

Machine Learning

CS-433

Martin Jaggi & Nicolas Flammarion
EPFL

Alternatives

Master

- CIVIL-459 – Data and AI for transportation
- MGT-418 – Convex optimization
- MATH-403 – Low-rank approximation techniques
- MATH-412 – Statistical ML
- MICRO-455 – Applied ML
- MICRO-401 – ML Programming
- DH-406 – ML for the Digital Humanities
- CS-430 – Intelligent Agents
- CS-439 – Optimization for ML
- CS-401 – Applied Data Analysis
- EE-490(h) – Lab in data science
- EE-556 – Mathematics of data
- CS-526 – Learning theory
- EE-559 – Deep Learning
- EE-558 – A Network Tour of Data Science
- MICRO-570 – Advanced ML
- CS-433 – Machine Learning

Bachelor

- CS-233 – Intro to ML
- CS-330 – Artificial Intelligence
- BIO-322 – Intro to ML for bioengineers

Seminars, Doctoral Courses and continued education

- ENG-704 – EECS Seminar: Advanced Topics in ML
- CS-723 – Topics in ML Systems
- EE-608 – Deep Learn. for Natural Language Proc.
- EE-618 – Theory and M. for Reinforcement Learning
- EE-613 – ML for engineers
- EE-621 – Adaptation and Learning
- EPFL Extension School – Applied Data Science: ML

Course Logistics

Assessment

- ✿ **Project 1 (10%), due Oct 26th**
- ✿ **Project 2 (30%), due Dec 17th**
- ✿ **Final exam (60%)**

Course Logistics

Lectures

tentative
schedule

	Date	Topics Covered	Lectures	Exercises	Projects
MJ	15/9	Introduction, Linear Regression	01a		
	17/9	Cost functions		Lab 1	
	22/9	Optimization			
	24/9	Optimization		Lab 2	Project 1 start
	29/10	Least Squares, Max Likelihood			
	01/10	Overfitting, Ridge Regression, Lasso		Lab 3	
	06/10	Generalization, Model Selection, and Validation			
	08/10	Bias-Variance decomposition		Lab 4	
	13/10	Classification			
	15/10	Logistic Regression		Lab 5	
NF	20/10	Generalized Linear Models			
	22/10	K-Nearest Neighbor		Lab 6	
	27/10	Support Vector Machines			Proj. 1 due 26.10.
	29/10	Kernel Regression		Lab 7	
	03/11	Neural Networks – Basics, Representation Power			Project 2 start
	07/11	Neural Networks – Backpropagation, Activation Functions		Lab 8	
	10/11	Neural Networks – CNN, Regularization, Data Augmentation, Dropout			
	12/11	Adversarial ML		Lab 9	
	17/11	Adversarial ML			
	19/11	GANs		Lab 10	
MJ	24/11	Unsupervised Learning, K-Means			
	26/11	Gaussian Mixture Models		Lab 11 & Q&A	
	01/12	EM algorithm			
	03/12	SVD and PCA		Lab 12	
	08/12	Matrix Factorizations			
	10/12	Text Representation Learning		Lab 13	
	15/12	Invited Guest Lecture			
	17/12	projects			Proj. 2 due 17.12.

Course Logistics

Lectures

Tuesday 2x45mins, Room: youtube

Thursday 2x45mins, Room: youtube

We provide PDF lecture notes on our webpage and GitHub

Exercises

Course Logistics

Thursday 14:15 - 16:00 - live interaction!

Room alternatives:

- On zoom (link see moodle)
- On campus (according to covid schedule, and lastnames)

[INF119](#) (A-C),

[INF2](#) (D-He),

[INJ218](#) (Ho-Me)

[INM202](#) (Mf-Se),

[INR219](#) (Sh-Z)

All labs and projects are in **Python**.

See the first lab to get started.

Code Repository for Labs: github.com/epfml/ML_course

Course Logistics

TA team

Maksym Andriushchenko

Arnout Devos

Semih Günel

Prakhar Gupta

Mahdi Hajibabaei

Sai Praneeth Karimireddy

Anastasiia Koloskova

Tao Lin

Zohreh Mostaani

Guillermo Ortiz Jimenez

Matteo Pagliardini

Scott Pesme

Aswin Suresh

Thijs Vogels

contact us: moodle forum!

Mohamed Chahed

Paul Griesser

Haitham Hammami

Wei Jiang

Stanislas Jouven

Maja Stamenkovic

Robin Zbinden

Course Logistics

Projects

- ✿ **Project 1 (10%), due Oct 26th**
- ✿ **Project 2 (30%), due Dec 17th**

Real-world problems, Python, Groups of 3 Students

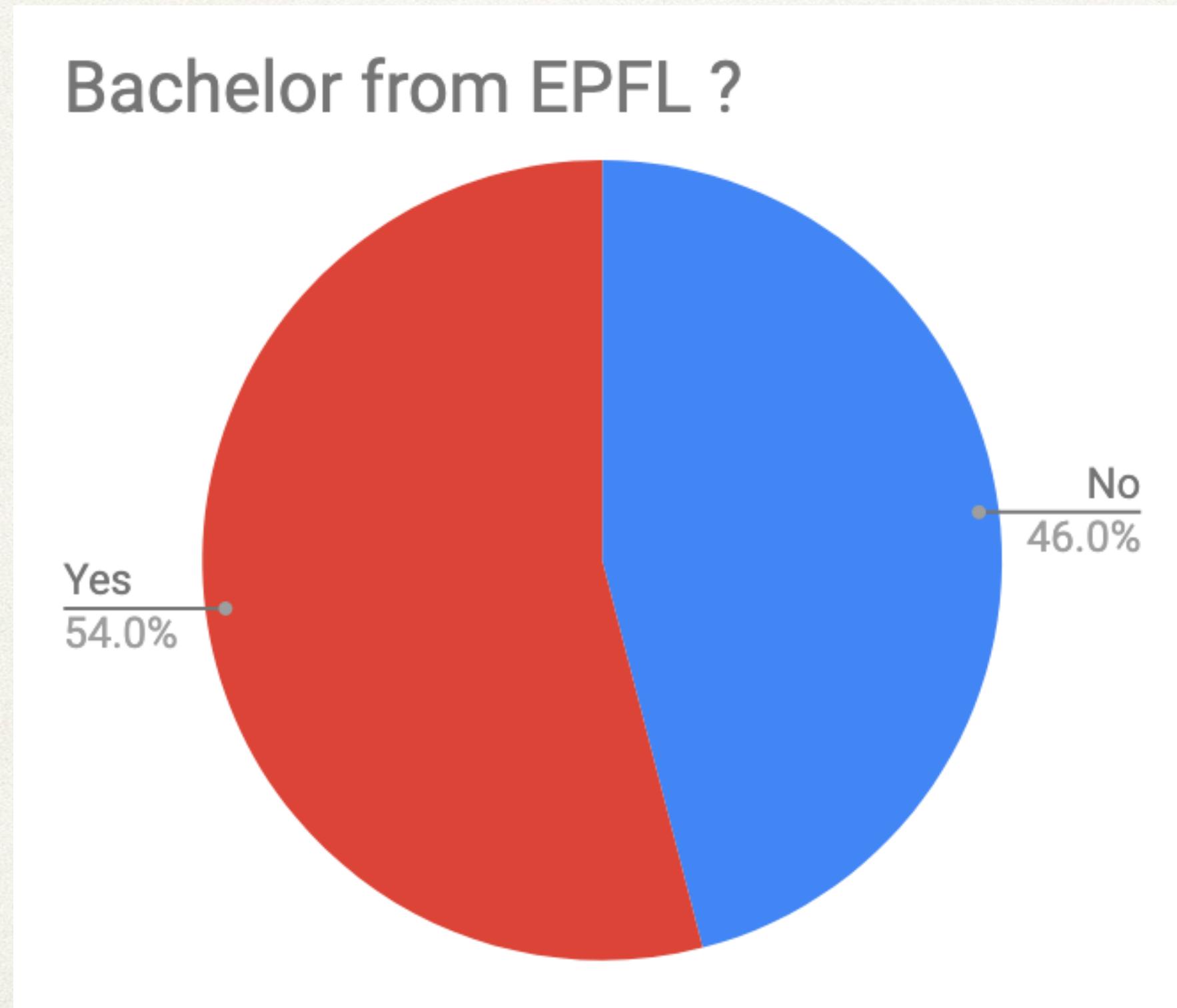
What to expect?

- ❖ **overview over ML**
- ❖ **basic understanding of most important ML methods and fundamental concepts**
- ❖ **experience how ML is done on a practical problem**

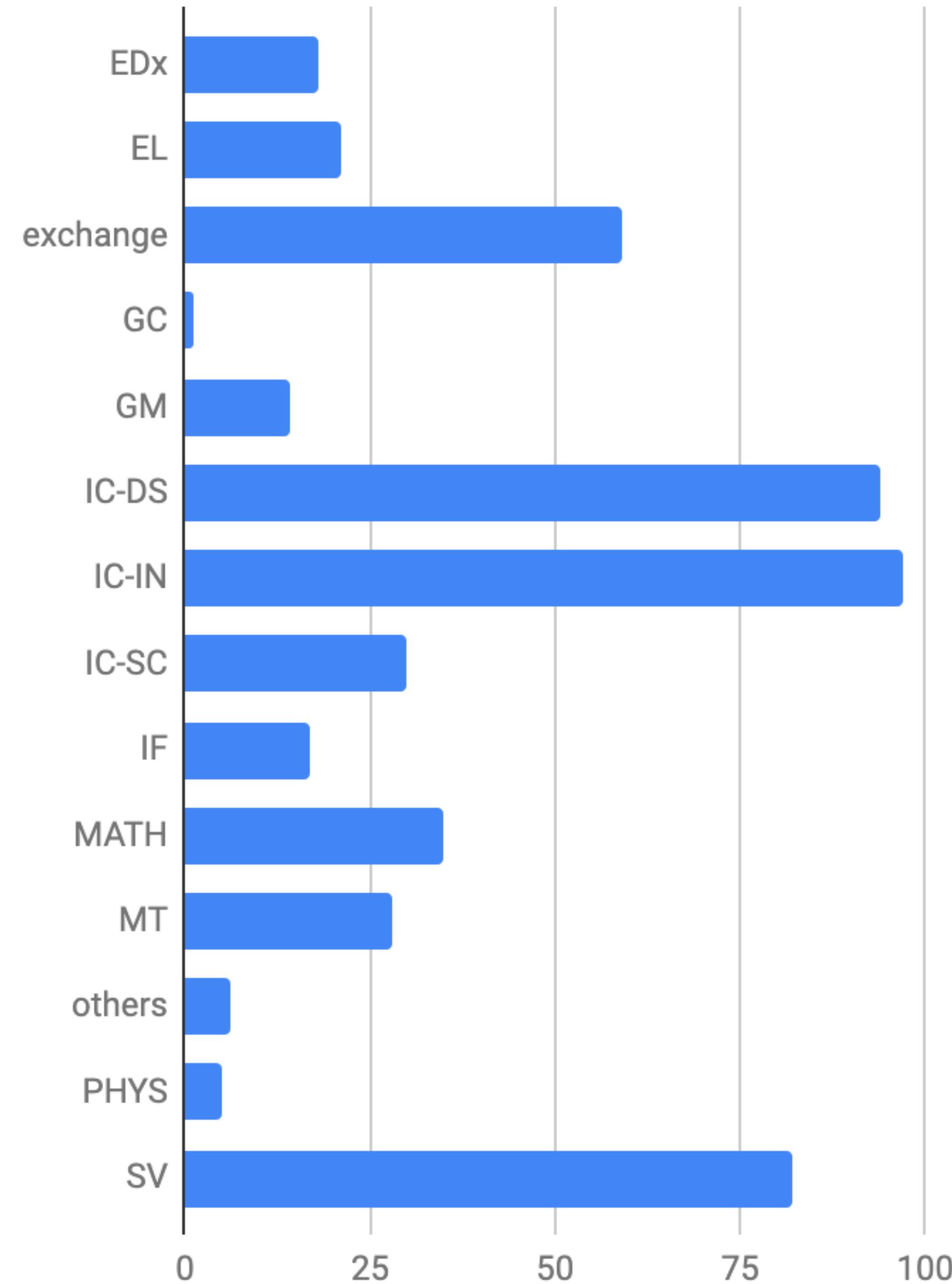
What not to expect?

