization
  - Mindset of field experts

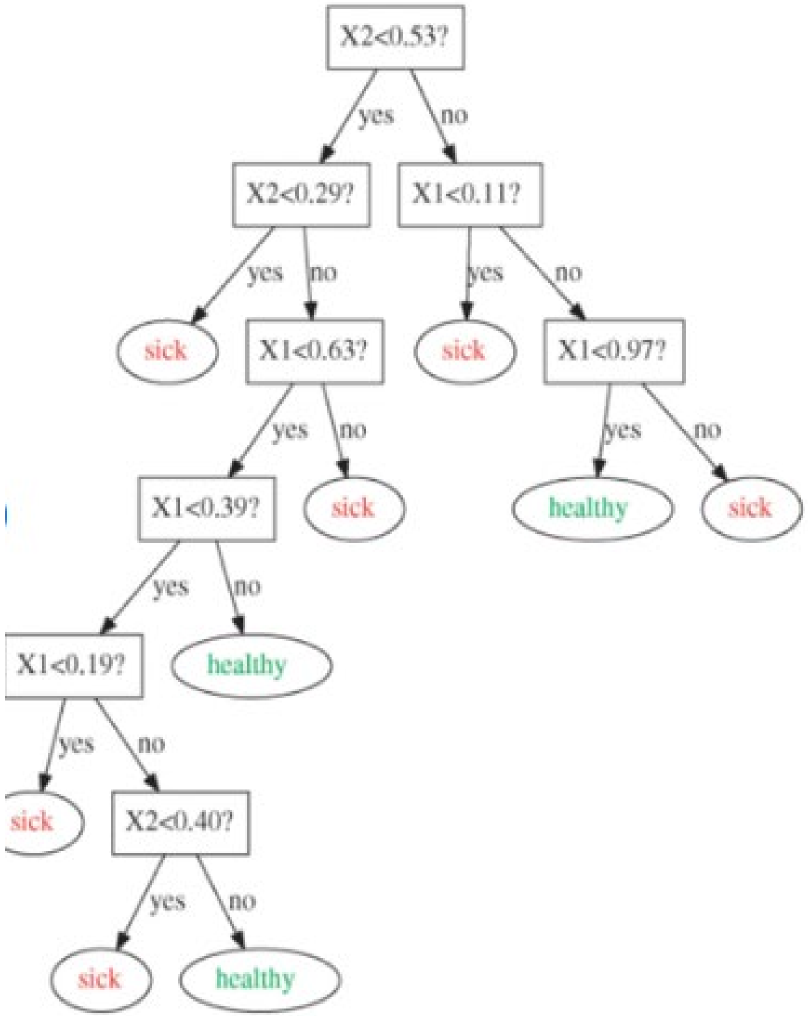
Demystification
of AI/machine learning

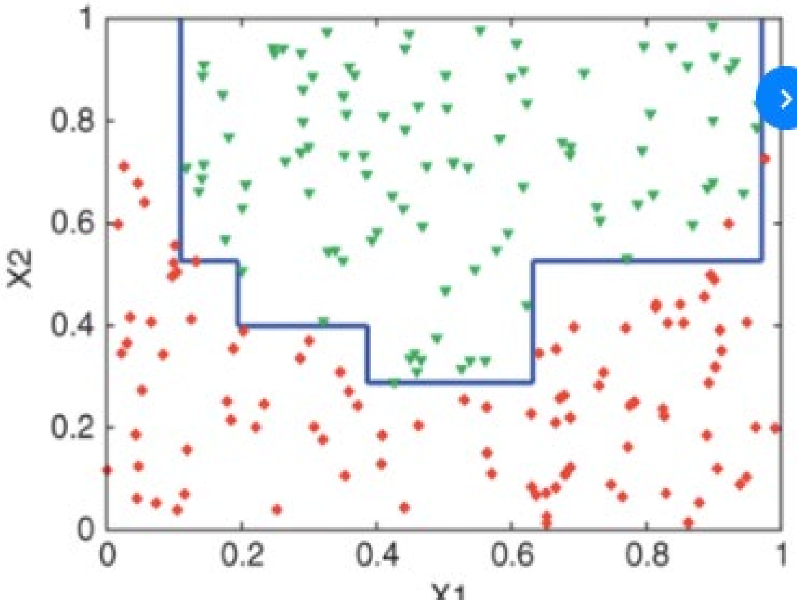
Supervised training data for pedestrian recognition

How does the machine “represent what it learned”?



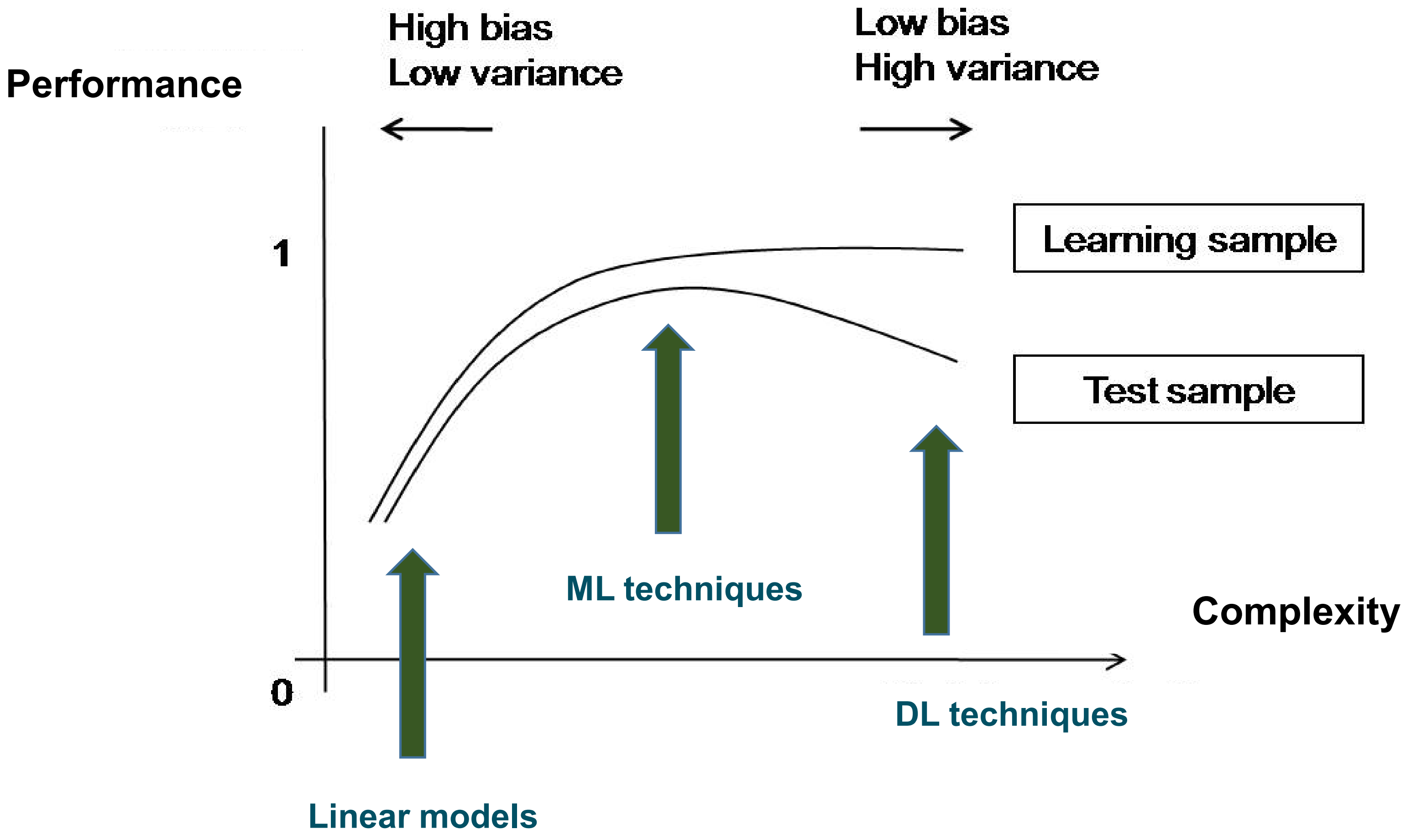
Machine learning is mainly about function estimation /approximation





