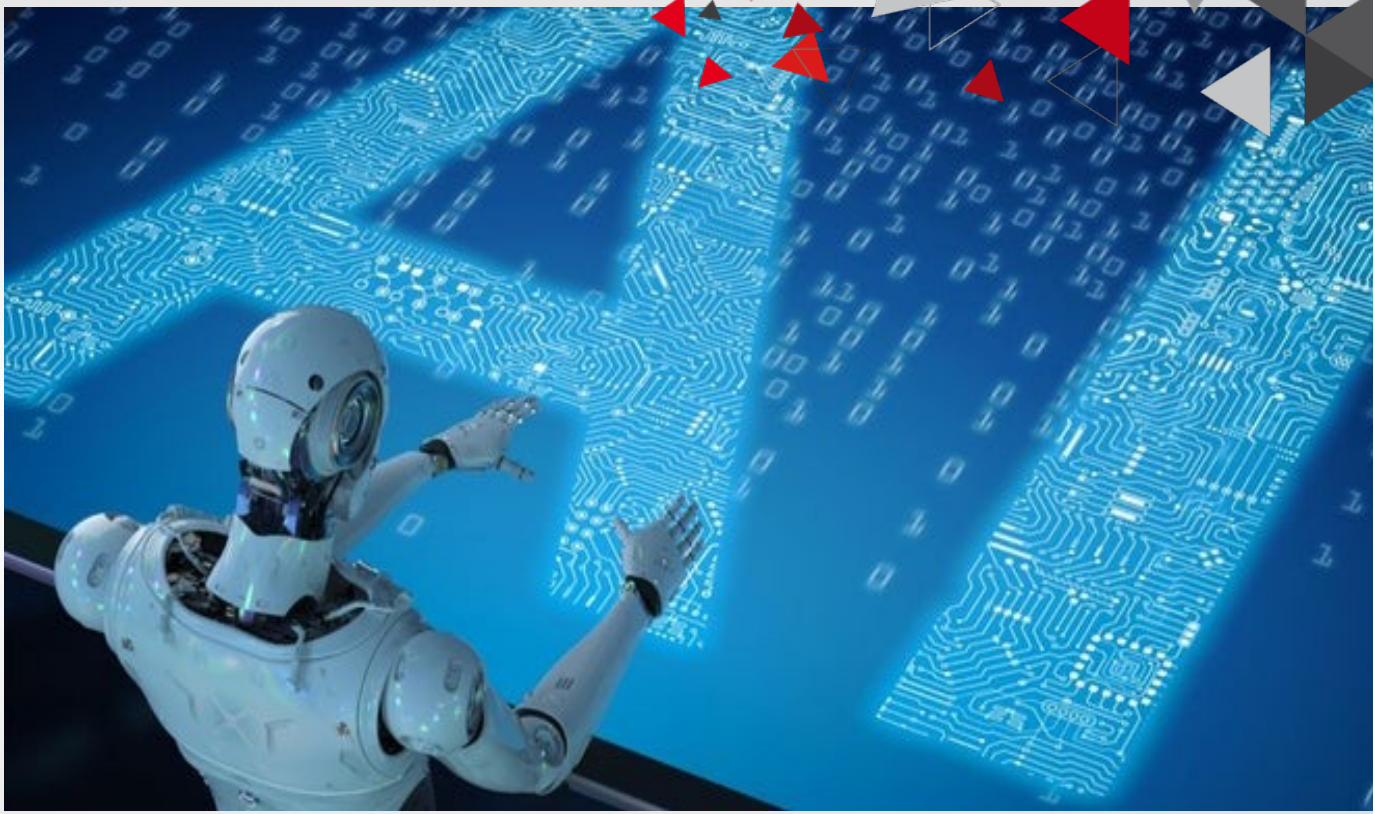
esprit
Se former autrement

Artificial Intelligence

And Ethics



United Nations Educational, Scientific and Cultural Organization
UNESCO Chair "Project-based learning" ESPRIT School of engineering, Tunisia
CONFÉRENCE DES GRANDES ÉCOLES
EXCELLENCE FOR A COMPLEX WORLD
EUR-ACE®
Cti
cdio
CONCEIVE DESIGN IMPLEMENT OPERATE

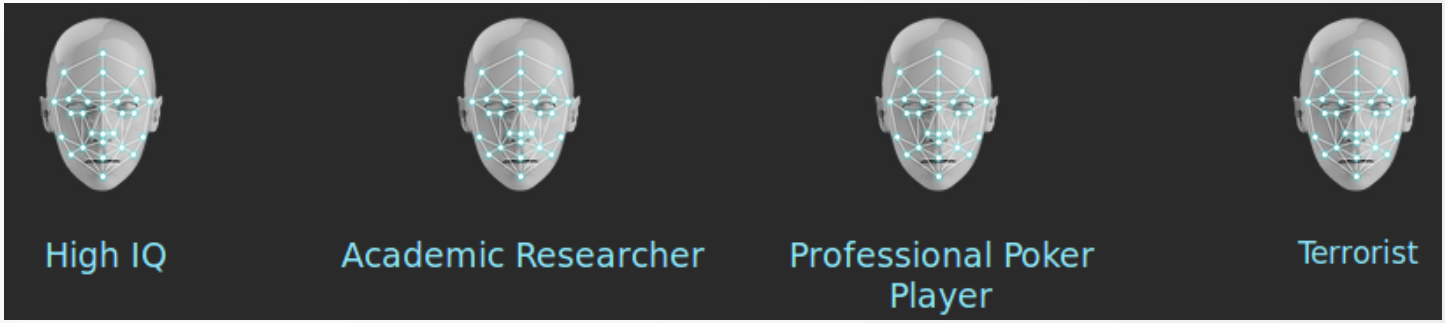
AI is changing our lives

Let’s start with some real applications

AI is changing our lives

Facial recognition for security

Facial personality analysis



AI is changing our lives

Google autonomous car

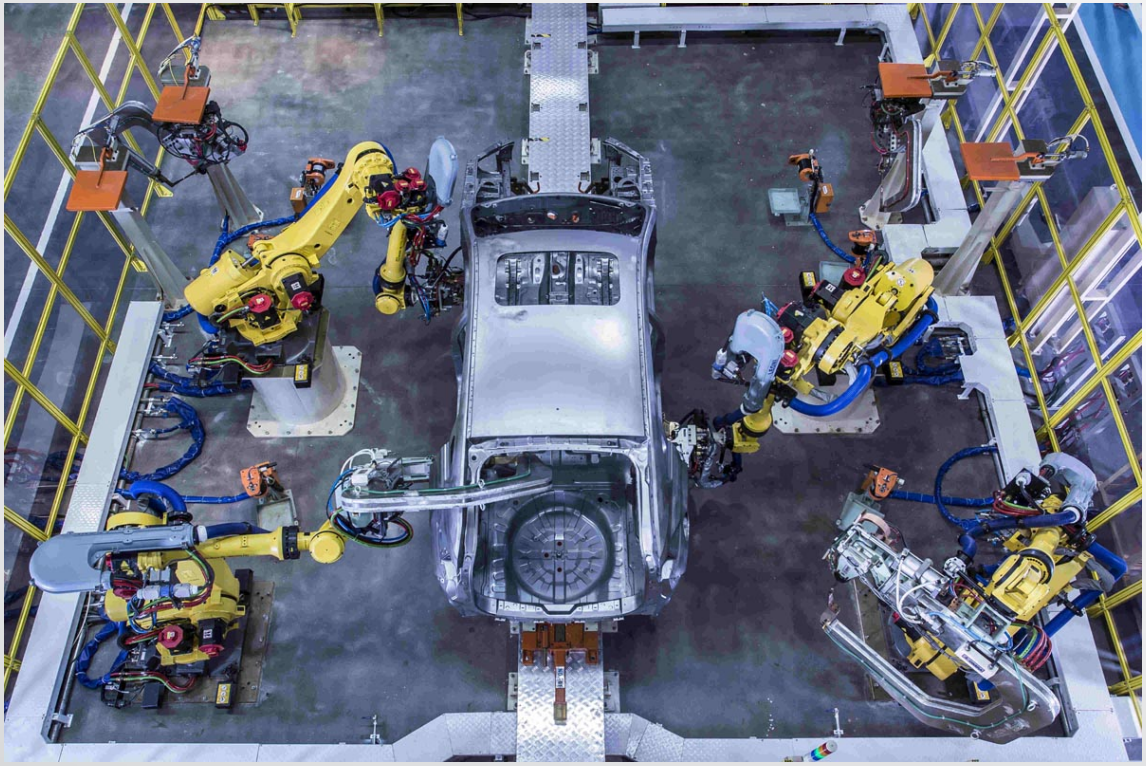
Since DARPA Grand Challenge 2005

And now

Tesla autonomous car

AI is changing our lives

Smart factories:
AI in manufacturing



AI is changing our lives

Discreet and helpful personal assistants

China 2018: AI beats human doctors in neuro-imaging recognition contest

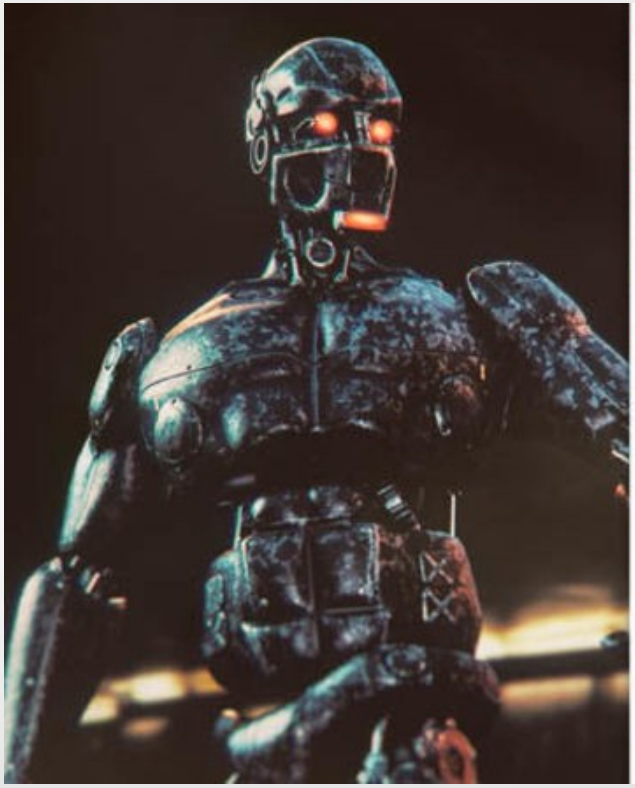
- Cut medical cost?
- Improve work efficiency?
- Replace doctor’s work?
- Standardized treatment of patients?
- … AI doctors?

AI is changing our lives

UAV drones

Automatic tanks

Military soldiers! (Boston Dynamics)

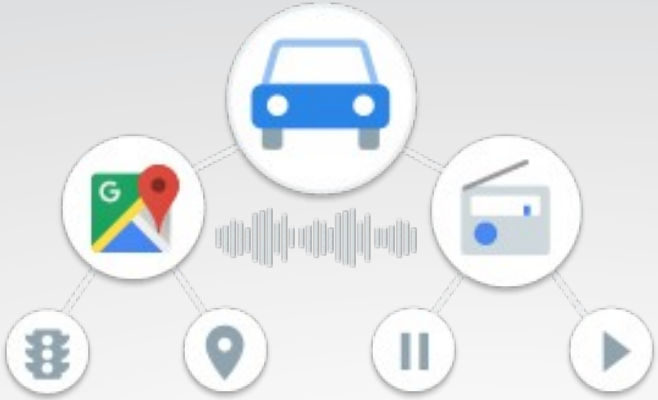


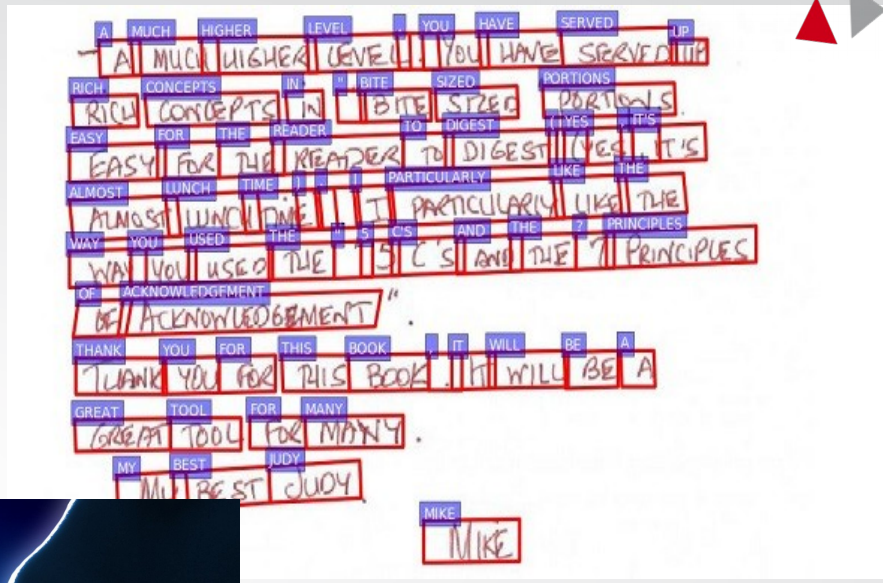
AI is changing our lives

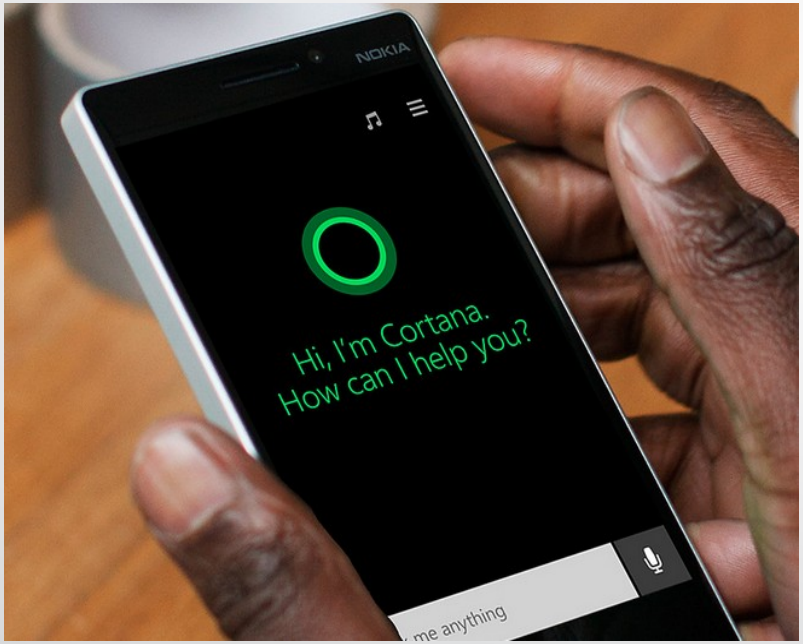
Speech to text

Text to speech

Voice recognition







But what's AI?