- ✿ instantly becoming an **expert** data scientist / ML researcher / statistician / large-scale ML specialist / ML software expert / domain expert
- ✿ learn about that newest **hack** you saw on blog *xyz*.
- ✿ build your own **self-aware** AI

Your colleagues here

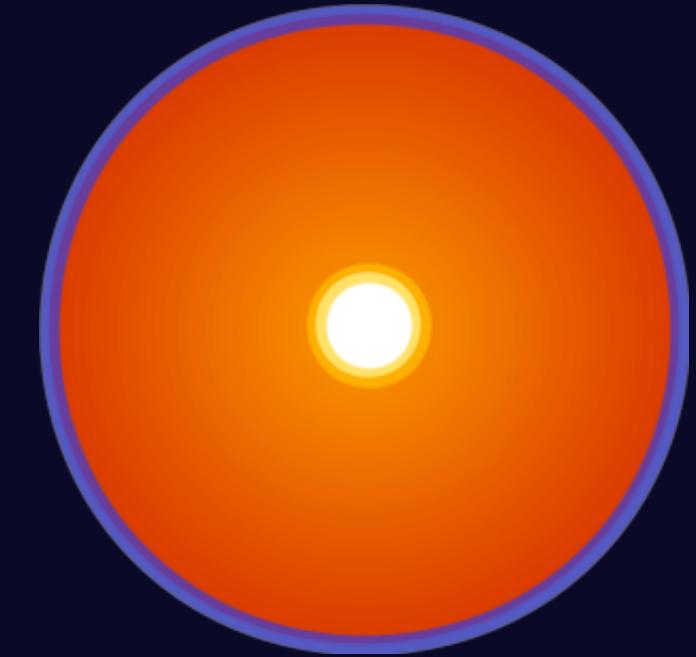


Count of Section



Learn Machine Learning with Machine Learning!

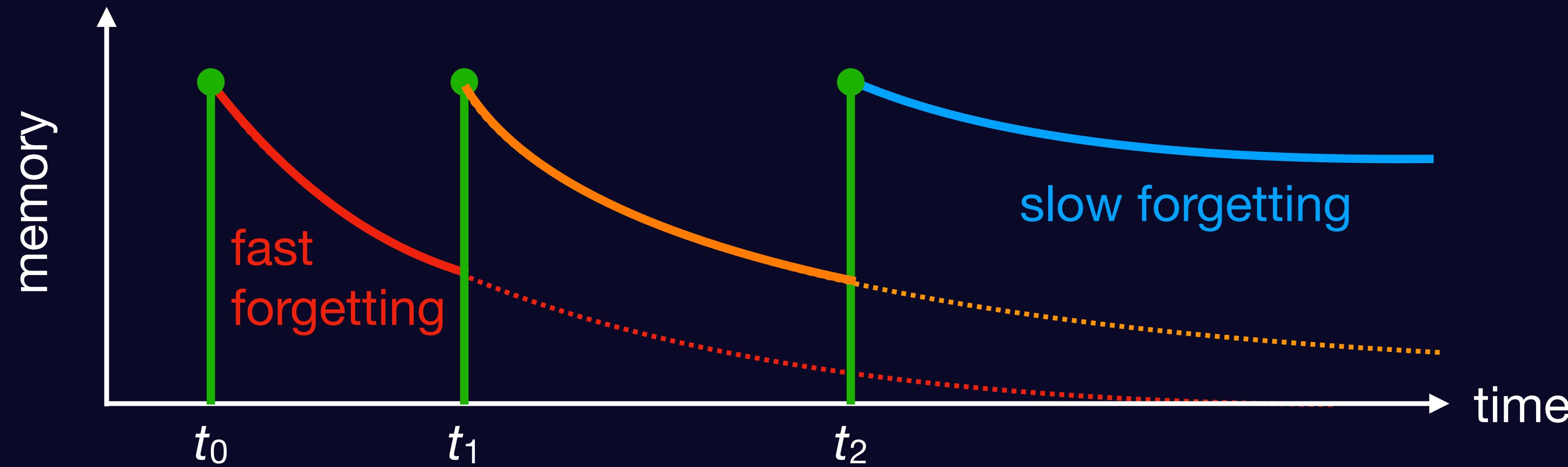
Ari is an AI tutor app that personalizes your revisions of the concepts in CS-433



ari9000.com

Spaced Repetition

You forget everything you learn exponentially fast.



Ari can determine optimal revision timing for you:

You will remember better with less effort!

Introduction

What is Machine Learning?

What is Machine Learning?

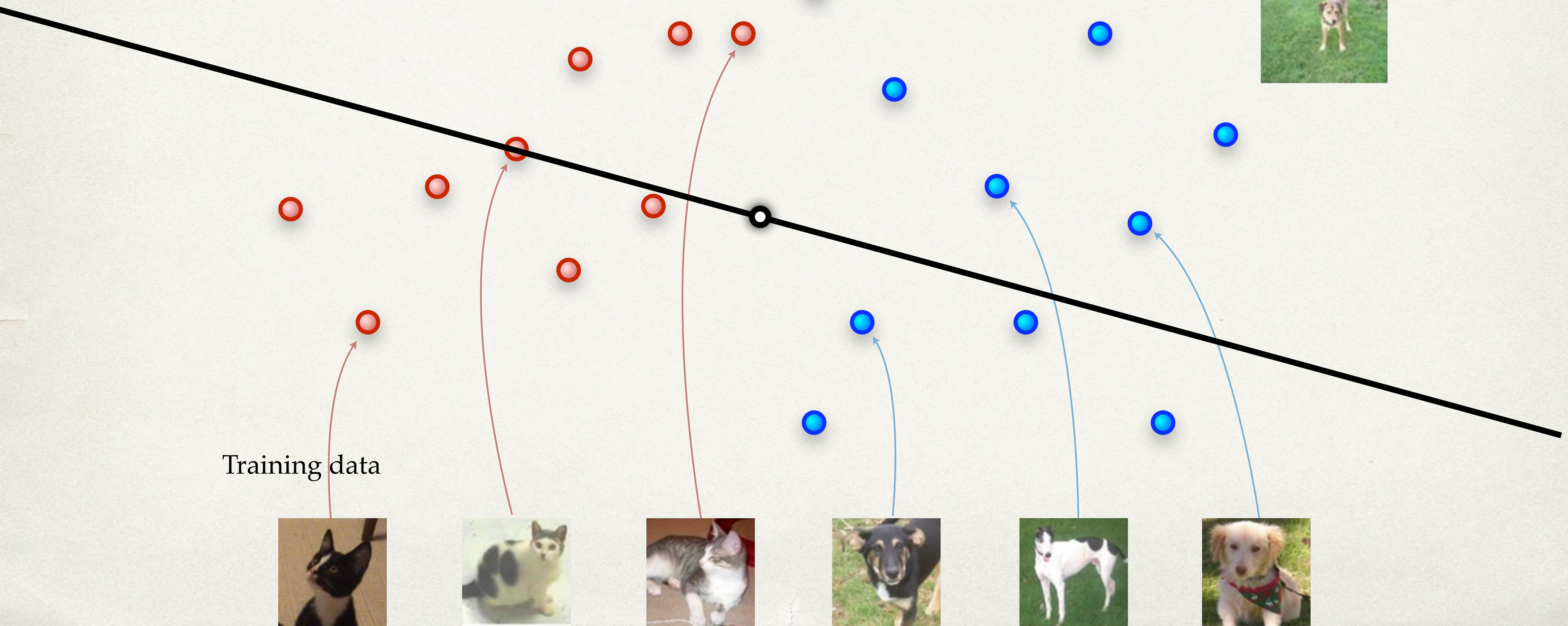
algorithms that can
learn from data



[image source](#)

