

Intro to data science

Development, use cases, and prospect

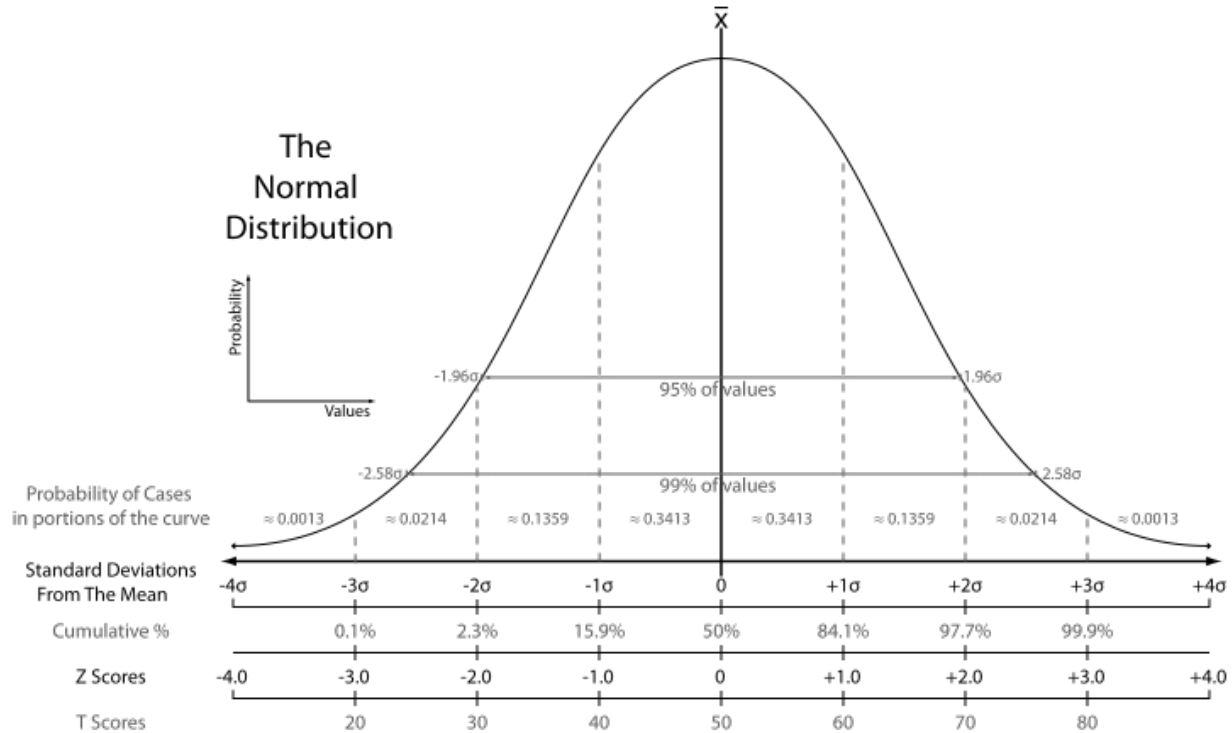
May 26, 2018
Thammasat University

Chatdanai Lumdee (Tua/ตั้ว)
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



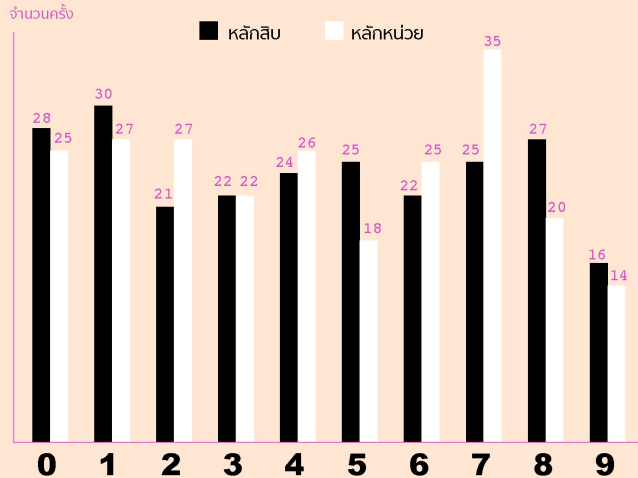


เล่น(กับ)หวย ด้วยสถิติ : สำรอง หวยไทยในรอบ 10 ปี

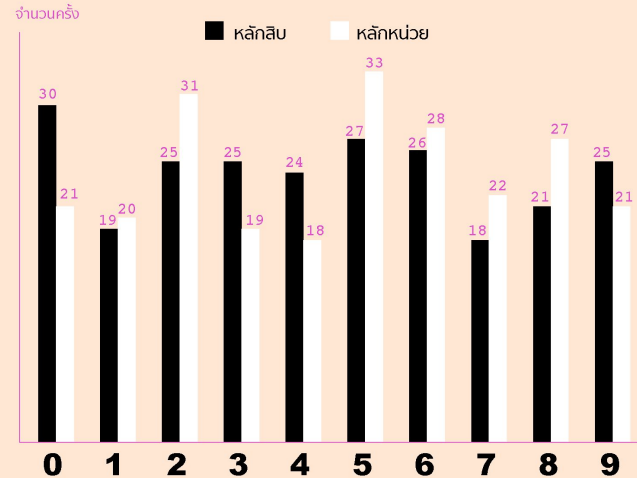
SEPTEMBER 16, 2017 by THANISARA GG RUANGDEJ



เลขท้าย 2 ตัวบน ที่เคยออกรางวัล (แยกตามหลัก)



เลขท้าย 2 ตัวล่าง ที่เคยออกรางวัล (แยกตามหลัก)



WHAT IS ACTUARIAL SCIENCE

"It's about calculating money-related risks"



Actuarial science involves:

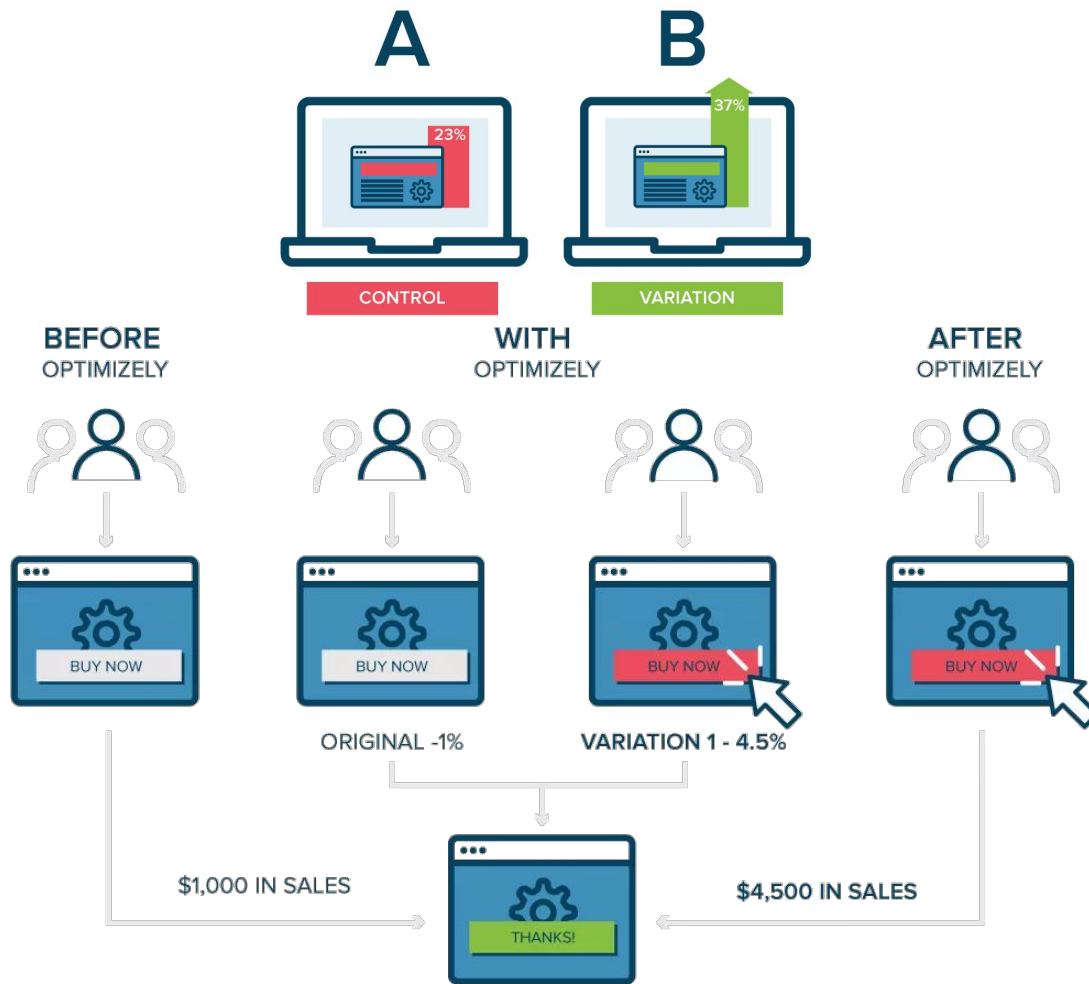
- Statistics
- Probability
- Knowledge of Business & Economics

... to make sound business decisions



Example:

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



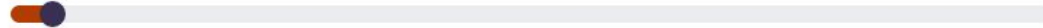
Settings

Solve for? Power Alpha n d

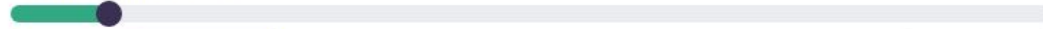
Power ($1-\beta = 0.8$)



