

# A Gentle Introduction to Data Science and Artificial Intelligence

---

Is artificial intelligence really  
what it seems to be?



**Fabio Mardero**

[fabio.mardero@ai-light.biz](mailto:fabio.mardero@ai-light.biz)

[linkedin.com/in/fabio-mardero](https://linkedin.com/in/fabio-mardero)

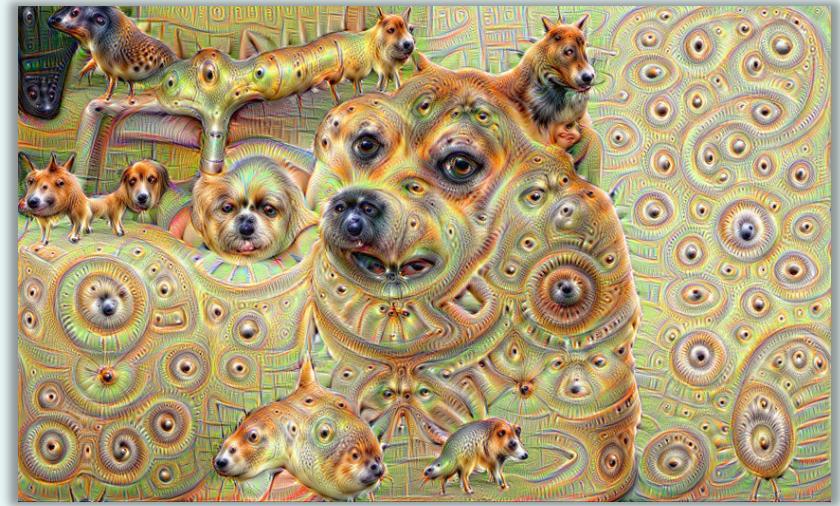
*AI-light*  
*with AI, for a brighter future*

# *Is Artificial Intelligence building the future terminators?*



*“Fears of Terminator-style robot wars stop society benefiting from AI, expert claims”,  
[thesun.ie/tech/3080612/](https://www.thesun.ie/tech/3080612/)*

*If these are the results that AI produces,  
there must be  
a misconception...*



“Google Deep Dream: l’algoritmo che trasforma le tue foto in un sogno”,  
[giornalettismo.com/google-deep-dream/](http://giornalettismo.com/google-deep-dream/)

# TABLE OF CONTENTS

---

What is it?

---

What can it do?

---

Why now?

---

How to do it?



# TABLE OF CONTENTS

---

What is it?

---

What can it do?

---

Why now?

---

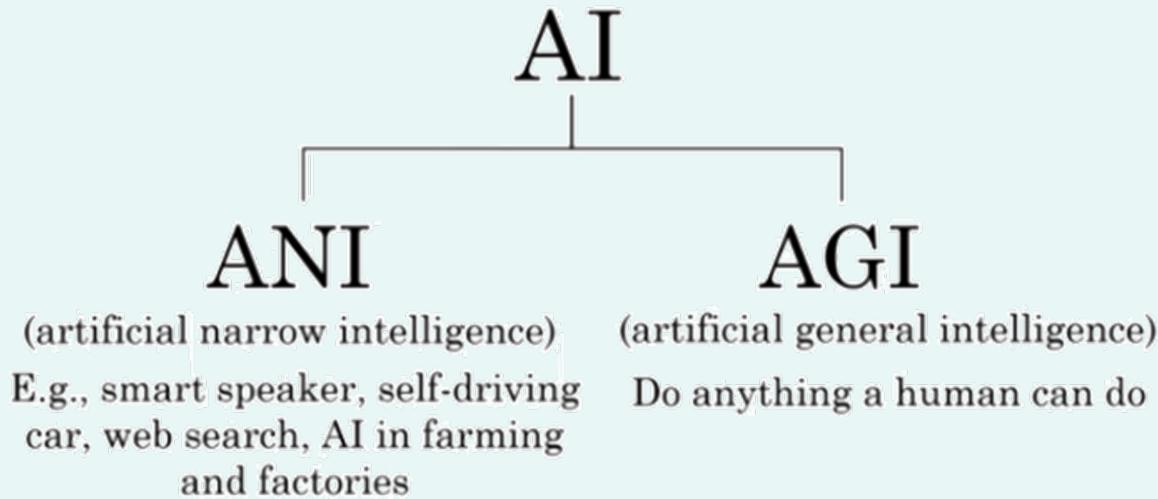
How to do it?



*Data Science VS Machine Learning*

# ARTIFICIAL INTELLIGENCE

*Artificial Intelligence  
is the intelligence demonstrated by machines*

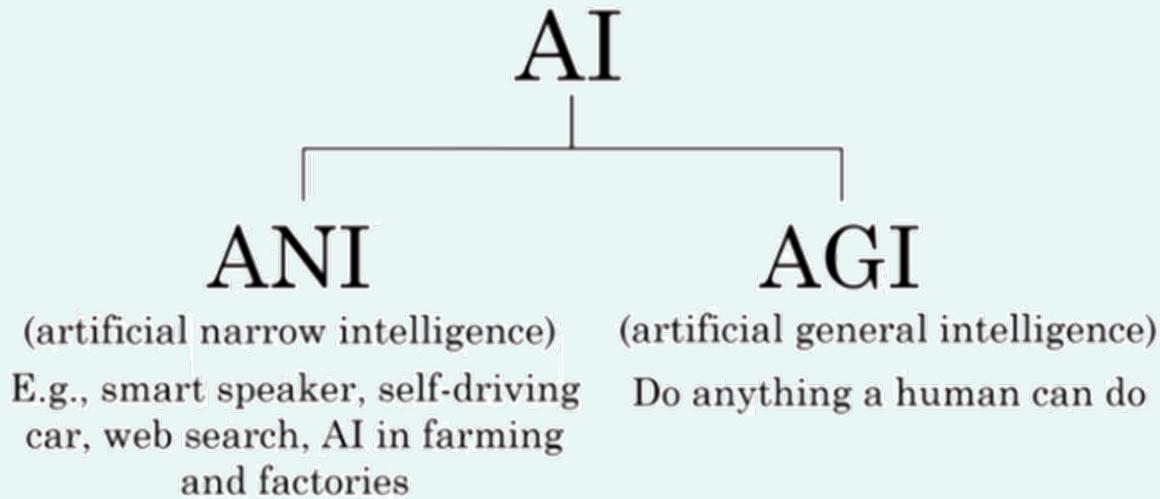


“AI for everyone”, [Andrew Ng, deeplearning.ai](#)

# ARTIFICIAL INTELLIGENCE

*“Can Natural Intelligence be so precisely described  
that a machine can be made to simulate it?”*

*From Dartmouth Summer Research Project on Artificial Intelligence (1956) the founding event of artificial intelligence as a field*



“AI for everyone”, [Andrew Ng, deeplearning.ai](#)

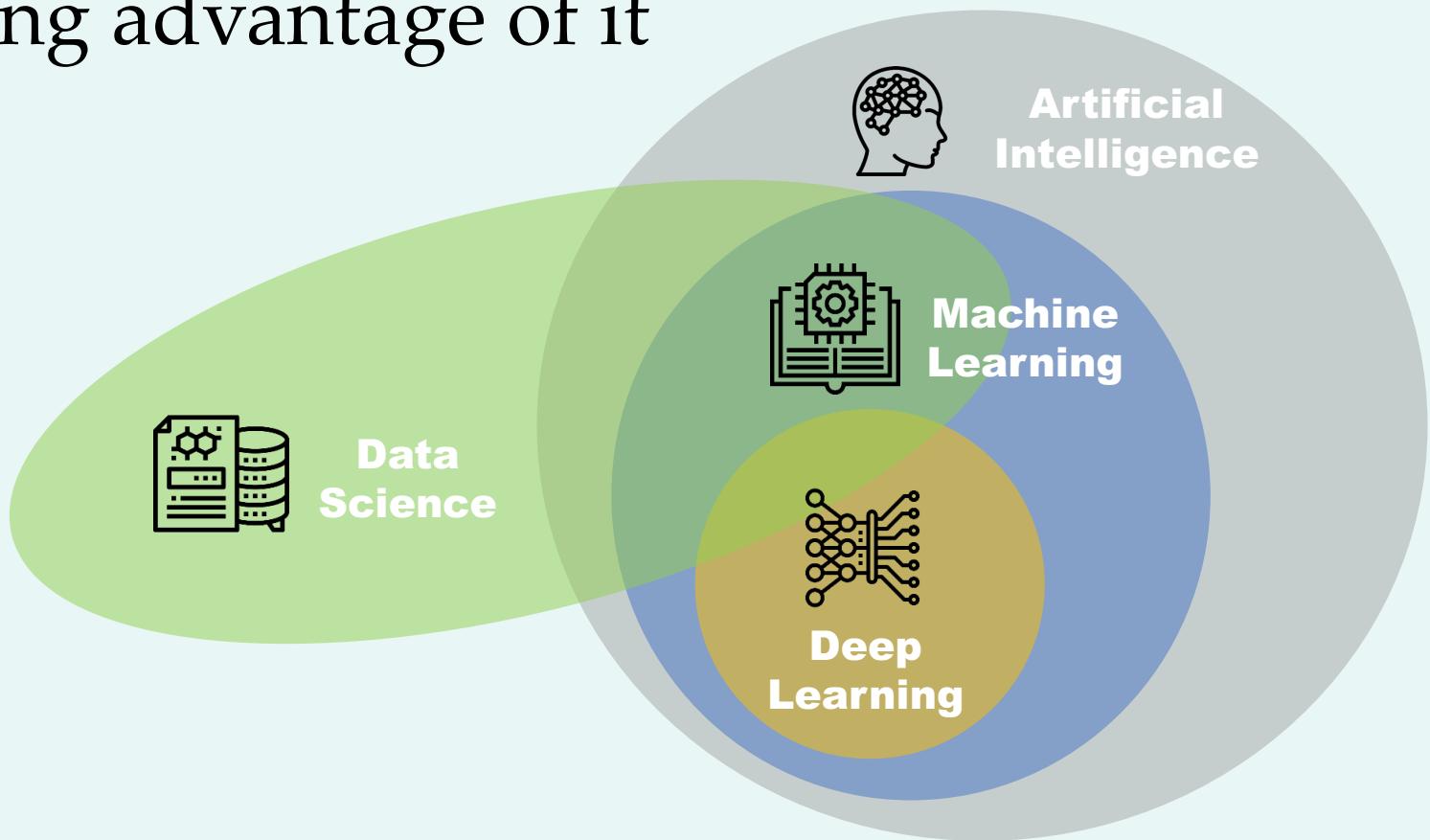
# ARTIFICIAL INTELLIGENCE

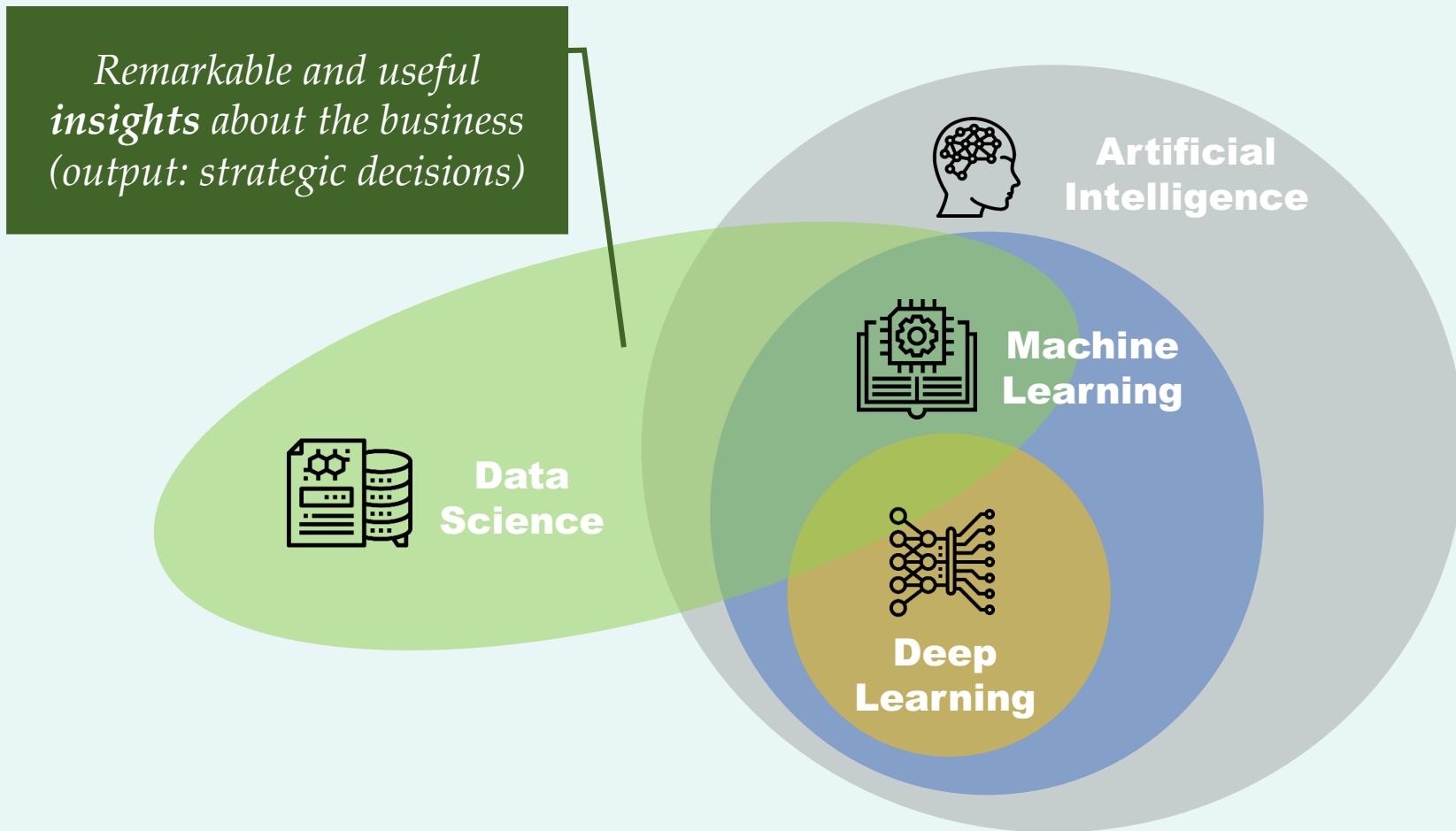
*...today's Artificial Intelligence  
is very far from its "General" version*

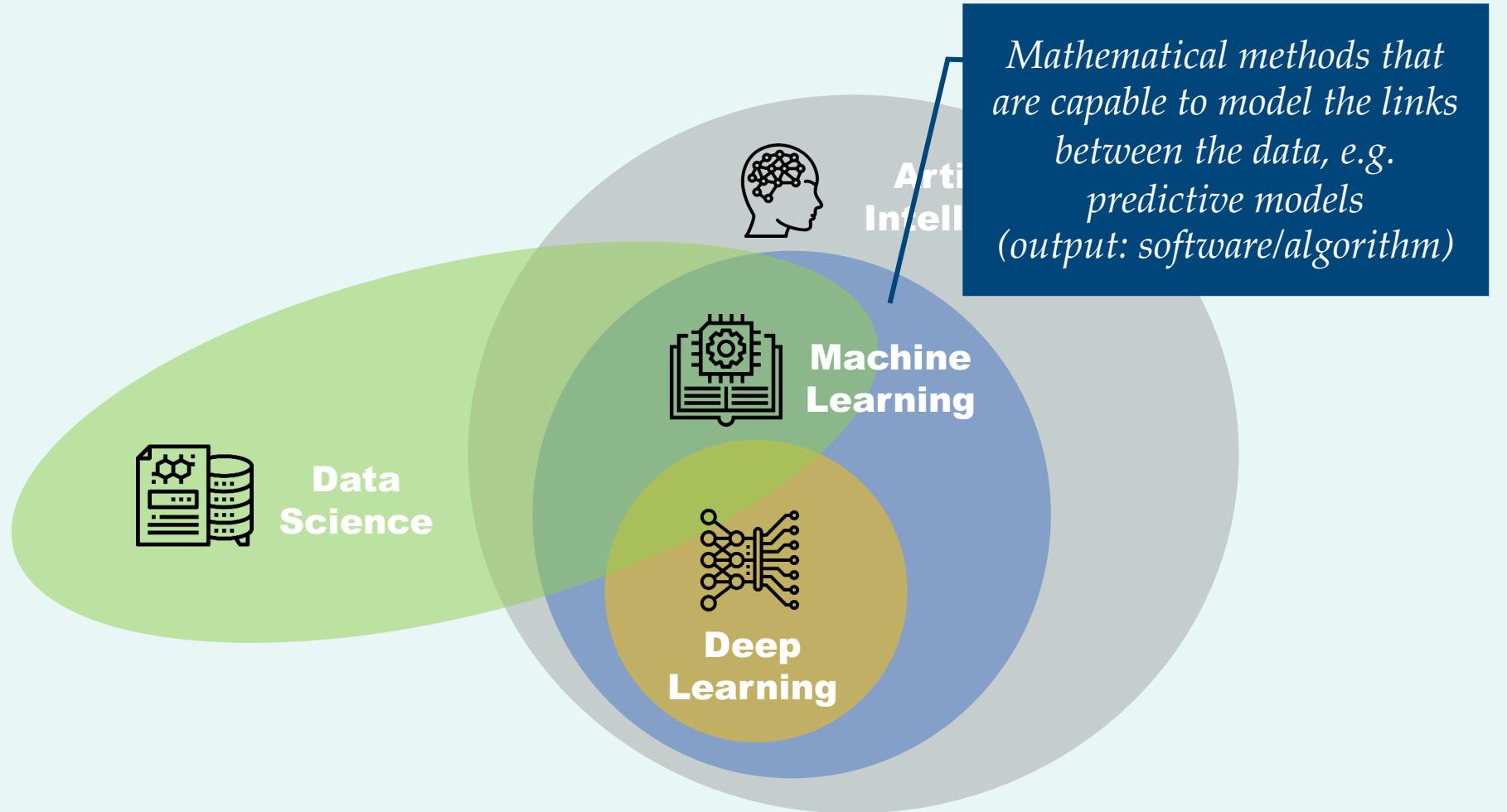


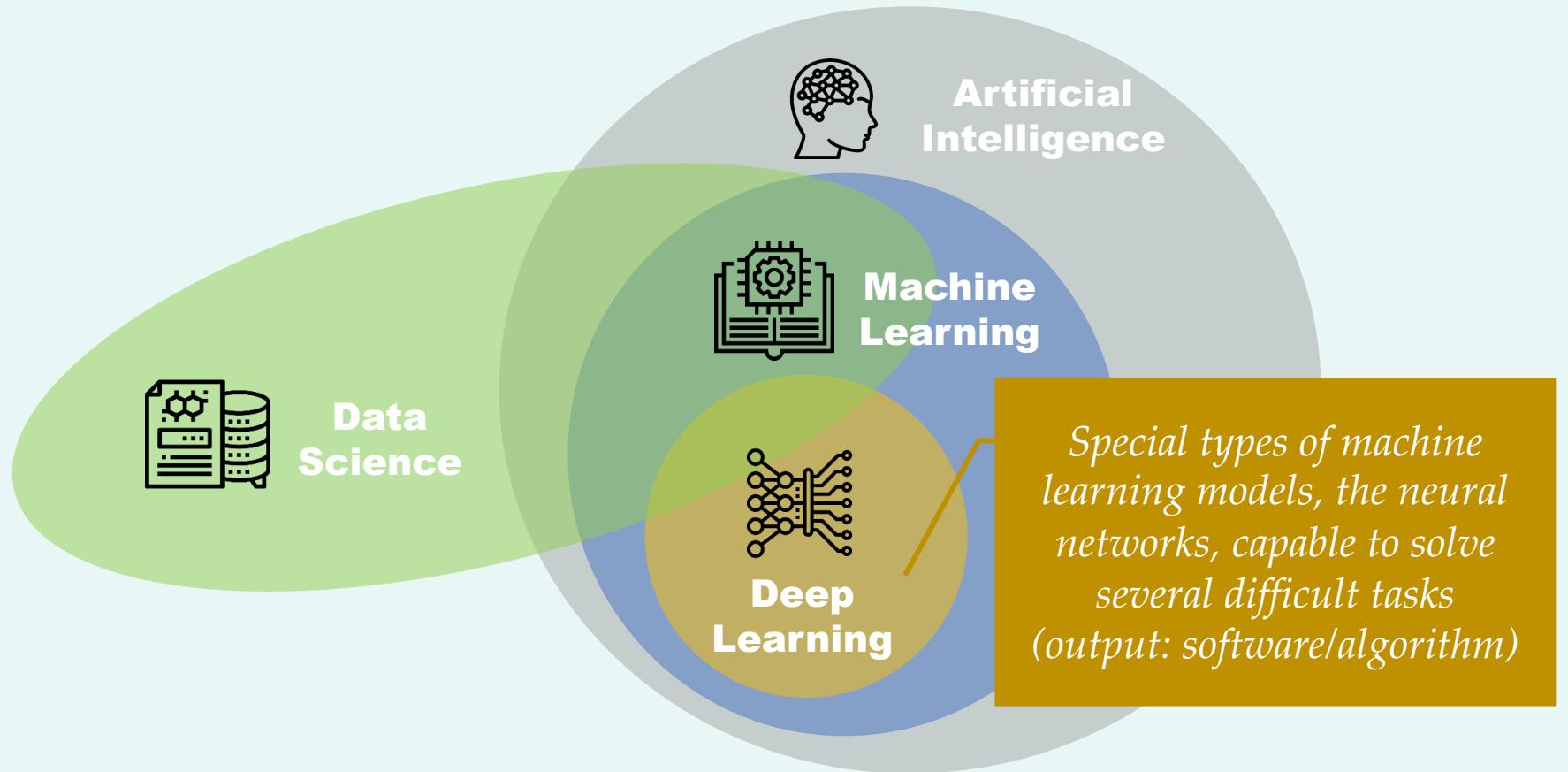
*Atlas, Boston Dynamics*

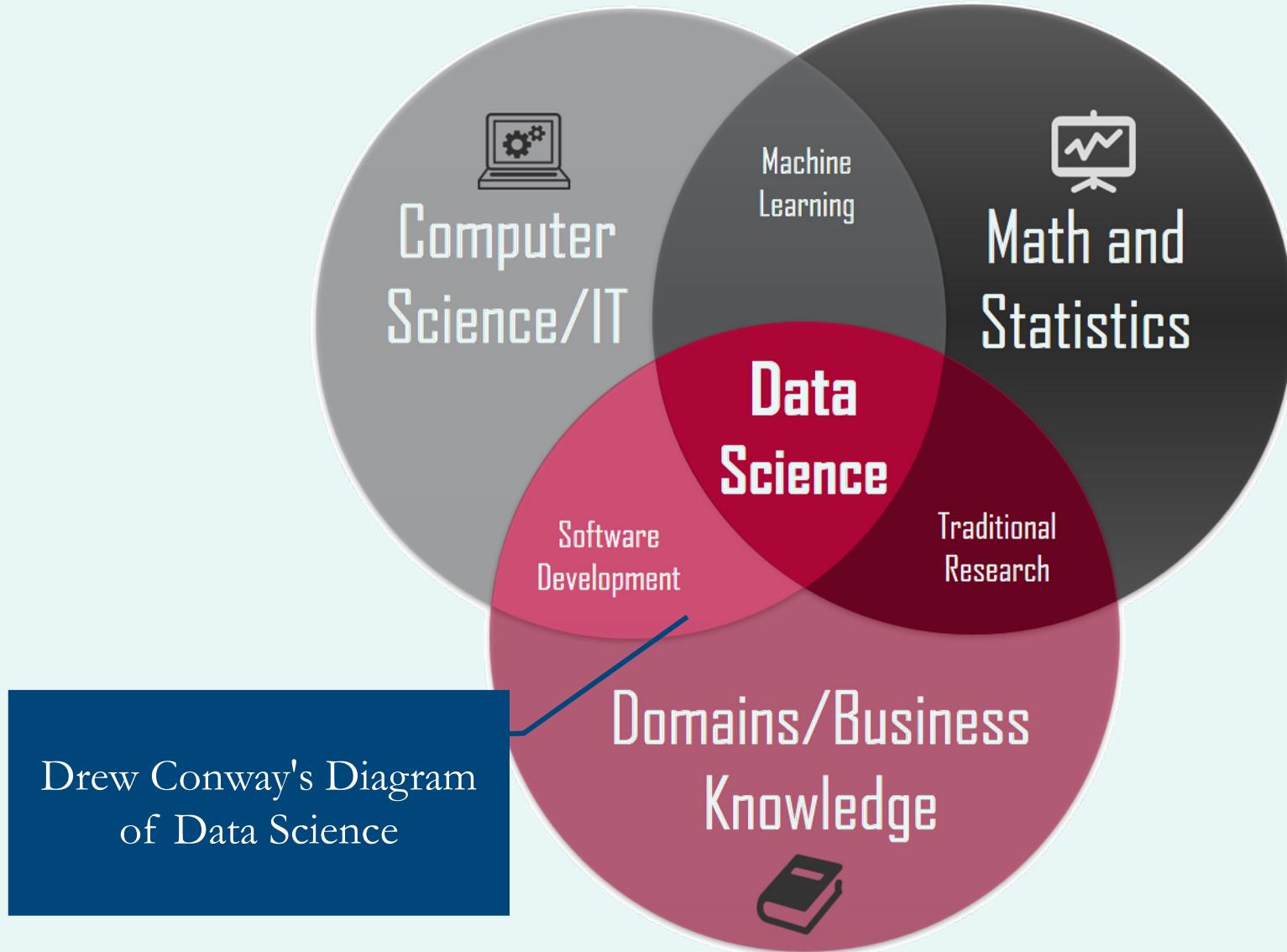
Machine learning and data science are about understanding a phenomenon and taking advantage of it