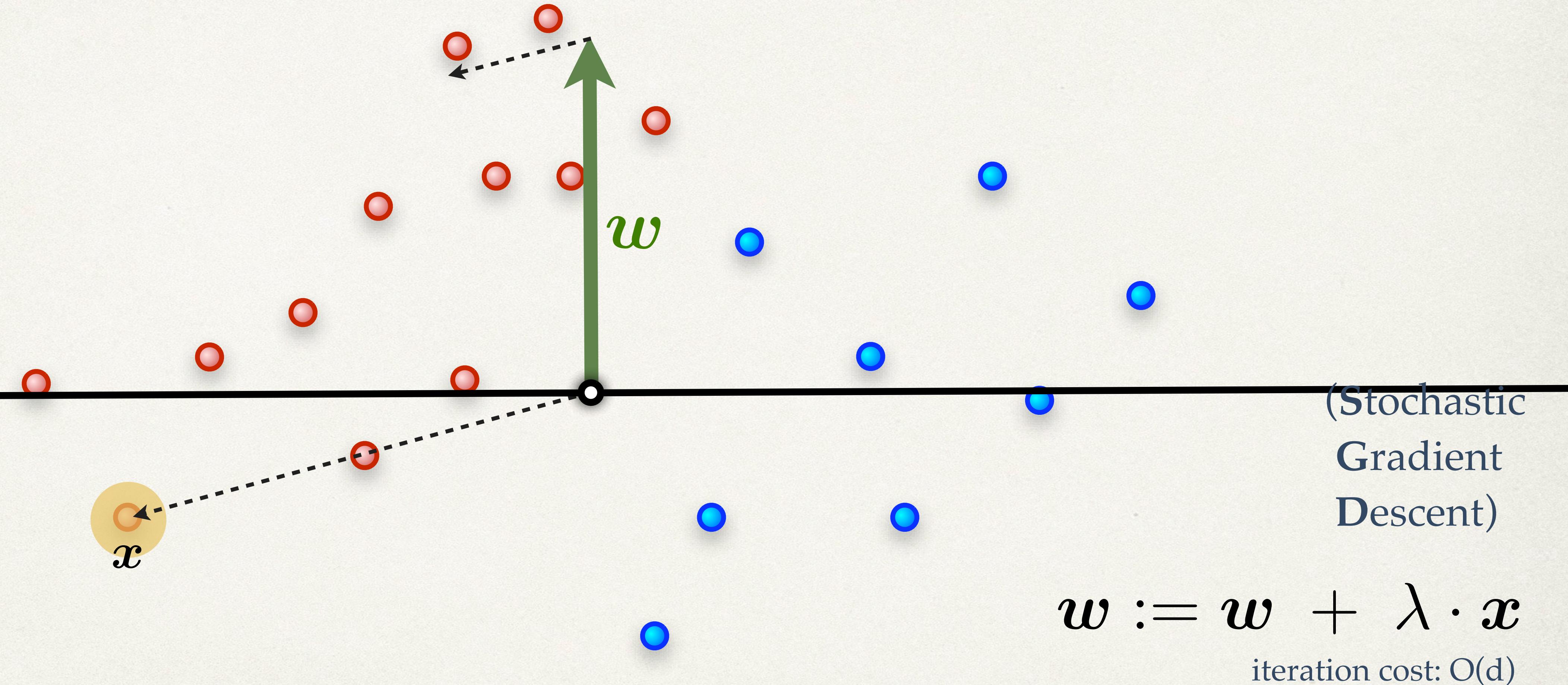
Classification

$$\boldsymbol{x}_i \in \mathbb{R}^d$$



The Learning Algorithm

$$\mathbf{x}_i \in \mathbb{R}^d$$



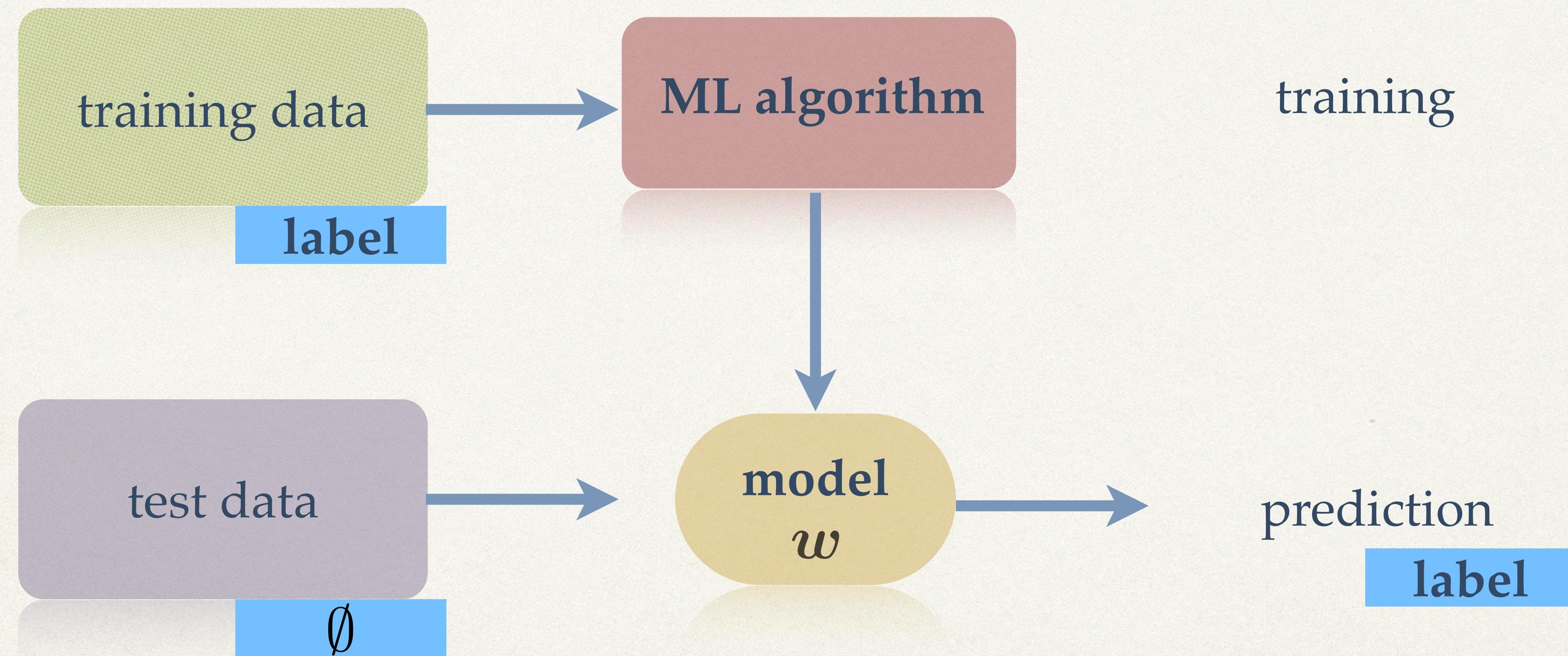
Perceptron

(Rosenblatt 1957)

Support-Vector-Machine

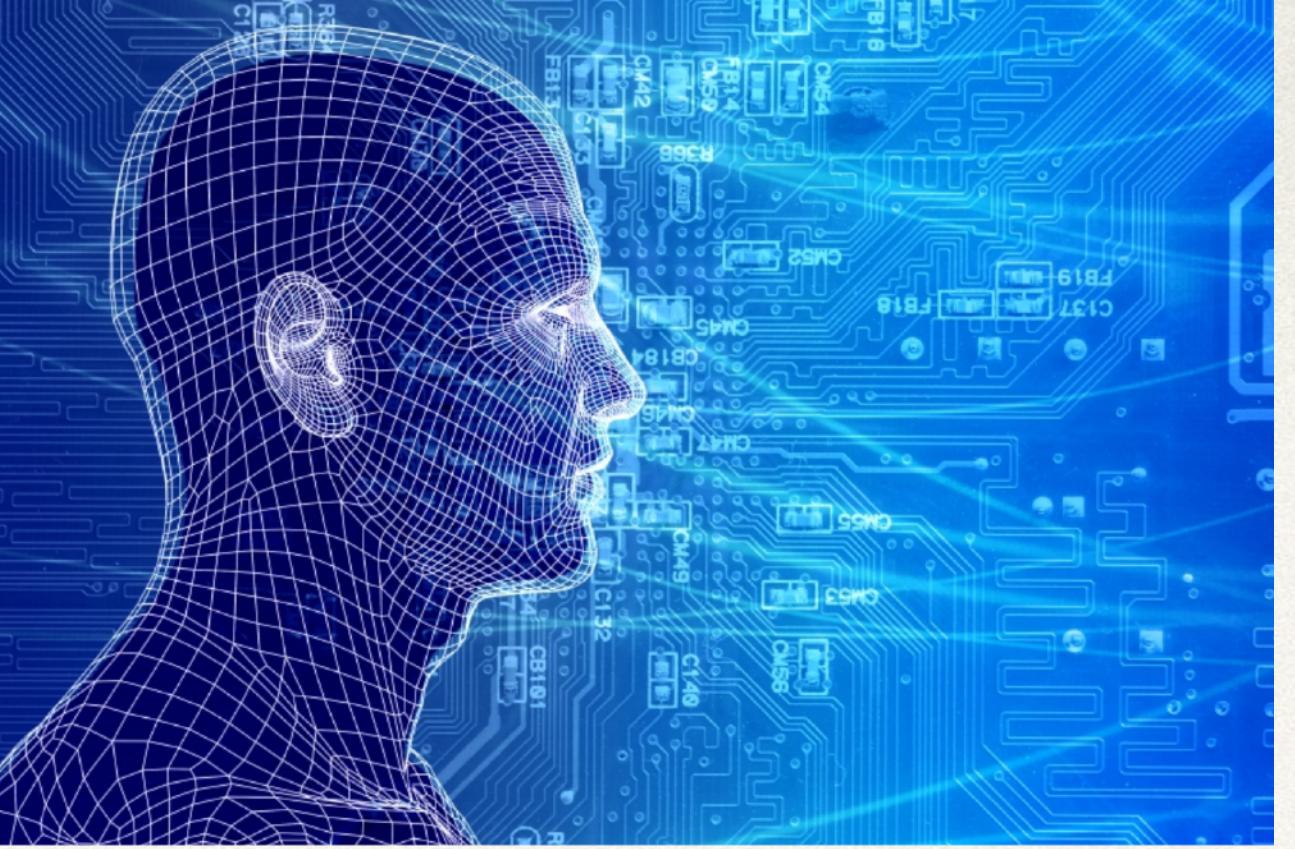
(Cortes & Vapnik 1995)

Machine Learning Fundamentals



what Evaluation Metric?





towards...
understanding intelligence
?

