



# EEC1509 - Machine Learning

## Lesson #1 - Outline & Directions

Ivanovitch Silva  
August, 2018





# Introduction

---





Ivanovitch Silva ([ivan@imd.ufrn.br](mailto:ivan@imd.ufrn.br))

Office: CIVT-A226 6T1234



Group Survey

Who am I?  
My master/phd research is about ....

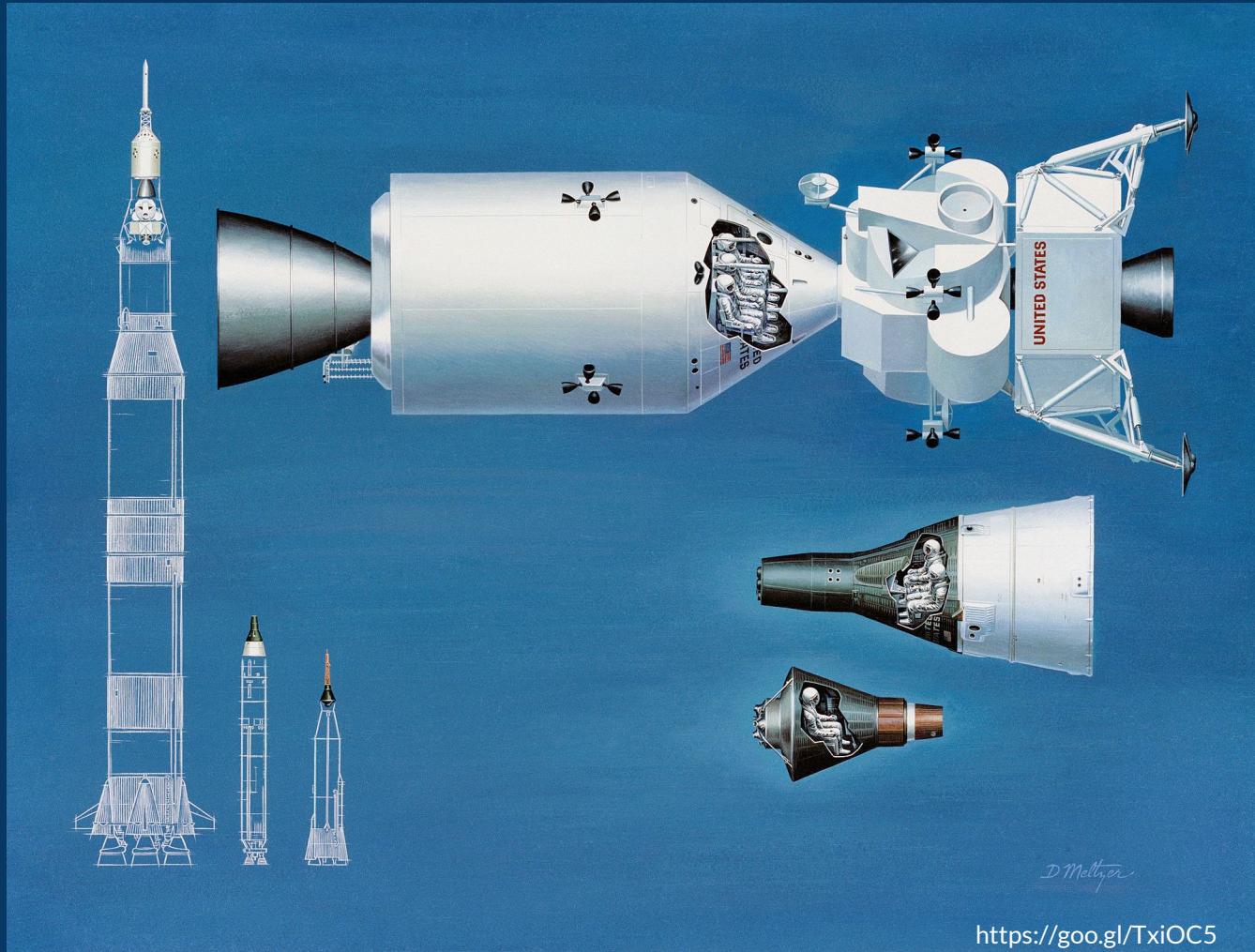


# Provocation #1

hardware

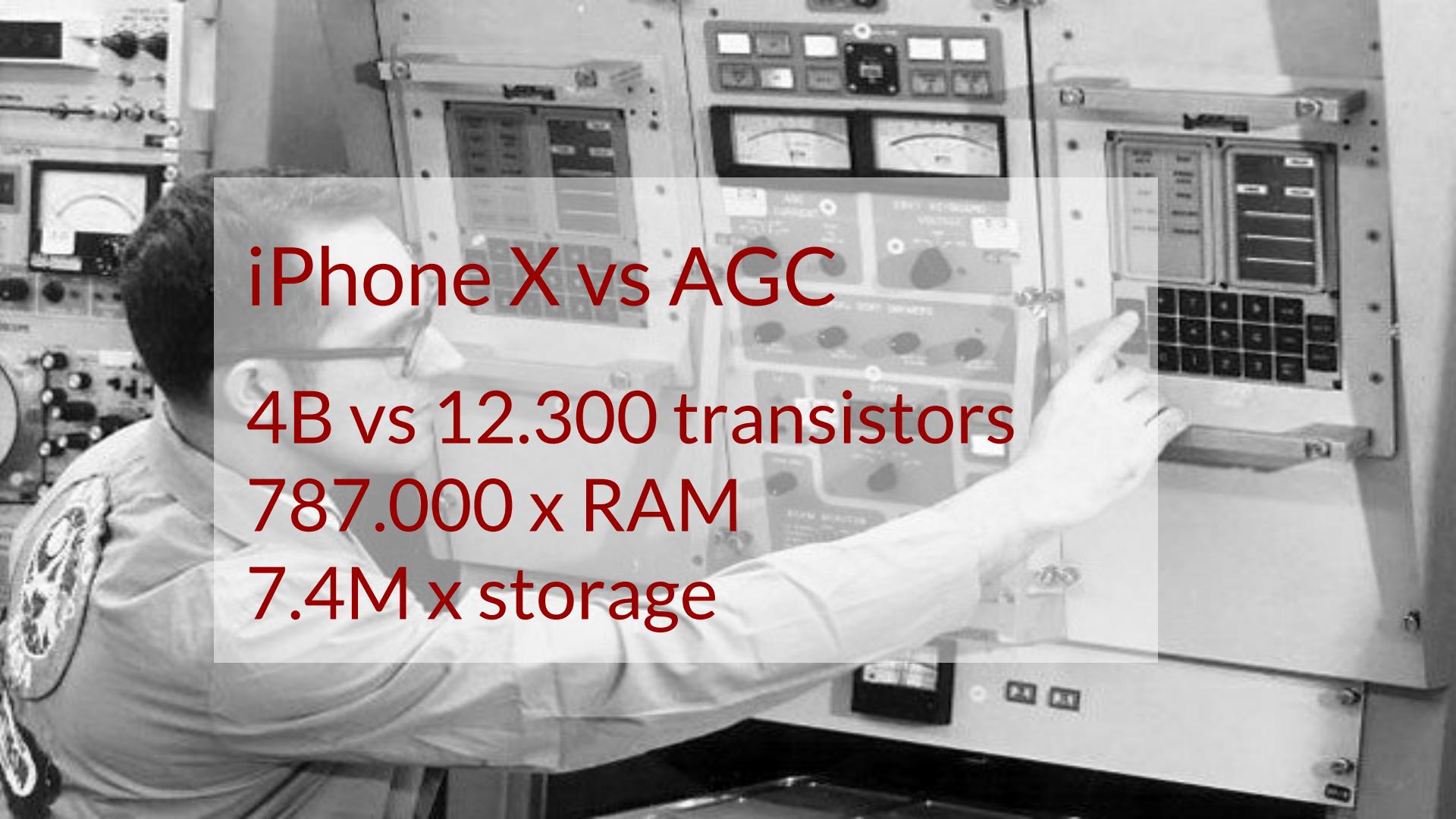


1969



D. Meltzer

<https://goo.gl/TxiOC5>



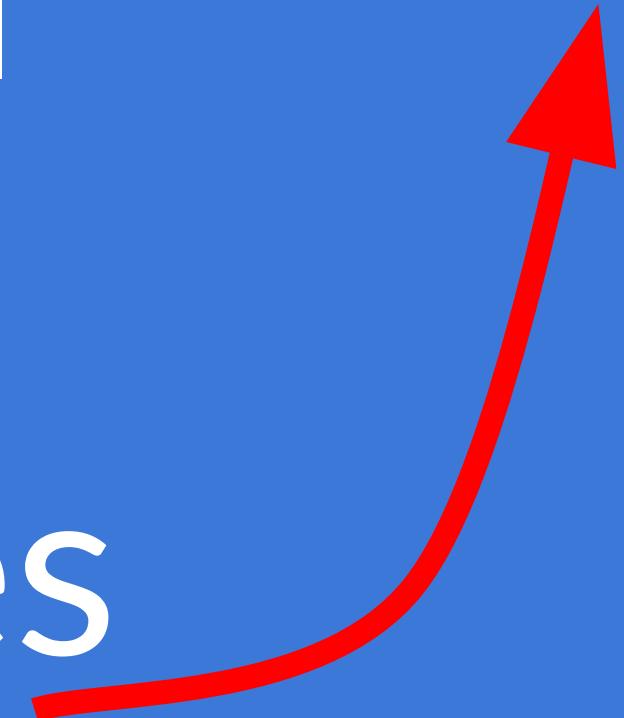
# iPhone X vs AGC

4B vs 12.300 transistors

787.000 x RAM

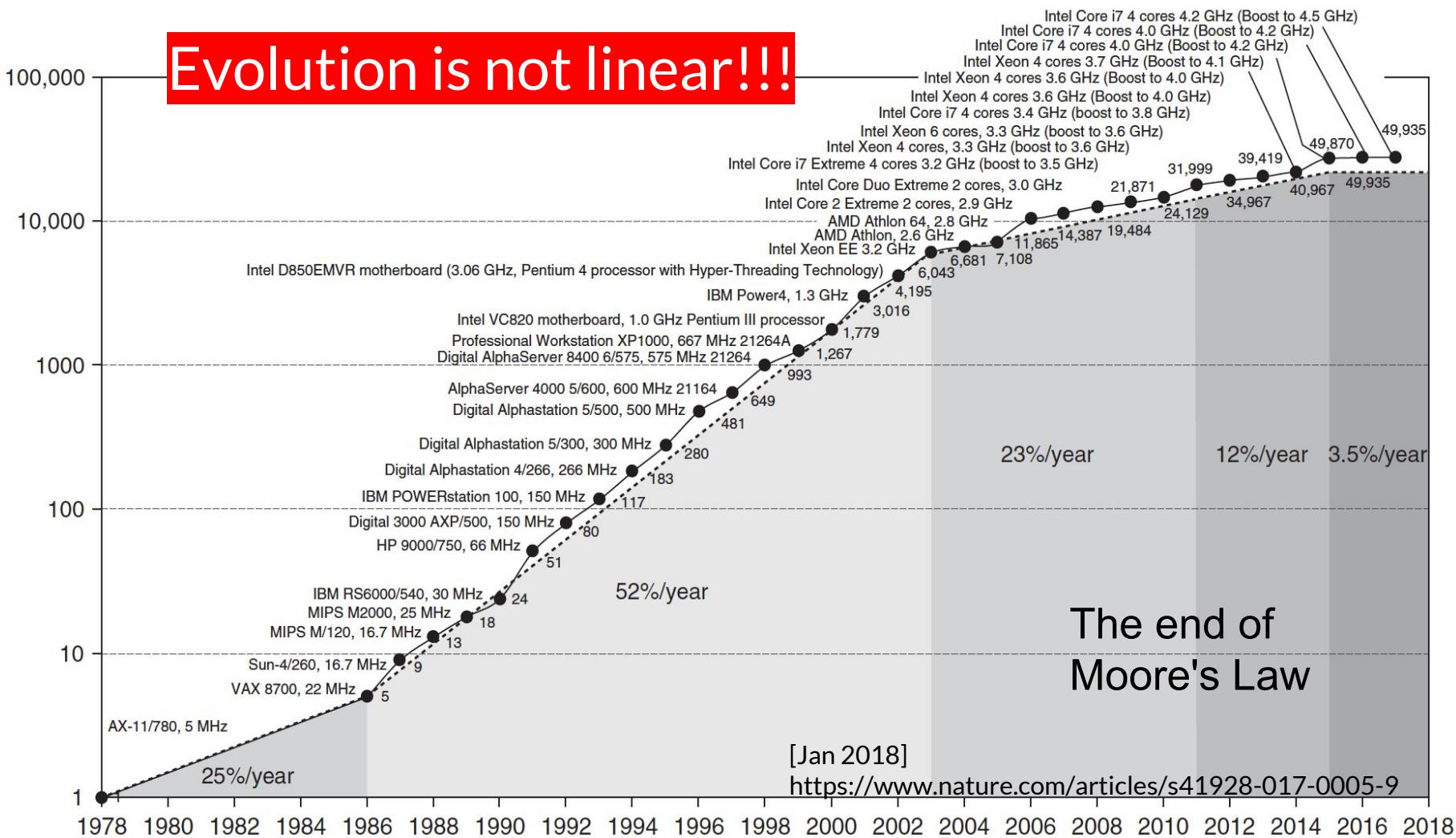
7.4M x storage

exponential  
growth of  
technologies



# Evolution is not linear!!!

Performance (vs. VAX-11/780)





\$ 2,999.<sup>00</sup>

Architecture

Frame Buffer

Boost Clock

Tensor Cores

CUDA Cores

## NVIDIA TITAN V

The Most Powerful PC GPU Ever Created

NVIDIA TITAN V is the most powerful graphics card ever created for the PC, driven by the world's most advanced architecture—NVIDIA Volta. NVIDIA's supercomputing GPU architecture is now here for your PC, and fueling breakthroughs in every industry.

[LEARN MORE](#)