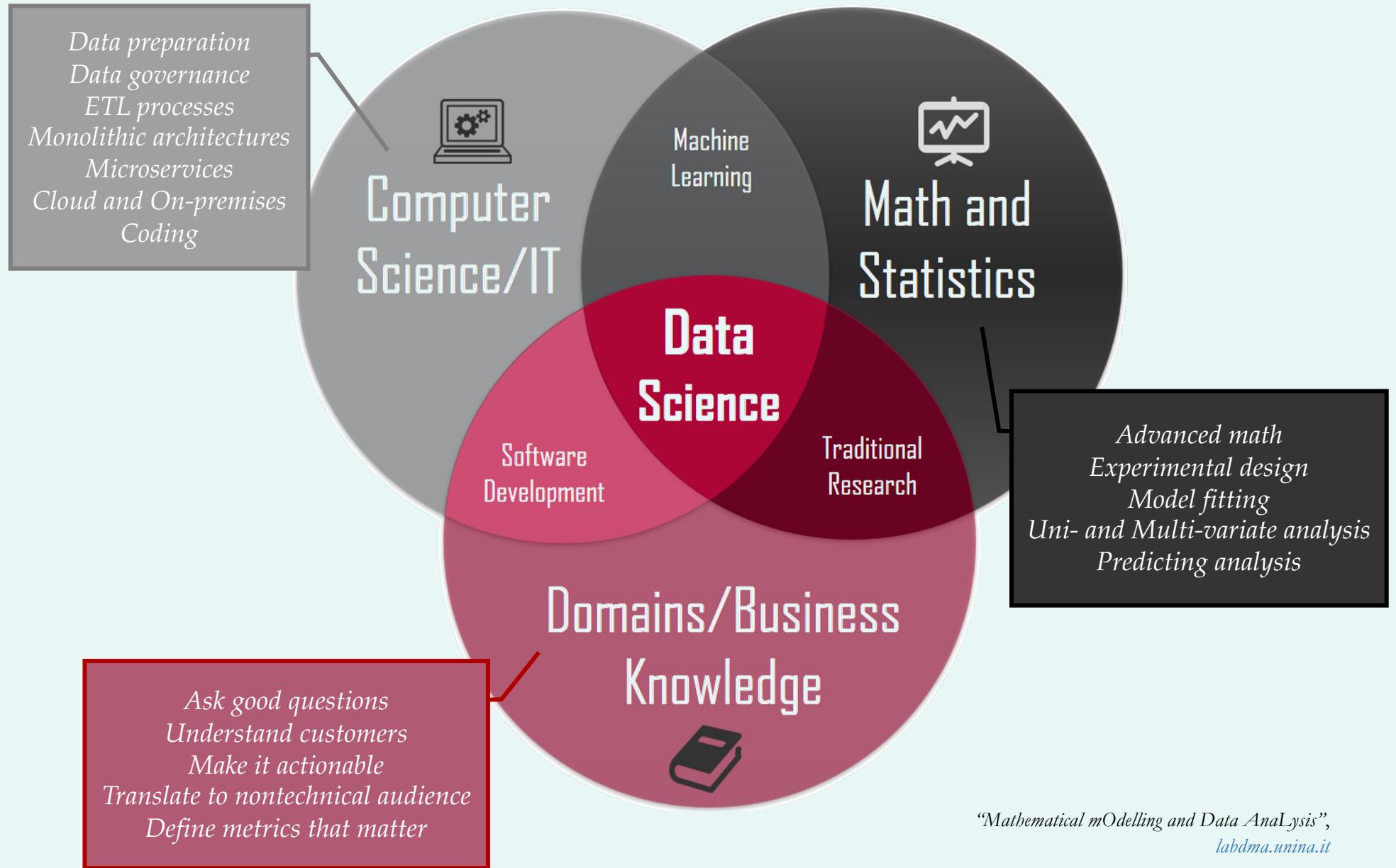








"*Mathematical mOdelling and Data AnaLysis*",  
[labdma.unina.it](http://labdma.unina.it)



# MACHINE LEARNING

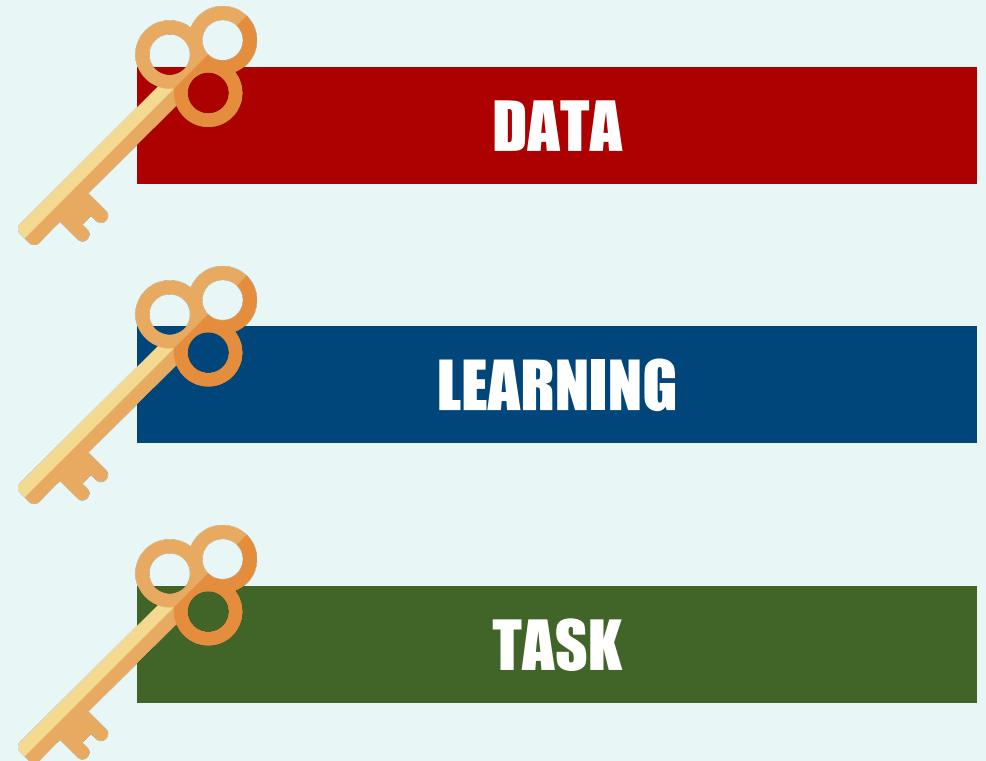
*Discipline that uses statistical methods  
to progressively improve, from previous information,  
the performance of an algorithm ("a model") in a given task*

# MACHINE LEARNING

*A machine learning model is able to “learn” from **data** in order to perform a given **task** in the best possible way*

# Machine Learning Model

*key elements:*

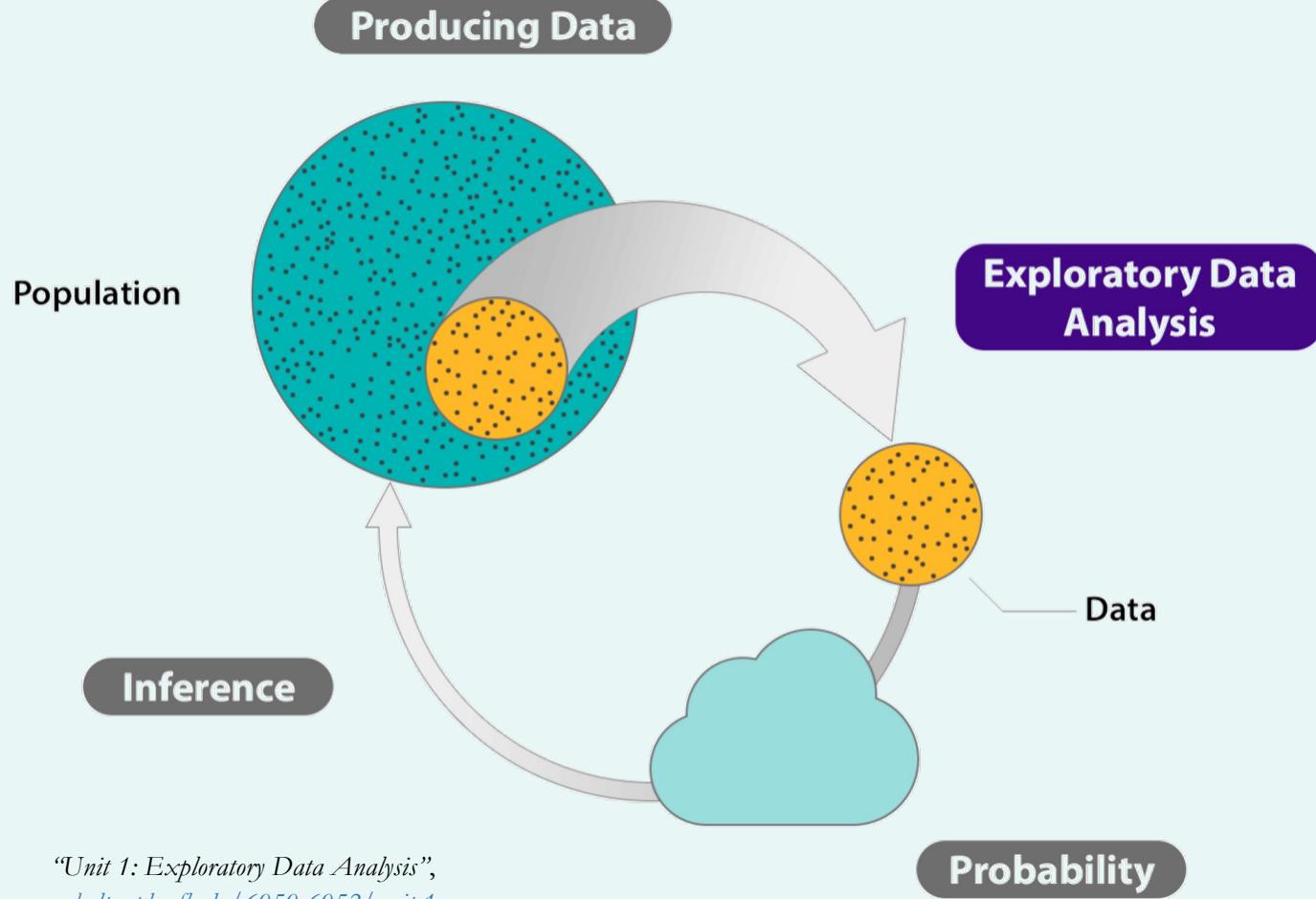


# DATA



Data is the information on which the machine learning model is adapted, so it must be as correct as possible

# DATA



"Unit 1: Exploratory Data Analysis",  
[bolt.mph.ufl.edu/6050-6052/unit-1](http://bolt.mph.ufl.edu/6050-6052/unit-1)

# DATA



*Could be in any format, including images, audio or text*



*The amount of data needed depends on the algorithm used and the difficulty of the task*

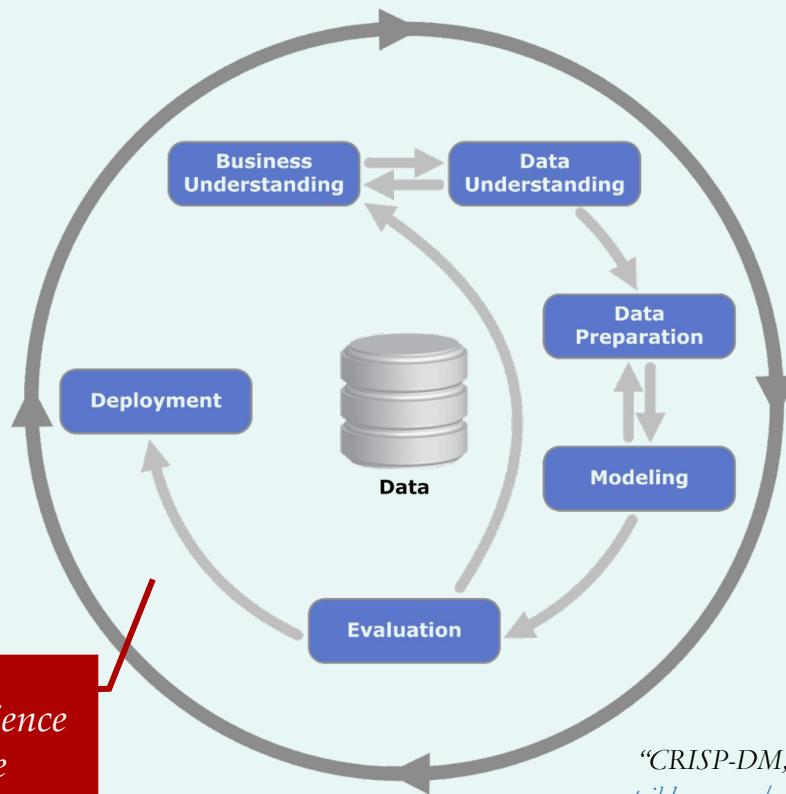


*Some data pre-processing may be needed before feeding it to the model*

# DATA

## CRISP-DM

*CRoss-Industry Standard Process for Data Mining*



*Describes a data science  
project life cycle*

"CRISP-DM, one AI/ML Lifecycle",  
[tribloom.com/crisp-dm-one-ai-ml-lifecycle](http://tribloom.com/crisp-dm-one-ai-ml-lifecycle)

# DATA

**CRISP-DM**  
*CRoss-Industry Standard Process for Data Mining*



EXPLORATIVE ANALYSIS



EXPLANATION AND PRESENTATION PHASE

# DATA

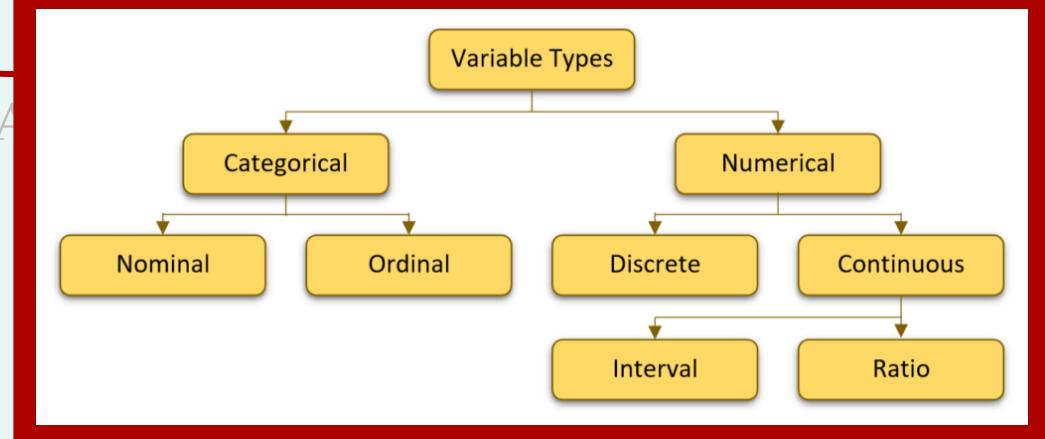
## CRISP-DM

*CRoss-Industry Standard Process for Data Mining*



### EXPLORATIVE ANALYSIS

EXPLANATION A



"Statistics | Types of Variables",

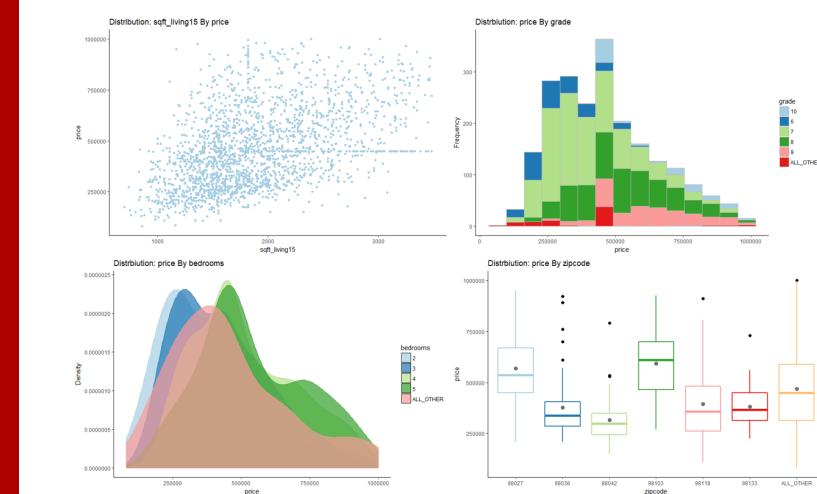
[k2analytics.co.in/statistics-types-of-variables](http://k2analytics.co.in/statistics-types-of-variables)

# DATA

Cross-Industry  
S

EXPLORATION

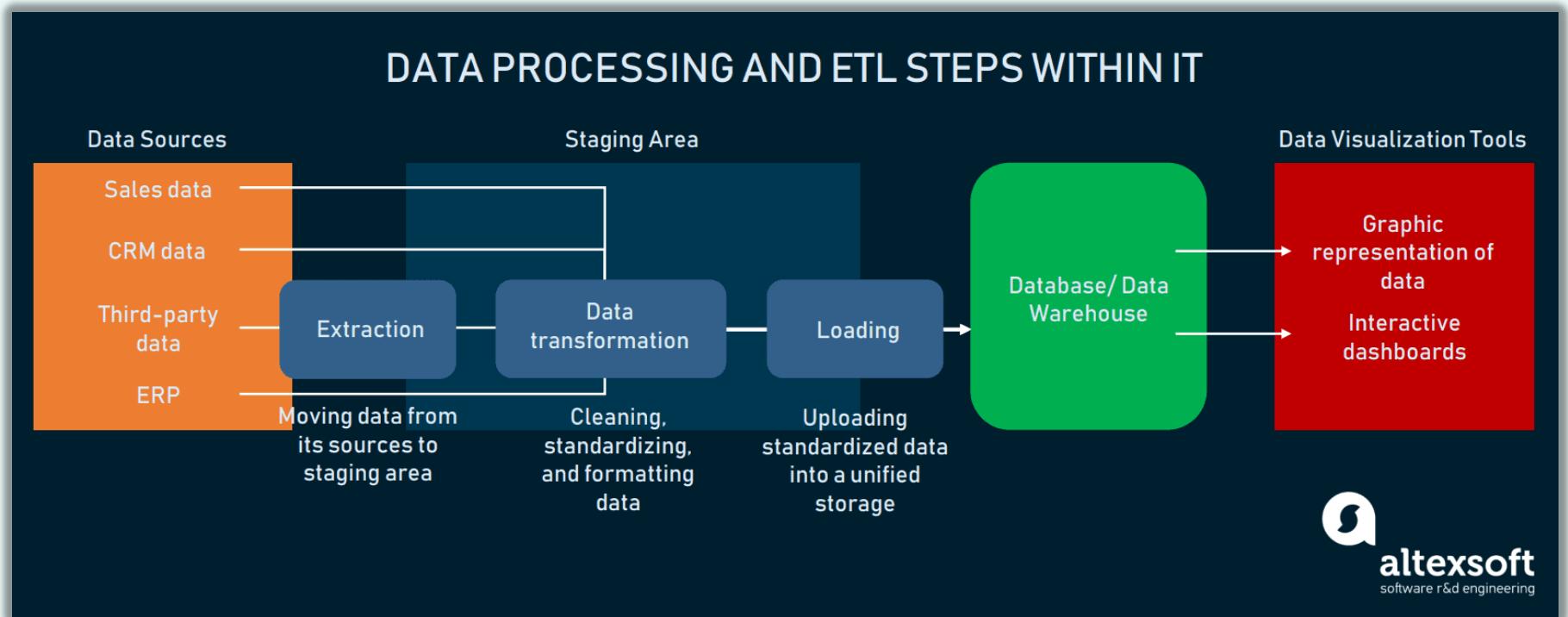
## Data Visualization



EXPLANATION AND PRESENTATION PHASE

"Auto-EDA - Automated Exploratory Data Analysis",  
[github.com/sourarbose1991/Auto-EDA](https://github.com/sourarbose1991/Auto-EDA)

# DATA



# DATA

## CRISP-DM

*CRoss-Industry Standard Process for Data Mining*

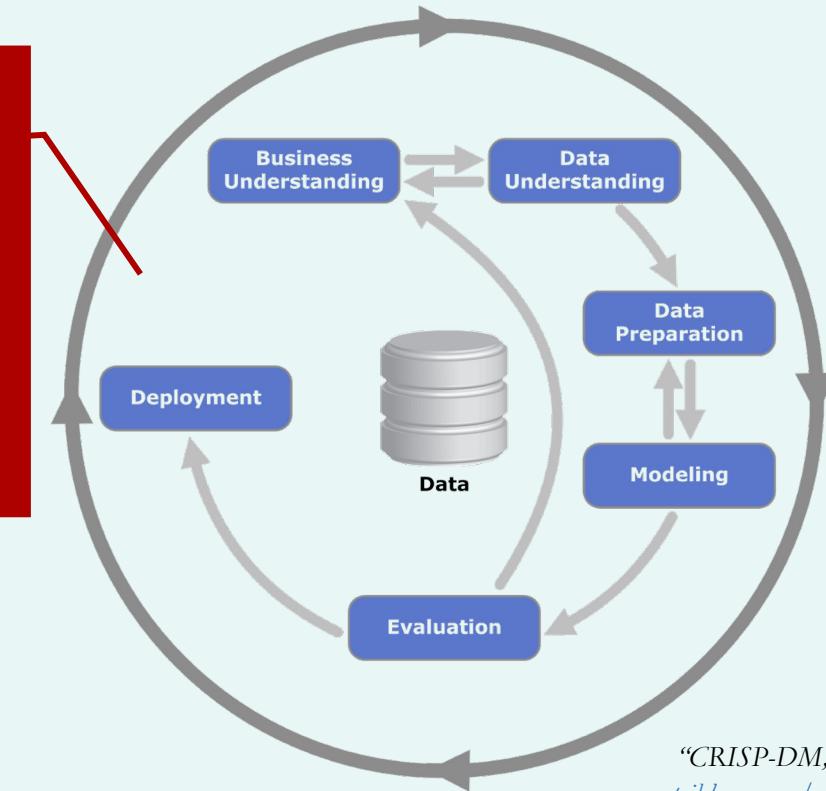


*Requires Extra Efforts:*

*Data Governance*

*Data Security*

*Data Compliance*

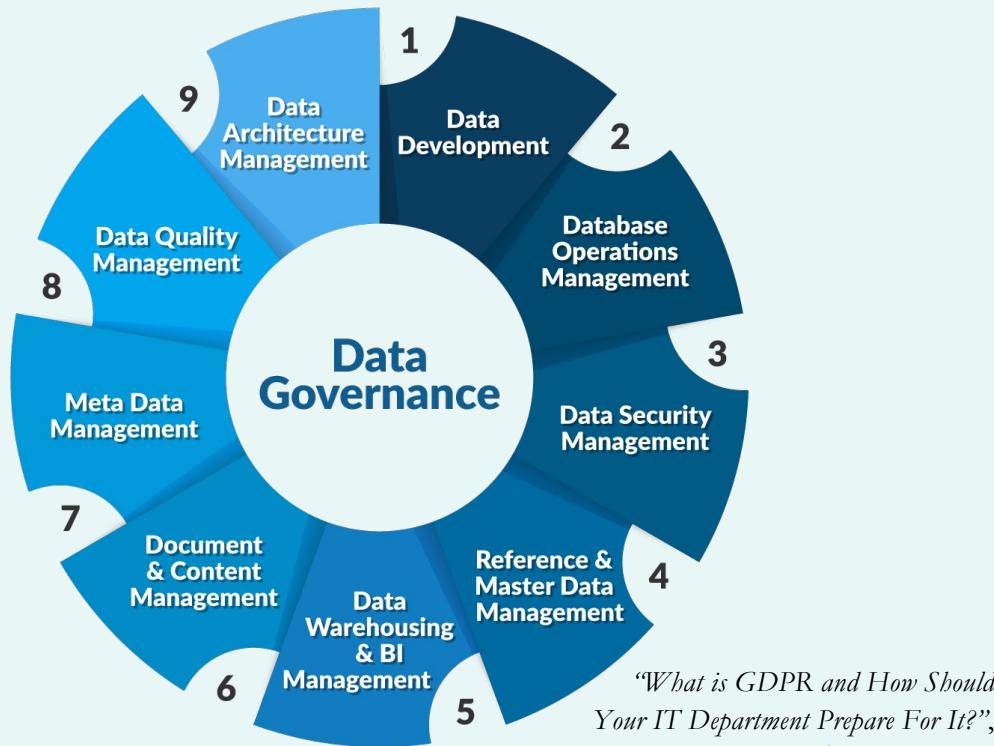


"CRISP-DM, one AI/ML Lifecycle",  
[tribloom.com/crisp-dm-one-ai-ml-lifecycle](http://tribloom.com/crisp-dm-one-ai-ml-lifecycle)

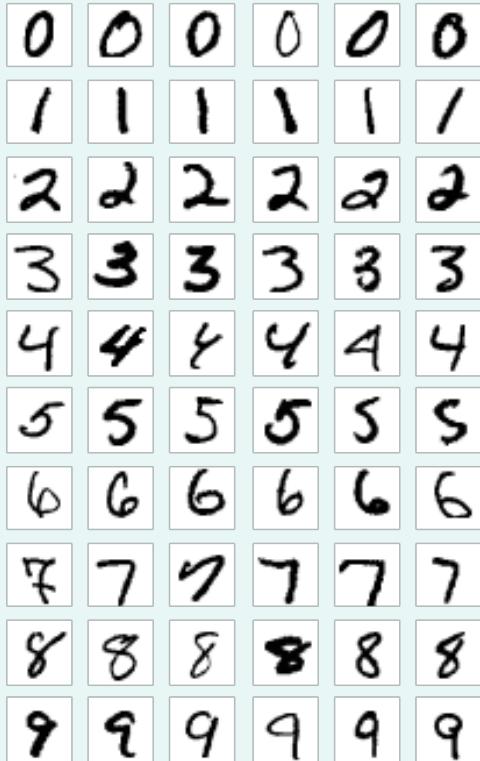
# DATA

## DAMA Wheel

*from the DAMA-DMBOK Functional Framework  
(DAta MAnagement Body of Knowledge Functional Framework)*



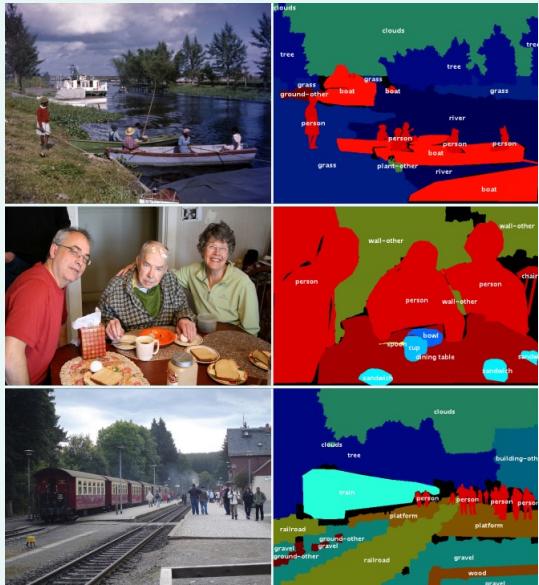
# DATA



**MNIST**  
*(modified National Institute of Standards and Technology database)  
dataset*

***Multi-class Classification***  
*70k labeled images (28x28 pixels)*  
*< 1% accuracy error (~2002)*

# DATA



## COCO (Common Objects in Context) dataset

*Object Detection and Segmentation*

*164k labeled images*

*~1.5 million object instances*

*> 40% mIoU (2020) for segmentation*

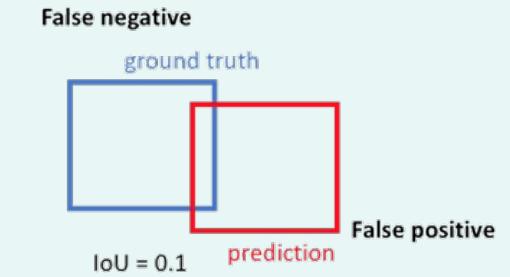
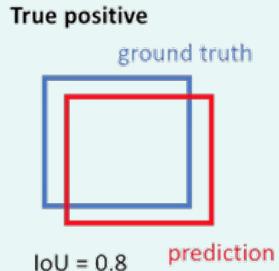
“COCO dataset”,  
[github.com/nightrome/cocostuff](https://github.com/nightrome/cocostuff)