Significance level ($\alpha = 0.05$)

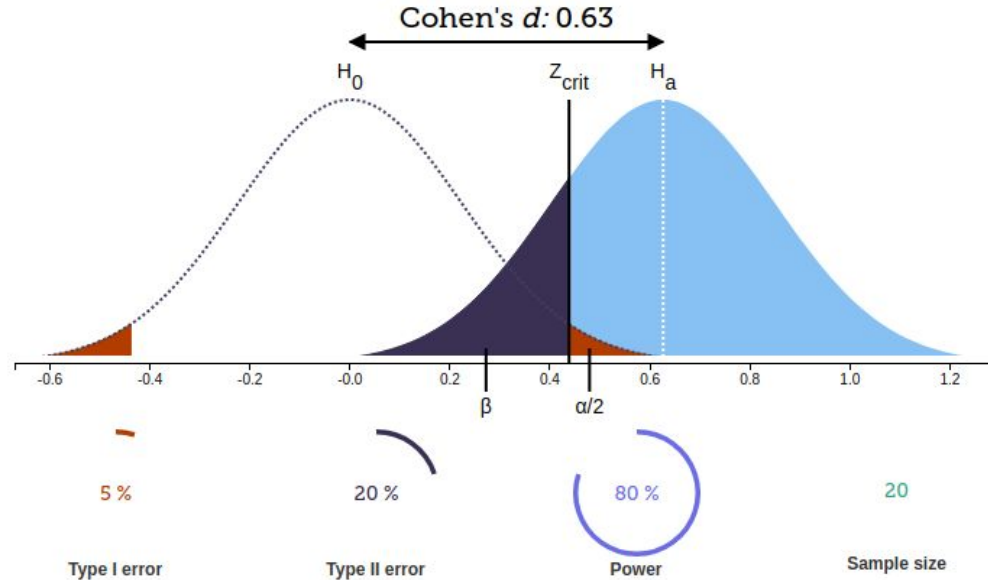


Sample size ($n = 20$)



One-tailed Two-tailed

Reset zoom



BIG DATA



THE WORLD OF DATA

NUMBER OF EMAILS SENT EVERY SECOND

2.9 MILLION



DATA CONSUMED BY HOUSEHOLDS EACH DAY

375 MEGABYTES



VIDEO UPLOADED TO YOUTUBE EVERY MINUTE

20 HOURS



DATA PER DAY PROCESSED BY GOOGLE

24 PETABYTES



TWEETS PER DAY

50 MILLION



TOTAL MINUTES SPENT ON FACEBOOK EACH MONTH

700 BILLION



DATA SENT AND RECEIVED BY MOBILE INTERNET USERS

1.3 EXABYTES



PRODUCTS ORDERED ON AMAZON PER SECOND

72.9 ITEMS



SOURCES: Cisco, eMarketer, McAfee, Facebook, Crunch, Twitter, 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 Sterenberg](#).



Search Stanford

10 results | clustering on | Search

Search The Web

10 results | clustering on | Search

Now



big data



big data industries

Remove

big data

Remove

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

Google Search

I'm Feeling Lucky

Report inappropriate predictions

Google

Say "Ok Google" 

24°C Pak Kret District

Eat & Drink Entertainment... Sports

Machine learning


Why Do Developers Find It Hard To Learn Machine Learning?

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

Forbes · 1 day ago


Data science

Docker for Data Science



Secure | <https://www.amazon.com>

Recommended for You in Kindle Books



NEW YORK TIMES BESTSELLER
ERNEST CLINE

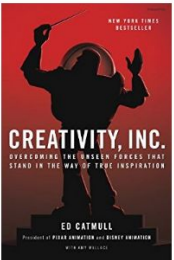
"Harari makes the biggest questions of history and of the modern world, and it is written in an engagingly child-like language."
—JAMES H. COOPER, author of *Love, Games and War*

Yuval Noah Harari

Sapiens

A Brief History of Humankind

NEW YORK TIMES BESTSELLER



ED CATMULL
President of Pixar Animation and Disney Animation

Yuval Noah Harari
New York Times Bestselling Author of *Sapiens*

Homo Deus

A Brief History of Tomorrow

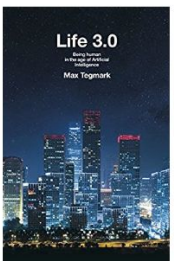
THE CLASSIC BESTSELLER
CARL SAGAN
COSMOS
WITH JOHN FRIEDMAN
AND NEIL DEGRASSA TYSON
WITH ILLUSTRATIONS BY ANN DRUYAN

CLASSIC
G
S
T

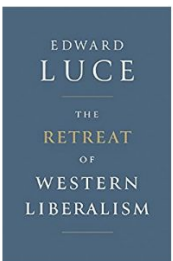
S
W
L
A
W
T
A

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

Life 3.0
How to Survive in the Age of AI
Max Tegmark



EDWARD LUCE
THE
RETREAT
OF
WESTERN
LIBERALISM

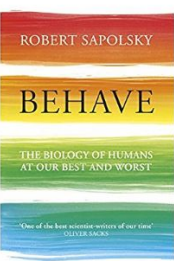


ROBERT SAPOLSKY

BEHAVE

THE BIOLOGY OF HUMANS AT OUR BEST AND WORST

"One of the best scientific writers of our time"
—DAVID SIKES




"This is that rare book that is not only brilliant but inspiring. You'll want to read it often for inspiration in the same way you'll want to read *Sapiens*, the author of *Enigmas* and *Humankind*."

the four

THE HIDDEN DNA OF AMAZON, APPLE, FACEBOOK AND GOOGLE

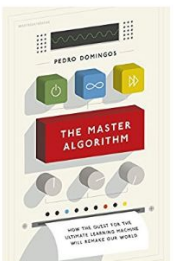
Scott Galloway



PERD DOMINGOS

THE MASTER ALGORITHM

HOW THE GOOGLES OF THE WORLD ARE MAKING MONEY AND SHAKING OUR WORLD

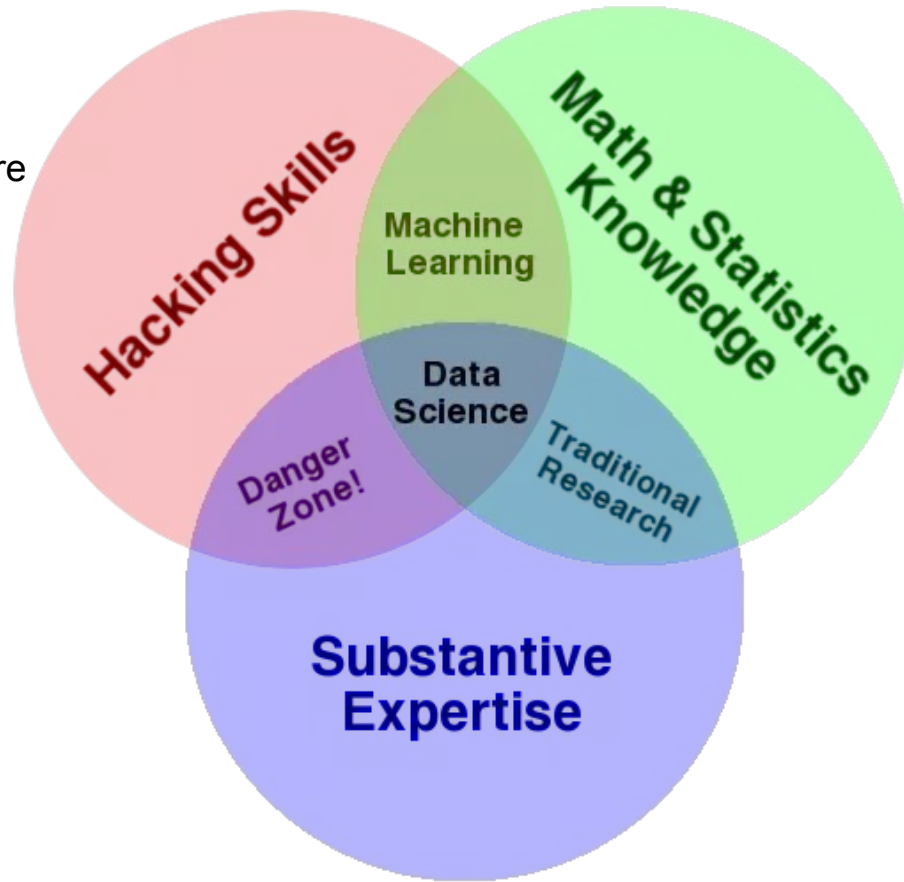


GA
Z
E
T
T
E

ST



Coding
Database
Infrastructure





Finance



Retail



Health



Media



Sports



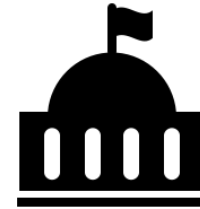
Logistics



Education

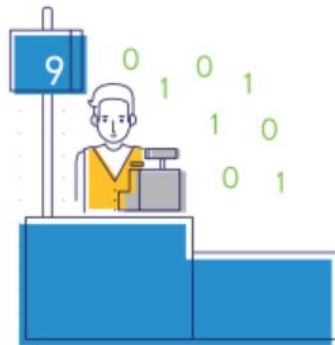


Government



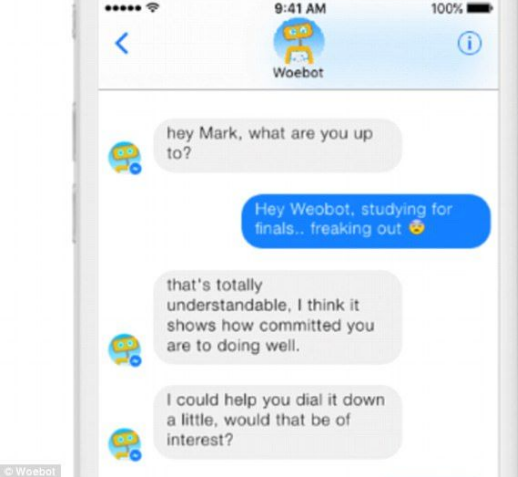
<https://www.datameer.com/product/industries-use-cases/>
<https://dataflog.com/read/8-industries-benefit-big-data-infographic/402>
<https://www.simplilearn.com/big-data-applications-in-industries-article>





A woman with curly hair, wearing a light-colored cardigan over a white top, is standing in a grocery store aisle. She is looking down at a brown paper shopping bag she is holding, and she is placing a clear plastic container of food into it. The background shows shelves stocked with various grocery items.

