

Intro to data science

Development, use cases, and prospect

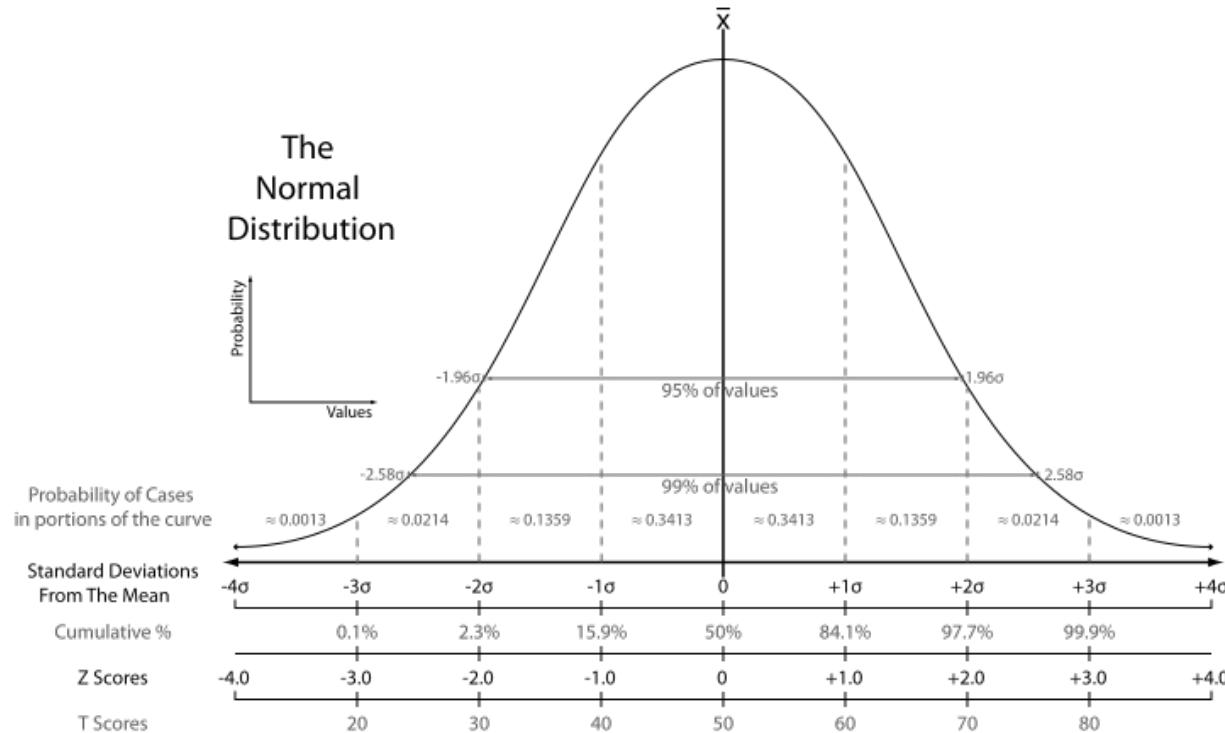
May 26, 2018
Thammasat University

Chatdanai Lumdee (Tua/ตัว)
Data scientist at Kasikorn Labs, Kasikorn Business-Technology Group
(and a physicist -- optics, nanotechnology)

Intro to data science

Development, use cases, and prospect

The Normal Distribution



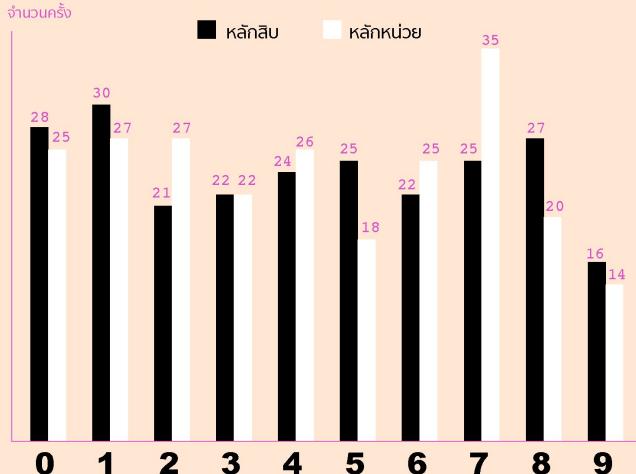


ເລີ່ມ(ກັບ)ຮວຍ ດ້ວຍສົດຕິ : ສໍາຮວຈ ຮວຍໄທຍໃນຮອບ 10 ປີ

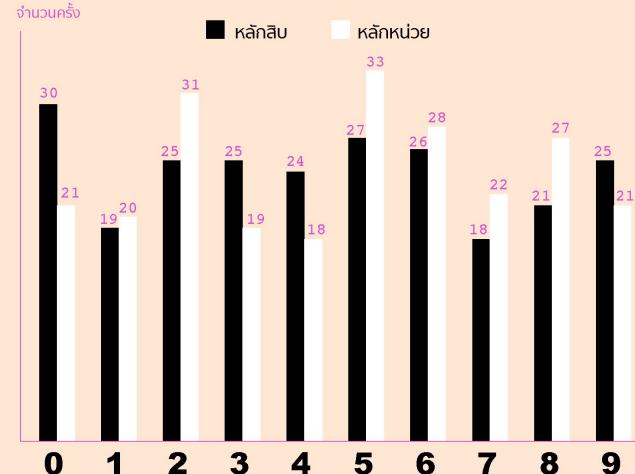
SEPTEMBER 16, 2017 by THANISARA GG RUANGDEJ



ເລີ່ມທ້າຍ 2 ຕ້າບນ ທີ່ເຄີຍອອກຮາງວັລ
(ແຢກຕາມຫລັກ)



ເລີ່ມທ້າຍ 2 ຕ້າລ່າງ ທີ່ເຄີຍອອກຮາງວັລ
(ແຢກຕາມຫລັກ)



WHAT IS ACTUARIAL SCIENCE

"It's about calculating money-related risks"

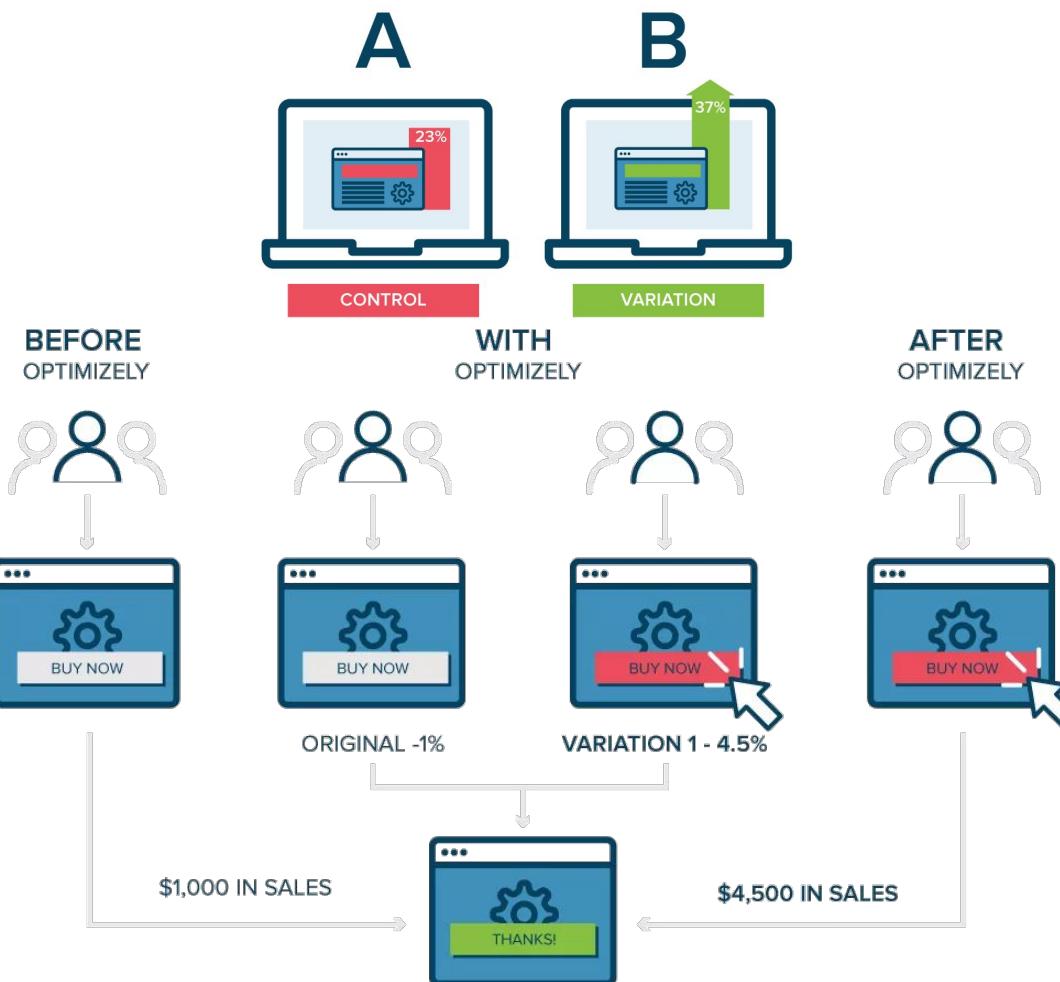


Actuarial science involves:

- Statistics
 - Probability
 - Knowledge of Business & Economics
- ... to make sound business decisions

Example:

The amount that you must pay for your insurance plan is calculated by an actuary!