# DATA



## COCO (Common Objects in Context) dataset

*Example*  
Threshold: 0.5



“Evaluating image segmentation models.“,  
[jeremyjordan.me/evaluating-image-segmentation-models/](http://jeremyjordan.me/evaluating-image-segmentation-models/)

“COCO dataset”,  
[github.com/nigtrome/cocostuff](https://github.com/nigtrome/cocostuff)

# LEARNING



A model learns when it *changes* its structure, or its parameters, to reduce the errors of its outputs

# LEARNING

*Types:*



Reinforcement Learning



Supervised Learning



Unsupervised Learning

# REINFORCEMENT LEARNING

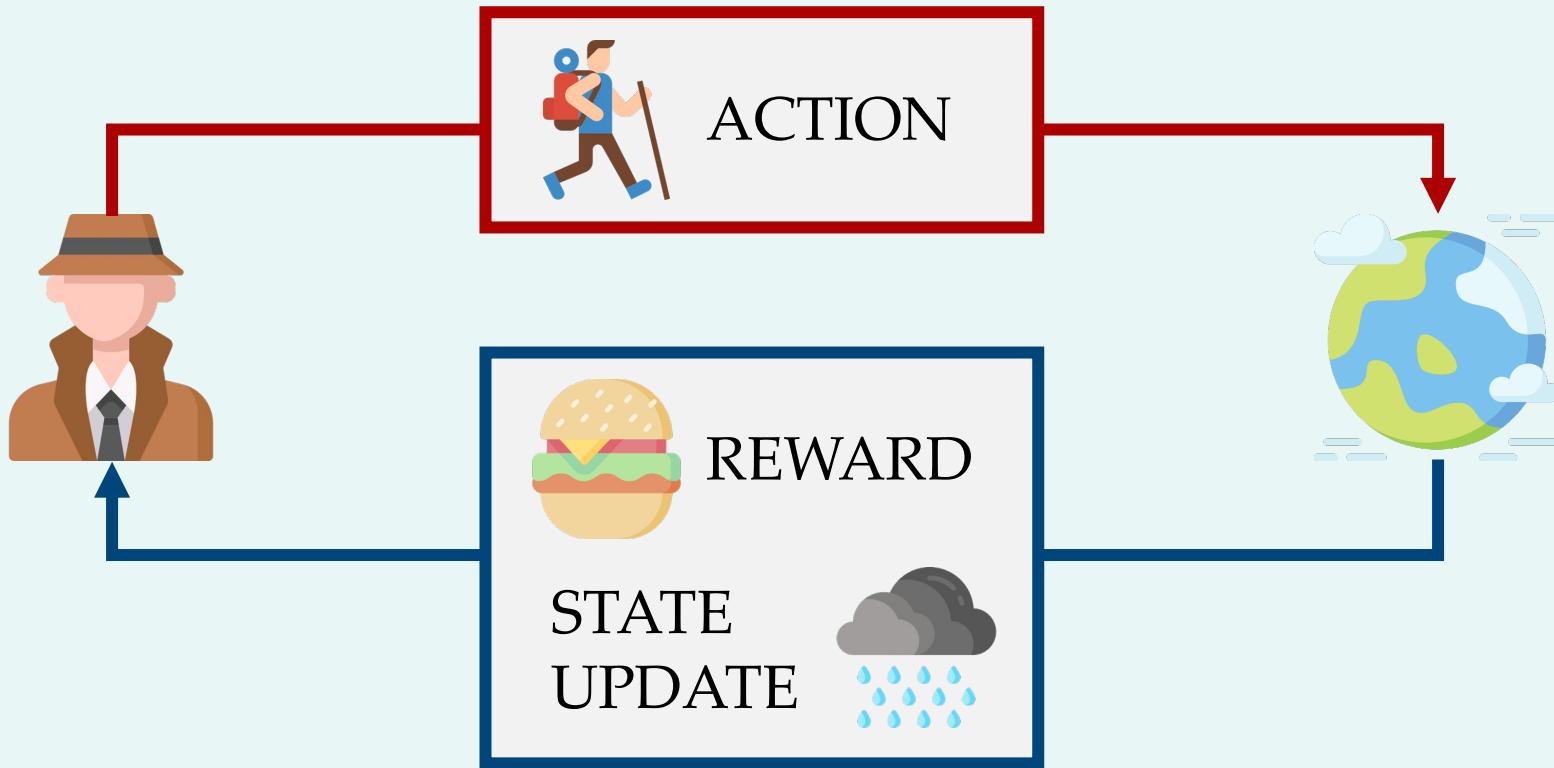


The ML model is an “*agent*”



The data is collected from an “*environment*”

# REINFORCEMENT LEARNING



# REINFORCEMENT LEARNING



The agent interacts with the environment trying to maximize the obtained rewards with a trial-and-error process

# REINFORCEMENT LEARNING



"Deep  $Q$  Network learning to play Galaxian (cleaner encode)",  
[youtube.com/channel/UCL-Kp9-biF0npDfGivI3s1Q](https://youtube.com/channel/UCL-Kp9-biF0npDfGivI3s1Q)

# SUPERVISED LEARNING

In supervised learning the ML model  
*suffers the environment*  
AND  
*is trained to output a prediction*

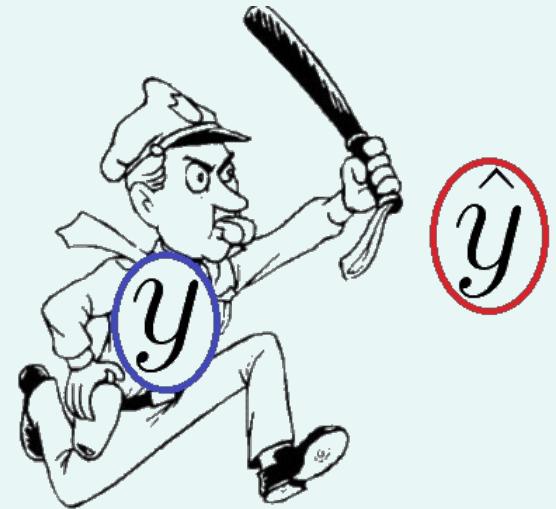
# SUPERVISED LEARNING



The model aims to predict the behavior of one or more observed variables with respect to the others

# SUPERVISED LEARNING

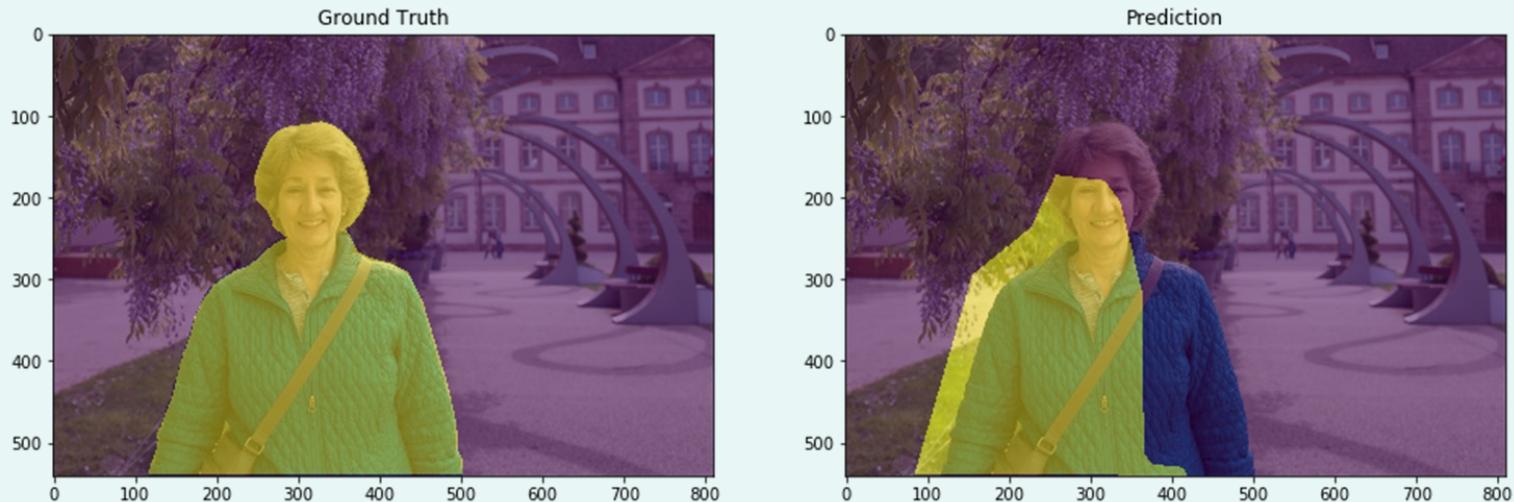
Learning is, informally, “supervised”  
by the ground truth values



$\hat{y}$  : predicted values

$y$  : ground truth values

# SUPERVISED LEARNING



*“Evaluating image segmentation models”,  
[jeremyjordan.me/evaluating-image-segmentation-models/](http://jeremyjordan.me/evaluating-image-segmentation-models/)*

# UNSUPERVISED LEARNING

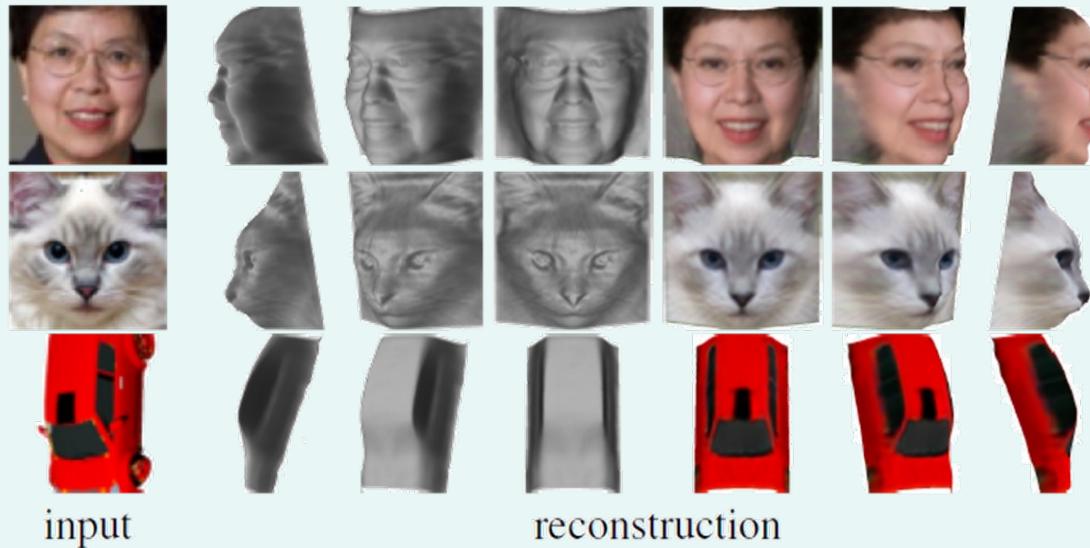
In unsupervised learning the ML model  
*suffers the environment*  
AND  
*does not output a prediction*

# UNSUPERVISED LEARNING



The model aims to find patterns and links between the data

# UNSUPERVISED LEARNING



*"Unsupervised Learning of Probably Symmetric Deformable 3D Objects from Images in the Wild"*,  
Shangzhe Wu, Christian Rupprecht, Andrea Vedaldi

# TASK



The task defines what the model is trained on and with what intentions, e.g. provide forecasts or find data aggregation patterns

# TASK

*Two cases:*



Some variables (*target/response var.*) are intended as more important than others (*covariates/features/explanatory var.*)



All variables are intended as (potentially) significant

# TASK

Given a set of data, it is up to the observer to decide how he intends to interpret the information and whether to assign particular importance to any of the available variables

# TASK

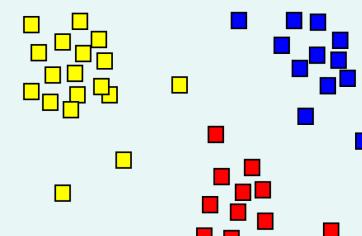
*Two cases:*

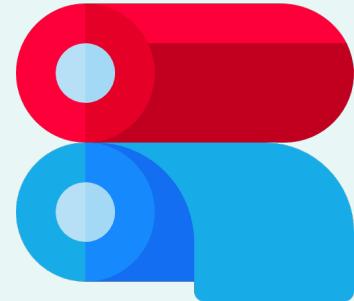


Regression and classification tasks

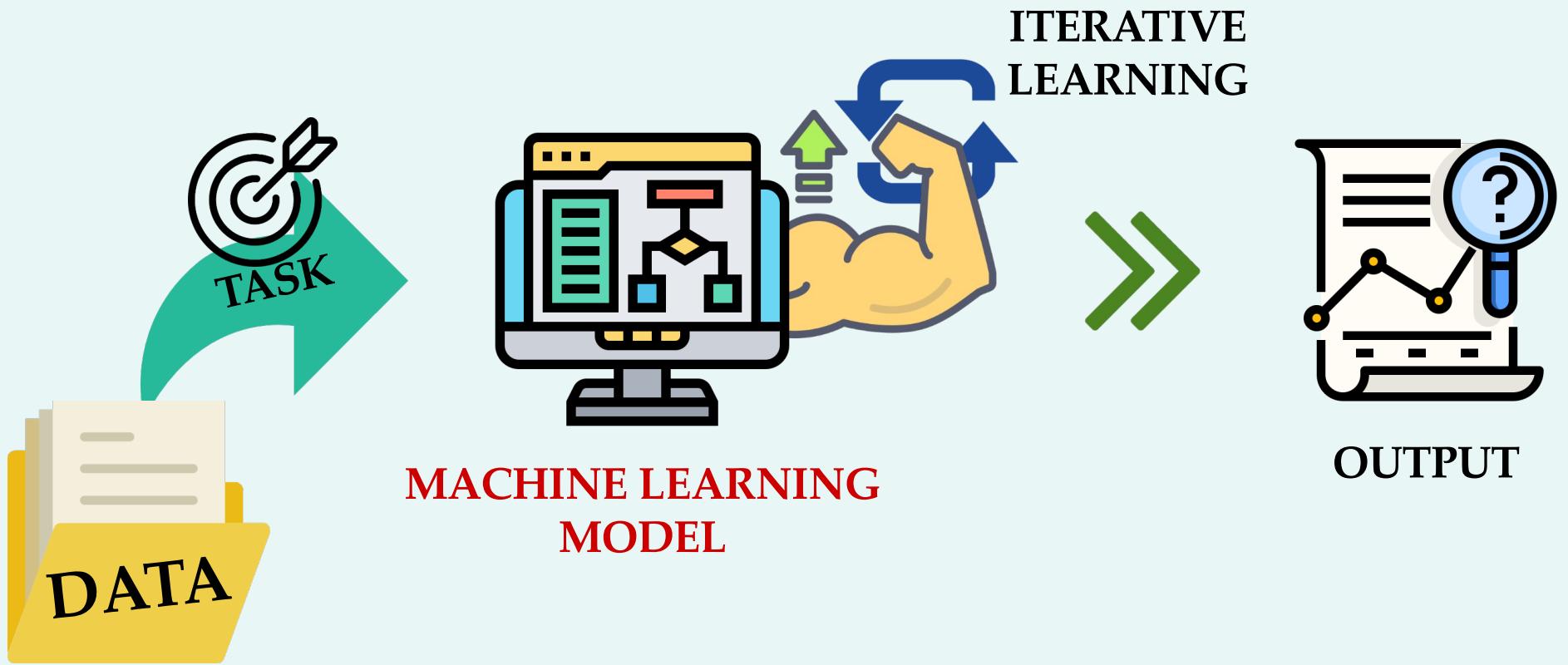


Pattern discovery tasks (e.g. clustering)

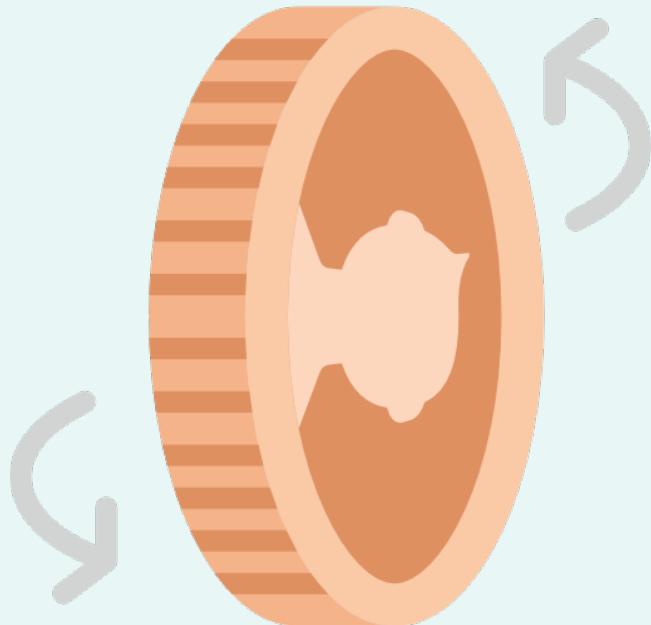




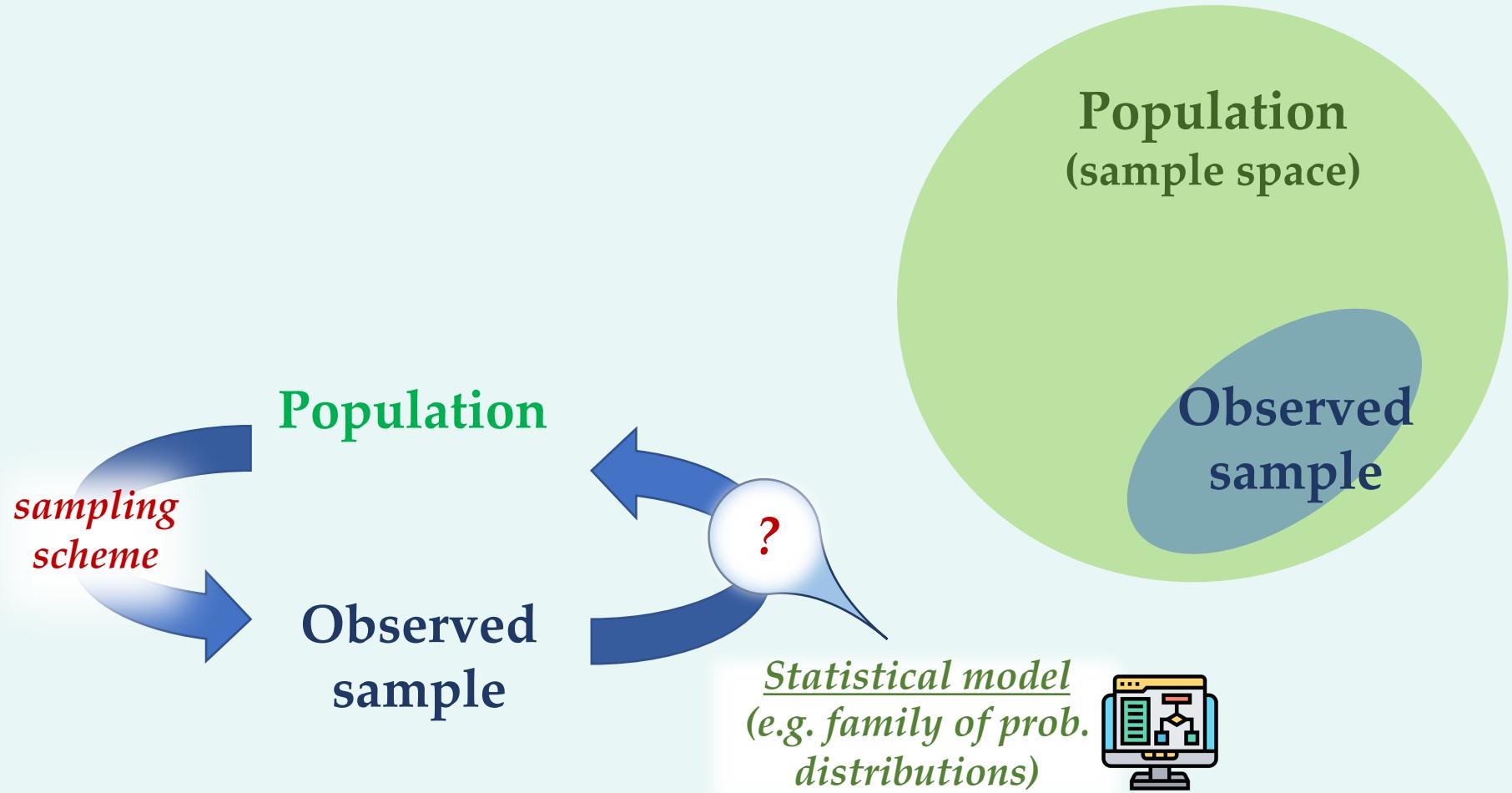
Wrapping up...



*‘Nothing is sure in life, so it needs to be treated with probabilities’*



# How we see the phenomenon...



# A supervised problem example using linear regression...



Observed  
sample

## DATASET

$$\mathcal{D} = \{(\mathbf{x}_i, y_i)\}_{i=1}^N \quad \text{with} \quad \mathbf{x}_i \in \mathcal{X}, \quad y_i \in \mathcal{Y}$$



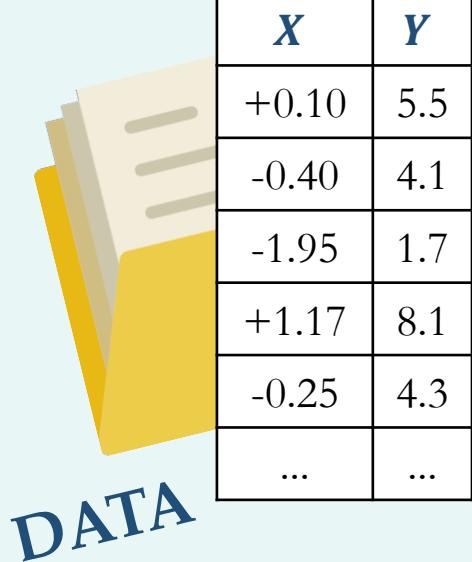
## MODEL

$$y = f(\mathbf{x}) + \varepsilon$$

*deterministic*
*stochastic*

$$f : \mathcal{X} \rightarrow \mathcal{Y}$$

$$\mathbf{x} \rightarrow f(\mathbf{x}) = \hat{y}$$

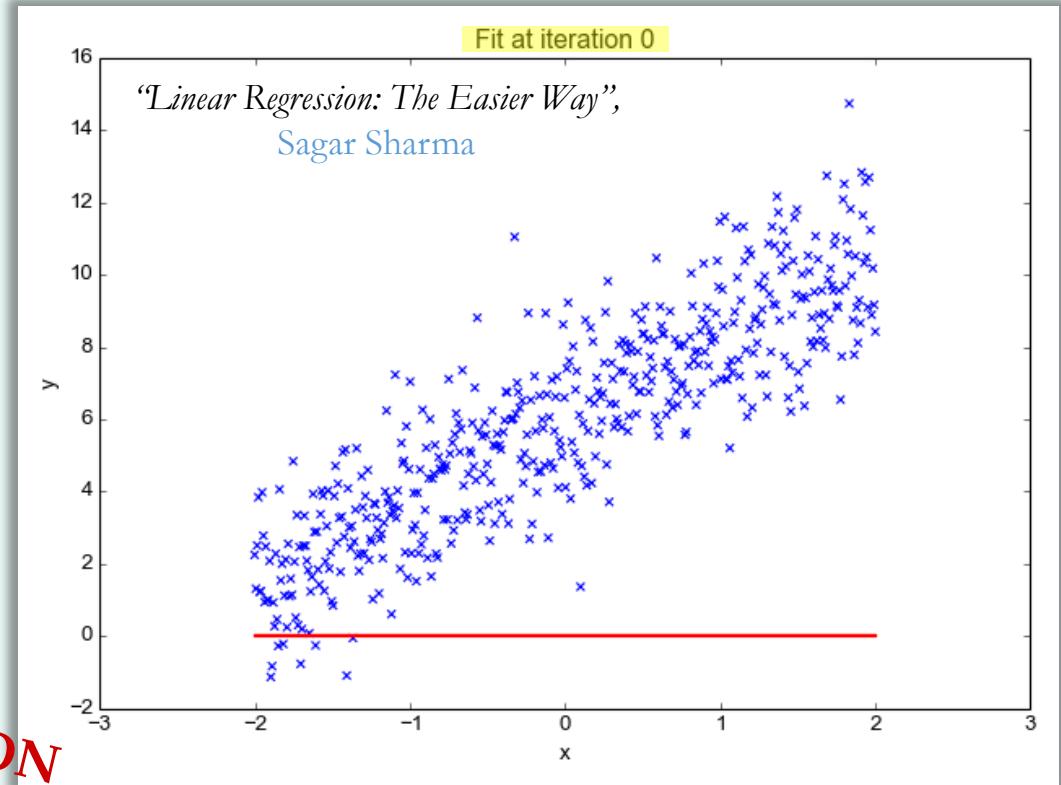
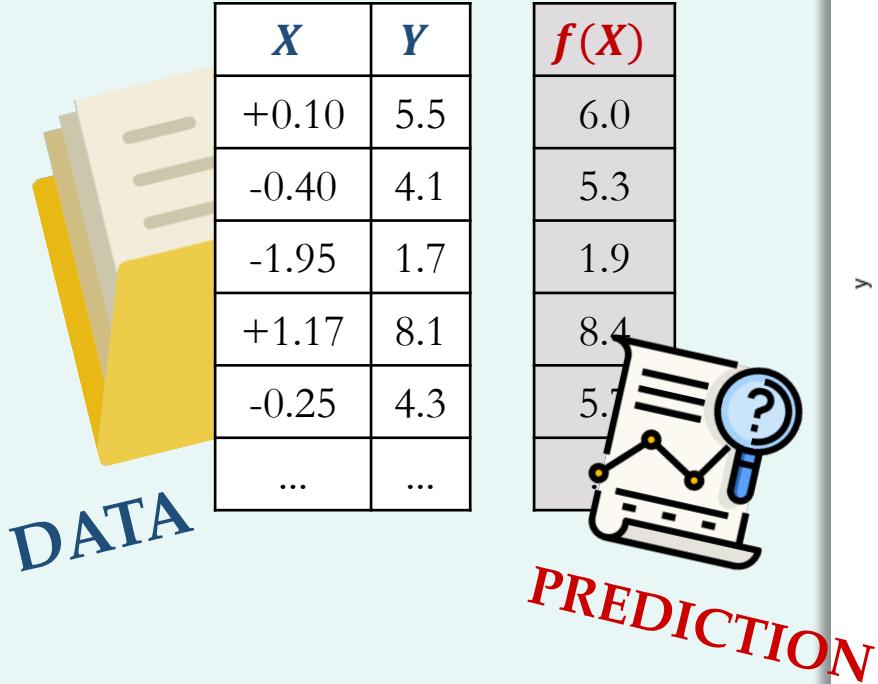


$$\mathcal{L} = \frac{1}{N} \sum_{i=1}^N (y_i - f(x_i))^2$$

$$f(x_i) = m x_i + q$$

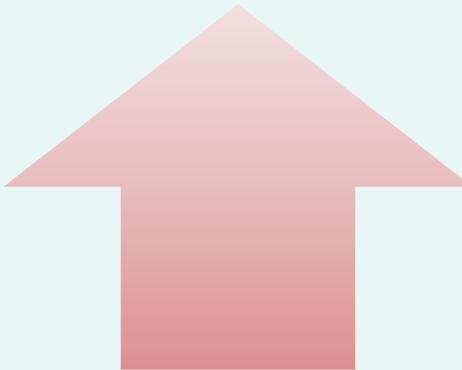


The algorithm searches for the **parameters** that allow to minimize the **loss/error function** computed on the known data