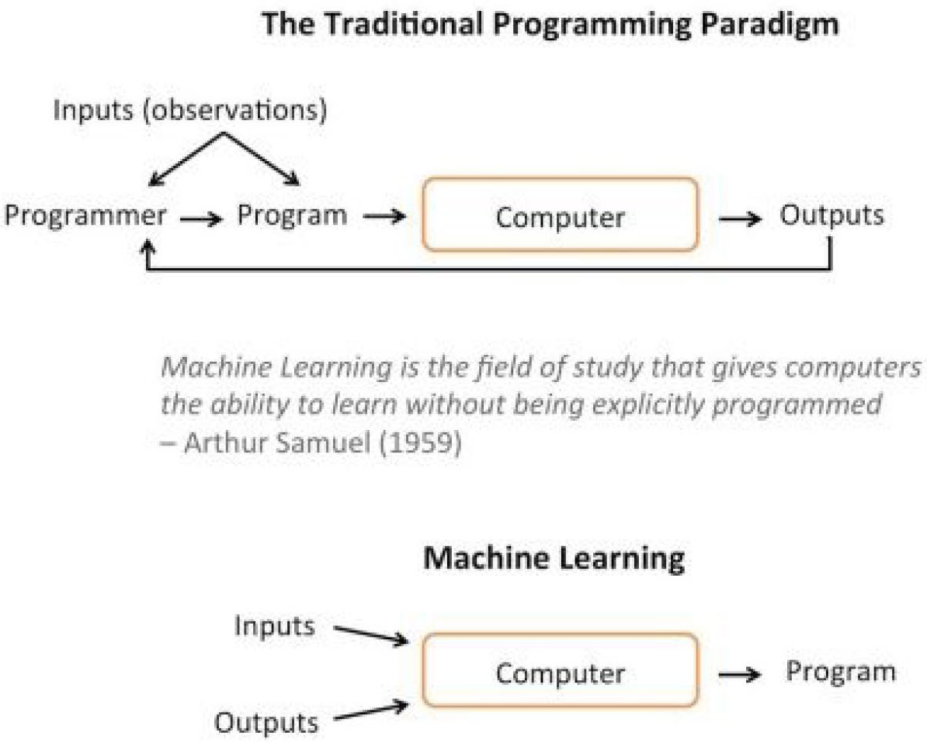
ALL OF MACHINE LEARNING IN ONE FIGURE

UNDERFITTING AND OVERFITTING, IT'S ALL ABOUT TRADE-OFF

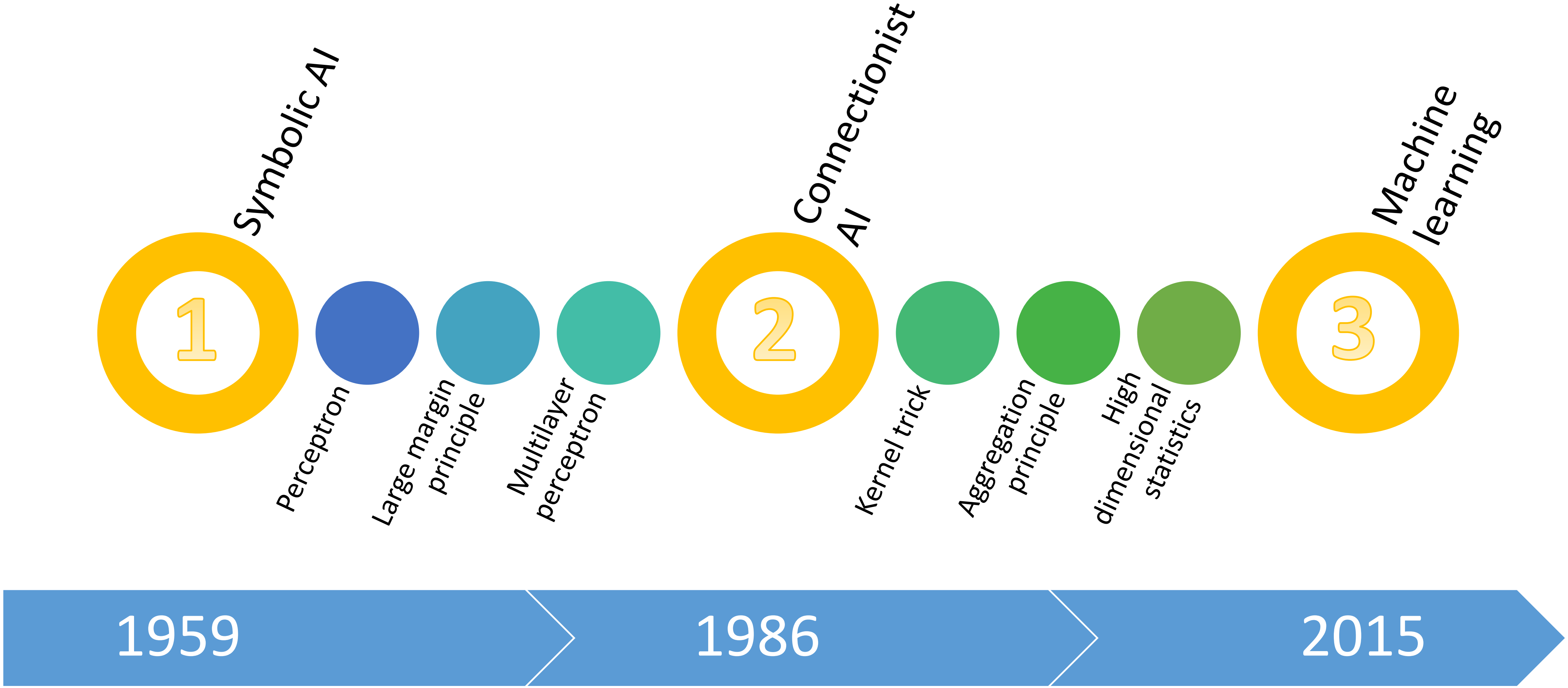


Historical perspective on AI

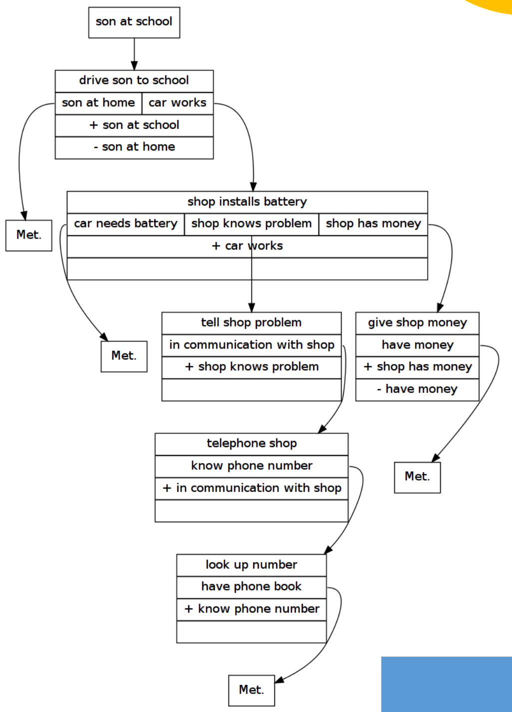
From Symbolic AI to Machine Learning



Three AI waves… and two AI winters



Winters explained



1

Symbolic AI

Only able to capture explicit knowledge

Lack of robustness

Computational burden

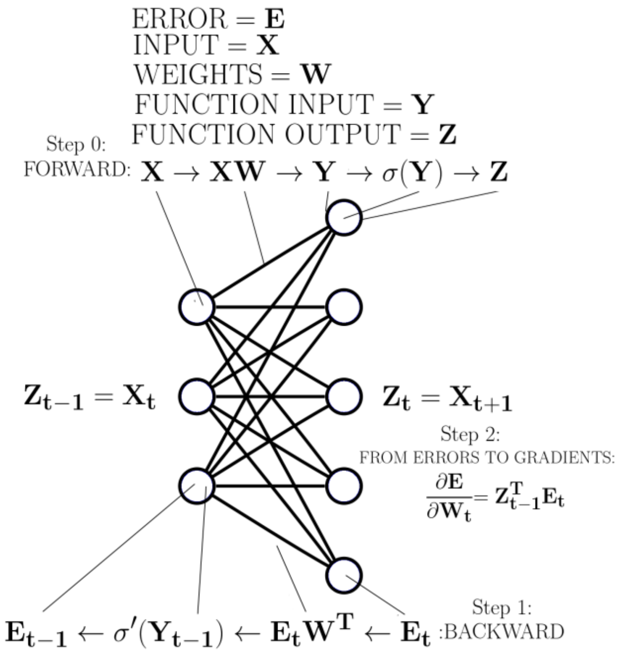
2

Connectionist AI

Lack of data

Lack of interpretability

Hard to maintain

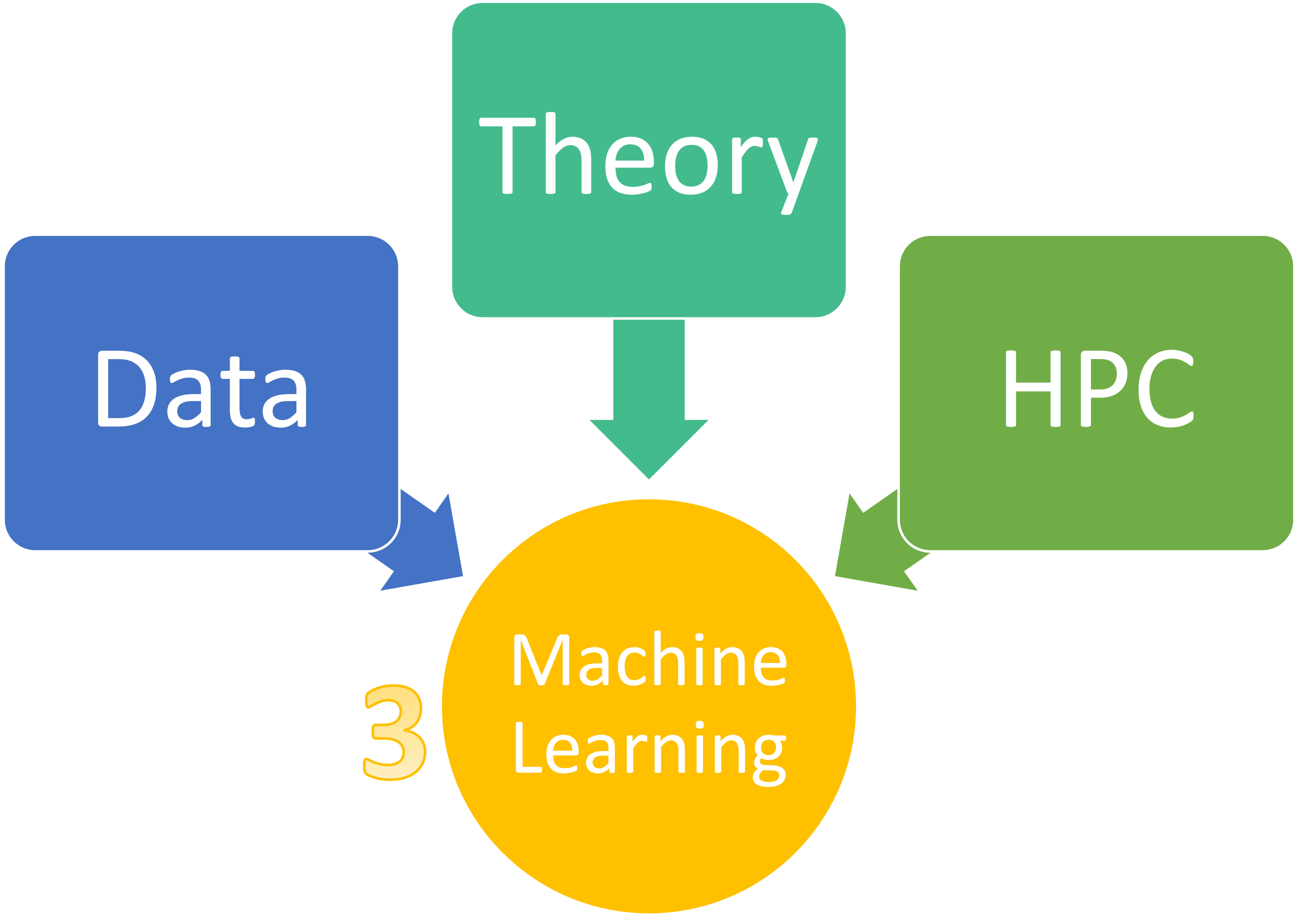


1959

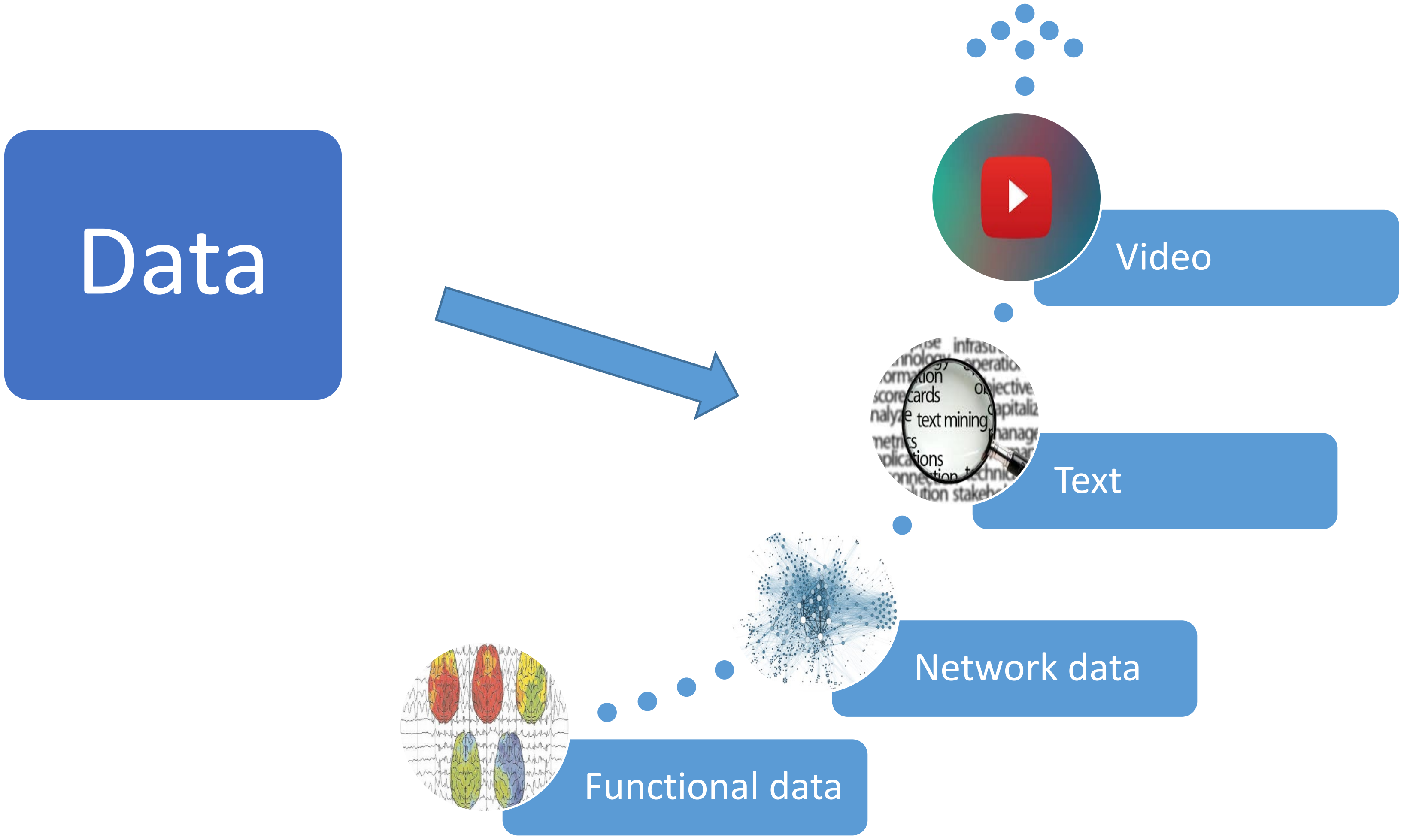
1986

The three drivers of the third AI wave

What is different now



All is data



Statistical Learning Theory: setup

Data

Sampled from the unknown distribution of (X,Y)

where