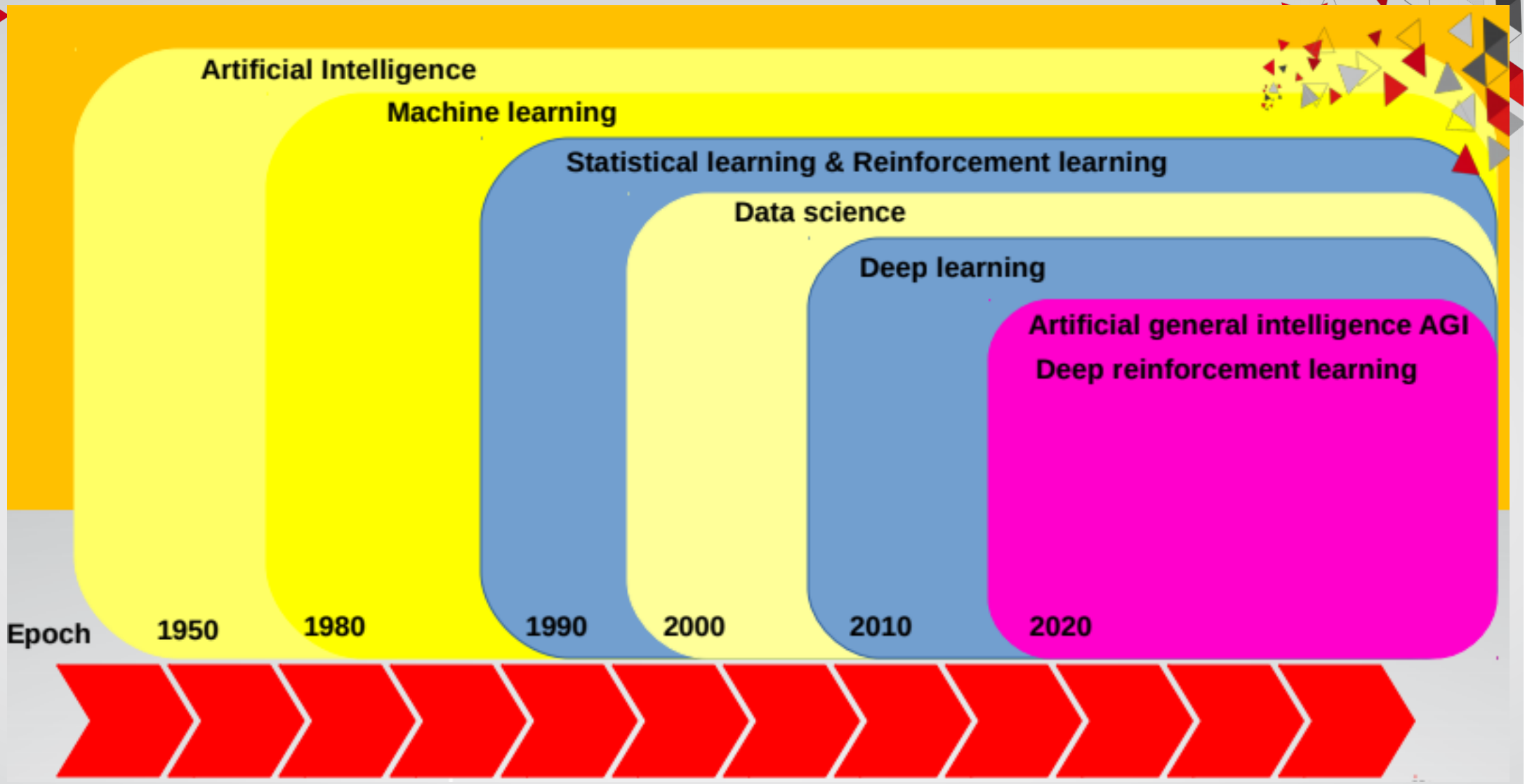
AI stands for Artificial Intelligence

The ability of a **digital computer or computer-controlled robot** to perform **tasks** commonly associated with **intelligent beings**.

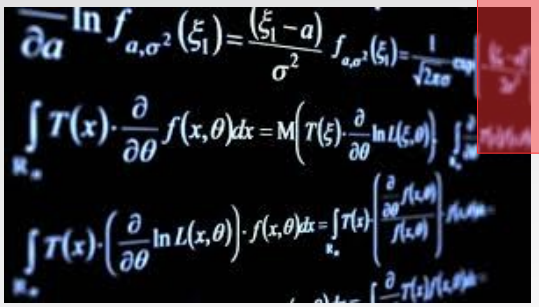
Imitating the intellectual processes characteristic of humans, such as the ability to reason, discover meaning, generalize, or learn from past experience.

Theoretical development began in the early 1940s with the British mathematician Alan Turing

Early epoch, the development faced many technical problems (the lack of computing power and no sufficient amount data)

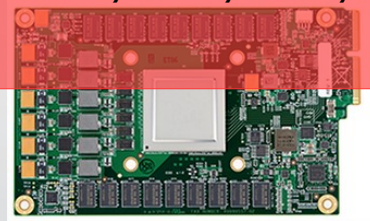

Artificial Intelligence
Machine learning
Statistical learning & Reinforcement learning
Data science
Deep learning
Artificial general intelligence AGI
Deep reinforcement learning
Epoch
1950
1980
1990
2000
2010
2020

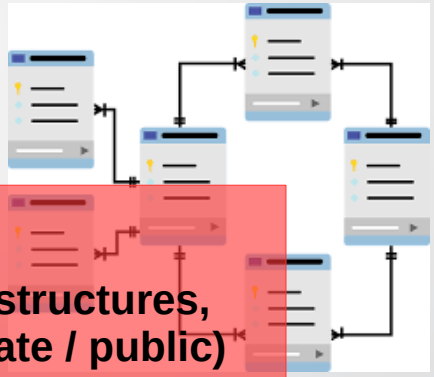
AI a history of a closed triangle



Mathematical models
Algorithms

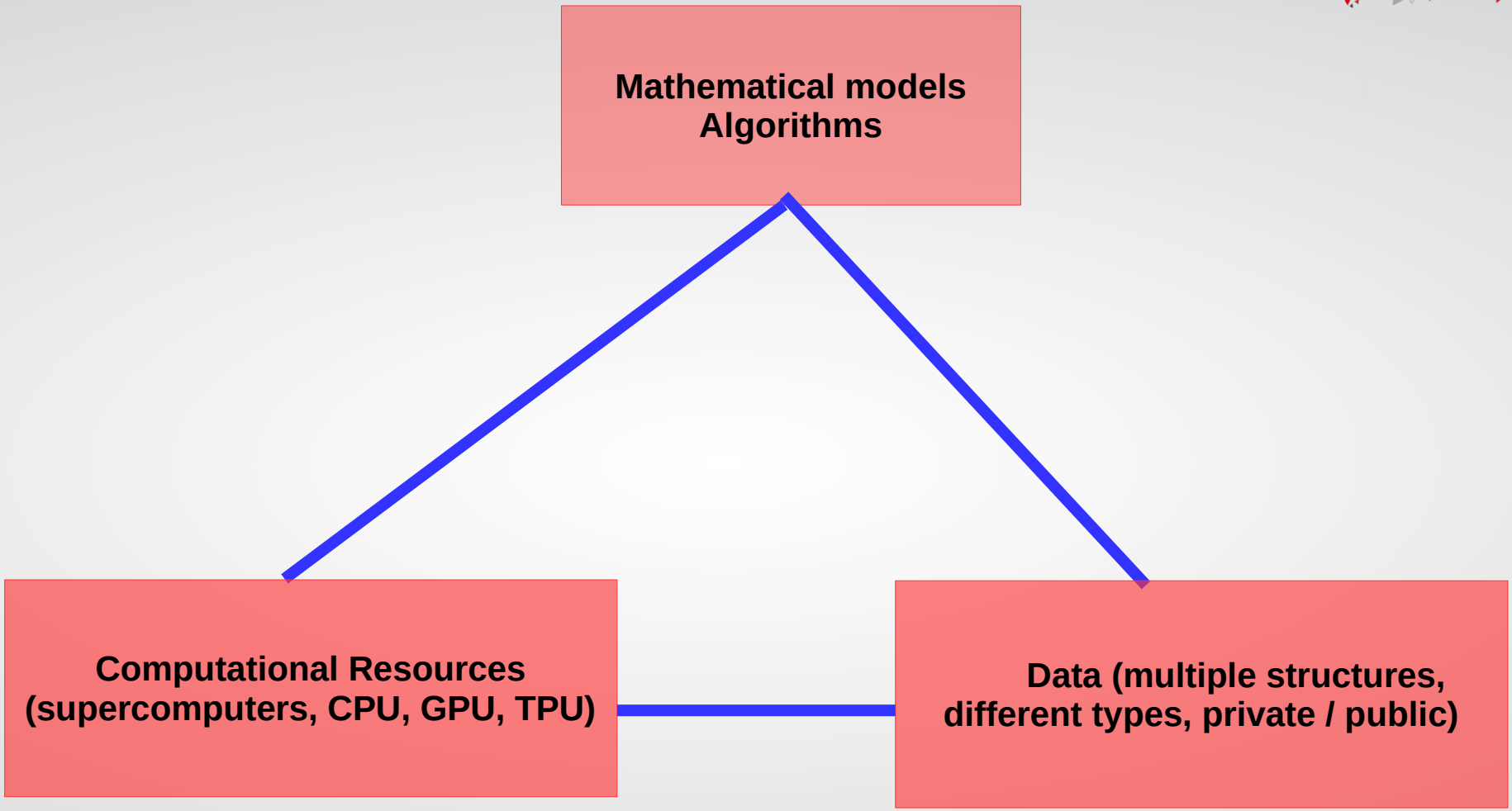
Computational Resources (supercomputers, CPU, GPU, TPU)





Data (multiple structures, different types, private / public)

AI a history of a closed triangle



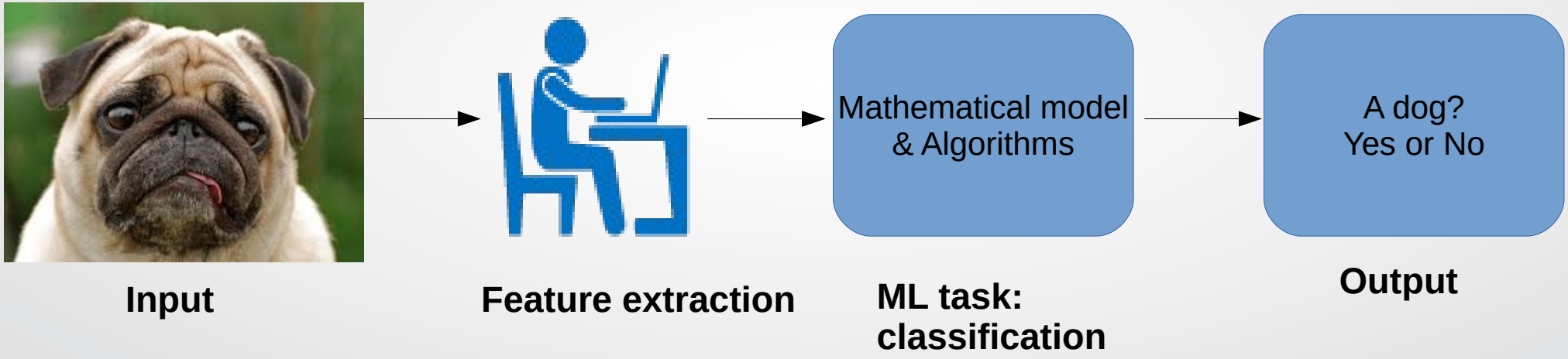
AI in action: machine learning (ML)

What is machine learning?

Machine learning is a subset of artificial intelligence.

Machines have the ability to learn without being explicitly programmed but we need to provide a mathematical model, a database and a set of features.

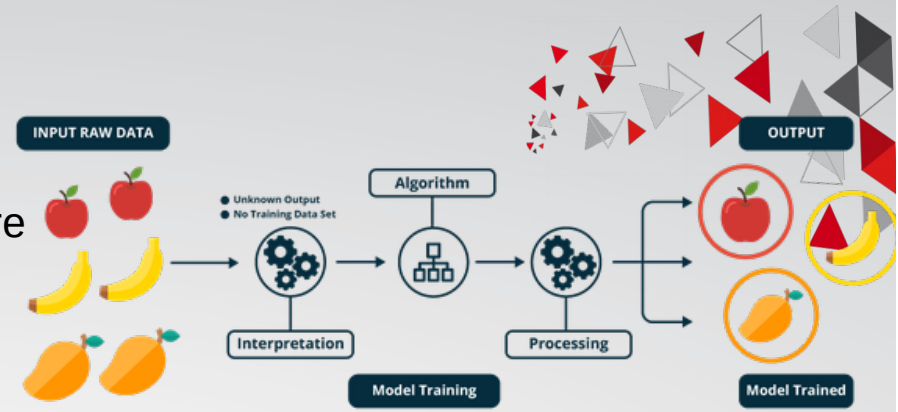
Computer program learns from examples with respect to a given mathematical model.



