AI, Machine Learning, Data Science

Alberto Paccanaro
EMAp – FGV

www.paccanarolab.org

Artificial Intelligence

Goal

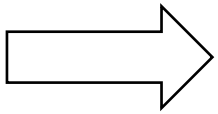
Build systems that execute tasks that require “intelligence”

1. Ability to reason (draw logical conclusions)
2. Ability to learn

1. Logic-based AI
(a.k.a. classical AI, rule-based AI)

Build system that are able to perform logical deductions

knowledge base
+
logic rules



deductions

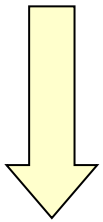
- Resolution principle (Robinson, 1965)
- Prolog (1972)

Strongly symbolic

Logic based AI

All men are mortal

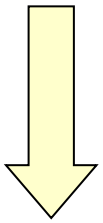
Socrates is a man



Socrates is mortal

All gruks are pake

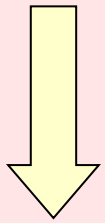
Grax is a gruk



Grax is pake

$\forall x \in X, P(x) = T$

$a \in X$



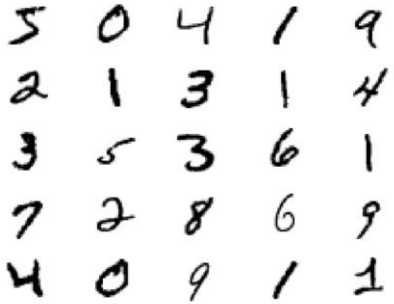
$P(a) = T$

Success stories

- **Expert Systems**
  - **Mycin –** identified bacteria responsible for an infection and suggested antibiotics and specific dosage for patients (~ 600 rules, 1970s)
  - **XCON (eXpert CONfigurer) –** chose components to obtain a system specified by the user for DEC Vax systems (~2500 rules, 1980s)

- **Methods for automated verification**
  - hardware
  - software

Did not work for other tasks



Machine Learning
(a.k.a. machine learning-based AI)

- **Build a mathematical model that depends on parameters**
- **Use the data to adjust the parameters in order to optimize performance (learning)**

Starting point: data

Strongly numerical

Several disciplines have been doing that …

Statistics

Applied mathematics

Engineering

Computer Science

Cognitive sciences

Regression
Matrix decompositions
Discriminant analysis
…

Neural Networks (mid-80s)
Kernel methods, SVM (mid-90s)
Random Forest(late-90s)
Graphical models (00s)
Latent Variable models (00s)
…

Machine Learning

Simple models
Few parameters

Complex models
Many parameters

Geoffrey Hinton

- Robust to noise in the data
- Ideal to integrate data from different sources

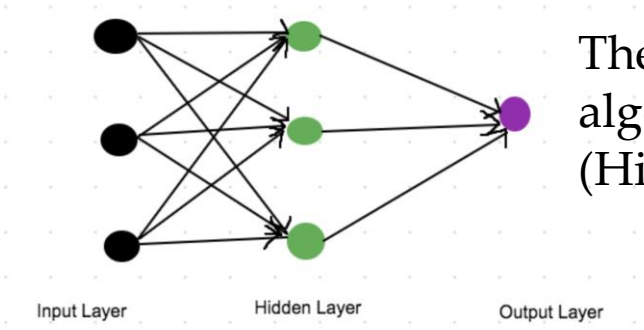
DATA IS THE CURRENCY

… what happened next … ?