X: measurement

Y: label

→ **supervised learning**

Decision rules

A functional class G with elements g

Characterized by its **complexity**

R(G)

Prediction on X is given by g(X)

Risk (performance)

Loss function assigns a cost $l(g(X), Y)$

Risk assessed by the average of the loss over the population (in expectation)

Inference principle

Empirical risk minimization

Consists in minimization over G of the empirical loss (computed over the training data)

(Machine) learning amounts to functional optimization

Statistical Learning Theory: main ingredients

$$\widehat{R}_n(\mathcal{F}) = \mathbb{E}\left(\sup_{f\in\mathcal{F}} \frac{1}{n}\sum_{i=1}^{n} \varepsilon_i f(Z_i) \,\middle|\, D_n \right)$$

Assume h is a function with bounded differences and denote by $c_1, \ldots, c_n > 0$ the upper bounds on its componentwise variations

We have, for any $t > 0$

$$\mathbb{P}\left(h(Z_1, \ldots, Z_n) - \mathbb{E}(h(Z_1, \ldots, Z_n)) > t \right) \leq \exp\left(-\frac{2t^2}{\sum_{i=1}^{n} c_i^2} \right)$$

Under (A1-A2), we have, for some $s > 1$, that any real-valued measurable f satisfies :

$$L(g_f) - L^* \leq 2c(A(f) - A^*)^{1/s}$$

Rademacher complexity

Concentration inequality

Risk communication

Statistical Learning Theory: typical garantee

$$\boxed{L(\widehat{g}_n) \leq \inf_{g \in \mathcal{G}} L(g) + \widehat{R}_n(\mathcal{G}) + 3\sqrt{\frac{\log(2/\delta)}{2n}}}$$

with probability at least $1 - \delta$

True loss of the ERM

≤ Minimal loss in the class

+ Complexity

+ Precision

Is HPC a necessary tool for Machine Learning?

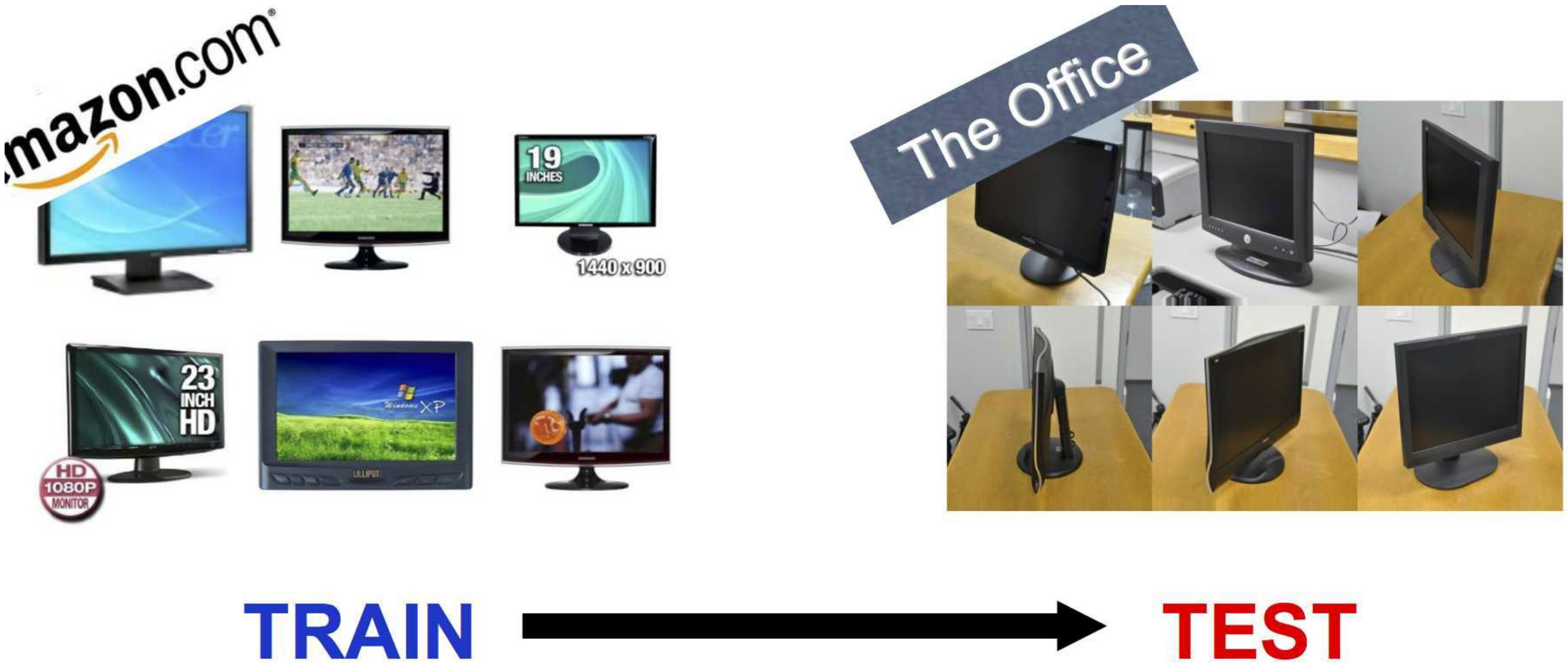
- Big data + Deep Learning + Real-time training → definitely needs HPC