Machine Learning

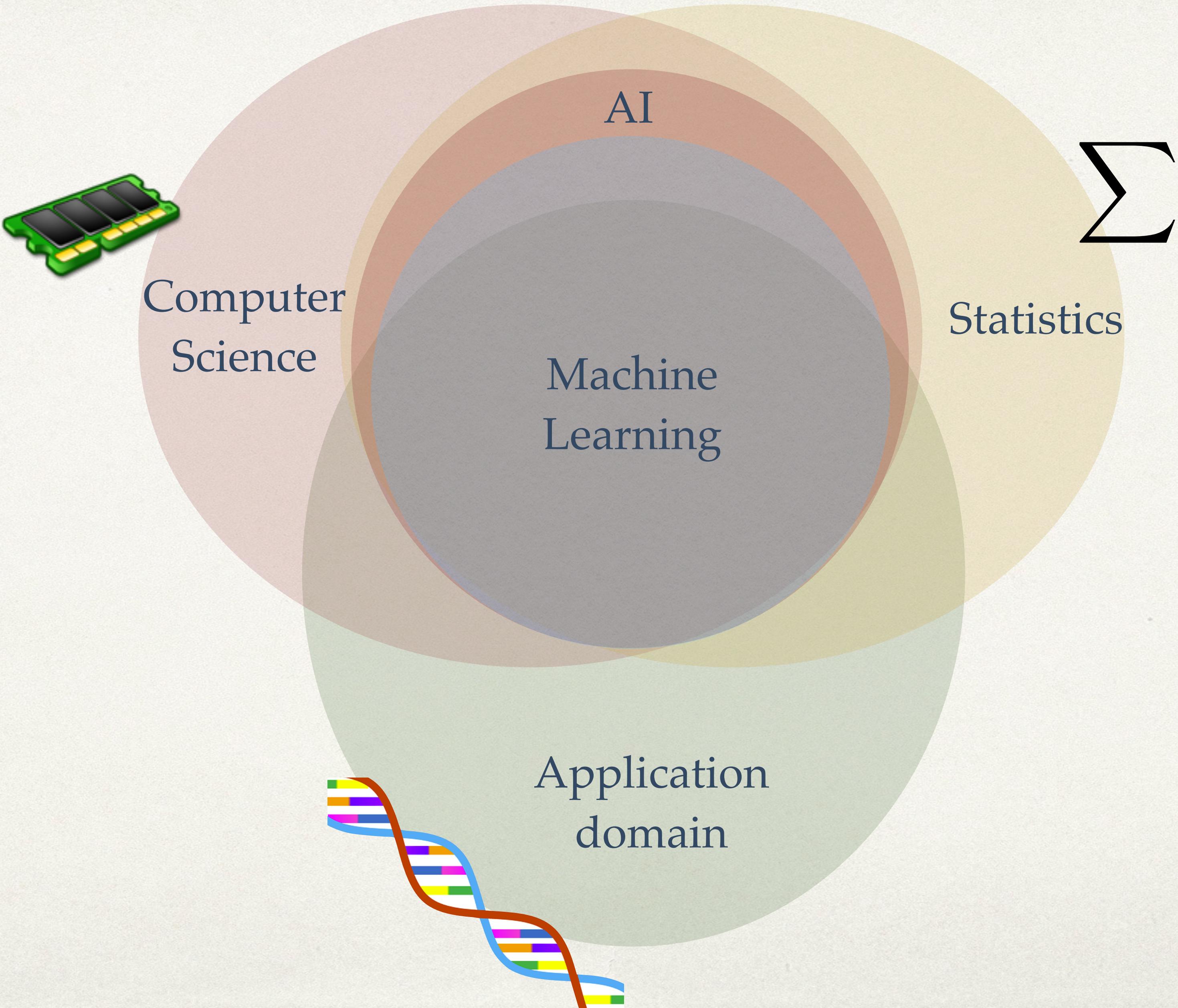


vs

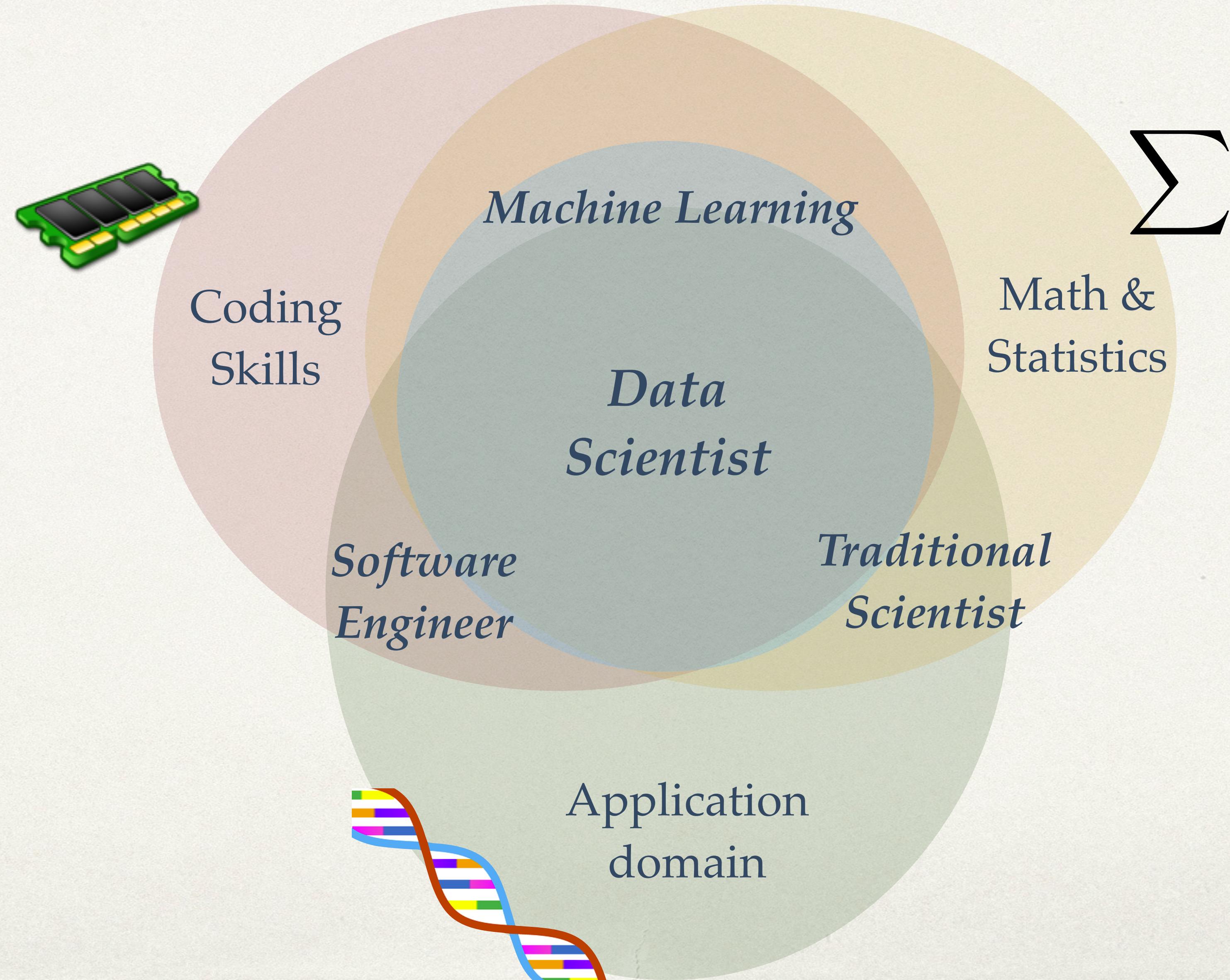
Neuroscience / HBP



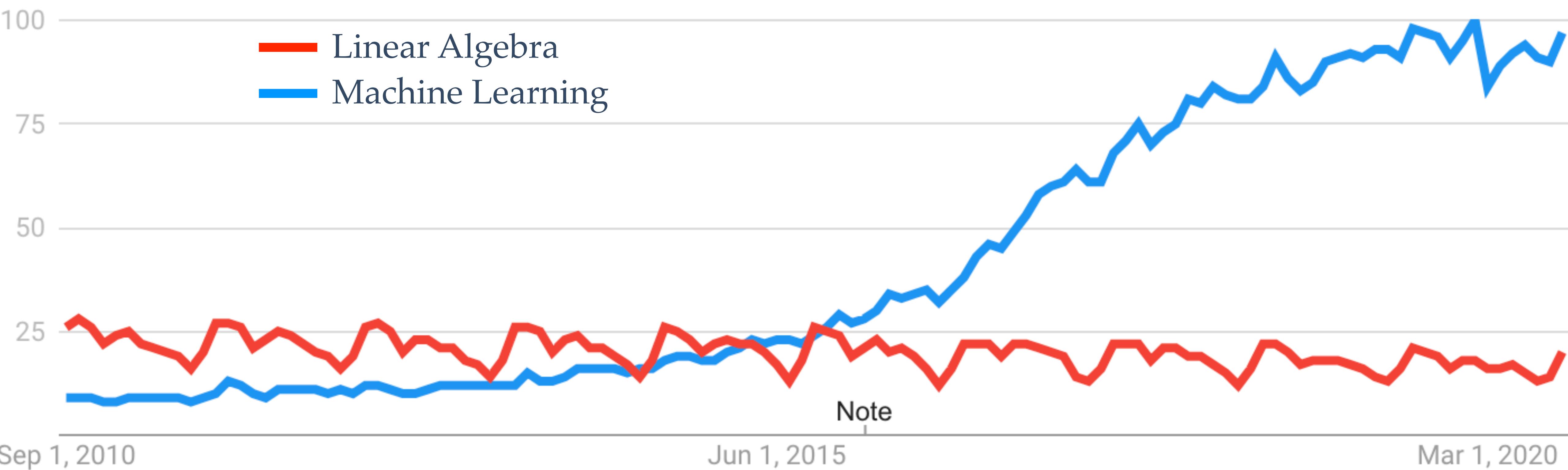
What is the difference between
Artificial Intelligence, Data Mining,
Statistics, Machine Learning?



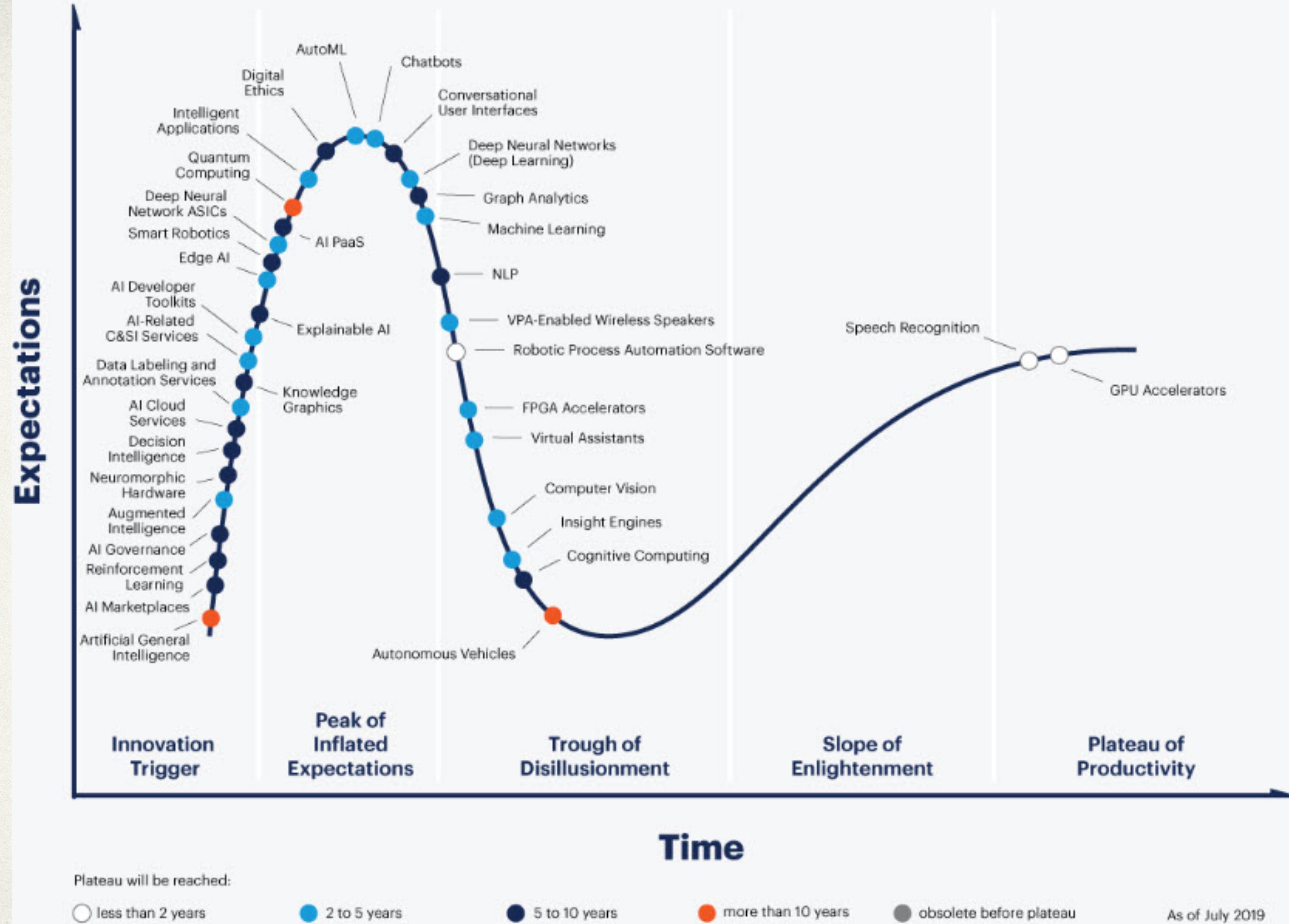
Job Skills



Evolution of Interest



source



SOURCE

why ML?

Applications

Industry Applications

- ✿ majority of industries, originally not 'digital':
 - ✿ agriculture, NGOs, 'sharing economy', logistics, delivery, services, manufacturing, sports, personalized health, call centers, entertainment, ...
- ✿ **not only** the 'usual suspects'