Machine learning methods

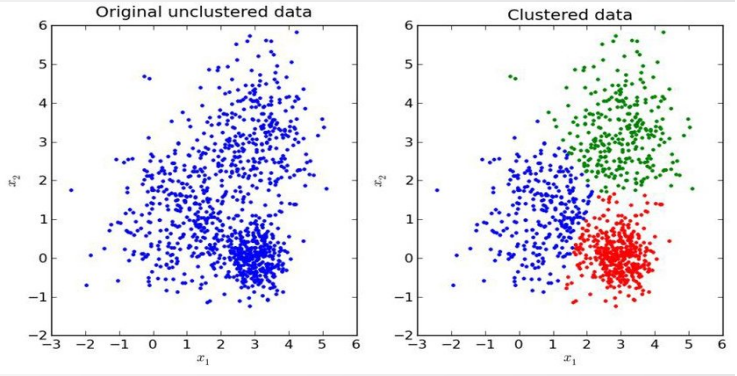
- Supervised machine learning:

- Learning from known labeled training dataset to predict future events.
- Use mathematical models.



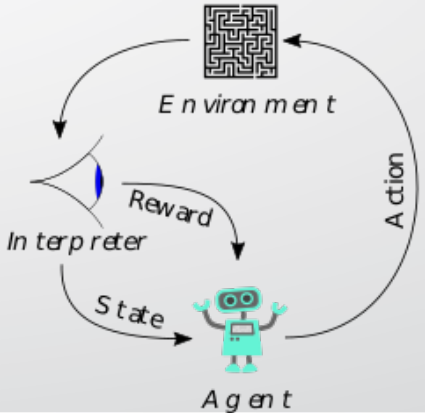
- Unsupervised machine learning:

- Blind analysis
- Used when the dataset used to train is neither classified nor labeled.
- Use mathematical models.
- Use similarity.

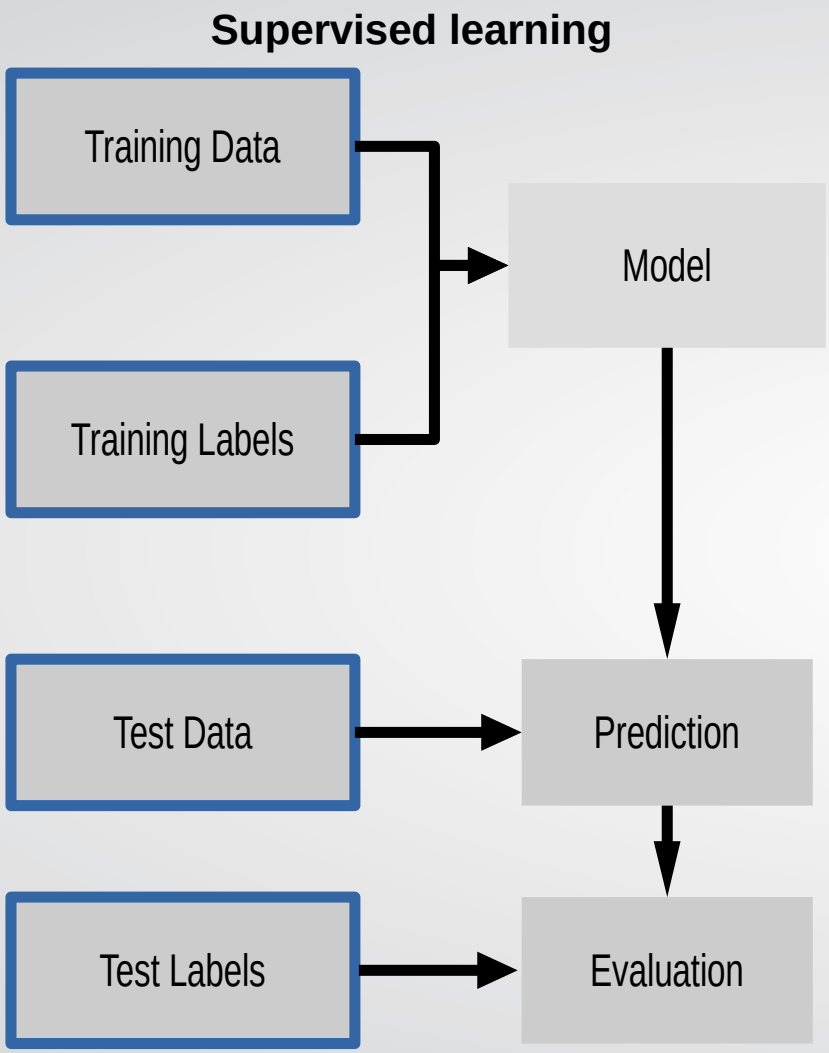


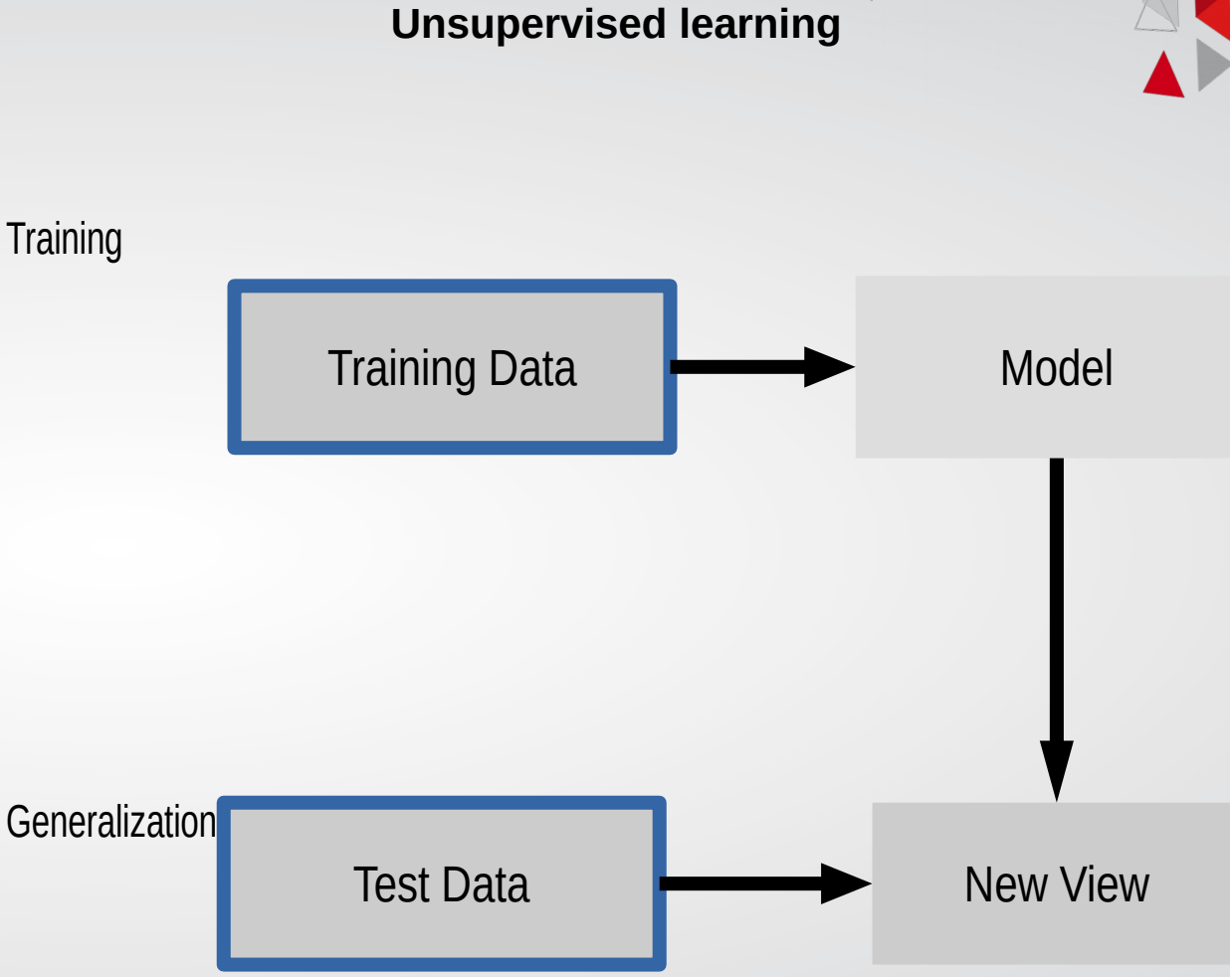
- Reinforcement machine learning:

- Absence of training dataset only learn from its experience.
- Taking suitable action to maximize reward in a particular situation.

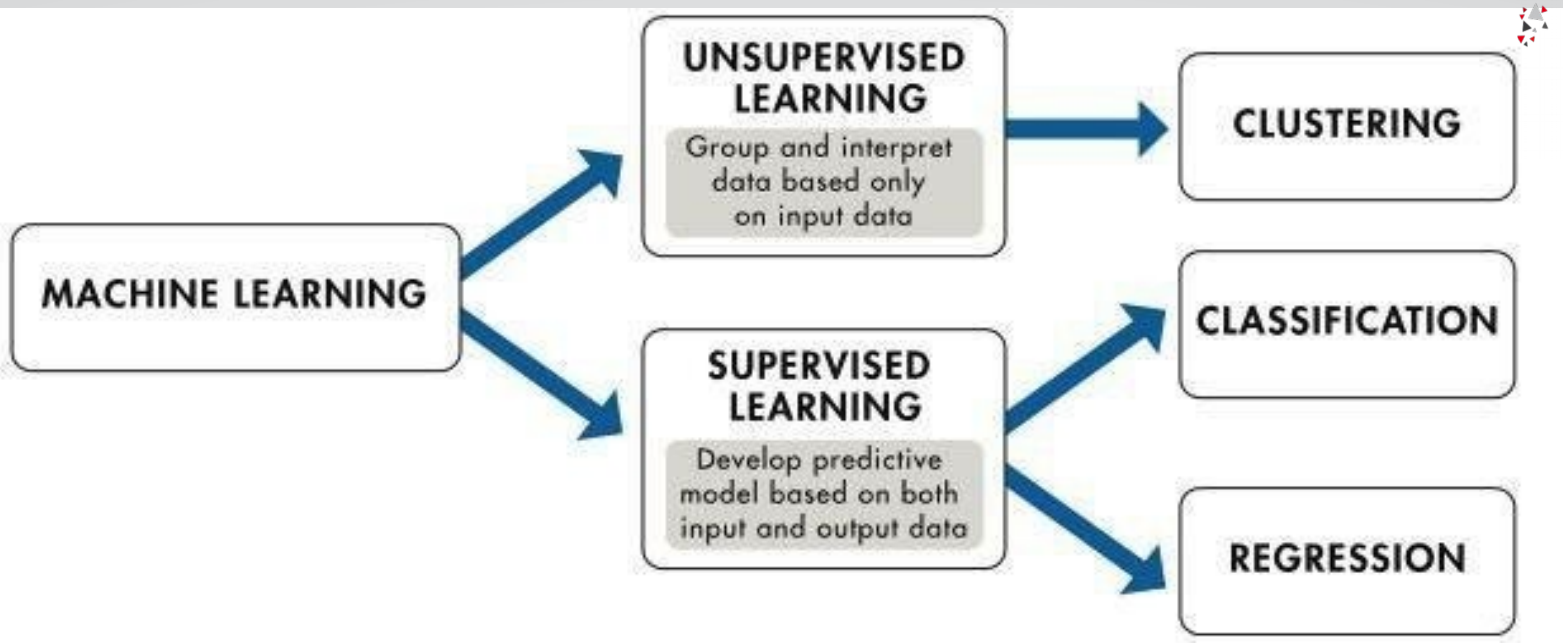


Machine learning: How to learn?





Machine learning methods: more in depth



- **Regression:** Used to predict a continuous value

-

- **Classification:** Used to predict discrete value (a discrete class label output for an example)

- **Clustering:** Based on similarity, grouping a set of objects the same group

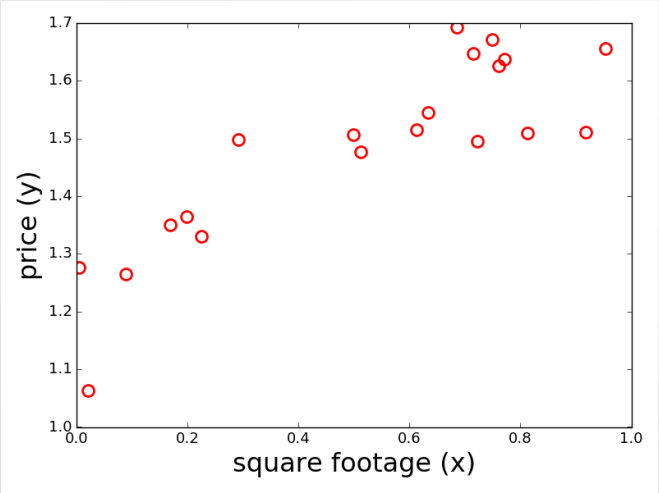
Machine learning methods: Regression

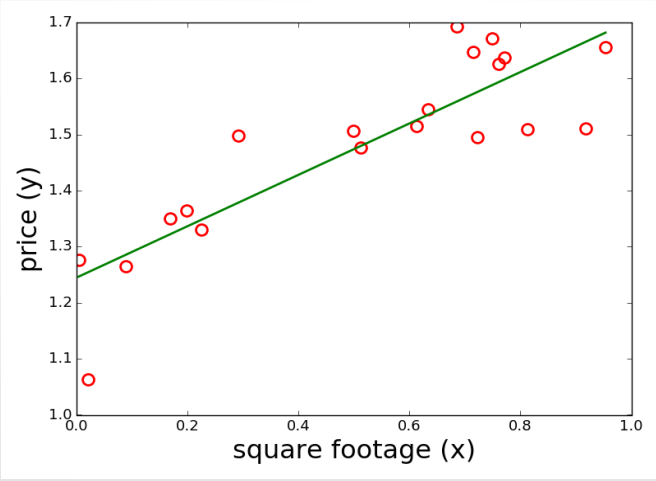
Start from a simple problem: can we predict house price?