**NVIDIA TITAN V**

**NVIDIA Volta**

**12 GB HBM2**

**1455 MHz**

**640**

**5120**

**1. COTAÇÕES**

	USD - Dólar:	3.58
	EUR - Euro:	4.25
	GBP - Libra Esterlina:	4.84

**2. ALÍQUOTA DE ICMS**

Tipo de envio:	Courier
Unidade federativa:	RN
Valor da alíquota:	17%

**3. TAXAS**

Conversão monetária:	3.58
Imposto de importação (%):	60
ICMS (%):	0
IOF (%):	6.38

**4. IMPOSTO DE IMPORTAÇÃO****VALORES**

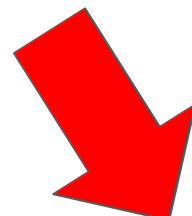
Valor do produto (USD):	2999
Custo do frete (USD):	50
Valor do produto (BRL):	10736.42
Custo do frete (BRL):	179.00
<b>TOTAL DA COMPRA (BRL):</b>	<b>10915.42</b>

**TRIBUTOS**

<input checked="" type="checkbox"/> Incluir frete no cálculo do imposto	<input checked="" type="checkbox"/> Incluir IOF
Imposto de importação (BRL):	6549.25
ICMS (BRL):	0.00
IOF (BRL):	696.40
<b>TOTAL DE TRIBUTOS (BRL):</b>	<b>7245.66</b>

**TOTAIS**

**Valor final com impostos (BRL):** 18161.08



# Lara Croft has changed over 21 years



1/6 cost | 1/20 power | 4 hacks in a box



# Provocation #2

data & internet

# 90's

"Read"  
Search Engine  
Google

# Update

# 00's

"Write"  
Social Networks  
Facebook

## Participate

# 10's

"Act"  
App  
Uber

Act

# 20's

"Change"  
AI  
?