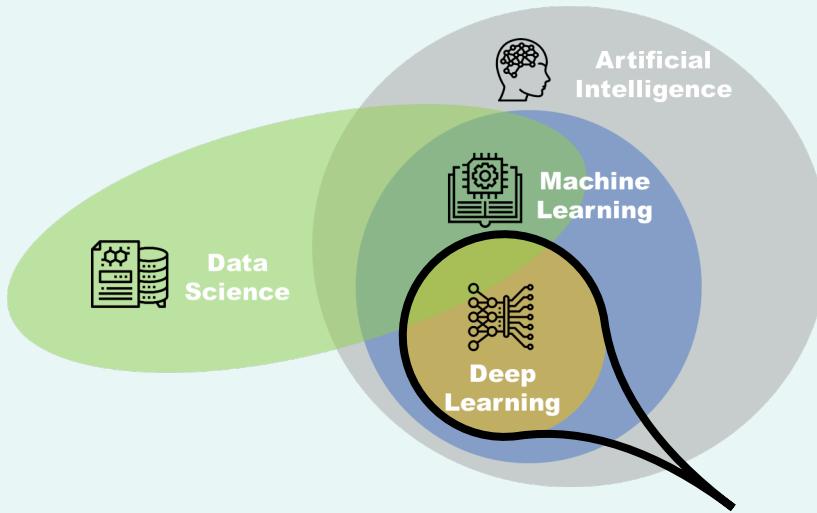


The algorithm searches for the parameters that allow to minimize the loss/error function computed on the known data

# Learning is about finding the “best” model parameters

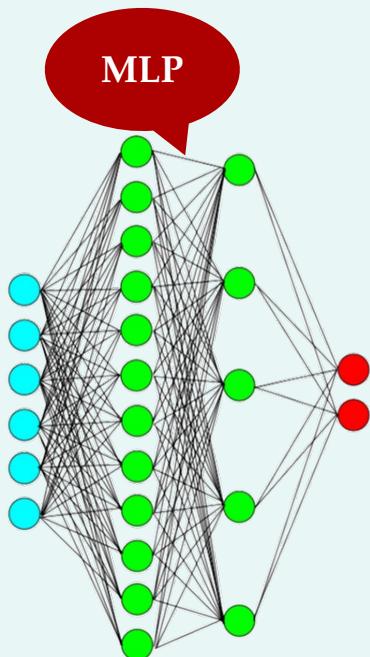


The algorithm searches for the **parameters** that allow to minimize  
**the loss/error function** computed on the known data

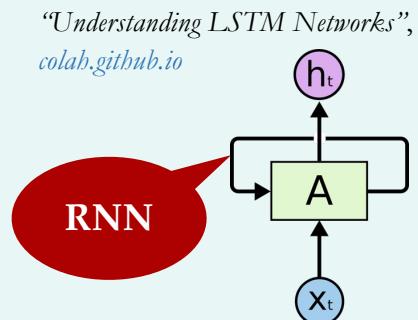


**Deep Neural Networks**  
are the most powerful tool (2020) when  
image/text/audio related problems  
need to be solved

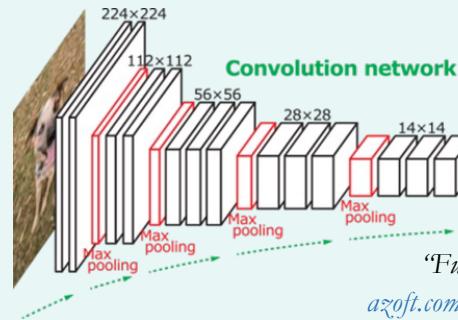
# Neural network can be built using different architectures



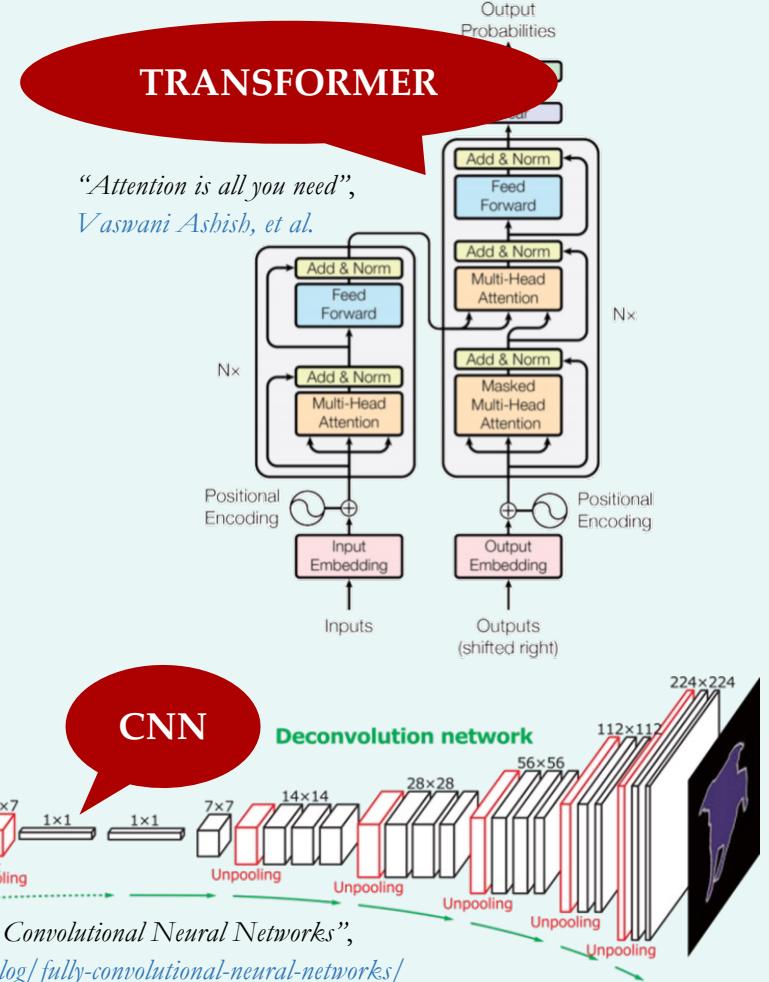
"Multilayer Perceptron",  
[blog.refu.co/wp-content/uploads/2009/05/mlp.png](http://blog.refu.co/wp-content/uploads/2009/05/mlp.png)



"Understanding LSTM Networks",  
[colab.github.io](https://colab.github.io)



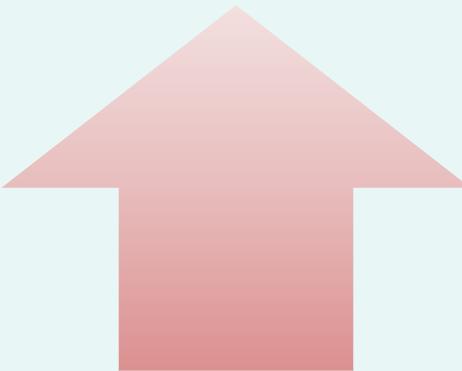
"Fully Convolutional Neural Networks",  
[azsoft.com/blog/fully-convolutional-neural-networks/](http://azsoft.com/blog/fully-convolutional-neural-networks/)



TRANSFORMER

"Attention is all you need",  
Vaswani Ashish, et al.

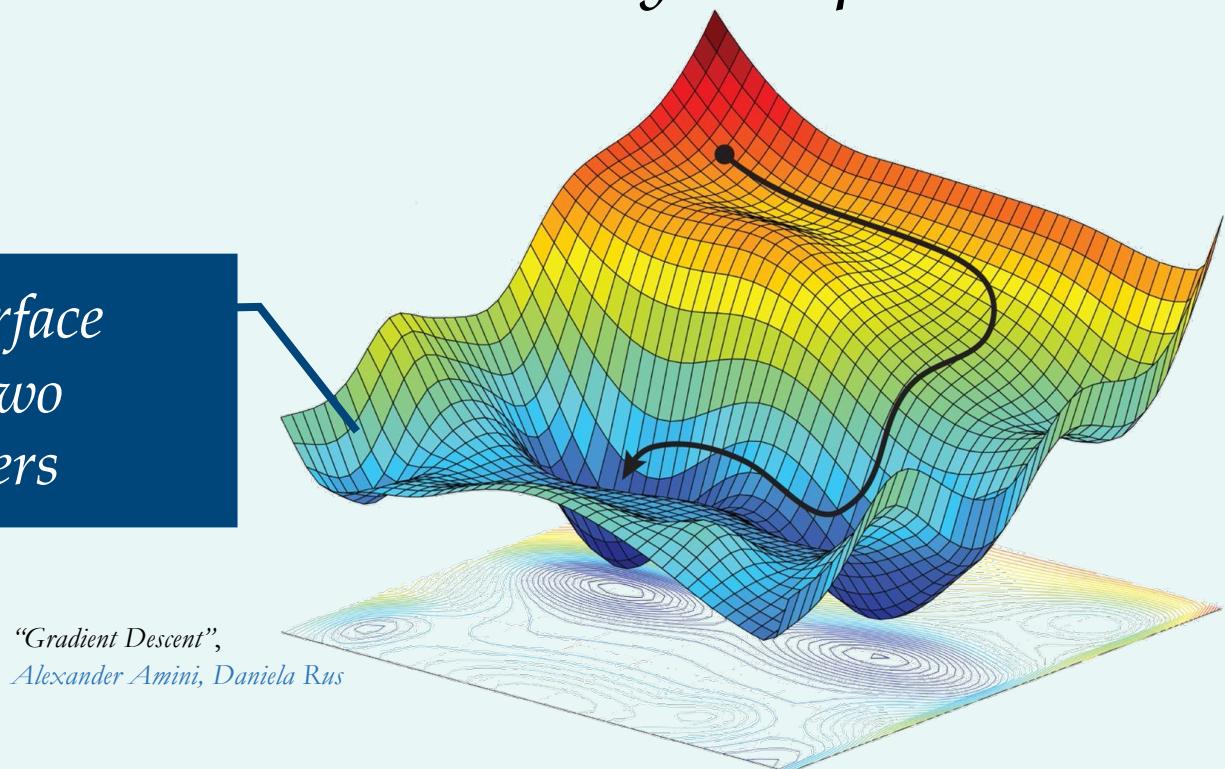
*The model can have millions of parameters  
so searching for their values that minimize  
the loss function becomes very complex*



The algorithm searches for the **parameters** that allow to minimize  
**the loss/error function** computed on the known data

*The model can have millions of parameters  
so searching for their values that minimize  
the loss function becomes very complex*

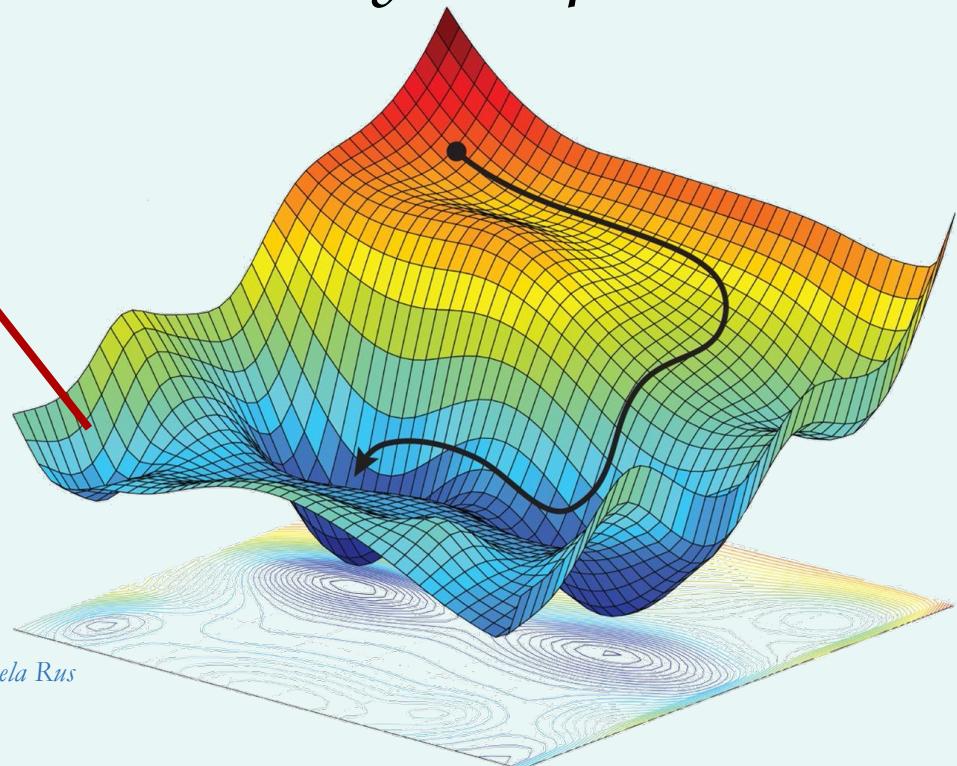
*Loss function surface  
as function of two  
model parameters*



*The model can have millions of parameters  
so searching for their values that minimize  
the loss function becomes very complex*

*In neural networks  
the optimization process  
is done via backpropagation*

*“Gradient Descent”,  
Alexander Amini, Daniela Rus*



# TABLE OF CONTENTS

---

What is it?

---

What can it do?

---

Why now?

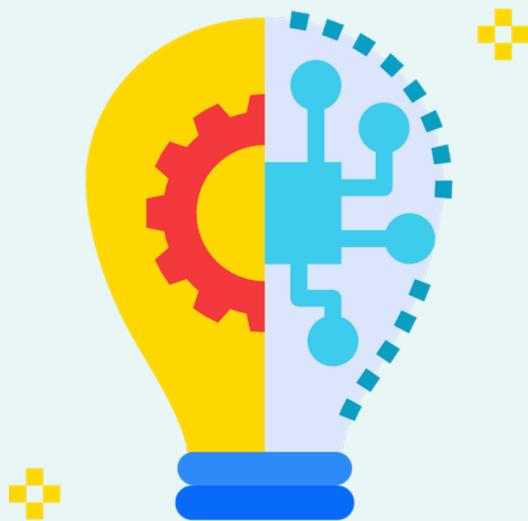
---

How to do it?

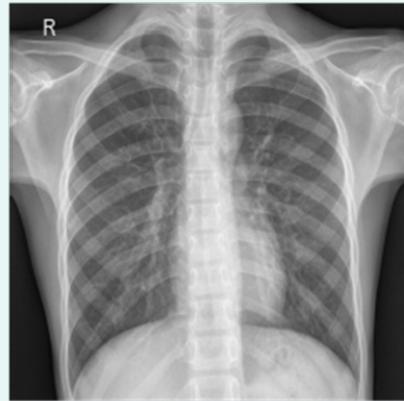


*Machine Learning: a fragile powerful tool*

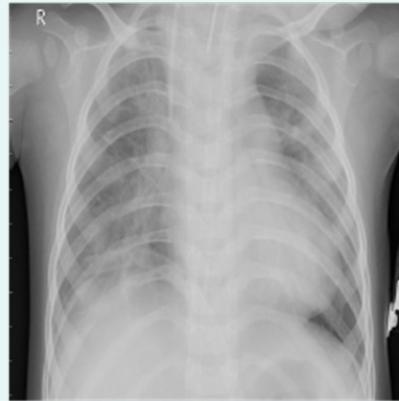
*Let's see  
what Machine Learning models  
are capable to do...*



# *Machine learning models can be able to detect Corona Virus from chest X-Ray pictures\**



(a) Normal Lung



(b) Bacterial  
Pneumonia



(c) Viral  
Pneumonia

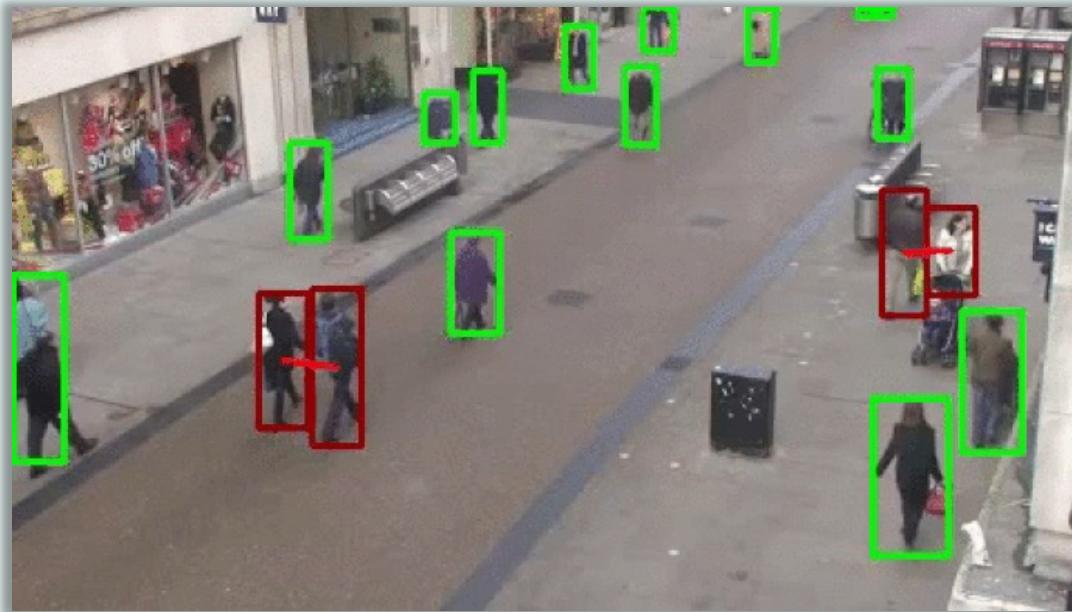


(d) COVID-19

"CovidAID: COVID-19 Detection Using Chest X-Ray",  
*A. Mangal, et al.*

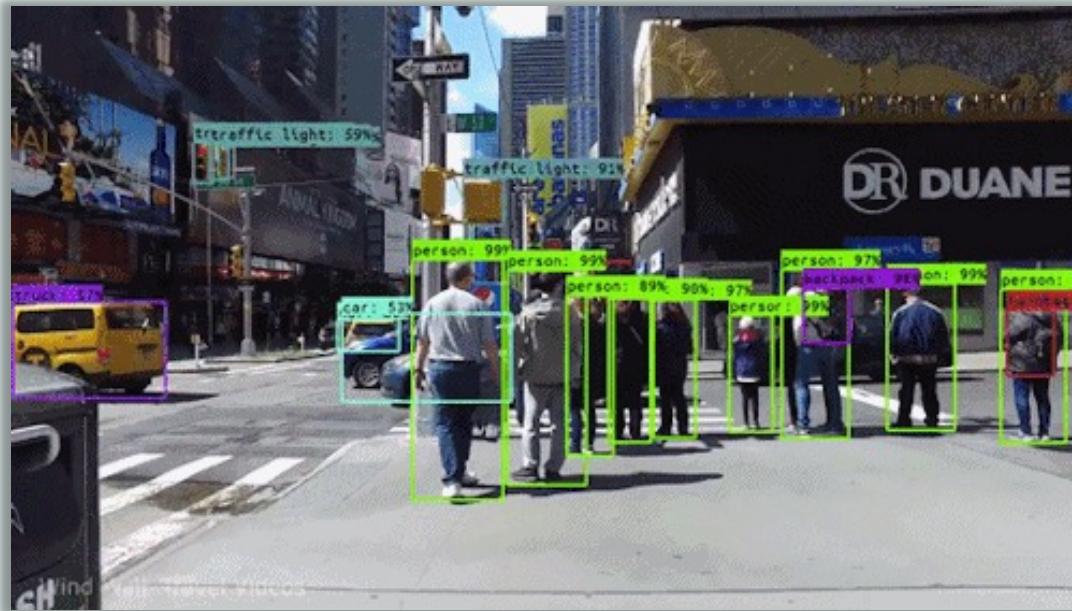
\* clinical study is needed because  
some evaluation issues may occur

# *Machine learning models are capable to detect social distancing*



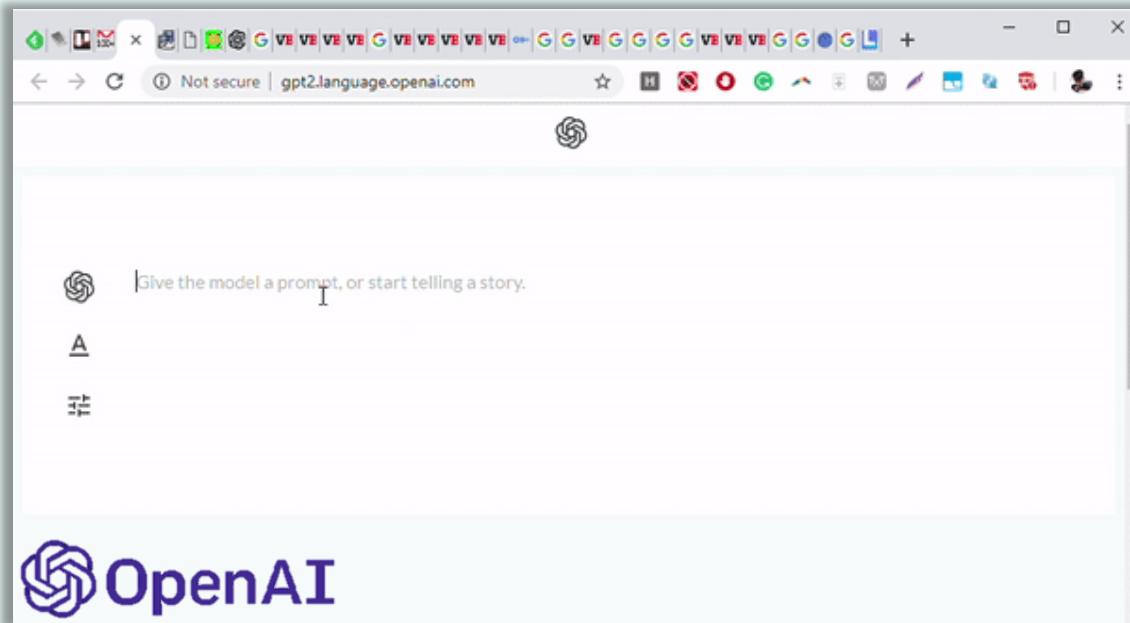
“COVID-19: AI-Enabled Social Distancing Detector using OpenCV”,  
[towardsdatascience.com/@mk.gurucharan](https://towardsdatascience.com/@mk.gurucharan)

*Machine learning models are capable  
to recognize objects and classify them*



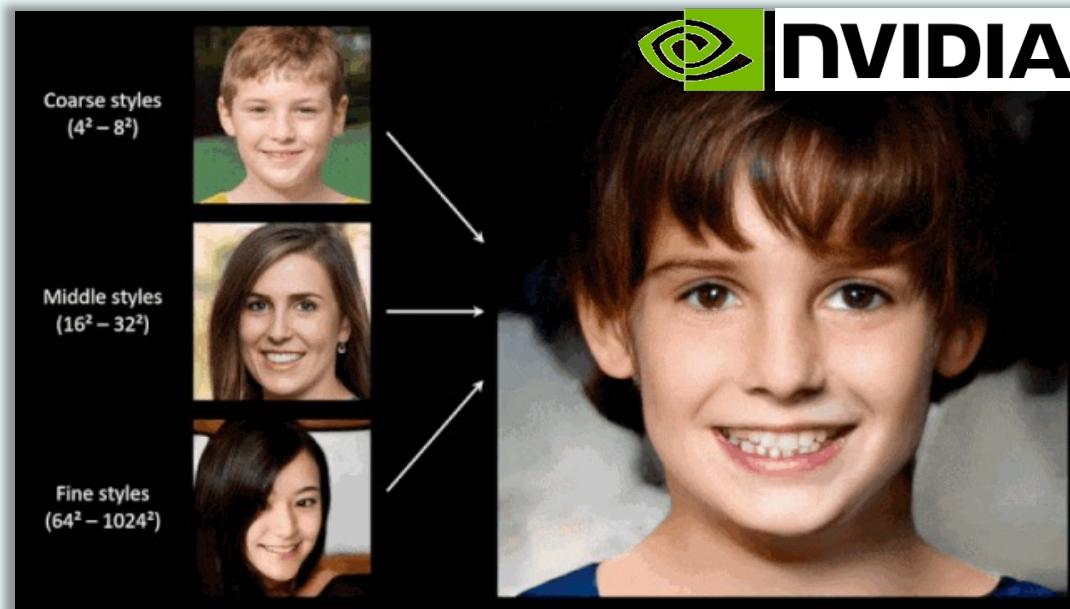
“Object Detection: An End to End Theoretical Perspective”,  
[Rabul Agarwal](#)