The
ARRIVAL OF THE DATA

+ BIGGER COMPUTERS

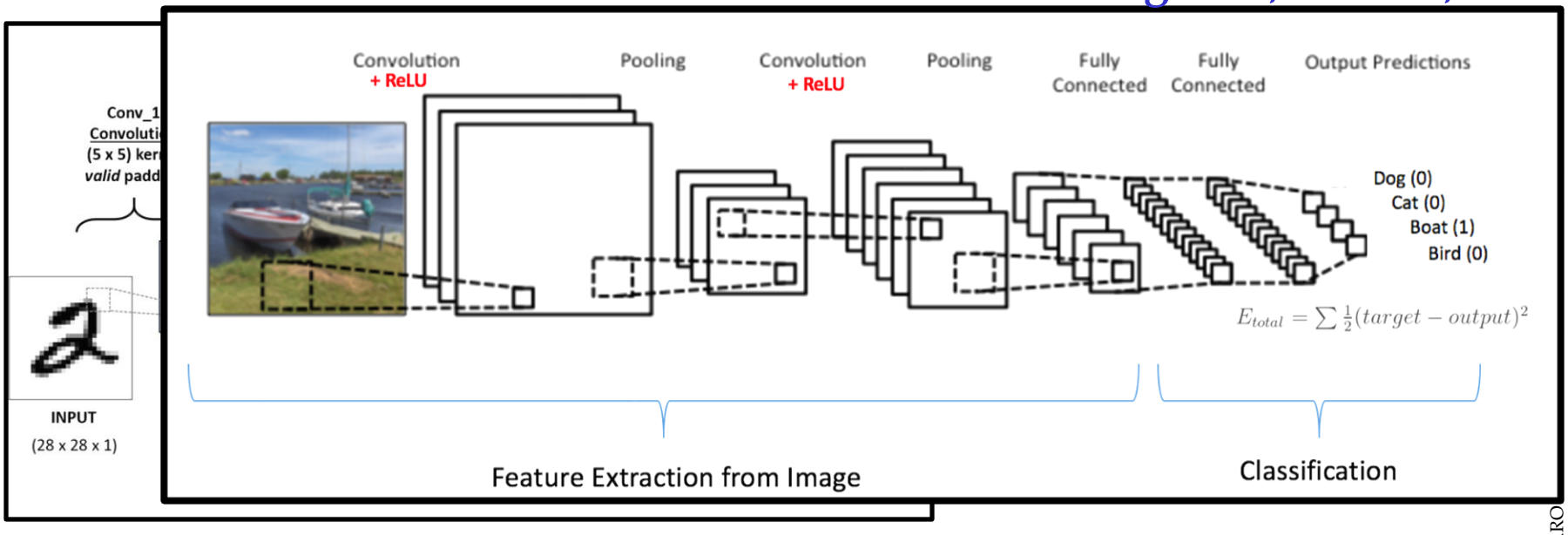
The evolution of Neural Networks



The backpropagation algorithm
(Hinton, 1986)

Deep Learning

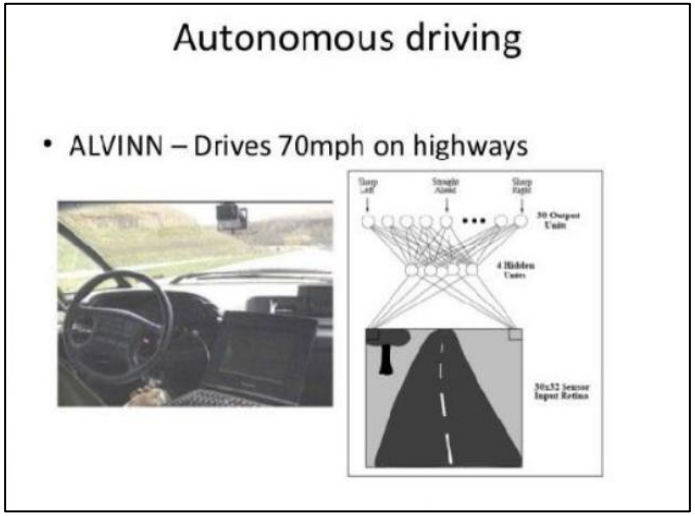
ImageNet, Hinton, 2013



GPT-3 (latest language model) contains 175 billion parameters required several thousand petaflop/s-days of compute

Autonomous vehicles

Uber ATG 2015



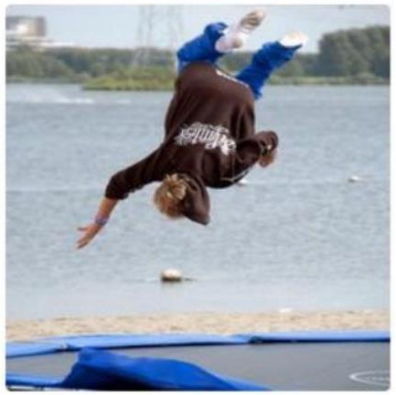
ALVINN CMU, 1989

Computer Vision
automatic image captioning

"man in black shirt is playing guitar."

