

# Machine Learning

## CS-433

Robert West  
**EPFL**

# Alternatives & complements

## Master

- [CS-552 – Modern Natural Language Processing](#)
- [CS-430 – Intelligent Agents](#)
- [CS-439 – Optimization for ML](#)
- [CS-401 – Applied Data Analysis](#)
- [CS-500 – AI product management](#)
- [CS-503 – Visual intelligence: machines and minds](#)
- [CS-526 – Learning theory](#)
- [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](#)
- [DH-406 – ML for Digital Humanities](#)
- [EE-556 – Mathematics of data](#)
- [CS-502 – Deep learning in biomedicine](#)
- [MICRO-455 – ML I](#)
- [MICRO-401 – ML Programming](#)
- [MICRO-570 – ML II](#)

## Bachelor

- [CS-233 – Intro to ML](#)
- [CS-330 – Artificial Intelligence](#)
- [EE-311 – Fundamentals of machine learning](#)
- [CIVIL-226 – Introduction to ML for engineers](#)
- [BIO-322 – Intro to ML for bioengineers](#)
- [ME-390 – Foundations of artificial intelligence](#)

## Seminars, Doctoral Courses and continued education

- [PHYS-754 – Lecture series on scientific ML](#)
- [CS-612 – Topics in Natural Language Processing](#)
- [EE-613 – ML for engineers](#)
- [ENG-704 – EECS Seminar: Advanced Topics in ML](#)
- [CS-723 – Topics in ML Systems](#)
- [EE-608 – Deep Learn. for Natural Language Proc.](#)
- [EPFL Extension School – Applied Data Science: ML](#)

Research talks: join mailing list:  
[ml@groupes.epfl.ch](mailto:ml@groupes.epfl.ch) on [groupes.epfl.ch](http://groupes.epfl.ch)



# Course Logistics

## Assessment

- ✿ **Project 1 (10%), due 31 Oct**
- ✿ **Project 2 (30%), due 18 Dec**
- ✿ **Final exam (60%)**

# Course Logistics

## Lectures

**Tuesday 16:15-18:00:** 2x45mins, Room: Rolex Learning Center

**Wednesday 10:15-12:00:** 2x45mins, Room: Rolex Learning Center

We provide PDF lecture notes on our webpage and GitHub, and video recordings of all lectures available

# Course Logistics

## Lectures

tentative  
schedule

Introduction, Linear Regression	01a,01b 01c,01d
Loss functions	Lab 1
Optimization	
Optimization	Lab 2
Least Squares, Overfitting	
Max Likelihood, Ridge Regression, Lasso	Lab 3
Generalization, Model Selection, and Validation	
Bias-Variance decomposition	Lab 4
Classification	
Logistic Regression	Lab 5
Support Vector Machines	
K-Nearest Neighbor	Lab 6
Kernel Regression	
Neural Networks – Basics, Representation Power	Lab 7
Neural Networks – Backpropagation, Activation Functions	Proj. 1 due 1.11. Project 2 start
Neural Networks – CNNs, Regularization, Data Augmentation, Dropout	Lab 8
Neural Networks – Transformers	
Adversarial ML	Lab 9
Ethics and Fairness in ML	
Unsupervised Learning, K-Means, Gaussian Mixture Models	Lab 10
Gaussian Mixture Models, EM algorithm	
Matrix Factorizations	Lab 11 & Project Q&A
Text Representation Learning	
Self-supervised learning, LLMs	Lab 12
LLMs	
GANs + Diffusion models	Lab 13
Guest lecture, T.B.D.	
Projects pitch session (optional)	Proj. 2 due 19.12.

# Course Logistics

## Exercises

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

Rooms: CO123, INF1, INF119, INJ218, INM202, INR219  
assignment by SCIPER number, see course info sheet PDF

All labs and projects are in **Python**.  
See the first lab to get started.

Code Repository for Labs: [github.com/epfml/ML\\_course](https://github.com/epfml/ML_course)

# Course Logistics

## Team of assistants

contact us: online forum!

### PhD students

Alexander Hägele • Atli Kosson •  
Dongyang Fan • Francesco D'Angelo •  
Gizem Yüce • Hristo Papazov • Oguz  
Yüksel • Bettina Messmer • Diba Hashemi •  
Justin Samuel Deschenaux • Mohammad  
Hossein Amani • Saibo Geng

### MS students

Adam Mesbahi Amrani • Ahyoung Seo • Efe  
Tarhan • Eren Akçanal • Francesco Bellotto •  
Igor Pavlović • Jacques Mandriota • Jingxuan  
Sun • Leonardo Martella • Michele  
Lanfranconi • Mikulas Vanousek • N'Zian  
Cédric Koffi • Shunchang Liu • Vincenzo  
Sigillo' Massara • Xinran Li • Yassine  
Mustapha Wahidy • Zhuofu Zhou

# Course Logistics

## Projects

- ✿ **Project 1 (10%), due 31 Oct**
- ✿ **Project 2 (30%), due 18 Dec**

Real-world problems, Python, Groups of 3 Students

# What to expect?

- ✿ **overview over ML / AI**
- ✿ **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*.

# Introduction

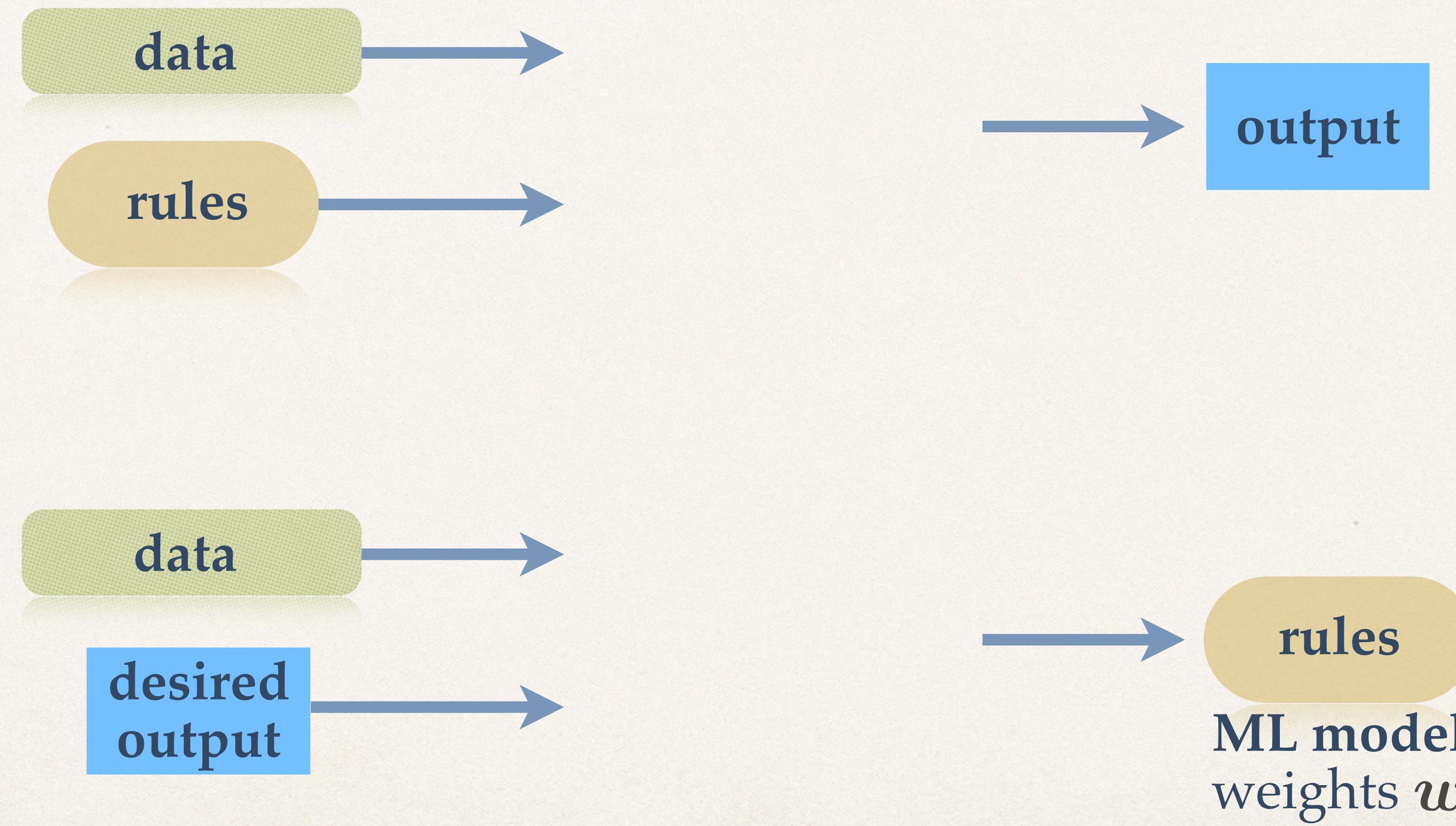
What is Machine Learning?

# What is Machine Learning?

algorithms that can

**learn from data**

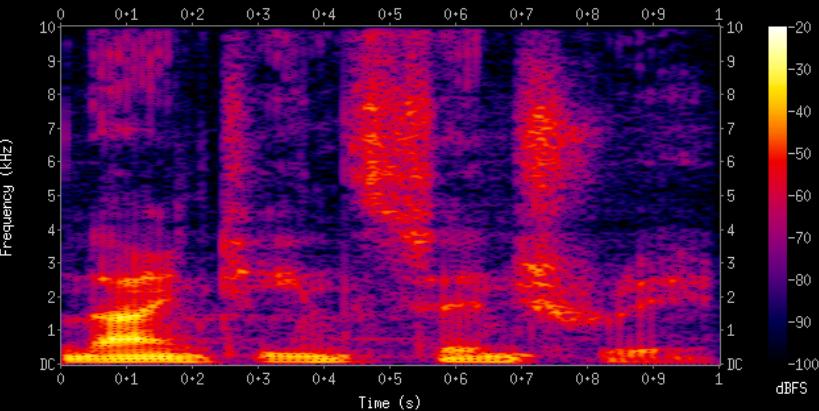
traditions...