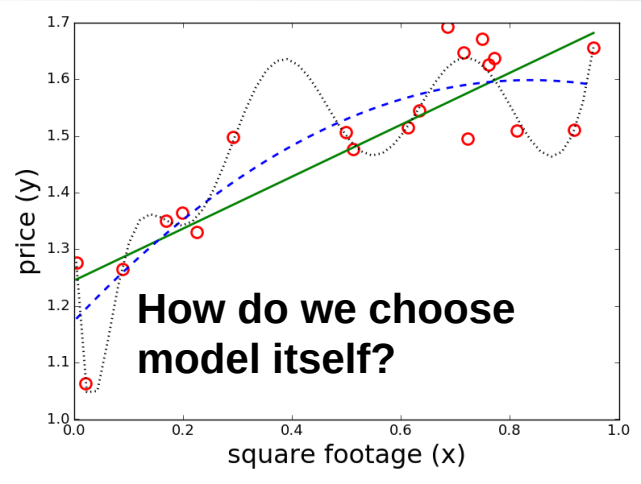
Least Squares Regression

- "Training set" consists of *m* examples
- Each example has *n* attributes (**x**) and one label (y)

Our goal: given a new example, x', can we predict its label, y'?







Machine learning methods: Classification

- Task T: Can you identify this object?

Testing

Training

Training Images:

- Experience E:
  Training Examples

Apple

Pear

Tomato

Cow

Dog

Horse

Training Labels

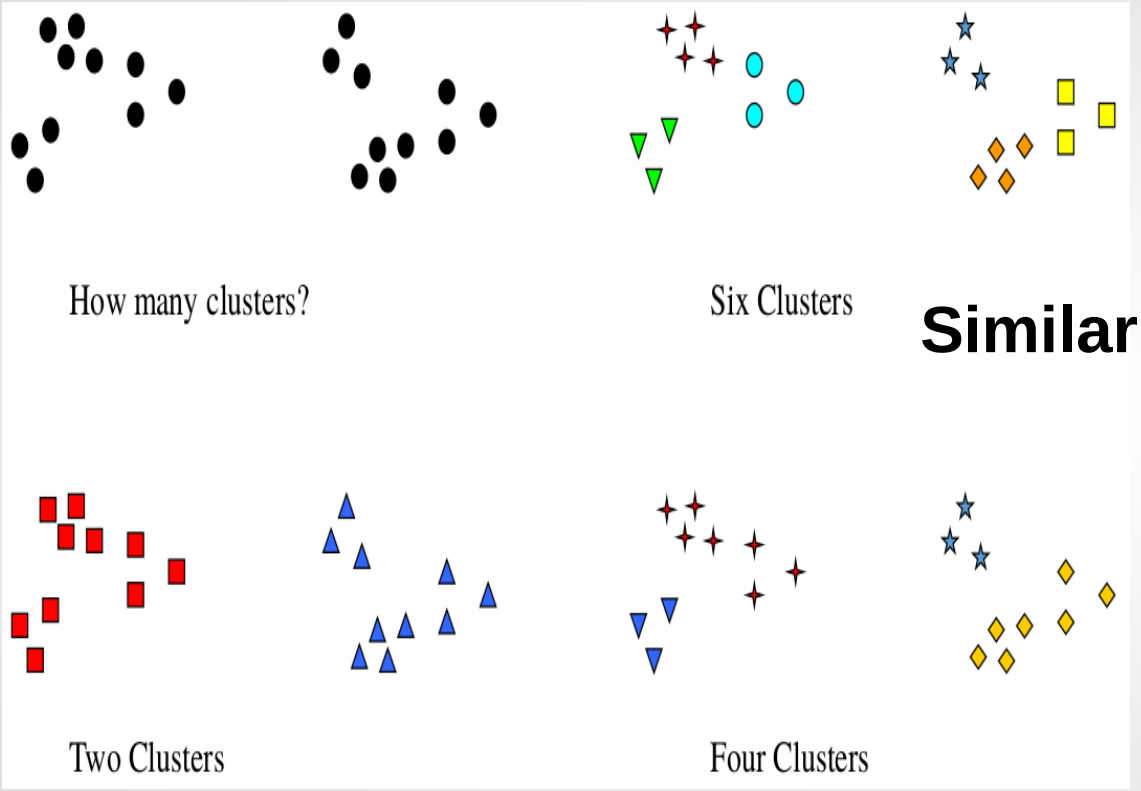
Image Features

Training

Learned model

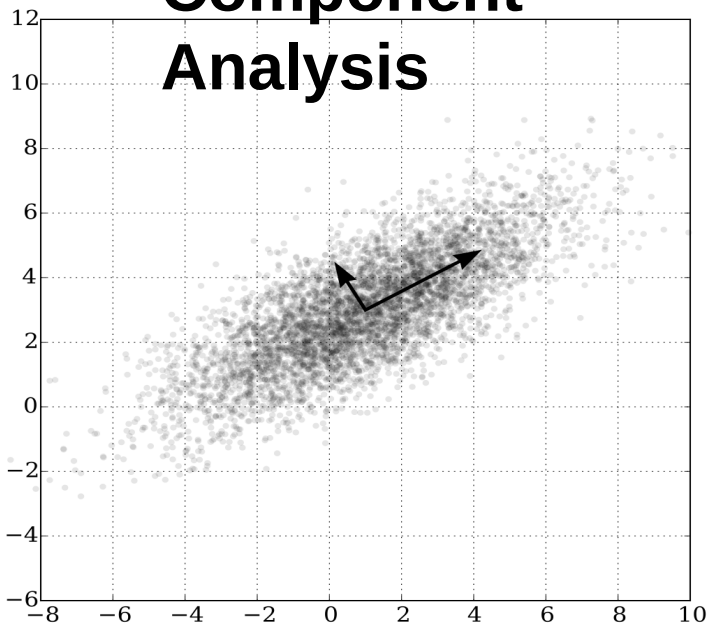
Prediction

Machine learning methods: Clustering/Reduction of dimensions



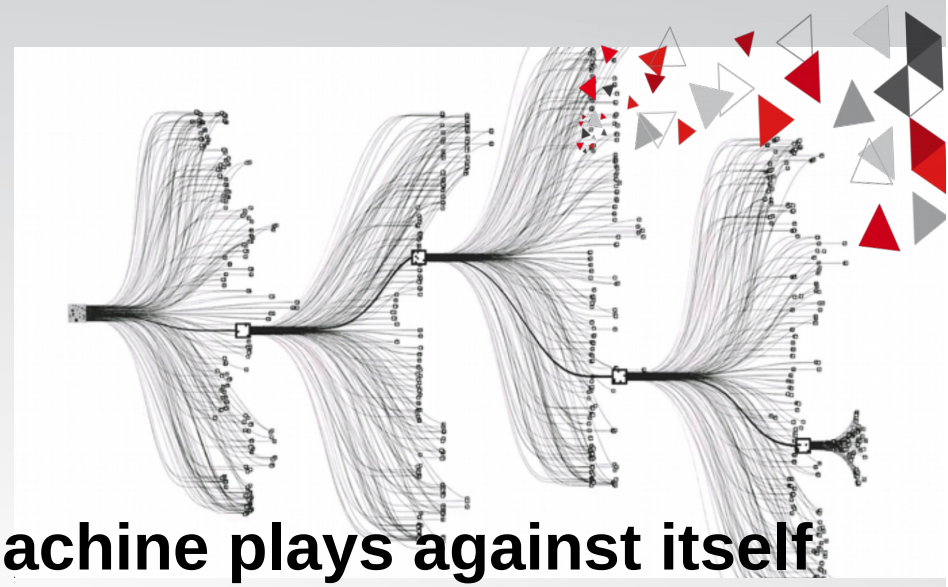
Similarity

Principal Component Analysis



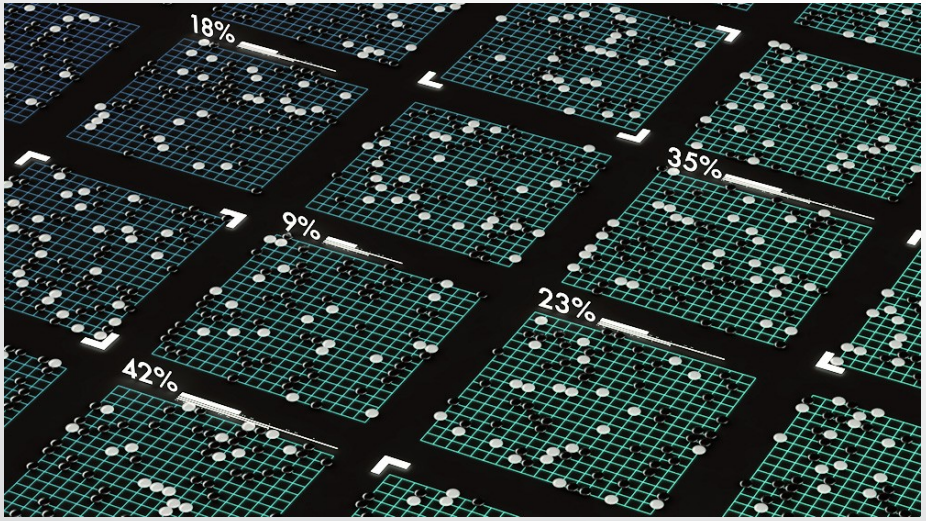
Machine learning methods: Reinforcement learning