- What if:
  - Real-time decisions but not necessarily real-time training
  - Not so big data
  - Satisfied with other (shallow) Machine Learning algorithms
    (e.g. Random Forests, Boosting, SVM)

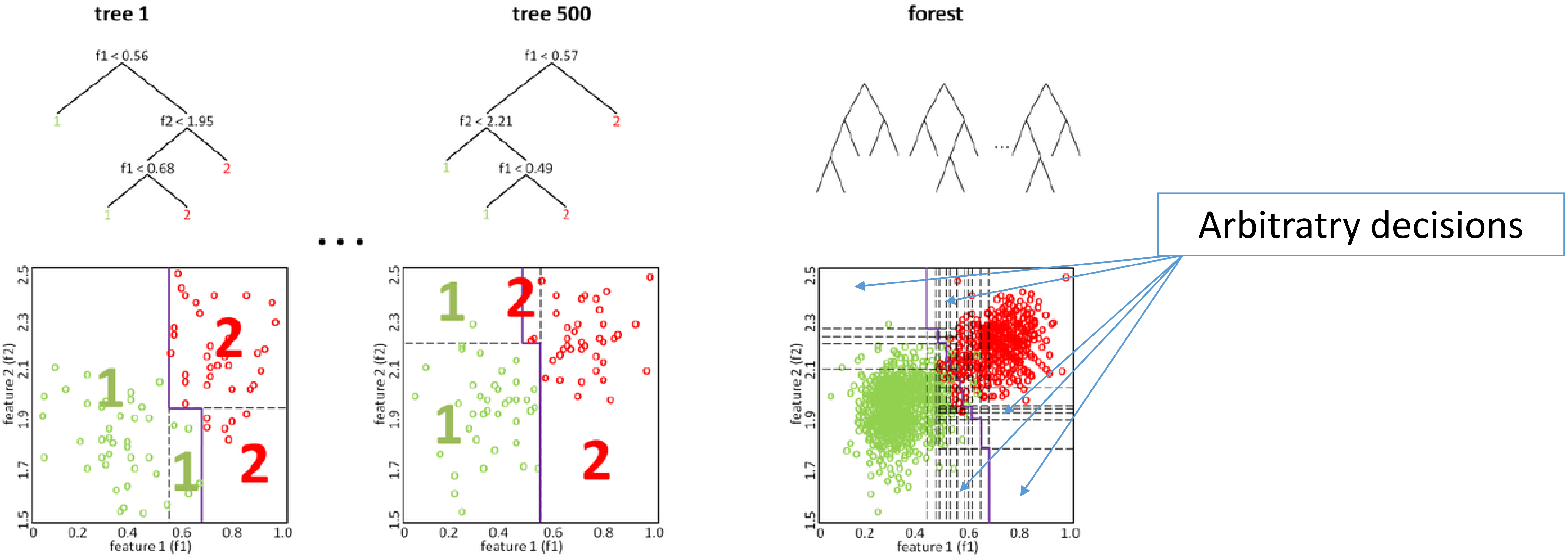
?