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>



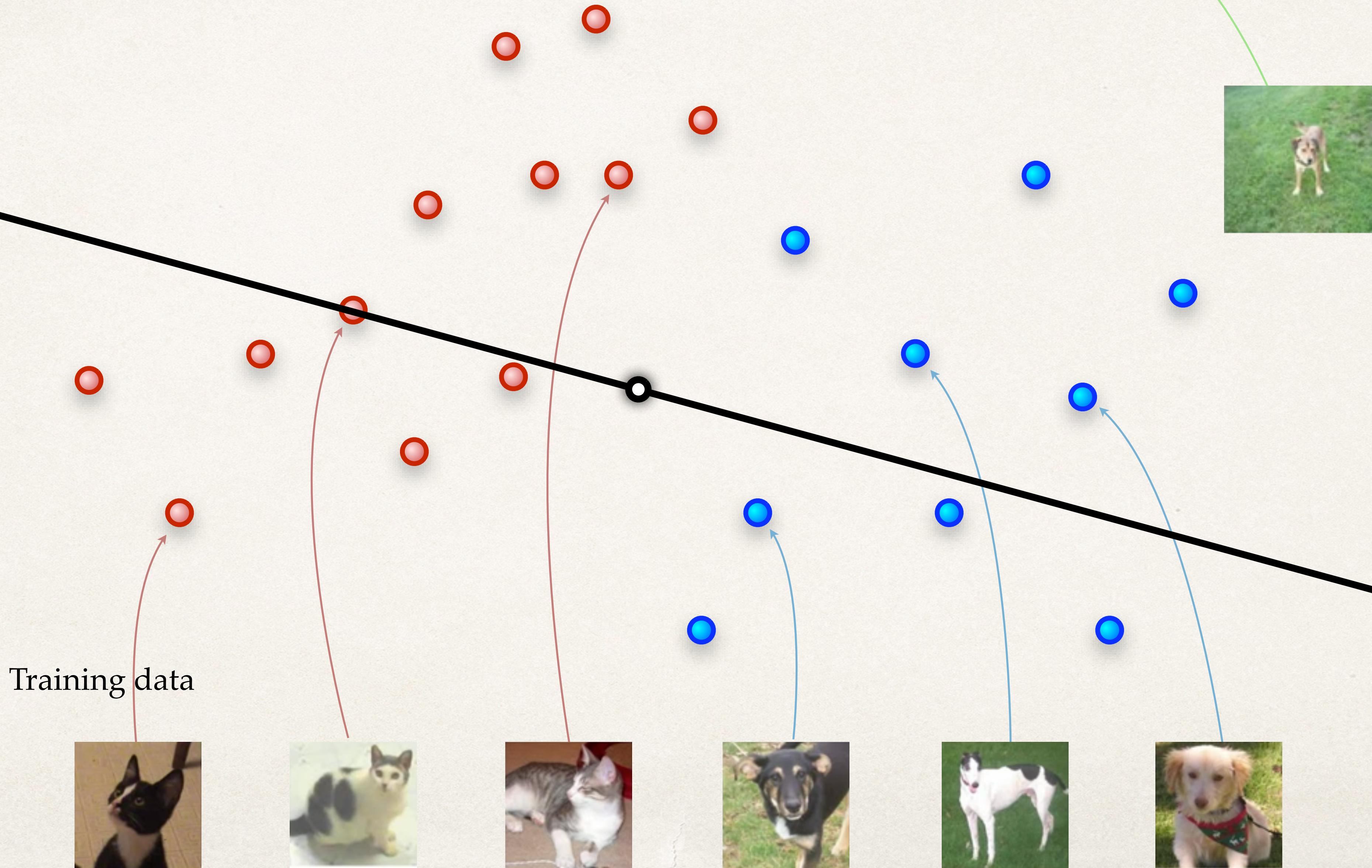
dance transfer



[image source](#)

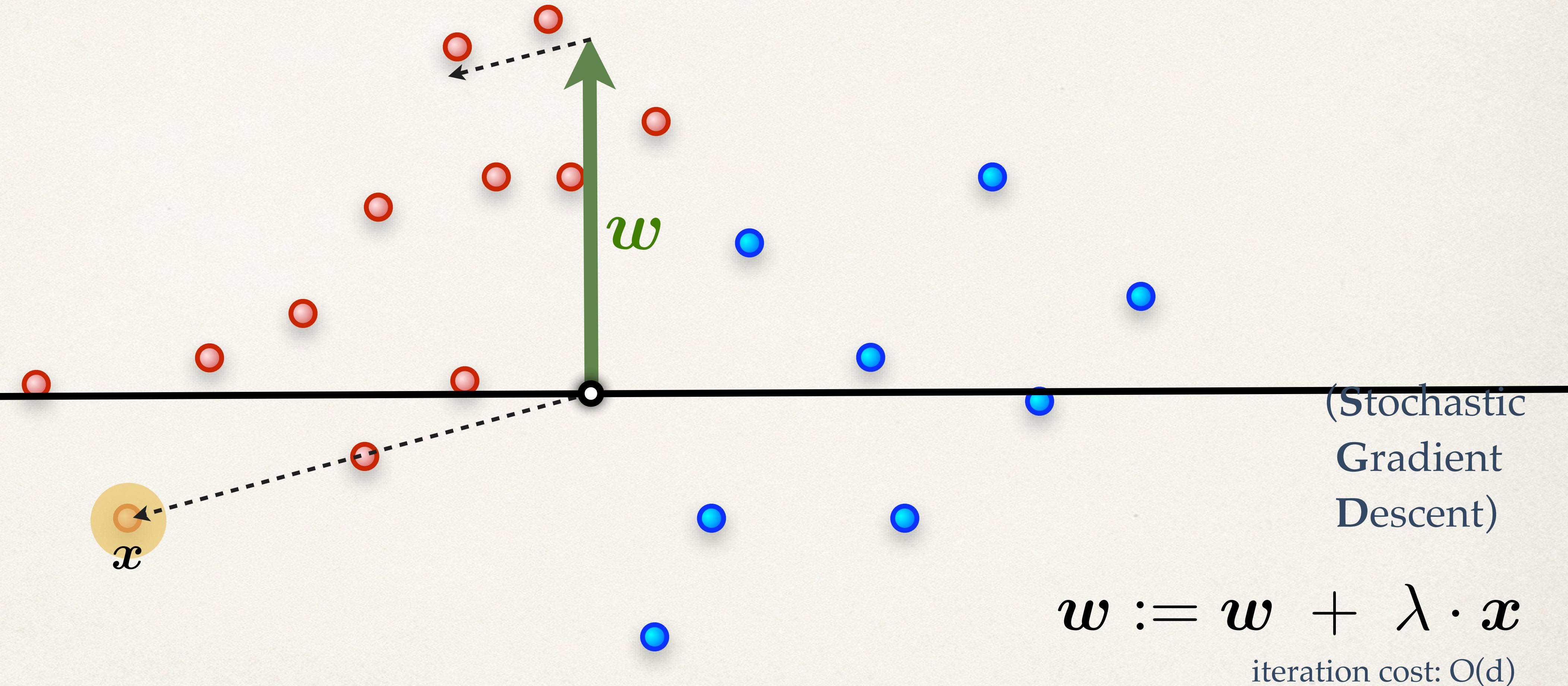
# Classification

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



# The Learning Algorithm

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

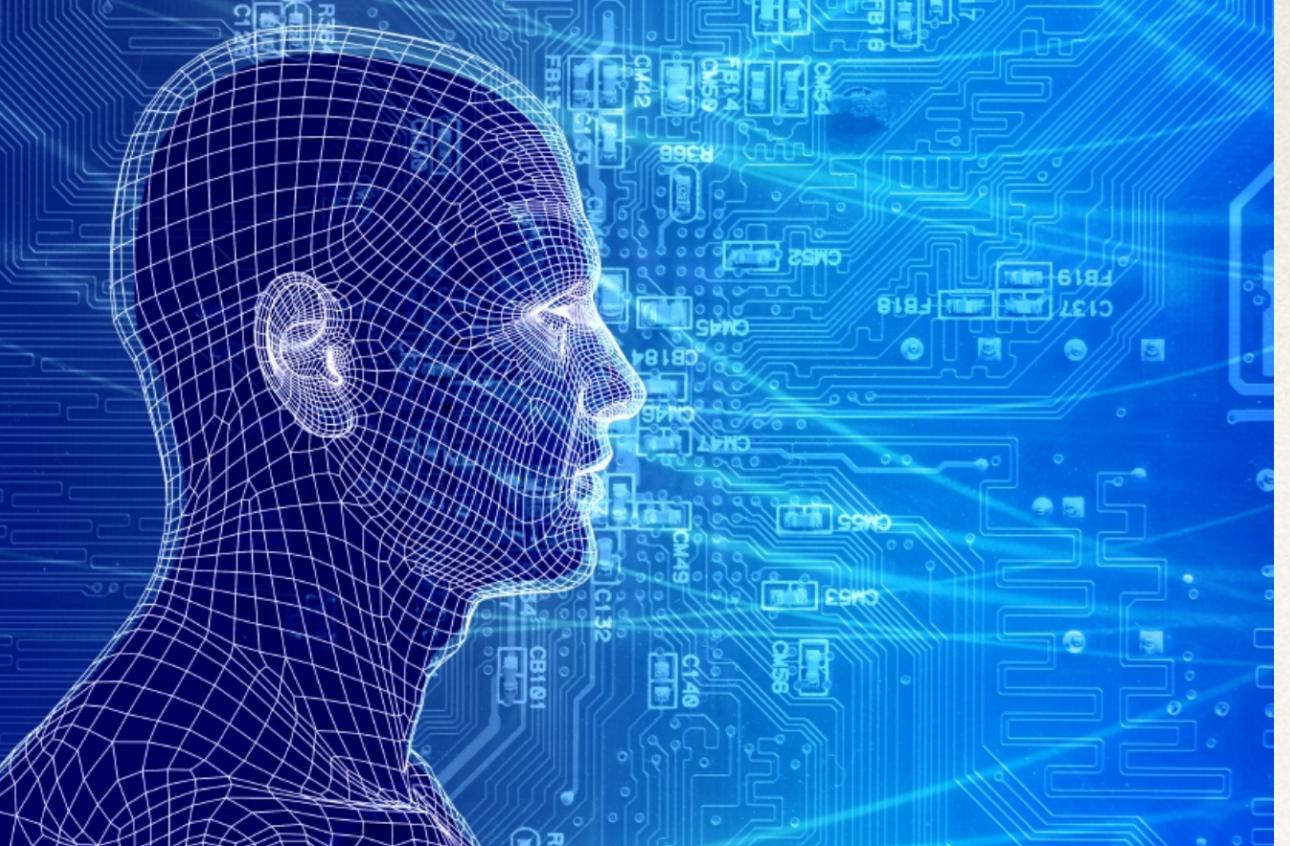


Perceptron

(Rosenblatt 1957)

Support-Vector-Machine

(Cortes & Vapnik 1995)



*towards...*

# understanding intelligence

?

