

Machine Learning

CS-433

Martin Jaggi & Nicolas Flammarion
EPFL

Alternatives

Master

- EE-559 – Deep Learning
- CIVIL-459 – Deep learn. for autonomous vehicles
- EE-411 – Fundamentals of inference and learning
- 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 Digital Humanities
- CS-430 – Intelligent Agents
- CS-439 – Optimization for ML
- CS-401 – Applied Data Analysis
- CS-449 – Systems for data science
- EE-556 – Mathematics of data
- CS-526 – Learning theory
- MICRO-570 – Advanced ML
- CS-433 – Machine Learning

Bachelor

- CS-233a – Intro to ML
CS-233b – 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
- EPFL Extension School – Applied Data Science: ML

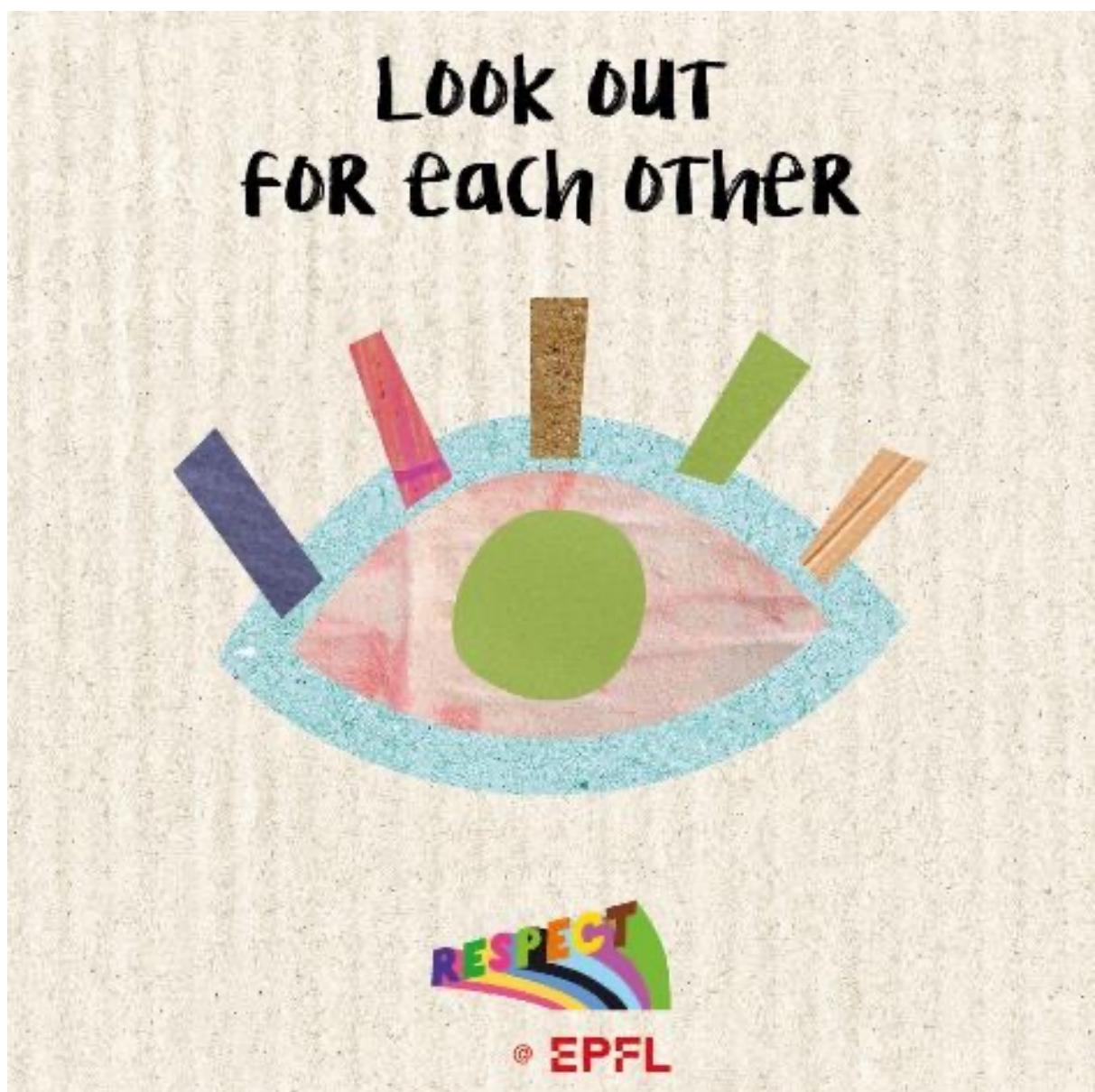
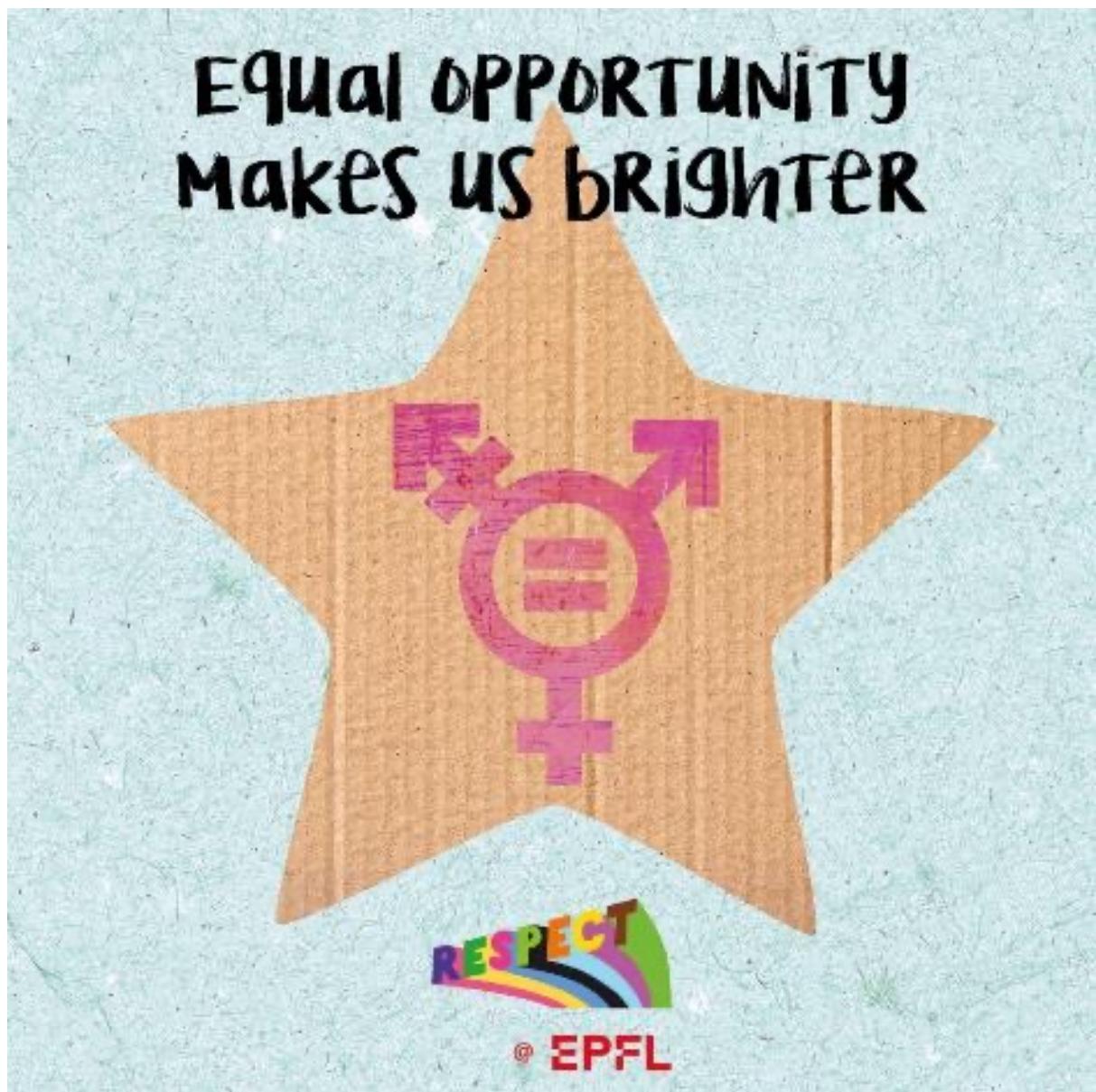
Talks: join mailing list:
ml@groupes.epfl.ch

Course Logistics

Assessment

- ✿ **Project 1 (10%), due Nov 1st**
- ✿ **Project 2 (30%), due Dec 23th**
- ✿ **Final exam (60%)**







Course Logistics

Lectures

tentative
schedule

Date	Topics Covered	Lectures	Exercises	Projects
21/9	Introduction, Linear Regression <u>01a,01b</u> <u>01c,01d</u>			
23/9	Cost functions		<u>Lab 1</u>	
28/9	Optimization			
30/9	Optimization		Lab 2	<u>Project 1 start</u>
05/10	Least Squares, Max Likelihood			
07/10	Overfitting, Ridge Regression, Lasso		Lab 3	
12/10	Generalization, Model Selection, and Validation			
14/10	Bias-Variance decomposition		Lab 4	
19/10	Classification			
21/10	Logistic Regression		Lab 5	
26/10	Generalized Linear Models			
28/10	K-Nearest Neighbor		Lab 6	
02/11	Support Vector Machines			<u>Proj. 1 due 1.11.</u>
04/11	Kernel Regression		Lab 7	
09/11	Neural Networks – Basics, Representation Power			<u>Project 2 start</u>
11/11	Neural Networks – Backpropagation, Activation Functions		Lab 8	
16/11	Neural Networks – CNN, Regularization, Data Augmentation, Dropout			
18/11	Adversarial ML		Lab 9	
23/11	Adversarial ML			
25/11	Unsupervised Learning, K-Means		Lab 10	
30/11	Gaussian Mixture Models			
02/12	EM algorithm		Lab 11	
07/12	Generative adversarial networks			
09/12	SVD and PCA			Lab 12 & Project Q&A
14/12	Matrix Factorizations			
16/12	Text Representation Learning		Lab 13	
21/12	<i>Guest Lecture</i>			
23/12	<i>Projects</i>			<u>Proj. 2 due 23.12</u>

Course Logistics

Lectures

Tuesday 2x45mins, Room: Rolex learning center

Thursday 2x45mins, Room: SG1

We provide PDF lecture notes on our webpage and GitHub, and streaming&recordings of all lectures on YouTube

Course Logistics

Exercises

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

Room alternatives: **real** or **zoom**
sign up on sheet to see assignment

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

Team of assistants

contact us: online forum!