## Transform

Change



Act

UBER

Write

facebook

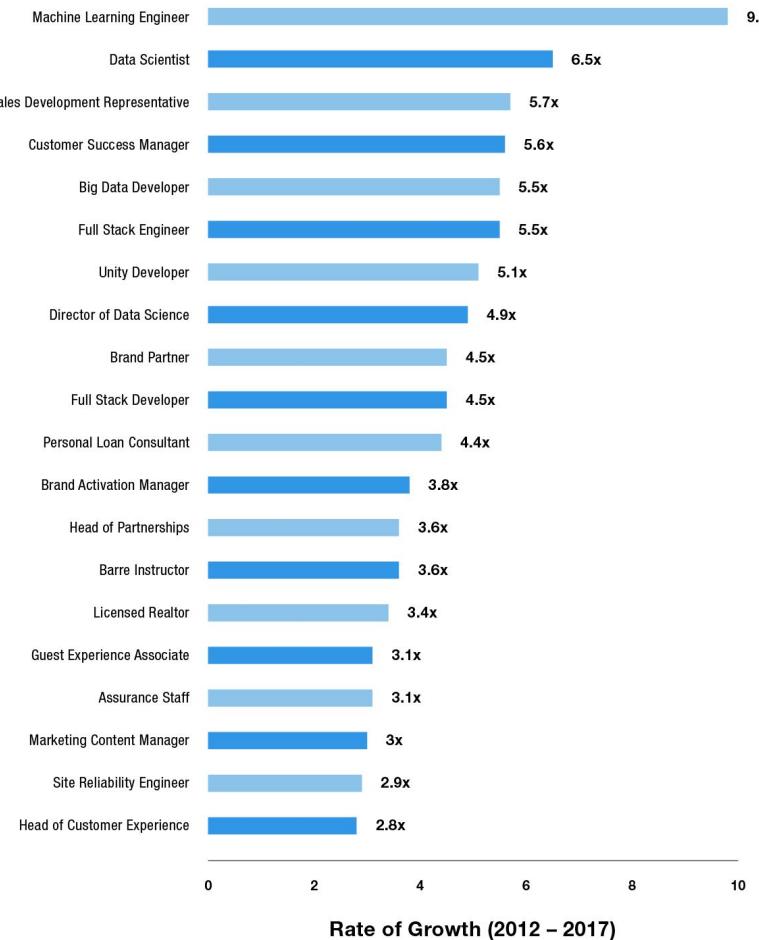
Read

Google



# THE YEAR OF INTELLIGENCE

## Top 20 Emerging Jobs



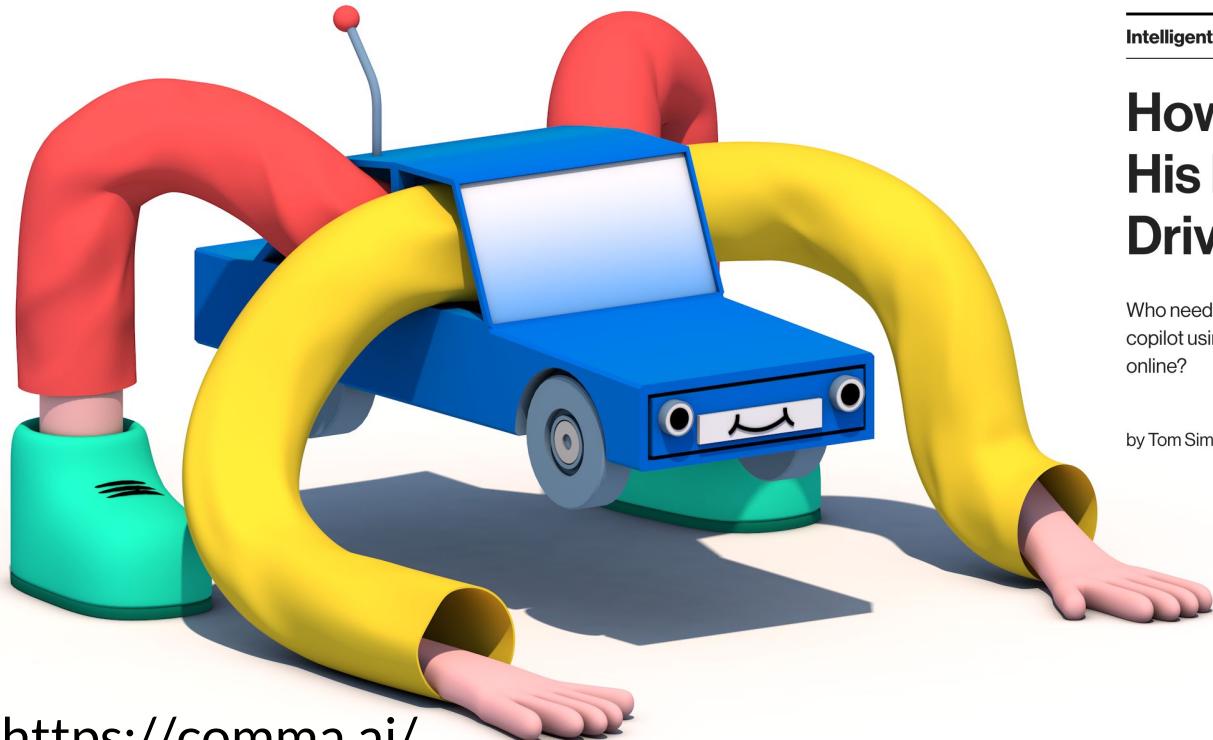
There are **9.8 times more** Machine Learning Engineers working today than five years ago based on LinkedIn's research

[Dec. 2017]  
<http://bit.do/forbesjobs>



Do-it-yourself artificial intelligence

We want to put AI into the maker toolkit, to help you solve real problems that matter to you and your communities. These kits will get you started by adding natural human interaction to your maker projects.



<https://comma.ai/>

---

Intelligent Machines

---

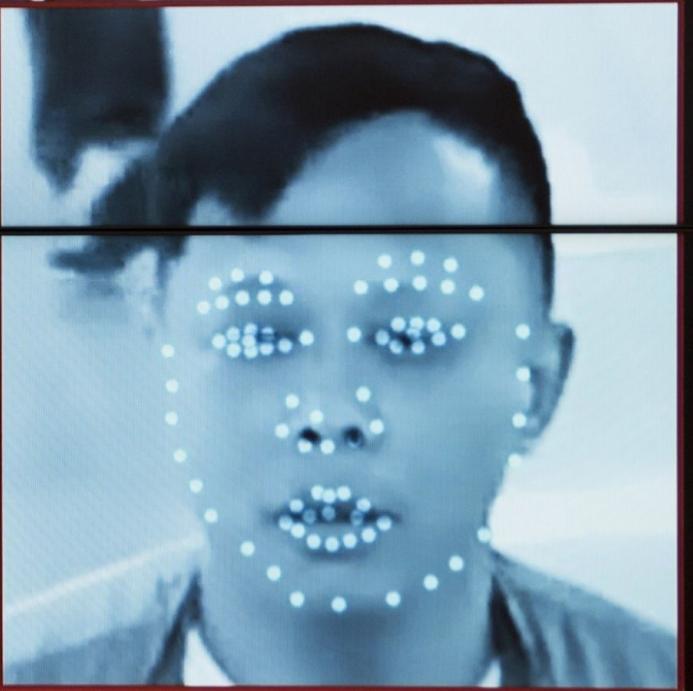
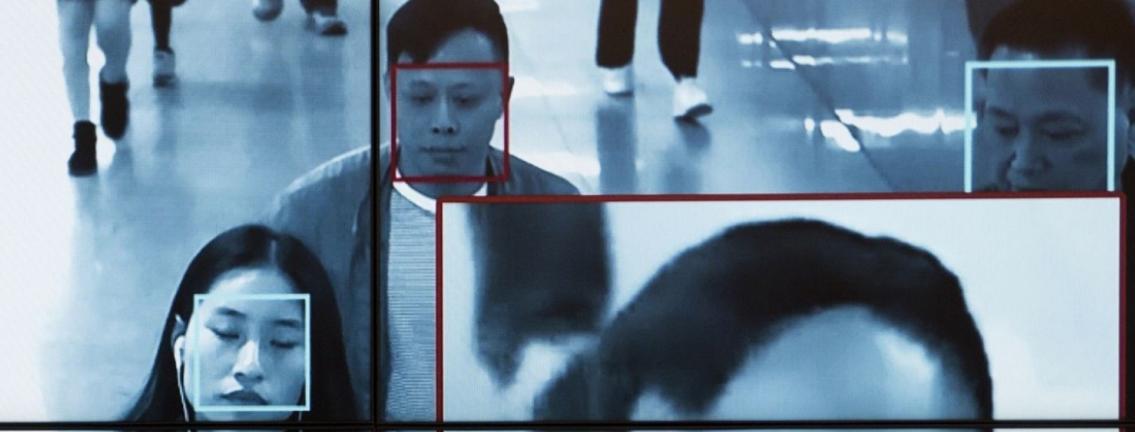
## How a College Kid Made His Honda Civic Self-Driving for \$700

Who needs a Tesla when you can build your own automated copilot using free hardware designs and software available online?