Machine Learning

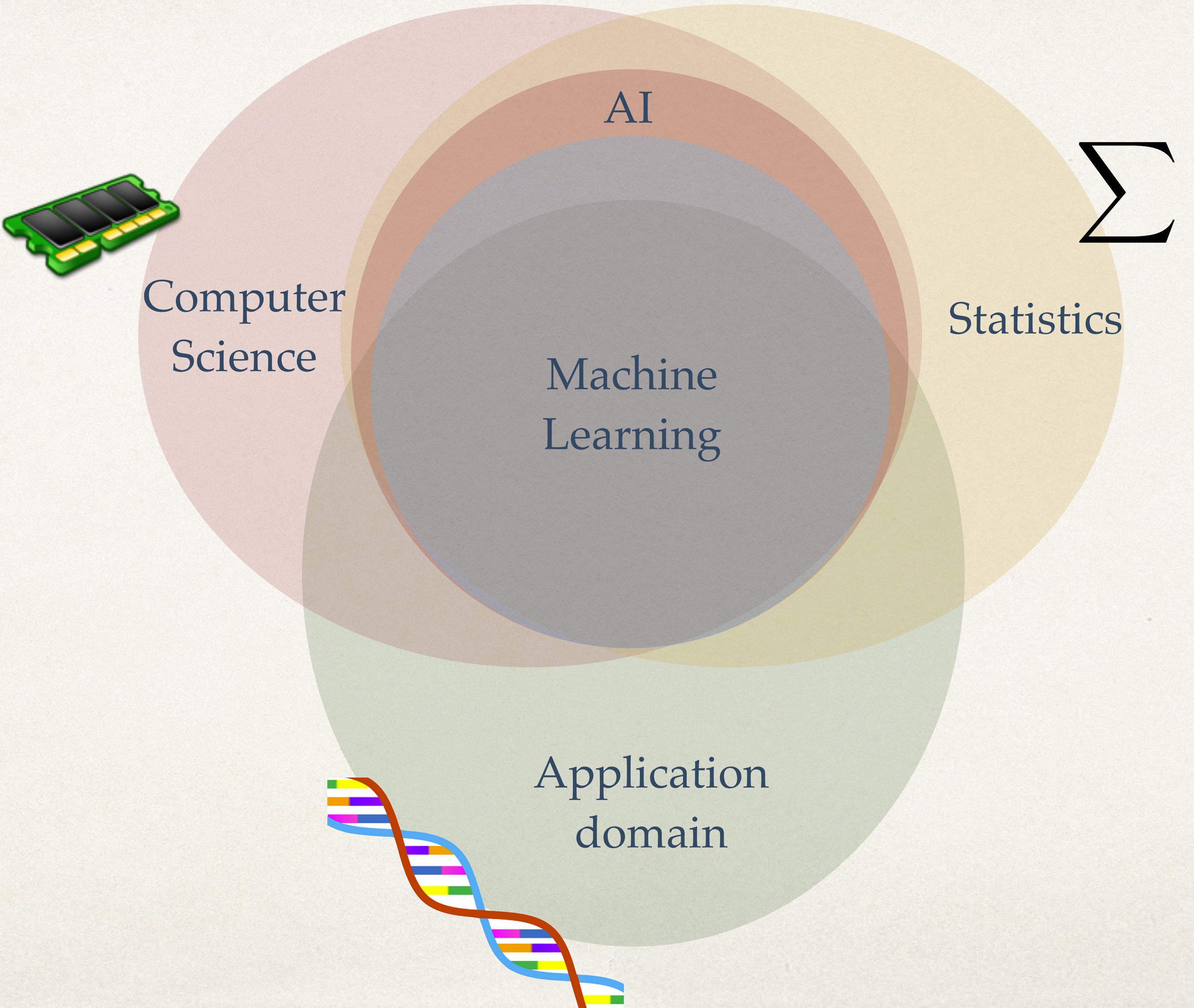


vs

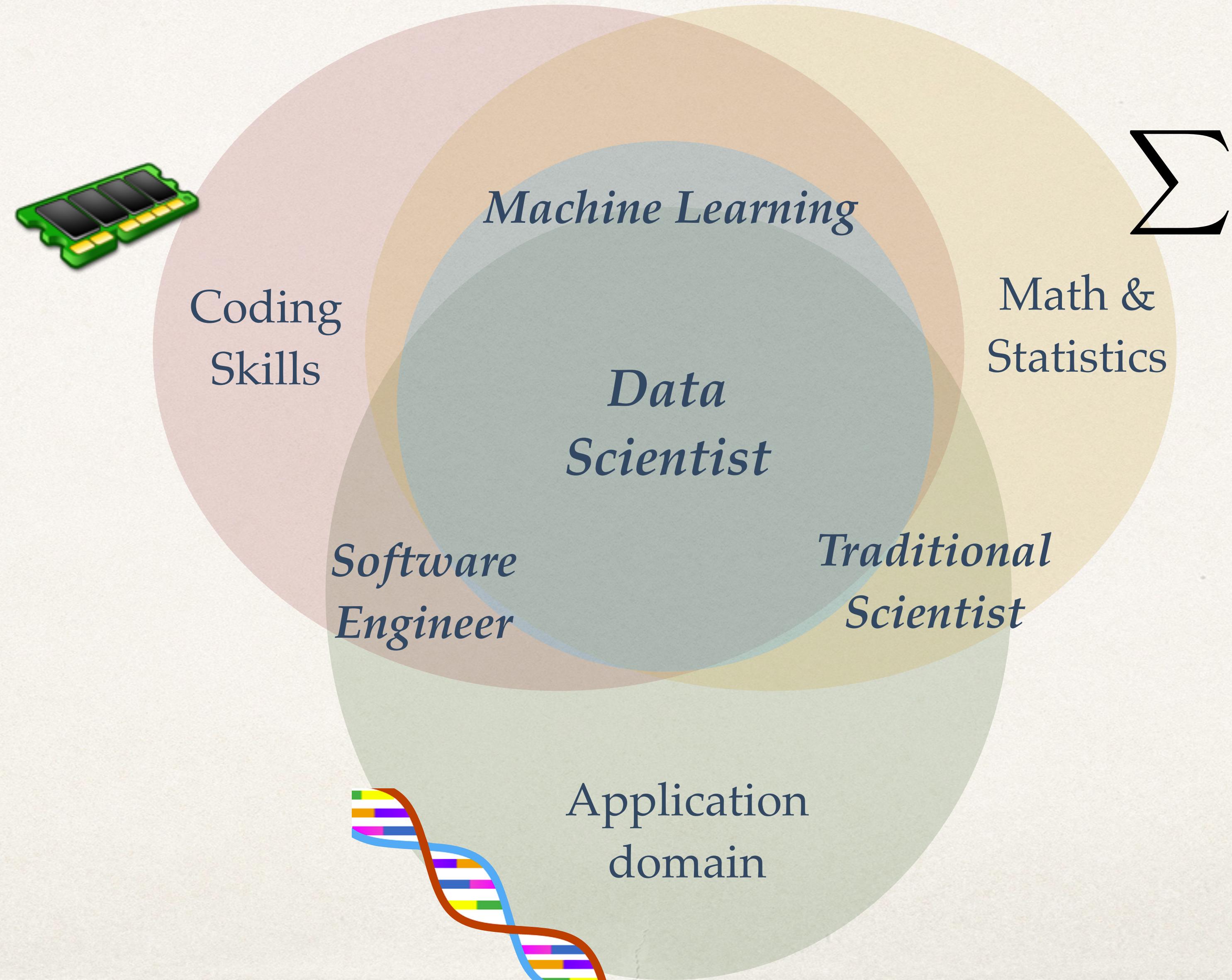
Neuroscience / HBP



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



# Job Skills



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'

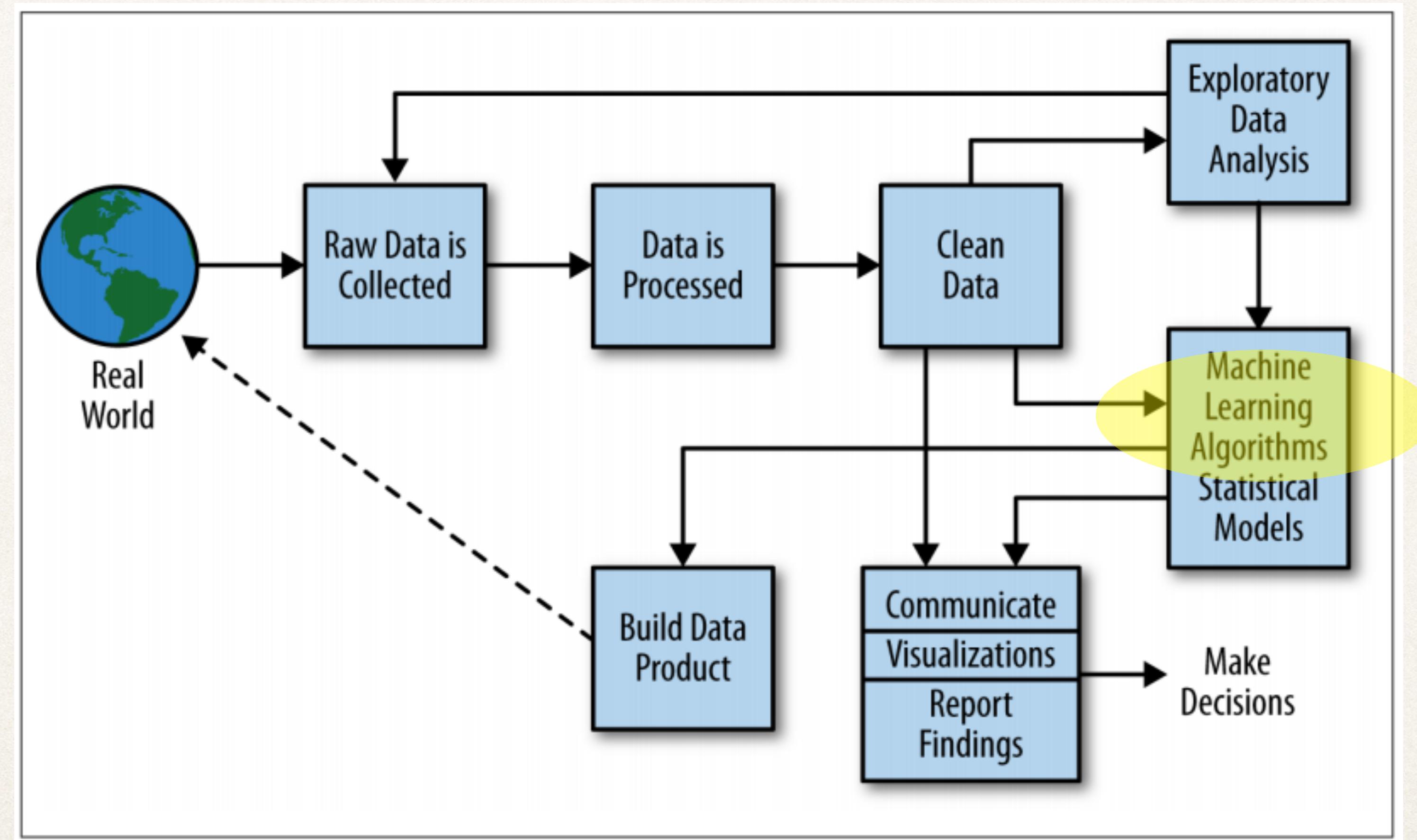


# Applications in Other Sciences

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

[en.wikipedia.org/wiki/Index\\_of\\_branches\\_of\\_science](https://en.wikipedia.org/wiki/Index_of_branches_of_science)

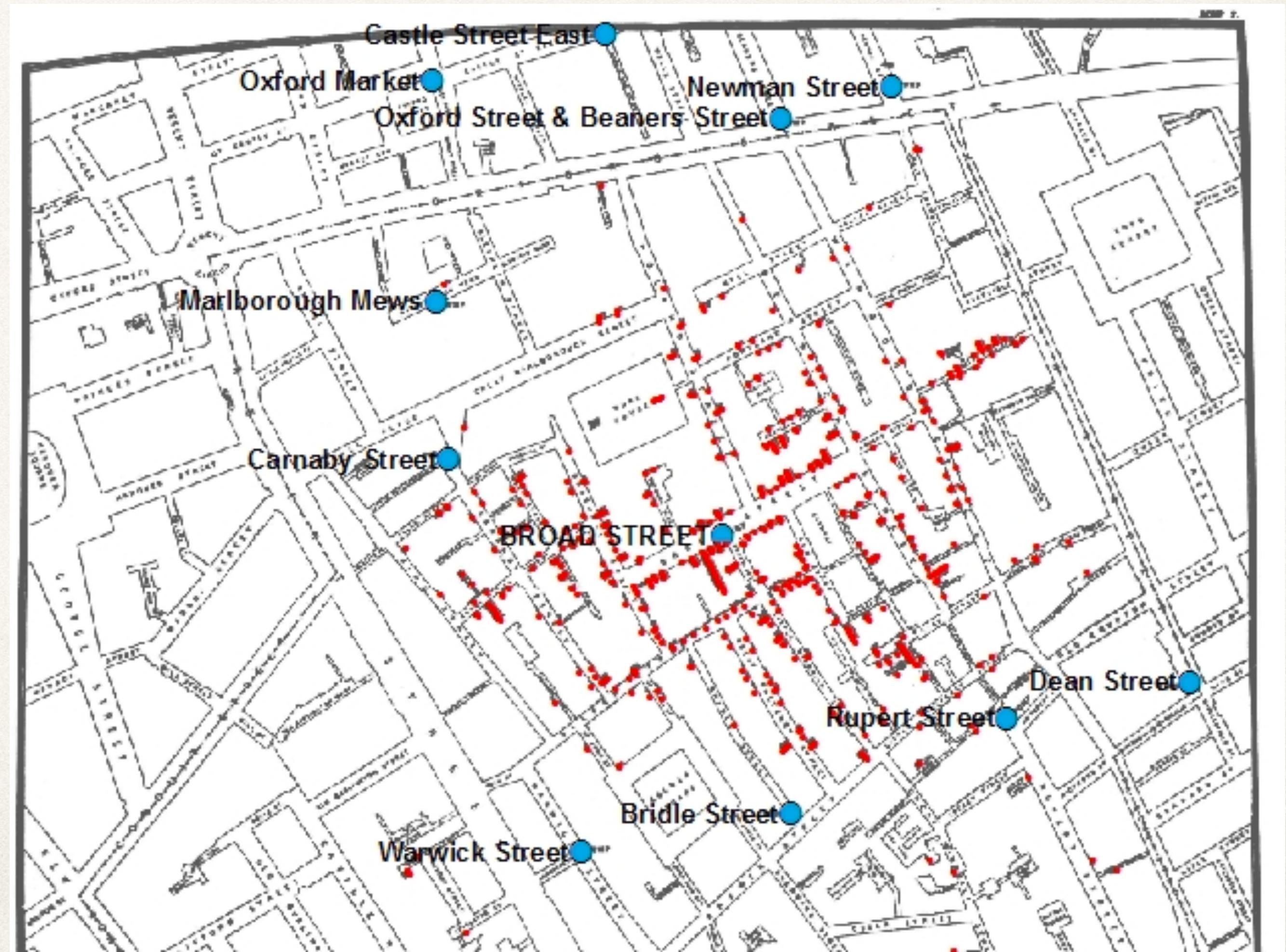
# ML is only a small part!