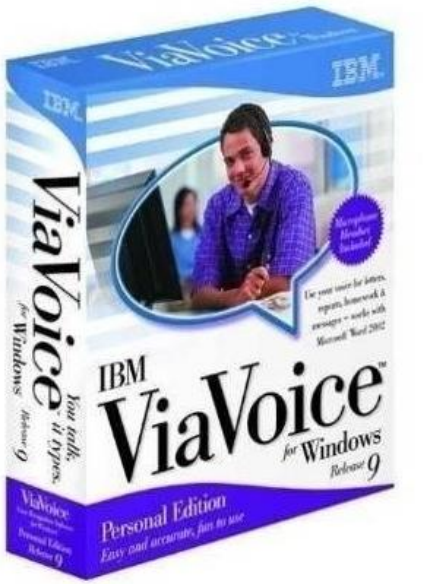
"construction worker in orange safety vest is working on road."

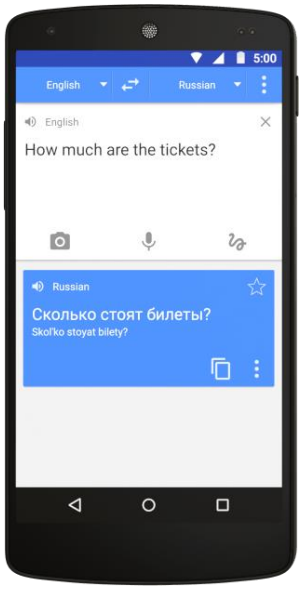
"two young girls are playing with legos toy."



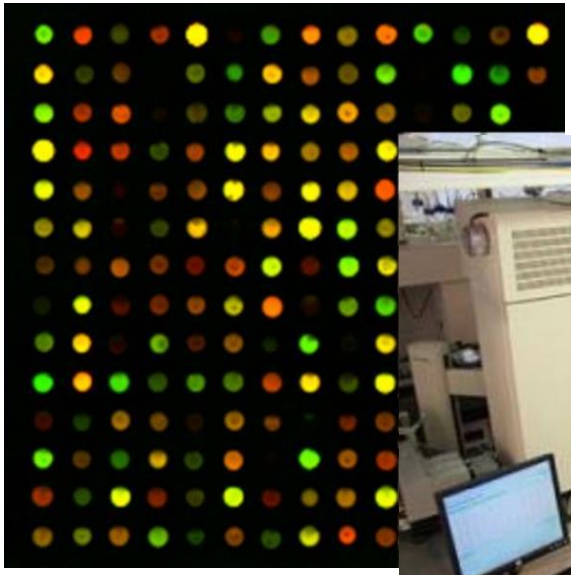
"boy is doing backflip on wakeboard."

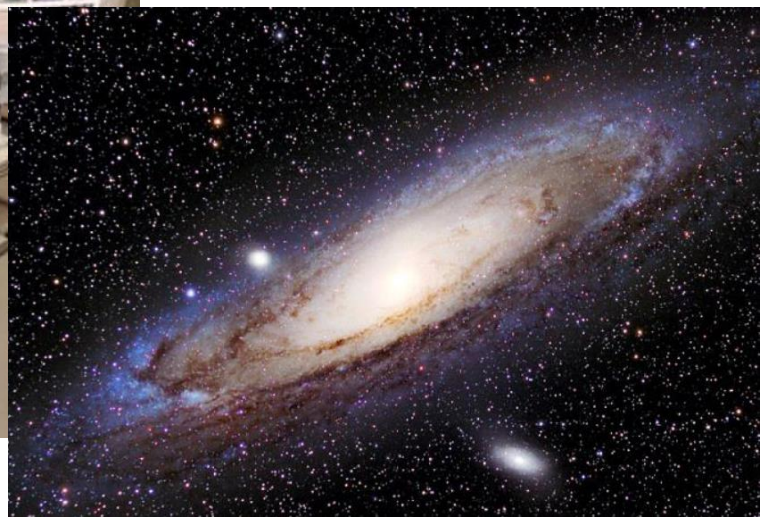
Automatic Speech Recognition
machine translation, dialog systems