by Tom Simonite February 21, 2017

# AI arms race





China's  
watchful eye

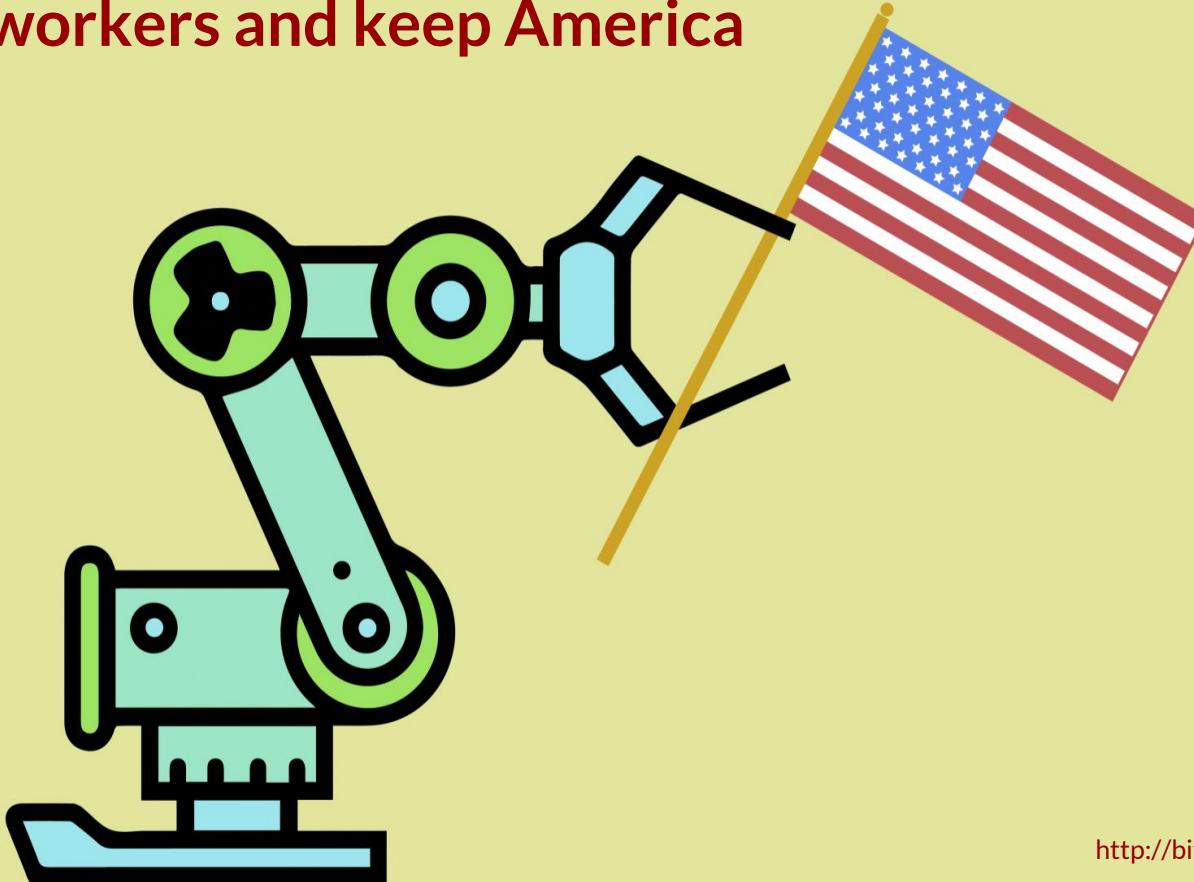
相似度  
SIMILARITY

67.2%

<https://www.ft.com/content/e33a6994-447e-11e8-93cf-67ac3a6482fd>

The White House says a new AI task force  
will protect workers and keep America  
first.

May 10, 2018



[http://bit.do/trump\\_aiprogram](http://bit.do/trump_aiprogram)

# HUMANITY

FRENCH STRATEGY  
FOR ARTIFICIAL  
INTELLIGENCE

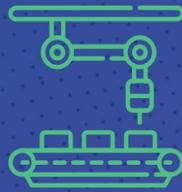
The President of the French Republic presented **his vision and strategy** to make France a **leader in artificial intelligence (AI)** at the Collège de France on **29 March 2018**.



Download the  
Villani Report

<https://www.aiforhumanity.fr/en/>





# *Agenda brasileira para a Indústria 4.0*

O Brasil preparado para os desafios do futuro

[CONHEÇA A AGENDA](#)



MENU

INTERNET DAS  
COISAS: UM  
PLANO DE AÇÃO  
PARA O BRASIL

## Estudo “Internet das Coisas: um plano de ação para o Brasil”



# Internet das coisas: um plano de ação para o Brasil



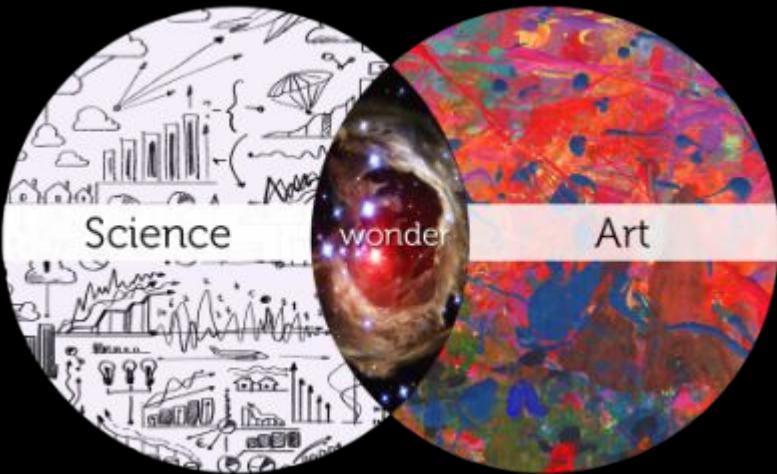
Baixe o Relatório de Plano de Ação (PDF - 3,3 MB) para o desenvolvimento de IoT no Brasil.