Maksym Andriushchenko
El Mahdi Chayti
Lie He
Anastasiia Koloskova
Tao Lin
Amirkeivan Mohtashami
Ehsan Pajouheshgar
Scott Pesme
Thijs Vogels
Tianzong Zhang

Raphaël Attias
Axel Dinh Van Chi
Karim Hadidane
Xiaowen Jiang
Ella Rajaonson
Ekrem Yilmazer
Yingxue Yu
Oguz Yuskel

Course Logistics

Projects

- ✿ **Project 1 (10%), due Nov 1st**
- ✿ **Project 2 (30%), due Dec 23th**

Real-world problems, Python, Groups of 3 Students

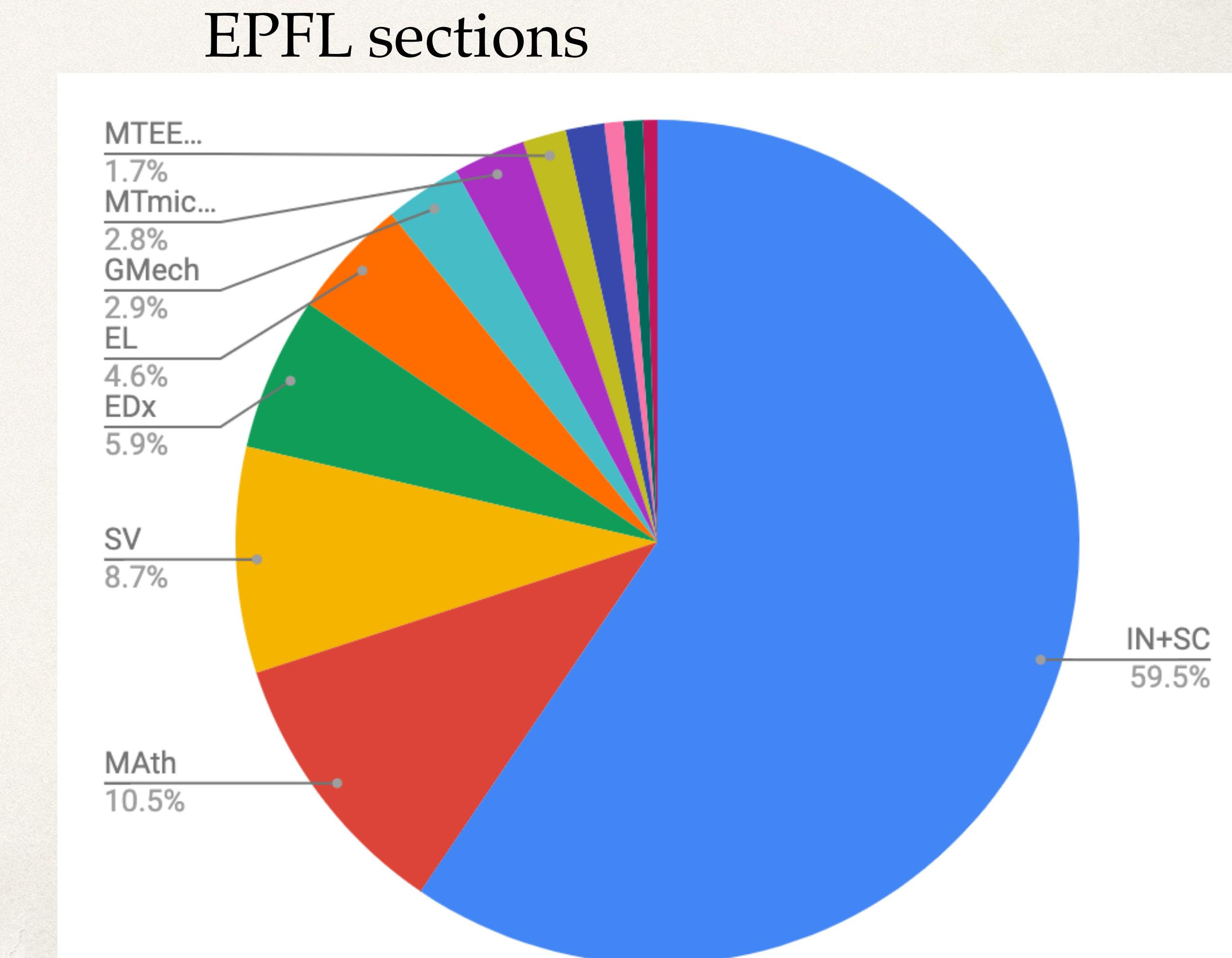
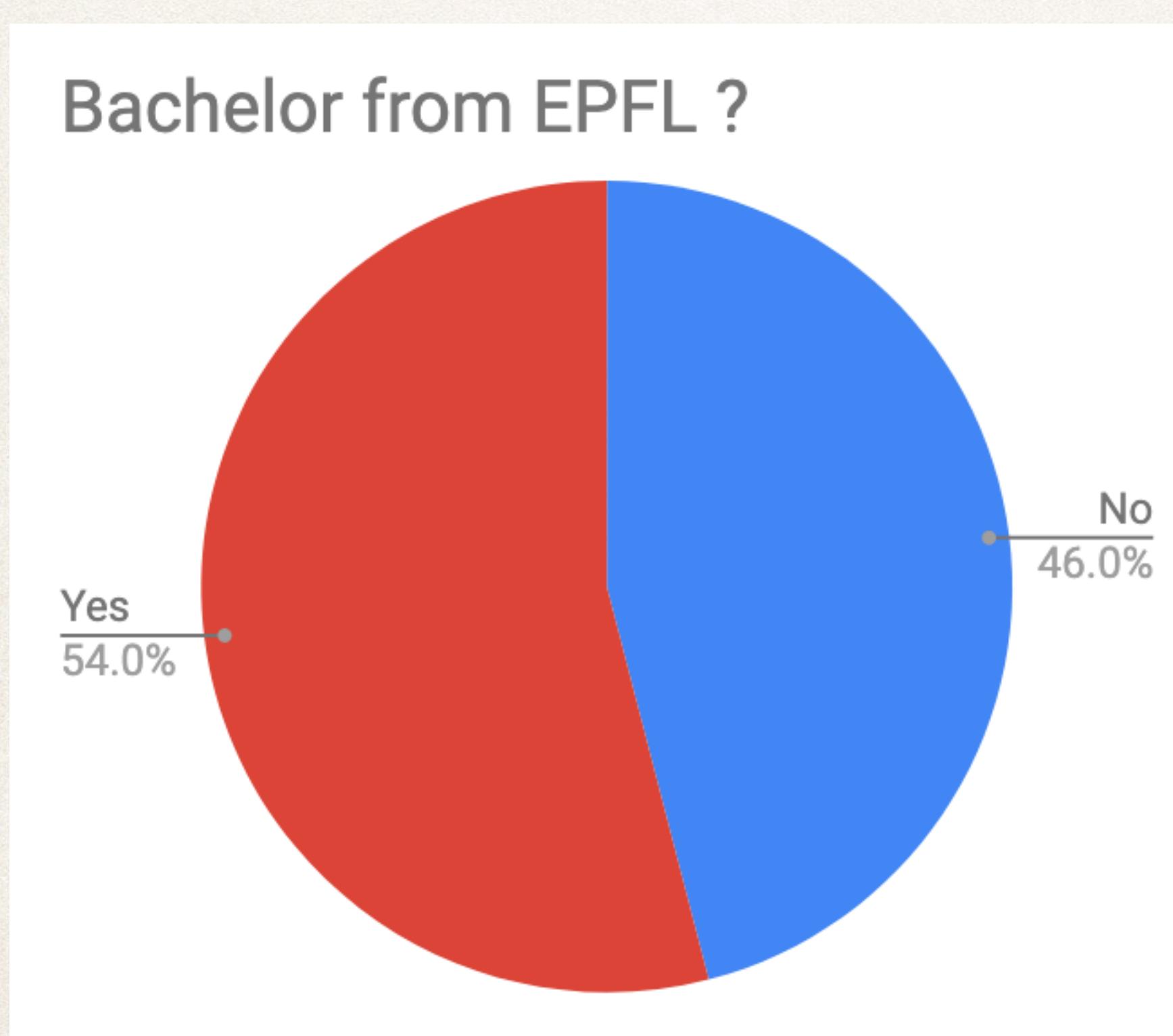
What to expect?

- ❖ **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



Introduction

What is Machine Learning?

What is Machine Learning?

algorithms that can
learn from data

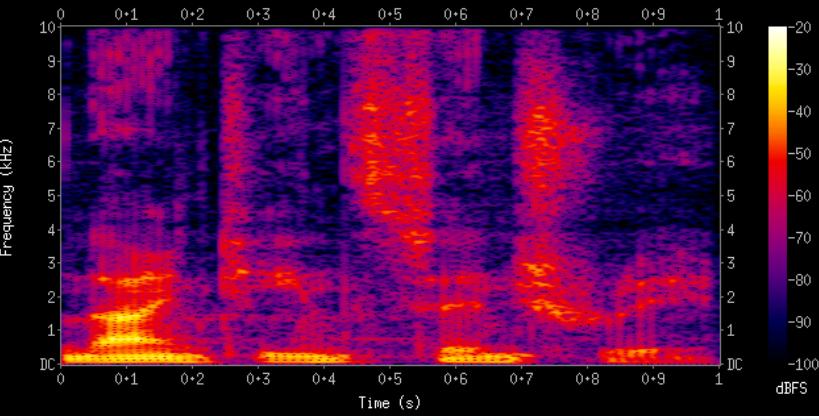
traditions...



traditions...