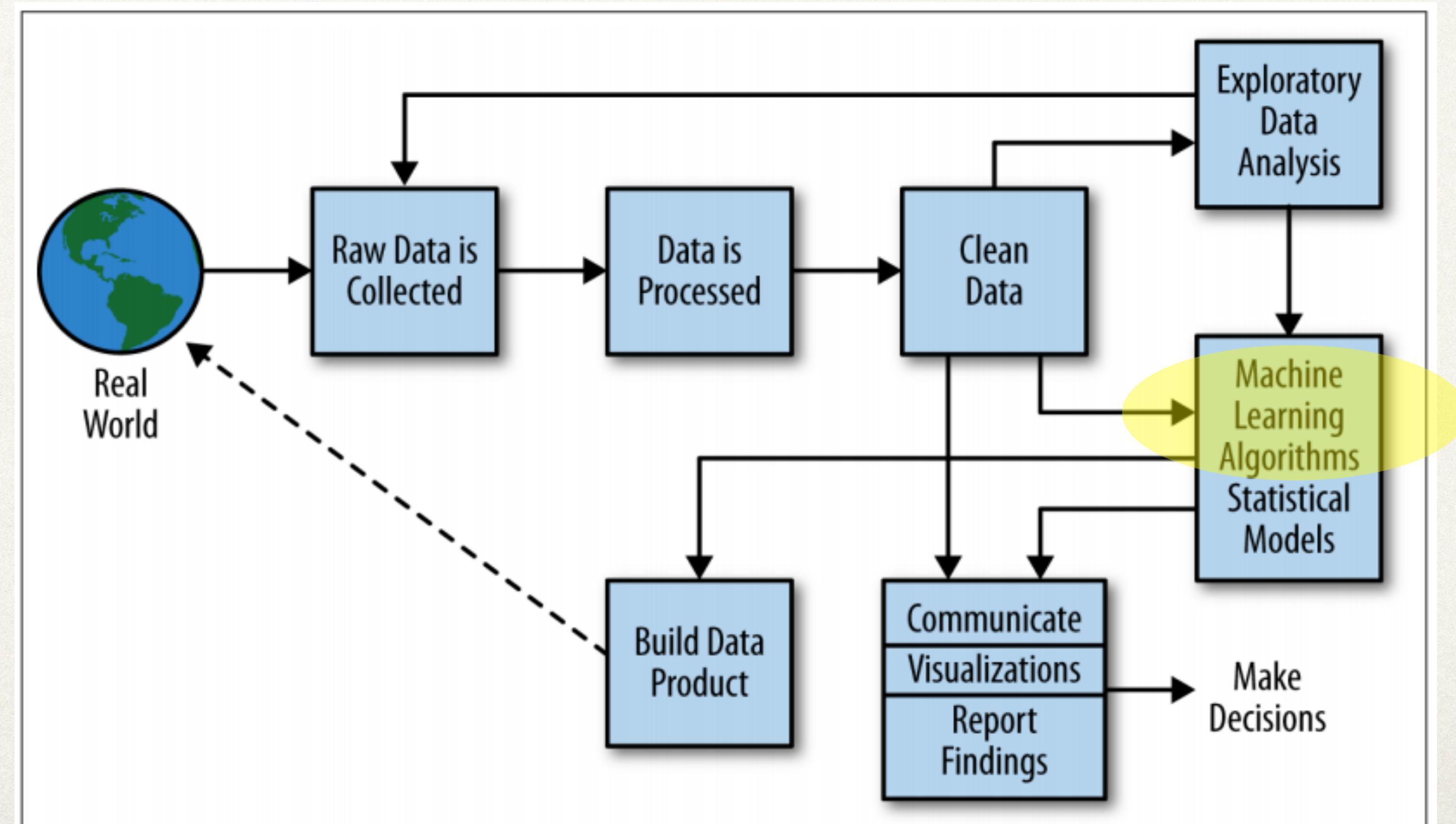


Applications in Other Sciences

- ✿ increasingly data driven
 - ✿ ... Psychology, Economics, Medicine, Social sciences
 - ✿ science of X → *digital* science of X

en.wikipedia.org/wiki/Index_of_branches_of_science

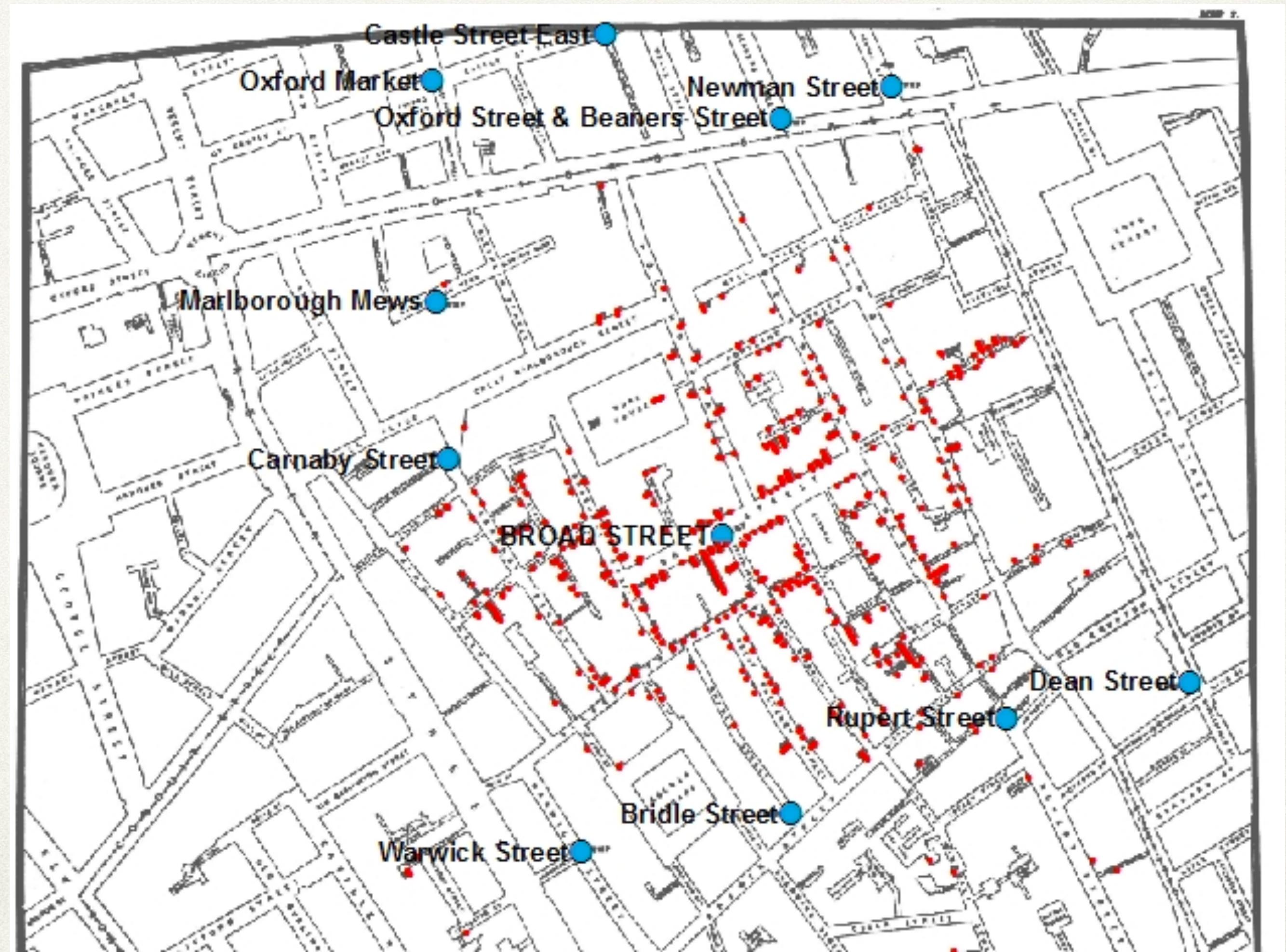
ML is only a small part!



History

- ❖ ML is not new!

London 1854
cholera outbreak



History

- ❖ ML is **not new!**
- ❖ the early days - 1950^{ies} and 1960^{ies}
 - ❖ Neural networks
 - ❖ Turing



What has changed?

1950s: 10^3 FLOPS

2017: 10^{17} FLOPS

“the embryo of an electronic computer that ... will be able to walk, talk, see, write, reproduce itself and be conscious of its existence.”

1958



Challenges

- ❖ Hype
 - ❖ cycles of AI popularity
- ❖ Data Ethics, Privacy, Fairness

[Alle Kategorien ansehen](#)Suche [Sport & Freizeit](#)[Erweiterte Suche](#)[Camping & Outdoor](#)[Fitness](#)[Fußball](#)[Golf](#)[Pulsuhren & GPS](#)[R](#)

Baseballschläger Aluminium "32" von Tysonz

von [Unbekannt](#)Noch keine Kundenrezensionen vorhanden: [Schreiben Sie die erste!](#)Preis: **EUR 21,90**

Alle Preisangaben inkl. MwSt.

Auf Lager.Verkauf und Versand durch [knockout24](#). Für weitere Informationen klicken Sie auf den Verkäufernamen.

Noch 4 Stück auf Lager.

[2 neu ab EUR 21,90](#)[Größeres Bild](#)[Für Kunden: Stellen Sie Ihre eigenen Bilder ein.](#)

Kunden, die diesen Artikel gekauft haben, kauften auch



[Pfeffer 2442 KO
Verteidigungsspray JET
40 ml](#) von Pfeffer KO

EUR 5,22



[Sturmhaube 3-loch
Acryl schwarz](#) von
Unbekannt

EUR 1,38



[Pfefferspray KO-FOG
40ML](#) von Pfeffer KO

EUR 5,19



[KO-CS
Verteidigungsspray 40
ml](#) von KO-CS

EUR 4,22

Challenges

- ❖ Hype
 - ❖ cycles of AI popularity
- ❖ Data Ethics, Privacy, Fairness
- ❖ Lack of Interpretability
 - ❖ example: medical applications of deep learning
- ❖ Social Implications of AI,
Threats from Super-human AI?
 - ❖ see Nick Bostrom, Yuval Harari

need: Scientific Method, Reproducible
Research, Open Source and Open Data

ML Applications by CS-433 Master Students

ML4science projects by your colleagues 2019

Machine Learning for Invisibility Cloak Architecture
Predicting PV Array Power Output Using an All-Sky Camera
Predicting bacteria evolution during cell cycle observations using a Recurrent Neural Network
Searching for Similar YouTube Channels
Stress Classification from Biosignals using Neural Networks
Crystal-Structure Descriptor for Binary Materials Based on Coordination Numbers
Polymers identification using nanopores
ML for Science – Generate images of Mesopotamian artifacts
Machine Learning for Toxicological Testing
Global gene expression analysis: determine hormone signalling activation in human breast cancer samples
Yeast Cell Segmentation with U_Net: Effects of Weight Maps and Attention Gates
Machine Learning for LHC performance optimization
Product Life Cycle Prediction Using Machine Learning
Clustering estrogen receptor-positive breast cancer tumors based on hormonal response type
Identifying the shape of the worms based on the hand-annotated images dataset
Applause Recognition of Live Concert Recordings for the MetaMedia Center
Machine Learning for EDS Data Decomposition
Twitter Astroturfing Detection
Removing Noise From Microscope Images Without Ground Truth
U-Net for Yeast Cell Segmentation
C elegans segmentation
FCN for neuronal semantic segmentation
Machine Learning Project 2 – Optimal Option Exit Strategies with Neural Networks
Bio Product Classifier
Segmenting yeast cells from microscope images
Instance Segmentation of Yeast Cells using Mask R-CNN
A machine learning approach to determine chemical shifts in NMR spectroscopy data
Detecting rooftop solar PV installations using CNN
A Physical-Interactive Pac-Man Game for Stroke Rehabilitation
Logistic Hairdressing
Prediction Forces for a Flapping Wing system using Linear Regression and Neural Networks Methods
Exploratory Analysis with MOOC data used for Blended Learning
Predicting Energy Building Consumption

Anomaly detection for energy consumption time series
Segmentation of Cracks on Laboratory Images
The _place_ (physical and conceptual) of advertising in video game magazines
Image Classification: Distinguish Murine from Human Cells in PDXs Models
PLIER for single-cell RNA sequencing data
Machine Learning project #2: Identifying Outer Divertor Legs On Images From Tokamak Experiments
Machine Learning: Gravitational Lens Finding
VNAV – No GPS, No Problem
Earthquake Detection from Seismological Data
Detecting Astroturfing Bots on Twitter
Feature identification of MANTIS data
ObsBox Instability Clustering
Photovoltaic Power Production in Swiss Communes
Detection of Strong Gravitational Lenses with Convolutional Neural Networks
Nano Manufacturing with ML and Raman Spectroscopy
Motion Prediction on Drosophila Using a Seq2seq Model
Segmentation of spinal cord images
Avalanche Intelligence
Pac-Man Error Project
Building Classification using Google Street-View
U-Net application for yeast cell segmentation
Landmark Based Visual Navigation for Drones
Classification: Crashes and Disruptions in Plasma Experiments
Automating Route Setting in Climbing via Deep Generative Models
Leading Edge Suction Prediction of a Dynamically Pitching Airfoil with Trailing Edge Flap
Aerodynamical parameters estimation from velocity, pressure and noise sensors data using temporal convolutional networks
Machine Learning for Science: Diffraction detection in rock structures
Classifying amino acid modifications using nanopore sequencing data
Microscopical Image Restoration of C. elegans using Noise2Noise
ML challenge : Strong Gravitational lensing
Instance segmentation
Invisibility cloak
Pacman collaboration with CHILI lab
Enchordings – Harmony Embeddings