Settings

Solve for? Power Alpha n d

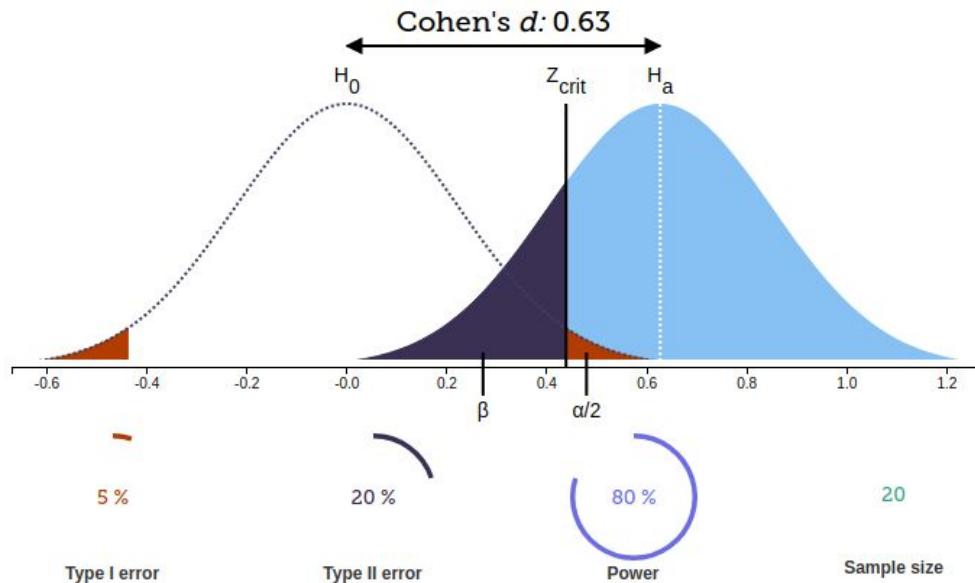
Power ($1-\beta = 0.8$)

Significance level ($\alpha = 0.05$)

Sample size ($n = 20$)

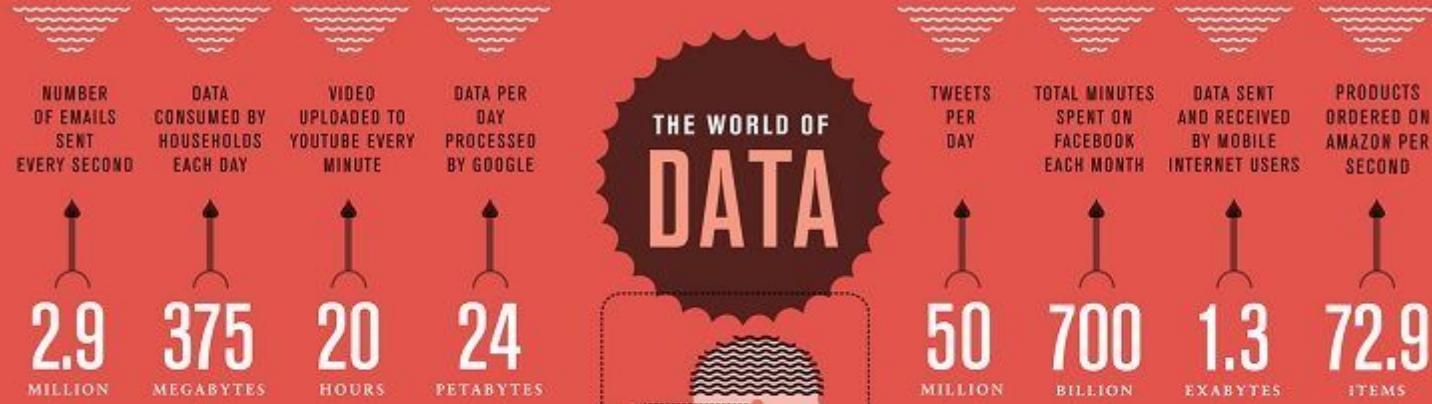
One-tailed Two-tailed

Reset zoom



BIG DATA





THE WORLD OF DATA

SOURCES: Cisco, comScore, Markit, Nielsen, Kantar, Google Trends, YouTube

IN THE 21ST CENTURY, we live a large part of our lives online. Almost everything we do is reduced to bits and sent through cables around the world at light speed. But just how much data are we generating? This is a look at just some of the massive amounts of information that human beings create every single day.

Beginning (~1996)

Google Search Engine

This is a demo of the Google Search Engine. Note, it is research in progress so expect some downtimes and malfunctions. You can find the older [Backrub web page here](#).

Google is being developed by [Larry Page](#) and [Sergey Brin](#) with very talented implementation help by [Scott Hassan](#) and [Alan Steremberg](#).



Search Stanford

10 results clustering on

Search The Web

10 results clustering on

Now

Google

big data



big data industries

big data

big data คือ

big data analytics คือ

big data meaning

big data analytics pdf

big data thailand

big data pantip

big data landscape 2017

big data ตัวอย่าง

[Report inappropriate predictions](#)

4G 1 08:02

Google

Say "Ok Google"

24°C Pak Kret District

Eat & Drink Entertainm... Sports

Machine learning

Why Do Developers Find It Hard To Learn Machine Learning?

Despite the interest and a sense of urgency in embracin...

F Forbes · 1 day ago

Data science

Docker for Data Science

Home Search Recent

Secure | https://www.amazon.com

Recommended for You in Kindle Books

READY PLAYER ONE by Ernest Cline

CREATIVITY, INC. by Ed Catmull

Homo Deus by Yuval Noah Harari

COSMOS by Carl Sagan

Inspired by your Wish List [See more](#)

Life 3.0 by Max Tegmark

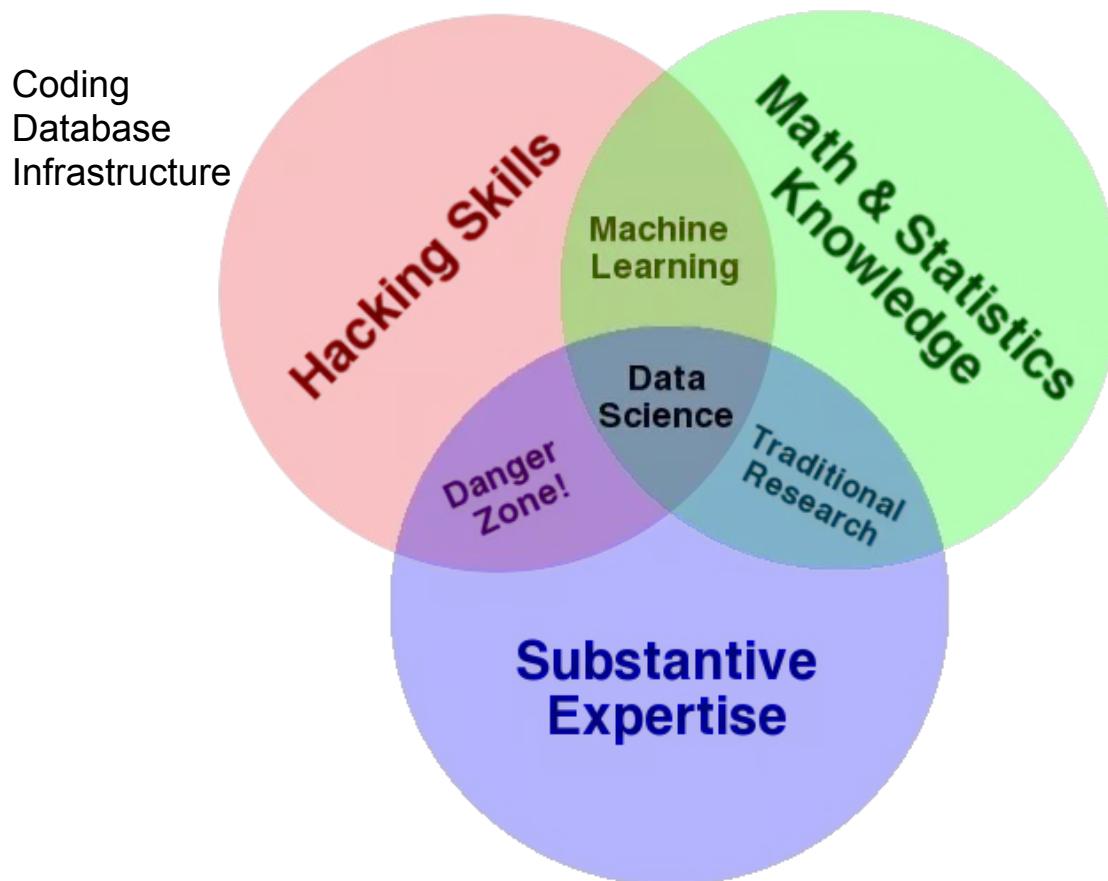
THE RETREAT OF WESTERN LIBERALISM by Edward Luce