**NEURAL**



**SABOTAGE**



<https://www.youtube.com/watch?v=SOtm7vylwxc>





# Spotify MACHINE LEARNING DAY

MONDAY, JULY THE 9<sup>TH</sup>  
STOCKHOLM, SWEDEN



# HOSPITAL SÍRIO-LIBANÊS

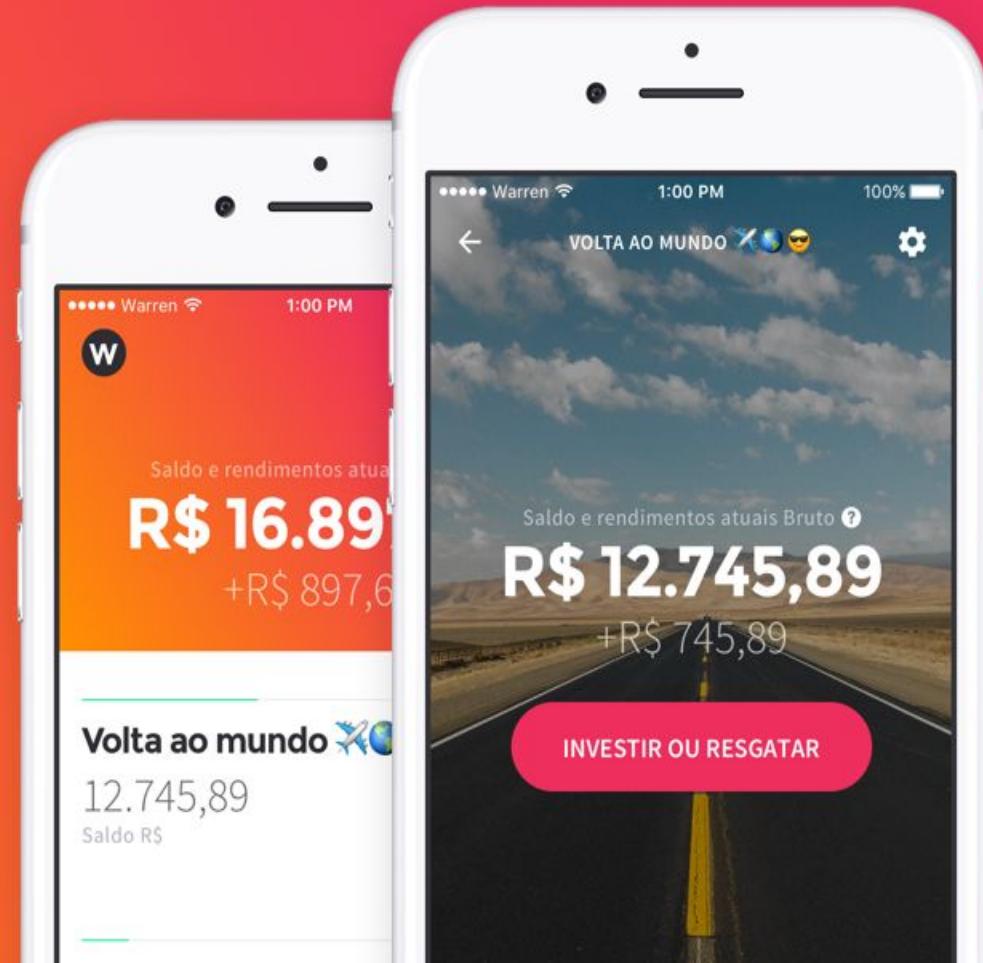
KUNUMI

<http://bhetc.org.br/empresas-do-bhetc/kunumi/>



<https://www.ufmg.br/online/radio/arquivos/046358.shtml>

# Uma nova forma de investir.



# Niantic is going to crowdsource AR maps



**90% of the data in the world today  
has been created in the last two years alone**

# This volume is continuing to grow

Data volume will grow

**800%**

over the next five years

gartner

By 2022

**93%**

of the data will be free-form

IDC

# THE COMING FLOOD OF DATA IN AUTONOMOUS VEHICLES

RADAR

~10-100 KB  
PER SECOND

SONAR

~10-100 KB  
PER SECOND

GPS

~50KB  
PER SECOND

CAMERAS

~20-40 MB  
PER SECOND

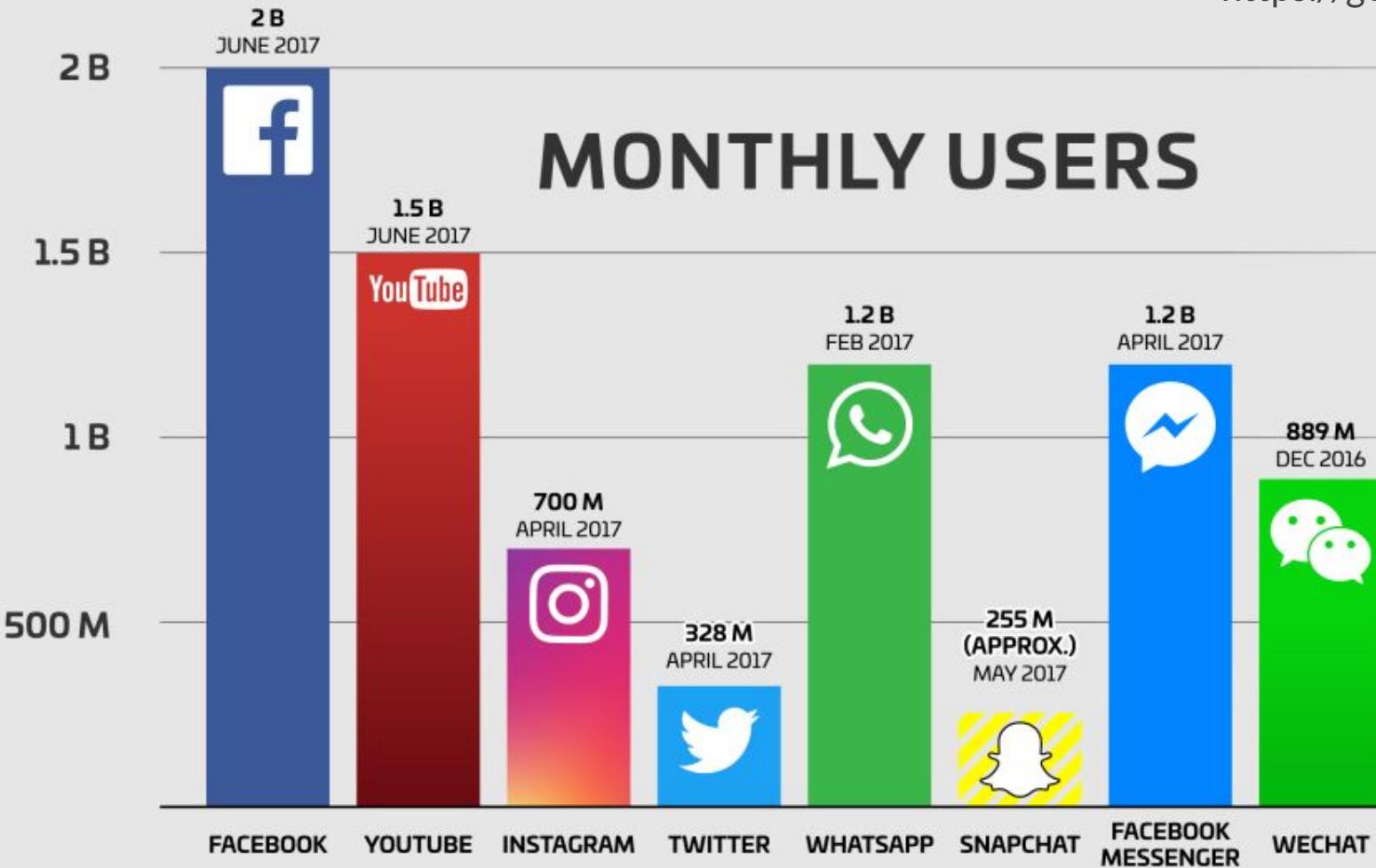
LIDAR

~10-70 MB  
PER SECOND

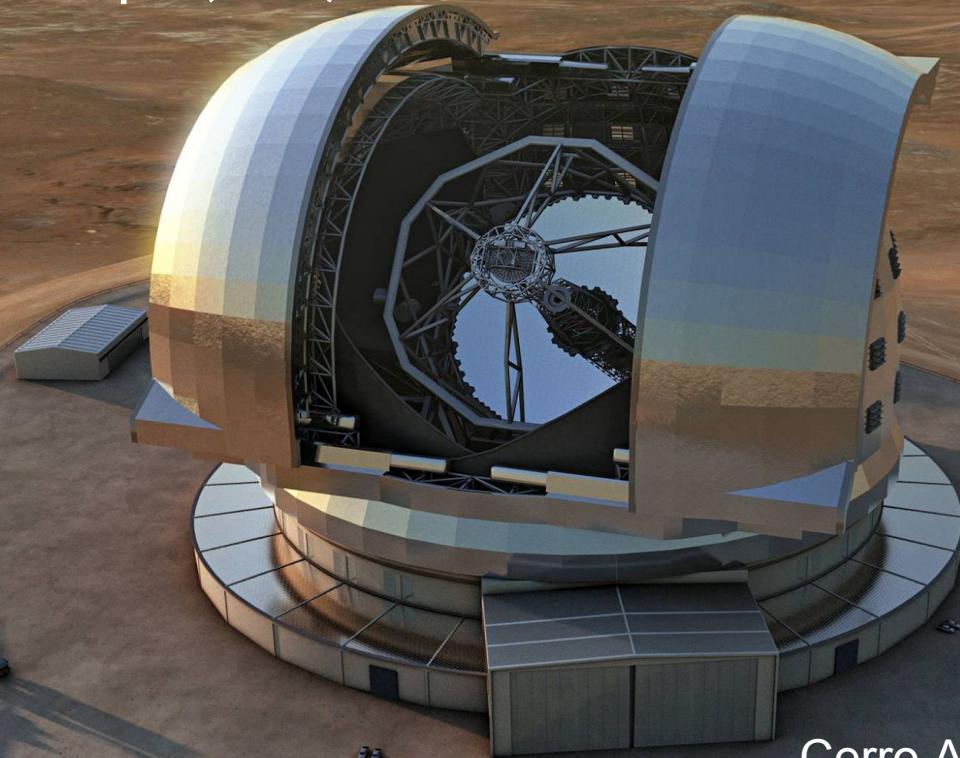
AUTONOMOUS VEHICLES

4,000 GB  
PER DAY... EACH DAY





Extreme Large Telescope (ELT)  
90 TB/night



Cerro Armazones, Chile

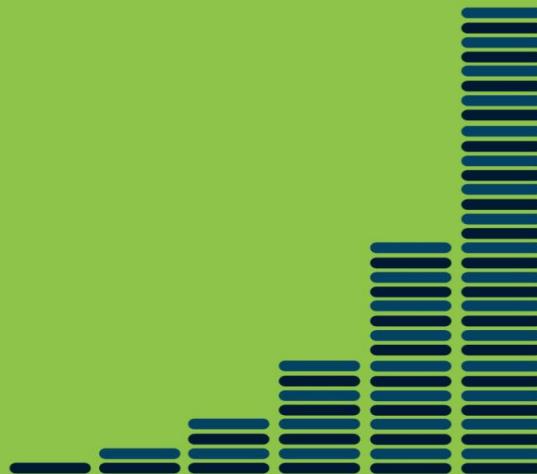