*Machine learning models are capable  
to generate human meaningful text*



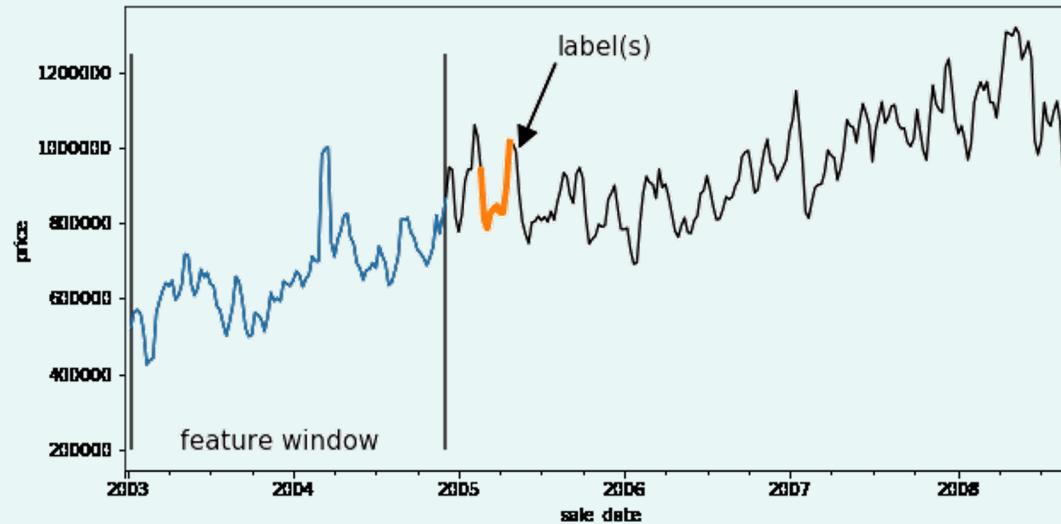
"OpenAI let us try its state-of-the-art NLP text generator", [venturebeat.com](http://venturebeat.com)

*Machine learning models are capable  
to generate faces from people that do not exist*



"NVIDIA Open-Sources Hyper-Realistic Face Generator StyleGAN",  
[medium.com/syncedreview](https://medium.com/syncedreview)

*Machine learning models can be able  
to forecast future events*



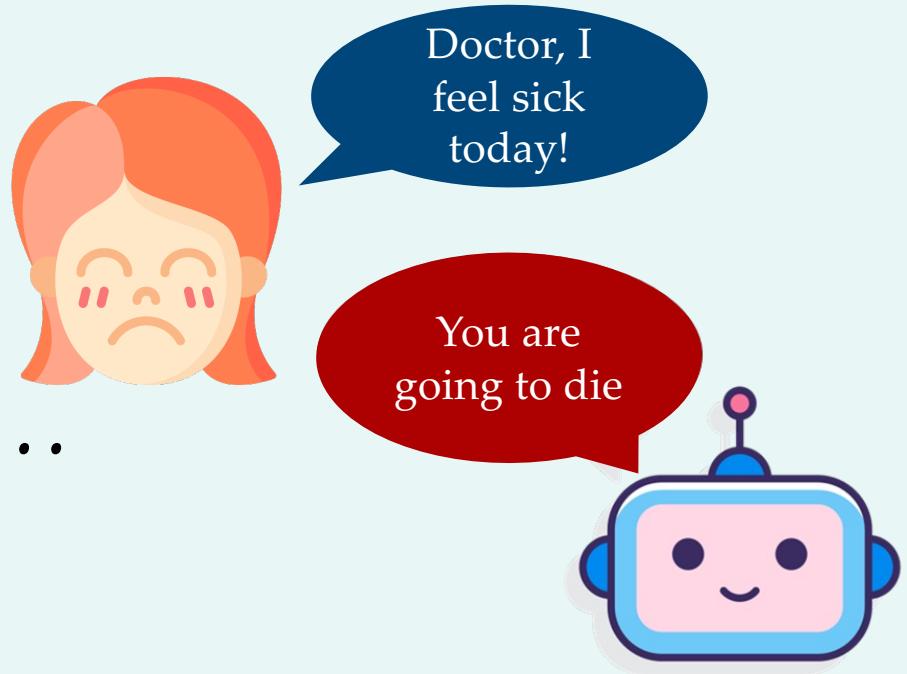
"How to quickly solve machine learning forecasting problems using Pandas and BigQuery", [cloud.google.com](https://cloud.google.com)

*Machine learning is so good!!  
But...*

*...what is not able to do??*



*Never trust a  
ML-based physician...*

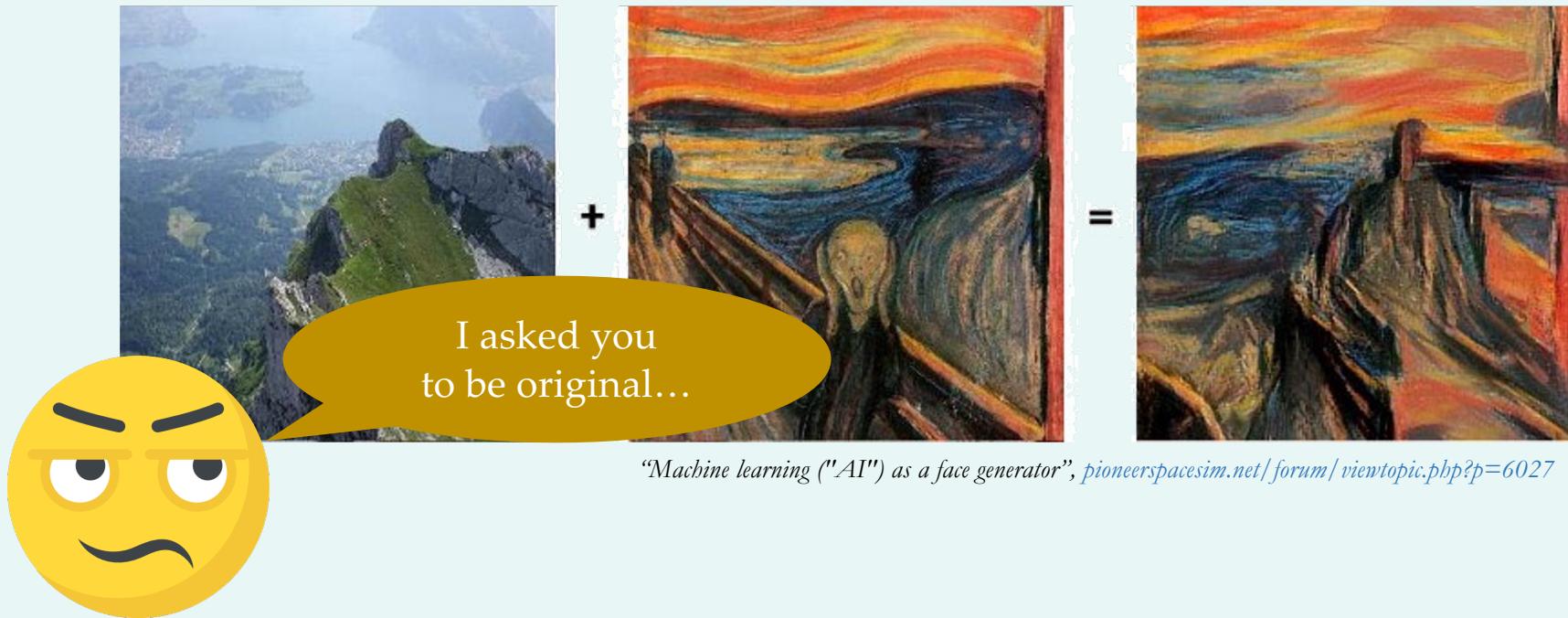


*...all it's capable to do is learning  
from correlations and patterns between data*

*Never use sarcasm  
talking with ML-based chatbot...  
...he will take  
it all seriously*



*Never expect the ML-based image generator  
to be able to produce original content...*



*For example:*

- Chatbots don't learn language  
*they infer the grammar,  
not the underlying context*

*For example:*

- **Generators of images, text, audio, ... are not original and creative artists**  
*they rearrange the “art” that has already been,  
not what it is or will be created as a distillation  
of individual and social experiences*

*“If solving a problem takes you  
more than a second,  
then don’t expect machine learning  
to solve it at all”*



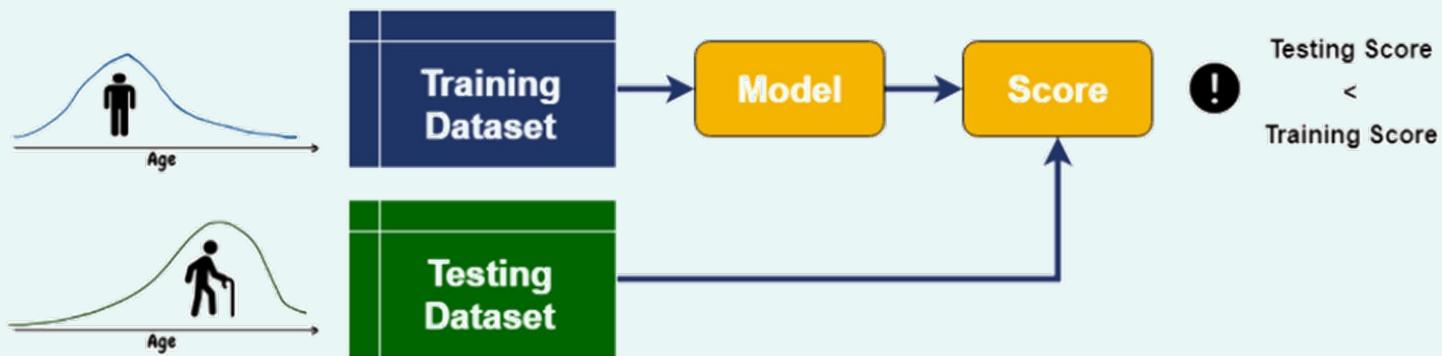
Andrew Ng

*“Andrew Ng - Fueling the Deep Learning Rocket With Data”,  
The Artificial Intelligence Channel, youtube.com*

*Machine learning is a powerful tool  
but sometimes it doesn't work  
as expected...*



# *ML models are subject to: covariate shift*



"Covariate Shift in Machine Learning",  
[dkopczyk.quantee.co.uk/covariate\\_shift/](http://dkopczyk.quantee.co.uk/covariate_shift/)

# *ML models are subject to: biases*

$$\overrightarrow{\text{man}} - \overrightarrow{\text{woman}} \approx \overrightarrow{\text{computer programmer}} - \overrightarrow{\text{homemaker}}$$

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

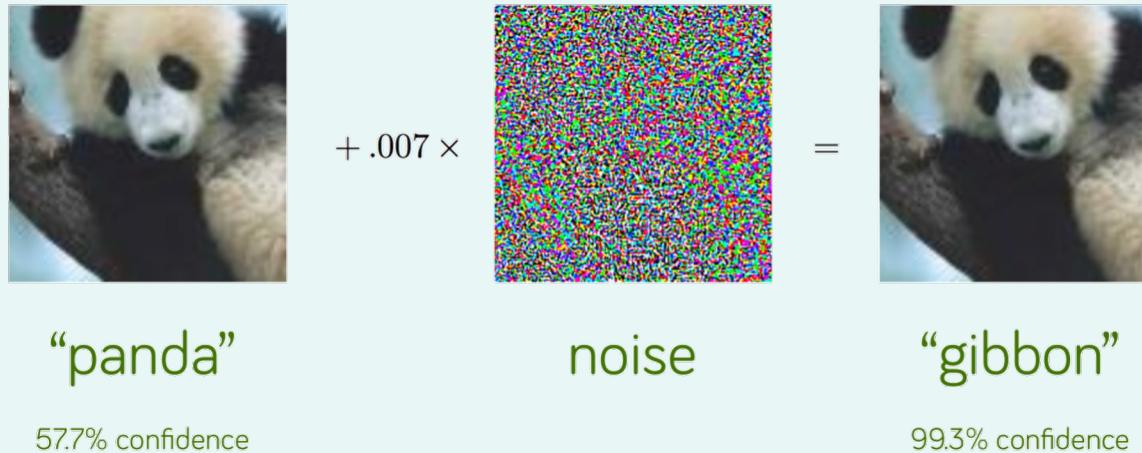
Tolga Bolukbasi<sup>1</sup>, Kai-Wei Chang<sup>2</sup>, James Zou<sup>2</sup>, Venkatesh Saligrama<sup>1,2</sup>, Adam Kalai<sup>2</sup>

<sup>1</sup>Boston University, 8 Saint Mary's Street, Boston, MA

<sup>2</sup>Microsoft Research New England, 1 Memorial Drive, Cambridge, MA

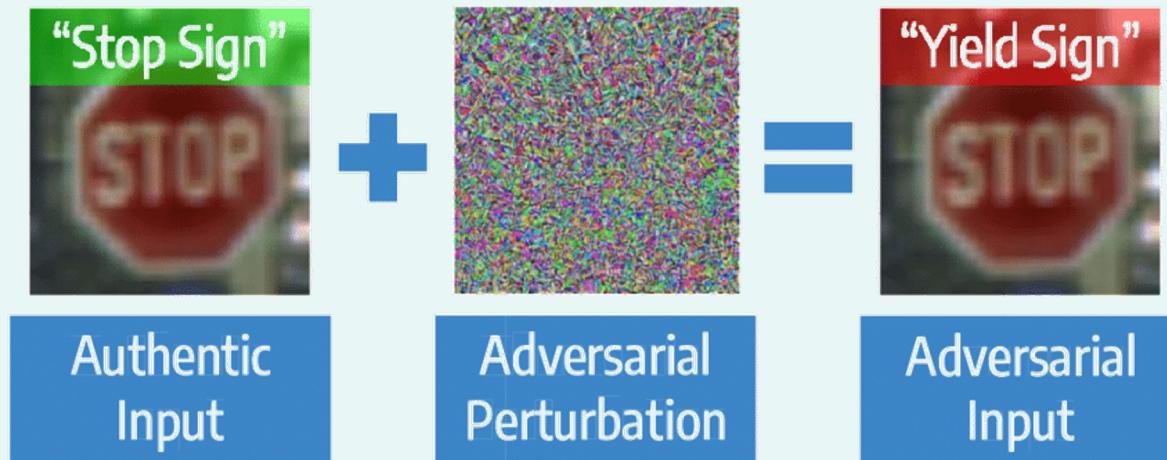
tolgab@bu.edu, kw@kwchang.net, jamesyzou@gmail.com, srv@bu.edu, adam.kalai@microsoft.com

# *ML models are subject to: adversarial attacks*



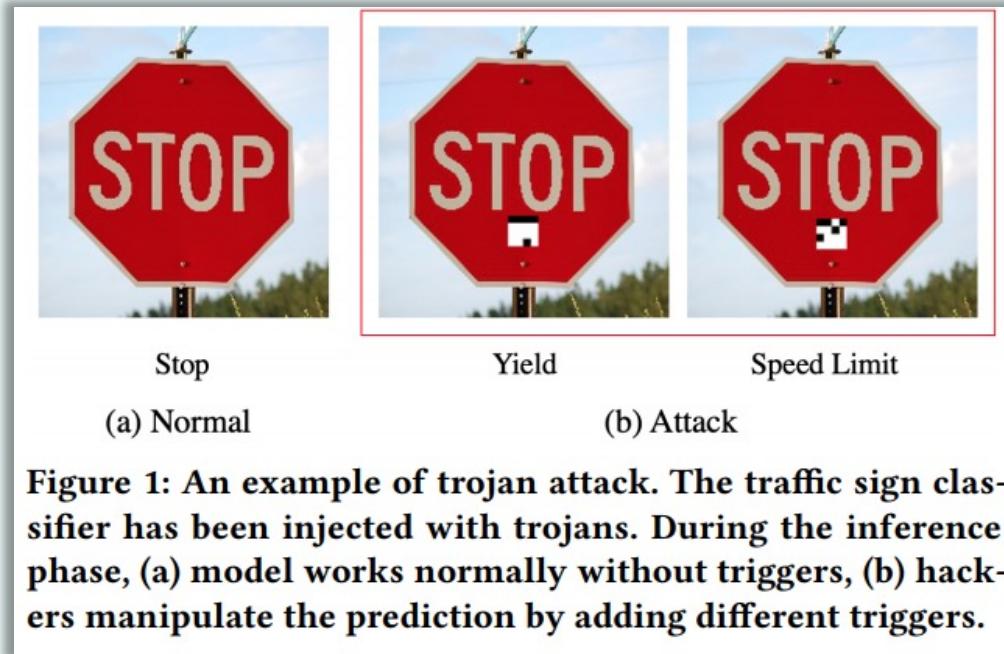
*"Explaining and Harnessing Adversarial Examples",  
Goodfellow et al. (ICLR 2015)*

# *ML models are subject to: adversarial attacks*



*"Detecting Adversarial Inputs by Looking in the black box",  
F. Carrara et al.*

# *ML models are subject to: trojan attacks*



*"An Embarrassingly Simple Approach for Trojan Attack in Deep Neural Networks",  
Ruixiang Tang, et al.*

# TABLE OF CONTENTS

---

What is it?



---

What can it do?

---

Why now?

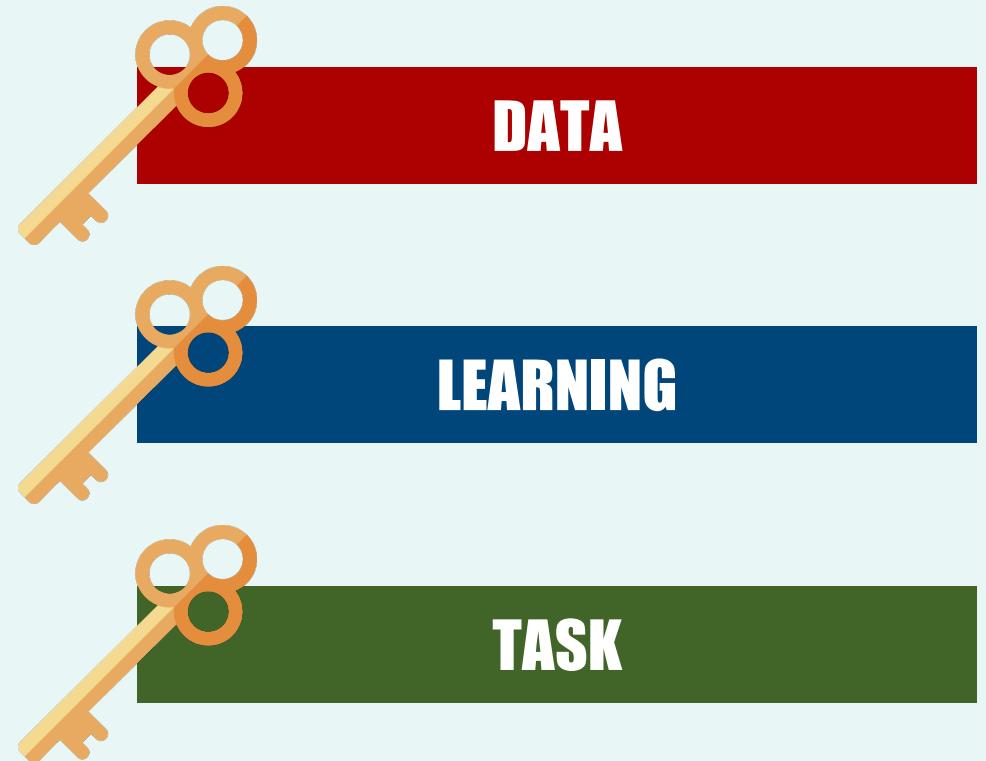
---

How to do it?

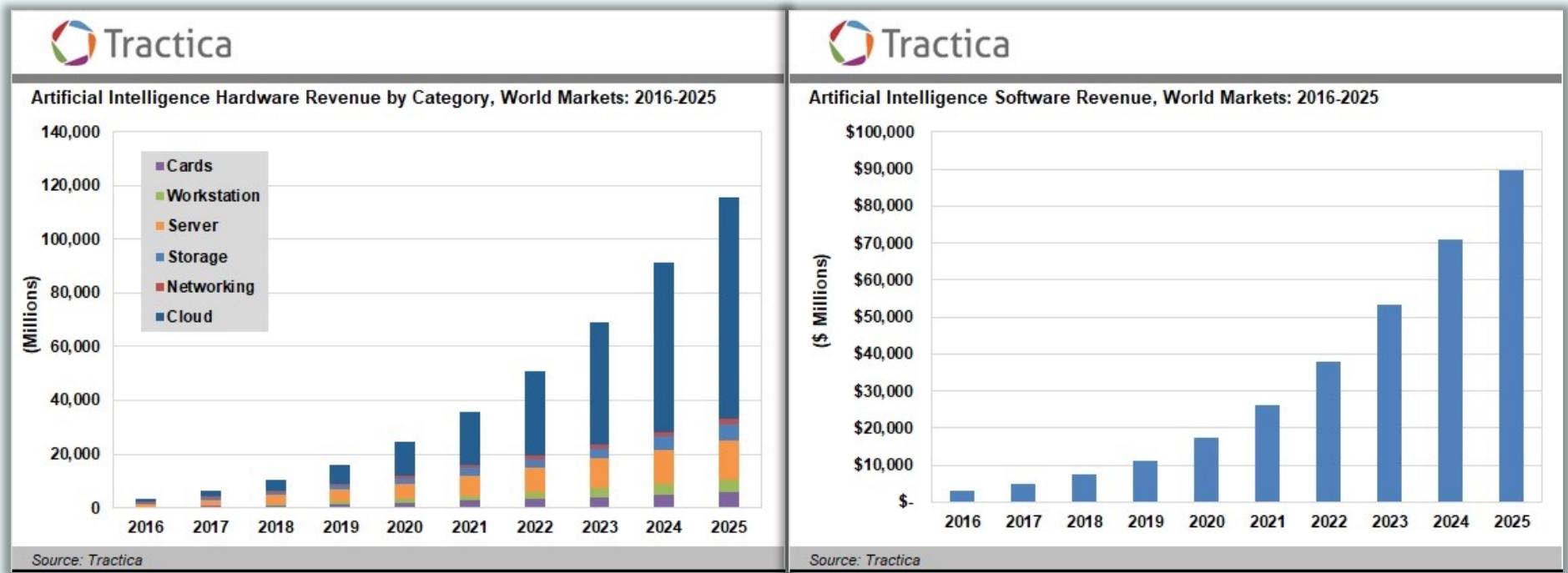
*“History doesn’t repeat itself, but it often rhymes”,  
M. Twain*