INTRODUCING
amazon go



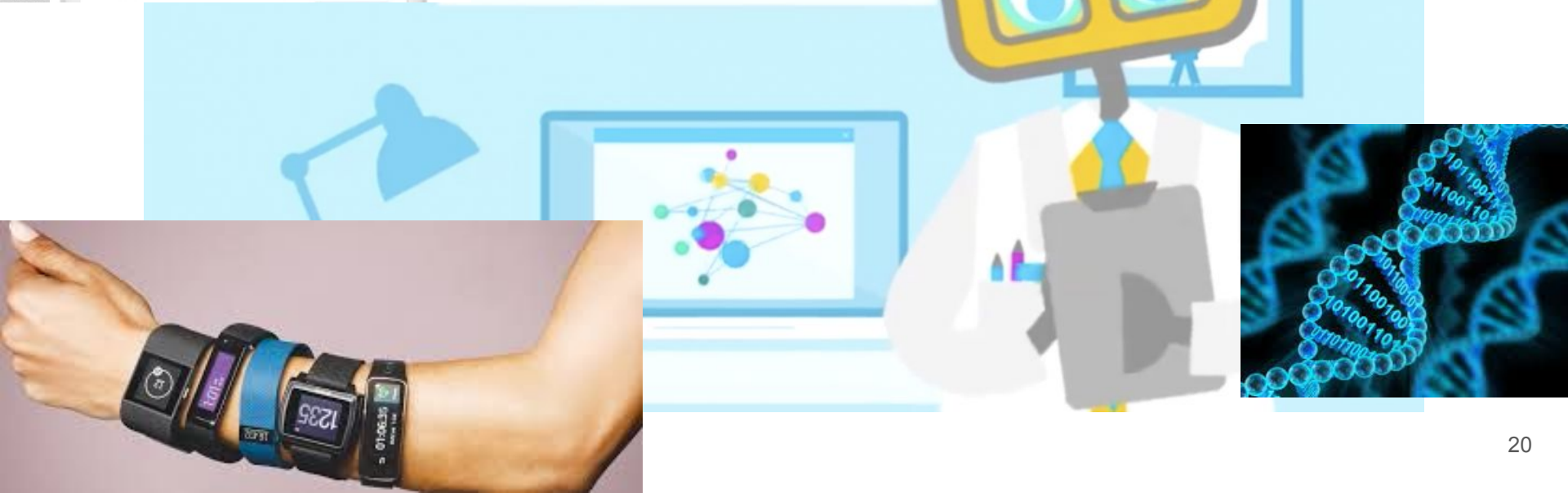
How Machine Learning Is Helping Us Predict Heart Disease and Diabetes

by Yannis Paschalidis

MAY 30, 2017



© Weebot





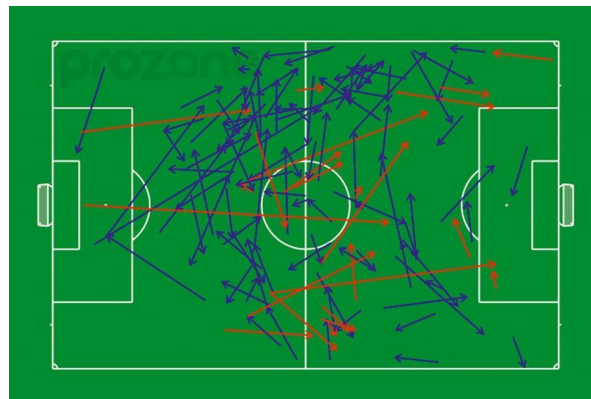
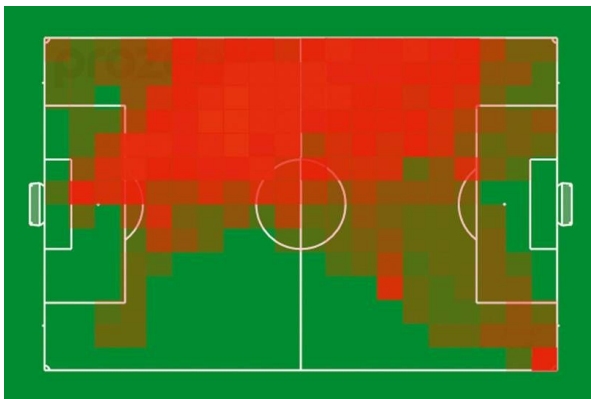




“Don’t take the risk of facing death”

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.





UEFA
CHAMPIONS
LEAGUE

FINAL



REAL MADRID **LIVERPOOL**

NSK Olympiyskiy Stadium - Kiev



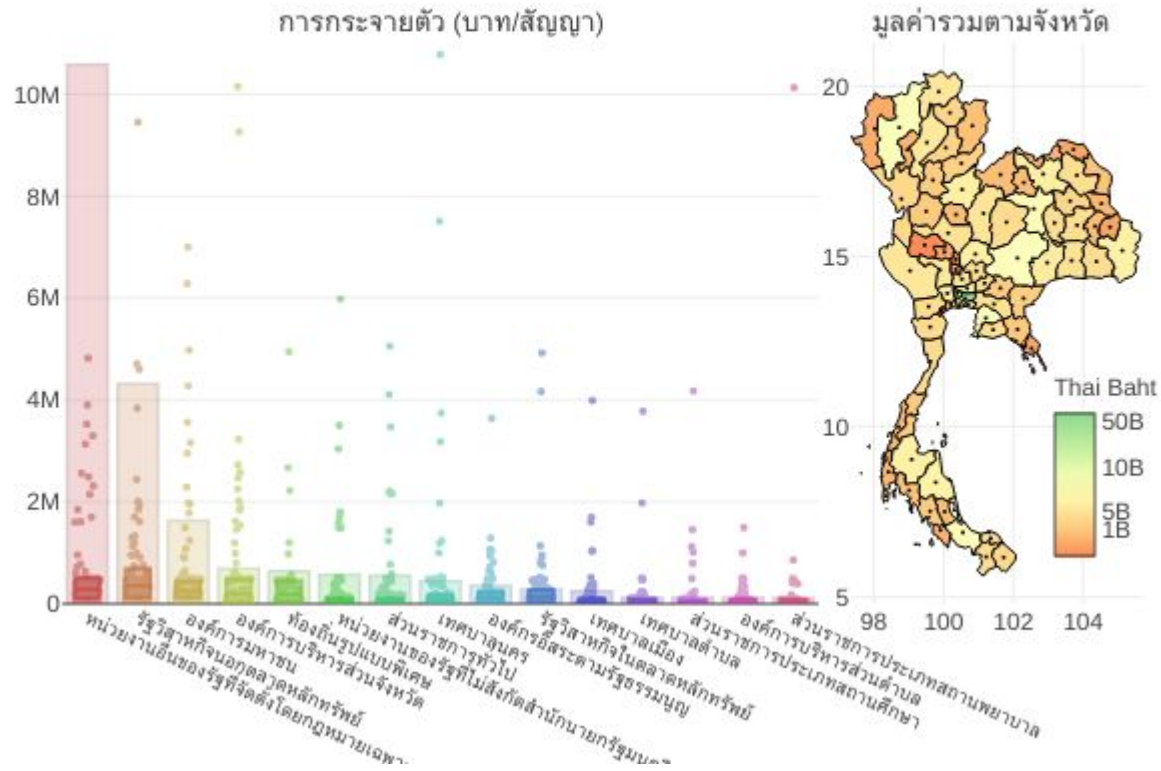




<https://www.reuters.com/article/us-india-technology-development/poor-nations-need-help-to-use-big-data-to-tackle-disease-poverty-expert-idUSKBN19V0UO>