BEHAVE by Robert Sapolsky

the four by Scott Galloway

THE MASTER ALGORITHM by Pedro Domingos







Finance



Retail



Health



Media



Sports



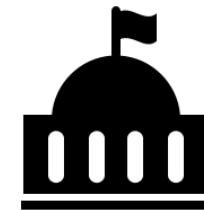
Logistics



Education



Government



<https://www.datameer.com/product/industries-use-cases/>

<https://datafloq.com/read/8-industries-benefit-big-data-infographic/402>

<https://www.simplilearn.com/big-data-applications-in-industries-article>

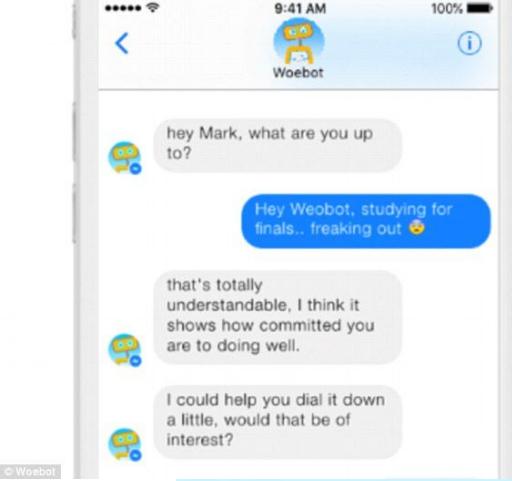
เทคโนโลยี และบริการ





INTRODUCING
amazon go





© Woebot



Harvard
Business
Review

ANALYTICS

How Machine Learning Is Helping Us Predict Heart Disease and Diabetes

by Yannis Paschalidis

MAY 30, 2017





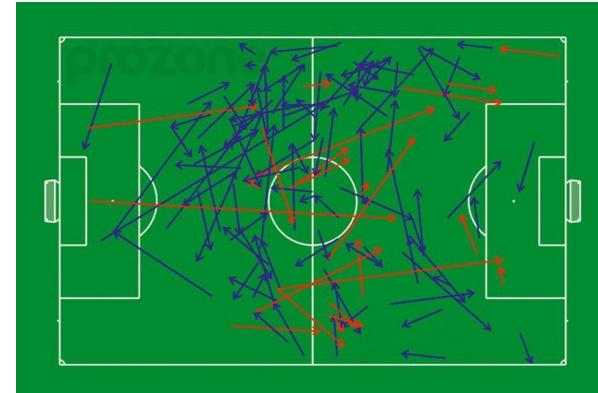
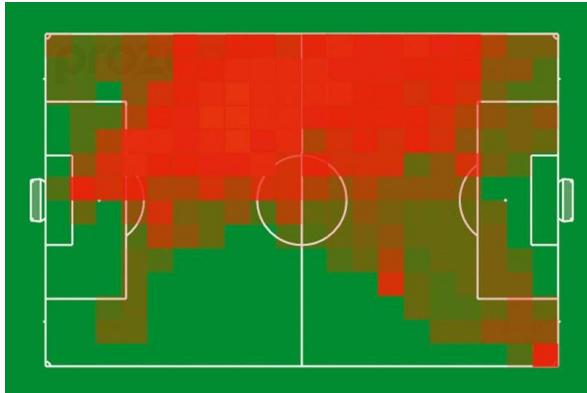




"Don't take the risk of facing death"

HOW BIG DATA IS CHANGING THE WORLD OF (DAILY FANTASY) FOOTBALL

Did you know that athletes are not only monitored by cameras on stadiums, but also by many quirky devices such as accelerometers, heart rate sensors and even local GPS-like systems? Indeed, Big Data and modern technologies are currently revolutionizing sports and even powering the Daily Fantasy Football industry.









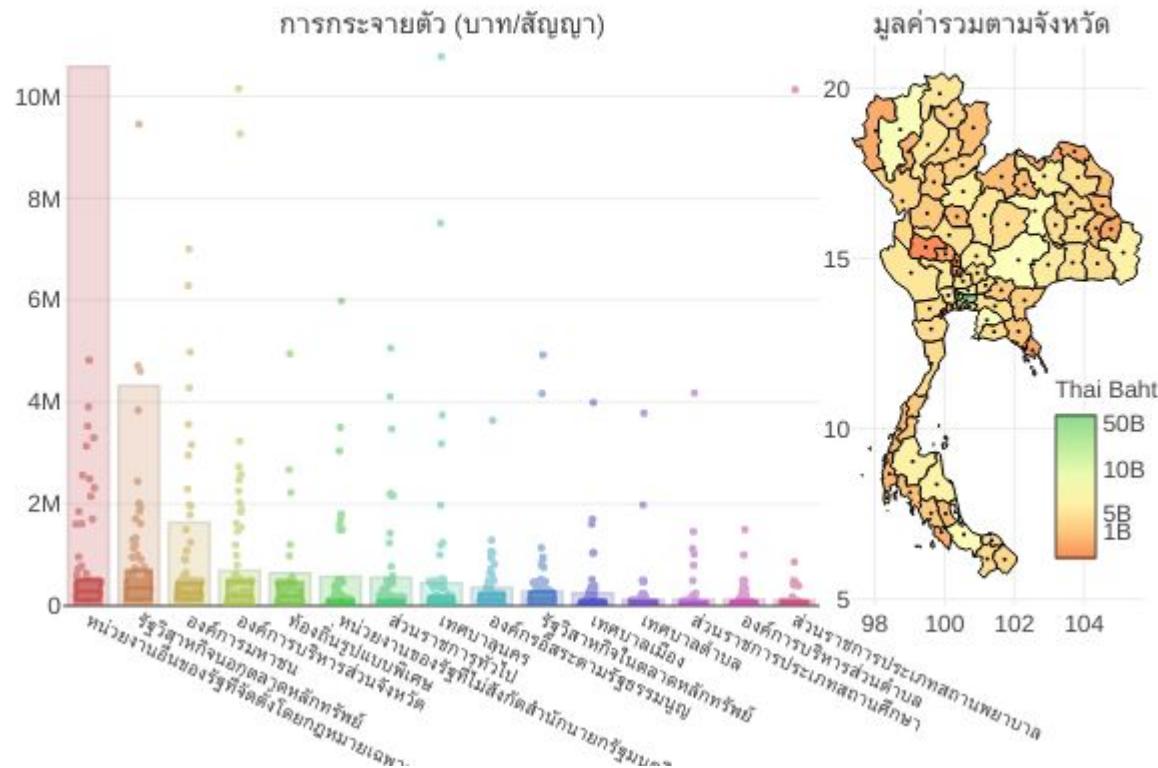


<https://www.reuters.com/article/us-india-technology-development/poor-nations-need-help-to-use-big-data-to-tackle-disease-poverty-expert-idUSKBN19V0UO>



Allan Sanders

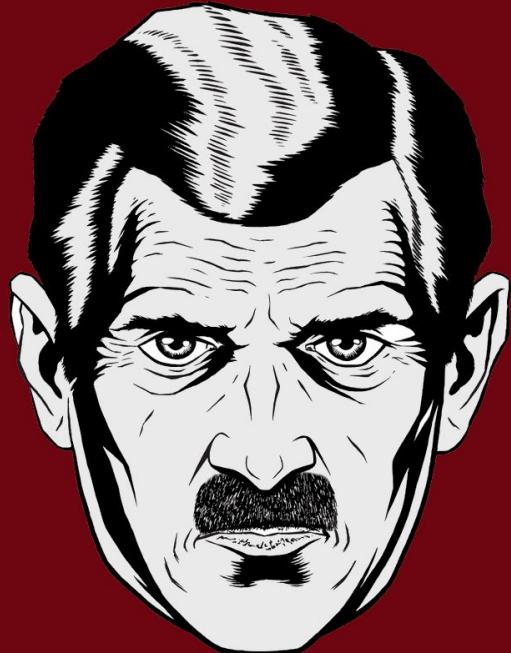
ข้อมูลคำสั่งจัดซื้อจัดจ้างโดยหน่วยงานรัฐบาล