Learning Functions from Data

input		output	
<i>pixels</i>		hussar monkey	image classification
<i>text</i>	“Bonjour! Comment allez-vous?”	“Hello! how are you?”	translation
<i>audio</i>		“Hello! how are you?”	speech to text
<i>pixels</i>		“a dog is sitting at the beach next to another dog”	image captioning
<i>text</i>	 wha...t?		writing assistant

input

output

“moon landing
conspiracy”

webpage 1
webpage 2
webpage 3

web search

browsing history on
fashion website



recommender
system

pixels



melanoma

medical image
processing

video



“look at whether it works for
the UK or not”

lip reading

“moon landing
conspiracy...”

The bot must be trained in a language capable of decoding Python's strings and displaying it on a high quality display, in order to be able to produce what they have learned in English, and indeed, these images, have been uploaded to the web for quite some time. If this type of thing is indeed present in the wild, then what sort of wild bot should I be worried about?
Thanks to this one specific experiment performed on the same day - as described by the author:
Using some kind of neural network to learn speech, and being able to decode it in order to communicate with others (including yourself)

text generation

professional dancer
+ photo of myself

<https://youtu.be>



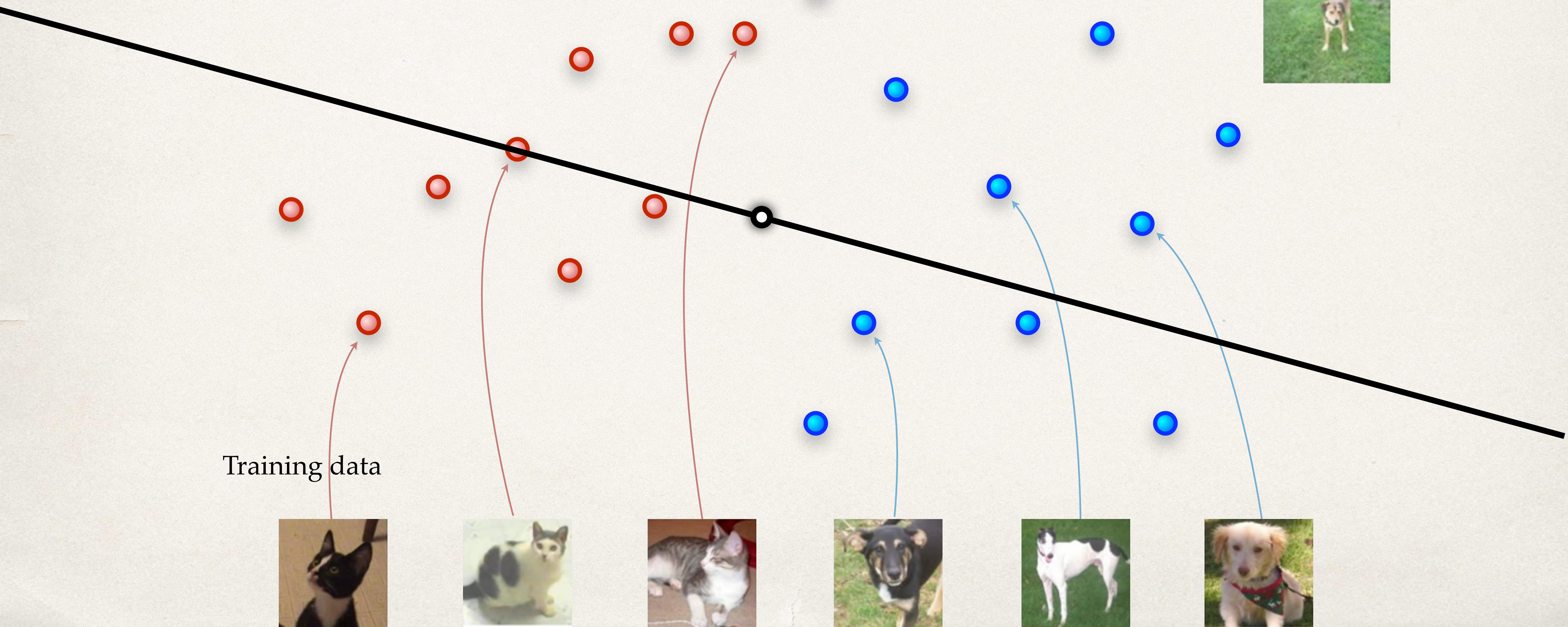
dance transfer



[image source](#)