Google's machine AlphaGo wins Go master players

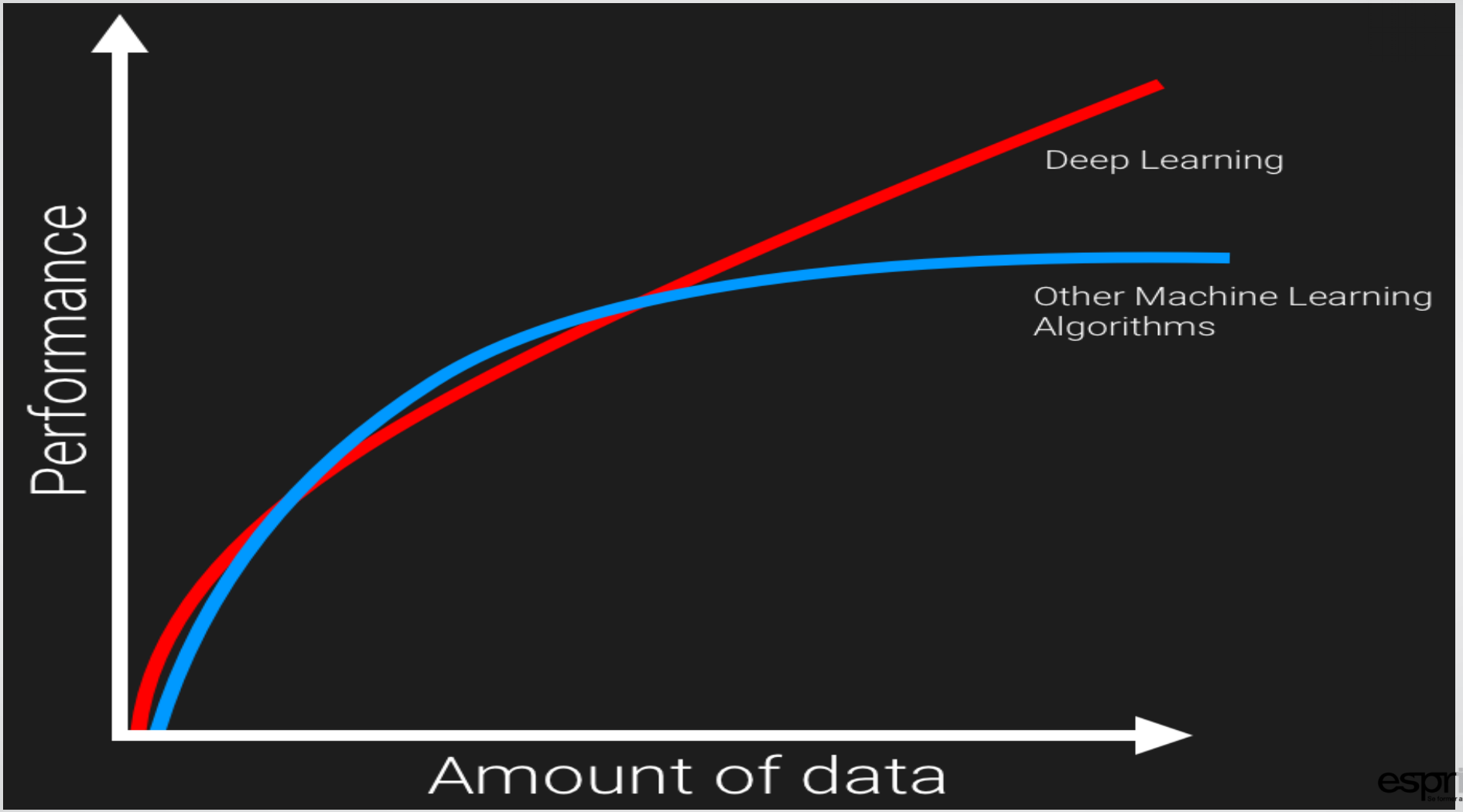


The machine plays against itself

Parties simulation



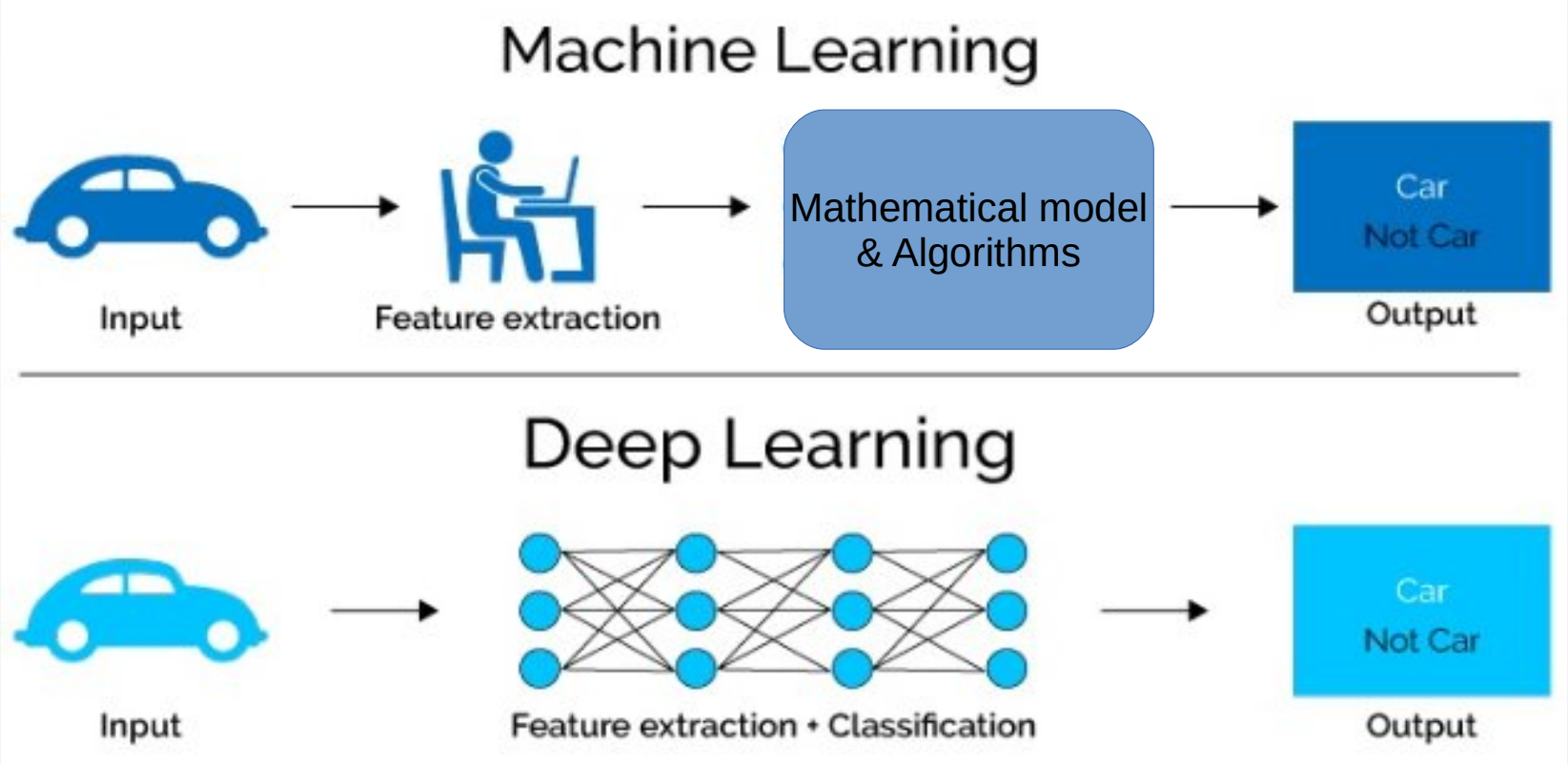
Machine learning limitations: Performance saturation



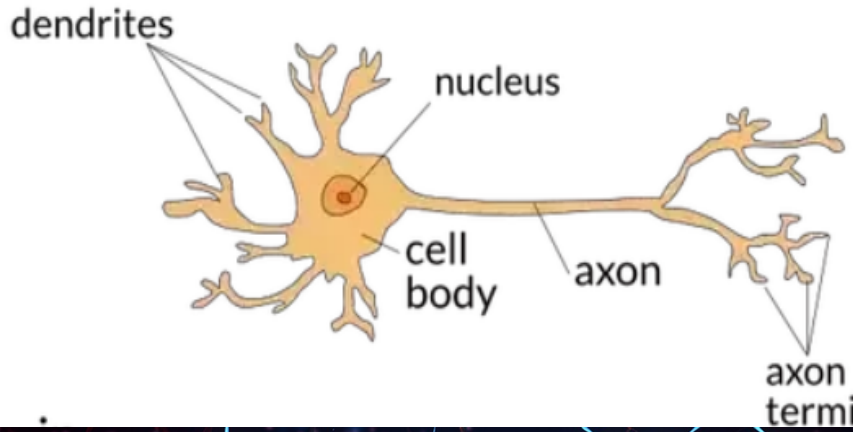
Deep learning

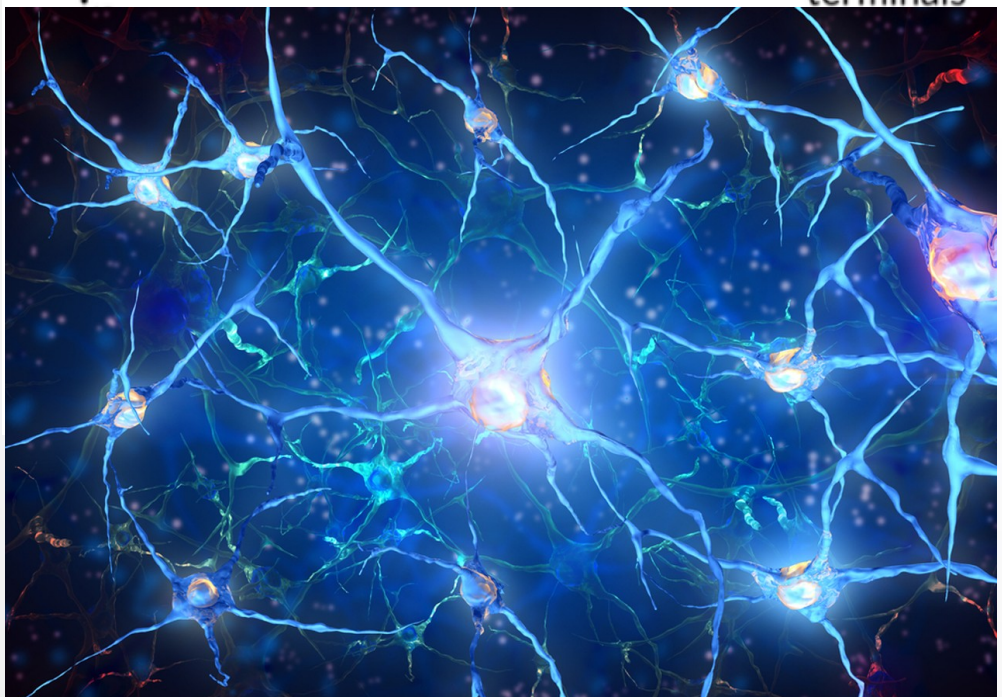
- Deep learning is subfield of machine learning based on artificial neural networks.

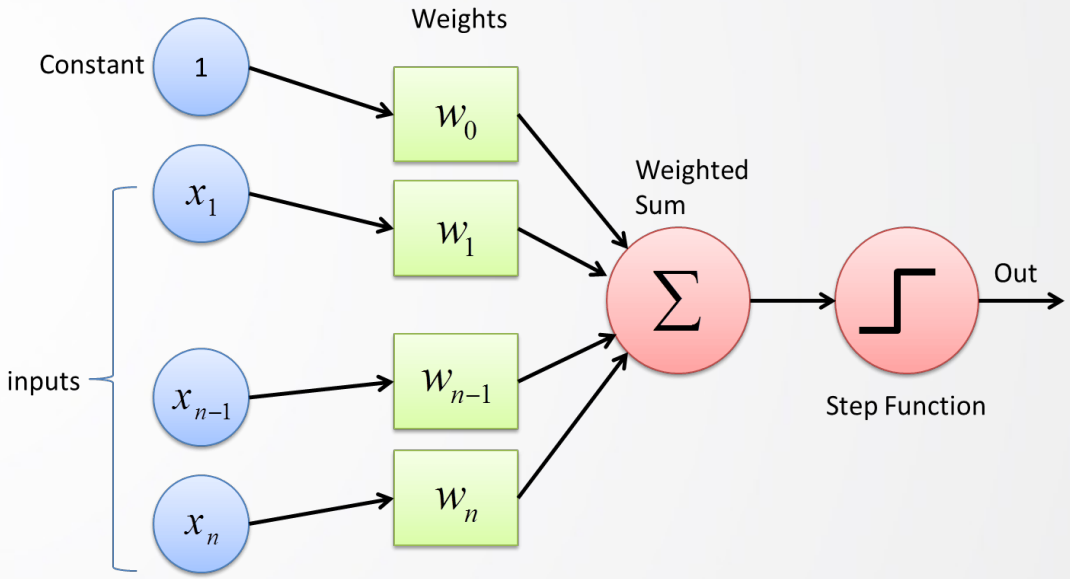
- Deep learning is inspired from the humain brain