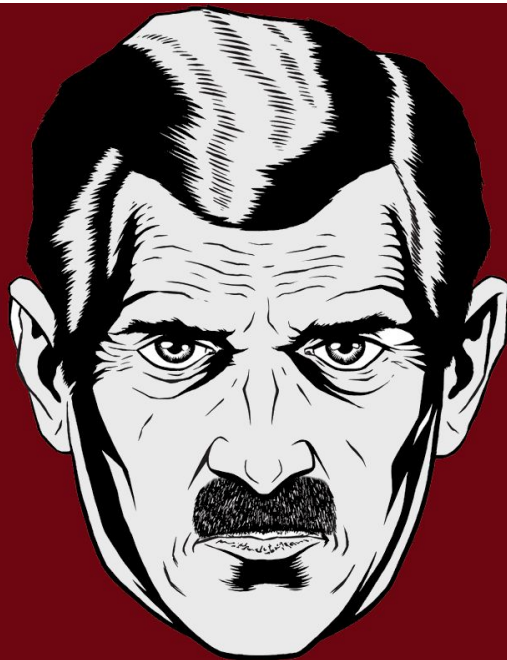


ข้อมูลคำสั่งจัดซื้อจัดจ้างโดยหน่วยงานรัฐบาล









**BIG BROTHER IS
WATCHING YOU**



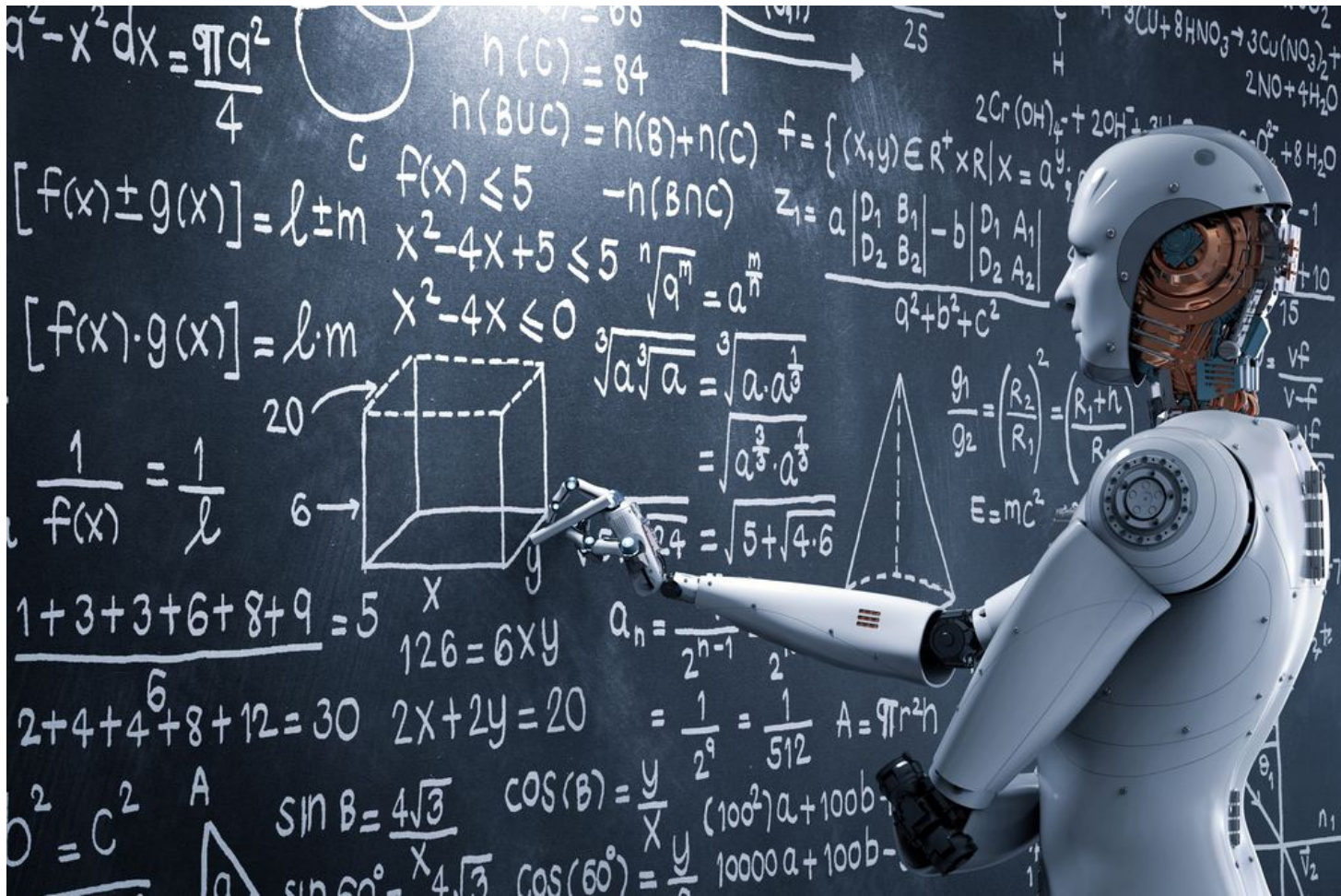
Intro to data science

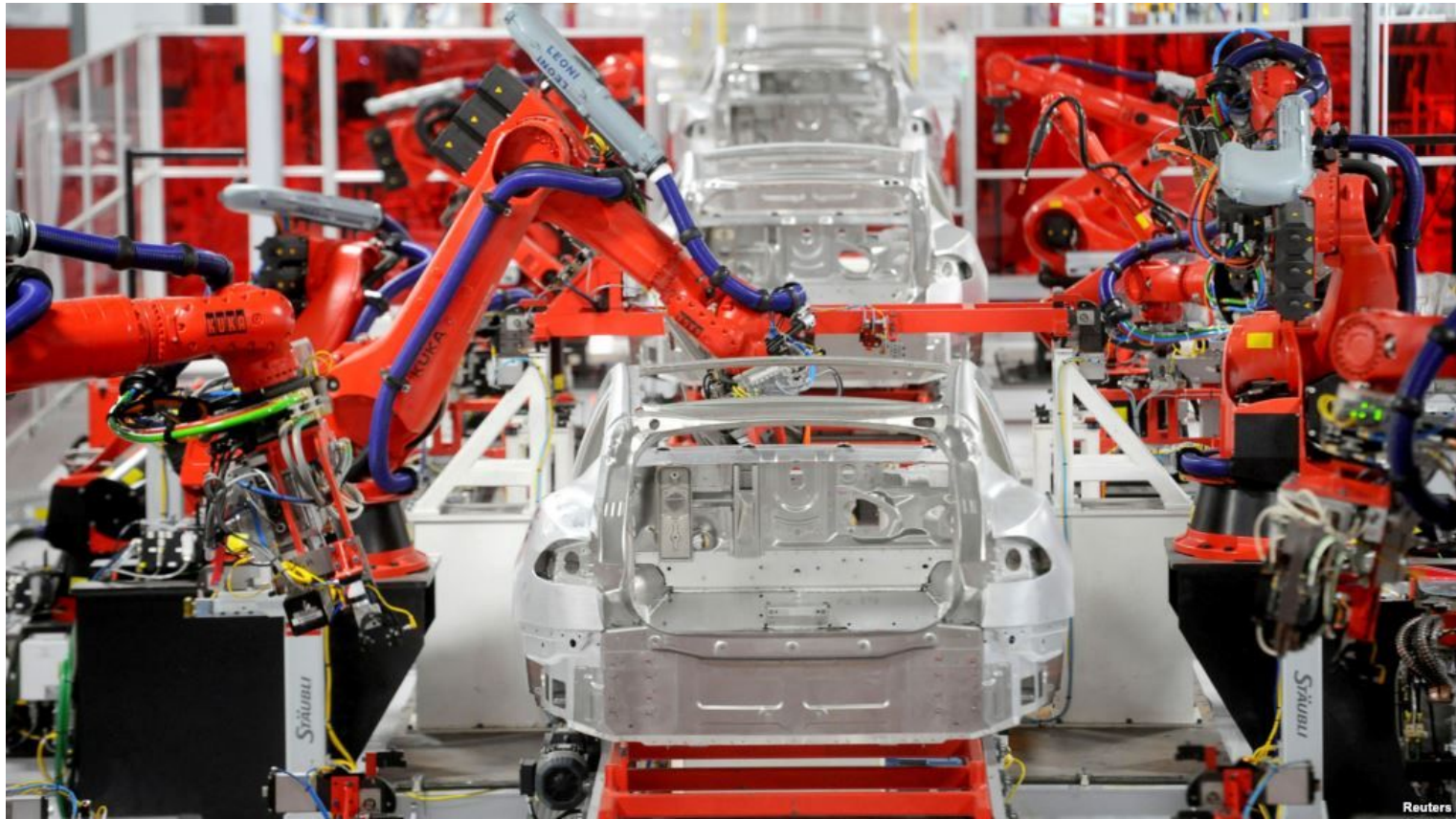
Development, use cases, and prospect



Intro to data science