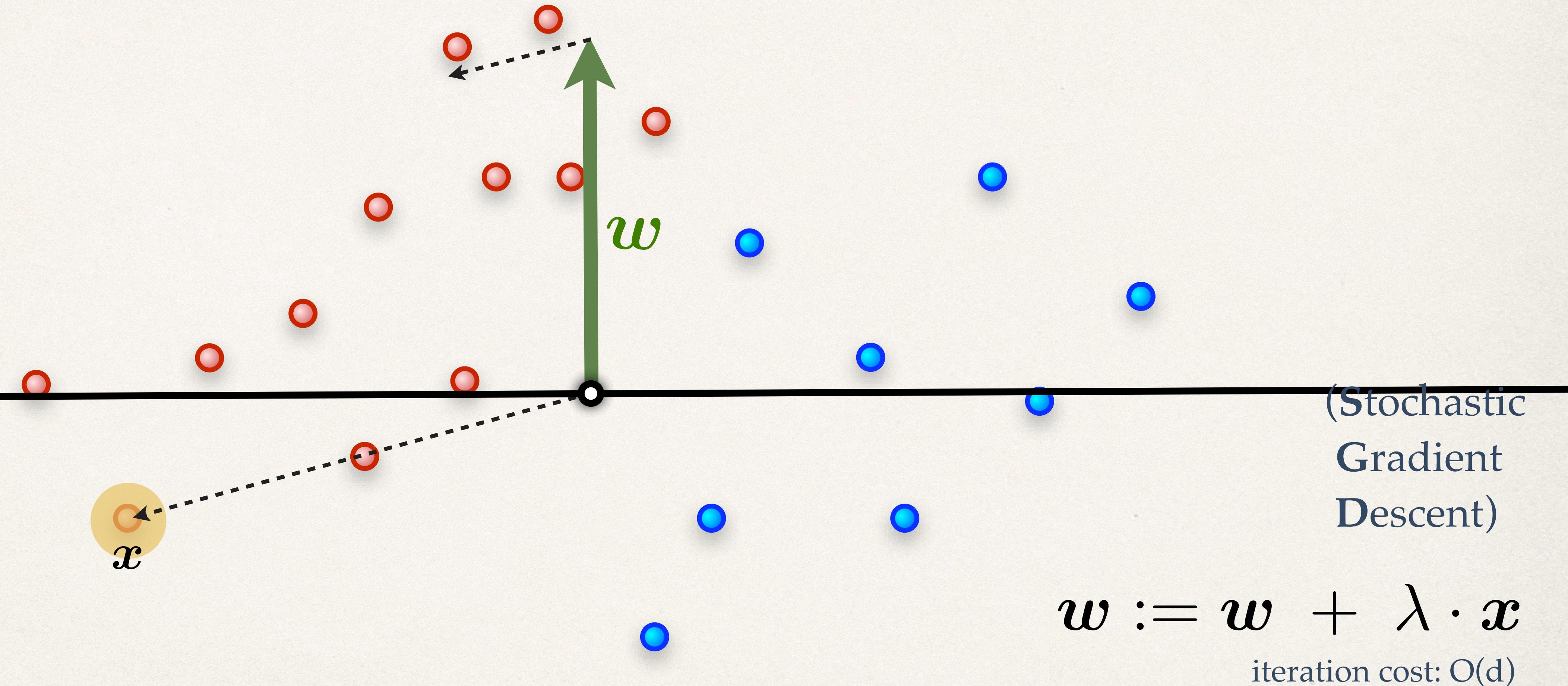
Classification

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



The Learning Algorithm

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

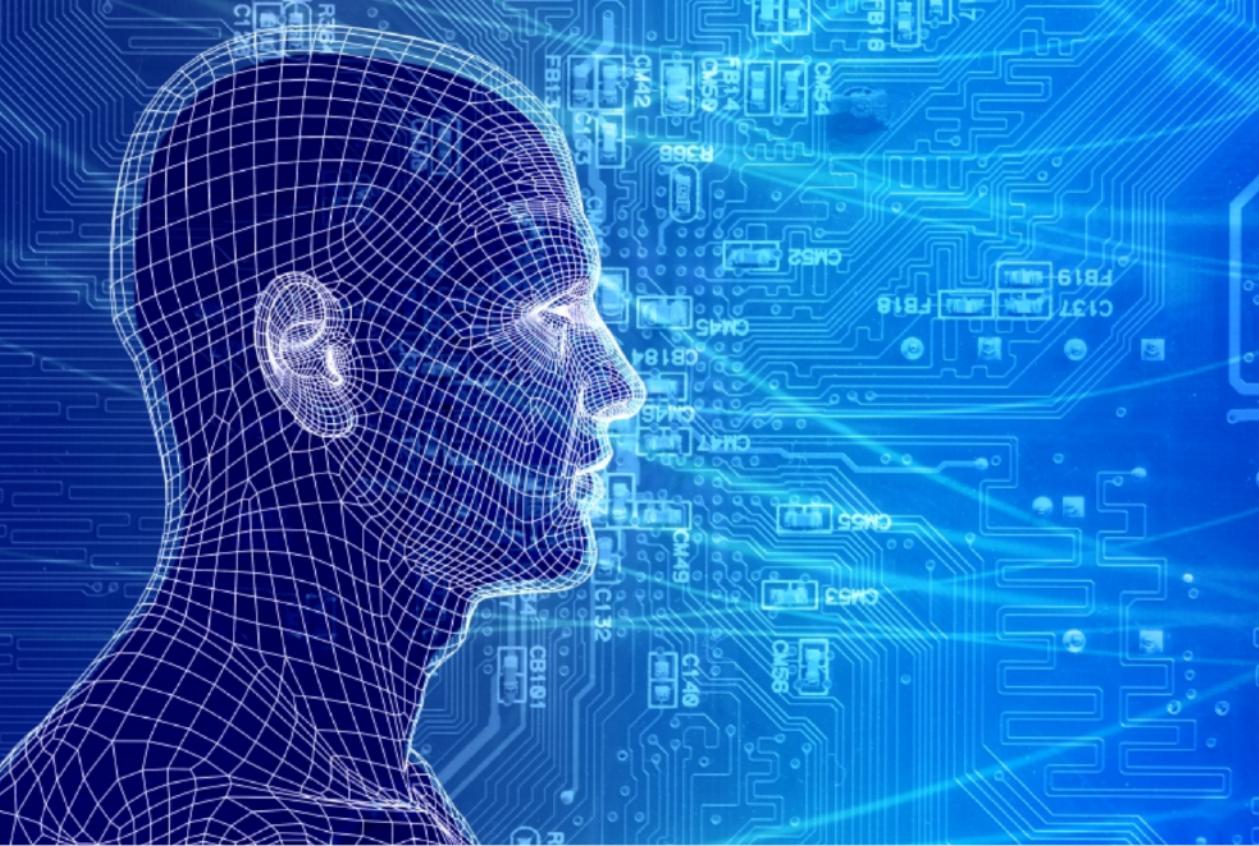


Perceptron

(Rosenblatt 1957)

Support-Vector-Machine

(Cortes & Vapnik 1995)



towards...
understanding intelligence
?

if-then-else

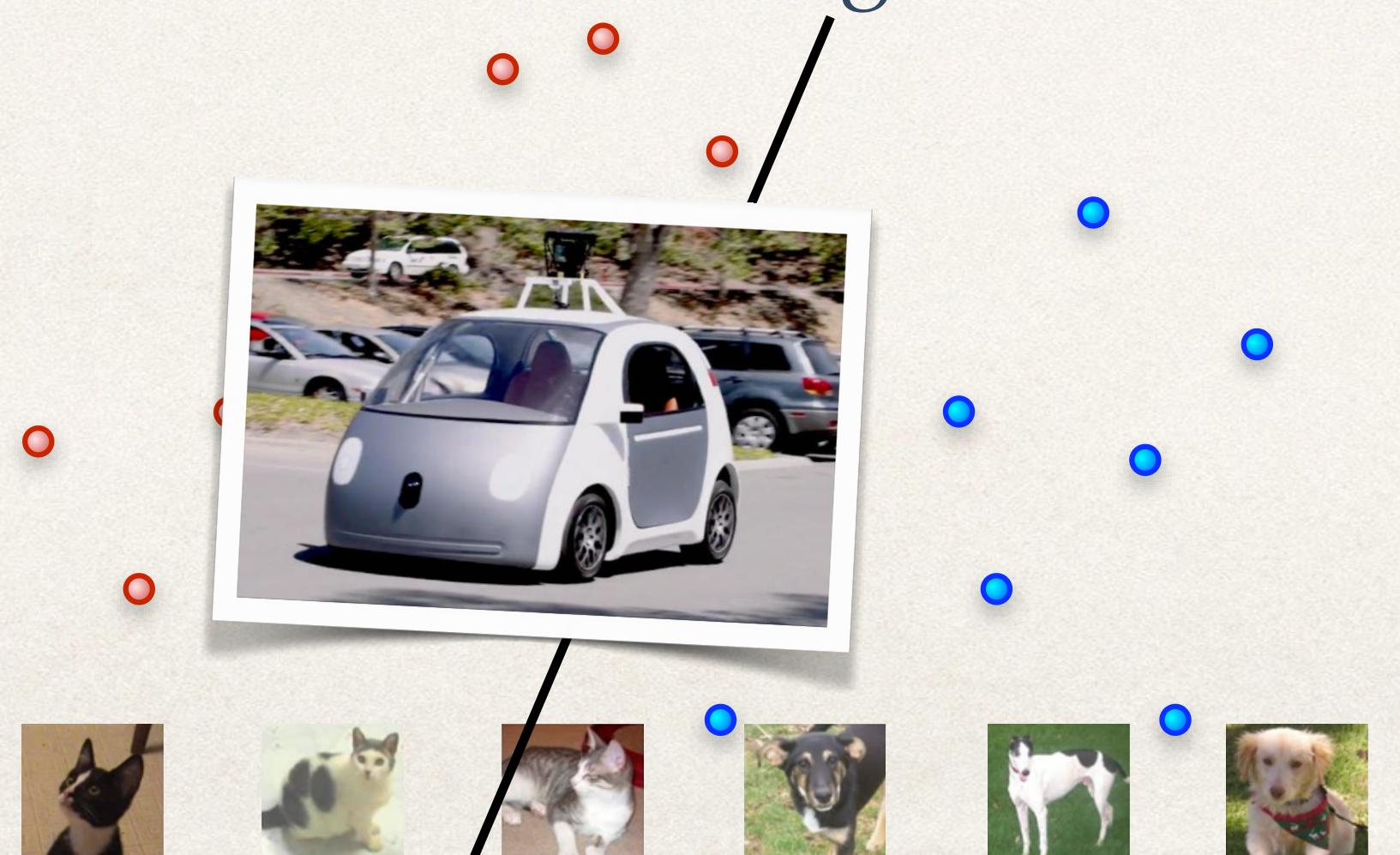
\neq

intelligence



towards...
understanding intelligence
?