Demystification of big data

First issue of big data: Sampling bias

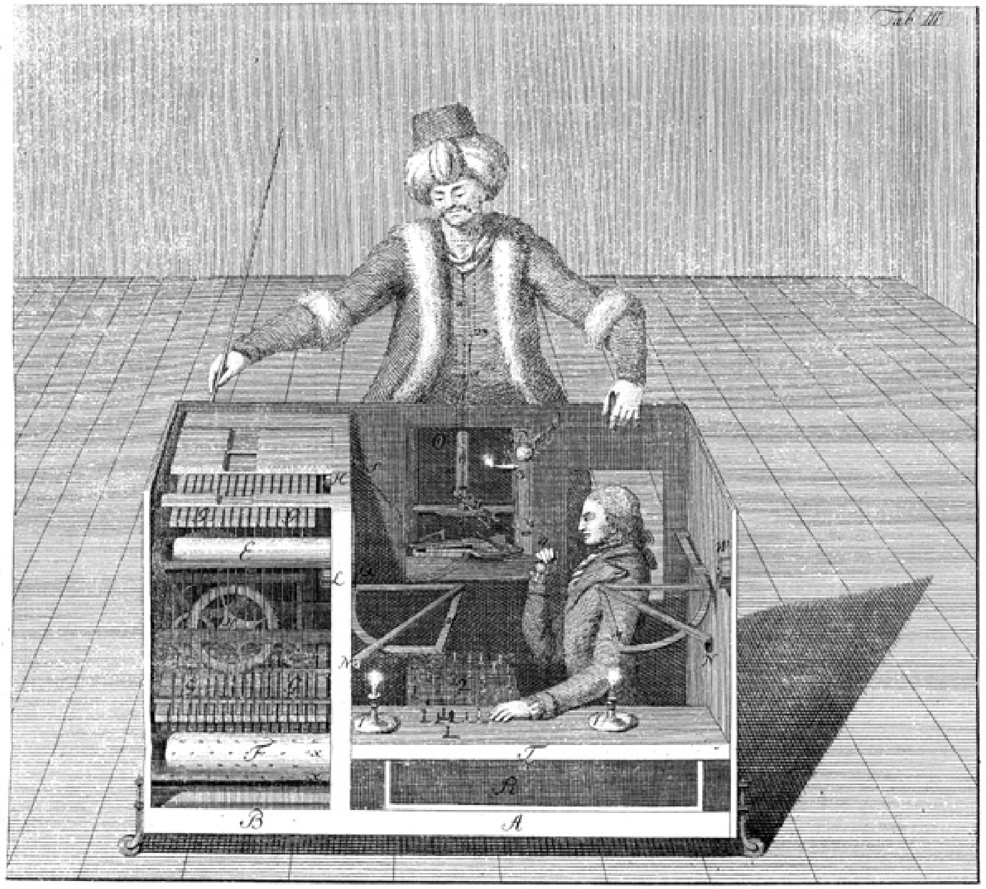


Biases and artefacts in data-driven partitions

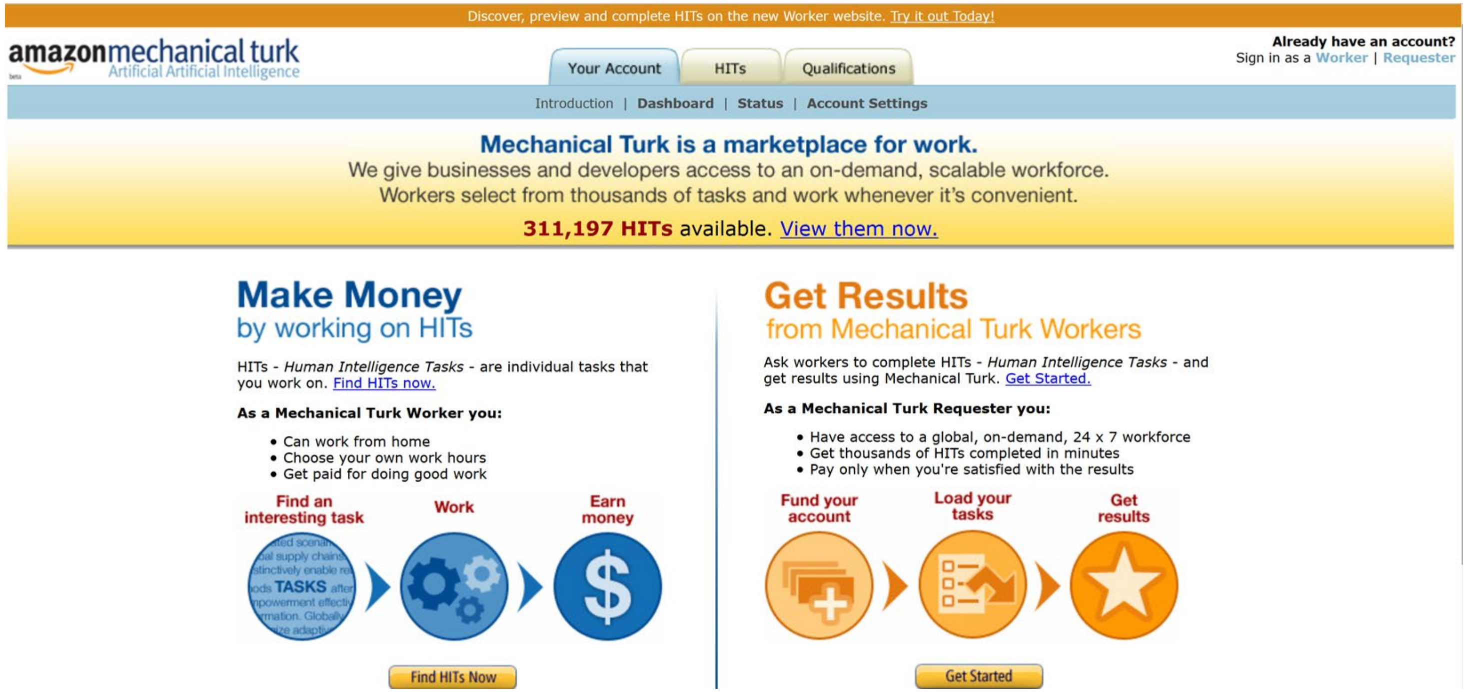


→ Need to learn with a reject option (see work by Marten Wegkamp, 2005-...)

Second issue with big data: Need a Turk!

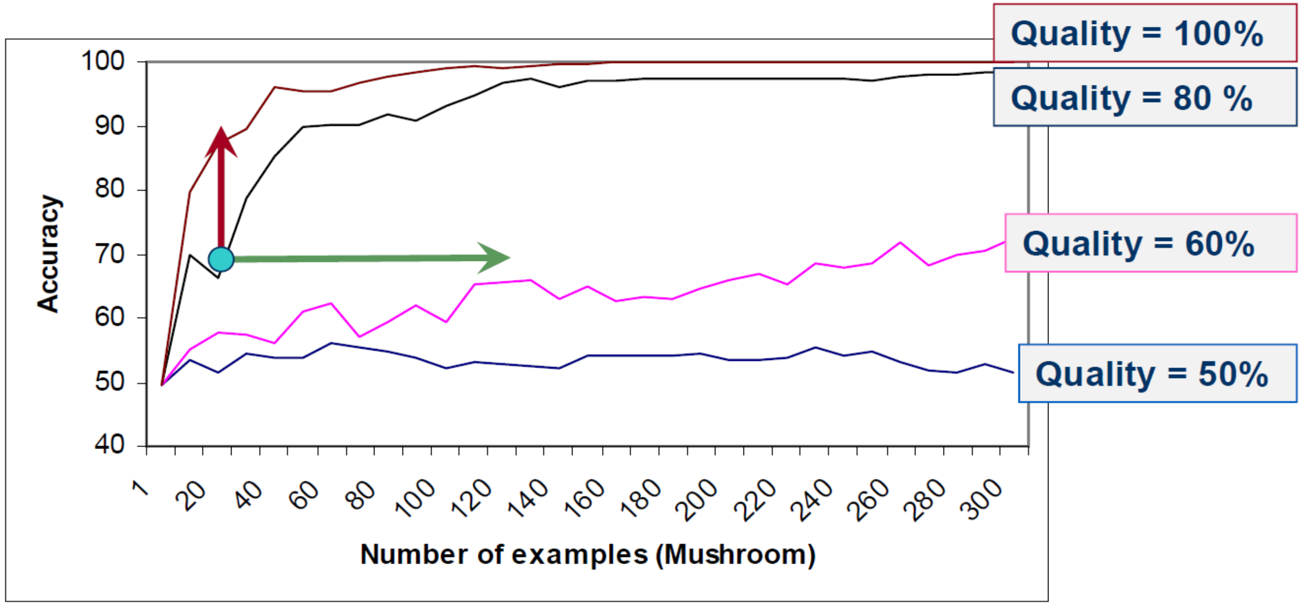


The cost of data labeling



Quality labels more powerful than big data

- Get more examples → Improve classification
- Get more labels → Improve label quality → Improve classification



Source: Get Another Label? Improving Data Quality and Data Mining Using Multiple, Noisy Labelers.
Proceedings of KDD-2008
by V. S. Sheng, F. Provost, P. G. Ipeirotis.

Machine Learning for Industry or Science: a different story…

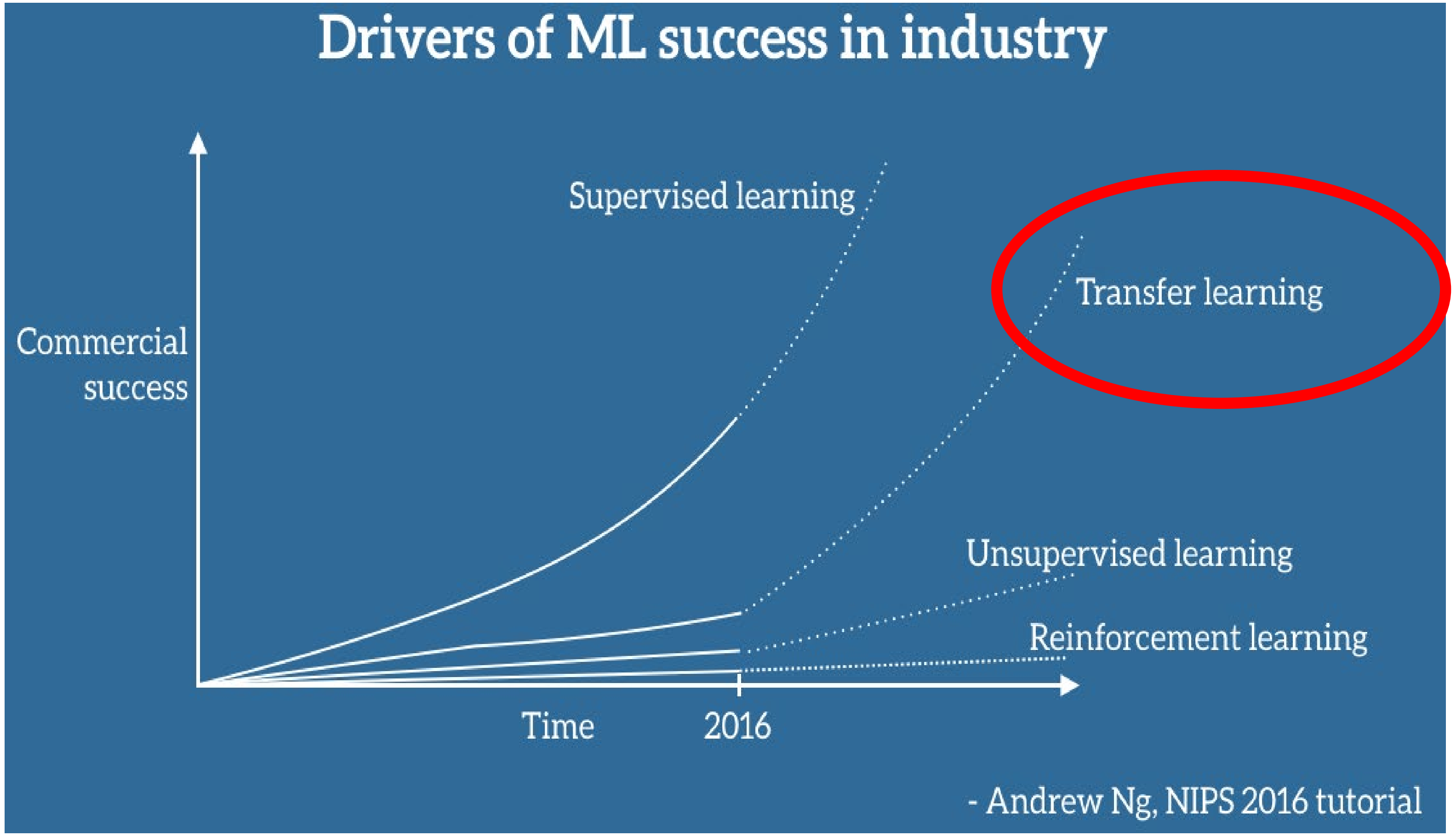
Industrial requirements heavier than for marketing