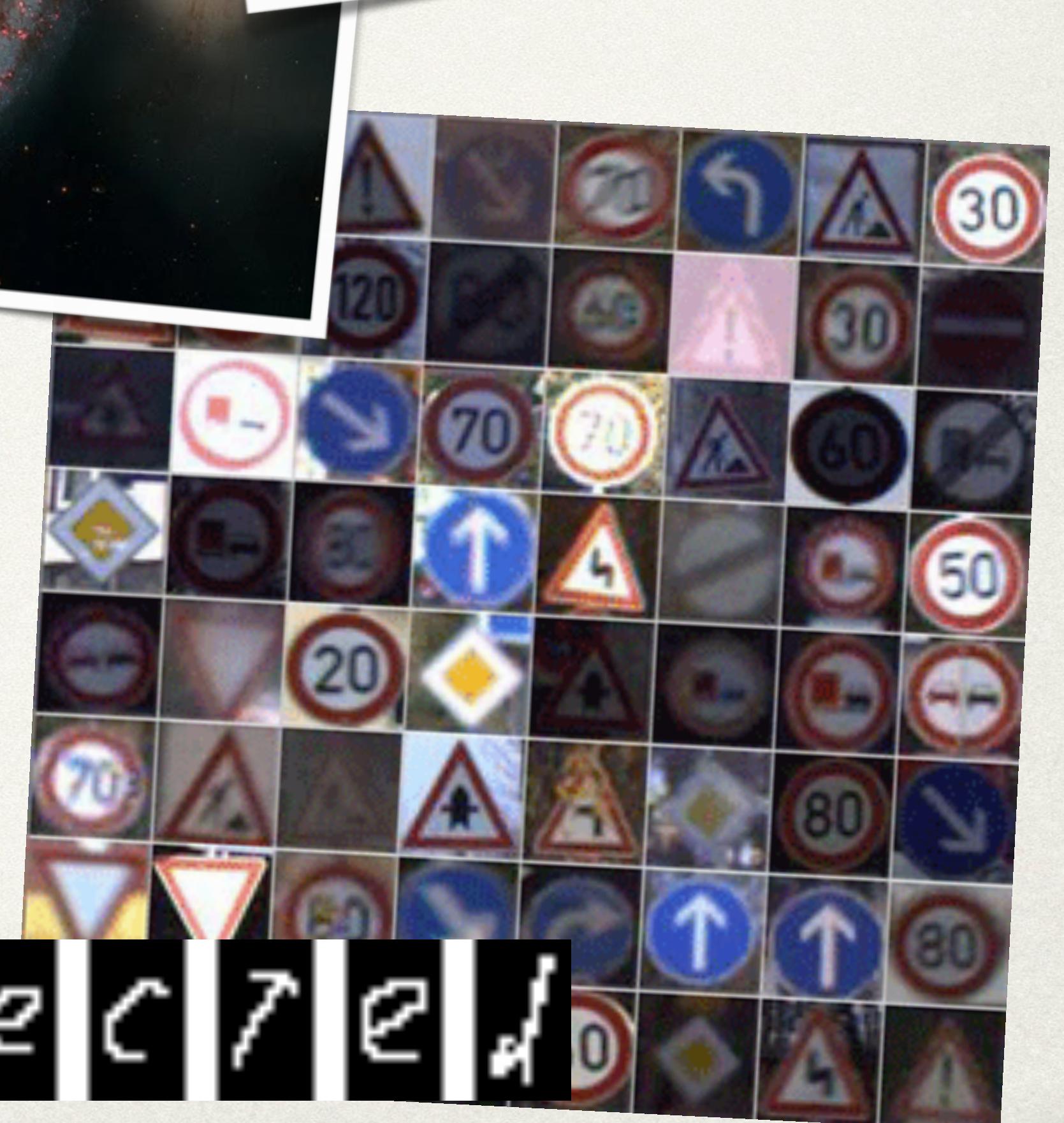
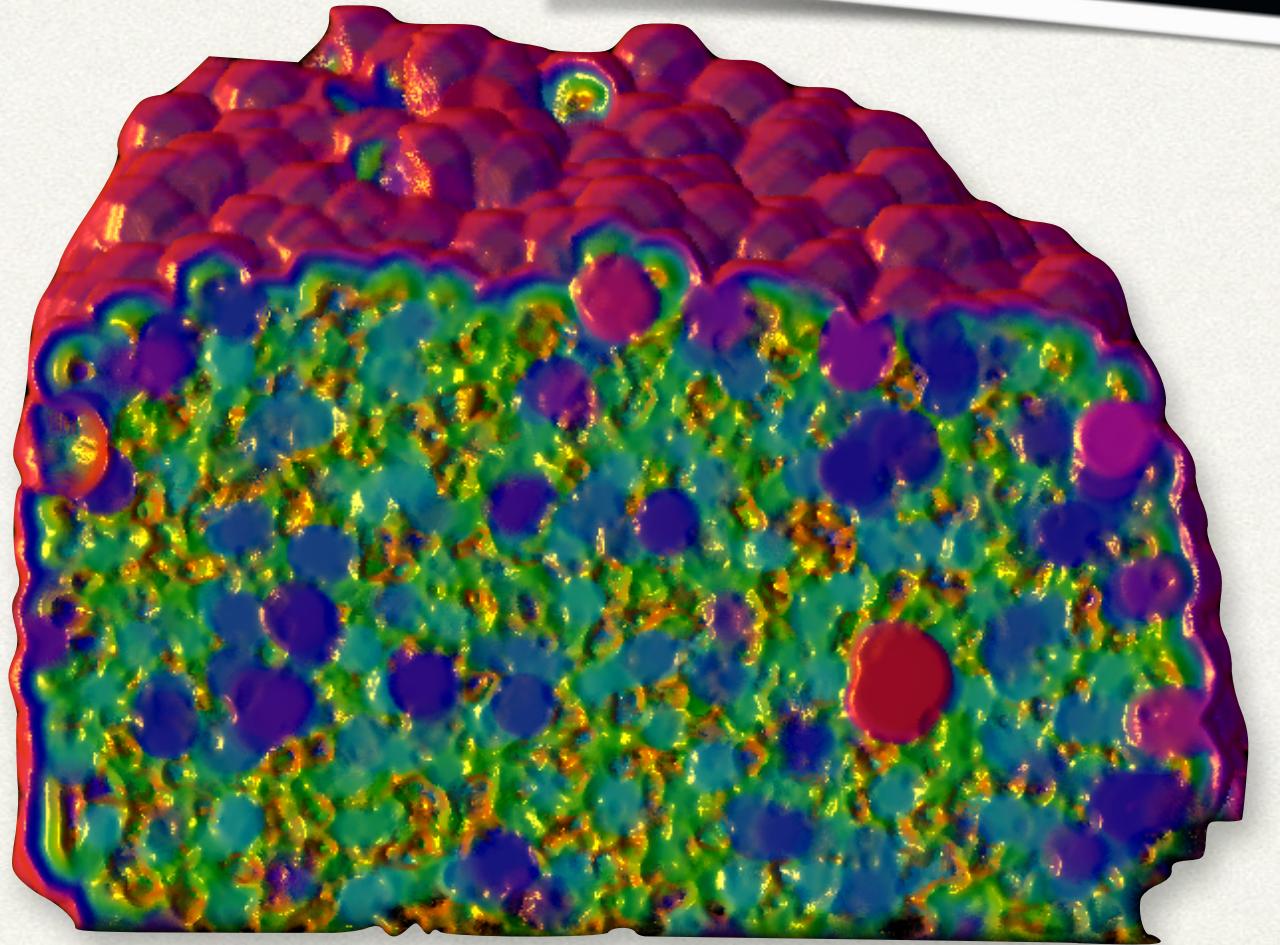
why ML?

Applications, cont.

Image Data

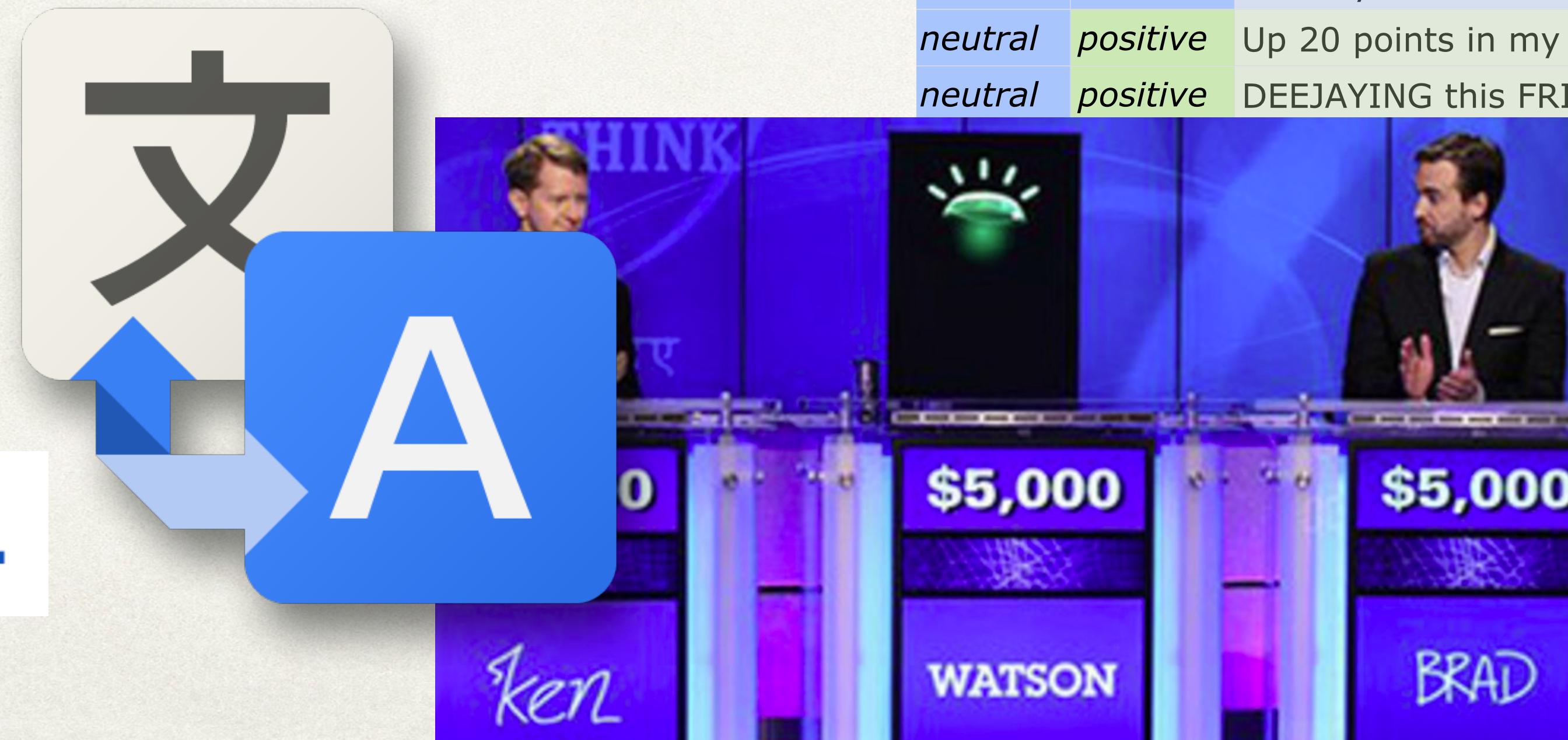
- ✿ Astronomy
- ✿ Face recognition
- ✿ 2D + 3D medical imaging
- ✿ OCR
- ✿ self-driving cars

how-old.net



Text Data

- ✿ Text Understanding & Text Generation
- ✿ Spam Detection
- ✿ User Content
- ✿ Medical Text
- ✿ Machine Translation



negative	neutral	But i wanna wear my Concord's tomorrow though but i don't feel like it
positive	neutral	Gonna watch Grey's Anatomy all day today and tomorrow(:
negative	neutral	@CoachVac heey do you know anything about UVA's fallll fest loll they invited me
neutral	neutral	@DustyEf when that sun is high in that Texas sky, I'll be buckin it to county fair. A
neutral	positive	Up 20 points in my money league with Vernon Davis and L. Fitz still to go tomorrow
neutral	positive	DEEJAYING this FRIDAY in THE FIRST CHOP it's CHRIS actual SMITH with a smash

signing that was scheduled for tomorrow at the Books A Million
oks like it! Was after El Clasico on Sunday. I didn't like her lol
ent for the 2nd time today!