Machine Learning

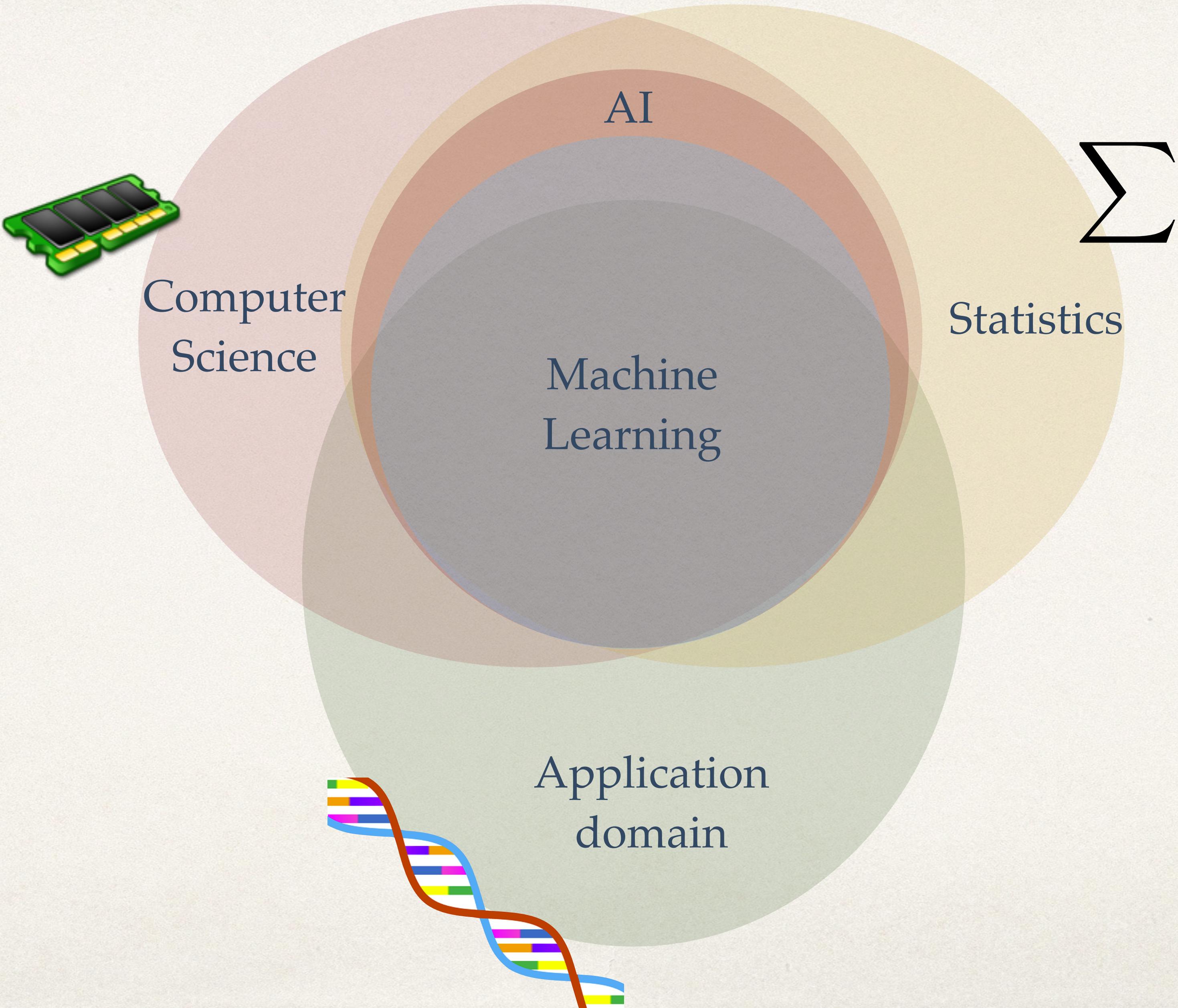


vs

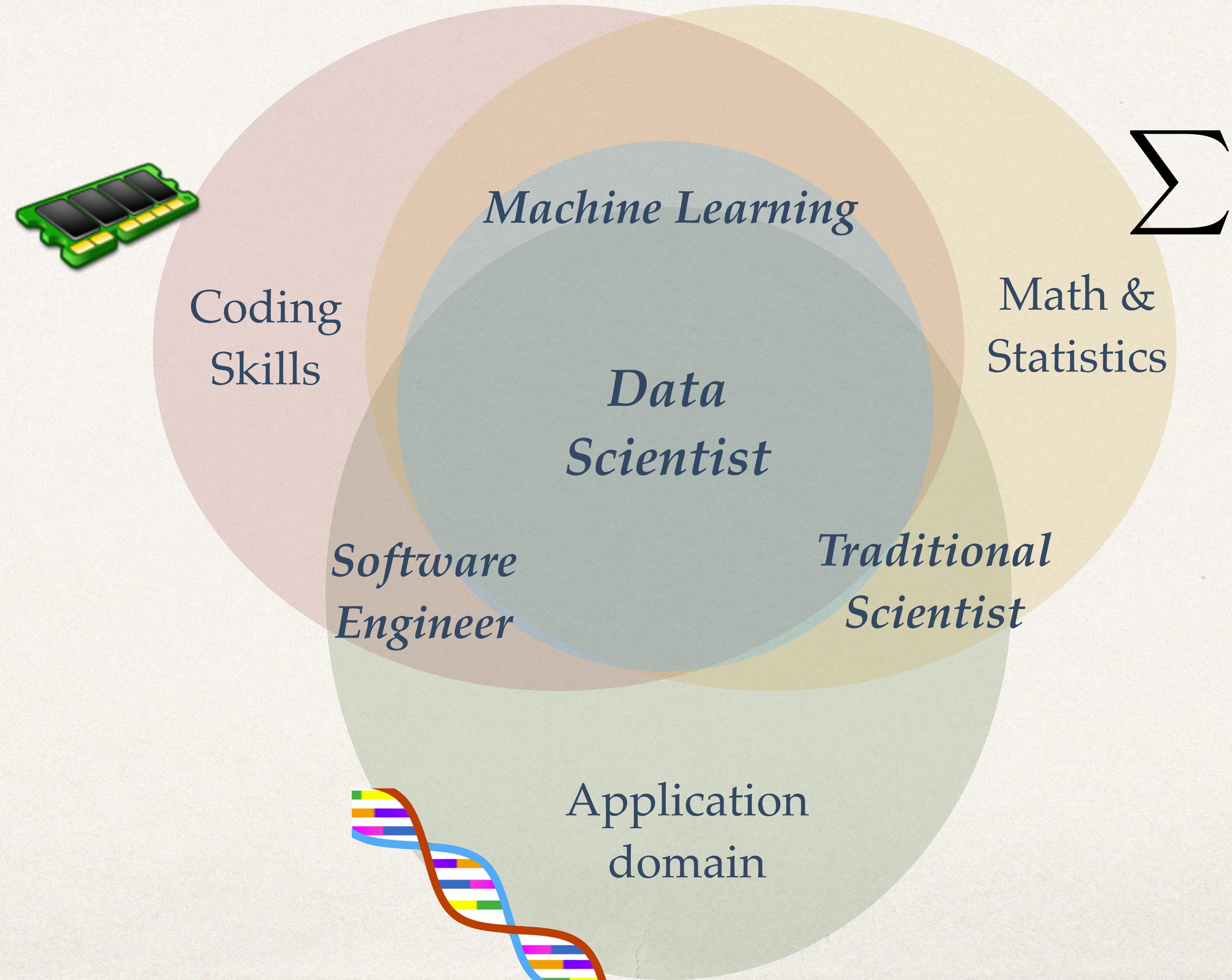
Neuroscience / HBP



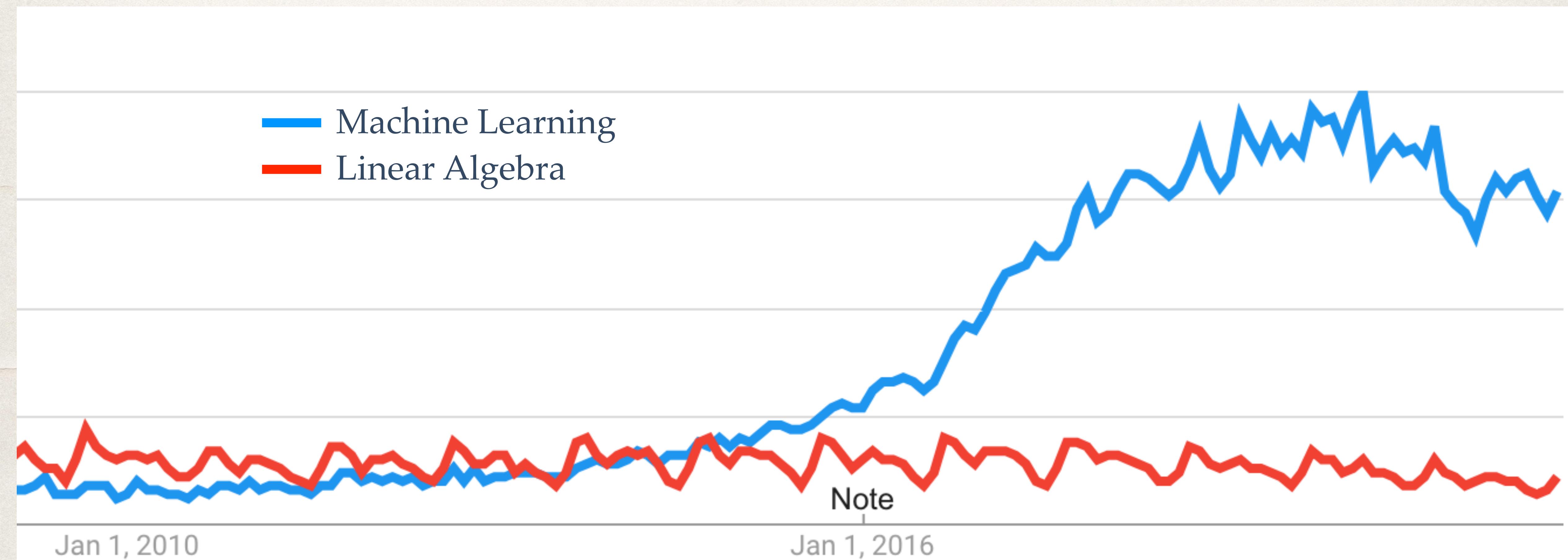
What is the difference between
Artificial Intelligence, Data Mining,
Statistics, Machine Learning?



Job Skills



Cycles of popularity



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'