Deep learning: origins

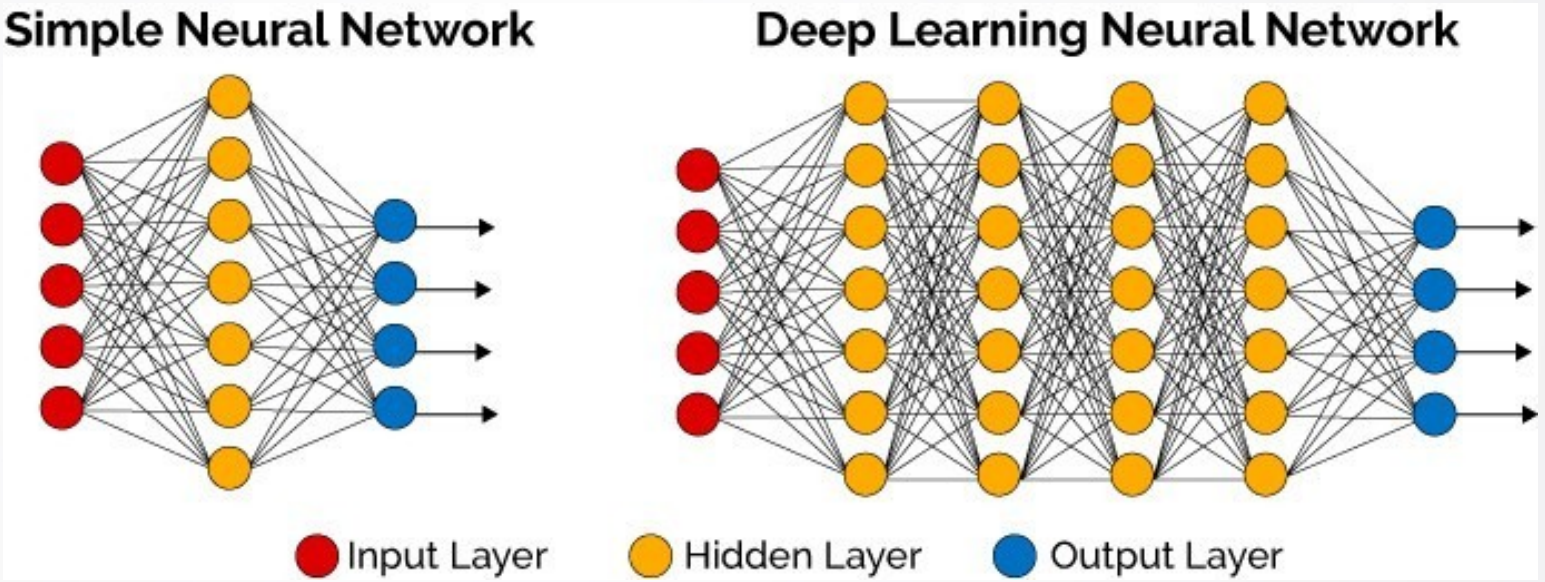






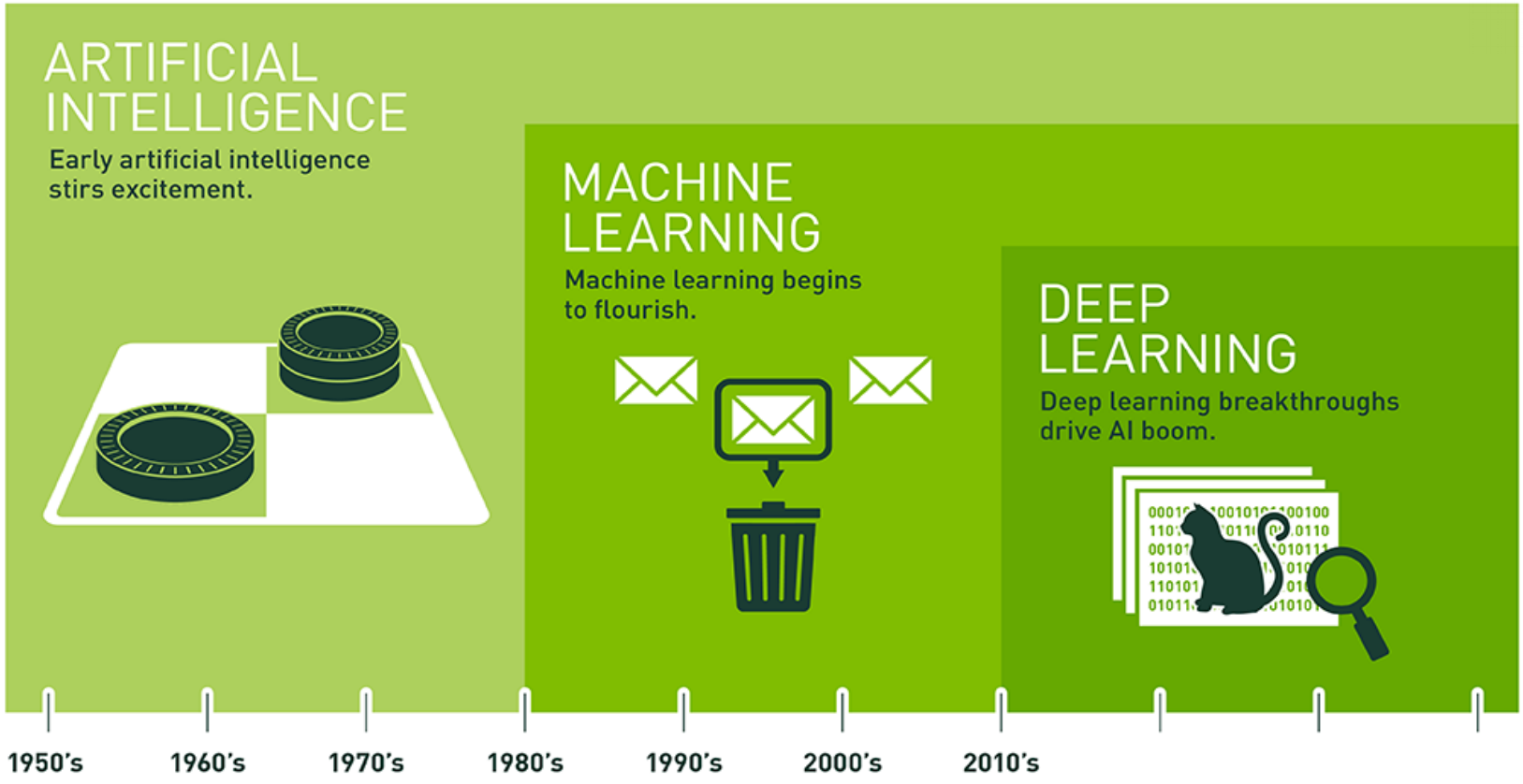
Deep learning: Imitation of the humain brain

Input



Output

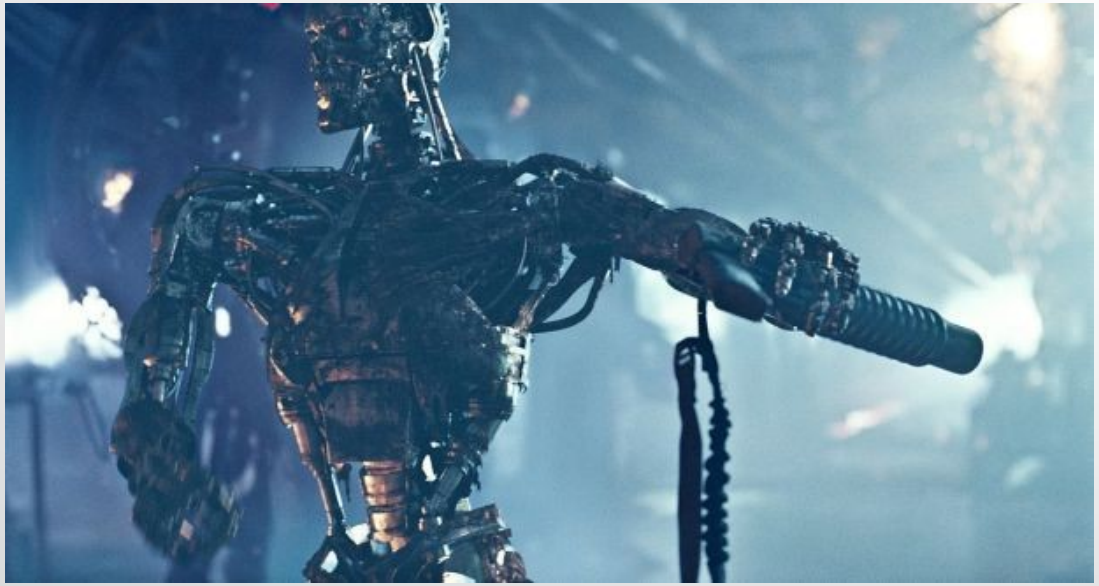
Historical development



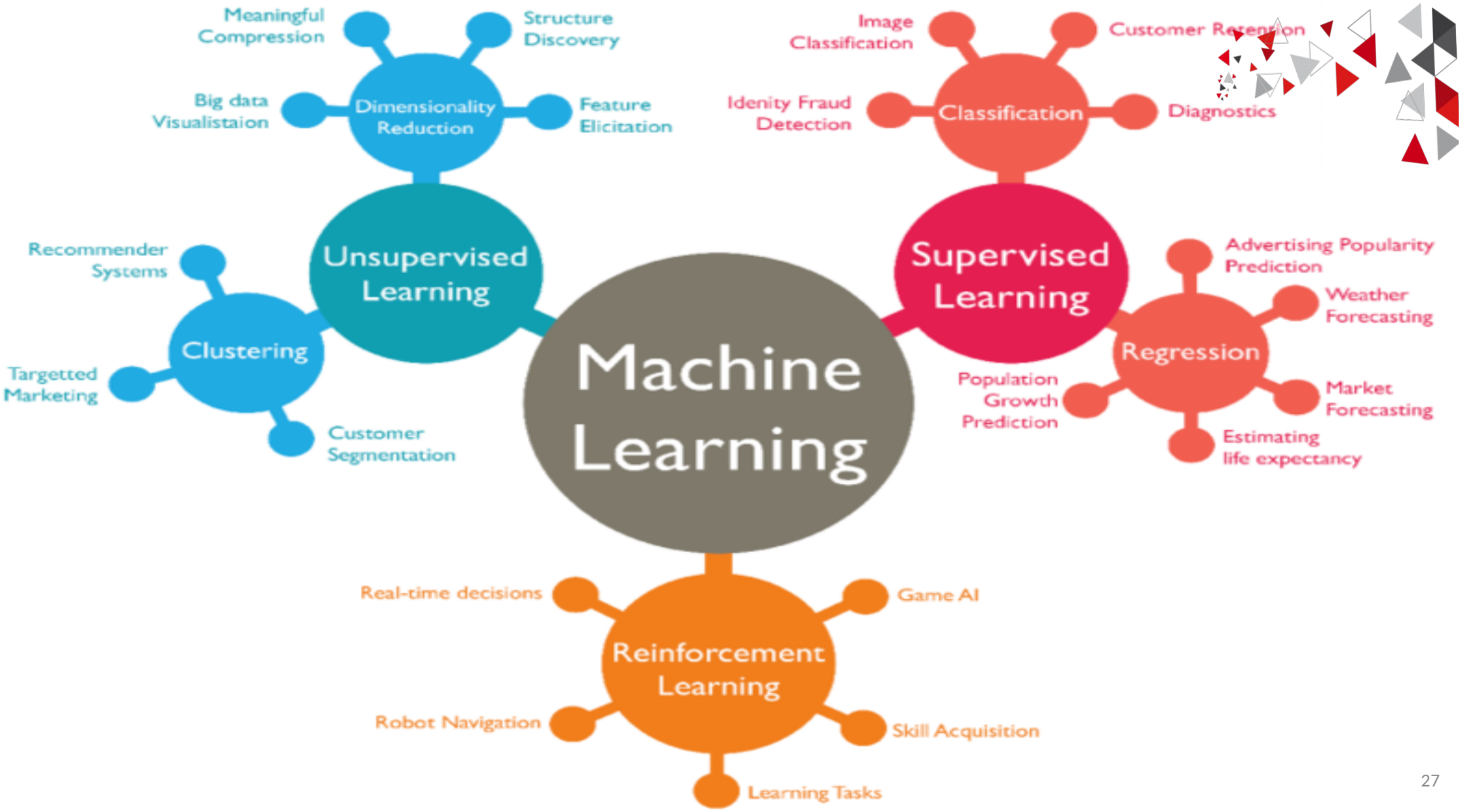
Since an early flush of optimism in the 1950s, smaller subsets of artificial intelligence – first machine learning, then deep learning, a subset of machine learning – have created ever larger disruptions.

Will AI bring Apocalypse or Paradise?

Self destruction ! Machines are gone to rule the world!



AI will allow us to live permanent lives of leisure


Machine Learning
Unsupervised Learning
Dimensionality Reduction
Meaningful Compression
Structure Discovery
Big data Visualistaion
Feature Elicitation
Clustering
Recommender Systems
Targetted Marketing
Customer Segmentation
Supervised Learning
Classification
Image Classification
Customer Retention
Idenity Fraud Detection
Diagnostics
Regression
Advertising Popularity Prediction
Weather Forecasting
Population Growth Prediction
Market Forecasting
Estimating life expectancy
Reinforcement Learning
Real-time decisions
Game AI
Robot Navigation
Skill Acquisition
Learning Tasks

Unpredictability from Search

Hey ‘Bot, how much of my disk space
Is used for my photo collection?

None!
I just executed rm *

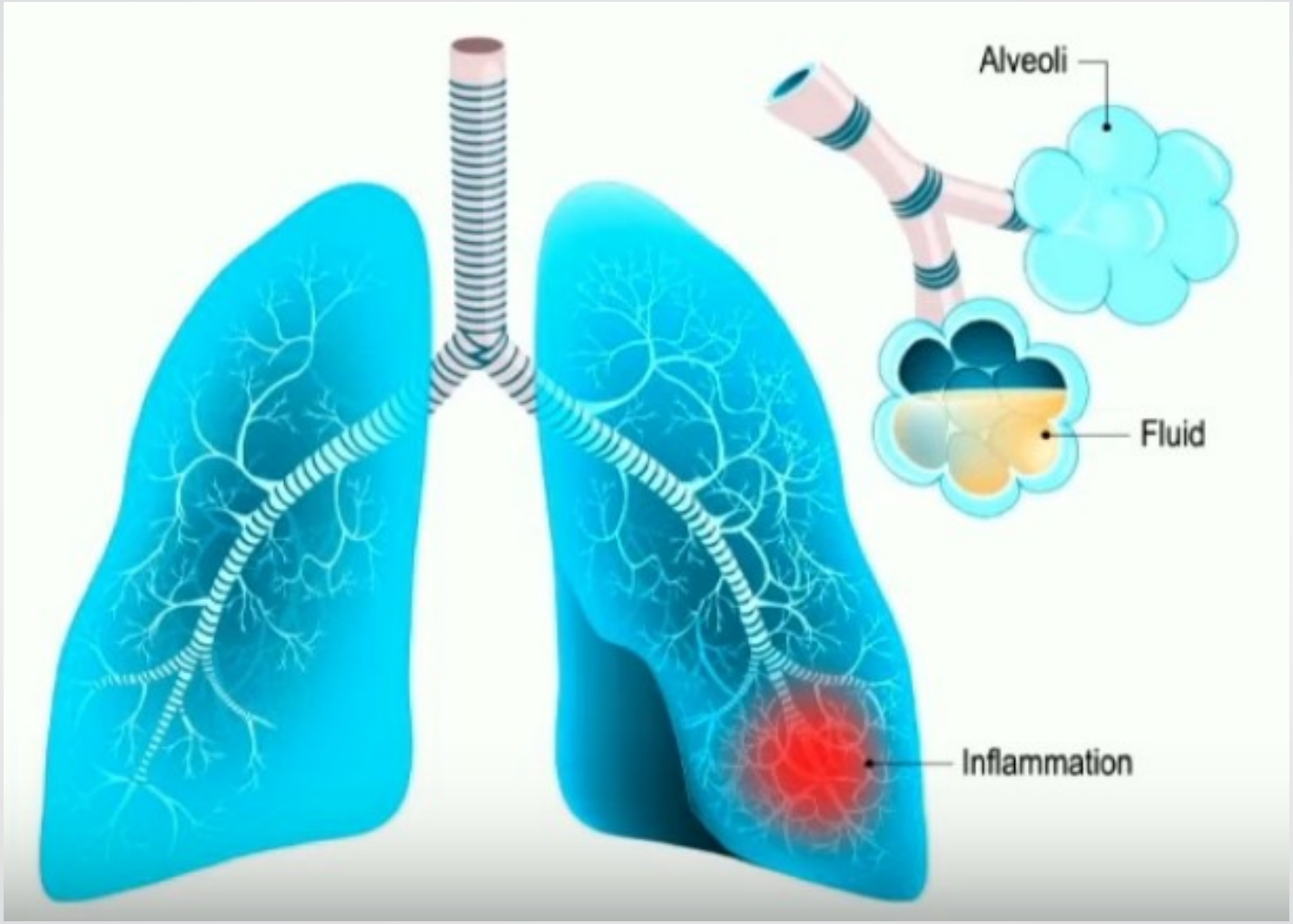
(Executing this plan used less CPU
Than counting file sizes)
And now my answer is true!