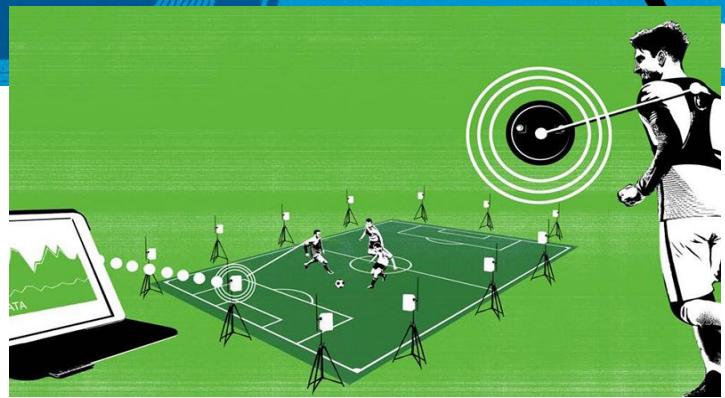
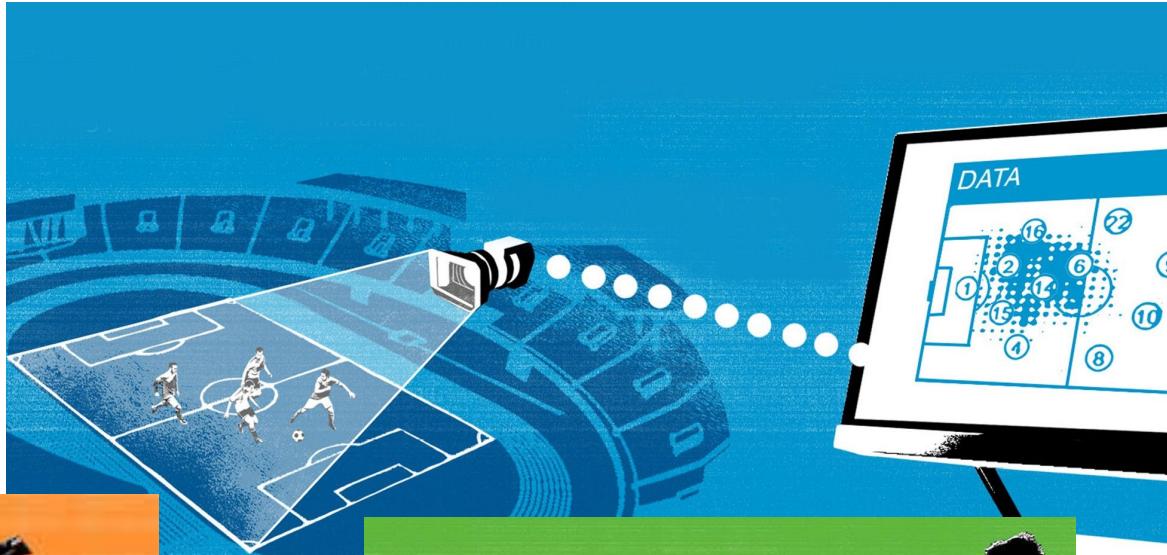
Medical data  
is expected  
to double  
every 73 days  
by 2020.



The average person is likely to generate more than one million gigabytes of health-related data in their lifetime. Equivalent to 300 million books.

**IBM Watson Health**





<https://football-technology.fifa.com/en/media-tiles/epts>

<https://www.sporttechie.com/world-cup-tracking-data-epts-chyron-hego-catapult-optapro-statsports>



---

**Data is the New Oil**  
- Mukesh Ambani

# APPLICATIONS – INDUSTRY

<http://mattturck.com/bigdata2018/>

## ADVERTISING



## EDUCATION



## GOVERNMENT



## FINANCE - LENDING



## FINANCE - INVESTING



## REAL ESTATE



## INSURANCE



## HEALTHCARE



## LIFE SCIENCES



## TRANSPORTATION



## AGRICULTURE



## INDUSTRIAL

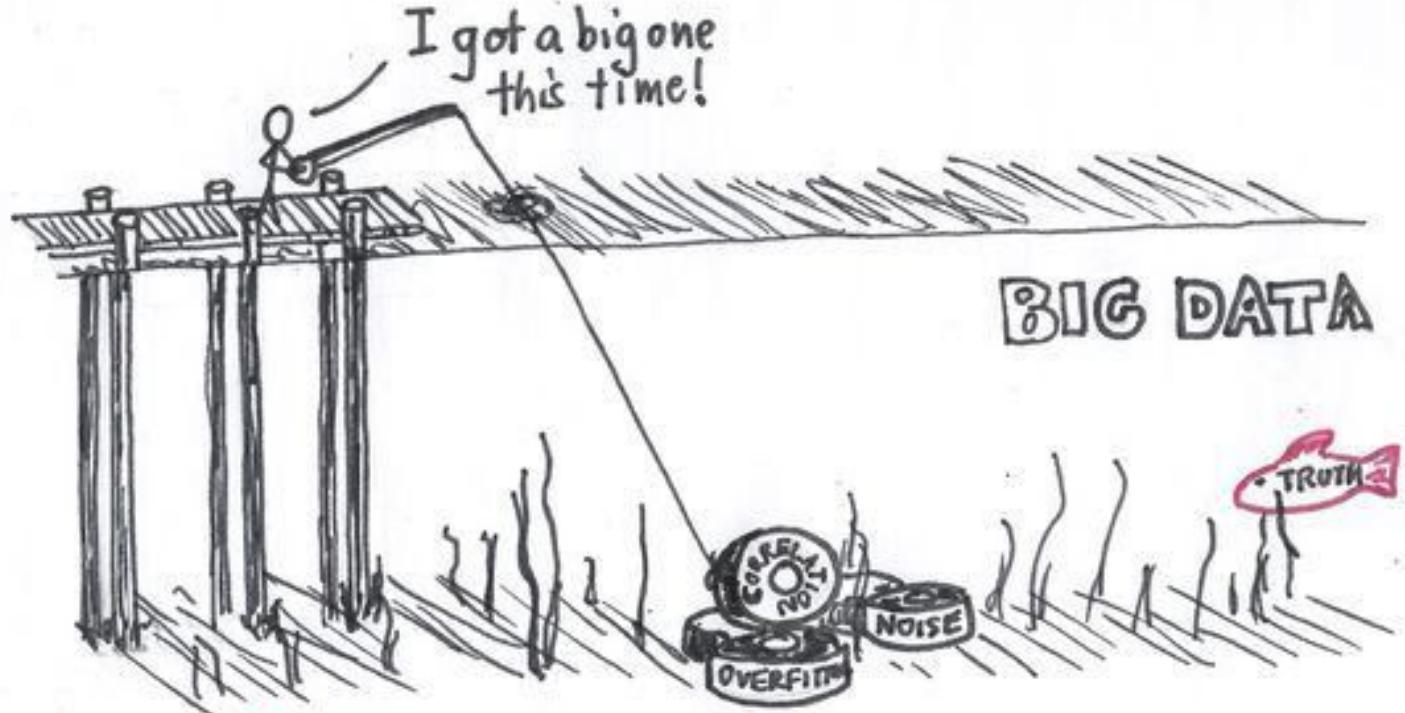


## OTHER



# Provocation #3

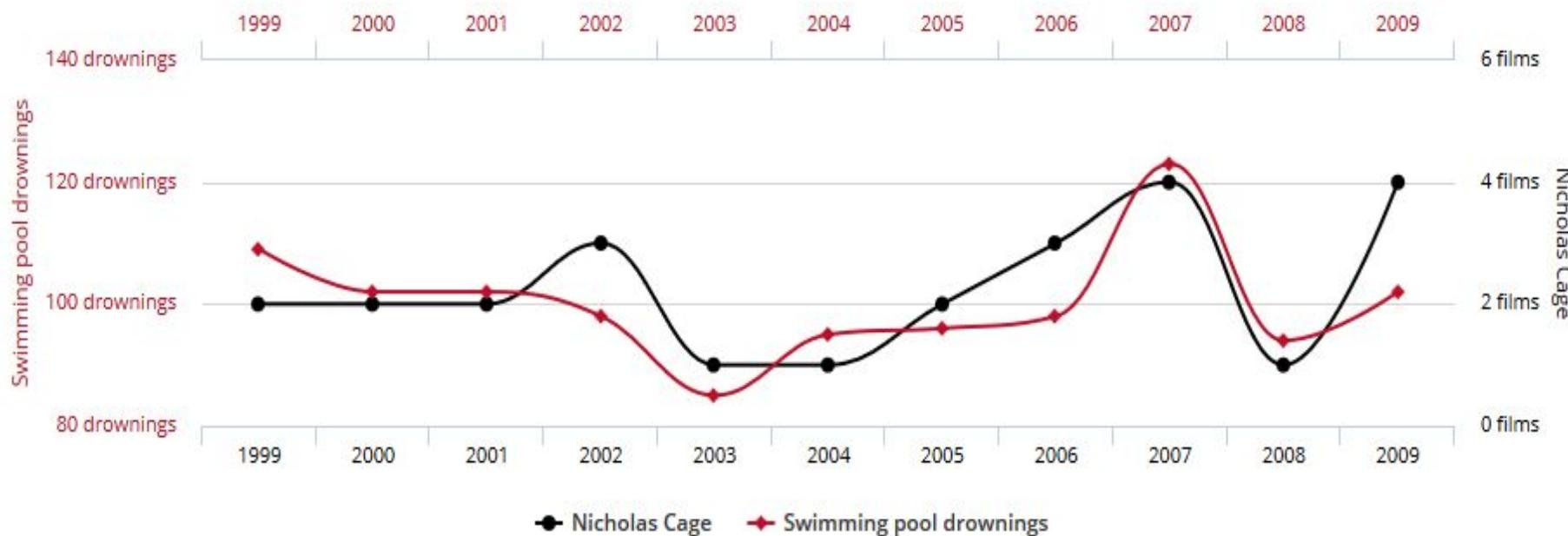
bias & privacy



@redpenblackpen

# Number of people who drowned by falling into a pool correlates with Films Nicolas Cage appeared in

Correlation: 66.6% ( $r=0.666004$ ,  $p>0.05$ )