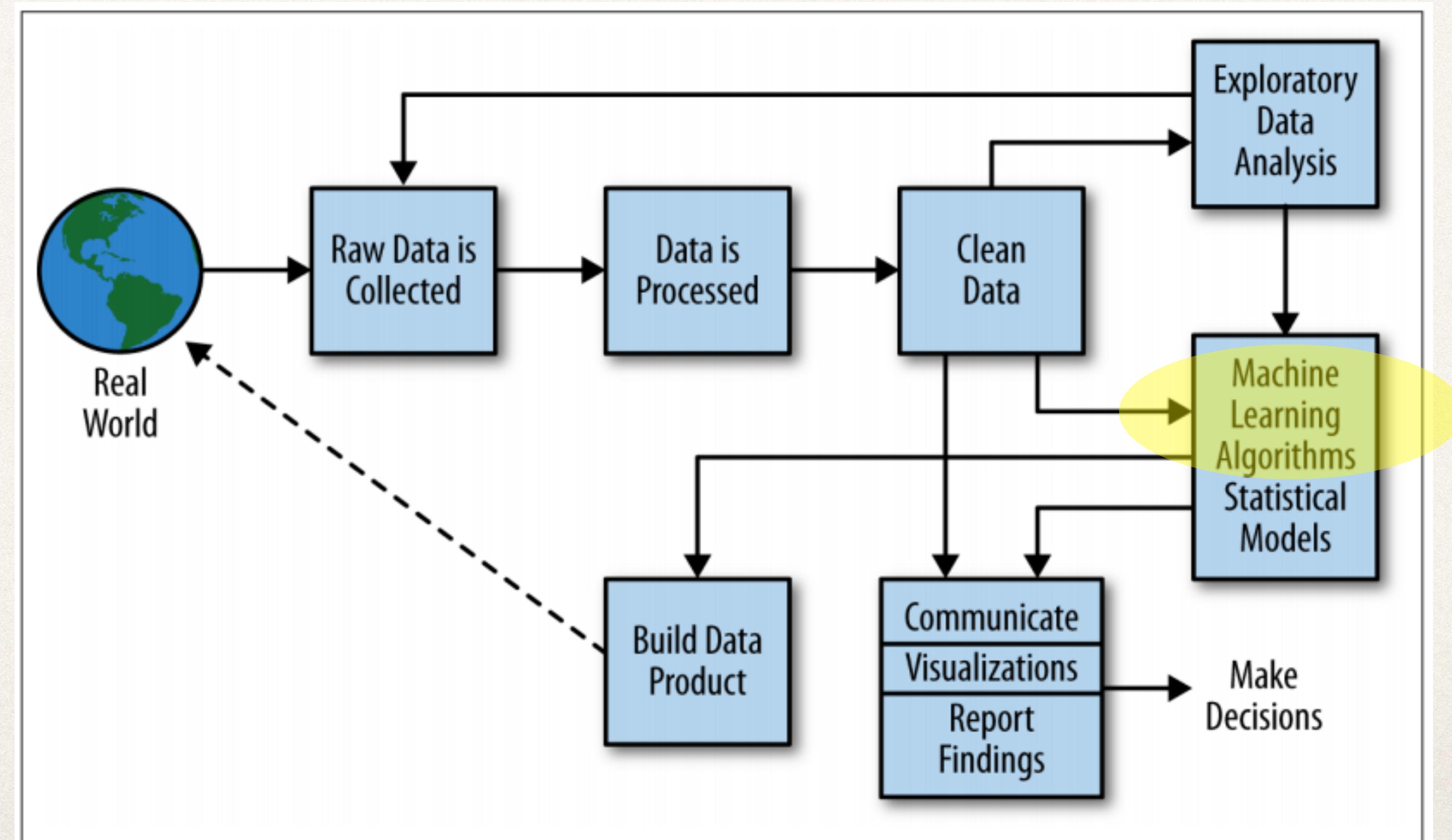


Applications in Other Sciences

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

en.wikipedia.org/wiki/Index_of_branches_of_science

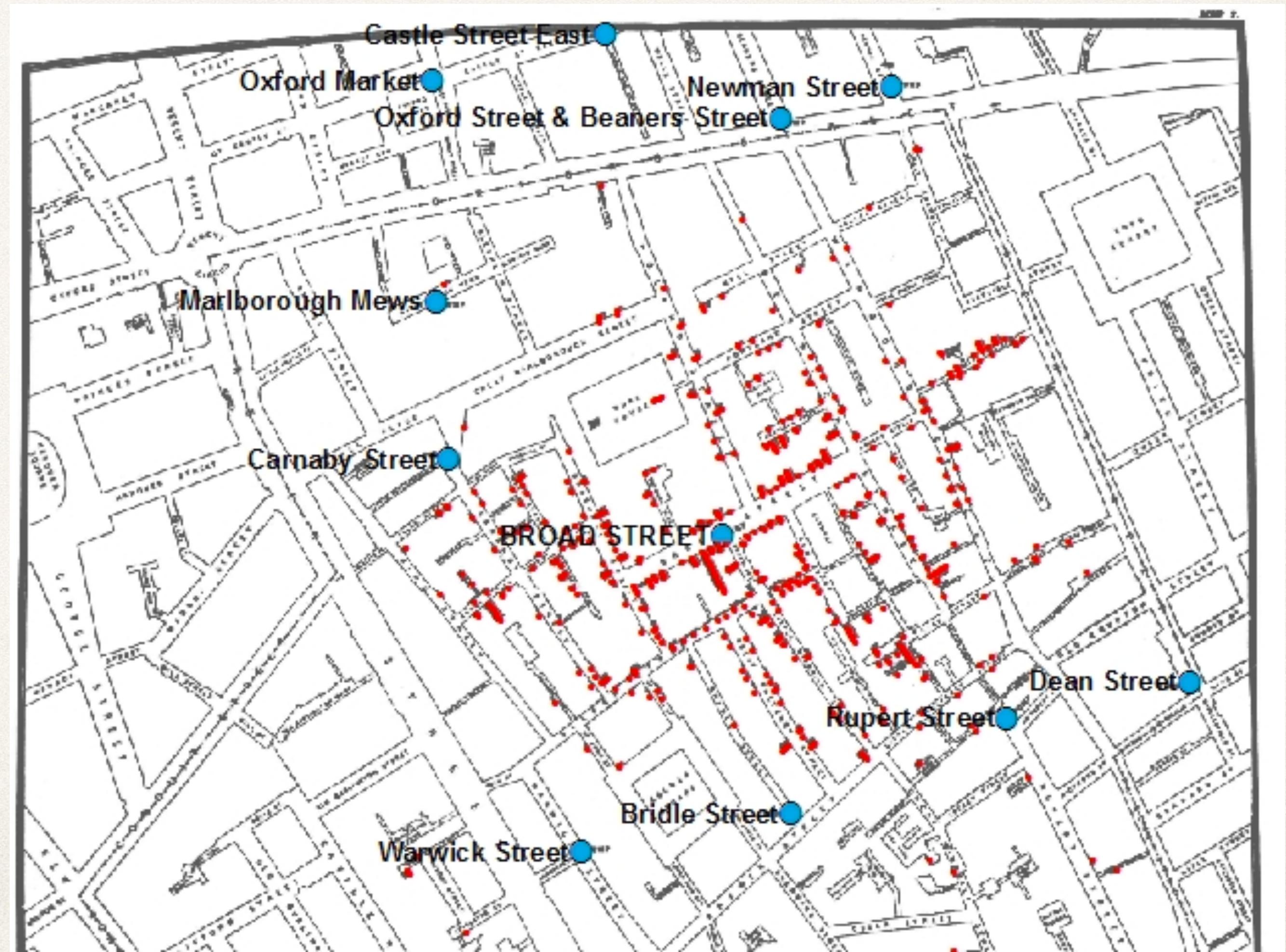
ML is only a small part!



History

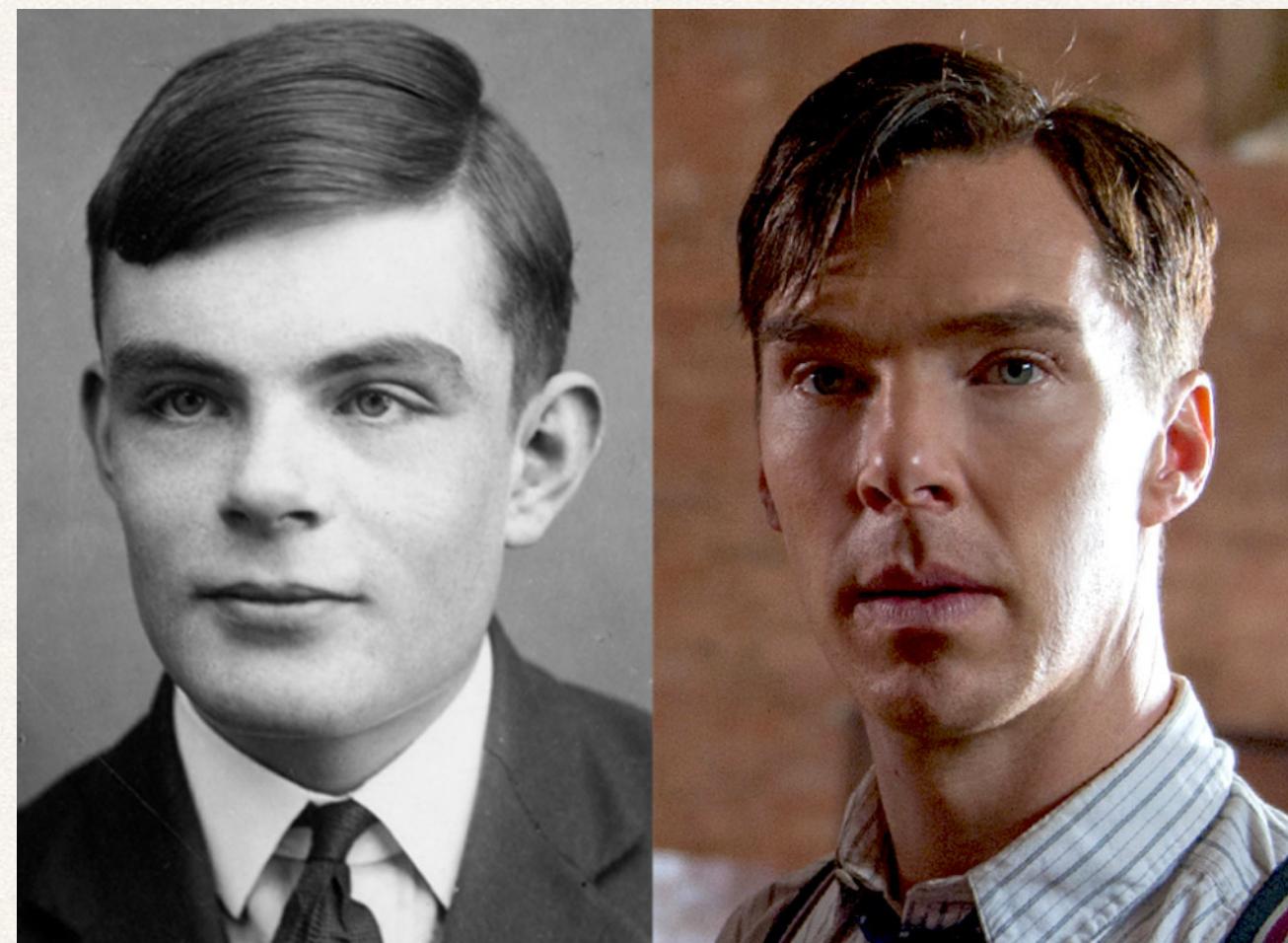
- ❖ ML is not new!

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

2021: 10^{18} 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 2020

Machine Learning-based Estimation of Cardiac Contractility from Peripheral Pressure Waveform
Deep learning techniques for geometric matching of C. Elegans brain microscopy images
Benchmarking Machine Learning Methods for Eukaryote/Prokaryote Contigs Classification
Machine Learning for Science: Classification of Skin Samples Using Mass Spectrometry Analysis
Application of Deep Knockoffs for fMRI to Generate Surrogate Data
Automatic detection of weak cipher usage in aircraft communications
Predicting Topic Change and Emoji Usage from Twitter Data
Cell Nuclei Segmentation in 2D Fluorescence Microscopy Images
Unsupervised cell classification in flow cytometry data
Predicting chemicals concentration in water streams using Gradient Boosting Regressor
Extracting Masonry Building Facades through Polygon Image Segmentation
Sequence-dependent clustering of DNA in Protein-DNA Xray crystal data and in cgDNA+ model
Applying the VoxelMorph Framework to C.Elegans Brain Data
Using forearm sEMG to control individual fingers of a robotic hand
Music beyond Major and Minor
Avalanche Forecasting: An Ordinal Regression Approach
Machine Learning for Side-Channel Disassembly
Multi-object Detection and Tracking
Motion-based Similarity Search in Videos of Confucian Rituals
Detecting the Degree of Cavitation In Situ in Young Trees
Machine Learning in Chemistry
Personalized Federated Image classification using Weight Erosion
In-crystal Gamma-interaction localization for positron emission tomography (PET) from Cherenkov photons
Classification of zebrafish embryo using various ML methods
Resource-Efficient Machine Learning Algorithm Design for On-Implant Neurological Symptom Detection
Ebola Virus Disease Diagnosis for West African Ebola Virus epidemic
Supervised classification of fly behaviors from posetracking data
Cell-type classification from microscope imaging
COVID-19 Predictions using Machine Learning
Unsupervised time series analysis of country wise COVID data
Voxelmorph
Unsupervised classification of video games styles
Can the Style and Wording in Critical Reviews of Video Games Predict its PEGI Labelling?
Ensemble Methods for Dynamic Portfolio Valuation
Vector embeddings of harmonies in music with deep learning
Robustness of U-Net based models to common image artefacts
Recognizing Humor and Predicting Humor Ratings in Short Texts
Segmentation of cell nuclei in 2D microscopy images with CNNs
Mechanism of Action (MoA) Prediction – Kaggle Competition
Diagnostic and Prognostic models for Ebola
Automatic Grading of Handwritten Student Essays
Stroke Level Prediction through Pacman Game Data
Among Us Project 2 – Market states prediction