Development, use cases, and prospect





Reuters

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



<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

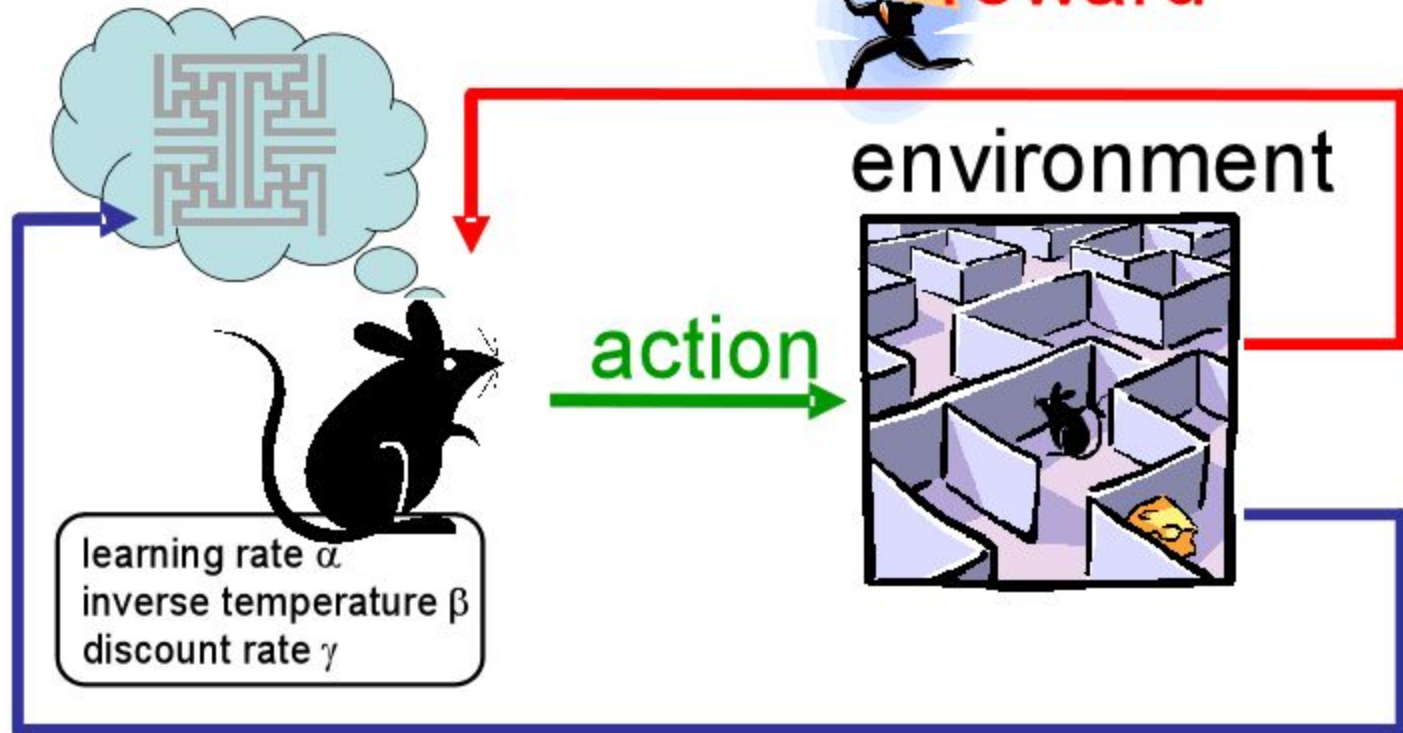
中国乌镇 围棋峰会

The Future of Go Summit in Wuzhen





internal state



observation

ax: ay: 1, 2
Deaths/HighScore: 24 / 61