Scientific Data Analysis
(e.g. Bioinformatics, Astronomy)





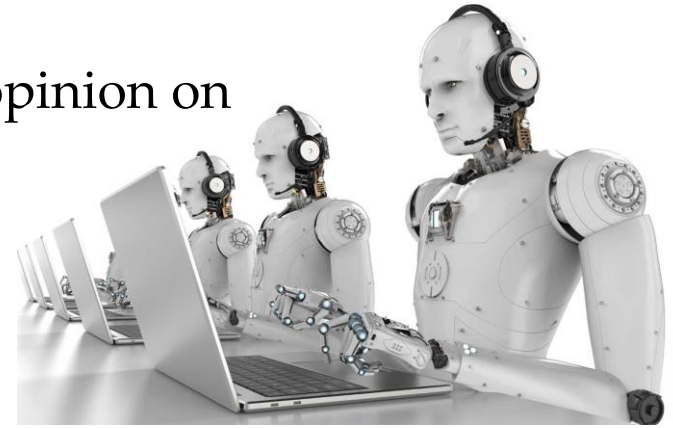
Economics, Finance

- Financial prediction
- Recommender systems
- Customer profiling
- Credit card fraud detection
- Credit risk assessment
- …

Solutions to new problems…

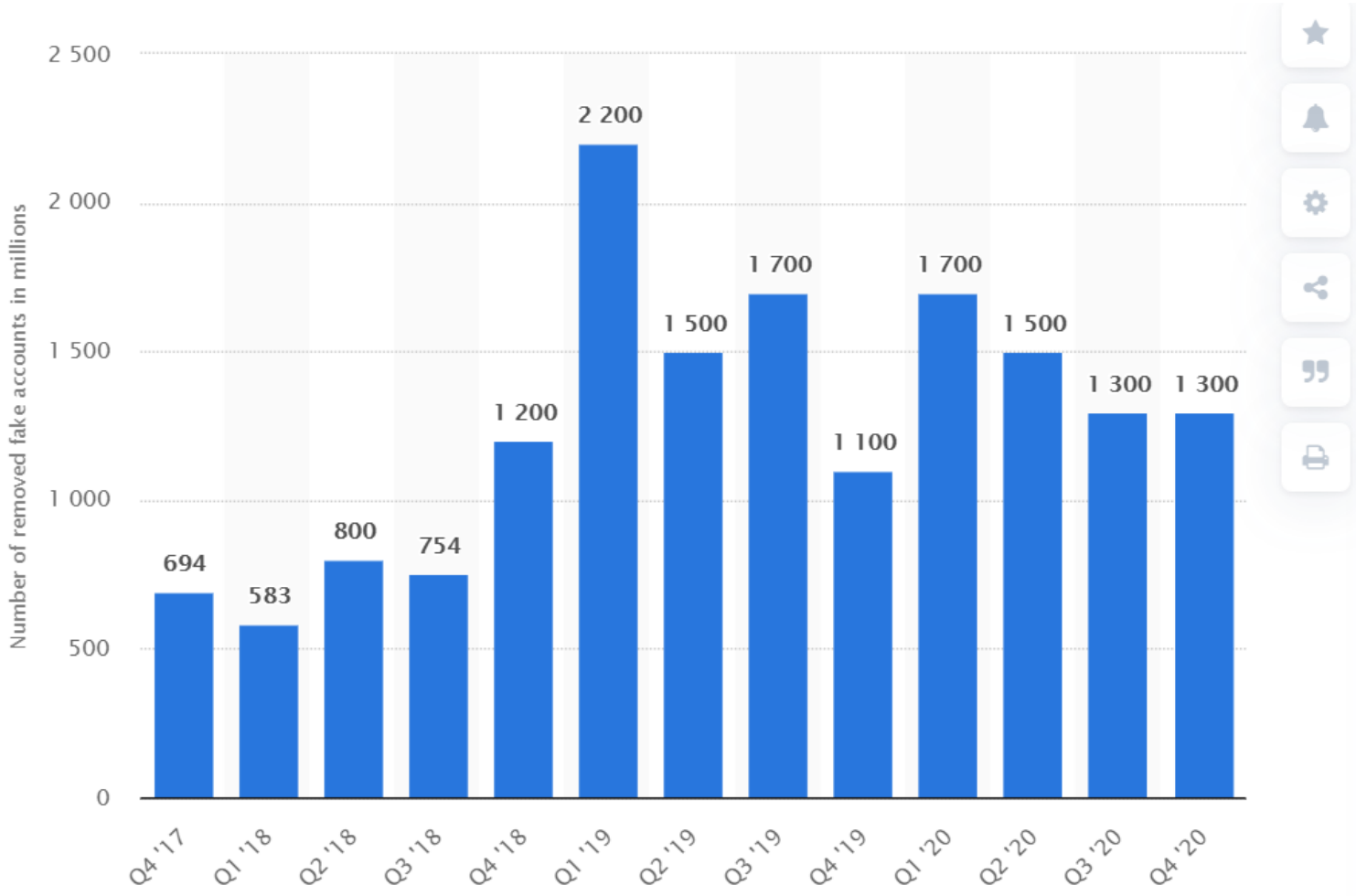
Bot or Not ?