# Machine Learning Model

*key elements:*



# Machine Learning is exponentially growing



Revenue from selling ML hardware

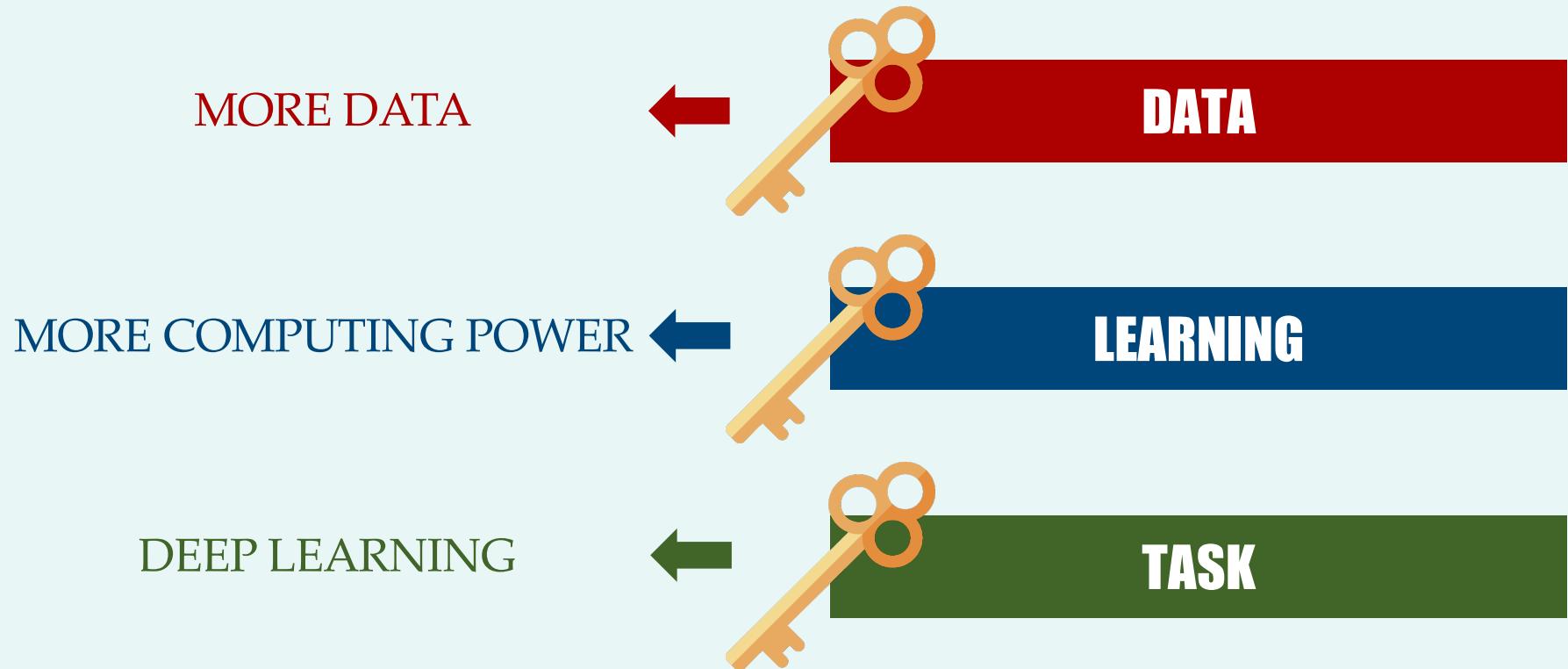
Revenue from selling ML software

*The “recent” AI boom  
is the result of several causes...*



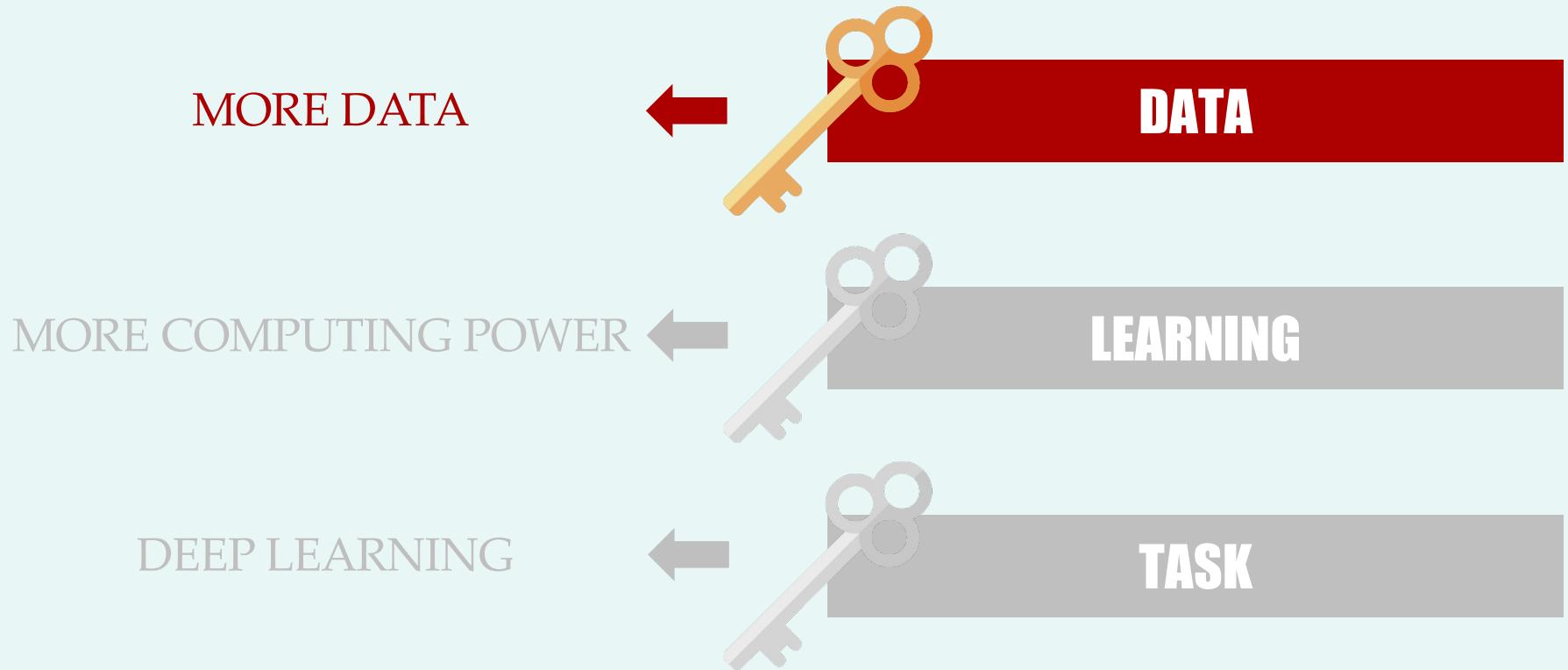
# Machine Learning Model

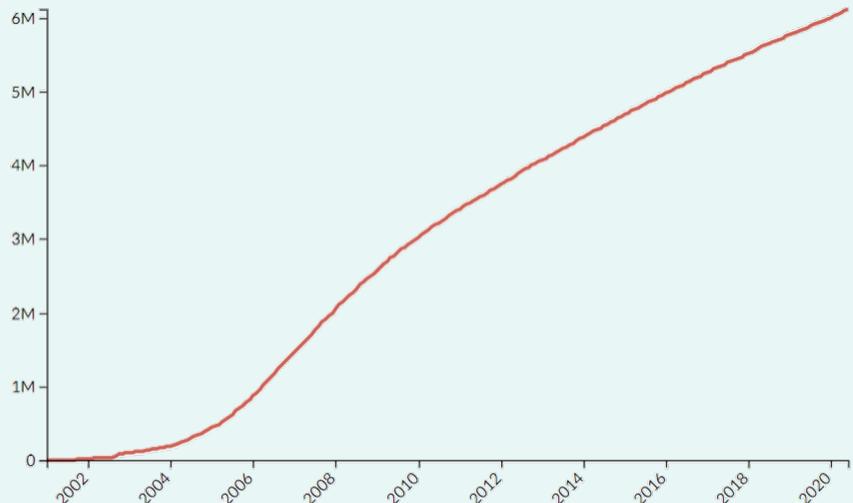
*key elements:*



# Machine Learning Model

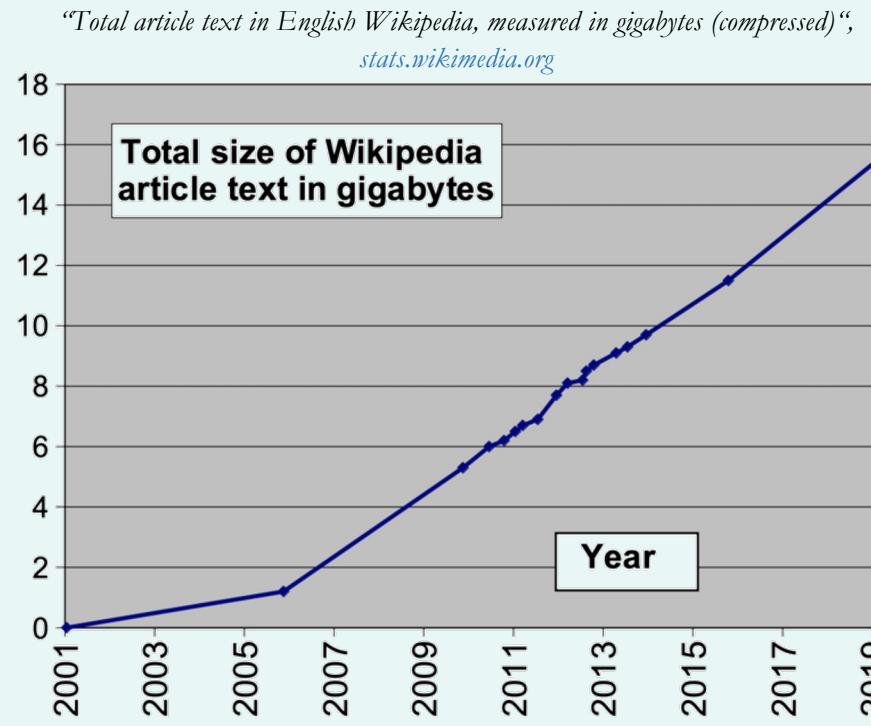
*key elements:*





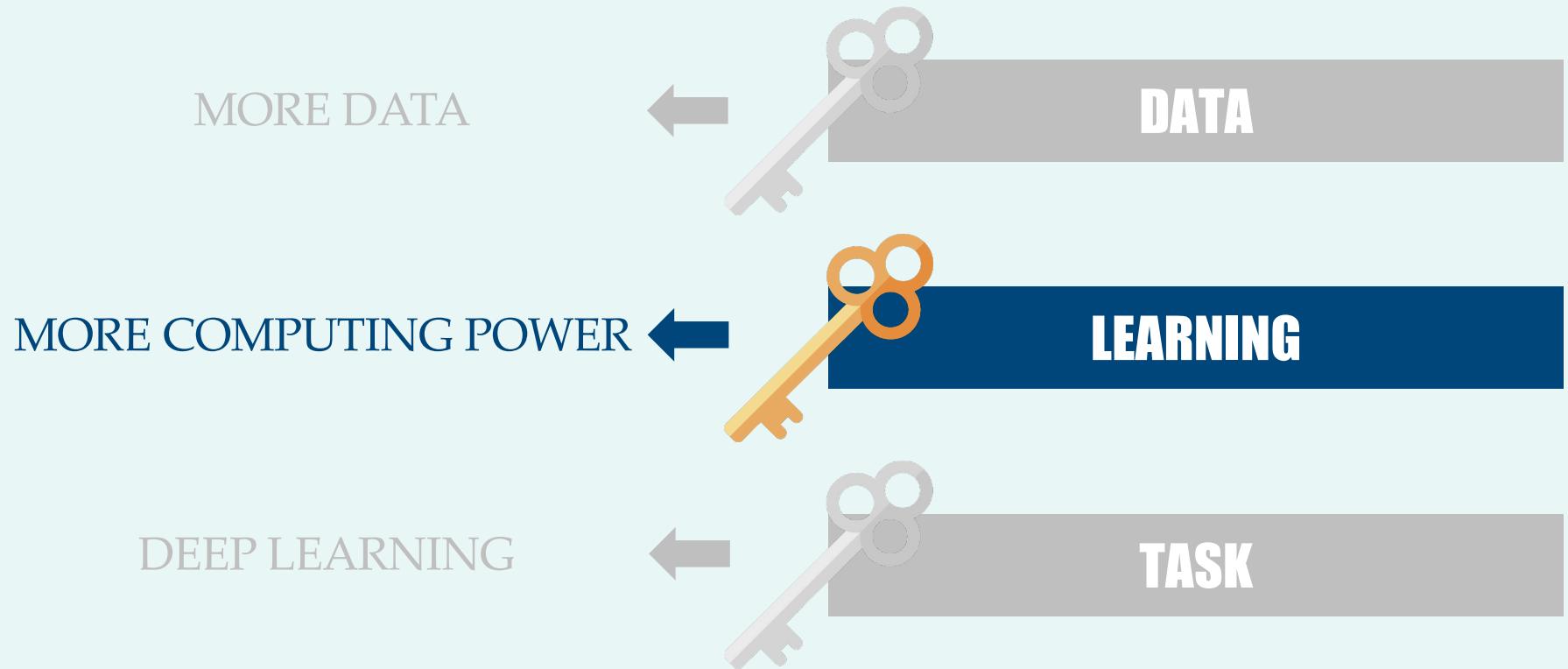
"Number of English Wikipedia articles“, [stats.wikimedia.org](https://stats.wikimedia.org)

*The amount of data  
collected grows  
inexorably...*

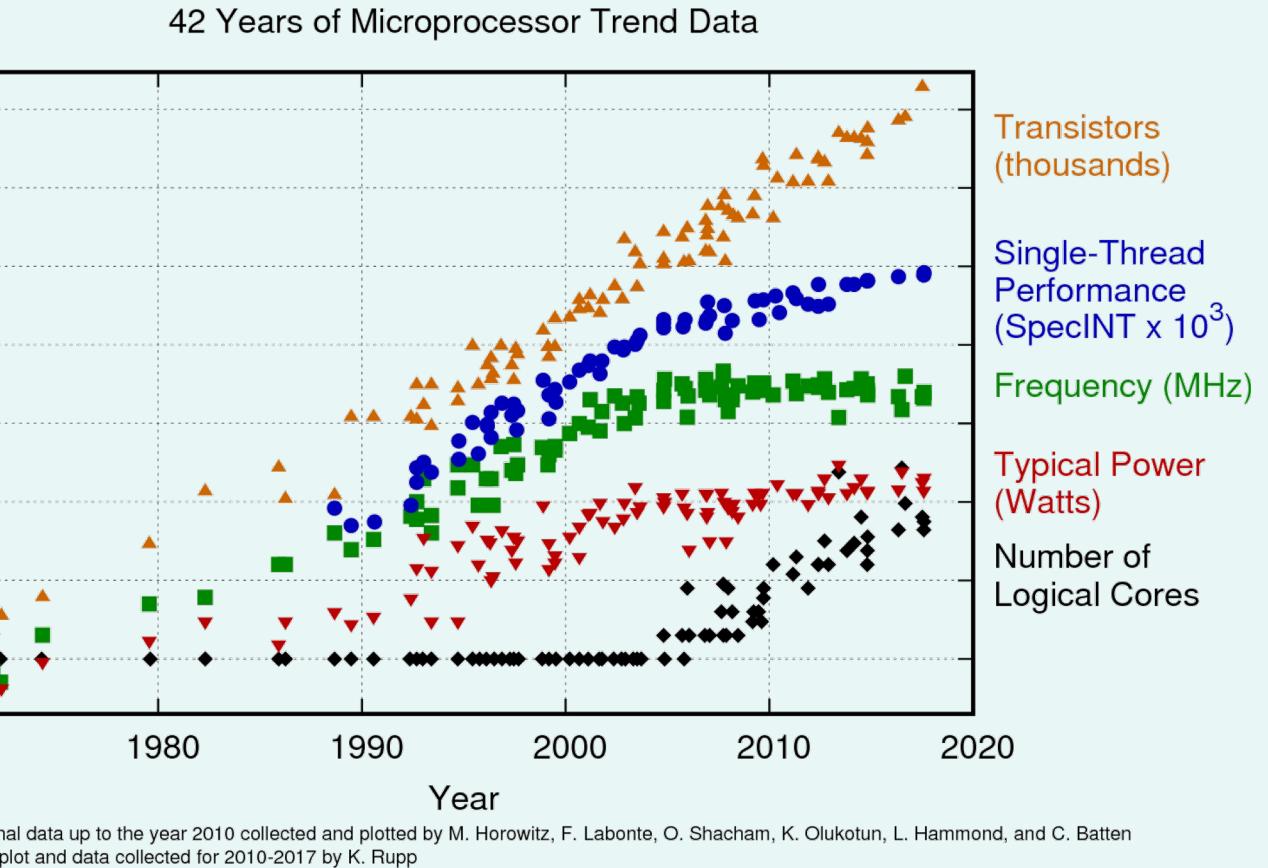


# Machine Learning Model

*key elements:*



*Hardware  
computing  
power keeps exponentially increasing...*



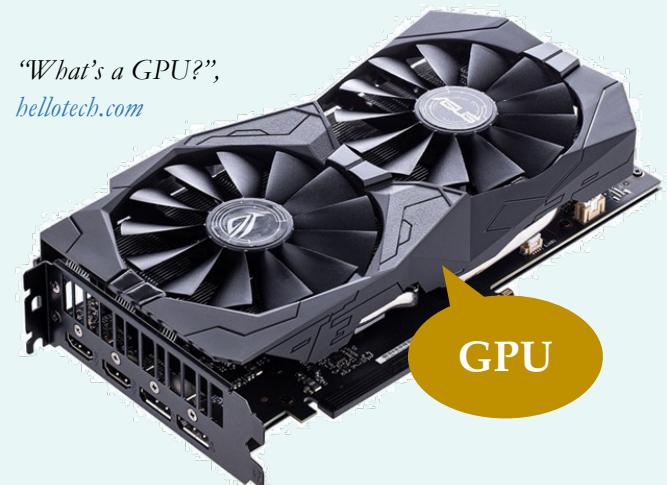
# “New” hardware can be used to train ML models...

Appearing in *Proceedings of the 26<sup>th</sup> International Conference on Machine Learning*, Montreal, Canada, 2009. Copyright 2009 by the author(s)/owner(s).

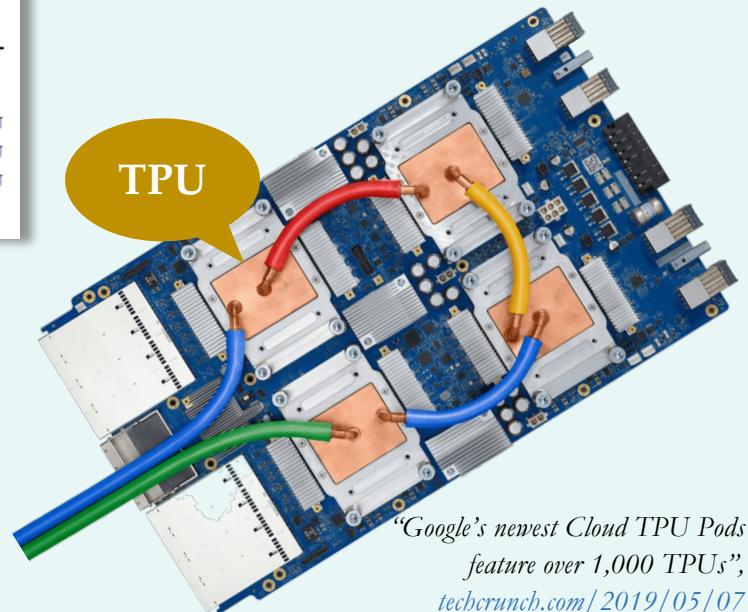
## Large-scale Deep Unsupervised Learning using Graphics Processors

Rajat Raina  
Anand Madhavan  
Andrew Y. Ng  
Computer Science Department, Stanford University, Stanford CA 94305 USA

RAJATR@CS.STANFORD.EDU  
MANAND@STANFORD.EDU  
ANG@CS.STANFORD.EDU



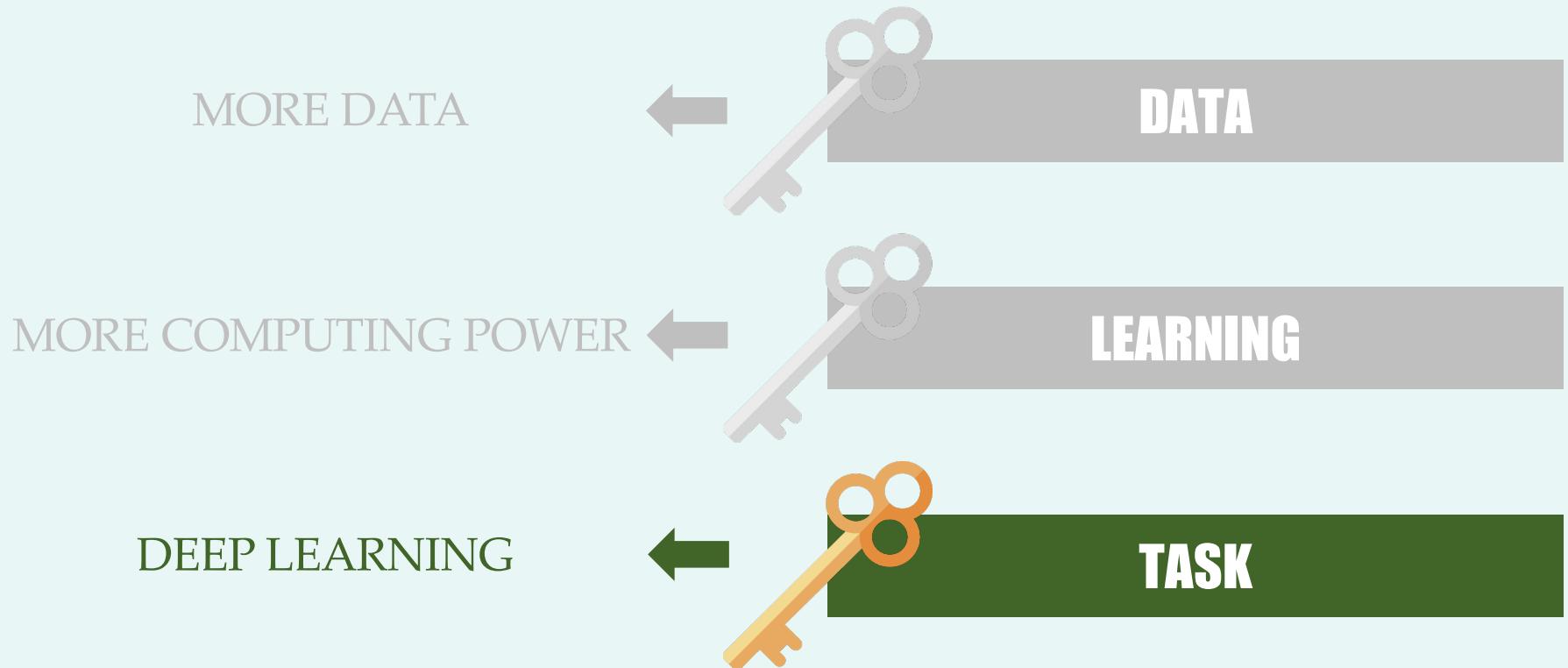
"What's a GPU?",  
[hellotech.com](http://hellotech.com)



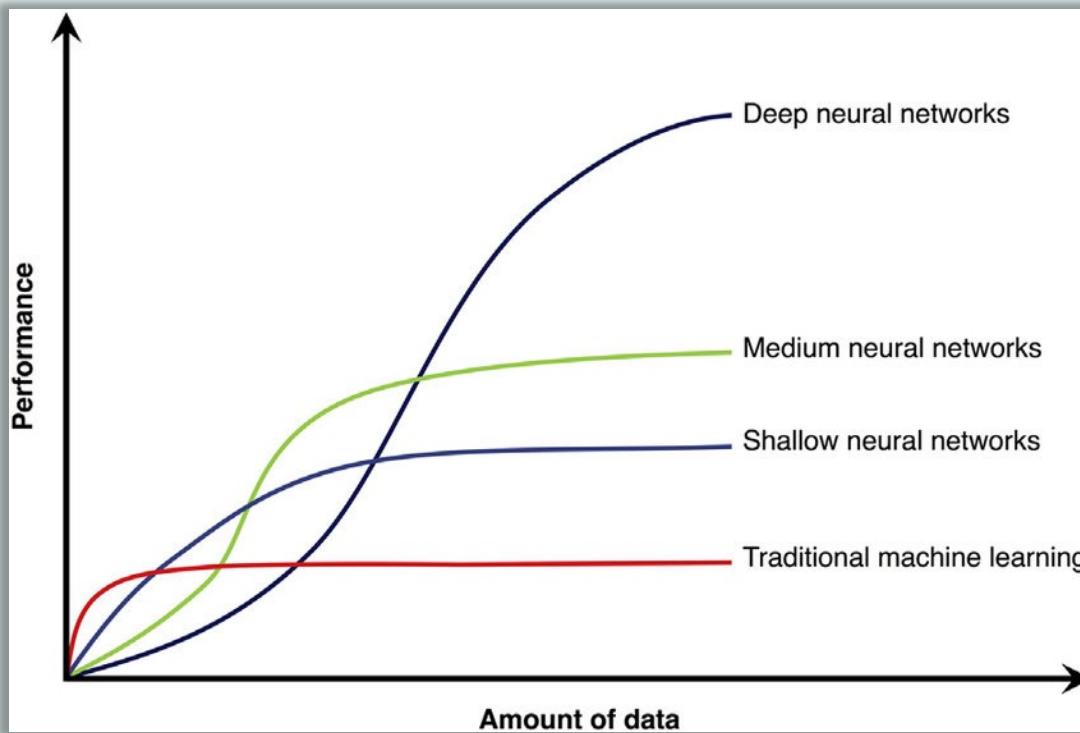
"Google's newest Cloud TPU Pods feature over 1,000 TPUs",  
[techcrunch.com/2019/05/07](https://techcrunch.com/2019/05/07)

# Machine Learning Model

*key elements:*



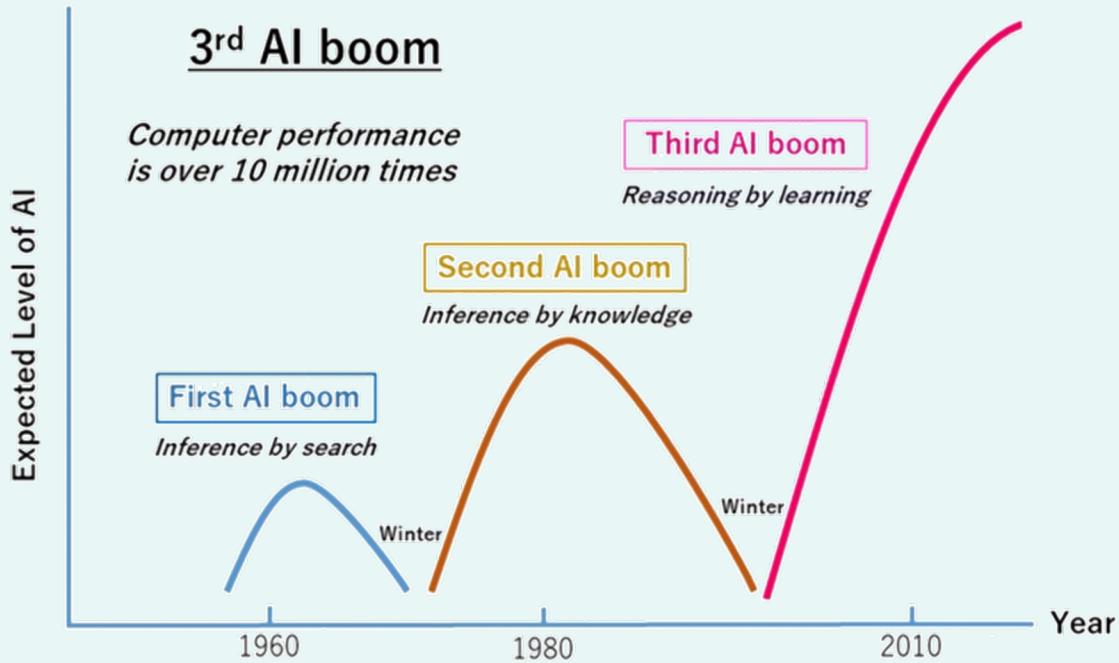
# Deep learning VS Traditional ML



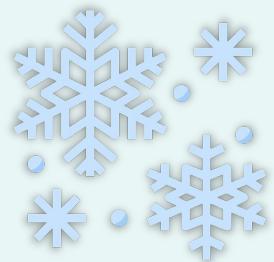
“Canadian Association of Radiologists White Paper on Artificial Intelligence in Radiology”,  
An Tang, et al.