17







\$24,000

Who is Stoker?
(FOR ONE WELCOME OUR
NEW COMPUTER OVERLORDS)
\$1,000

\$77,147

Who is Bram
Stoker?
\$17,973

\$21,600

WHO IS
BRAM STOKER?
\$5600

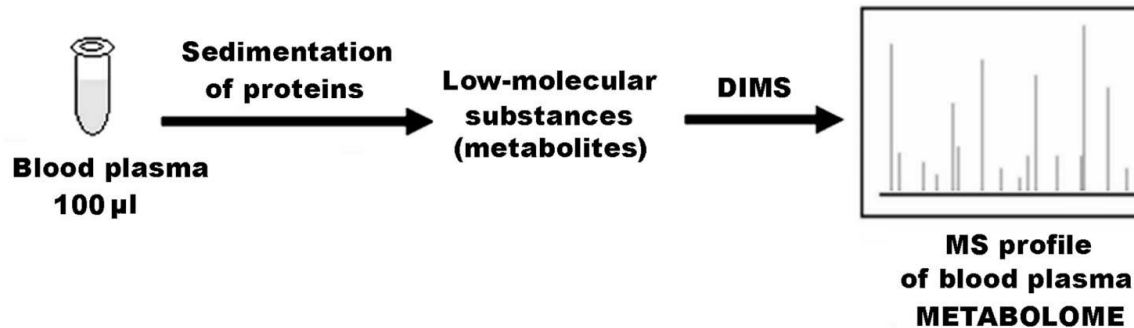


IBM Watson Health

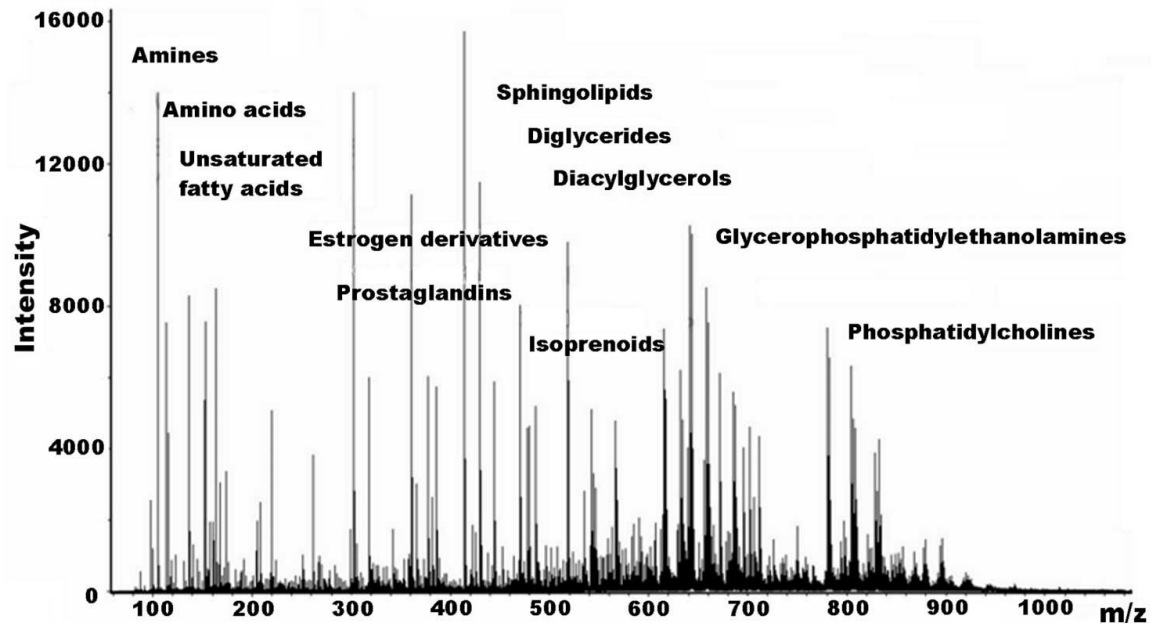


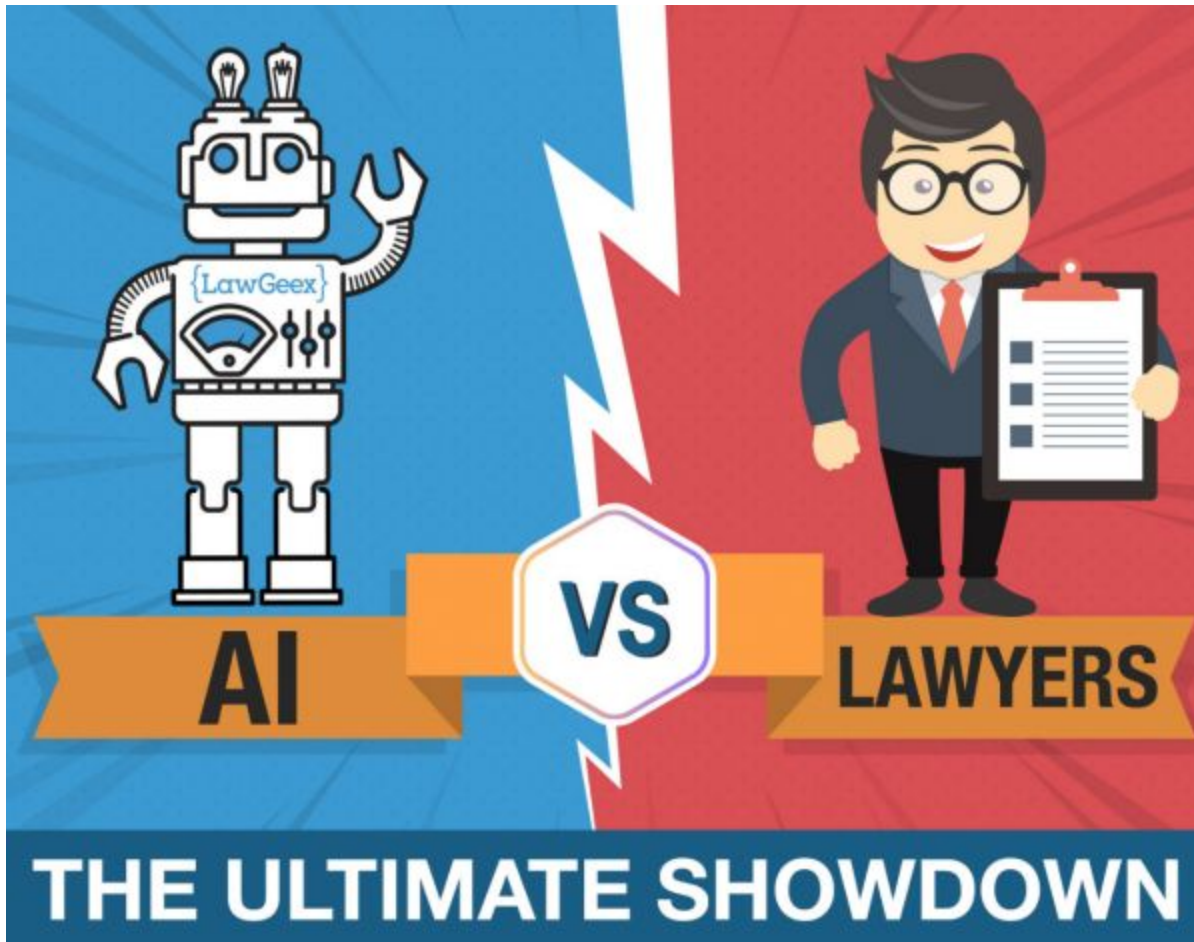