<https://medium.com/@chatdanailumdee/open-data-science-thailand-peeking-on-government-spending-f07d7fc1b4d8>







**BIG BROTHER IS
WATCHING YOU**



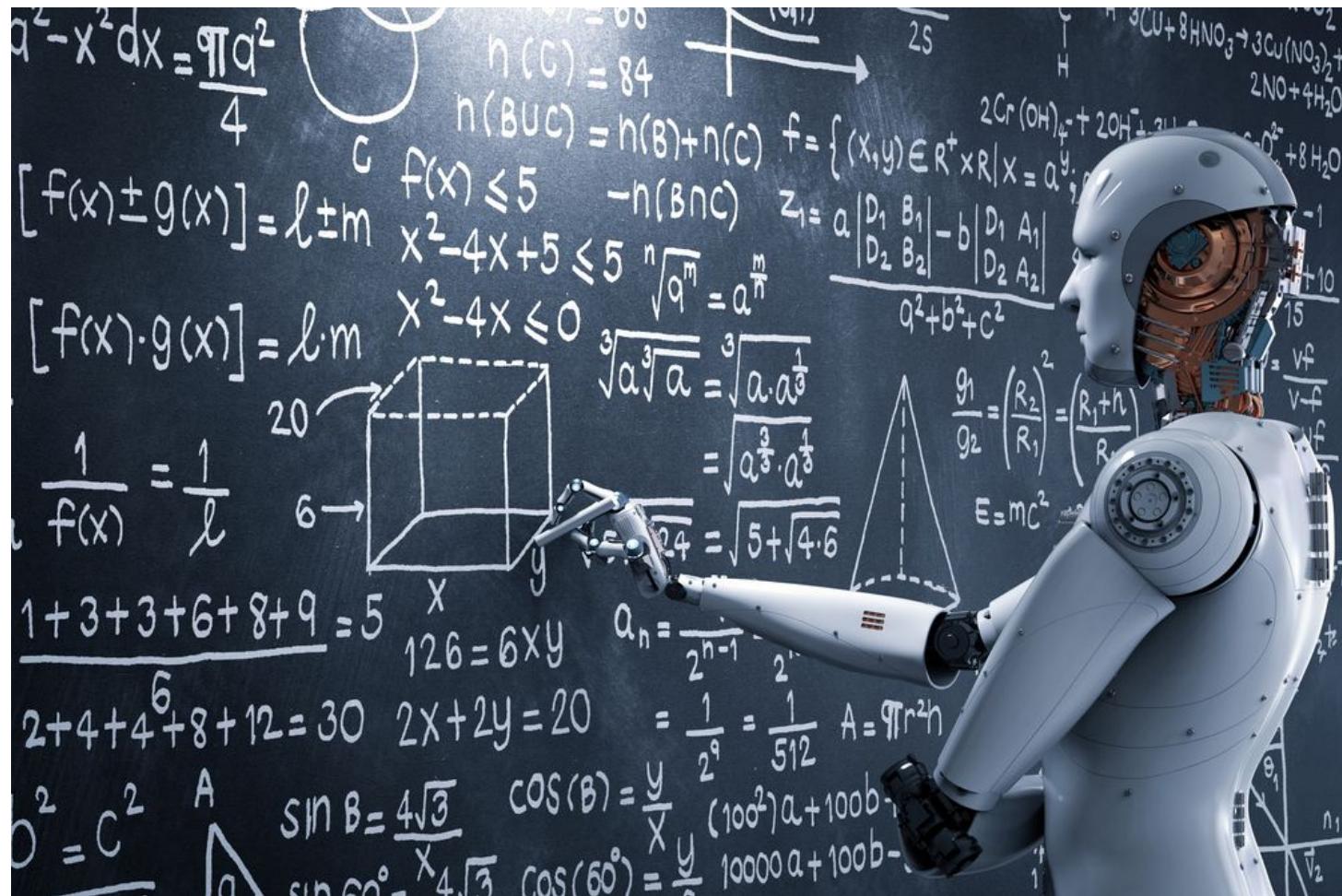
Intro to data science

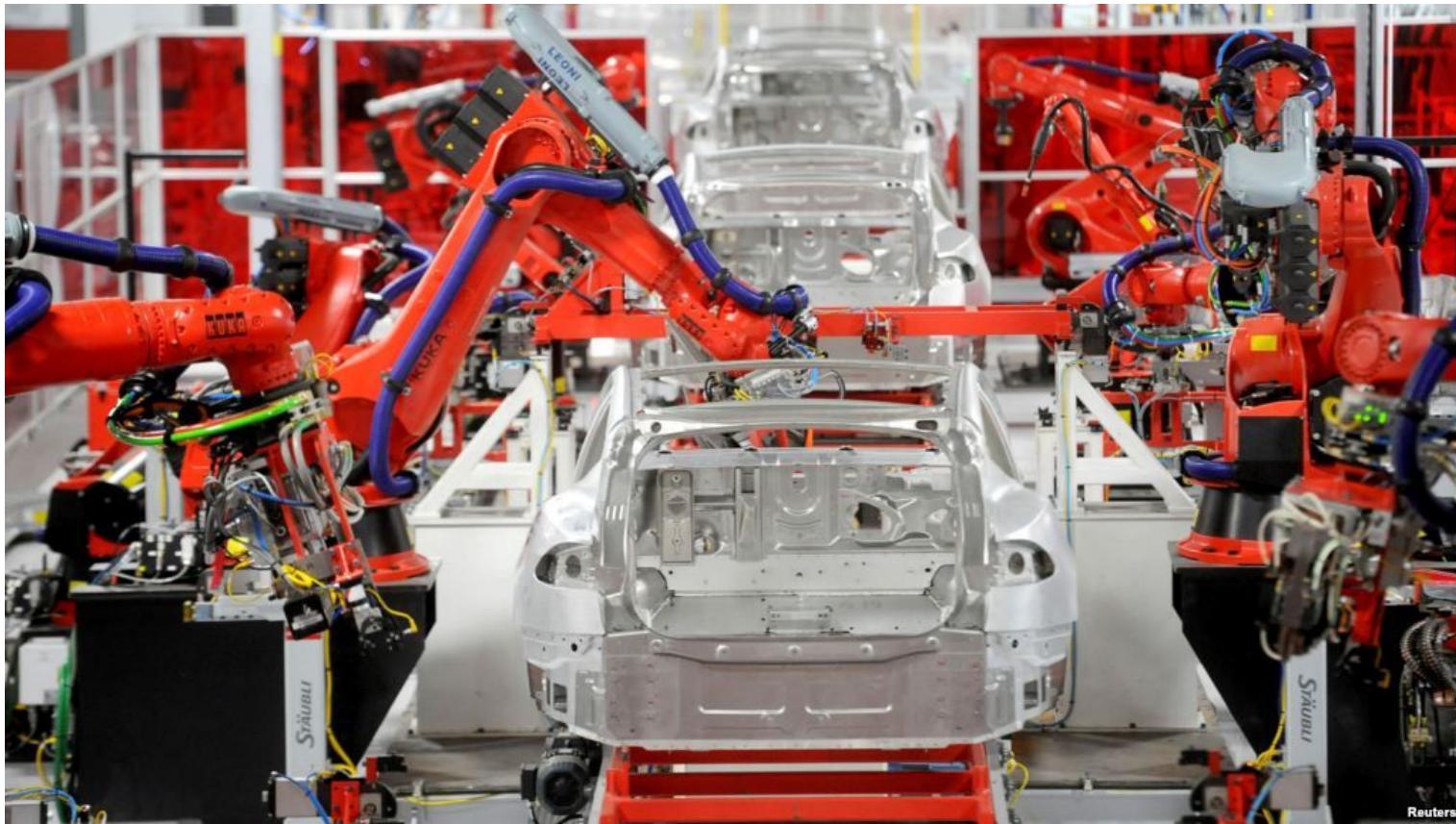
Development, use cases, and prospect



Intro to data science

Development, use cases, and prospect





<https://www.voanews.com/a/tesla-model-3-assembly-line-strategy/3823841.html>



ACM Chess Challenge

Garry Kasparov vs



<https://www.scientificamerican.com/article/20-years-after-deep-blue-how-ai-has-advanced-since-conquering-chess/>