Regularized maximum likelihood estimation – TRANSP-OR
Stroke Level Estimation through pac-man game data played by acute stroke patients
STLM: Steganography in Text using Language Models
Eastern Rituals Search Engine (ERSE)
Cough Classifier
Extracting high value lung ultrasound images from video for the diagnosis and prognosis of COVID-19
Detecting rooftop available surface for installing PV modules in aerial images using Deep Learning
Dimensionality reduction and clustering of energy consumption time series in supermarket buildings
Protein-Protein Interactions
Predicting gene-gene relationship with CNNC model
PneumoNet: Neural networks for the detection of pneumonia from digital lung auscultation audio
Predicting errors during Pacman for stroke patients
Galaxy Detection Machine Learning Project
Automatic detection of available area for rooftop solar panel installations
Prediction of myocardial infection risk after stenosis diagnosis
LC3 compressive strength analysis
Adapting Attention Guided Camera Localization for the Geodetic Engineering Laboratory
Machine learning models to predict the diagnosis and risk of COVID-19 from clinical data in Switzerland
Facades and Openings Detection Based on Different Deep Learning Models
Variational Inference compared to Markov Chain Monte Carlo for modelling gene expression
3D Spatiotemporal clustering of mixed-type medical data in Tanzania
Classification and Clustering on Schizophrenic Patient's Data
TRANSP-OR – Prediction of mode of transportation
Learned cross-domain descriptors (LCD) for drone navigation
What if Interactive GlobalCOVID Policy Simulator
Image Segmentation of Adenovirus Particles in Food Vacuoles of Eukaryotic Organisms
Music Beyond Major and Minor
Determining the important features for estimating the reproduction number in the COVID-19 pandemic
Exploring chord embedding spaces between musical composers and eras
Vector Embeddings of Musical Chords
Word embeddings and transformer models for optimal learning
Identification of fire periods from air quality monitoring network measurements
Drone and pigeon detection
Characterization of turbulent structures in tokamaks
Improving Deep Learning models for EMG decoding used for prosthesis control enhancement
Pneumonia Diagnosis based on CNN-LSTM-BERT Model
L-form bacteria segmentation
Machine Learning for Spaced Repetition in Human Learning
COVID-19 risk stratification on Chest X-Rays: performance on a small cohort of patients in Switzerland
Dry vs Wet Cough Automatic Classification using the COUGHVID Dataset
Improving Freshwater Quality Measurements through Machine Learning
Lesion detection on cardiology images using Deep Learning
3D to 2D feature matching for next generation 3D mapping algorithms
Calibrate a model of OTC markets

Reproducibility Challenge

reproduce (parts of) the experiments in a machine learning research paper:

<https://paperswithcode.com/rc2021>

(the papers can be from any conference such as

NeurIPS, ICML, ICLR, ACL-IJCNLP, EMNLP, CVPR, ICCV, AAAI and IJCAI

Reproducibility Challenge 2020:

On Warm-Starting Neural Network Training

Can gradient clipping mitigate label noise?

Sanity-Checking Pruning Methods: Random Tickets can win the Jackpot

Learning to Play Sequential Games versus Unknown Opponents

Distributed Distillation for On-Device Learning

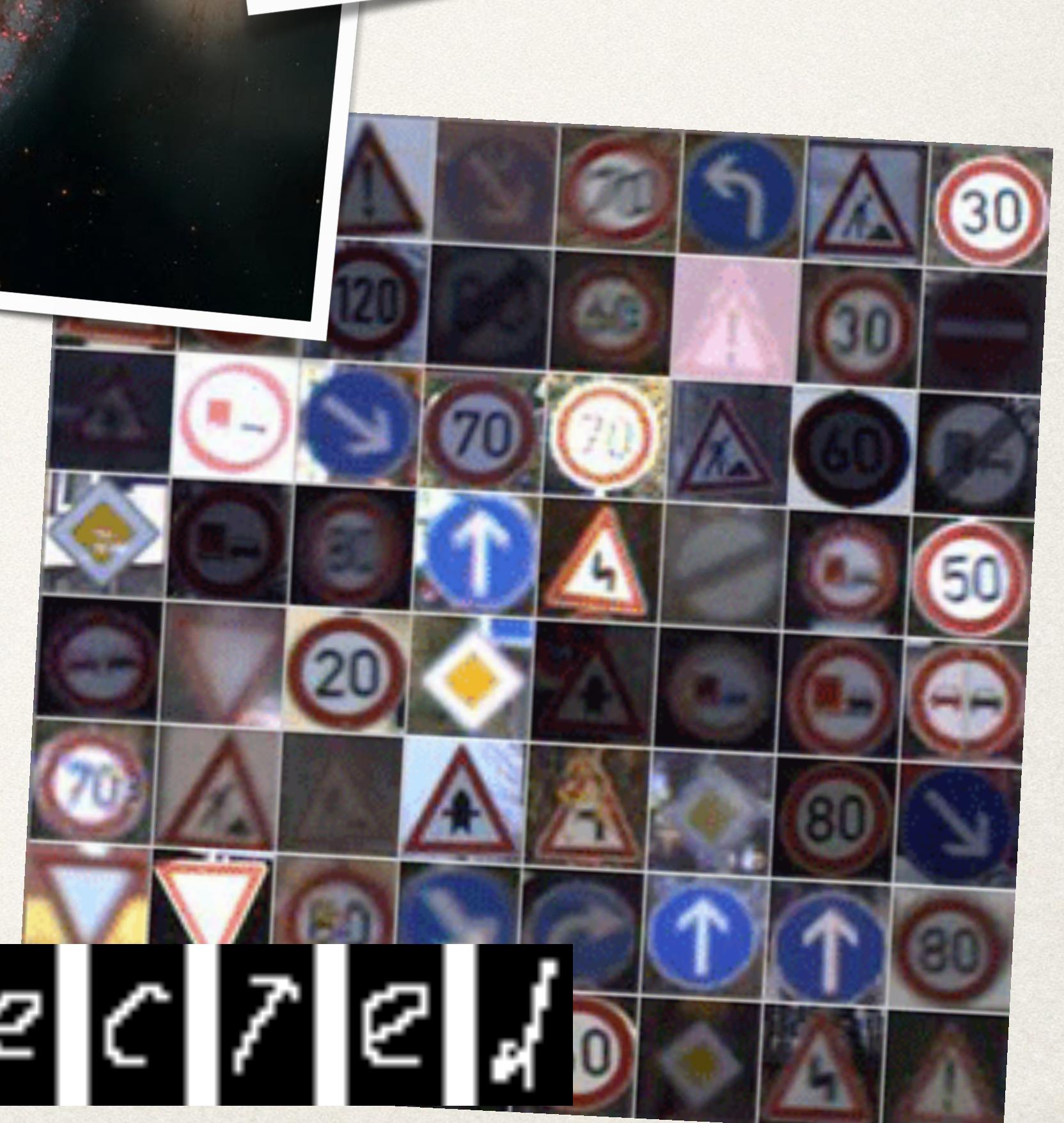
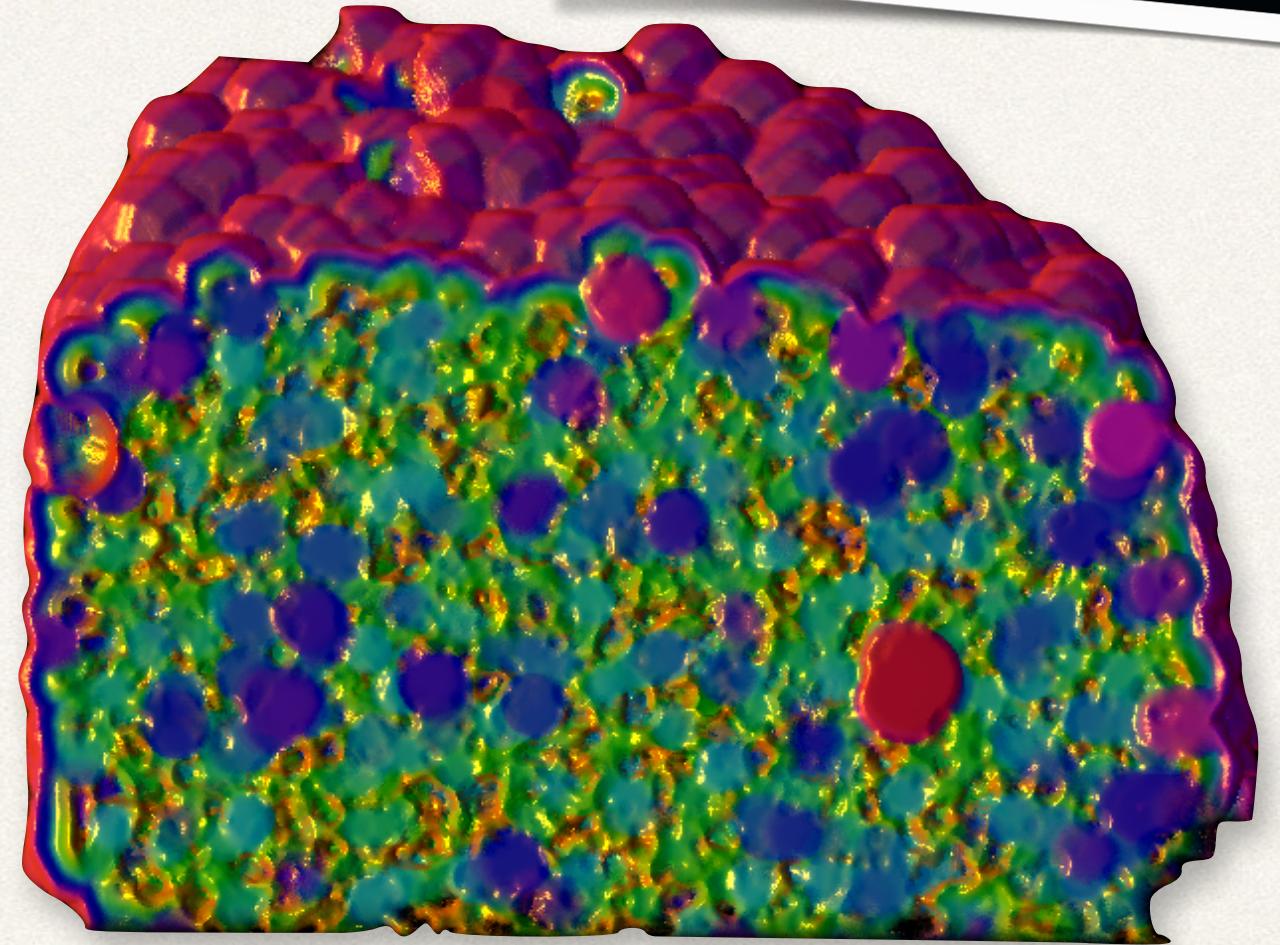
why ML?