Who is Bram
Stoker?
\$ 17,973

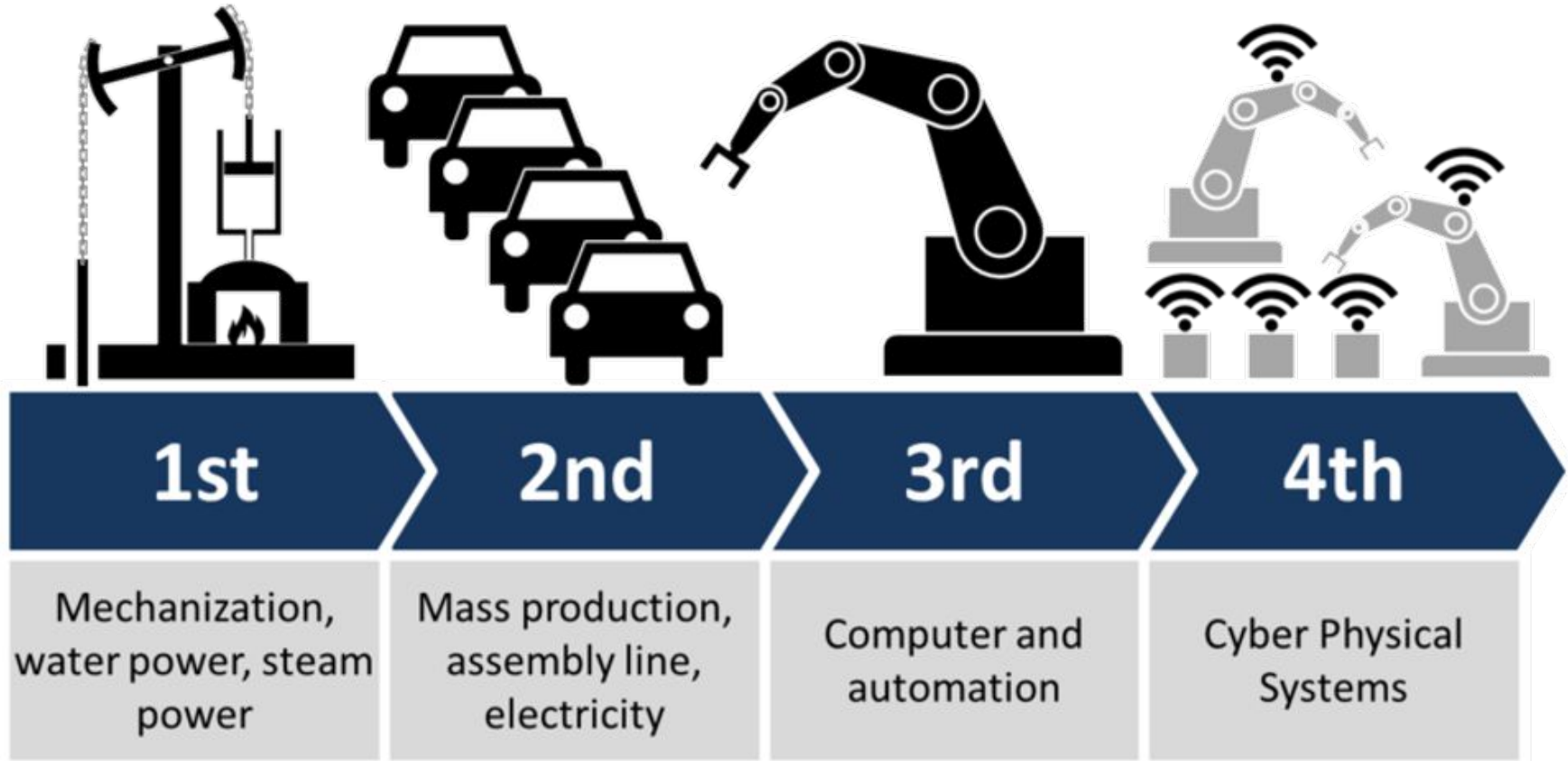
A



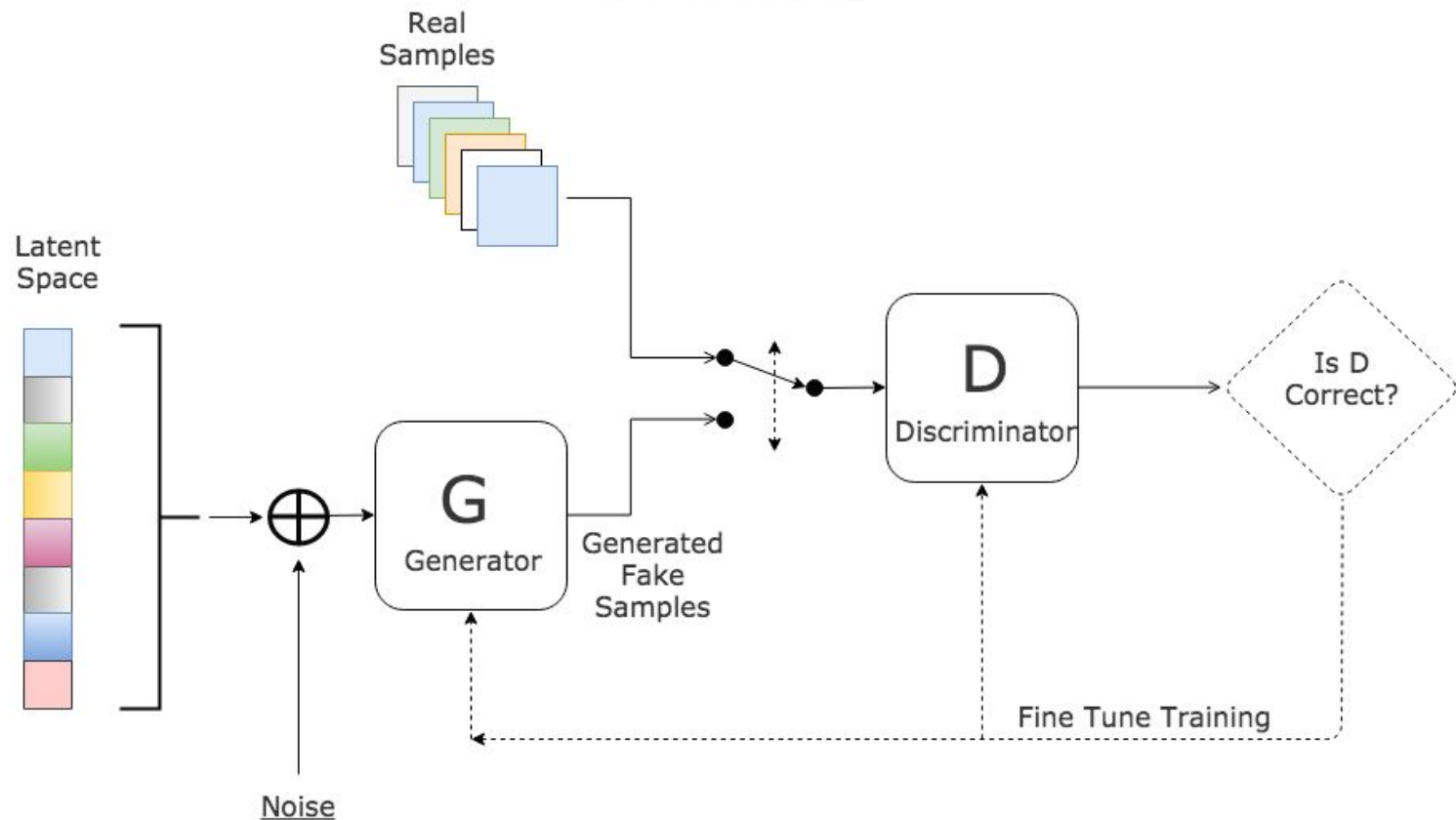
B

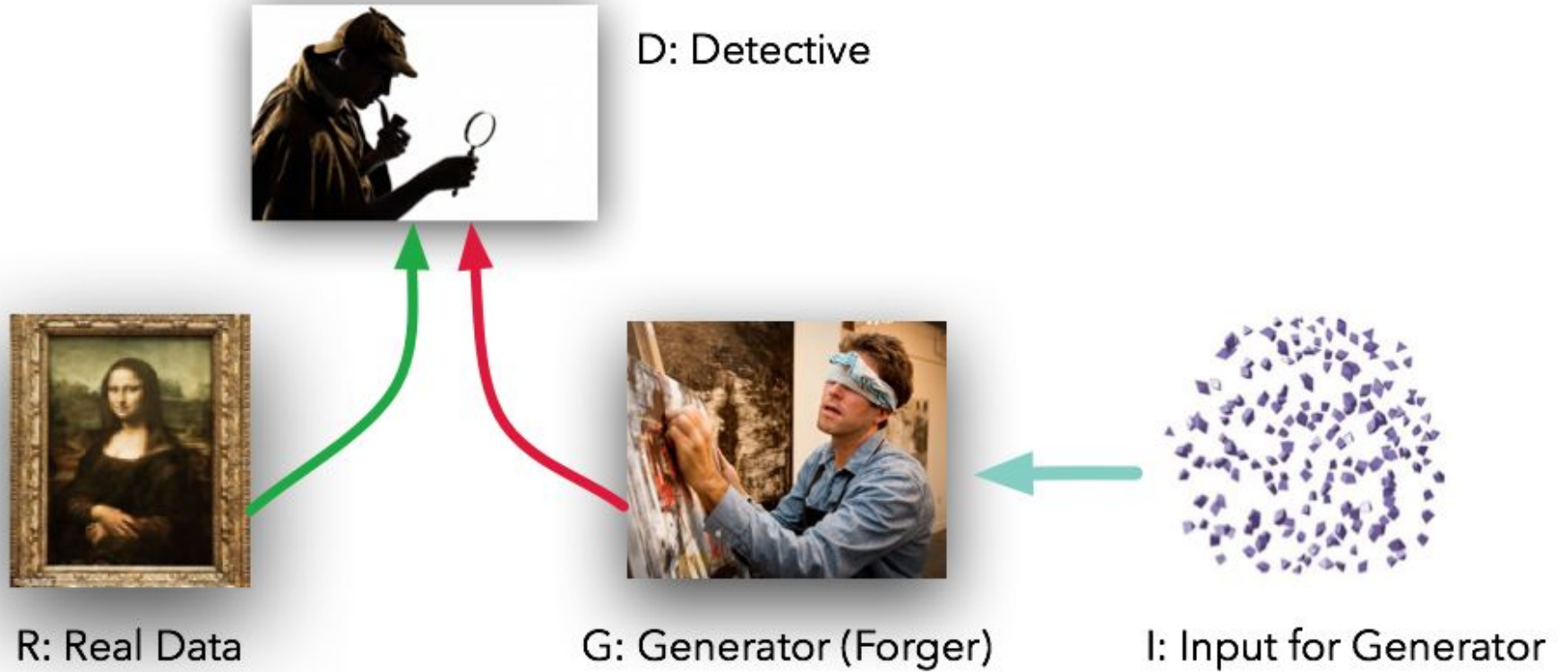




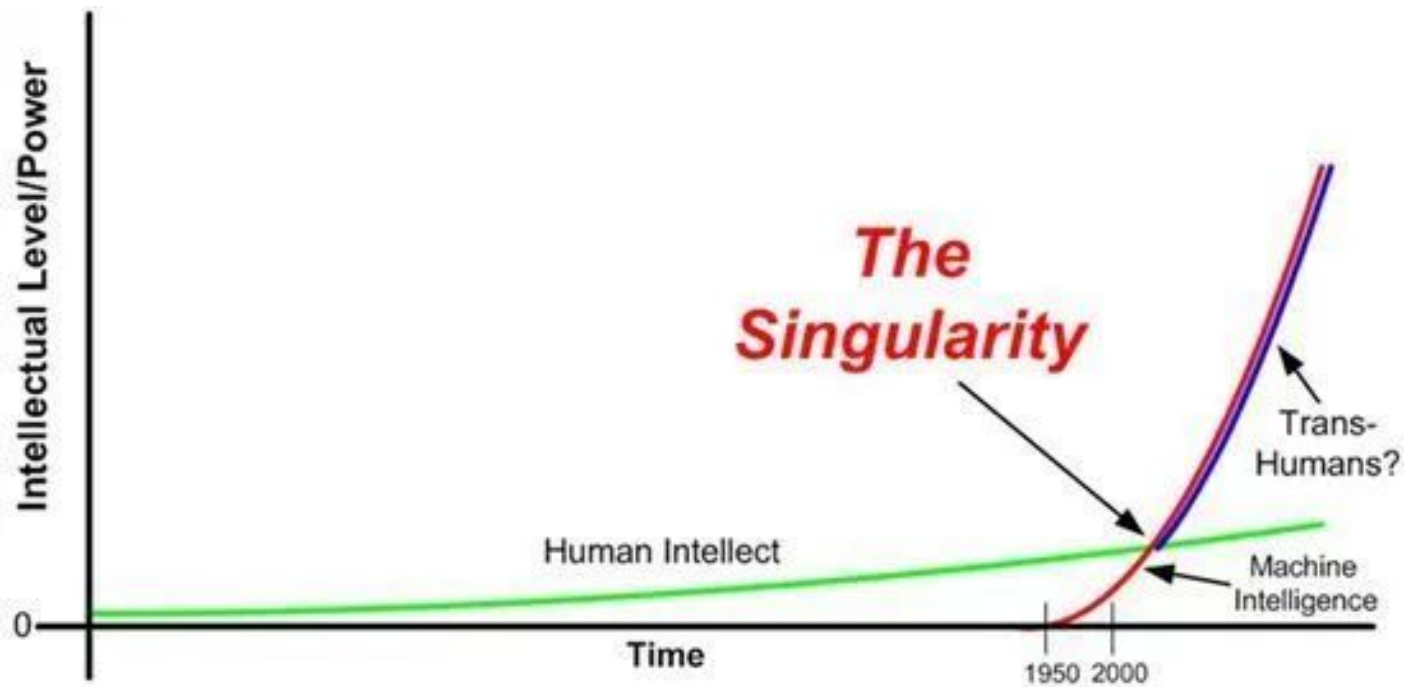


Generative Adversarial Network

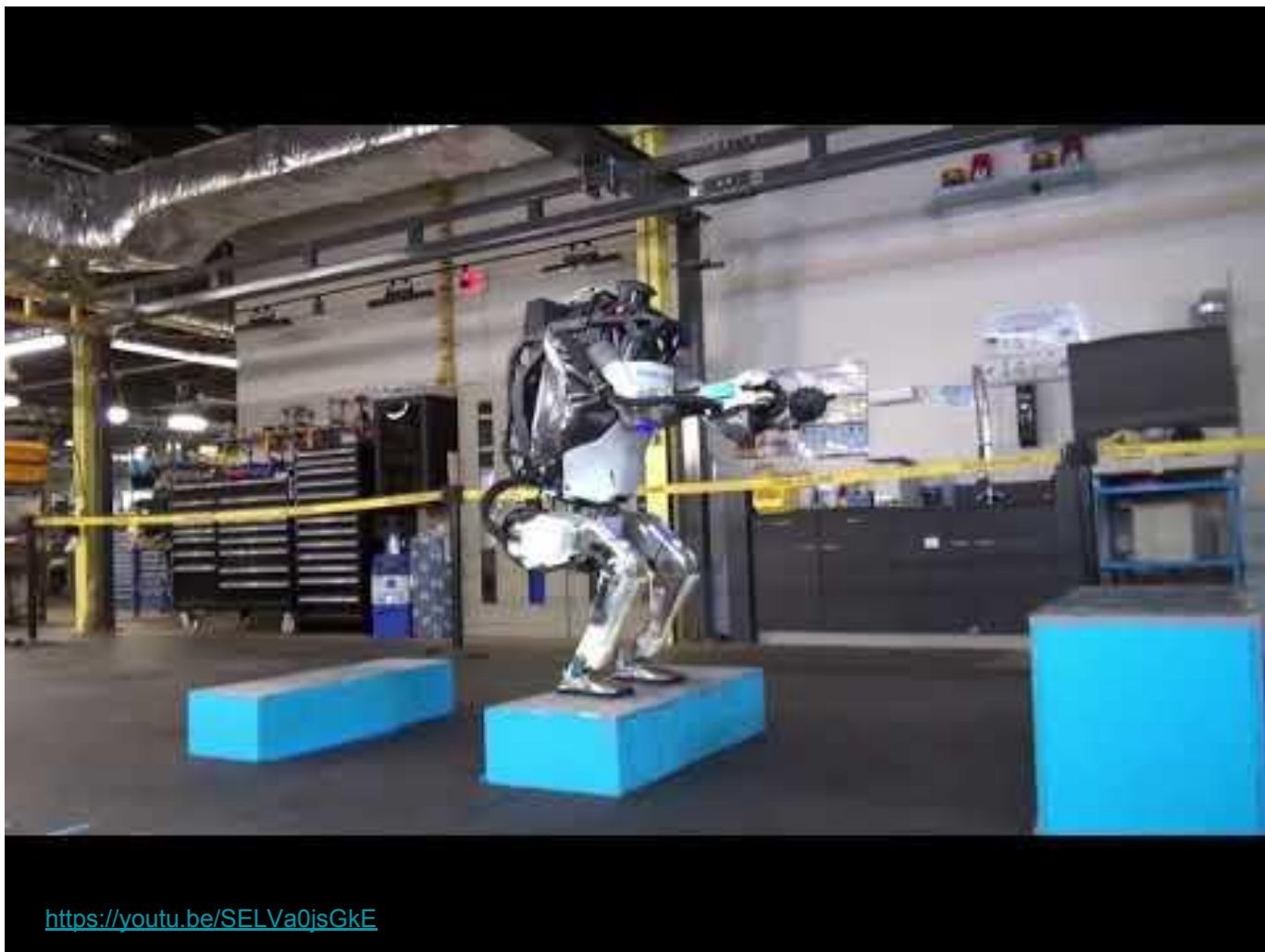












<https://youtu.be/SELVa0jsGkE>