中国围棋协会

Google

浙江省体育局

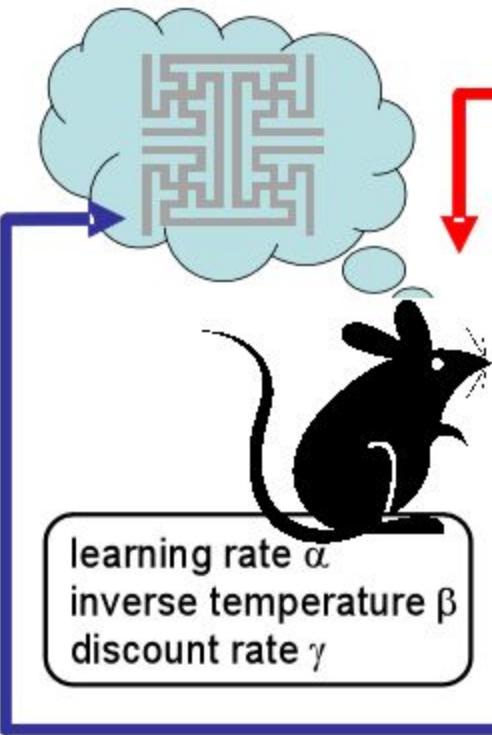
中国乌镇 围棋峰会

The Future of Go Summit in Wuzhen





internal state



reward

environment



action

observation

xx:yy: 1, 2
Deaths/HighScore: 24 / 61

17

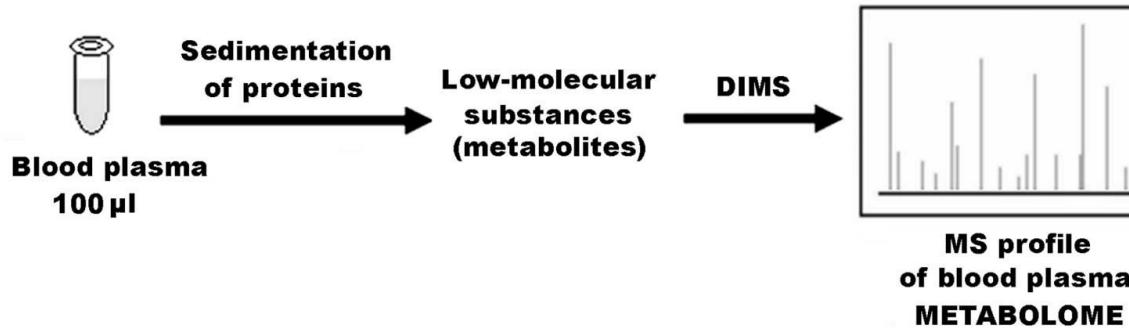