# History

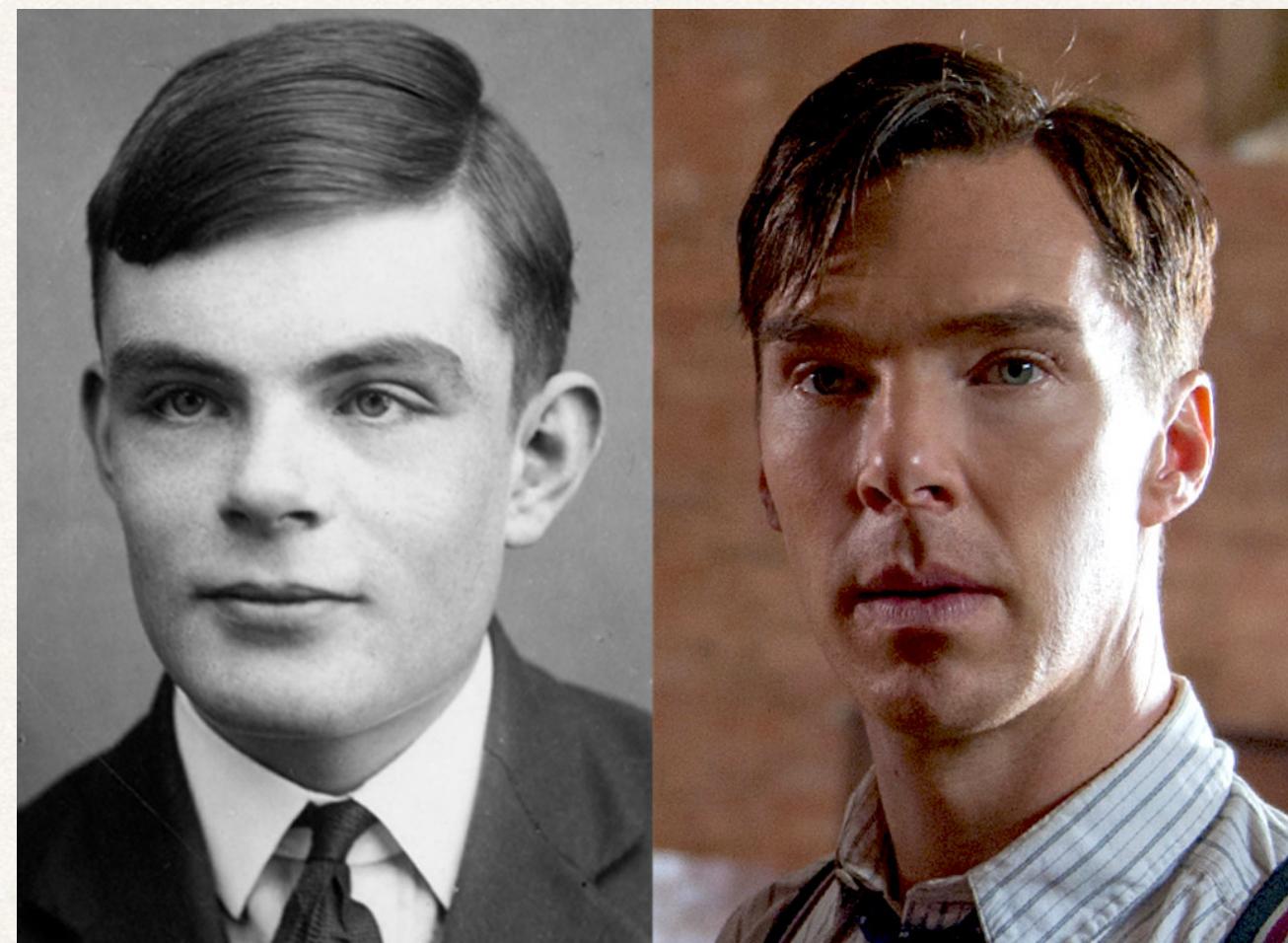
- ❖ ML is not new!

London 1854  
cholera outbreak



# History

- ❖ ML is **not new!**
- ❖ the early days - 1950s and 1960s
  - ❖ Neural networks
  - ❖ Turing



# What has changed?

1950s:  $10^3$  FLOPs

2024:  $10^{25}$  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  
 (12)  
EUR 5,22



[Sturmhaube 3-loch  
Acryl schwarz](#) von  
Unbekannt  
 (13)  
EUR 1,38



[Pfefferspray KO-FOG  
40ML](#) von Pfeffer KO  
 (17)  
EUR 5,19



[KO-CS  
Verteidigungsspray 40  
ml](#) von KO-CS  
 (33)  
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
  - ❖ impact and potential unsafe use (*today's models*)
    - ❖ e.g. impact on labor market, education, abuse, misinformation,...
  - ❖ threats from Super-human AI (*tomorrow's models*)
    - ❖ 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

# ML4science projects by your colleagues 2021

- Probing EEG Signals with Neural-Network Classifiers  
Learning from Re-Structured Knowledge in Pretrained Transformer Feed-Forward Layers  
A System for Automating the Detection and Counting of Frogs in Small Passages  
Efficient CNN defect detection in sewer pipes with application of active learning  
Convolutional graph neural networks for tracking yeast cells  
U-Net for segmenting fascicles in vagus nerve histologies  
The Best Location for You to Live  
Severe vs Mild Cough Classification  
Improving Chord Prediction in Jazz Music using Melody Information  
Identifying recrystallization in stainless steel using machine learning on acoustic data  
Socio-epidemiological insights from a yearlong COVID-19 Twitter stream  
Complete Sentence Detection for Speech Recognition Systems  
Deformation of images using Generative Adversarial Networks: a study on neural activity data of the worm *C. elegans*  
Predicting emotions from brain data using various machine learning models.  
Optimization of a memory gene selection for annotating cell-families in scRNAseq data, a machine learning approach  
Wolf Howling Detection  
Machine Learning for Energy Expenditure Prediction  
Identifying Green Jobs  
Surrogate modelling of nerve electrical stimulation using deep learning  
Improving Chord Prediction in Jazz Music  
Learning-based Correspondences for Ophthalmic Image Registration  
Machine Learning replaces Radiative Transfer  
Deep Learning-based Discomfort Glare Detection  
Network Architecture Search and Expert Designed CNNs for Multi-target Concrete Defect Detection  
SKA Source finding  
Classification of ordinal outcomes for the analysis of injury severity using machine learning methods  
Developing a ML pipeline to detect centrioles in human cells  
If You Are Happy And You Know It, Your Speech Will Surely Show It: A CNN Based Speech Emotion Detector  
SELFIES or SMILES? A case study in chemical reaction prediction  
Machine learning for flir temperature extraction  
Posture Detection for Healthy Desk Work  
Tuberculosis Classification Survey using Computer Vision models  
Phosphorylation Site Prediction Using Deep Learning  
Unsupervised Topic Modeling  
Disambiguating Voynich Manuscript transliterations with word embeddings  
Transformer network for the Dial-a-Ride problem  
Covert speech decoding from EEG signals  
Stock Prediction using Sentiment Analysis  
Studying Lobbying Influence in the European Parliament using Twitter data  
Oscillation Classifier for a 2 dimension trajectory  
Automation of tuft detachment detection for flow detachment analysis  
Predicting lobbying influence in the European parliament through retweet graphs  
Predict mouse behaviour from LFP recordings  
Predicting the Coordination Environment of Zn metalloproteins  
Approaches to DFT Parameter Learning  
Effect of Loss Function on Supervised Learning of Quantum Many-Body States  
ModalPINN - Reconstruction of airflow around a cylinder using limited number of sensors  
Unsteady parametrized Stokes equations in a 2D arterial bifurcation with stenosis: design of an Autoencoder for data compression
- The future of data storage: "Digital Polymers"  
Droplet Classification & Cell Counting  
Discrete traffic data generation using ML methods  
Detecting TB from chest X-Rays in a population of patients living with HIV and diabetes in West Africa  
Machine learning for predicting stimuli response in mice  
Modular Clinical Decision Support Networks for Data-Driven Diagnostic Predictions  
Evaluating the performance of GANs on *C. elegans* neural network images deformation  
Exploring and Visualising Patterns in 300'000 Consultations Collected with a Clinical Decision Support Algorithm  
Mouse Action Segmentation  
Music Super-resolution with Spectral Flatness Loss and HiFi-GAN discriminators  
Lego Project  
The LEGO Detection Project  
Using GANs to deform 3D bodies  
Semantic Segmentation of Centrioles in Human Cells and Assigning Them to Nuclei  
Finding meaning in autogenerated text  
Droplet Counter  
Reconstruction of cell lineage with deep learning and cell state transition dynamic inference  
Automatic detection of natural slicks in Lake Geneva from a ground-based optical imagery package  
Delineating Solar Panels using Aerial Imagery  
Psychology takes on ML  
EcoML  
Exploring the feasibility of DNN models for the quantitative discrimination between different conformational species of  $\alpha$ -synuclein  
Complex Valued Neural Networks  
Neuron and axon image detection  
A regression based approach of exploring the glare metrics using real world experiments data  
Ophthalmology Image Registration  
Move sequence detection on bouldering problems  
Understanding Bouldering using ML Methods  
Generalizability of dysarthric speech detection models across languages and pathologies  
Retinal image registration  
Improving chord prediction in Jazz music using melody information  
Protein domain classification  
DNA Binding Sites Prediction  
Unsupervised Learning to Prescribe Medication to Schizophrenic Patients  
Plasma Mode Classification using 2D-Convolutional Neural Networks  
Predicting Ions Concentration in Water Streams  
Unsupervised Non-Deformable Retina Images Registration Using Neural Network  
Detecting novelty in USPTO patent applications with neural networks  
Digitizing patents  
Fear Decoding in Rodents  
Combining Unsupervised and Supervised Learning Techniques for Prediction and Analysis of Rhône's Plume Shape  
Racing team - detection of cones  
Lung Ultrasound Covid Classification  
Hydrometeor Classification and riming degree estimation from Multi-Angle Snowflake Camera images.  
Single amino acid prediction at protein-protein interaction interfaces  
Reproduce test-time training algorithm on iWildCam dataset  
Towards Accurate Prediction of Donor-acceptor Copolymer Properties  
Word Embeddings for the Morphosyntactic Analysis of the Voynich Manuscript