Political bots are influencing public opinion on major political events worldwide



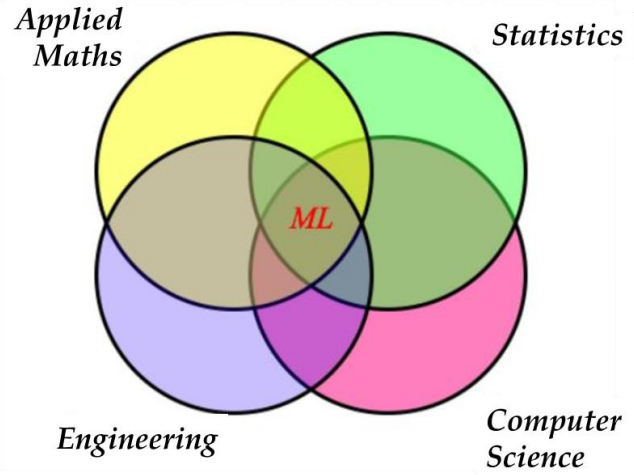
Brazil is the 4th country in the world in terms of Twitter users.

Global number of fake accounts deletd by Facebook 2017-2020

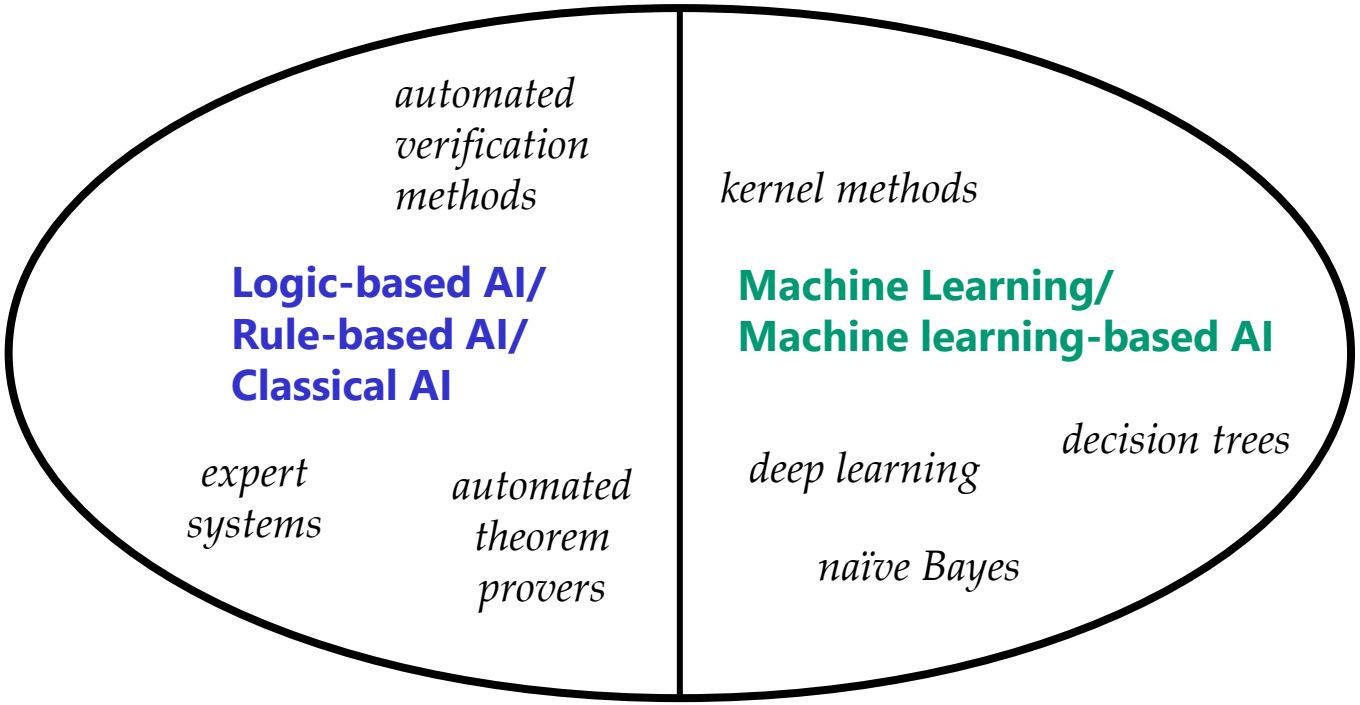


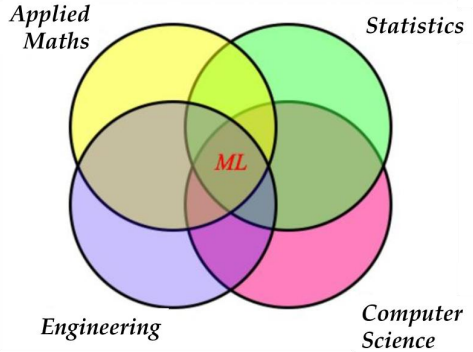
Additional Information

Show source



A.I.