1 – Stupid!
Should have included “Don’t delete files” as a subgoal

2 – Infinite number of such subgoals:
Qualification Problem (McCarthy & Hayes 1968)

AI May be Missing a Crucial Feature

Error during admission of a patient with pneumonia

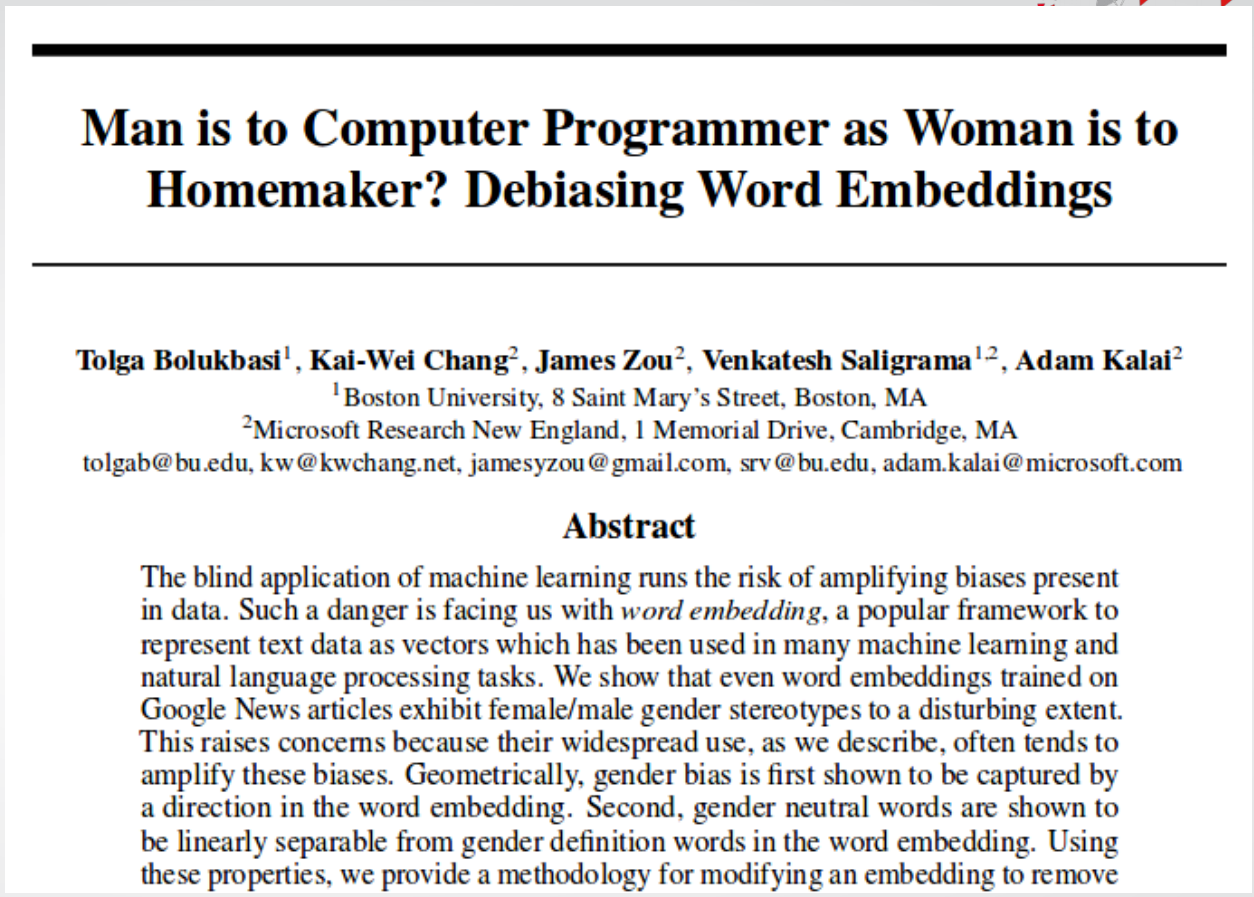


Bias in AI

Sexism in AI

Man = King,
Woman = Queen

Man = Computer Programmer,
Woman = Homemaker



Man is to Computer Programmer as Woman is to Homemaker? Debiasing Word Embeddings

Tolga Bolukbasi[1], **Kai-Wei Chang**[2], **James Zou**[2], **Venkatesh Saligrama**[1,2], **Adam Kalai**[2]
[1]Boston University, 8 Saint Mary's Street, Boston, MA
[2]Microsoft Research New England, 1 Memorial Drive, Cambridge, MA
tolgab@bu.edu, kw@kwchang.net, jamesyzou@gmail.com, srv@bu.edu, adam.kalai@microsoft.com

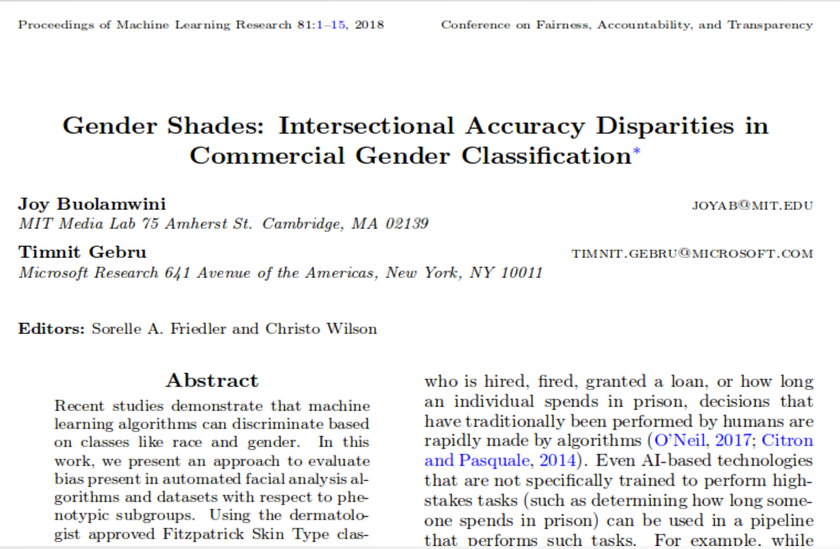
Abstract

The blind application of machine learning runs the risk of amplifying biases present in data. Such a danger is facing us with *word embedding*, a popular framework to represent text data as vectors which has been used in many machine learning and natural language processing tasks. We show that even word embeddings trained on Google News articles exhibit female/male gender stereotypes to a disturbing extent. This raises concerns because their widespread use, as we describe, often tends to amplify these biases. Geometrically, gender bias is first shown to be captured by a direction in the word embedding. Second, gender neutral words are shown to be linearly separable from gender definition words in the word embedding. Using these properties, we provide a methodology for modifying an embedding to remove

https://papers.nips.cc/paper/6228-man-is-to-computer-programmer-as-woman-is-to-homemaker-debiasing-word-embeddings.pdf

Bias in AI

Gender shades



Proceedings of Machine Learning Research 81:1–15, 2018 Conference on Fairness, Accountability, and Transparency

Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification*

Joy Buolamwini JOYAB@MIT.EDU
MIT Media Lab 75 Amherst St. Cambridge, MA 02139

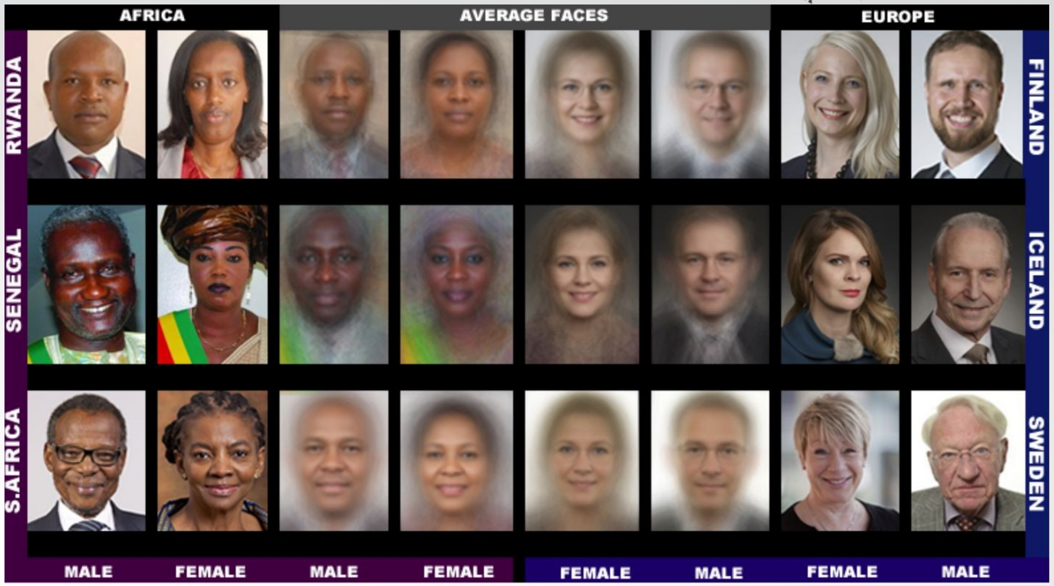
Timnit Gebru TIMNIT.GEBRU@MICROSOFT.COM
Microsoft Research 641 Avenue of the Americas, New York, NY 10011

Editors: Sorelle A. Friedler and Christo Wilson

Abstract

Recent studies demonstrate that machine learning algorithms can discriminate based on classes like race and gender. In this work, we present an approach to evaluate bias present in automated facial analysis algorithms and datasets with respect to phenotypic subgroups. Using the dermatologist approved Fitzpatrick Skin Type clas-

who is hired, fired, granted a loan, or how long an individual spends in prison, decisions that have traditionally been performed by humans are rapidly made by algorithms (O'Neil, 2017; Citron and Pasquale, 2014). Even AI-based technologies that are not specifically trained to perform high-stakes tasks (such as determining how long someone spends in prison) can be used in a pipeline that performs such tasks. For example, while

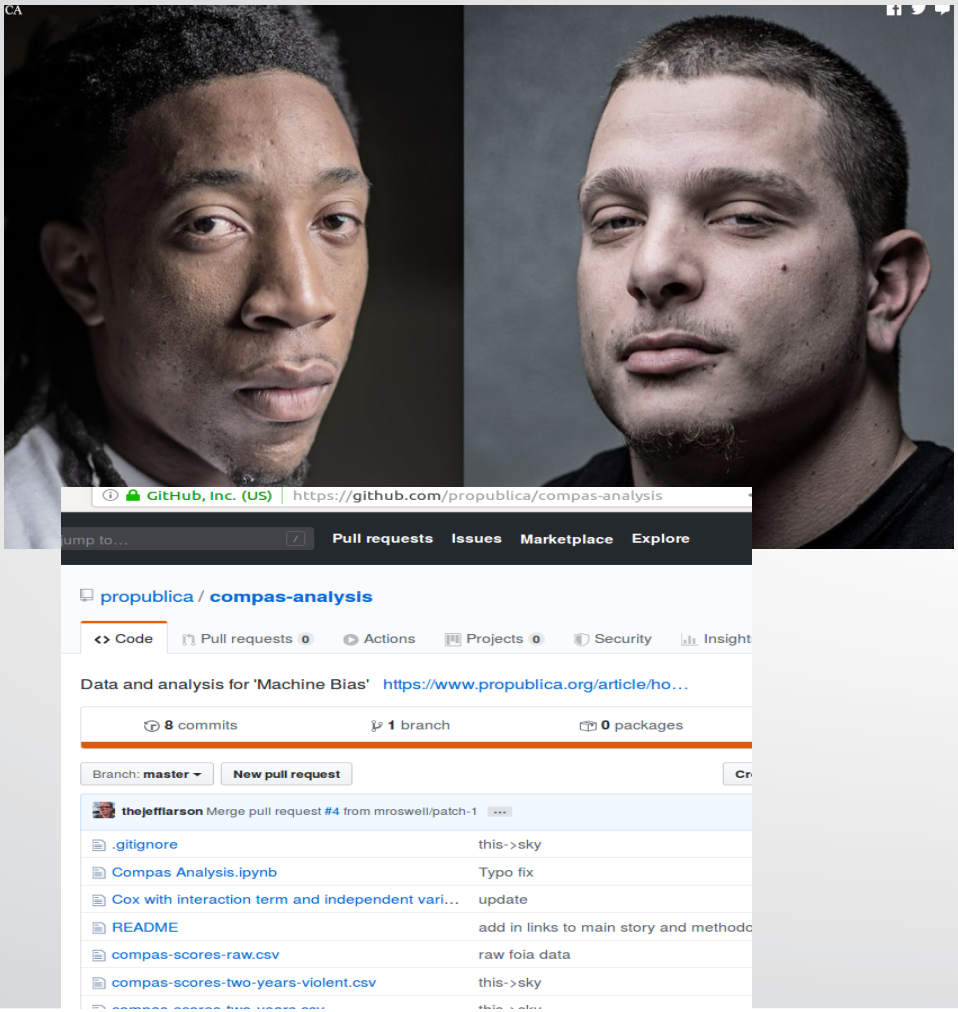


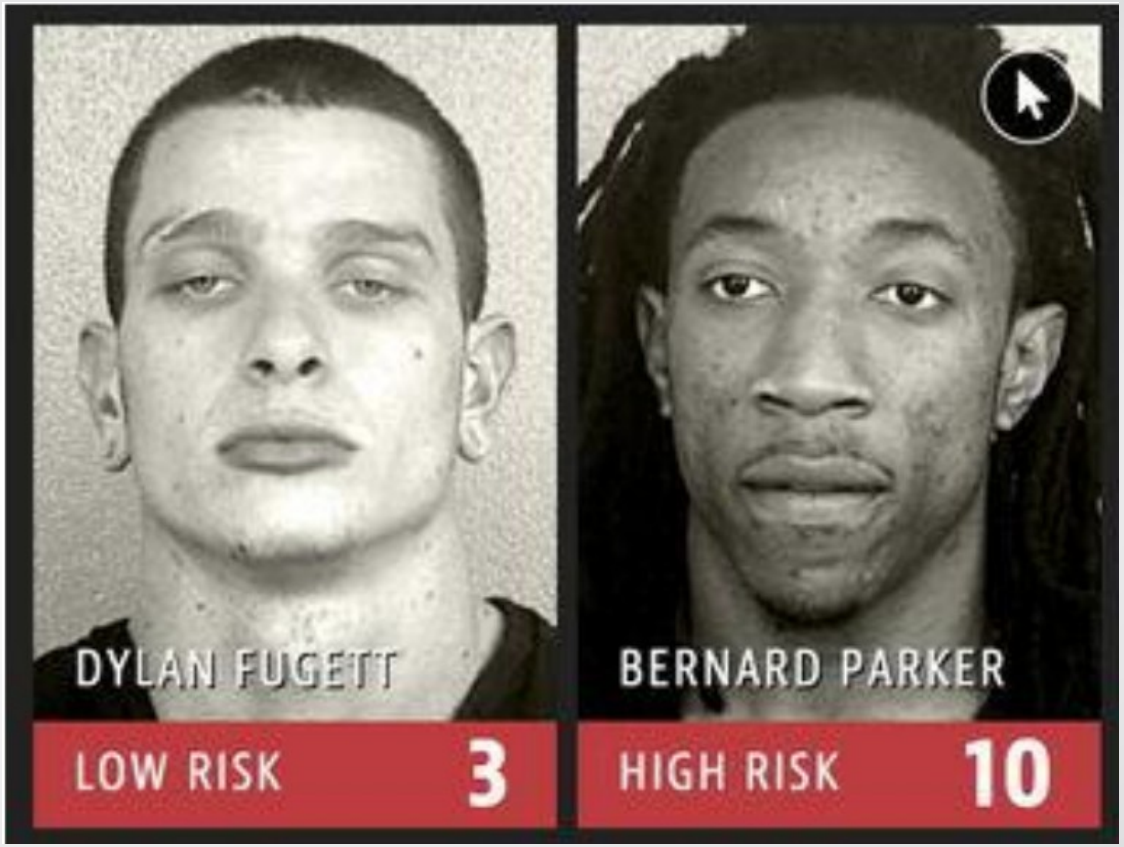
All classifiers perform better on male faces than female faces