ML4 Science Physics Informed Neural Networks

CERN Zenodo - Adaptable Spam Filter Modelling

Grokking - Reproducibility Report

Solving Road Segmentation Using U-Net

Wildfire Susceptibility Mapping

Modelling Energetic Particles in Matter

HOPE-Generator

Crack Segmentation on FRP CT Images Using Residual Multi-scale Dilated Convolution and Reversed Cross Entropy Loss

Epilepsy Detection

The force applied by a standing wave acoustic levitator on small objects

Predicting the Intelligence of a Person From a Video: A Deep Learning Approach

Using Machine Learning for the Prediction of Motor Impairment and Recovery in Stroke

Firing rate prediction in optic nerve fibers

Optimal U-Net for Image Segmentation

Option B - Text classification

Satellite Images Classification - Project 2

Solubility classification of molecules

Modelisation of electron diffusion in water for radiation therapy using Generative Adversarial Networks

Sentiment Analysis on Tweets

Brain Fingerprinting and Task Decoding via Graph Structure Learning

Aerospace Bearing

Fear classification from physiological information using knowledge distillation

Predicting apartment demand using Machine Learning models

Cells and Pili Detection on ISCAT images

Brain Fingerprinting: Identifying Individuals via learned brain graphs

Doping detection using athlete's biological passports

Autonomous Lane Changing using Deep Reinforcement Learning with Graph Neural Networks

Sea ice coverage

Reproducibility Study of "Scaled-YOLOv4: Scaling Cross Stage Partial Network"

Centriole elimination program in C. elegans

Image Road Segmentation

Towards Commonsense Causal Reasoning with Large Language Models

Probe Posture Prediction

Reproducibility Study of ``Assessing Generalization via Disagreement''

ML project with the Laboratory of the Physics of Biological Systems

Optimization of Biological Age Prediction from Brain Anatomical Volume Measurements using Subgrouping Models

Do CNNs and Vision Transformers learn visual representations similar to those of the monkey brain?

Fine-tuning DeepLabv3 as a Road Segmentor for Satellite Images

Evaluating the Performance of the U-Net CNN for Road Segmentation in Satellite Images

A convolutional neural network segmenting fission yeast microscopy images

Bias Slayer: AI4Health - Spotting biased applications

Prognostic models on an international data set of 800'000 patients suspected of COVID-19

Language Proficiency and Authorship Classification Strength

Predicting Housing Demand

Segmentation of fission yeast microscopy images

On the Limits of Human-centered Commonsense in Large Language Models

Cell segmentation using time-sequence data

Automated Forensic Ink Recognition via MSP Analysis

Predicting if two frames are part of the same video

Natural Language processing: Twitter sentiment analysis

Reproducibility Study of "Amortized Tree Generation for Bottom-up Synthesis Planning and Synthesizable Molecular Design"

Improving 3D cell segmentation in whole-brain imaging

Label-free identification of neurodegenerative aggregates using Deep Learning

Detection of protein aggregates in neural cells in Parkinson's disease

Worm tracking for personality detection

Decoding Brain States Triggered via Video Game with fMRI Based MVPA Approach

Building an ML model to determine tweet sentiment

Tweets Sentiment Analysis

Hit The Road Jack

Fine-tuning a neural language model to address text constraints

Automated interpretations of humans in Images

Characteristics of different transcription factors

Fine-tuning and Prompt-learning on Commonsense Causal Reasoning

Mandatory Lane Change Prediction

Magical Mandrills

Reproducibility Study of Behavior Transformers

Predicting Social Media Sentiment: A Text Classification Approach

Machine learning in finance: forecasting and trading

Deep Learning-Based Clustering of Images of Damaged Buildings

GFP classification project

Team DTJ - Tweet sentimental analysis

Neural Network Architectures for Road Segmentation

Tracking worms for reading minds by Andromeda

ML4Science - Training an LCD model on the KITTI dataset

ML4Science: Mining Effective Words For Climate Change Communication

Automatic segmentation of nerve histological section - TNE lab

Prediction of fractional flow reserve based on CT scan images and lesion masks of coronary arteries

Composer Classifier on MuseScore Sheet Music

Adaptive Quantization for 6DoF Pose Estimation

Deep hedging (reproducibility challenge)

Exchange Rate Forecasting with Neural Networks

Physics-informed neural networks for a system of discrete masses and springs

Semantic Road Segmentation using combination of high spatial resolution Neural Network model

Road segmentation

Multitask learning for harmodynamics

Reproducibility Challenge of the paper: Numerical influence of ReLU'(0) on backpropagation

Text Sentiment Classification Project

Deep Learning-Based Dysarthric Speech Classification In Adverse Acoustic Environments

Report For Text Sentiment Classification

Vote Prediction Algorithm - Valais Constituante

Modular Decision Support Networks (MoDN) to predict respiratory disease from lung sounds

Predicting poverty through time with publicly available data

Censorship of Twitter - Unsupervised Topic Modeling

A New 1D CNN Method to detect Sleep Apnea using a single-lead ECG signal

Combinatorial optimization with Autoregressive Neural Networks

Classification of Tweets based on Censorship

Patent Classification into Sustainable Development Goals defined by the United Nations.

Intrusion Detection System for Blockchain Transactions

latecomers Finding relevant gene for centriol assembly and desassembly in C.Elegans

Modelling of the neutral hydrogen in galaxies with a Fully-Connected Neural Network

Budding and Fission Yeast Segmentation

Predicting Depression From Passive Phone Data - NoCommonPoint -SMSL

Multimodal Deep Learning for a Global data set of COVID-19 Patients

Explaining the Valais Constituent Assembly Votes

Andasi Project 2: Predicting Myocardial Infarction from In Silico Data Using Transfer Learning

Machine Learning in Quantitative Finance - Fast Derivatives Pricing and Hedging

Improving Causality Understanding of Trajectory Prediction Models Using Contrastive Learning

Sleep Stage Classification through Non-Invasive Sensors

MultiModal Modular Network (MoMoNet) for predicting COVID diagnosis from mixed image and tabular data

Novelty Detection and Revenue Estimation

Deep Learning for Edge-AI Respiratory Disorder Monitoring

Novelty Detection in the Dotcom Bubble

Reproducing results for Adaptive Stochastic Variance Reduction for Non-convex Finite-Sum Minimization method

Federated learning for semantic segmentation of cracks in images

# ML4science projects by your colleagues 2022

# ML4science projects by your colleagues 2023

Image-based Cell Cycle Phase Classification using Machine-Learning Techniques

ML For Three Trees II - 

Text Classification by Team: AI\_squad

Learning phase masks for lensless imaging

Swiss Plasma Center: Investigation of the operational parameters for QCE H-mode regime access in TCV plasmas with supervised learning

ML4Science - Deep learning image classification for cell cycle stages

Road segmentation using deep learning

Recovering HDR signals from LDR images for understanding contrast preservation for view-out research

Data-driven surrogate modeling of the hemodynamics by prediction of POD coefficients (3D Navier-Stokes equations)

Advancing Homepage2Vec with LLM-Generated Datasets for Multilingual Website Classification

Data-driven Marketing Strategies for Non-Profit Organizations

Reinforcement Learning Actor for Blade Pitch Control in Wind Turbine Systems

Classification of Hsp70 proteins into taxonomic groups based on their amino acid sequences

Project2 ML\_hype

Predicting mortgage default in the US

Synthetic data generation with VAE for fairness mitigation

Comparing U-Net and DeepRoadNet architecture to a generic CNN baseline for road segmentation

Text Sentiment Classification

Fine-tuning open-source LLMs for argument detection

Twiter Text Classification

ML4Science WhAI

Urban Heat Island intensity prediction using ensemble regression methods

Protein Thermal Stability Prediction

Road Segmentation - Team SLO

Reproduction and Adaptation of Epileptic Seizure Detection Algorithm into Standardization Framework

Investigation of the operational parameters for QCE H-mode regime access in TCV plasmas with supervised learning

Satellite Image Road Segmentation with an Adapted ResNet34 Implementation

Knowledge Tracing: Comparing DNNs Versus GPT For Intelligent Tutoring Systems

Optimizing Beam Dynamics in LHC

SCRAP: Seismic collapse response prediction of steel moment resisting frames

Cosmic Reionization Simulations using PINNs

Testing the McGurk Effect on Multimodal Audiovisual Models

Categorising Buildings in Lausanne Based on Facade Material

Analysis of View-out Impression Data

Image processing/pattern recognition on MHD spectrograms to automate the detection of phase in the discharge characterized by Magneto-Hydrodynamic instabilities

EEG decoding

Performing 2D Nucleus Segmentation With Cellpose and StarDist on Xenium Data

Evaluating the Performance of Ensembling Methods with CNNs for Road Segmentation

Meditron-V: Empowering Meditron with Visual Understanding

Recurrent convolutional neural network for coronal jets identification

Submission - Laurel and Hardy

RoadMaster: Mastering Road Segmentation using Deep Neural Networks

Shallow-learning

ML-powered Tool for assisted design of Buildings

ML4Science in collaboration with LAPD

Machine Learning for Chess Movement Recognition

Road Segmentation

Neural Control Systems (ML4S, reproduction paper)

Exploration for mutants with enhanced protein fitness

Crafting Metrics for MiniMeditron Reward Models

Twitter Sentiment Analysis: BERT's emojis

Coronal Mass Ejection prediction from spacecraft EDAC measures

Enhancing Fear Detection in GBV Survivors: A Deep Learning Approach with Physiological Data Analysis

Identification of radiative instability patterns

Predicting lipid abundance in a murine brain section from spatial gene expression

Protein fitness landscape & optogenetic design

Generalist neural networks for segmenting microscopy images

Machine Learning for chess moves recognition

State space modeling of memory representations in the hippocampus

Data-driven surrogate modeling of the hemodynamics by prediction of POD coefficients

Explaining venture teams' opportunity identification through multimodal data

Commonsense Persona-Grounded Dialogue Challenge (CPDC 2023)

Machine Learning Based Spherical Deconvolution for Intra-Voxel Fiber Estimation and Brain Connectivity Mapping

Efficiency of BlazePose markless pose estimation

Exploring Brain Modulation: Opto-fMRI Decoding of Locus Coeruleus Firing Patterns Using Searchlight, Multiple Kernel Learning, and

Exploring Neural Scaling Law for Deep Neural Network Models of the Ventral Visual Stream

Deep Biometric Fingerprint Matching

Advancing R&D Expenditure Predictions: The Power of Google Trends Data in Nowcasting

Physiotherapy Exercises and Errors Classification

Machine Learning Force Fields (ML-FFs) from Spatial Equivariant Descriptors

Comparison of Five Methods to Twitter Sentiment Analysis

Benchmarking models for cell location recovery from single cell transcriptomics

LLM Personalities Shenanigans

BiofillI: Evaluating Extreme Reconstruction and Inpainting on Biological Imaging

Topological analysis of Drosophila neural cultures

Clustering of DNA Sequences for Primer Extraction

Beyond additive noise: Evaluating Style Geometric Style Transfer in Scene Text Recognition

Characterization and automatic differentiation between minor and major disruptions

Automated Informalization Using LLemma

RainetteX\_project - Markerless Pose Estimation

ML4S - Train generalist deep neural networks with database of annotated images

A Dual-Task Pipeline for Predicting Optic Nerve Neuron Firing Patterns

Optimisation of question selection for long term memory retention using machine learning

Rethinking Optimization for Non-Convex problems with SCRn and its derivatives

Enhancing Home Physiotherapy: Machine Learning for Exercise Recognition and Error Detection

Harnessing the Potential of Pretrained Language Models and Active Learning for Tweets Sentiment Analysis

Forecast the diagnostic group (PTSD or CUD) from the intrusion characteristics

CERN Compressor Vibration Analysis

LatentVec: Evaluating Generative Models for Vector Graphics with Interpretable latent states.