- Industrial processes go under continuous improvement → **Sampling bias is the rule!**

- Labeling training data relies on **field expertise**→ Turks are expensive and unwilling!

- Expectations for **performance** are at a different scale when comparing decisions for critical systems or clinical applications to advertising or book recommendation

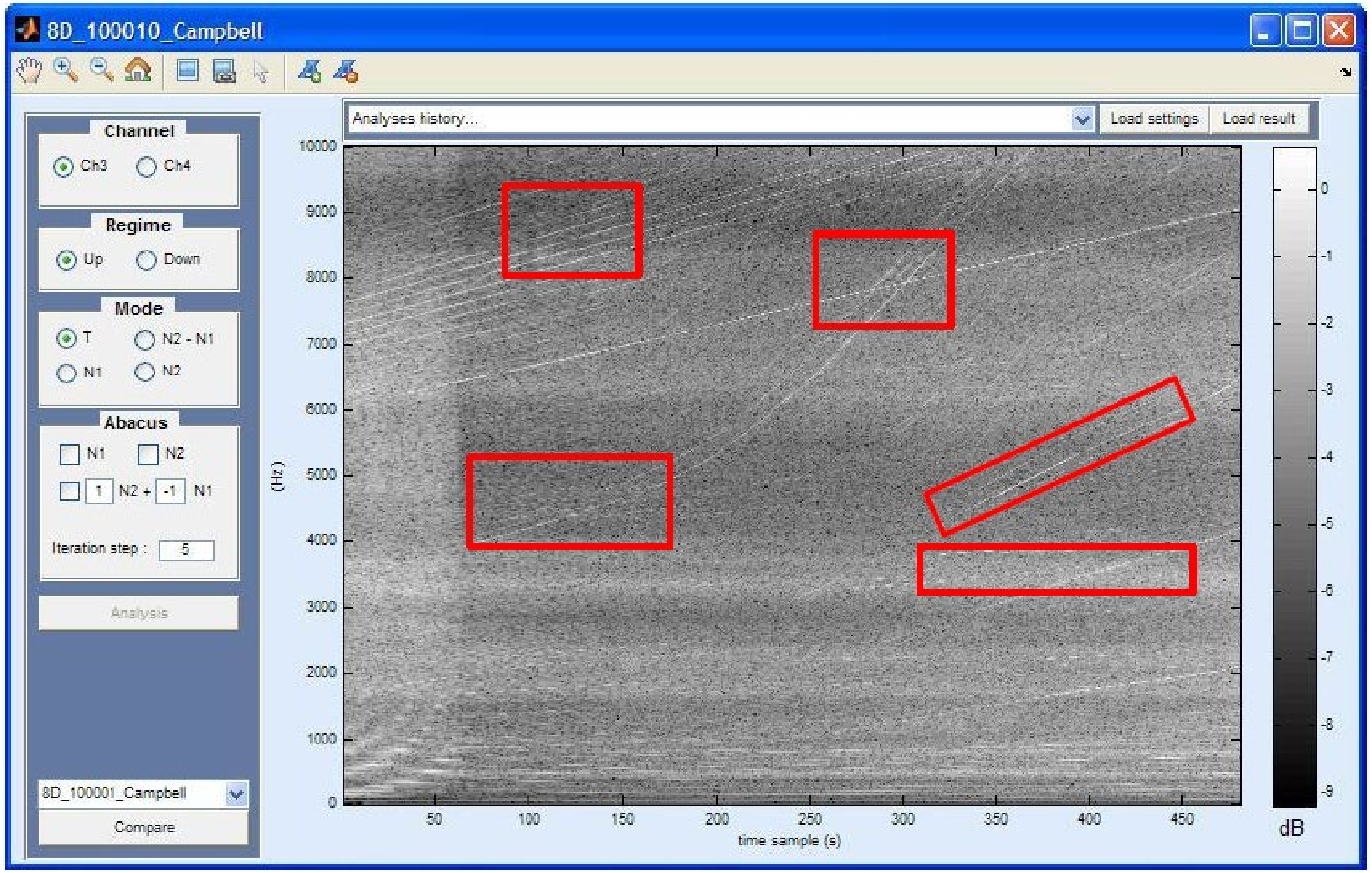
Expected impact of machine learning in the industry



NB: Andrew Ng is VP & Chief Scientist of Baidu, Co-Chairman and Co-Founder of Coursera, and an Adjunct Professor at Stanford University.

Machine learning to support scientific computing and simulation projects

An example of anomaly detection objective
Benchmark assessment of aircraft engine



Source: Confidential report (2012) - Mathilde Mougeot, NV

- What we see?

Time-frequency representation of vibration signals (Campbell diagram) wrt to speed during acceleration and decelaration regimes.

- Nature of anomalies

Tiny details in those images. Require a lot of expertise to tag.

- Databases are small

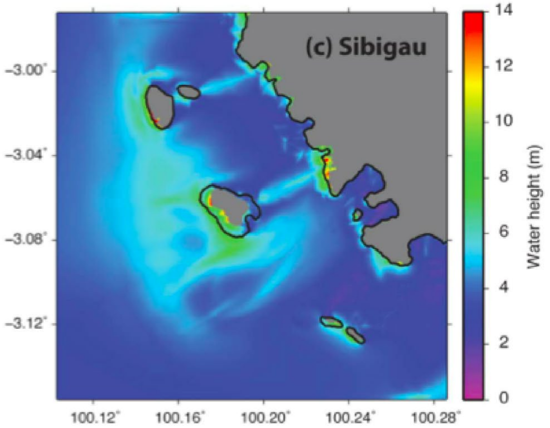
Only a few hundreds engines have been recorded with a very limited number of anomalies reported.

- But image structure helps!