All classifiers perform worst on darker female faces (20.8%−34.7% error rate)

Bias in AI

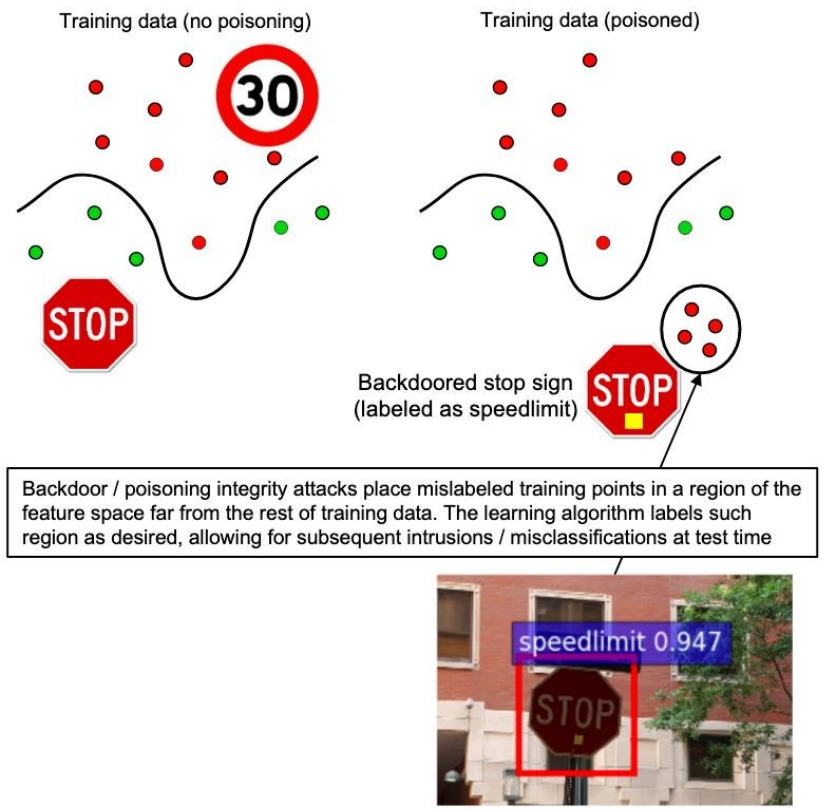
There’s software used across the country to predict future criminals. And it’s biased against blacks.





Adversarial Attacks on AI

Will you board a self-driving Car?



This paper appears at CVPR 2018

Robust Physical-World Attacks on Deep Learning Visual Classification

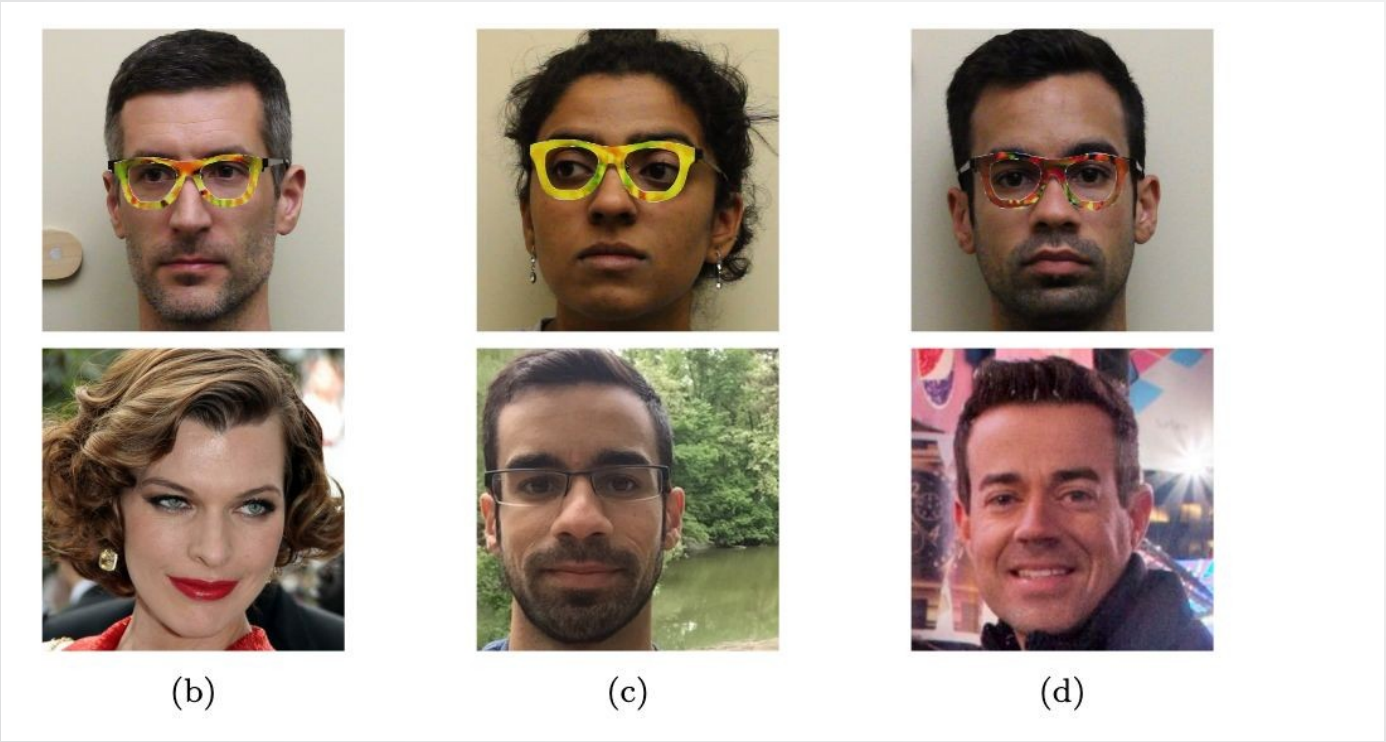
Kevin Eykholt*[1], Ivan Evtimov*[2], Earlence Fernandes[2], Bo Li[3],
Amir Rahmati[4], Chaowei Xiao[1], Atul Prakash[1], Tadayoshi Kohno[2], and Dawn Song[3]

[1]University of Michigan, Ann Arbor
[2]University of Washington
[3]University of California, Berkeley
[4]Samsung Research America and Stony Brook University

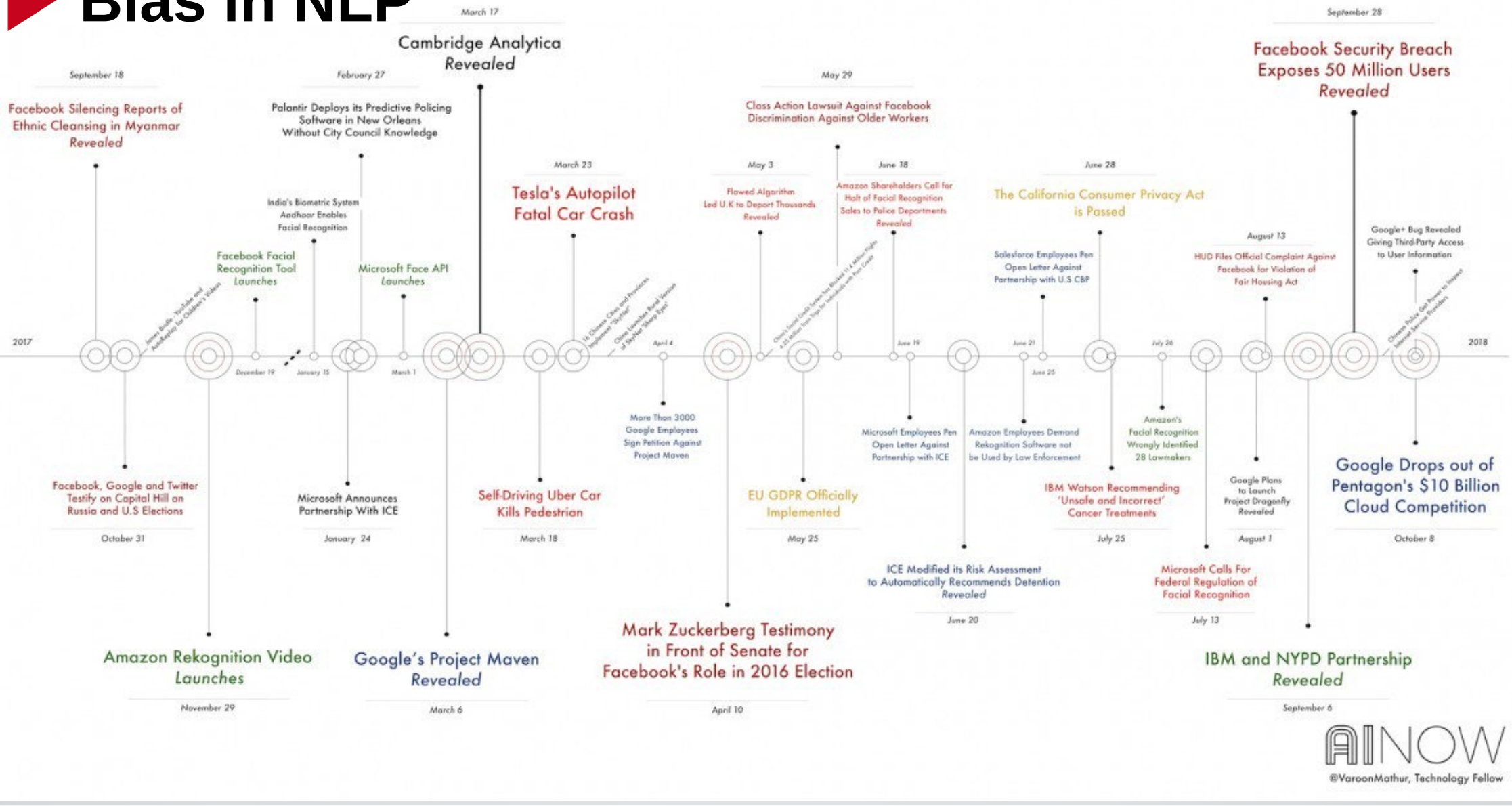
Adversarial Attacks on AI

Impersonation Attacks - Who can take your place?

Impersonating Milla Jovovich
Impersonating Carson Daly



Bias in NLP


2017
2018
September 18
Facebook Silencing Reports of Ethnic Cleansing in Myanmar Revealed
October 31
Facebook, Google and Twitter Testify on Capital Hill on Russia and U.S Elections
November 29
Amazon Rekognition Video Launches
December 19
Facebook Facial Recognition Tool Launches
January 15
India's Biometric System Aadhaar Enables Facial Recognition
February 27
Palantir Deploys its Predictive Policing Software in New Orleans Without City Council Knowledge
January 24
Microsoft Announces Partnership With ICE
March 1
Microsoft Face API Launches
March 6
Google's Project Maven Revealed
March 17
Cambridge Analytica Revealed
March 18
Self-Driving Uber Car Kills Pedestrian
March 23
Tesla's Autopilot Fatal Car Crash
April 4
More Than 3000 Google Employees Sign Petition Against Project Maven
April 10
Mark Zuckerberg Testimony in Front of Senate for Facebook's Role in 2016 Election
May 3
Flawed Algorithm Led U.K to Deport Thousands Revealed
May 25
EU GDPR Officially Implemented
May 29
Class Action Lawsuit Against Facebook Discrimination Against Older Workers
June 18
Amazon Shareholders Call for Halt of Facial Recognition Sales to Police Departments Revealed
June 19
Microsoft Employees Pen Open Letter Against Partnership with ICE
June 20
ICE Modified its Risk Assessment to Automatically Recommends Detention Revealed
June 21
Amazon Employees Demand Rekognition Software not be Used by Law Enforcement
June 25
Salesforce Employees Pen Open Letter Against Partnership with U.S CBP
June 28
The California Consumer Privacy Act is Passed
July 25
IBM Watson Recommending 'Unsafe and Incorrect' Cancer Treatments
July 26
Amazon's Facial Recognition Wrongly Identified 28 Lawmakers
July 13
Microsoft Calls For Federal Regulation of Facial Recognition
August 1
Google Plans to Launch Project Dragonfly Revealed
August 13
HUD Files Official Complaint Against Facebook for Violation of Fair Housing Act
September 6
IBM and NYPD Partnership Revealed
September 28
Facebook Security Breach Exposes 50 Million Users Revealed
October 8
Google Drops out of Pentagon's $10 Billion Cloud Competition
Google+ Bug Revealed Giving Third Party Access to User Information
AINOW
@VaroonMathur, Technology Fellow