basketball Game tomorrow at 6:00 pm Then Football Senior night

@Young__Assassin VS @jamievarner set for TUF 16 Finale on t

lide thru sometime this weekend ill have somethin yu can sip o

absolutely-- I meant out of the Bachmann, Perry, Santorum, H

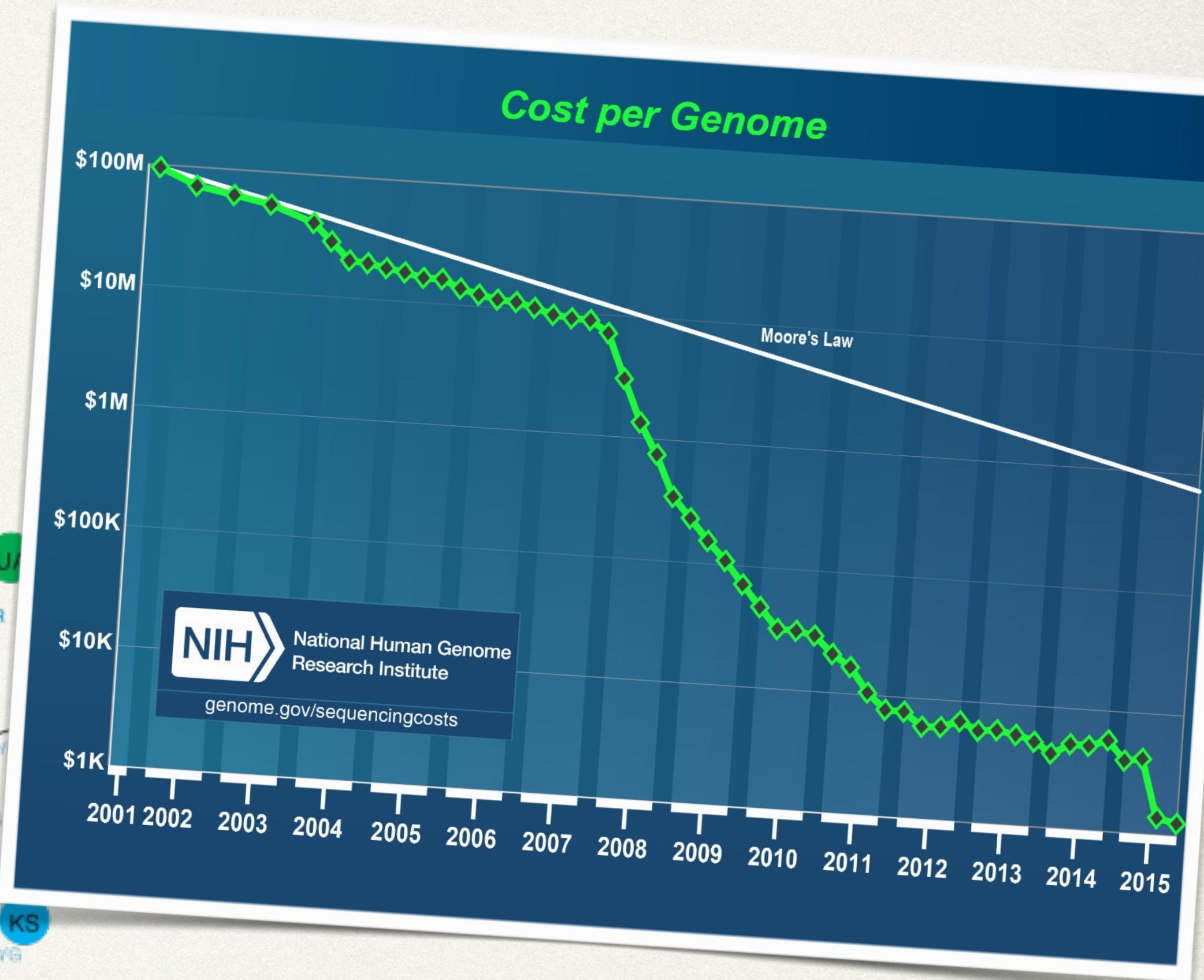
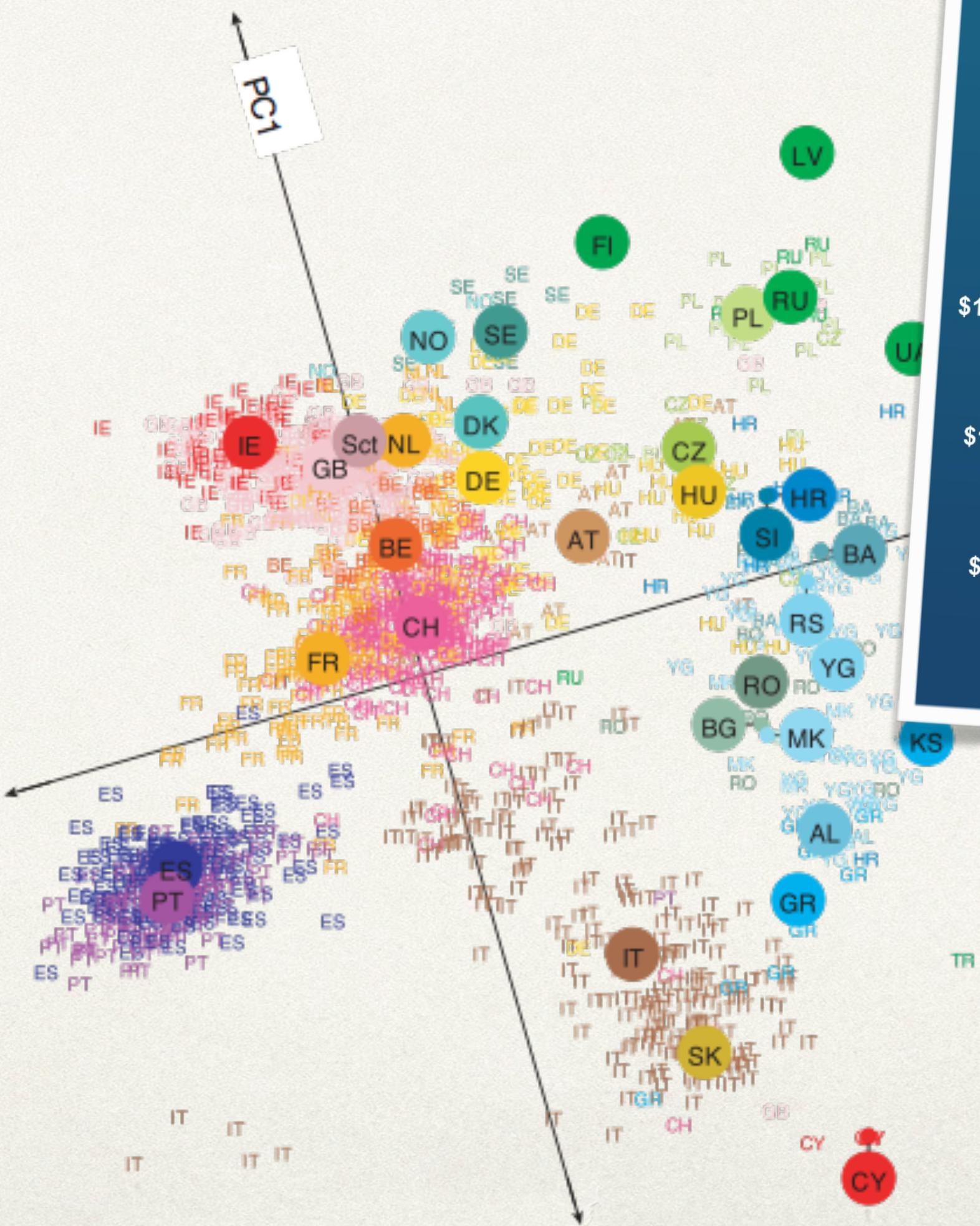
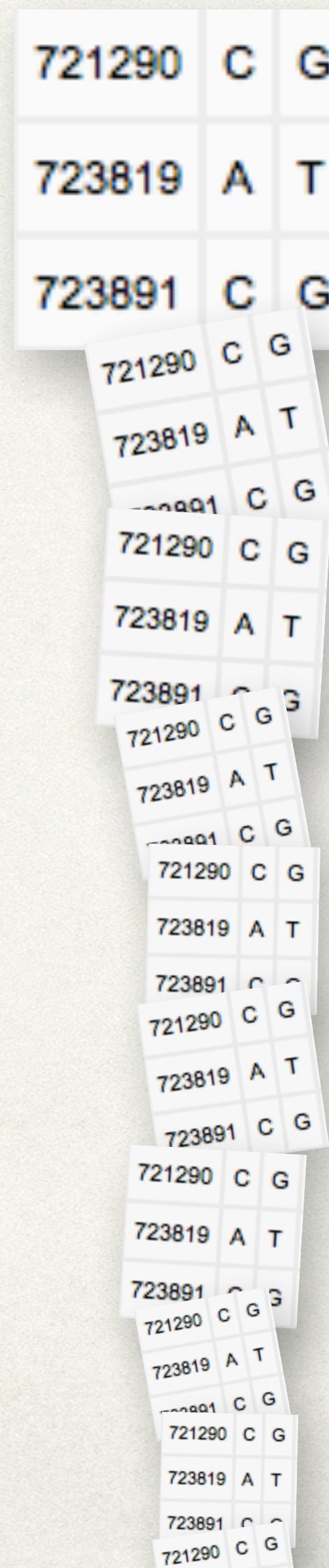
re Levein discussion on Wed. Can't keep changing boss, but he