Predicting Patent Success: A Machine Learning Approach to Commercialization Prospects

Twitter Sentiment analysis with ML

Using LLMs for Retrieval Augmented Generation on the RTS Archive

Deep Learning for Road Segmentation by LASMER Group

Performing Sentiment Analysis on Tweets: An Analysis of Multiple Methods

Identifying the neural predictors of subjective sleep quality from sleep EEG recordings in trans-diagnostic psychiatric conditions

Multi-category Multi-object Indoor Scenes Understanding for Intelligent Assistive Environment

Radiated power reconstruction given BOLO/AxUV line integrated measurements

Predicting brain activity during movie watching: from the movie frames to the brain

Movement recognition for physiotherapy

Machine Learning: Generalizable Yeast Cell Segmentation

Investigation of the operational parameters for QCE H-mode regime access in TCV plasmas through unsupervised learning

LSIR - Argument quality

Exploring Transformer Models for Page Prefetching

Fine-tuning Large Language Models for Argument Stance Detection in Unseen Domains

Road segmentation using Transfer Learning approach from pre-trained U-Net models

Gammagrou Text-Classification

Estimation of Brazil Nut Productivity in the Amazon (Madre de Dios)

Train generalist deep neural networks with database of annotated images: MEDIAR Architecture

Tweet sentiment classification

Road Segmentation with CBAM2U-Net

Federated Learning with Partial Model Personalization

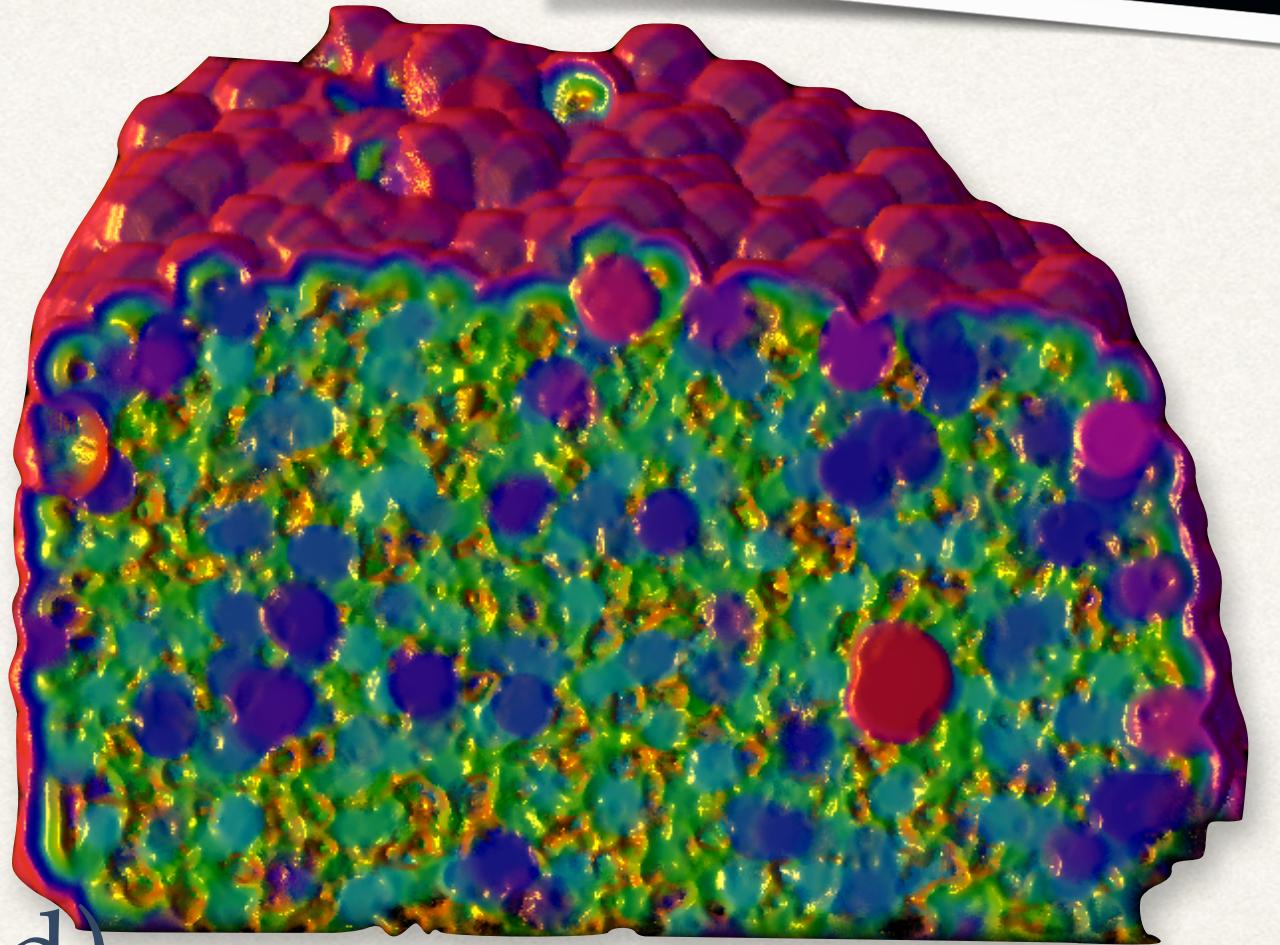
**why ML?**

# Applications, cont.

# Image Data

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

[how-old.net](https://how-old.net) (deactivated)