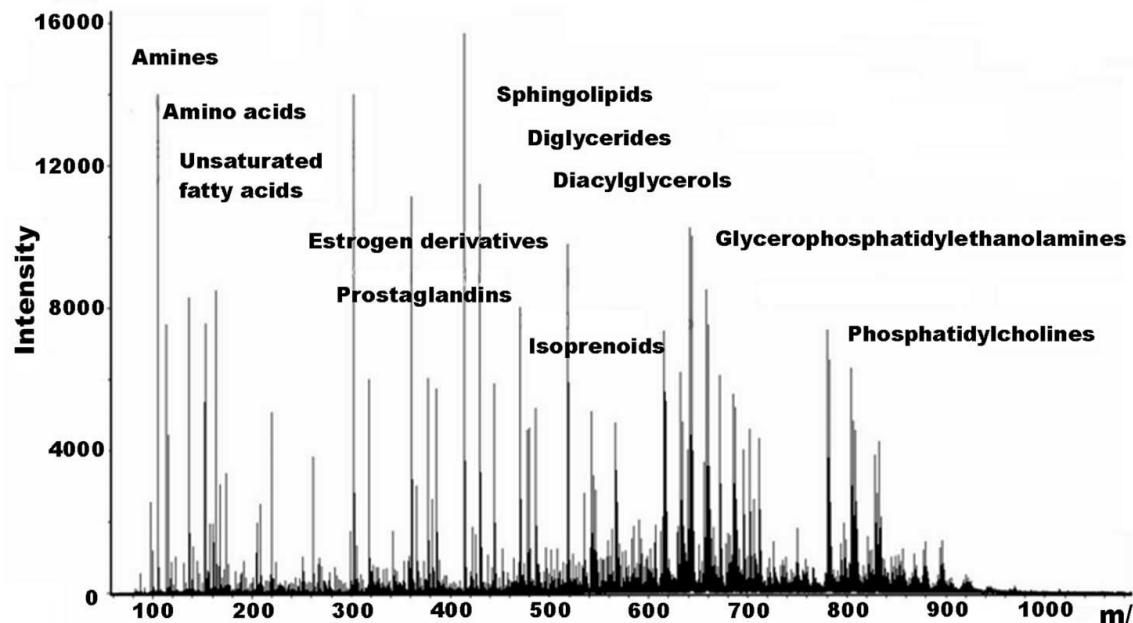


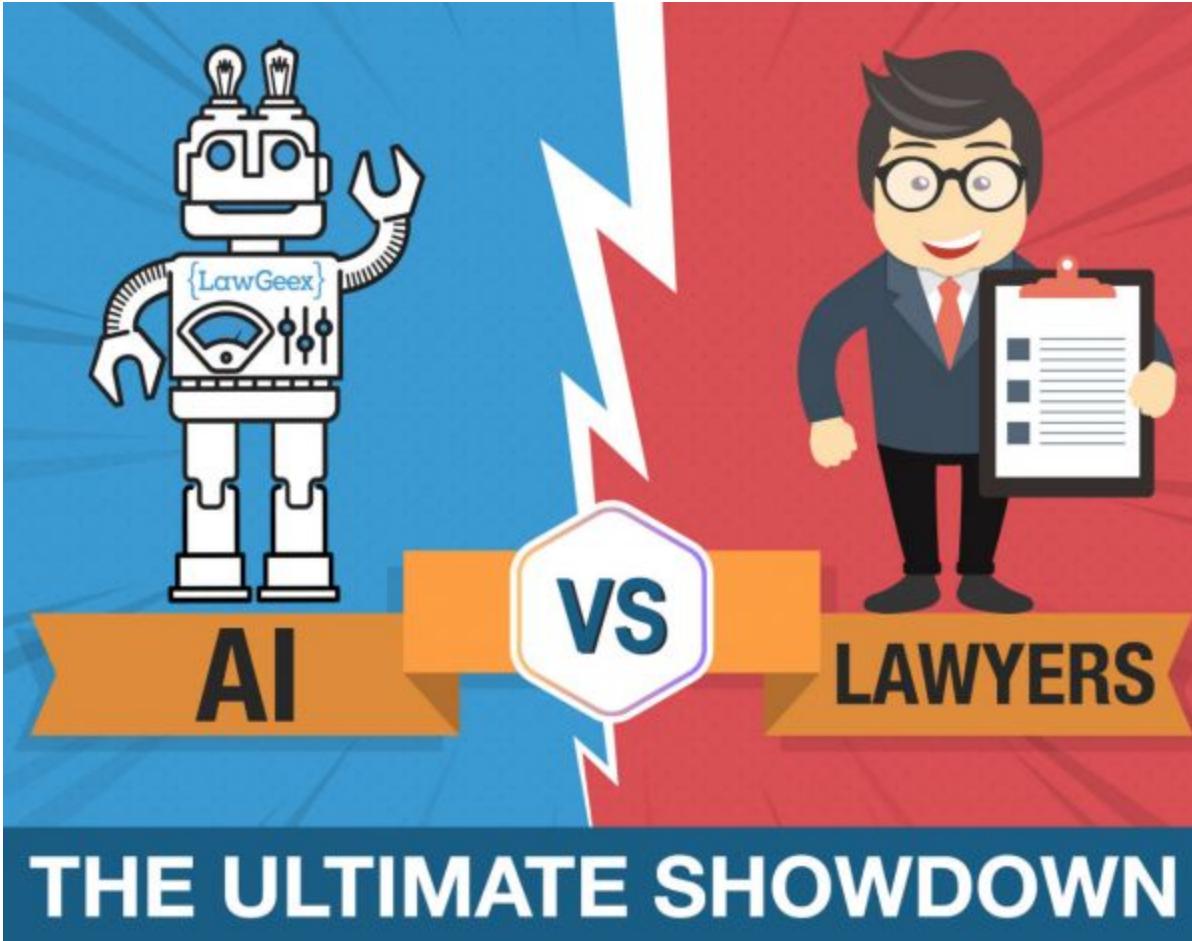


A

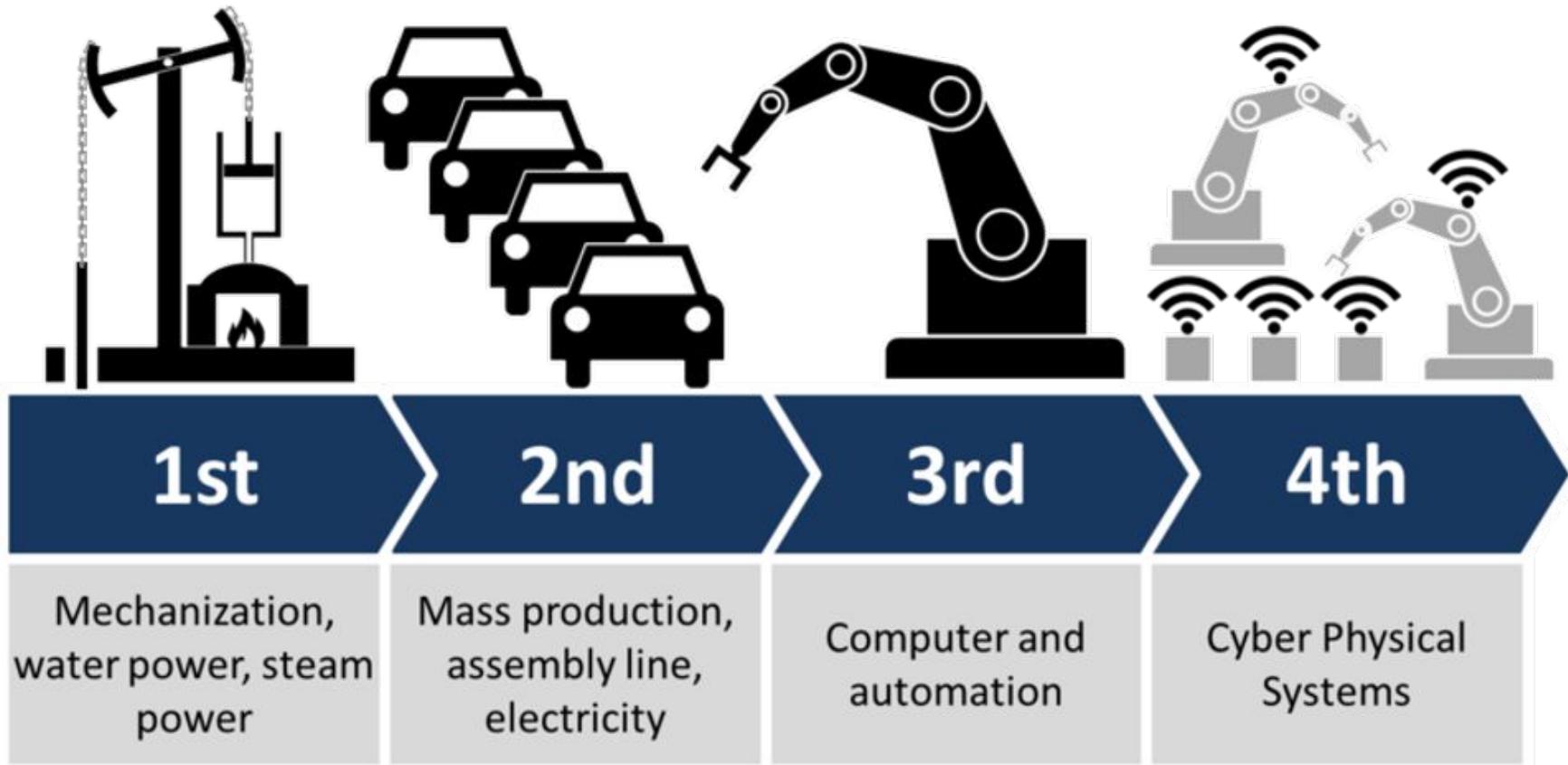


B

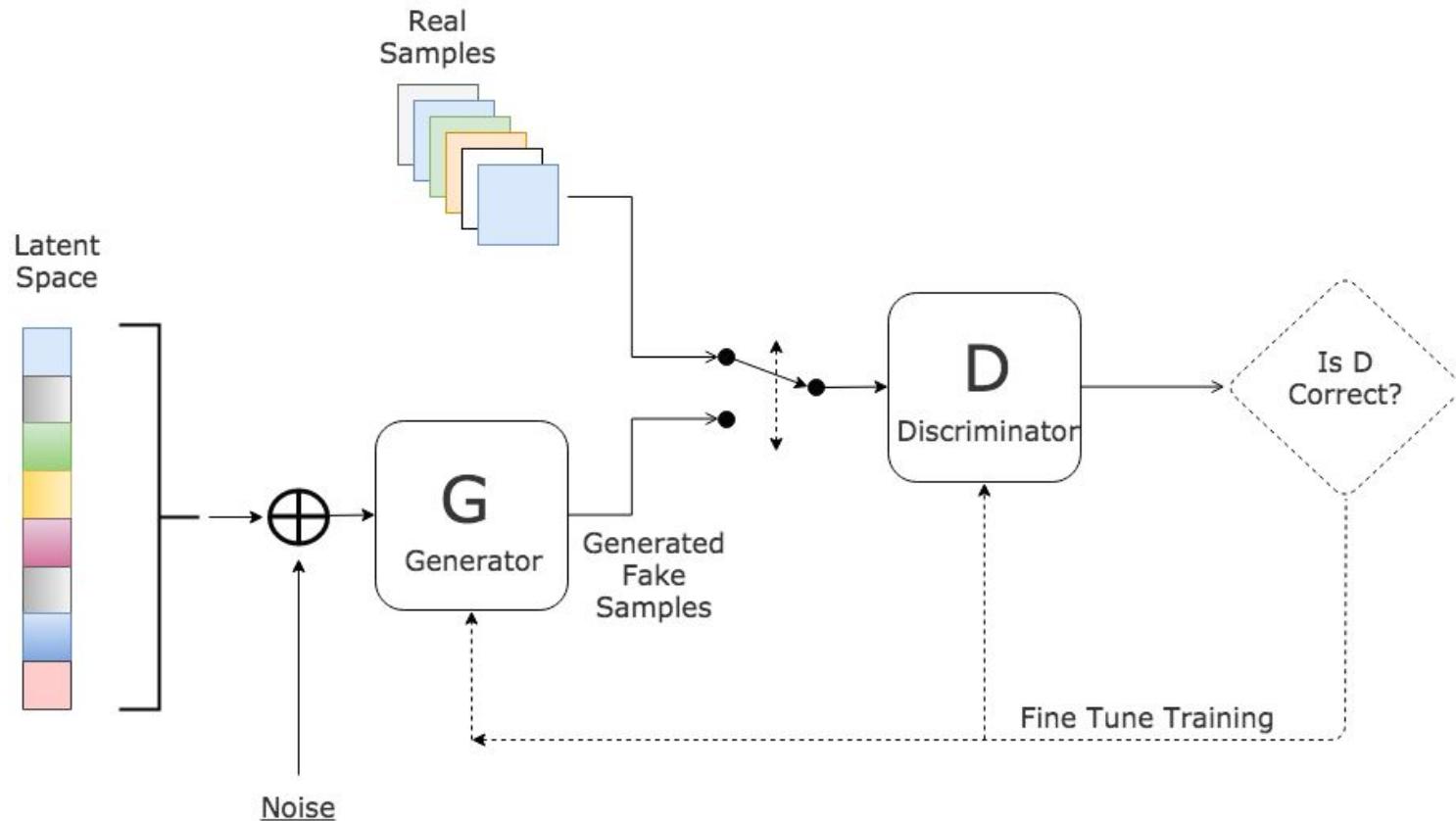




THE ULTIMATE SHOWDOWN



Generative Adversarial Network





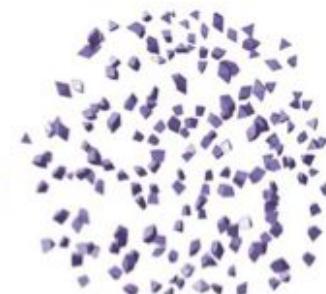