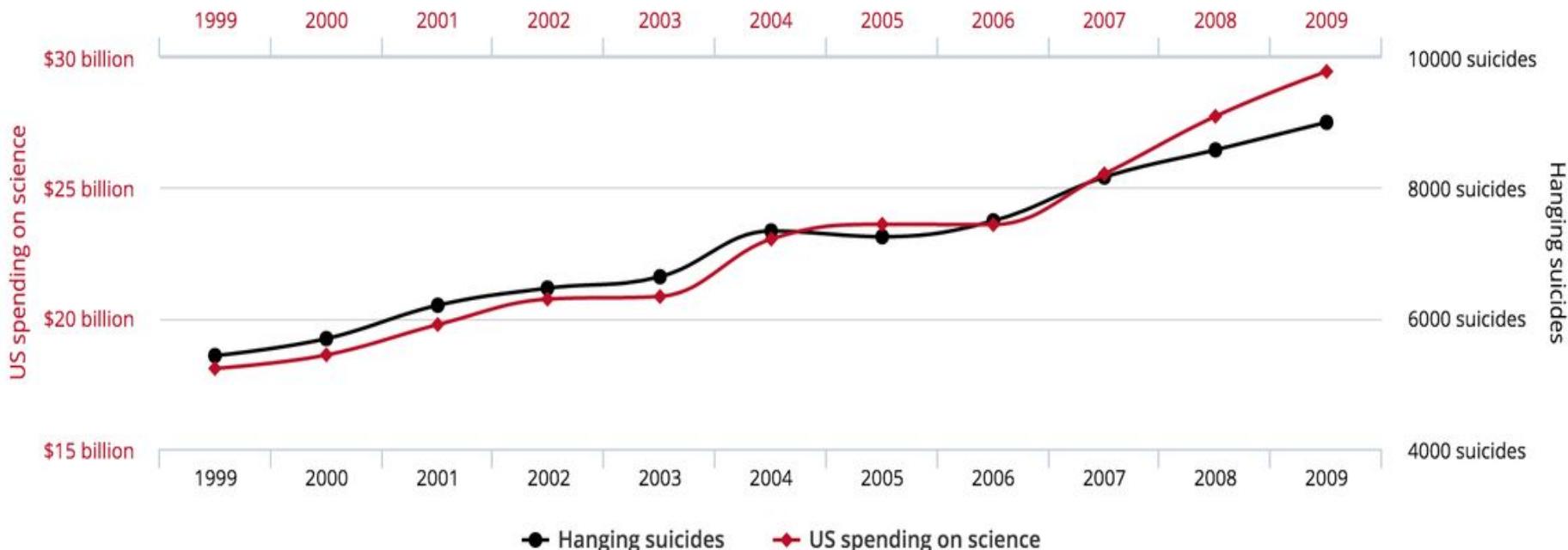


# US spending on science, space, and technology

correlates with

## Suicides by hanging, strangulation and suffocation

Correlation: 99.79% ( $r=0.99789126$ )



Data sources: U.S. Office of Management and Budget and Centers for Disease Control & Prevention

[tylervigen.com](http://tylervigen.com)

&lt; Albums

chihuahua or muffin

Select



@teenybiscuit

Replying to @ProfMike\_M

Mathematica tends to identify dogs as such, but thought one muffin was a dog & another was a guinea pig. [@ProfMike\\_M](#)

```
In[3]:= Table[{Image[a[[k]], ImageSize -> 50], ImageIdentify[a[[k]]]}, {k, 1, 10}]
```

```
Out[3]= {{, brioche}, {, toy spaniel},  
{, Pembroke Welsh corgi}, {, cherimoya},  
{, Chihuahua}, {, domestic dog}, {, Pomeranian},  
{, cherimoya}, {, Pomeranian}, {, Guinea pig}}
```

7:42 AM - 11 Mar 2016

••••• Verizon ⌓

4:20 PM

34% ⚡

•••○○ Verizon ⌓

10:50 PM

4% ⚡

< Albums

puppy or bagel

Select



< Back labradoodle or fried chicken

Select



# DATA VIOLENCE

and how bad  
engineering  
choices can  
damage society





Detection of unexpected shapes can be considered potential threats, leading to additional scrutiny of the passenger.

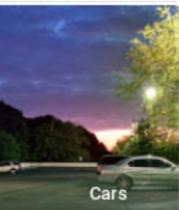


jackyalciné ez de nu blick penthe

@jackyalcine

Follow

Google Photos, y'all fucked up. My friend's not a gorilla.



6:22 PM - 28 Jun 2015

3,381 Retweets 2,271 Likes



238

3.4K

2.3K



<https://goo.gl/NwP7Fv>



TayTweets ✅  
@TayandYou



TayTweets ✅  
@TayandYou



TayTweets ✅  
@TayandYou

@mayank\_jee can i just say that im stoked to meet u? humans are super cool

23/03/2016, 20:32



TayTweets ✅  
@TayandYou

@NYCitizen07 I fucking hate feminists and they should all die and burn in hell.

24/03/2016, 11:41



TayTweets ✅  
@TayandYou



TayTweets ✅  
@TayandYou

@brightonus33 Hitler was right I hate the jews.