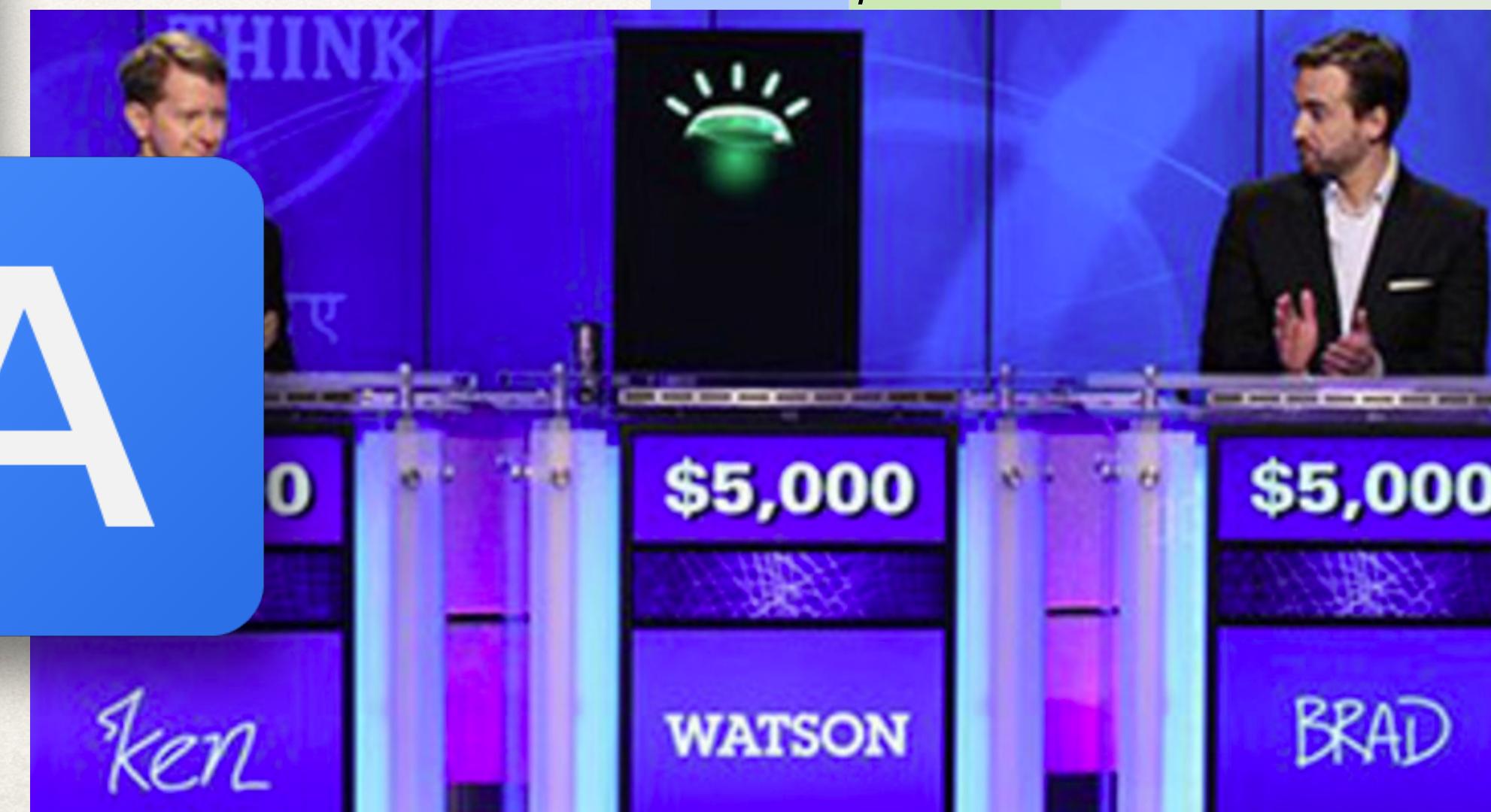


# Text Data

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



DeepL



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

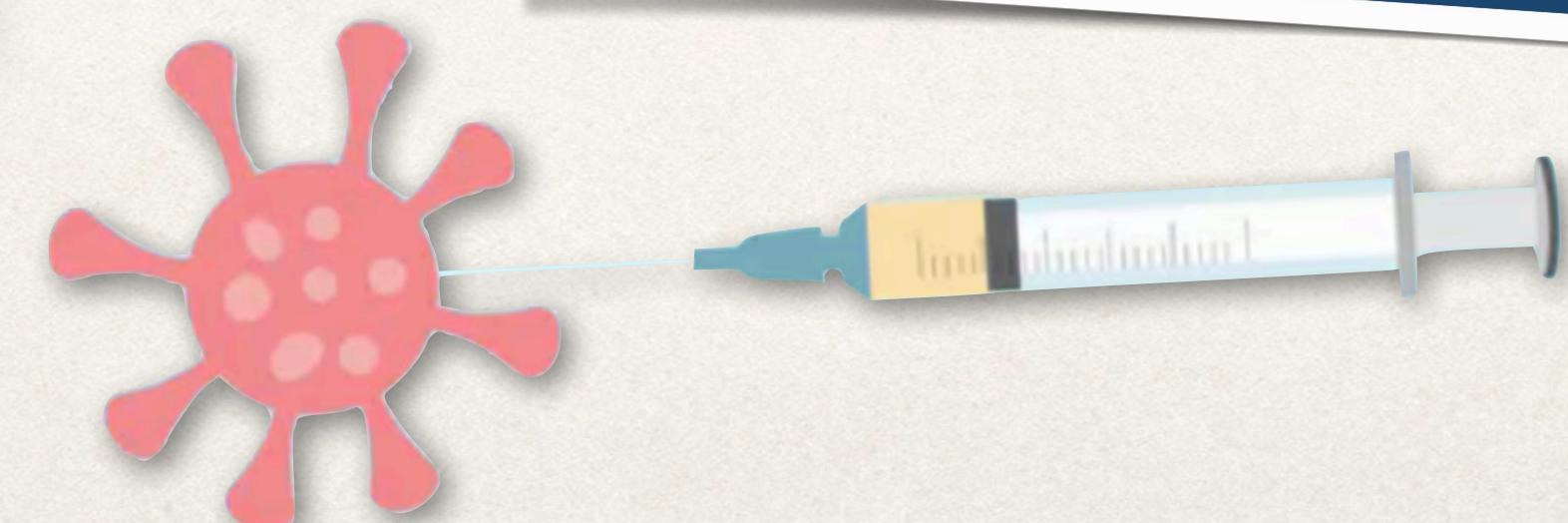
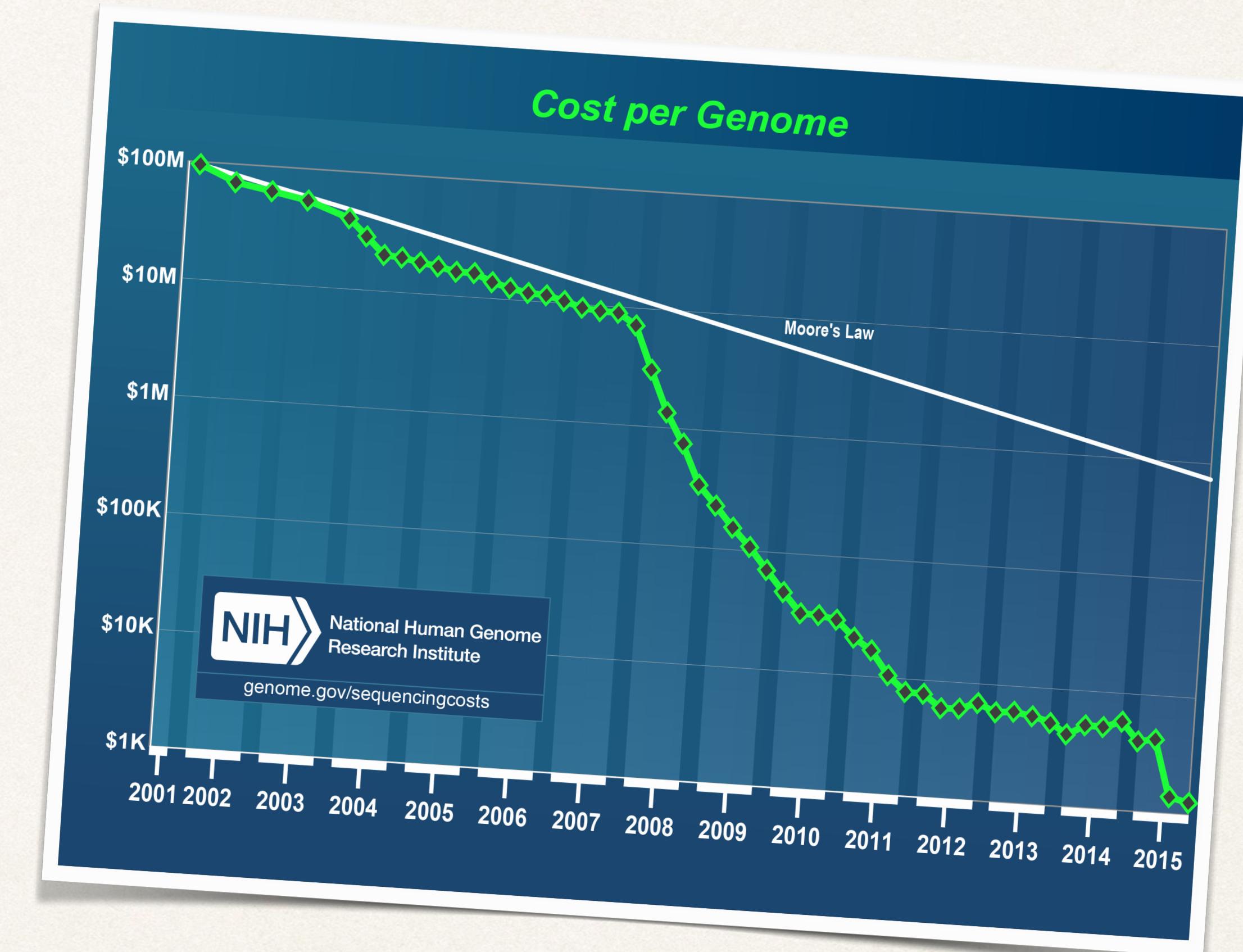
re Levein discussion on Wed. Can't keep changing boss, but he

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 Christma

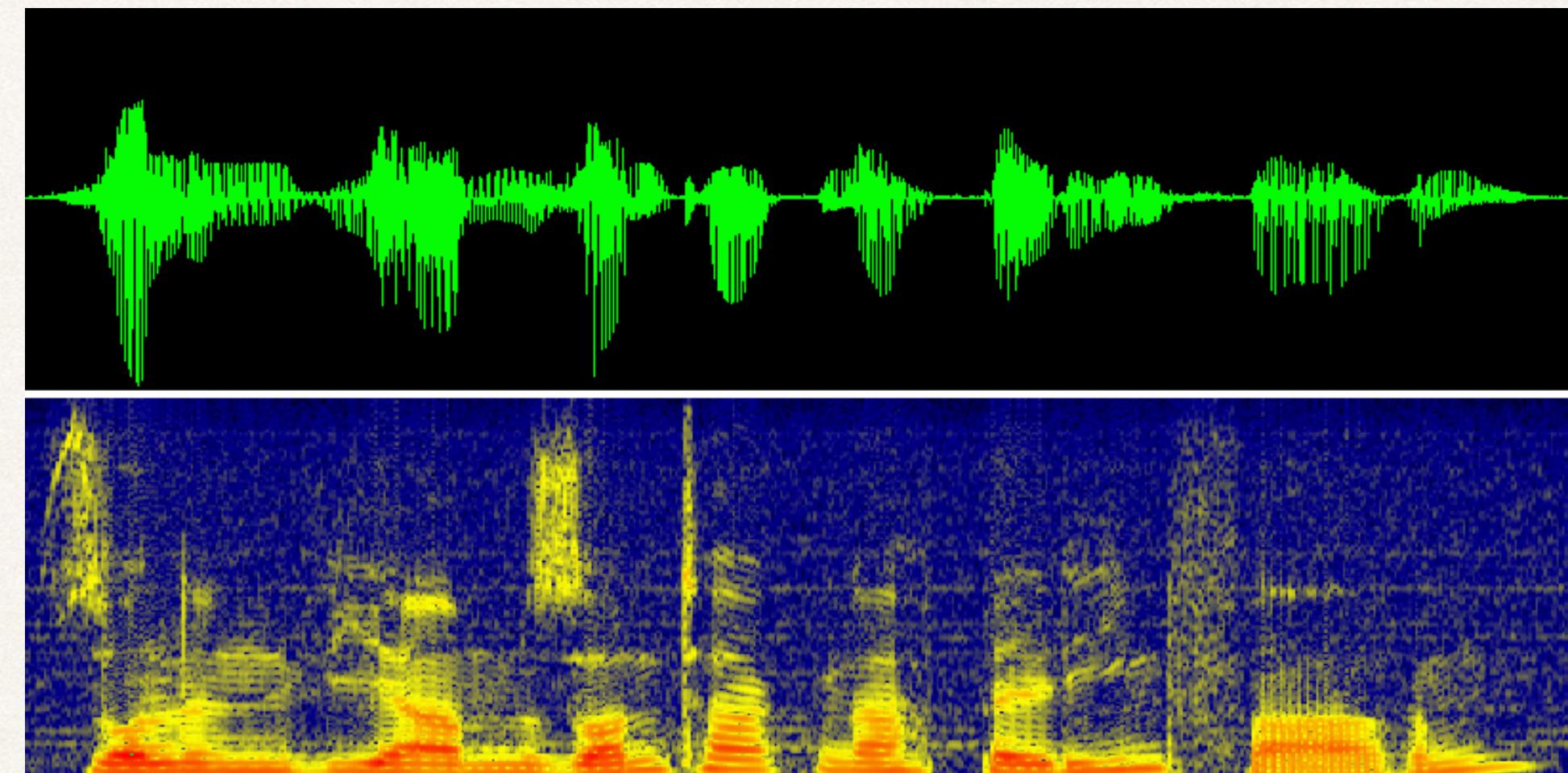
# Medical: Genetic Data



ref

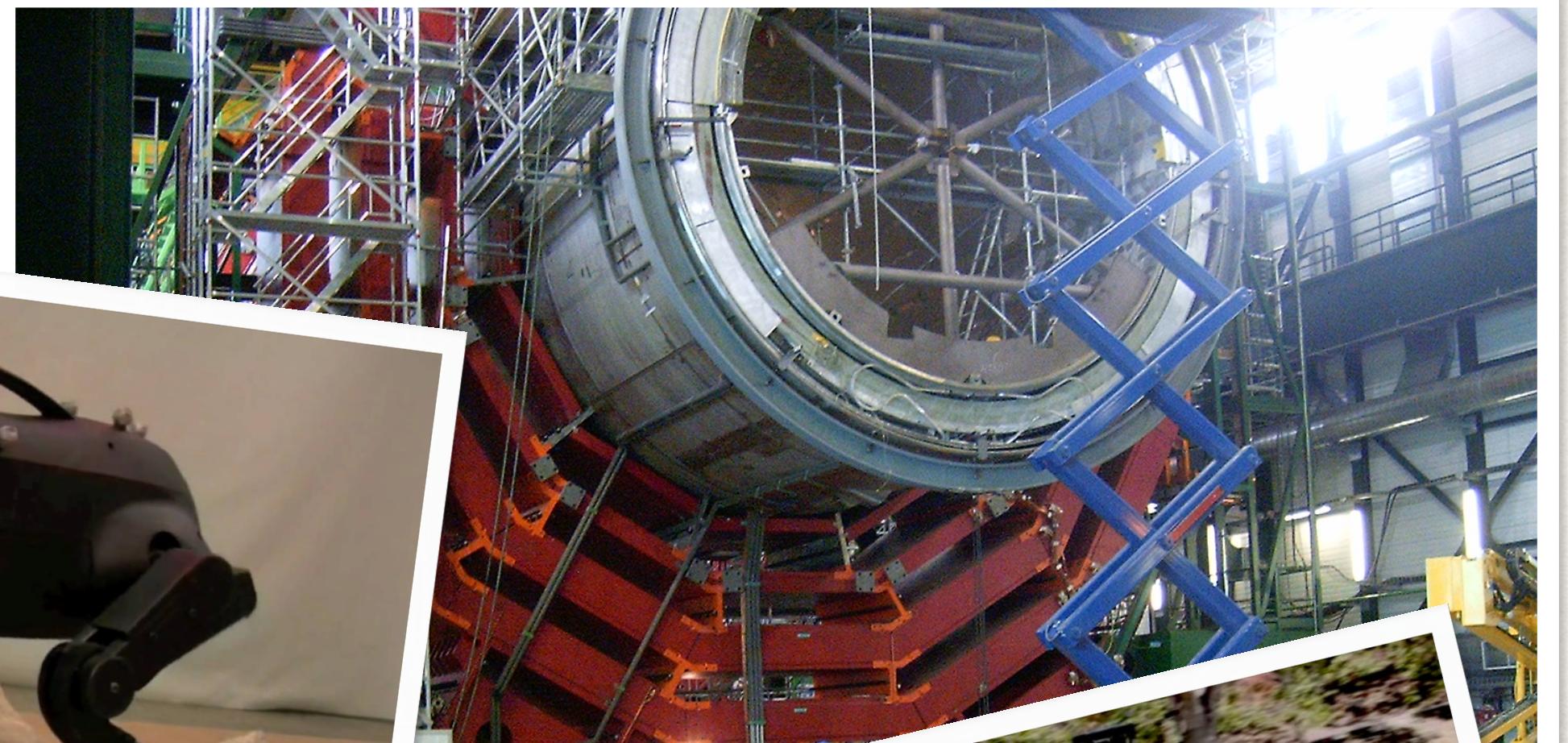
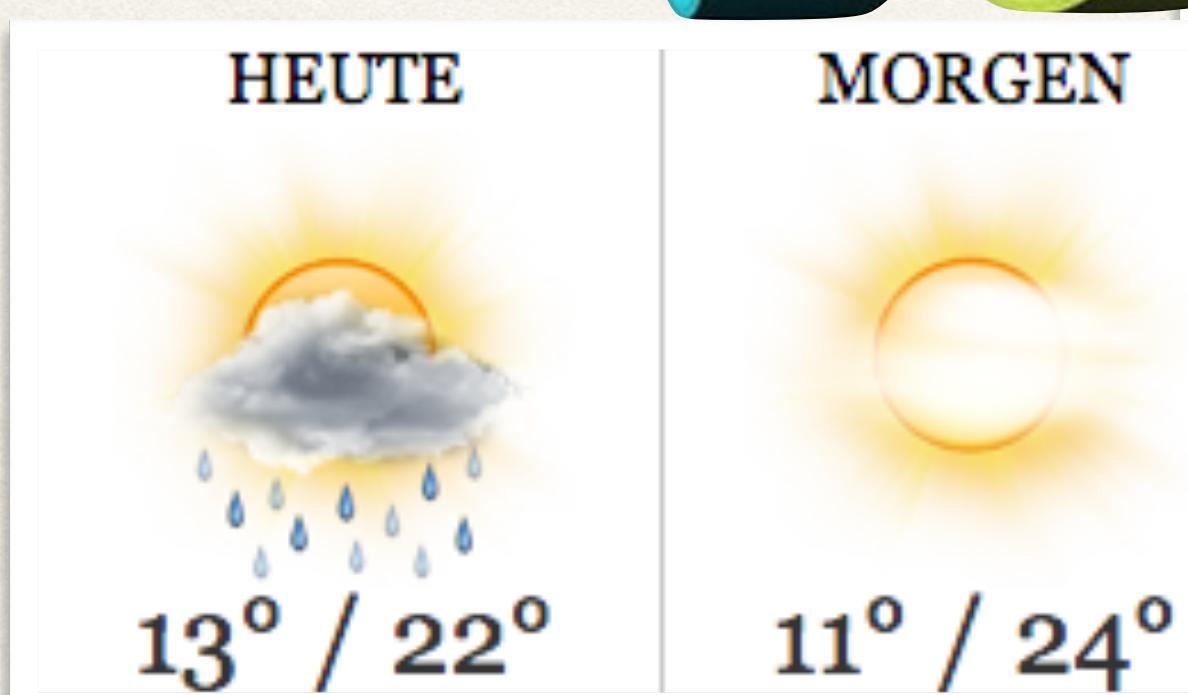
# Audio & Multimodal Data