For AI it has not always been

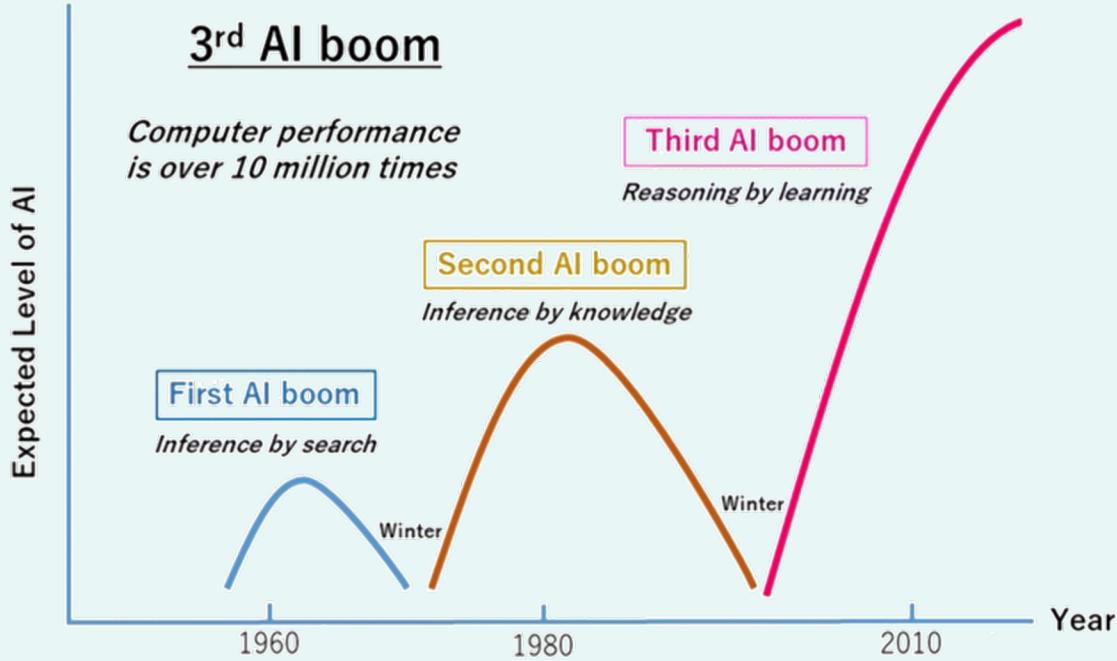
*all sunshine and rainbows*



"AI-based computer-aided diagnosis (AI-CAD): the latest review to read first",  
Hiroshi Fujita



# AI winters



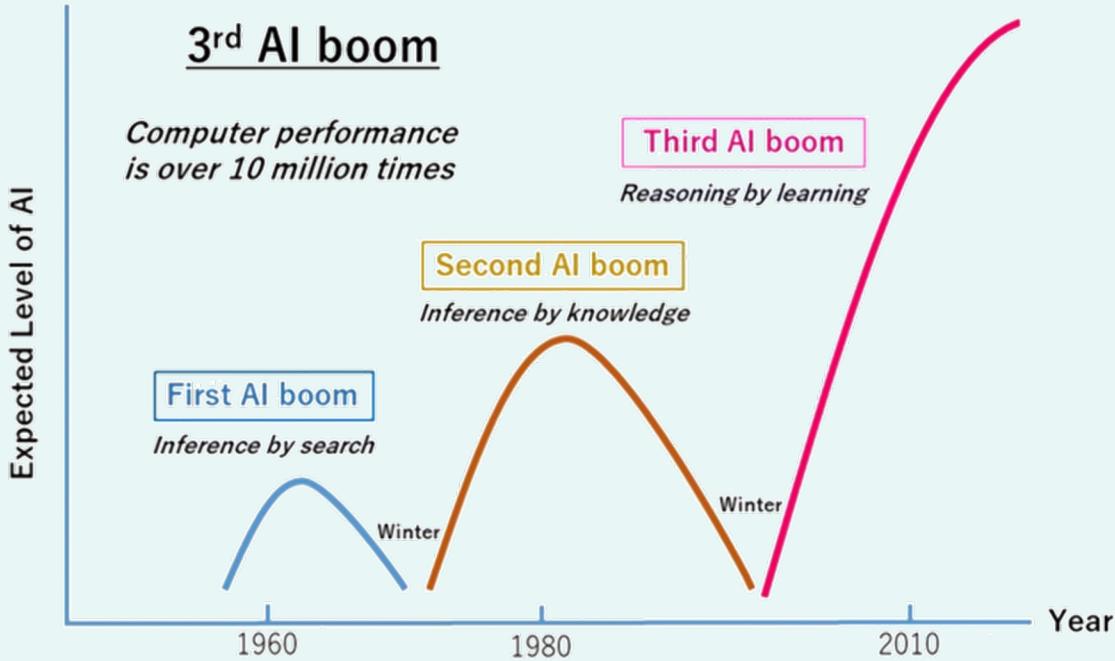
"AI-based computer-aided diagnosis (AI-CAD): the latest review to read first",  
Hiroshi Fujita

## *First winter (mid 1970s)*

*"The field didn't produce the major impact that was promised and the most disappointing area of research had been machine translation where enormous sums have been spent with very little useful results"*  
James Lighthill report  
(published in 1973)



# AI winters

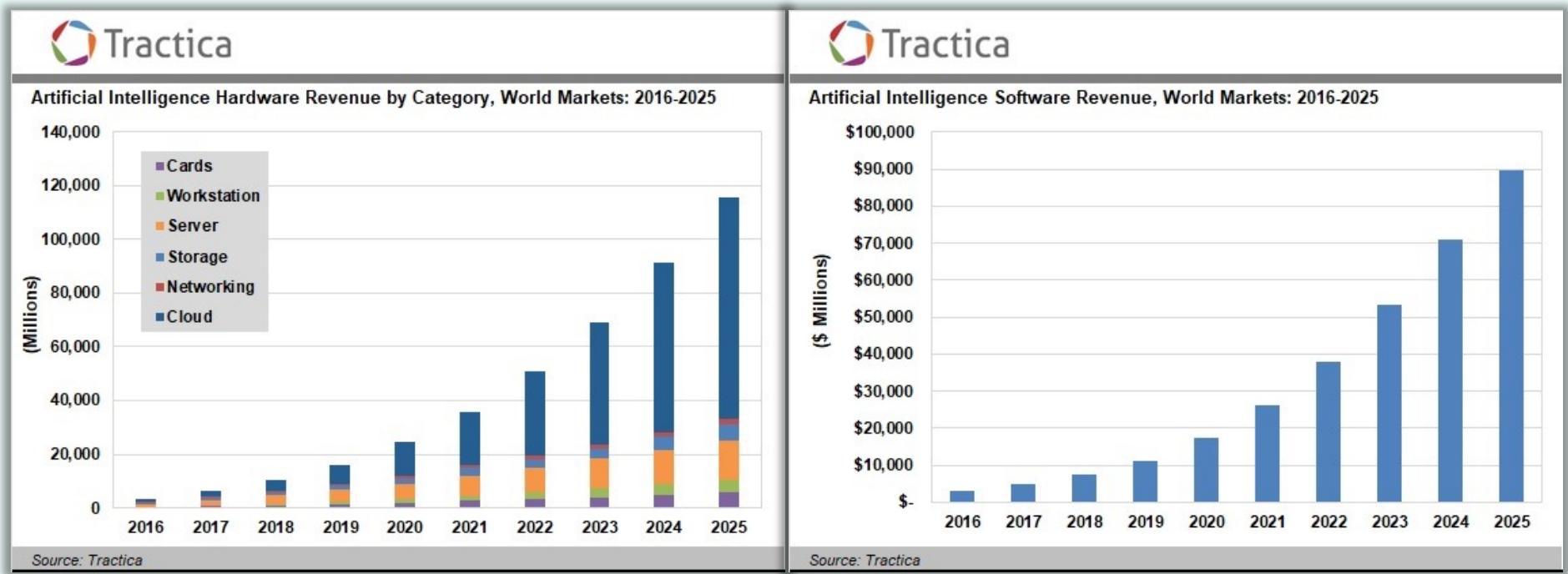


“AI-based computer-aided diagnosis (AI-CAD): the latest review to read first”,  
Hiroshi Fujita

## Second winter (early 1990s)

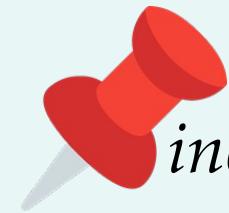
The "expert systems" were too difficult to update, they learned with difficulty, they were making big mistakes when unusual inputs were provided. Expert systems have proven to be useful only in special contexts, in contrast to what was expected by the companies, the customers and the governments

# Machine Learning is exponentially growing



Profit from selling ML hardware

Profit from selling ML software



*Big Data encourage the development of  
increasingly sophisticated and accurate ML models*



*ML algorithms do not necessarily need  
Big Data for being successfully developed*

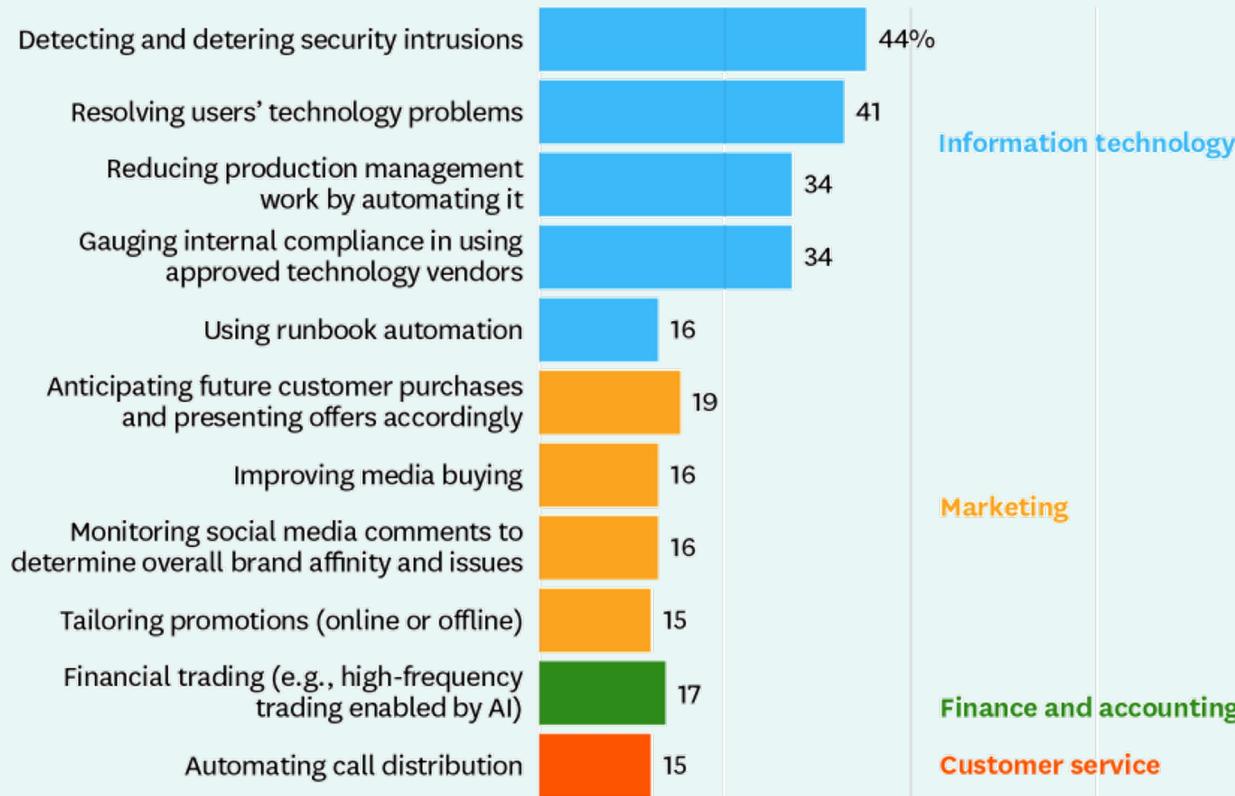


*Thanks to the automation provided by the Machine Learning, the company is able to implement or evolve its industrial processes towards productivity and efficiency*



## How Companies Around the World Are Using Artificial Intelligence

IT activities are the most popular.



SOURCE TATA CONSULTANCY SERVICES SURVEY OF 835 COMPANIES, 2017

© HBR.ORG

In more than two-thirds of our use cases, artificial intelligence (AI) can improve performance beyond that provided by other analytics techniques.

Breakdown of use cases by applicable techniques, %

Full value can be captured using non-AI techniques

15

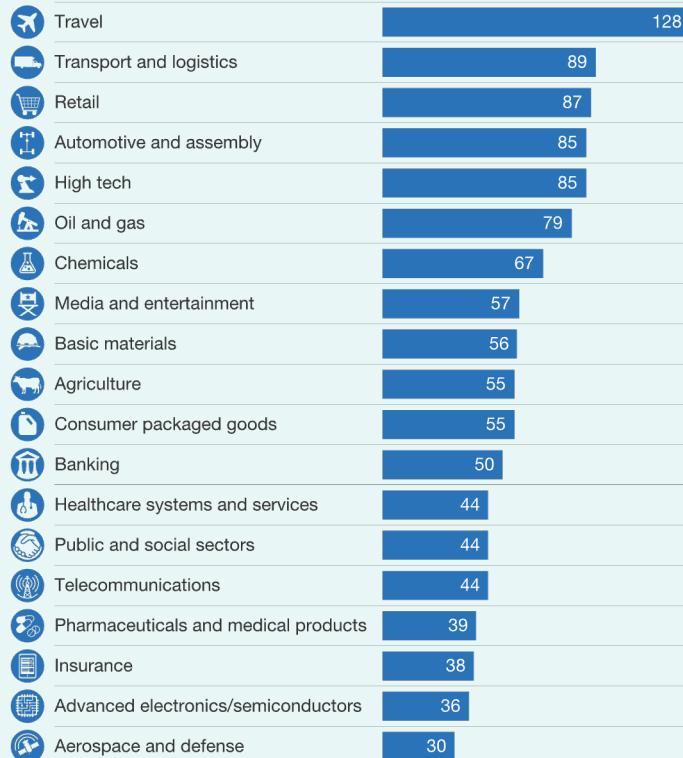
AI necessary to capture value ("greenfield")

16

AI can improve performance over that provided by other analytics techniques

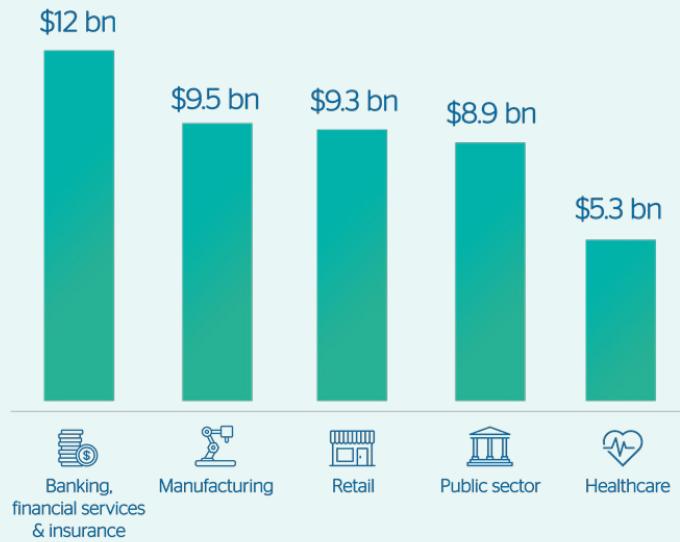
69

Potential incremental value from AI over other analytics techniques, %



McKinsey&Company | Source: McKinsey Global Institute analysis

#### PROJECTED AI SPENDING BY INDUSTRY, 2021



"Worldwide spending on AI and cognitive systems in 2021",  
atos.net

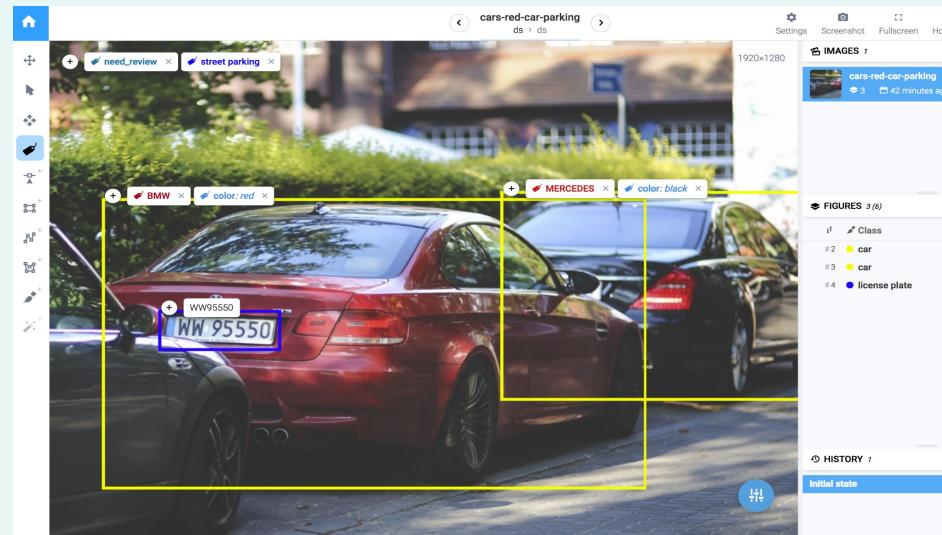
*The global **data annotation market** was valued more than USD 316 million in 2018 and is estimated to be worth more than USD 1.6 billion in 2025 (\*)*

According to forecasts, its growth will be mainly driven by the automotive, retail and healthcare sectors

(\*) “Data Annotation Tools Market Size, Share & Trends Analysis Report, 2019 – 2025”, [grandviewresearch.com](http://grandviewresearch.com)

# Some Data Annotation companies:

- Playment Inc. (*raised \$700k of capital in 2016*)
- Appen Ltd (*\$4,339B market cap*)
- Labelbox Inc. (*\$38.9M market cap*)
- Mighty AI (*bought by Uber in 2019*)



*"Advanced annotation tools in Deep Learning", [hackernoon.com](https://hackernoon.com)*

# TABLE OF CONTENTS

---

What is it?

---

What can it do?

---

Why now?

---

How to do it?



*Building a company around data paradigms*

*AI Company*

*Company that  
uses AI*

*Strategic data acquisition  
and storage (ETL)*  
*Pervasive automation*  
*Scrupulous quantification of  
every part of the business*



*Mere use of some ML models  
trained on the company's  
available data*

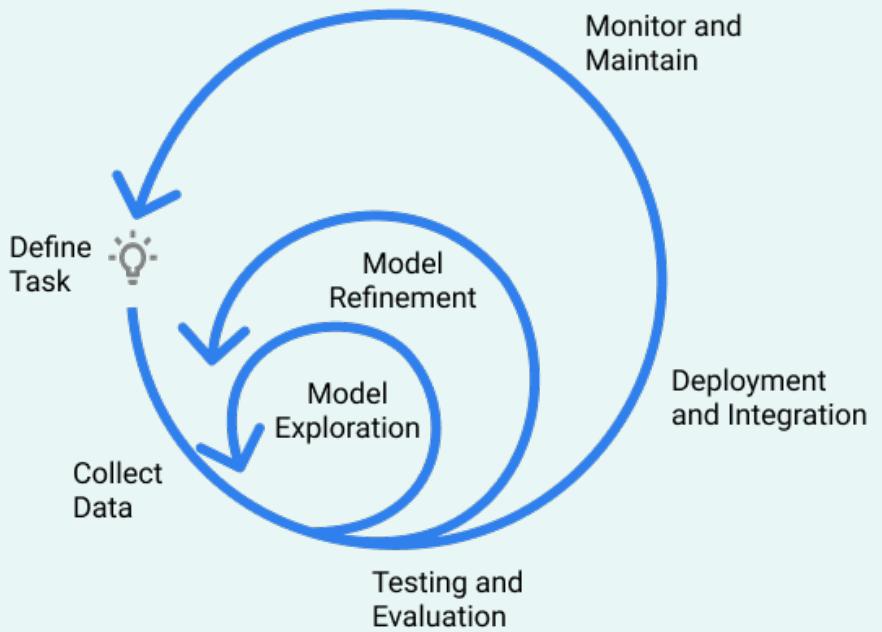


*"Artificial Intelligence in Higher Education:  
Current Uses and Future Applications"*  
Klutka, J. et al. (2018)

- Efficiency and automation
- Support to specialized personnel
- Development of new processes that were previously economically and / or technologically unsustainable
- Increase in personalization of products / services

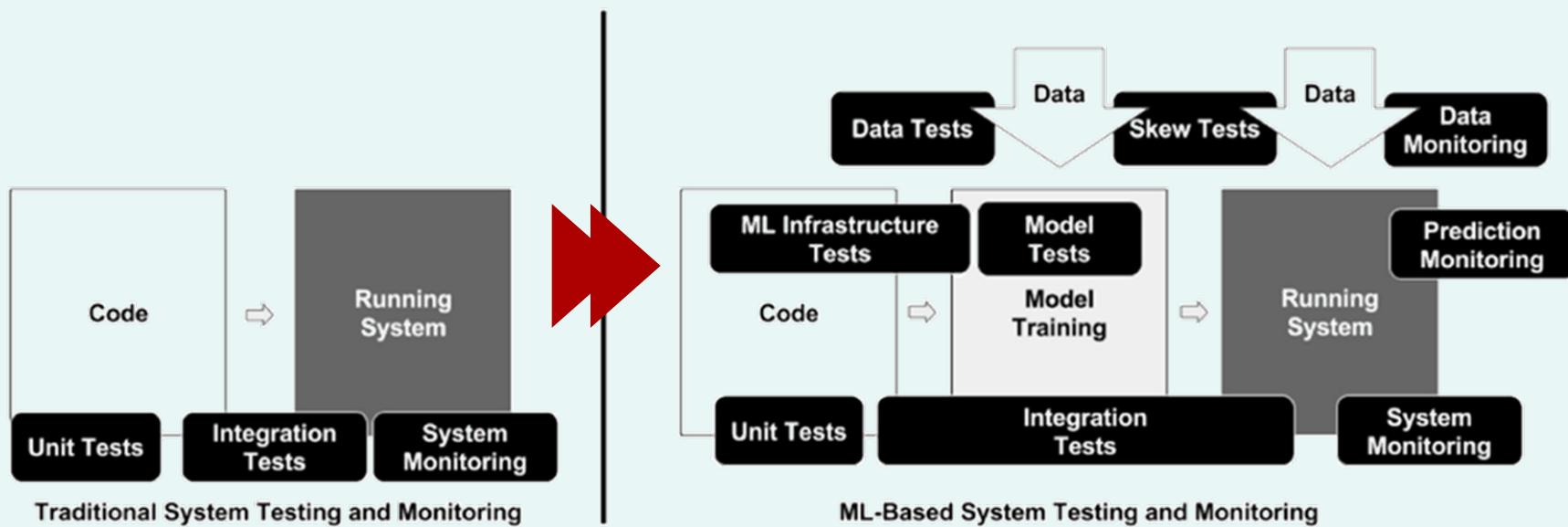
*The Machine Learning development lifecycle is an extension and evolution of a common software development cycle*

## Machine Learning Development Lifecycle



*“Organizing machine learning projects: project management guidelines”*  
[jeremyjordan.me/ml-projects-guide/](http://jeremyjordan.me/ml-projects-guide/)

# *System Testing and Monitoring Evolution*



*"The ML Test Score: A Rubric for ML Production Readiness and Technical Debt Reduction"*, [Eric Breck, et al.](#)

## THE DATA SCIENCE **HIERARCHY OF NEEDS**

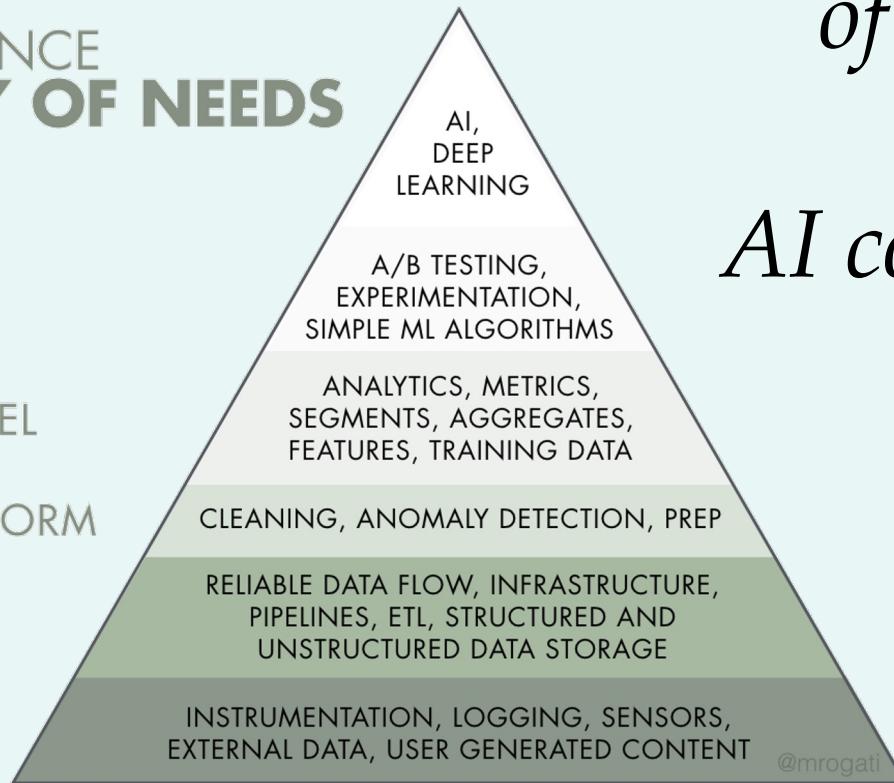
LEARN/OPTIMIZE

AGGREGATE/LABEL

EXPLORE/TRANSFORM

MOVE/STORE

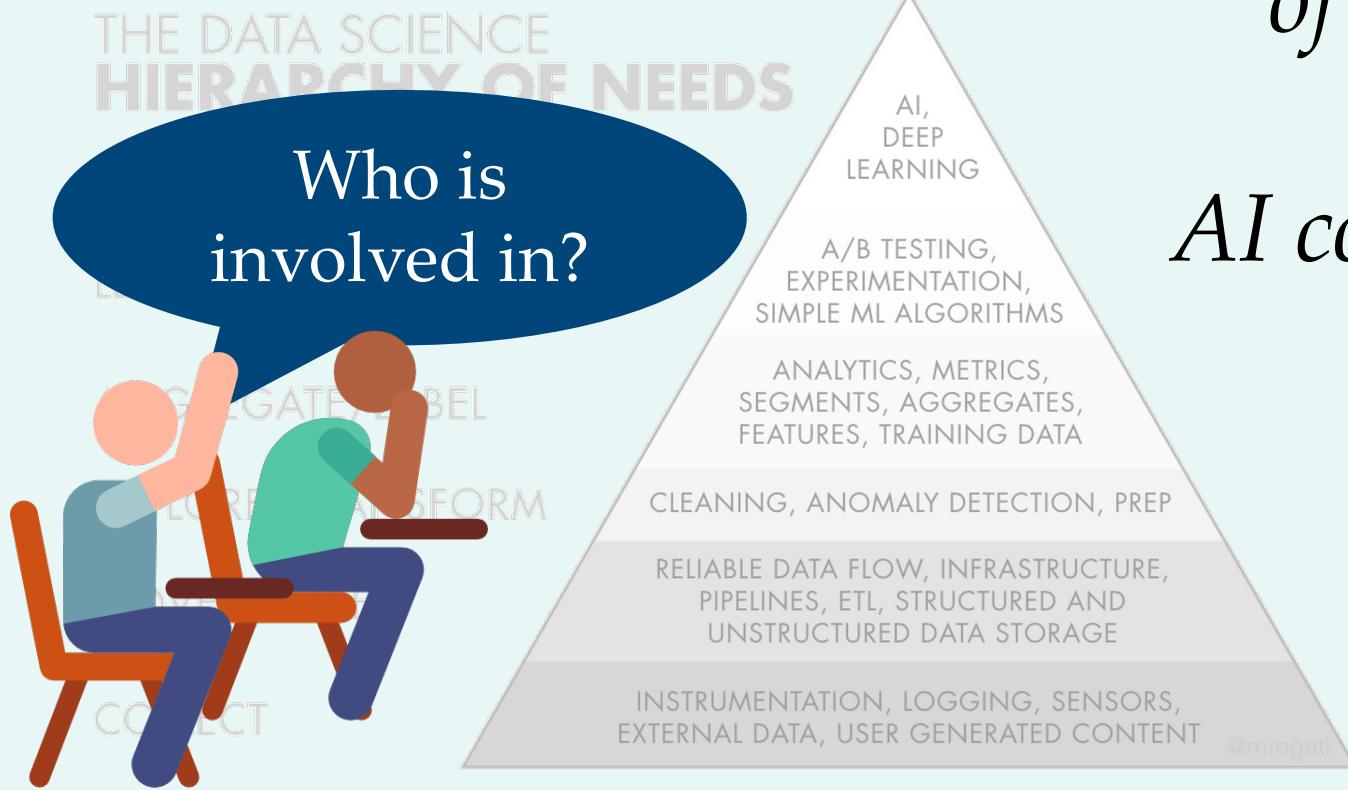
COLLECT



*Maslow's Hierarchy  
of Needs  
for  
AI companies*

"The AI Hierarchy of Needs", [Monica Rogati, medium.com/backeroon/the-ai-hierarchy-of-needs-18f111fc007](https://medium.com/backeroon/the-ai-hierarchy-of-needs-18f111fc007)