R: Real Data



D: Detective



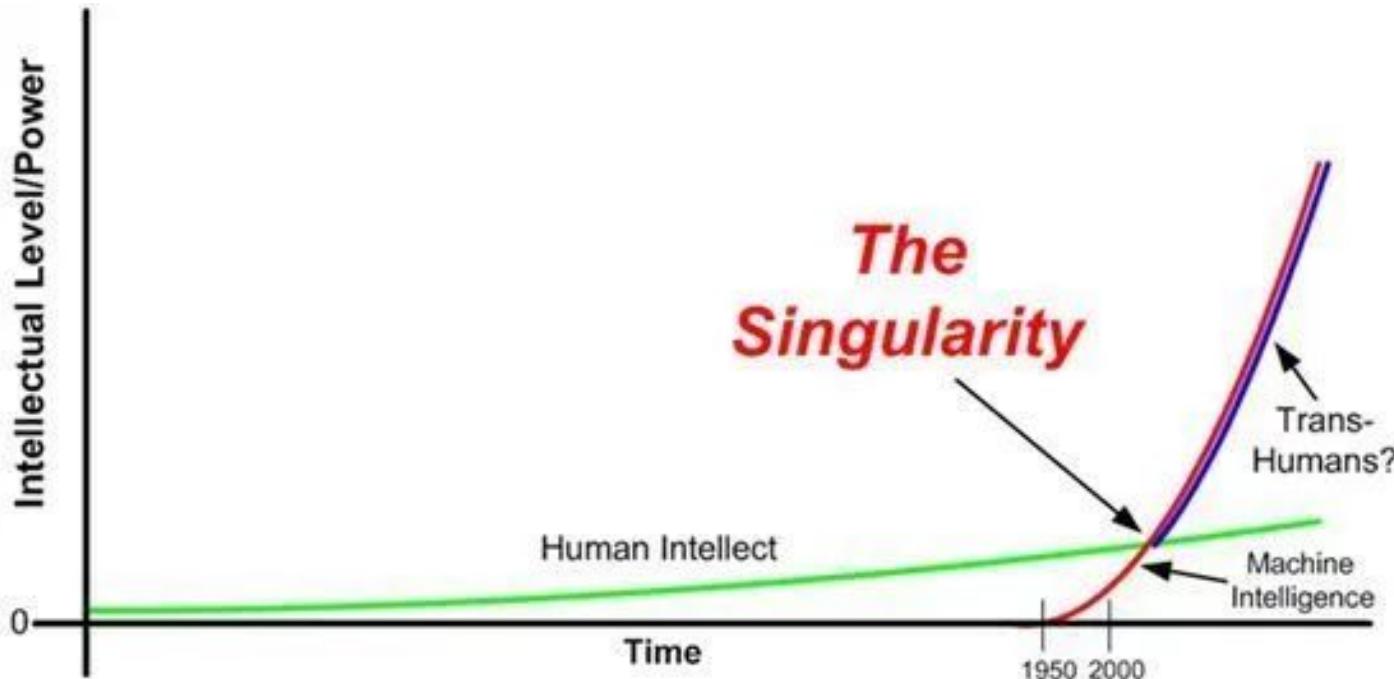
G: Generator (Forger)



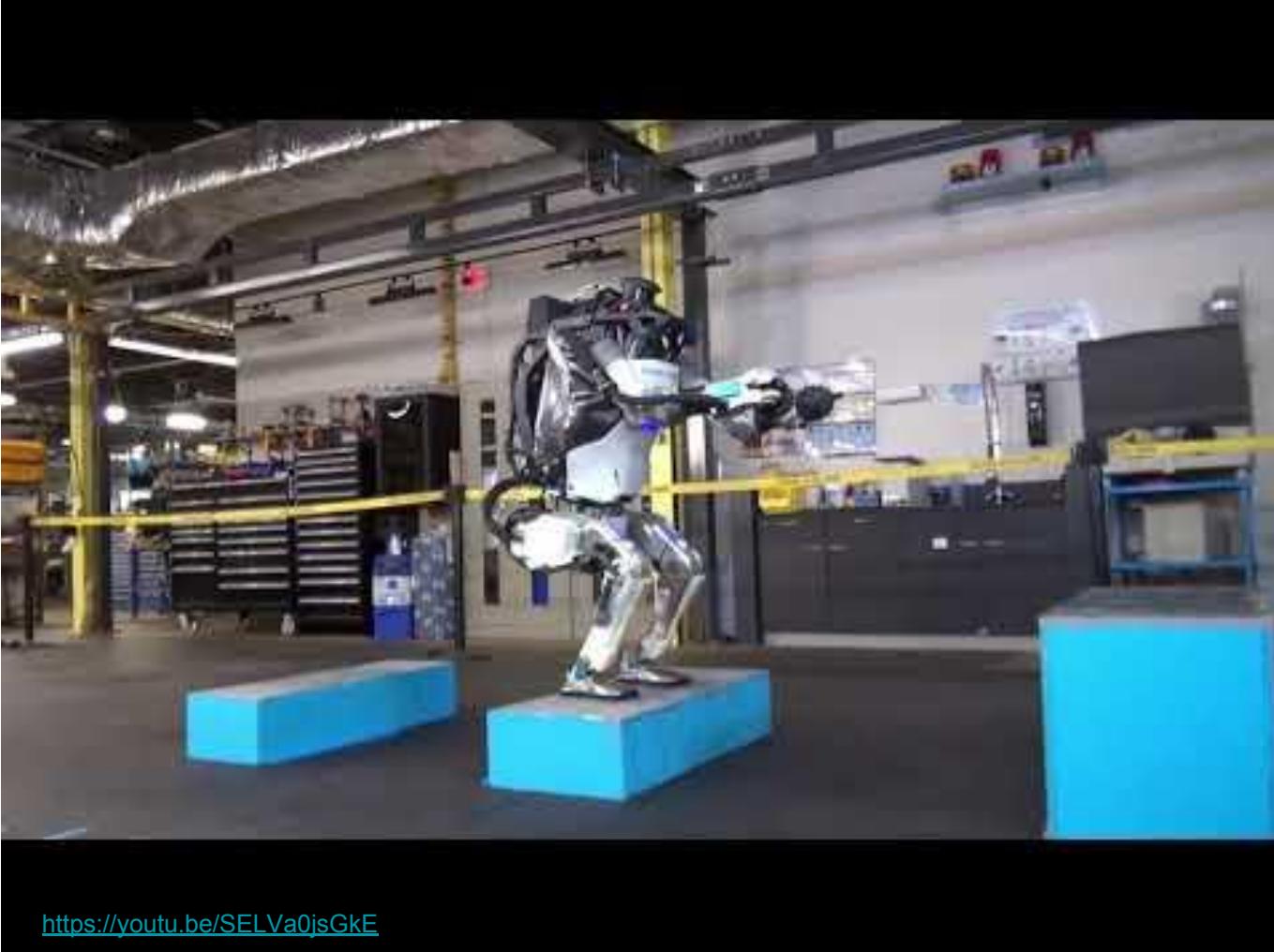
I: Input for Generator



<https://hothardware.com/news/nvidia-neural-network-generates-photorealistic-faces-disturbing-results>