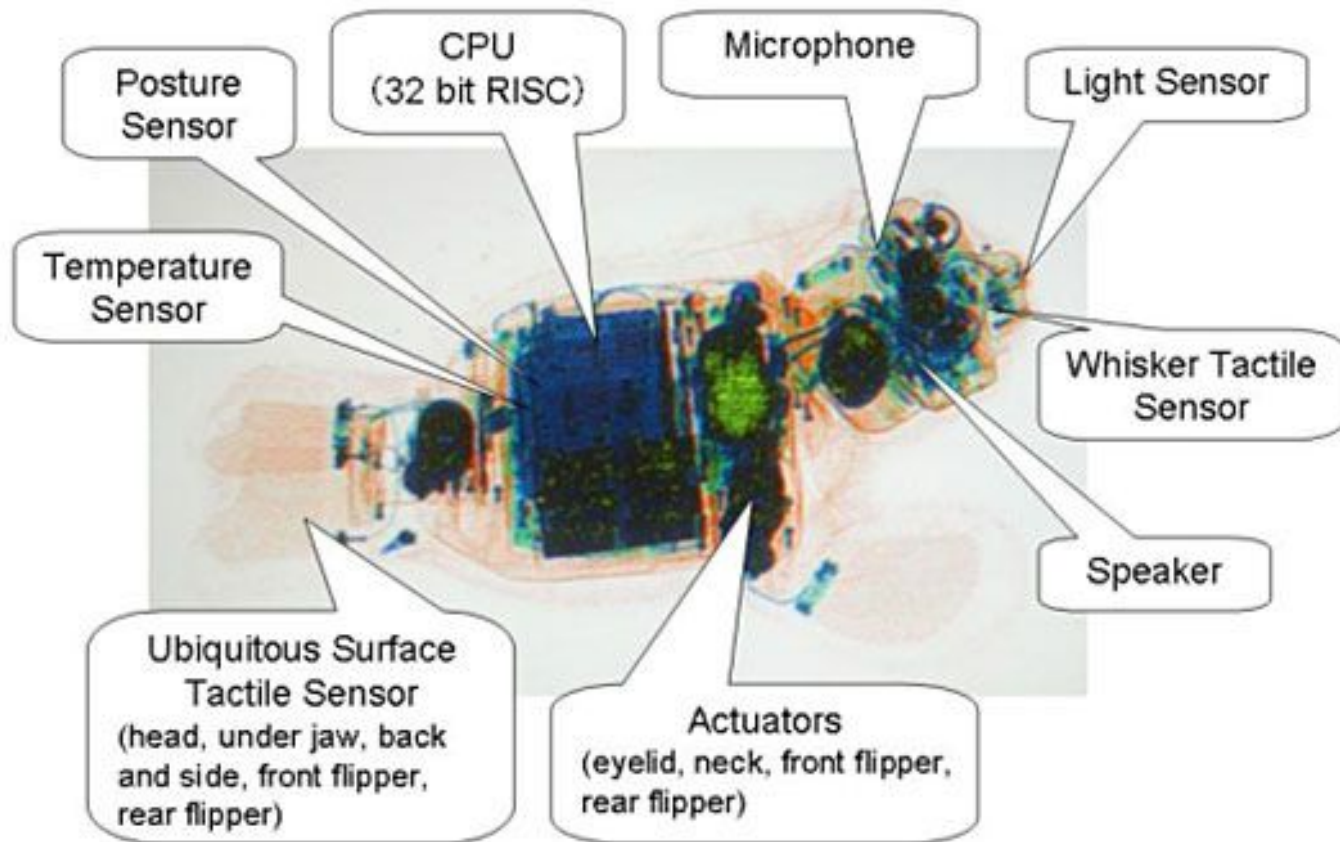
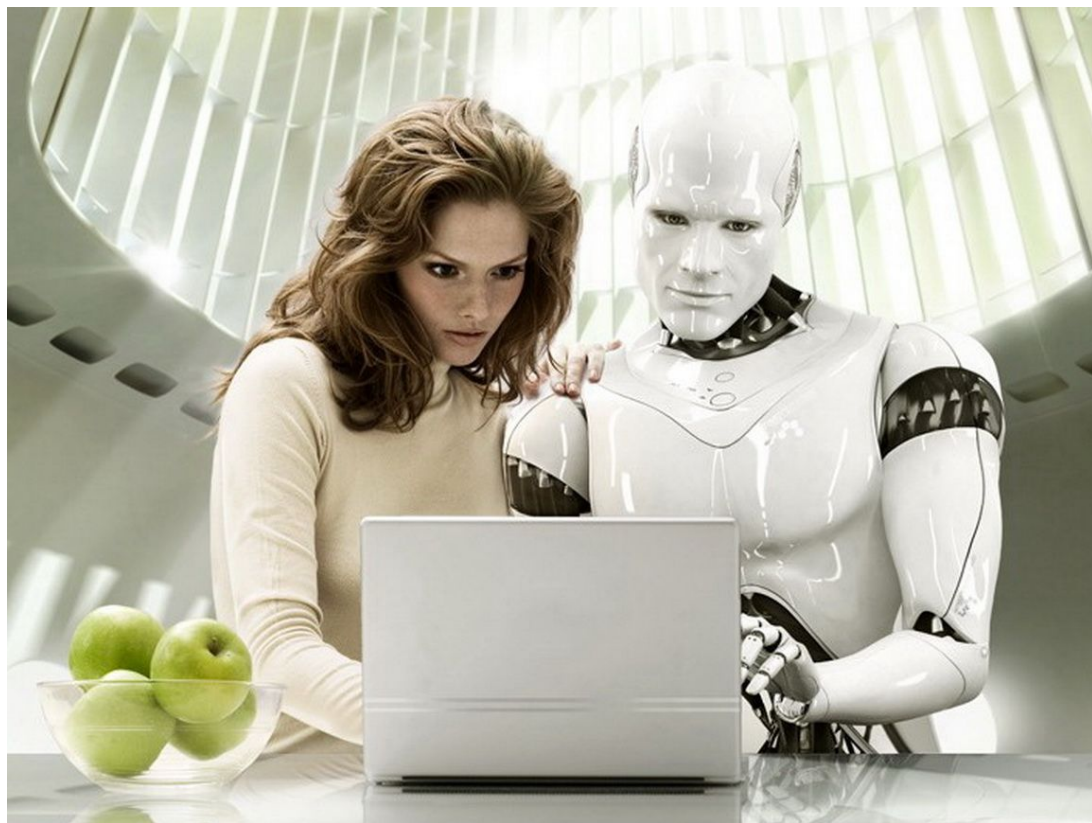


Figure 2: Paro's function







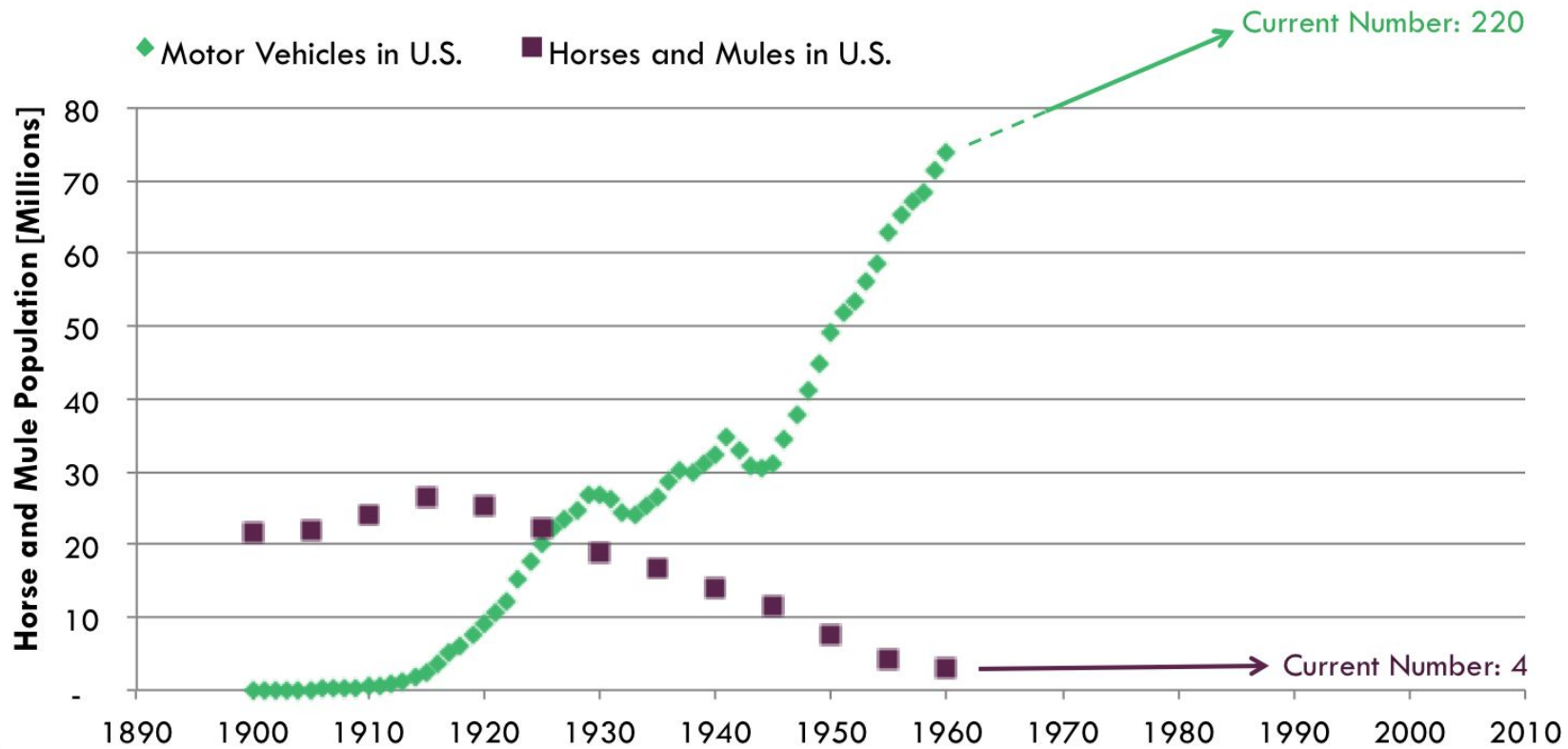
Humans Need Not Apply



<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

Watch

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