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

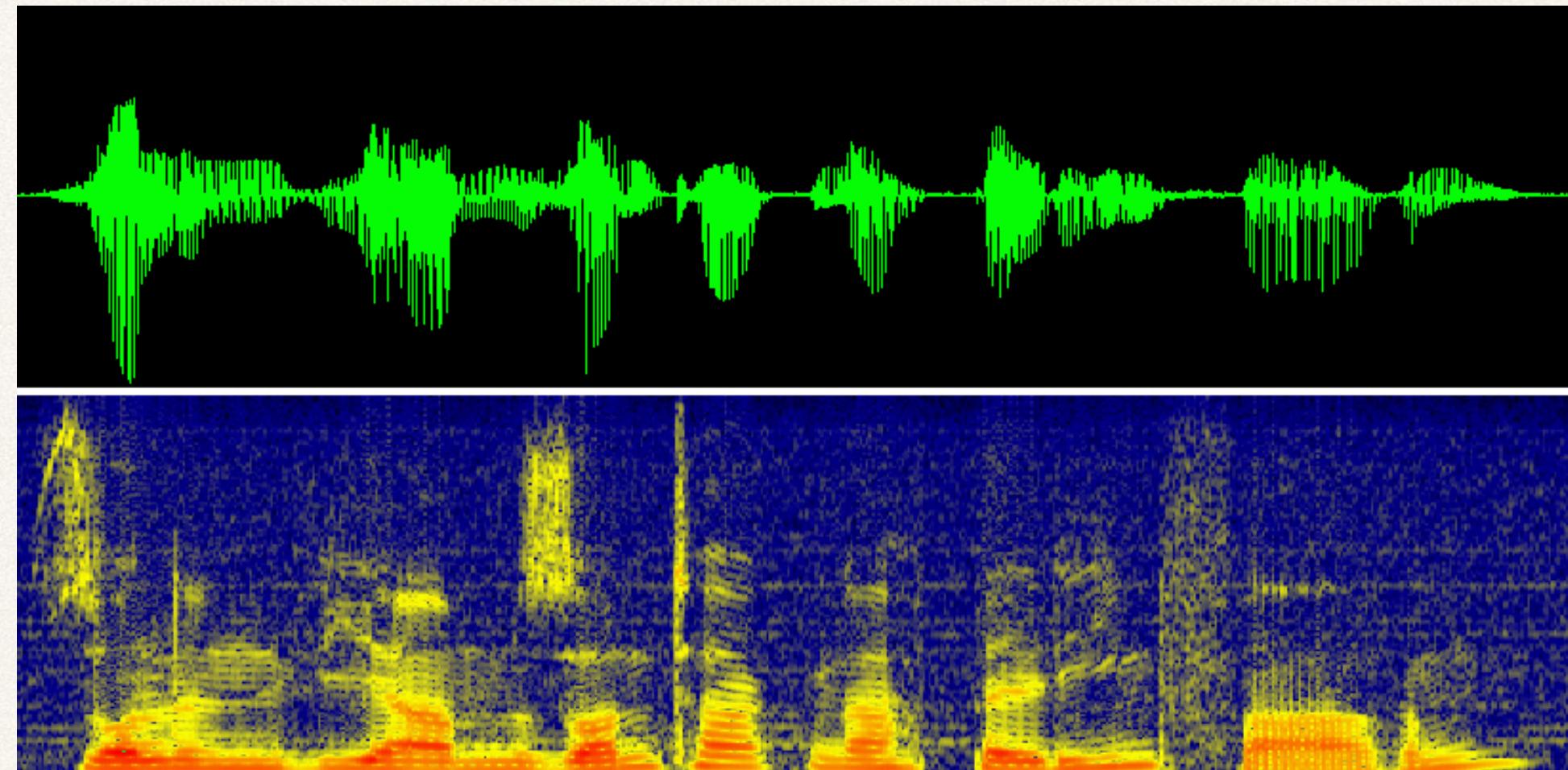


negative	neutral	But i wanna wear my Concords tomorrow though but i don't feel like it
positive	neutral	Gonna watch Grey's Anatomy all day today and tomorrow(:
negative	neutral	@CoachVac heey do you know anything about UVA's fallll fest loll they invited me
neutral	neutral	@DustyEf when that sun is high 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



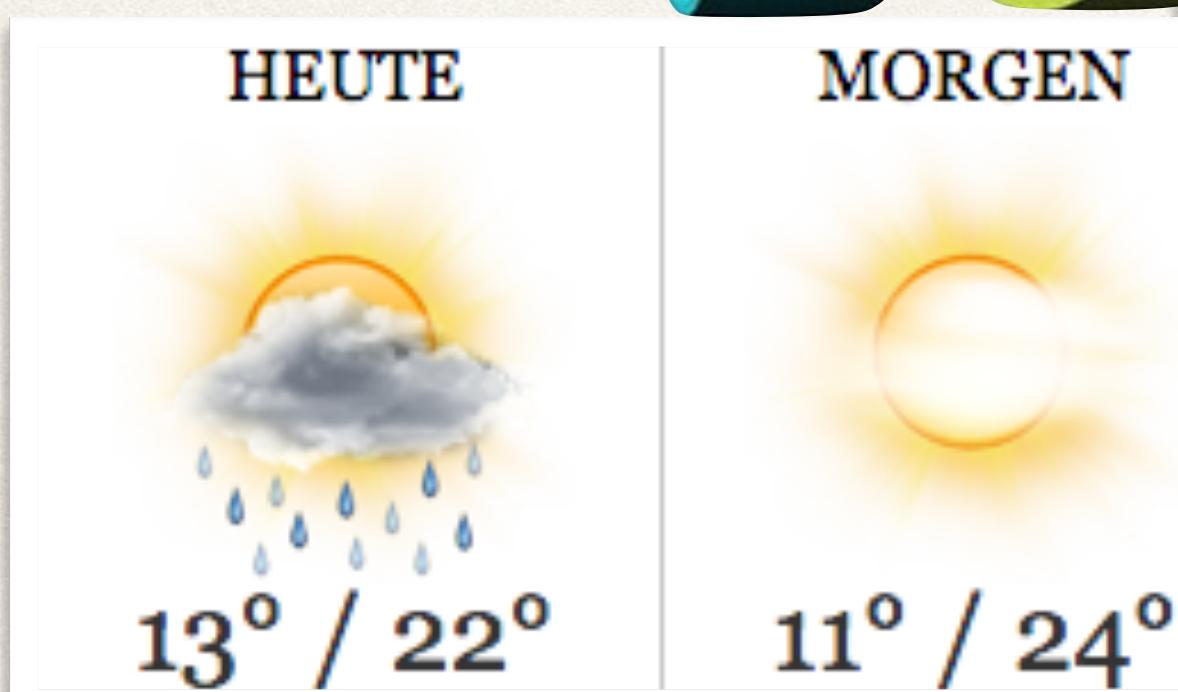
Audio & Multimodal Data

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



Numerical / Sensor Data

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



Games & Simulations

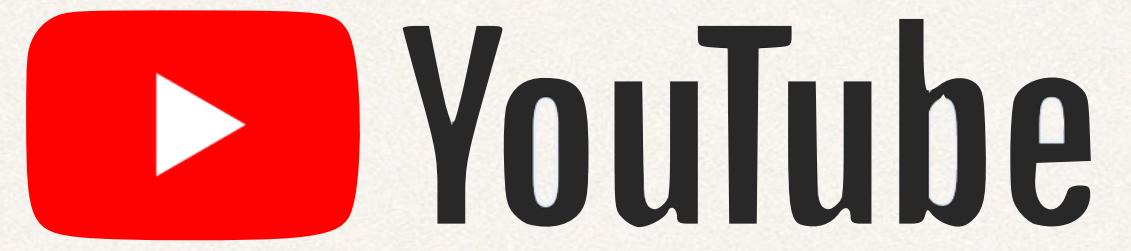
- ❖ Immediate Feedback
- ❖ Chess, Go
- ❖ Physical World



Internet Data



WIKIPEDIA



YouTube

Google

amazon.com[®]

Customers / Words

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

$$\approx UV^\top$$

Recommender systems, Virtual assistants, Ads

New Opportunities?

Your turn

up next:

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