<https://youtu.be/SELVa0jsGkE>



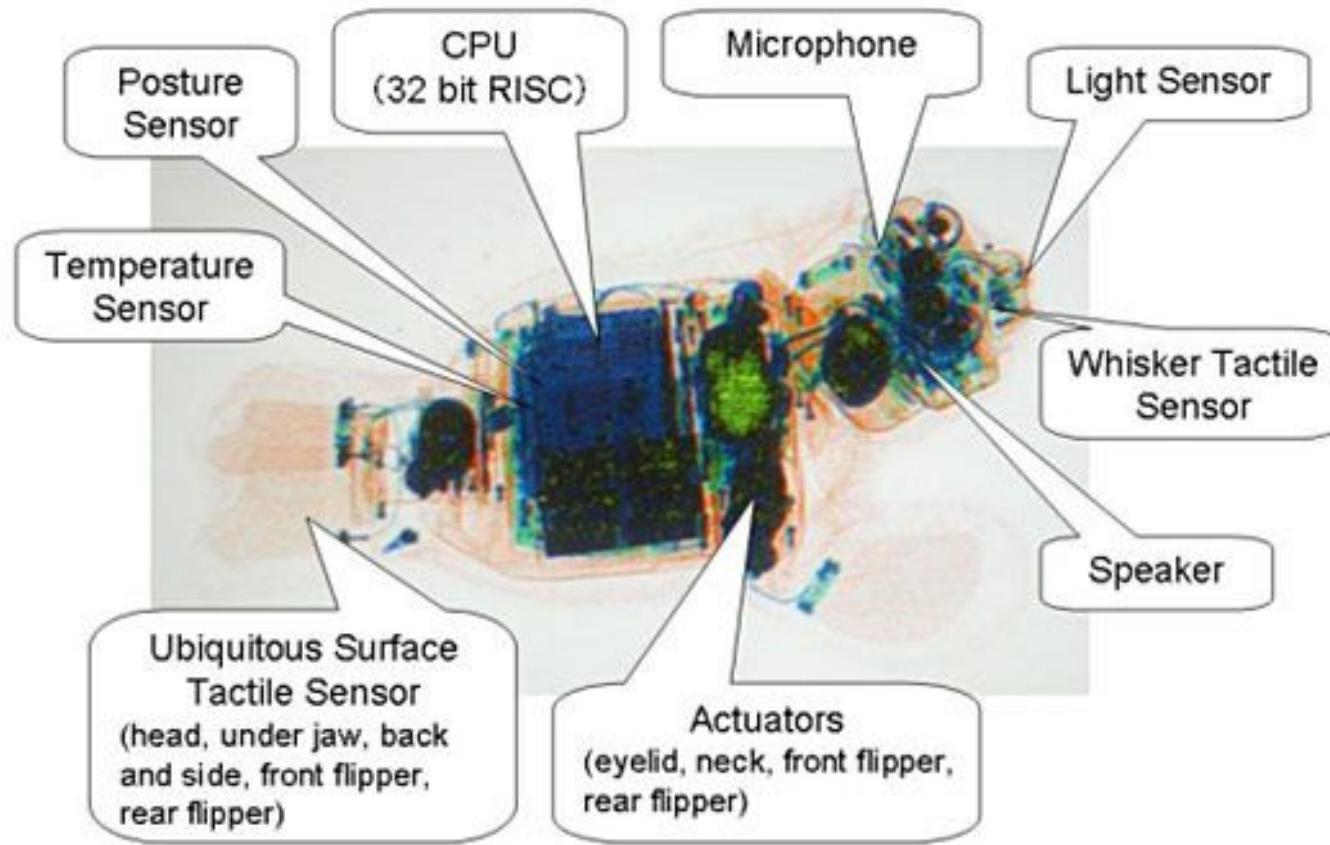
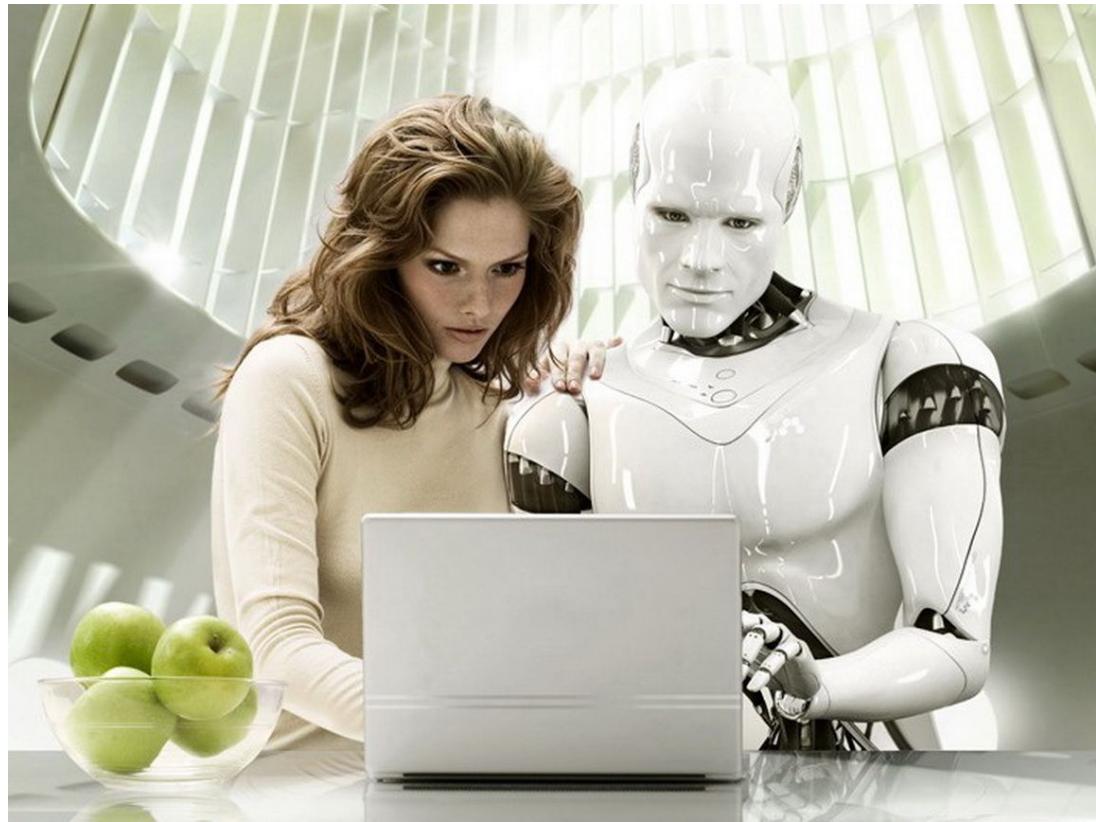


Figure 2: Paro's function





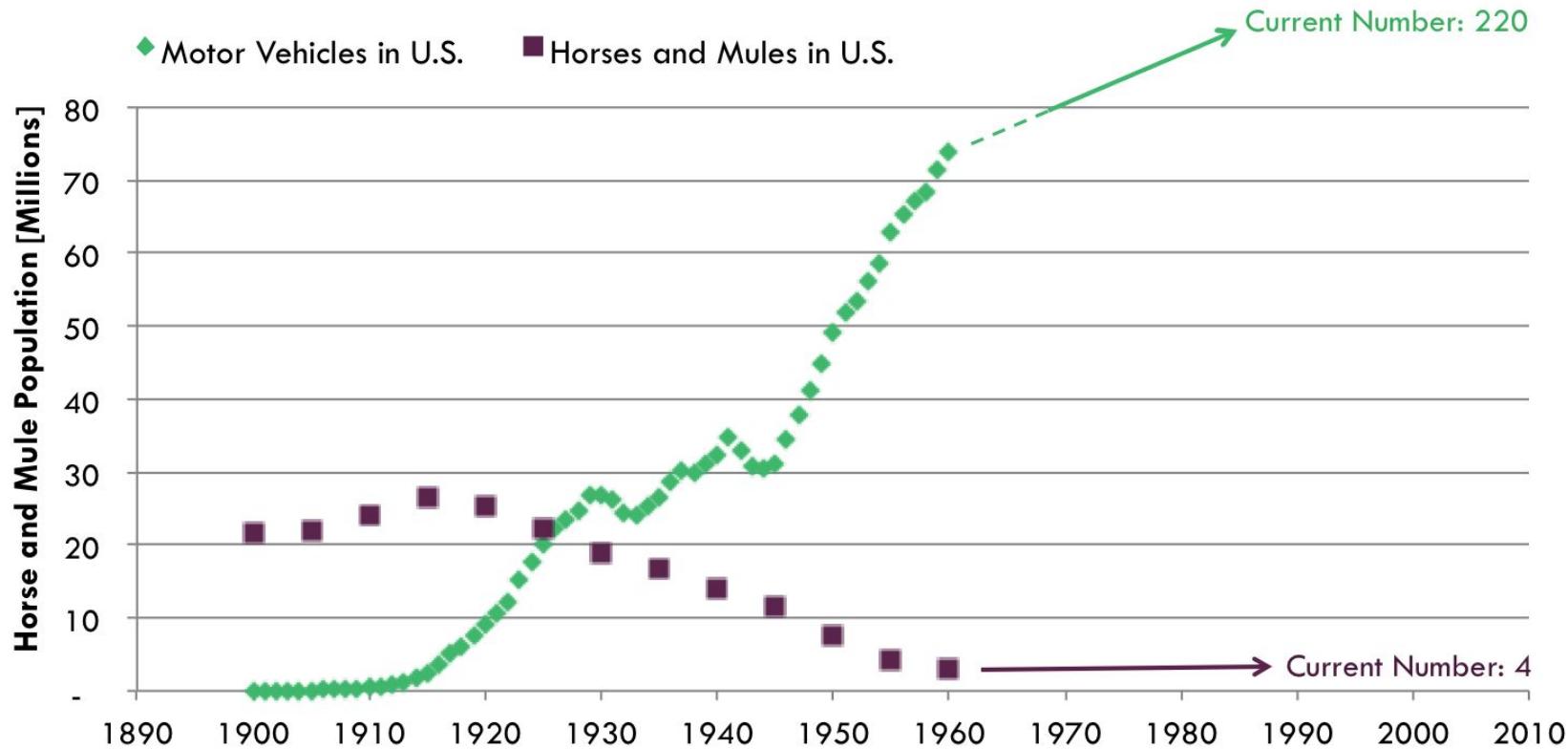




<https://www.youtube.com/watch?v=7Pq-S557XQU>



Motor Vehicle Registration and Number of Horses and Mules in the U.S.



Source: ARK Investment Management LLC, Bureau of Transportation Statistics, Bloomberg Intelligence, EVvolumes.com, IDC



Find me @ <https://clumdee.github.io/>



This repository

Search

Pull requests

Issues

Marketplace

Explore

[RudrakshTuwani / Football-Data-Analysis-and-Prediction](#)

6



PREMIER LEAGUE PREDICTION!!

```
# Total number of students.  
n_matches = data.shape[0]  
  
# Calculate number of features.  
n_features = data.shape[1] - 1  
  
# Calculate matches won by home team.  
n_homewins = len(data[data.FTR == 'H'])  
  
# Calculate win rate for home team.  
win_rate = (float(n_homewins) / (n_matches)) * 100  
  
# Print the results  
print "Total number of matches: {}".format(n_matches)  
print "Number of features: {}".format(n_features)  
print "Number of matches won by home team: {}".format(n_homewins)  
print "Win rate of home team: {:.2f}%".format(win_rate)
```



Data science workflow