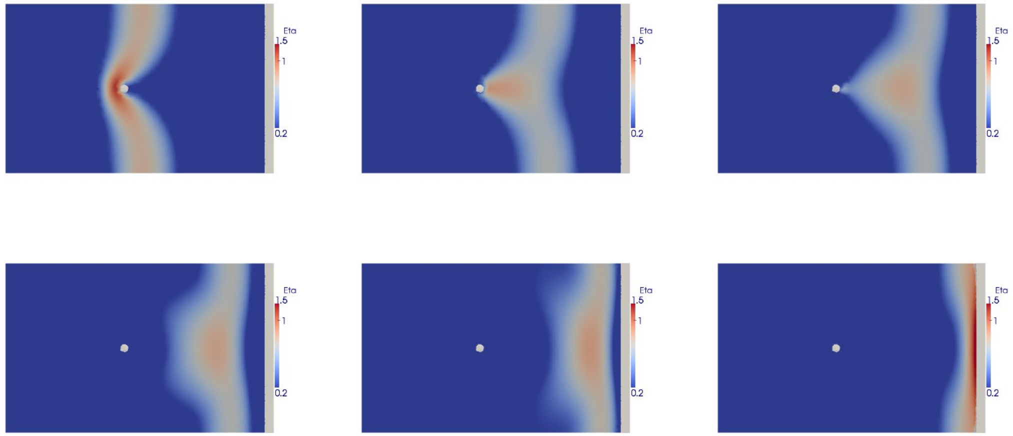
Anomaly detectors can be built using adapted representations of such signals and basic nearest neighbors in feature space.

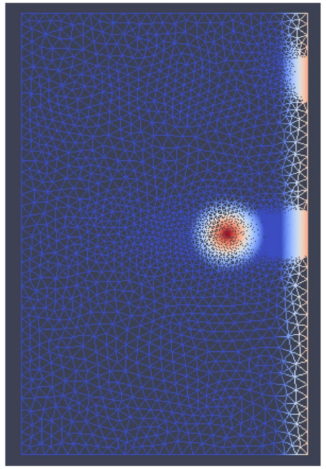
Example of project (1/2)

Tsunami run-up amplification



2010 Sumatra tsunami and the Mentawai Islands (Hill et al., 2012)



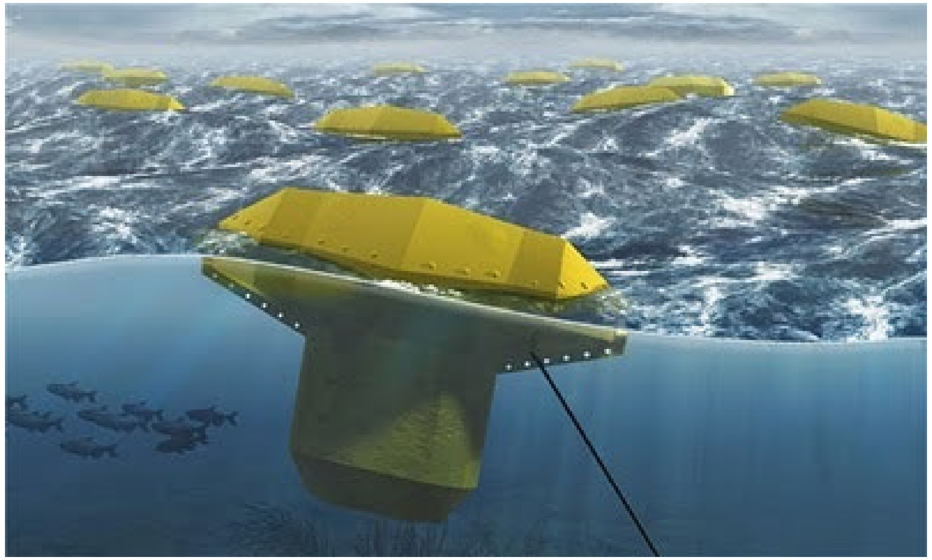


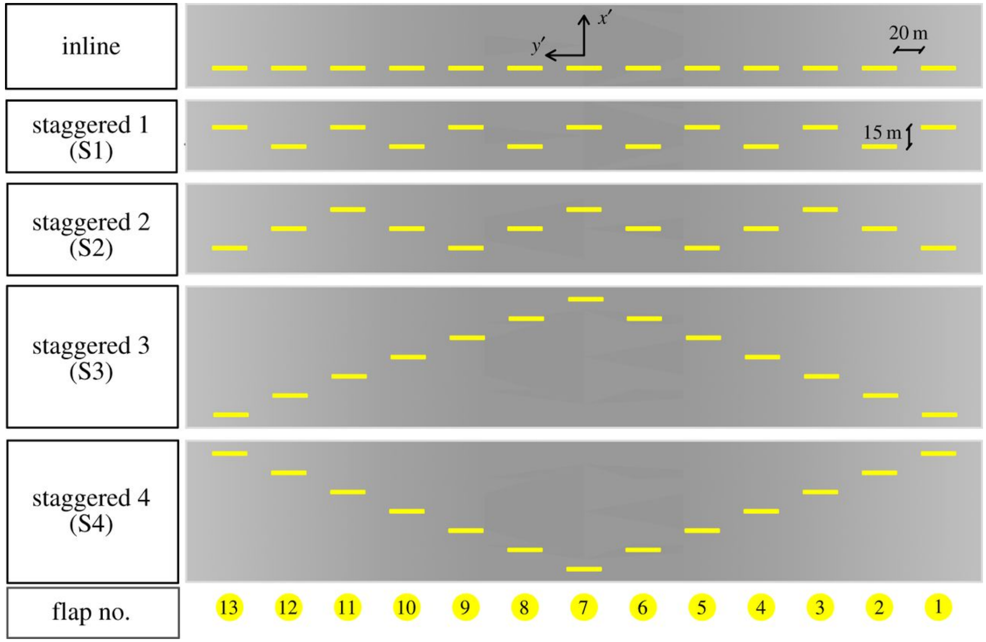
Adaptive mesh grid of the VOLNA solver

From:

Themistoklis S. Stefanakis, Emile Contal, Nicolas Vayatis, Frédéric Dias, and Costas E. Synolakis (2014). **Can Small Islands Protect Nearby Coasts From Tsunamis? An Active Experimental Design Approach .** Proceedings of the Royal Society-A, 470: 20140575.

Example of project (2/2)
System design for WEC farms





From:
Dripta Sarkar, Emile Contal, Nicolas Vayatis, Frederic Dias (2015).
A Machine Learning Approach to the Analysis of Wave Energy Converters.
Proceedings of OMAE 2015.

Take-home messages

Machine learning achieves some kind of regression in high dimensional (or structured) spaces

- Heavily relies on mathematics to model complex data and formulate the task-related optimization problem
- AI-based technologies may outperform humans in certain *well-defined* prediction tasks: detection, recognition, planning, etc.
- Missing piece: few studies on control actions (after prediction)
- Strong AI not for tomorrow… still need to define the search space and the objective…

Scaling up and industrialization of AI modules in science and industry raises scientific challenges

- Sciences (life sciences, engineering sciences, social sciences, physics…) and Industry (energy, healthcare, banking, defense…) will not benefit of supervised learning 'as-is'
- Naive implementation of AI has/may/will lead to industrial disasters
- The main risk: to be driven by a method and not by the problem to be solved in *its* context
- Secondary risk: believe too much in training data and Proof-of-Concept
- Eventually: for the previous reasons, the risk to be out of the game, and there will be a game!