vember 02, 1958 Elvis gave a party at his hotel before going o

ot to then kick back n party everyday like its Fri

oo exited about Vancouver tomorrow! I'm like a kid at Christmas

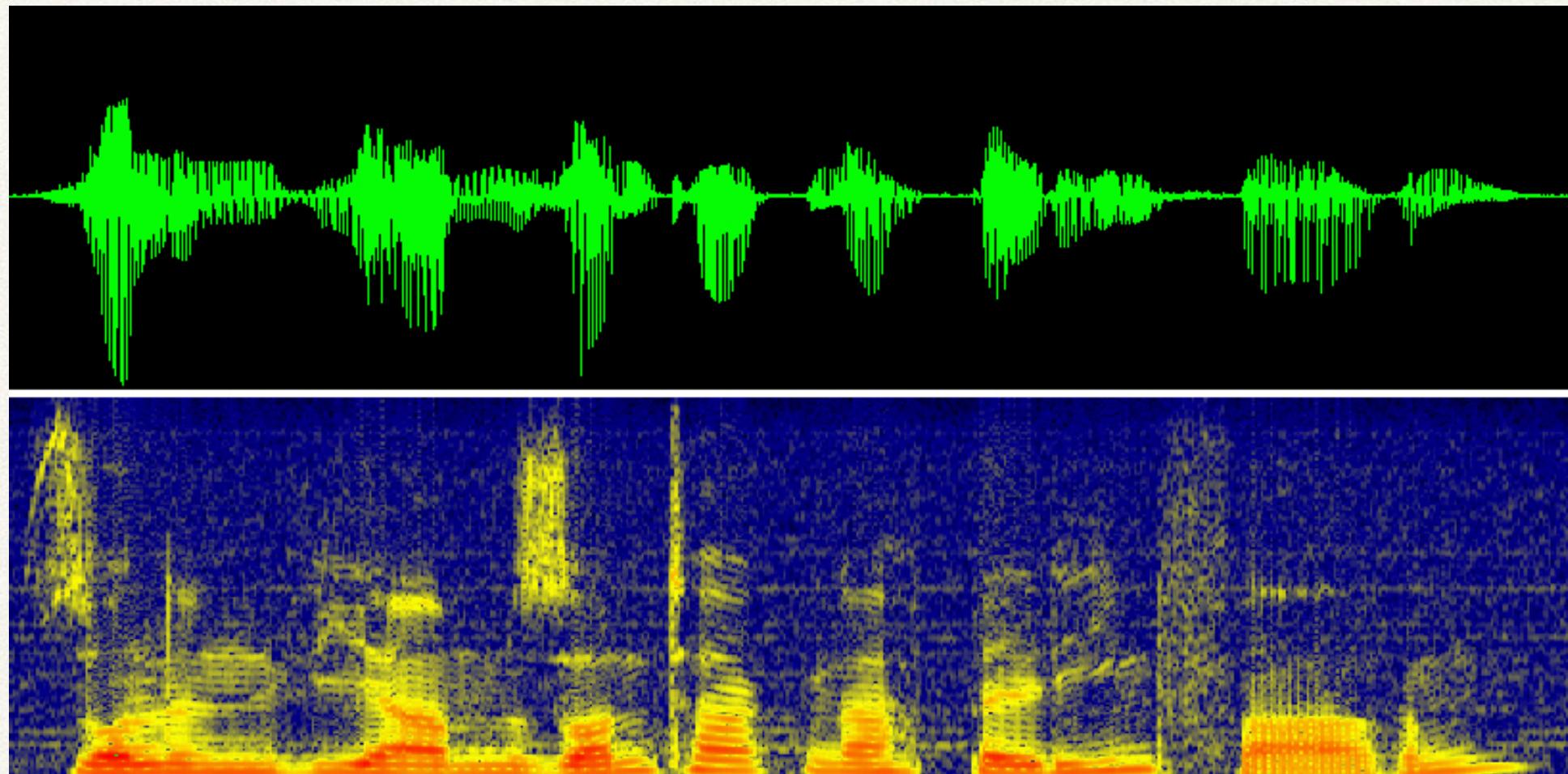
Medical: Genetic Data



ref

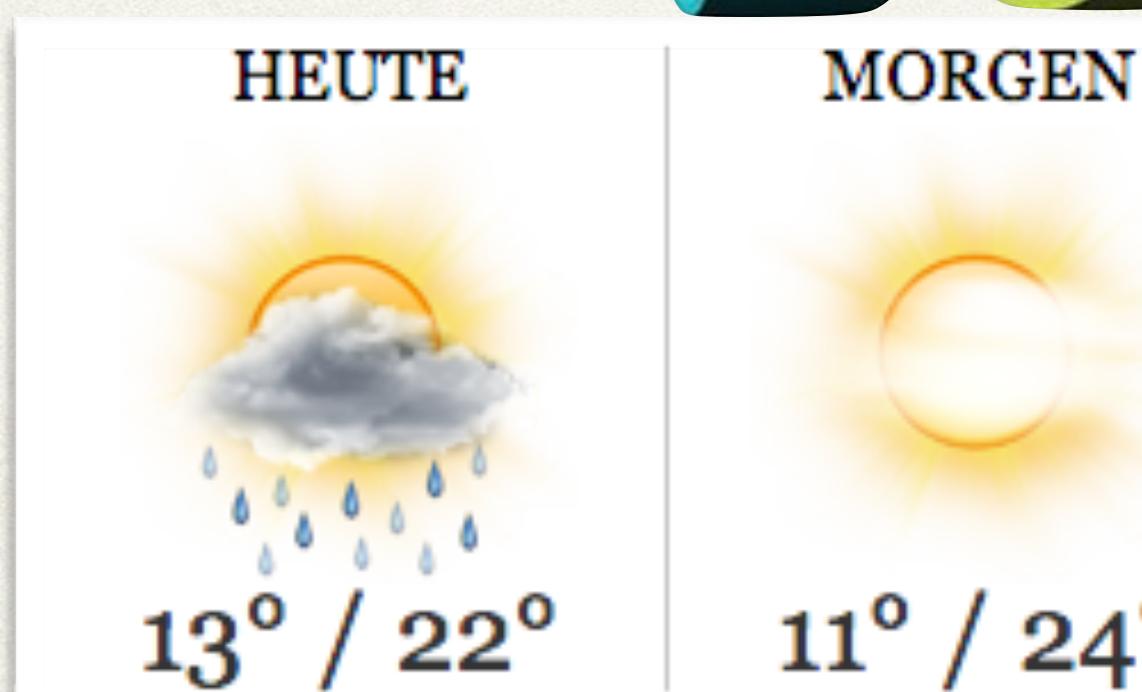
Audio & Multimodal Data

- ❖ Hearing aids
- ❖ Voice Recognition
- ❖ Automatic Translation
- ❖ Lip Reading
- ❖ Video Analysis



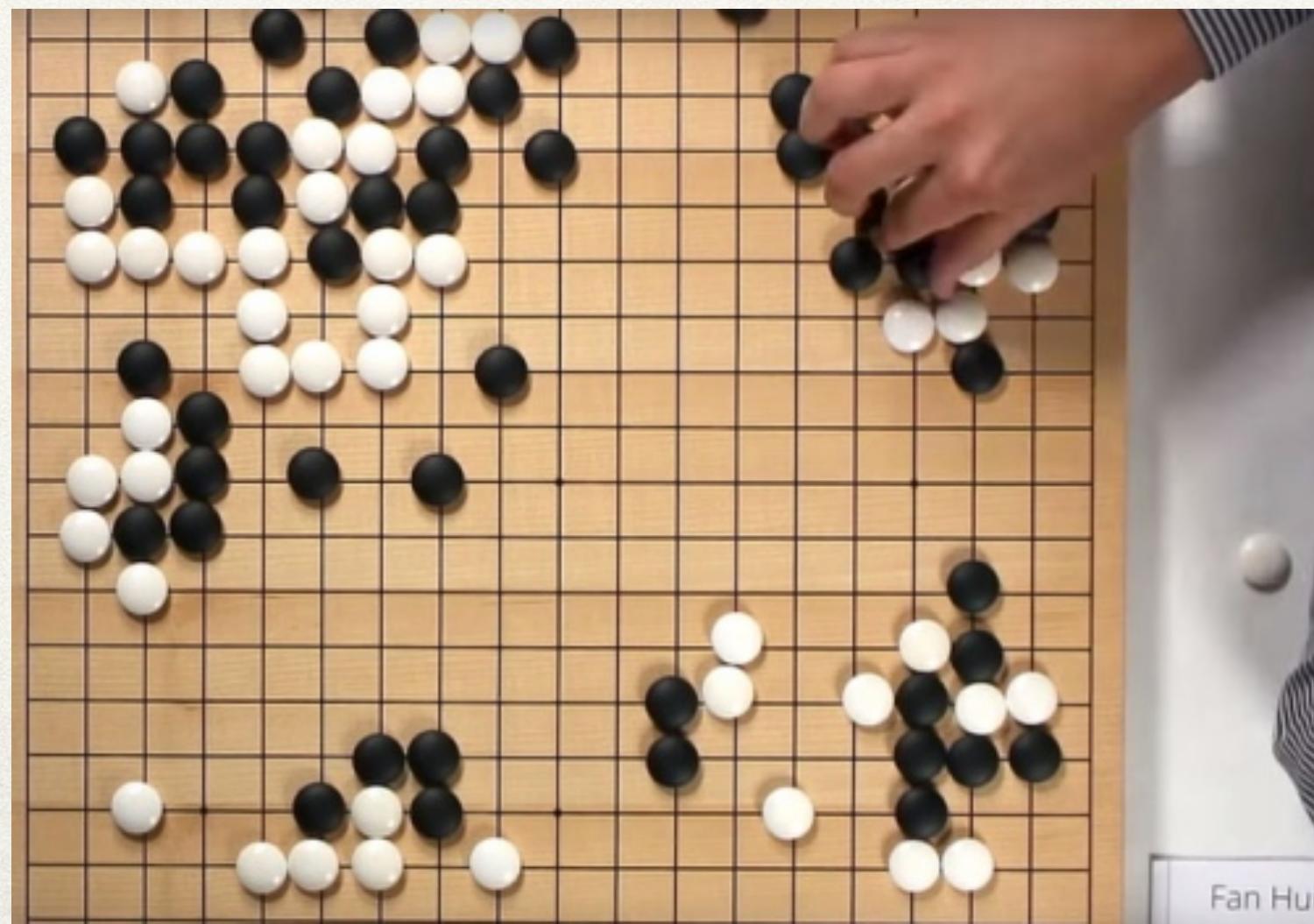
Numerical / Sensor Data

- ❖ Cern
- ❖ Astronomy / Telescopes
- ❖ Fitness Trackers
- ❖ Weather Forecast
- ❖ Robotics
- ❖ Kinect



Games & Simulations

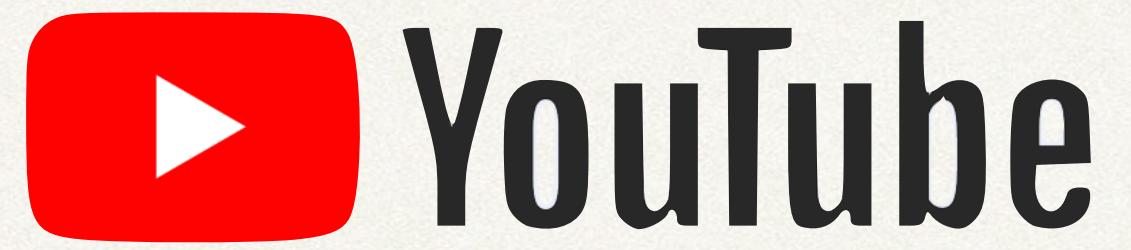
- ❖ Immediate Feedback
- ❖ Chess, Go
- ❖ Physical World



Internet Data



WIKIPEDIA



YouTube

Google

amazon.com[®]

Customers / Words

		Products / Words		
		★ ★	★	★ ★
		★	★	★
		★	★	★
		★	★	★
Customers / Words		★ ★	★	★ ★
		★	★	★
		★	★	★
		★	★	★
		★	★	★

$$\approx UV^\top$$

Recommender systems, Virtual assistants, Ads

New Opportunities?

Your turn

up next:

- ✿ Regression
- ✿ Linear Regression
- ✿ Classification
- ✿ ... fundamental concepts of ML