DATA SCIENCE

Data visualization

Statistical analysis

Data storing

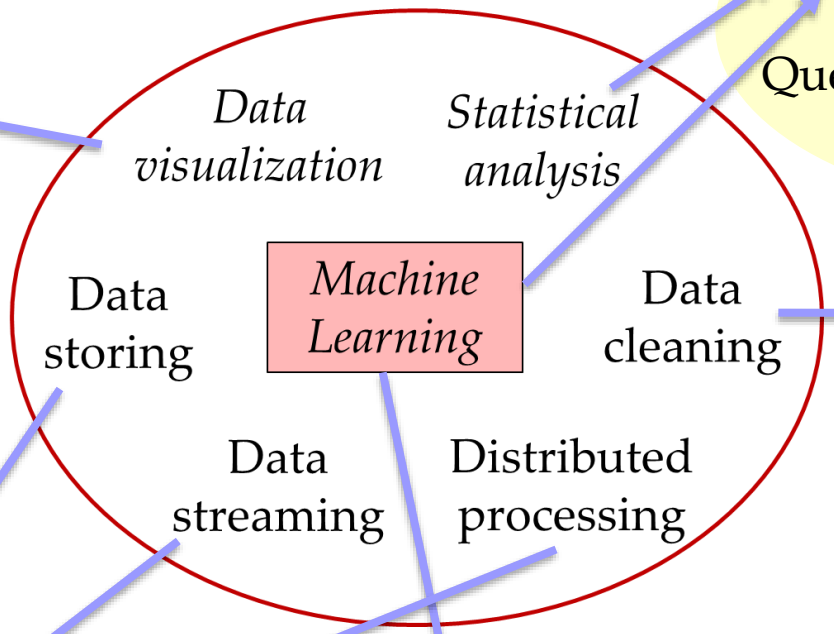
Machine Learning

Data cleaning

Data streaming

Distributed processing

The professionals for the job



Data Scientist

MSc, PhD
CS/AppMath/Stats
Question+data → formulates model

Data vis. analyst/ visualization artist

BSc CS
(Software Eng.)
implements the visualization

Feature Engineer

BSc CS (Software Eng.)
imputation, PCA, pre-processing

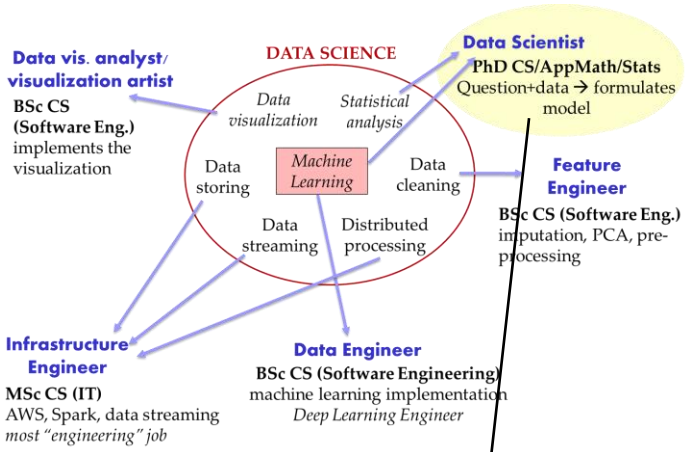
Infrastructure Engineer

MSc CS (IT)
AWS, Spark, data streaming
most "engineering" job

Data Engineer

BSc CS (Software Engineering)
machine learning implementation
Deep Learning Engineer

Data Scientist

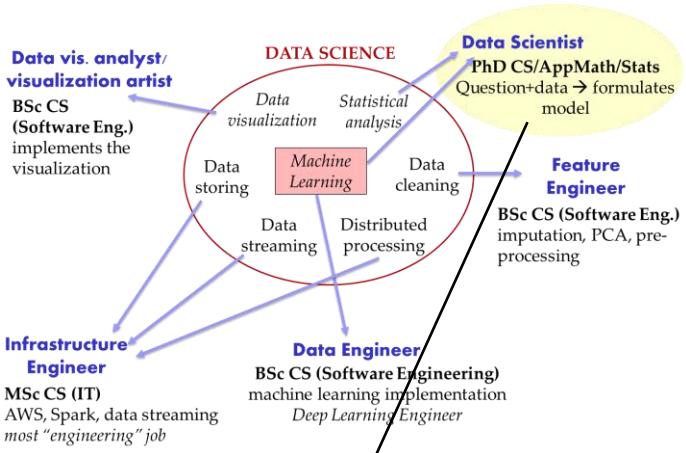


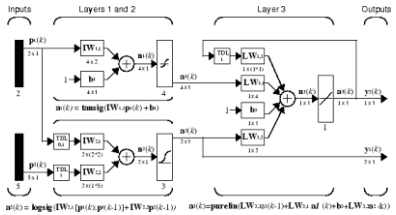
Application Domain Expert

1. Understanding **which questions** can be answered
2. **Which data** are needed

1. **Deep understanding** of the problem domain
2. Able to **interact with an expert** in the application domain
3. Is **not an expert** in the application domain

Partnership in research





feedback
(more data)

data

question

predictions

Application Domain Expert

The future ☺

- Better solutions to old problems
- Solutions to new problems

Societal impacts: both positive and negative

- Productivity
- Employment
- Health care
- Cities and transport
- Energy, climate, sustainability
- Political systems
- Government control
- Financial systems
- Education
- Crime, security
- Privacy
- Warfare – autonomous weapons and defence systems
- Human relationships