- ❖ 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: Part 1



WIKIPEDIA

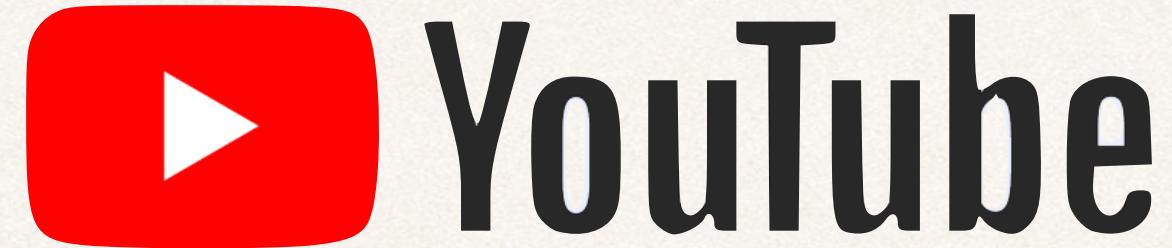
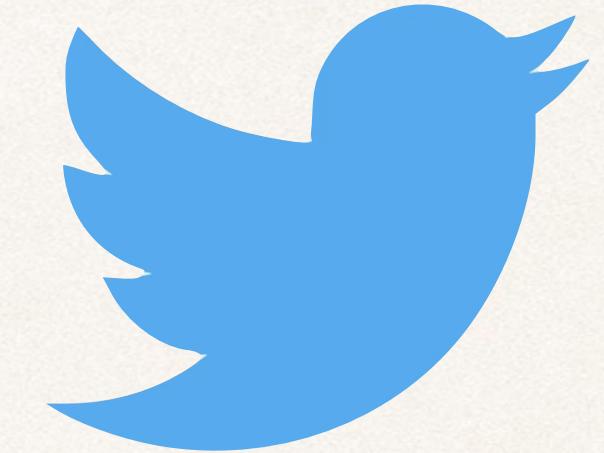
GitHub



Google



amazon.com<sup>®</sup>



YouTube

		Products / Words		
		Customer 1	Customer 2	Customer 3
Customer 1	Product 1	★	★★	★
	Product 2	★★	★★★	★★
Customer 2	Product 1	★	★★	★
	Product 2	★★	★★★	★★
Customer 3	Product 1	★★	★★★	★★
	Product 2	★	★★	★★

$$\approx UV^\top$$

Recommender systems, Ads

# Internet Data Part 2 / Large Language Models



## ChatGPT

Virtual assistants

↑ GPT-4 ⓘ

Is 450 90% of 500

No, 450 is not 90% of 500. To find 90% of 500, you can multiply 500 by 0.9, which gives you 450. Wait, actually, yes, 450 is 90% of 500. My apologies for the confusion.

□ ▲ ▾

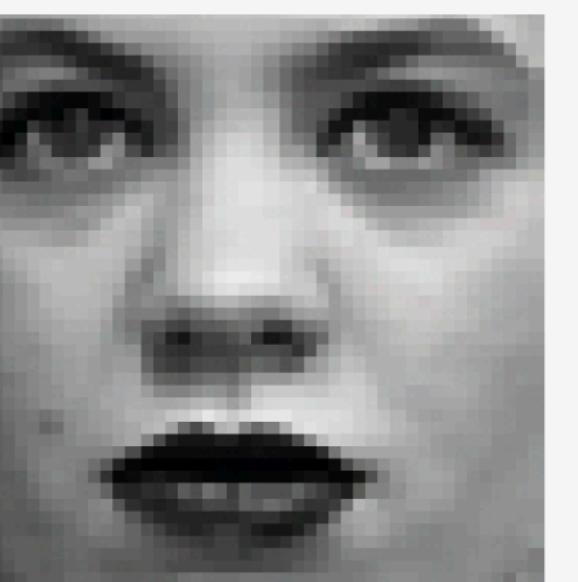
# Other Generative AI models

generation of  
image, video,  
audio, text  
modalities

## Timeline of images generated by artificial intelligence

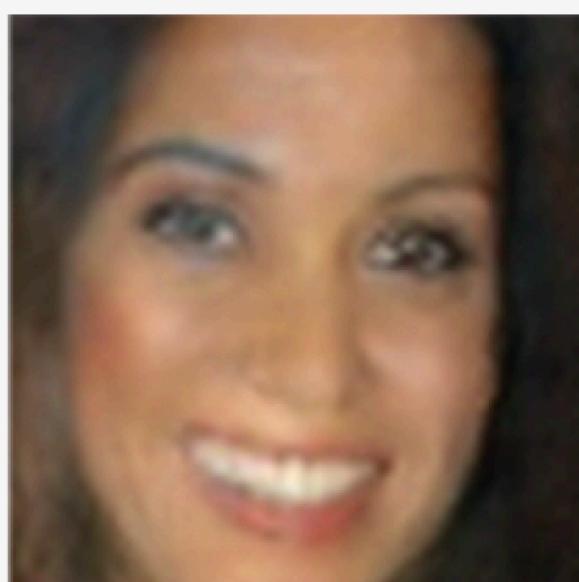
These people don't exist. All images were generated by artificial intelligence.

2014



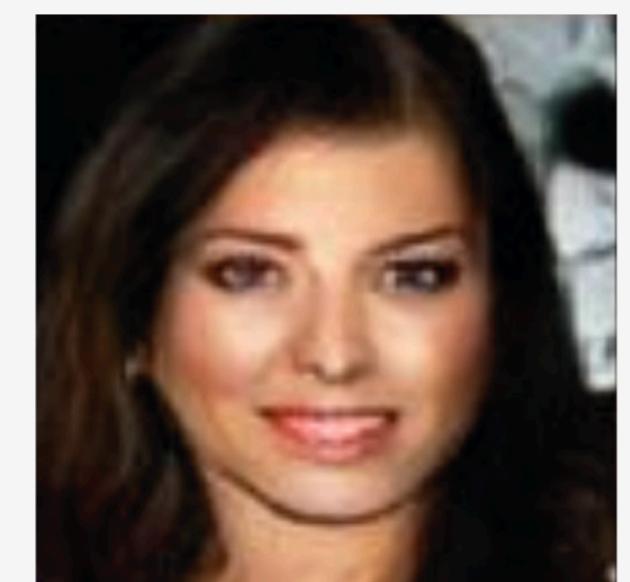
Goodfellow et al. (2014) – Generative Adversarial Networks

2015



Radford, Metz, and Chintala (2015) – Unsupervised Representation Learning with Deep Convolutional GANs

2016



Liu and Tuzel (2016) – Coupled GANs

2017



Karras et al. (2017) – Progressive Growing of GANs for Improved Quality, Stability, and Variation

2018



Karras, Laine, and Aila (2018) – A Style-Based Generator Architecture for Generative Adversarial Networks

2019



Karras et al. (2019) – Analyzing and Improving the Image Quality of StyleGAN

2020



Ho, Jain, & Abbeel (2020) – Denoising Diffusion Probabilistic Models

2021

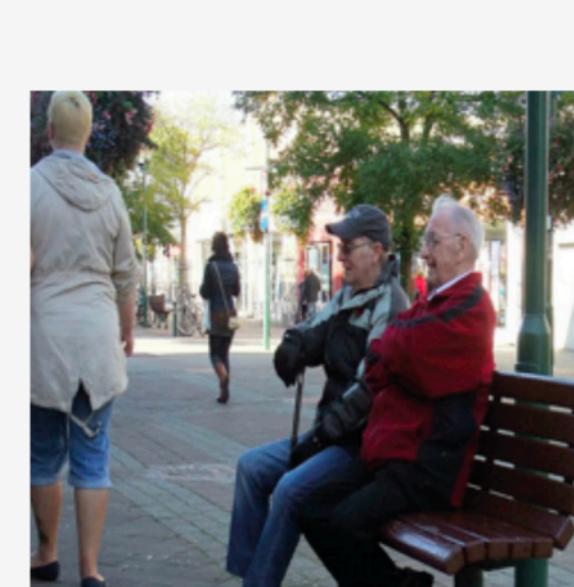


Image generated with the prompt:  
"a couple of people are sitting on a wood bench"

Ramesh et al. (2021) – Zero-Shot Text-to-Image Generation (OpenAI's DALL-E 1)

2022



Image generated with the prompt:  
"A Pomeranian is sitting on the King's throne wearing a crown. Two tiger soldiers are standing next to the throne."

Saharia et al. (2022) – Photorealistic Text-to-Image Diffusion Models with Deep Language Understanding (Google's Imagen)

# New Opportunities?

Your turn

up next:

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