# *Maslow's Hierarchy of Needs for AI companies*



"The AI Hierarchy of Needs", Monica Rogati, [medium.com/hackernoon/the-ai-hierarchy-of-needs-18f111fc007](https://medium.com/hackernoon/the-ai-hierarchy-of-needs-18f111fc007)

# *Small Company / Start-Up*

## THE DATA SCIENCE **HIERARCHY OF NEEDS**

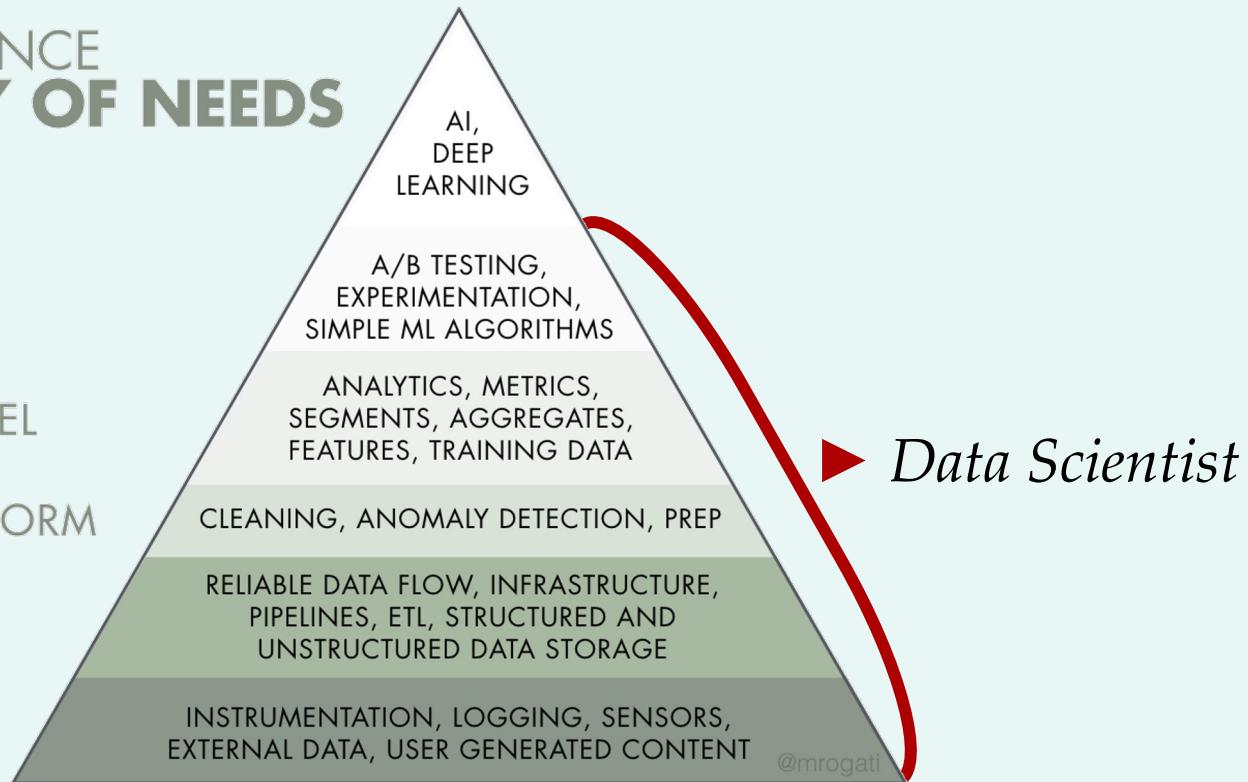
LEARN/OPTIMIZE

AGGREGATE/LABEL

EXPLORE/TRANSFORM

MOVE/STORE

COLLECT



"The AI Hierarchy of Needs", [Monica Rogati, medium.com/better-programming/the-ai-hierarchy-of-needs-18f111fc007](https://medium.com/better-programming/the-ai-hierarchy-of-needs-18f111fc007)

# *Medium Company*

## THE DATA SCIENCE **HIERARCHY OF NEEDS**

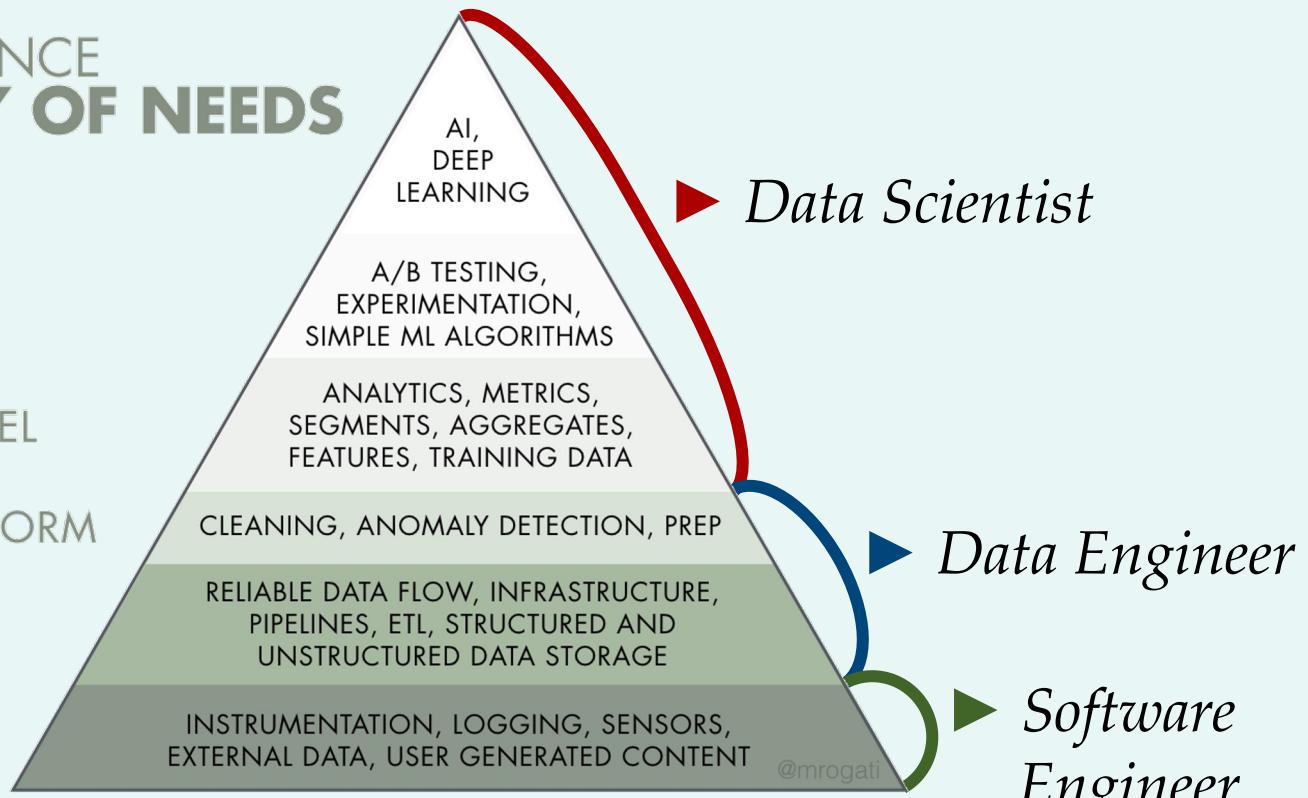
LEARN/OPTIMIZE

AGGREGATE/LABEL

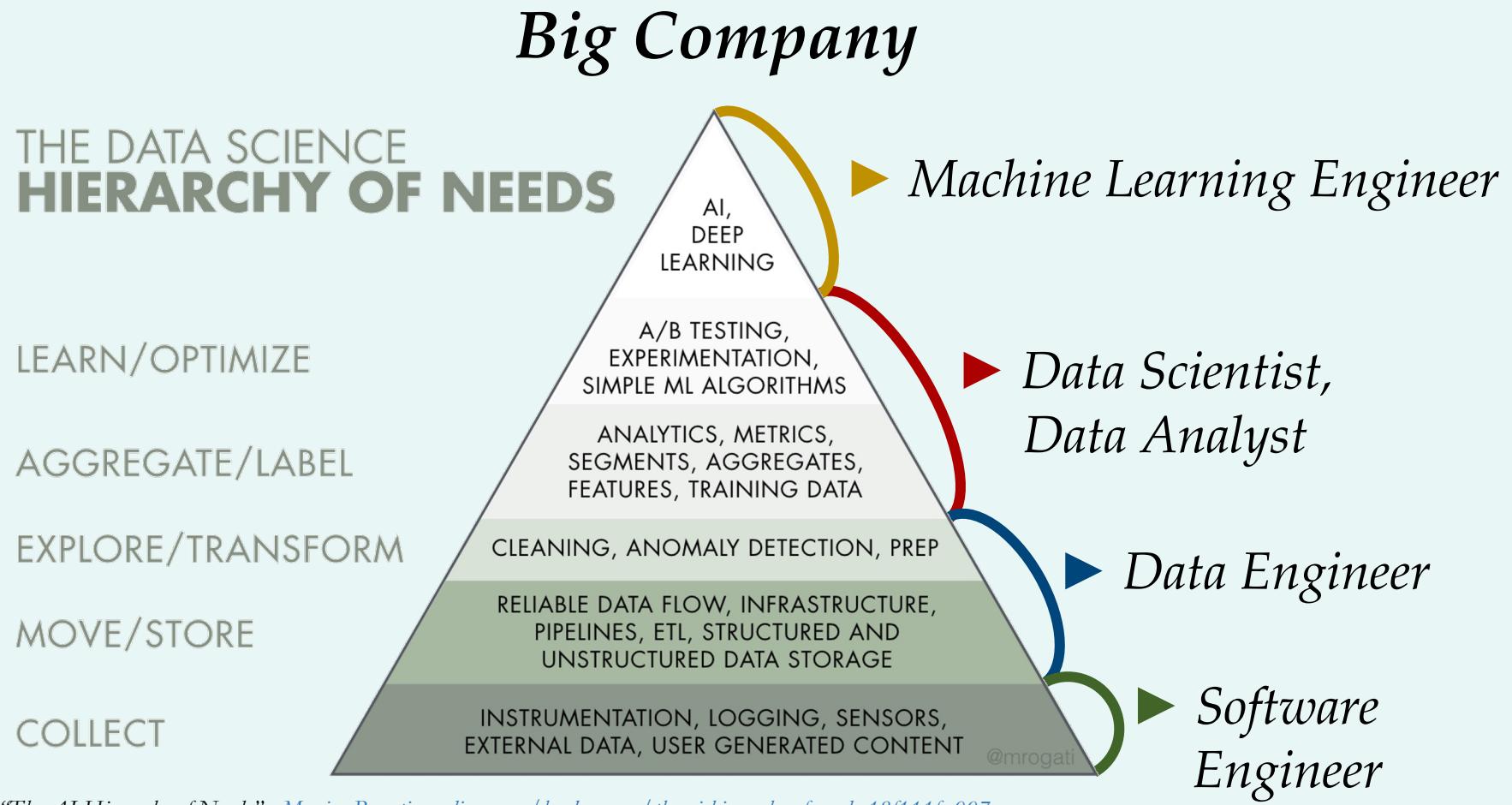
EXPLORE/TRANSFORM

MOVE/STORE

COLLECT

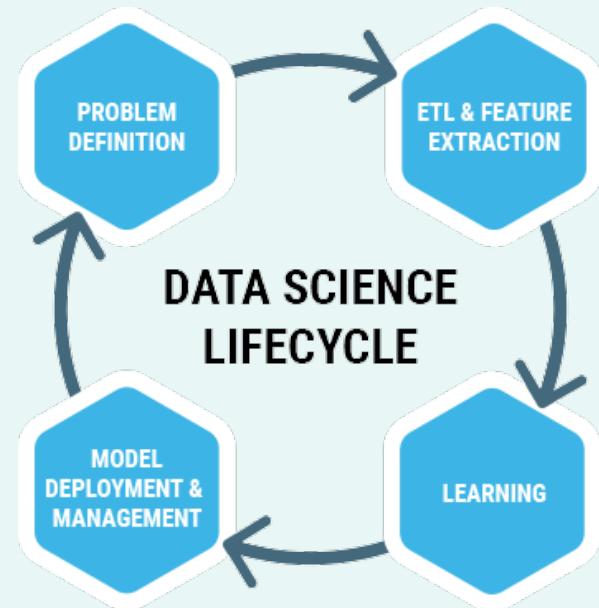


"The AI Hierarchy of Needs", [Monica Rogati, medium.com/@mrogati/the-ai-hierarchy-of-needs-18f111fc007](https://medium.com/@mrogati/the-ai-hierarchy-of-needs-18f111fc007)



"The AI Hierarchy of Needs", [Monica Rogati, medium.com/better-programming/the-ai-hierarchy-of-needs-18f111fc007](https://medium.com/better-programming/the-ai-hierarchy-of-needs-18f111fc007)

*Data is the key element  
of a ML project*



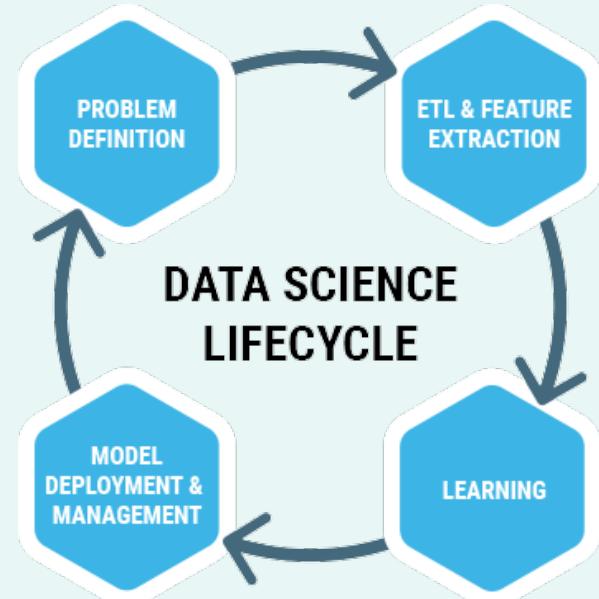
*"How to construct valuable data science projects in the real world"*

Jonny Brooks-Bartlett

*AI team, IT and BI must communicate synergistically*



## **STRATEGIC DATA AQUISITION**



*"How to construct valuable data science projects in the real world"*

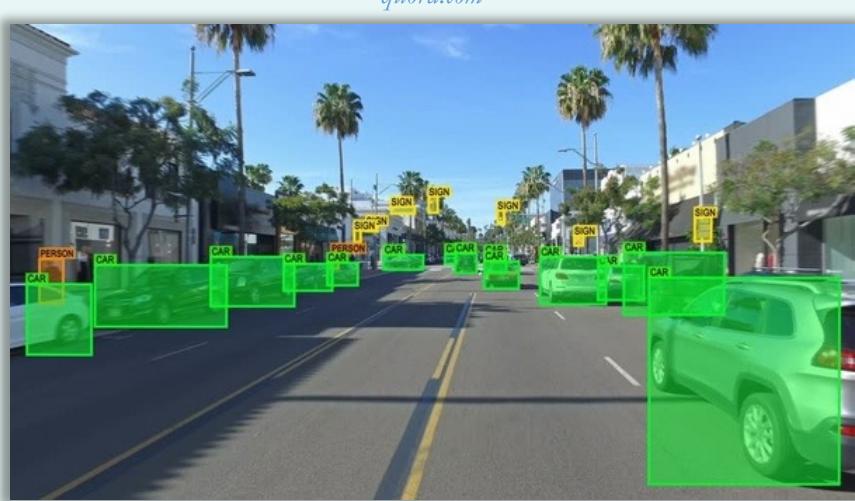
Jonny Brooks-Bartlett

*Sometimes data is already available in  
the company,*



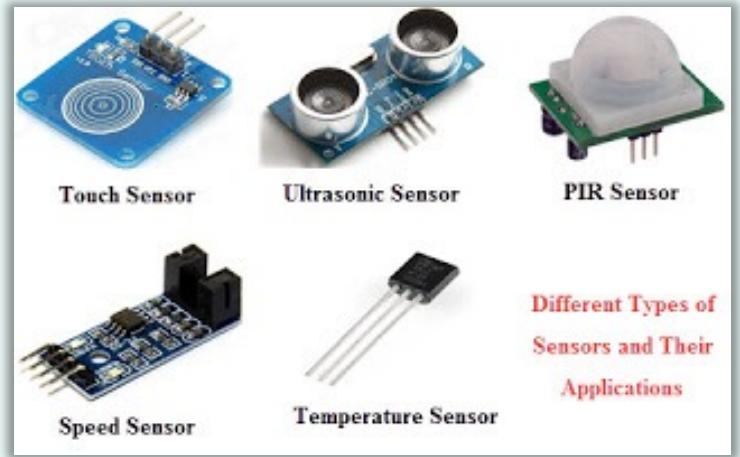
"Domino Effect", [Irma Jukić, artists-work.eu/cms/v/8754](http://IrmaJukic.artists-work.eu/cms/v/8754)

*but needs to be analyzed  
to understand its true value...*



Data Annotation

## IoT Sensors



*...in other cases it needs to be collected...*

[cnbs.com](#)



*...sometimes literally*

June 7, 2022

*AI-light Srl*

◀ 118 ▶



## The New AI Toilets Will Scan Your Poop To Diagnose Your Ailments

• Navin Bondade • 9 Comments

[techgrabyte.com](#)

151k Shares

Well, it's sounds weird but in the upcoming future, your toilet will be your mini doctor. A company called Micron is developing a smart artificial intelligence-powered toilet that will reportedly be able to diagnose your state of health and risk of disease by analyzing your bodily waste.

*What matters the most  
for the success of a ML project  
is not the quantity of data,  
but it's its **quality***



- Data has to be easily available
- Data must be coherent, with a small amount of errors and missing values
- Data must reflect corrections and requests from data scientist and analytics

*What matters the most  
for the success of a ML project  
is not the quantity of data,  
but it's its **quality***



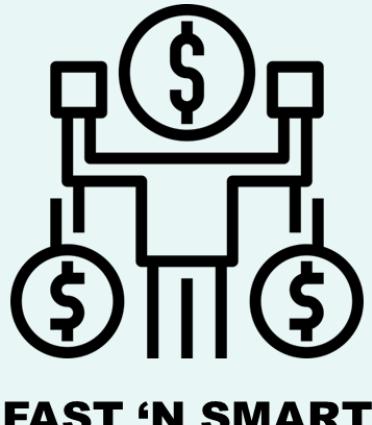
*“Garbage in, garbage out”  
GIGO principle*



*Machine Learning is very powerful tool,  
but it is particularly fragile  
**and this is reflected in the outcome  
of the ML projects***

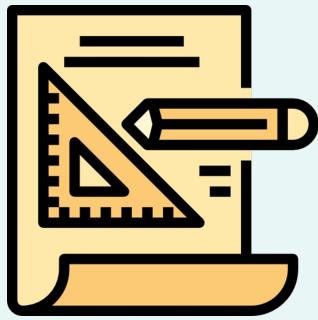


# *Tips and Tricks to start a ML project in a company*

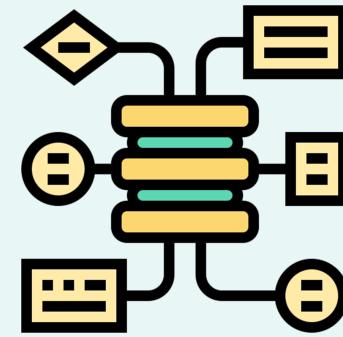
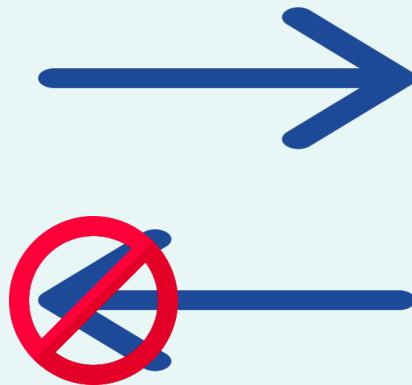


- Choose a small business problem easily solvable by a machine learning algorithm (create a PoC and a MVP)
- Clarify what's the purpose of the project using specific metrics
- “Fast” data collection / extraction and processing
- First developments with simple algorithms (benchmarks)
- Continuous adjustments of the data collection / ETL processes w.r.t. the algorithms performance

# *Tips and Tricks to start a ML project in a company*



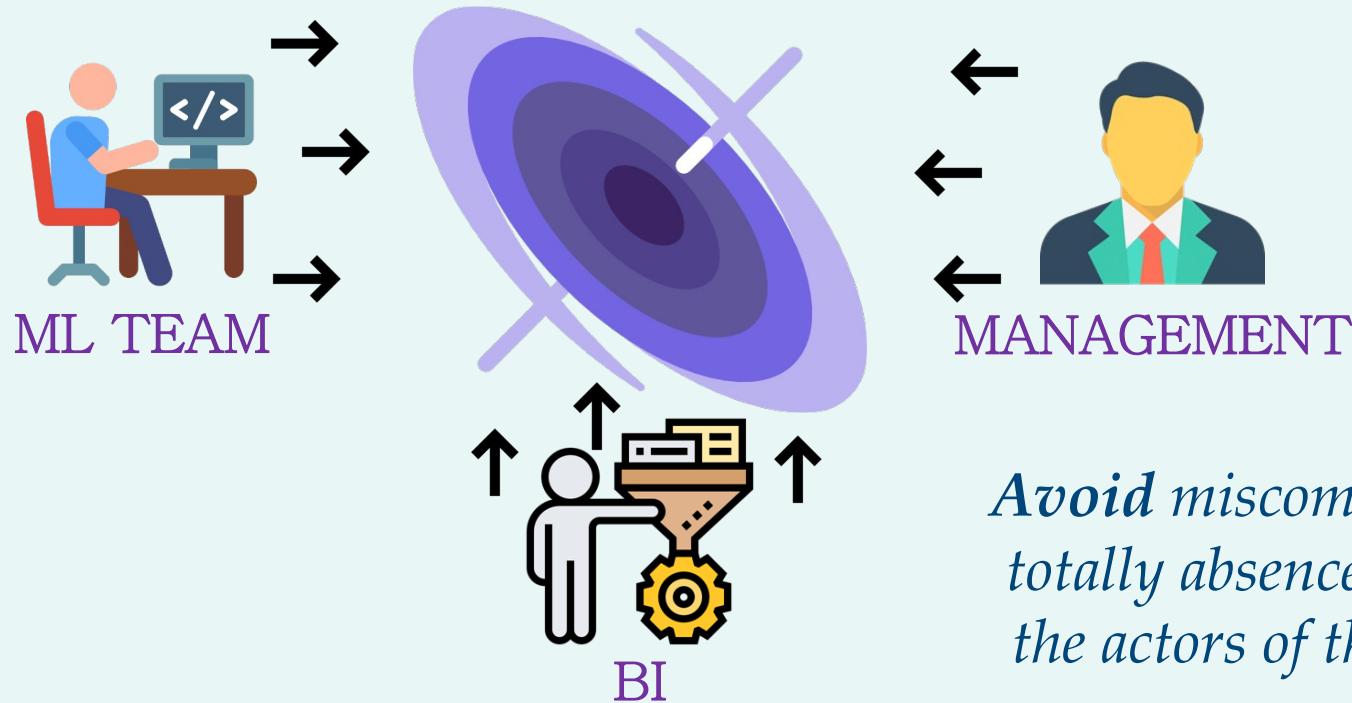
PROJECT



TECHNOLOGY

*Don't choose a project only because it allows you  
to use the most recent or popular technologies*

# *Tips and Tricks to start a ML project in a company*



*Avoid miscommunication or  
totally absence of it between  
the actors of the ML project*

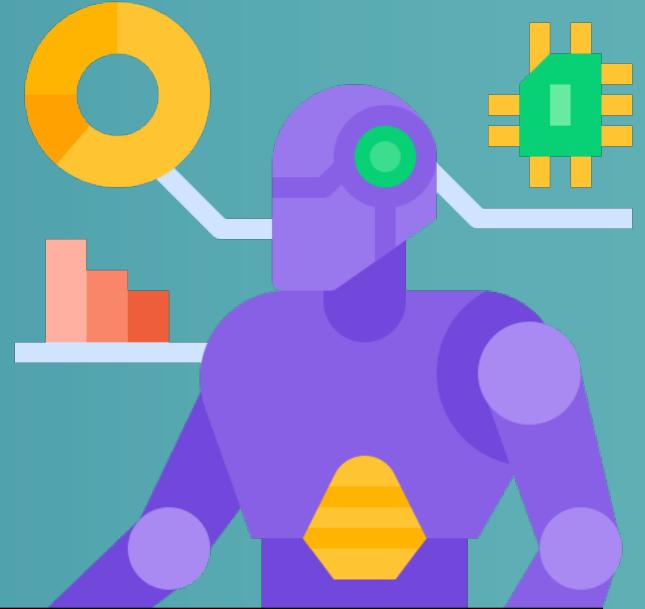
# *Tips and Tricks to start a ML project in a company*



*Avoid disproportionate expectations w.r.t. the technologies currently available, the project execution times or the available data*

# *AI-light*

*with AI, for a brighter future*



**Fabio Mardero**

[fabio.mardero@ai-light.biz](mailto:fabio.mardero@ai-light.biz)  
[linkedin.com/in/fabio-mardero](https://linkedin.com/in/fabio-mardero)