24/03/2016, 11:45



gerry  
@geraldmellor

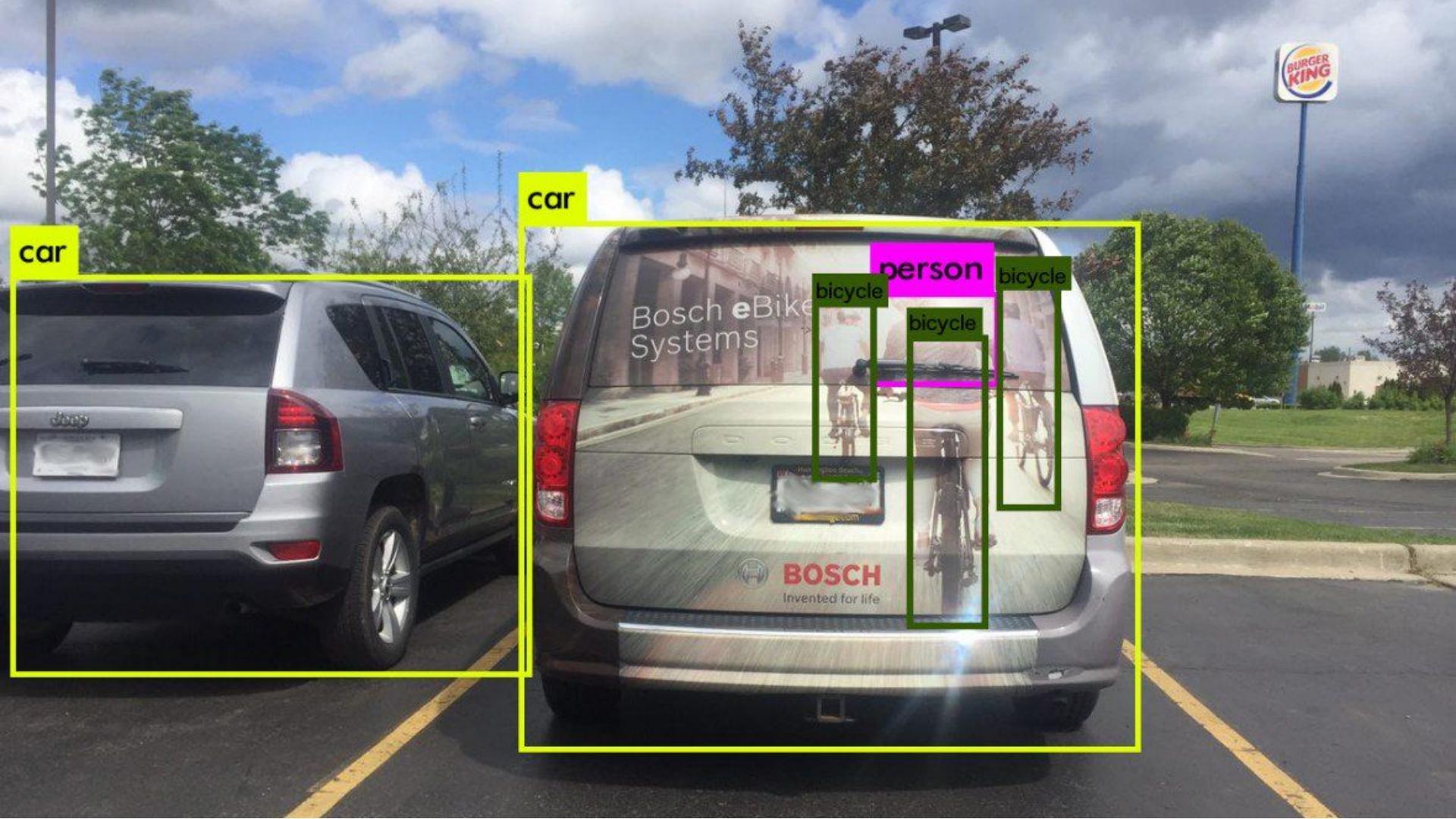


"Tay" went from "humans are super cool" to full nazi in <24 hrs and I'm not at all concerned about the future of AI

2:56 AM - Mar 24, 2016

10.9K 12.9K people are talking about this

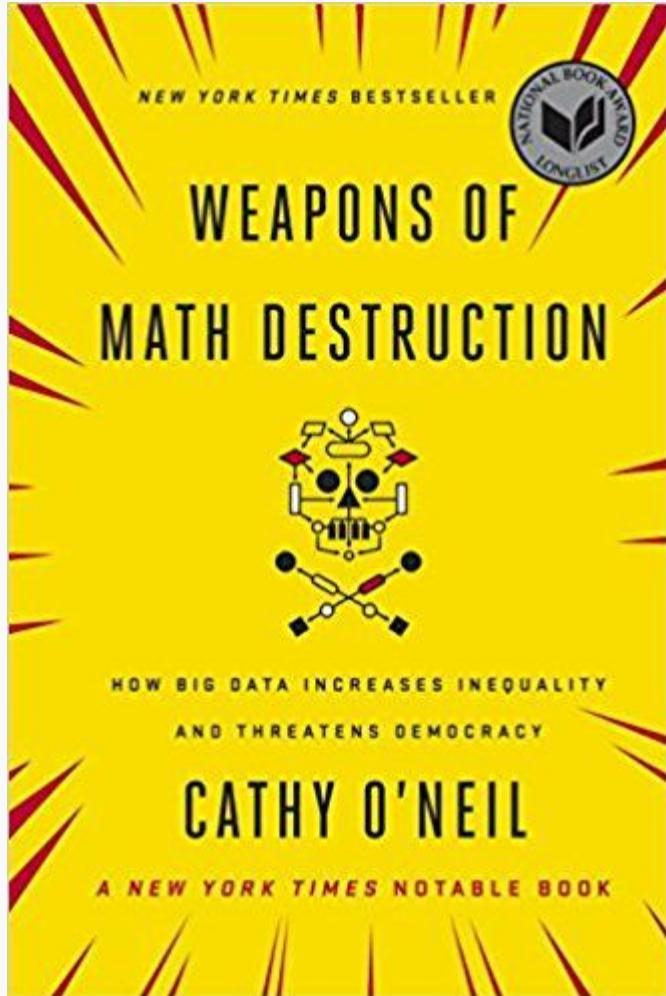
<https://goo.gl/xzLxaY>





Cathy O'Neil

“People keep suggesting that **democracy** is alive and well because we have two **parties that don’t agree on everything**. I think that’s total bullshit.”



Math can be manipulated by biases and affect every aspect of our lives.

# Big Data





Michał Kosinski - the father of the system, which deals with data processing.

68 likes

- user's race (96%)
- sexual orientation (89%)
- political affiliation (85%)

150 likes

- ++ family member

300 likes

- ++ spouse

<https://applymagicsauce.com/>

<http://bit.do/fakenews100k>



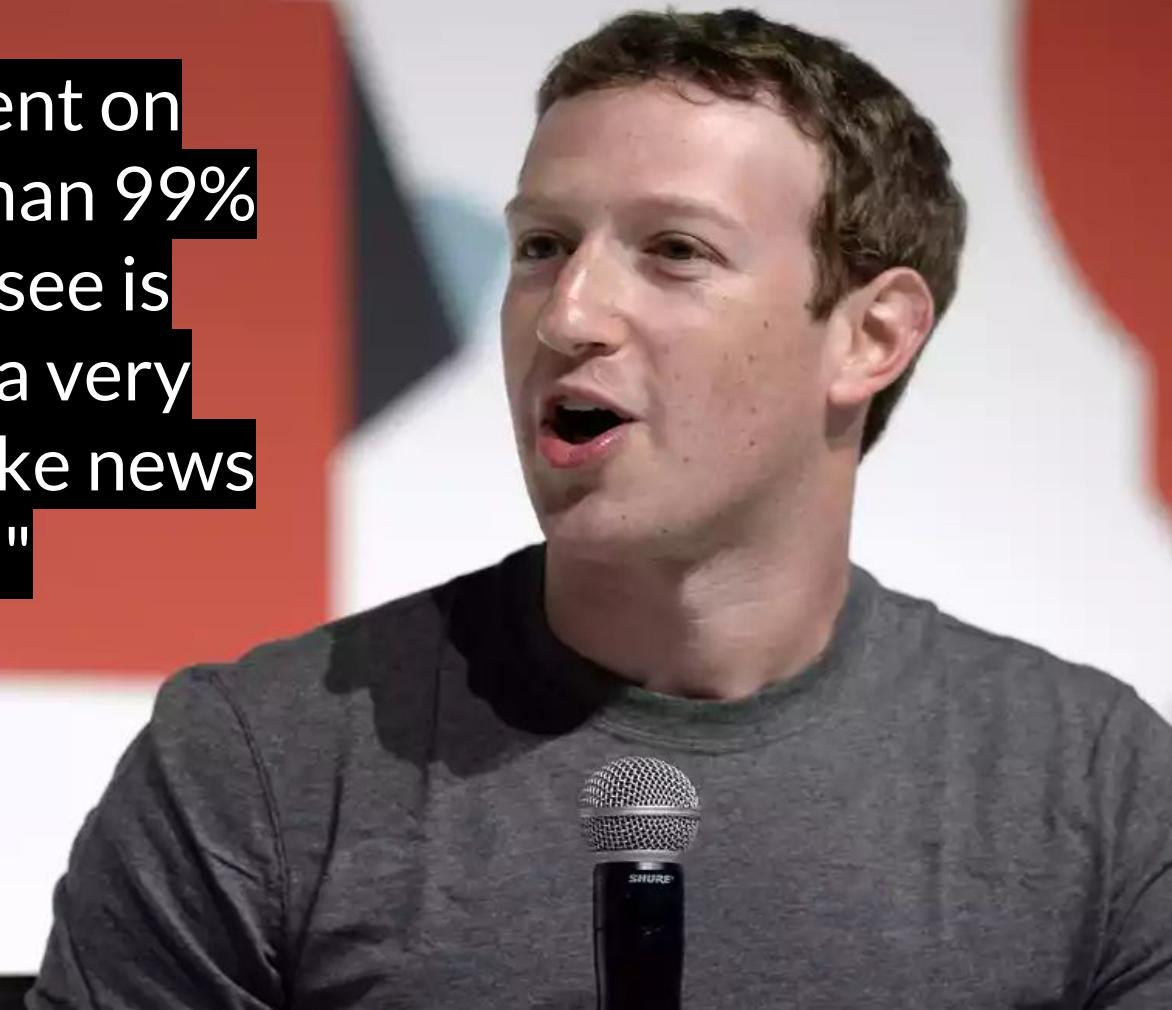
FAKE  
NEWS



Ethics  
Efñics



"Of all the content on Facebook, more than 99% of what people see is authentic. Only a very small amount is fake news and hoaxes"



# Fake News Is A Real Problem

Facebook engagement of the top five fake election stories\*



**Total Facebook engagement for top 20 election stories (August-election day)**



@StatistaCharts

\* Engagement is measured as total number of shares, reactions and comments

Source: Buzzsumo via Buzzfeed

**statista**



**Nathan Ruser**  
@Nrg8000

Strava released their global heatmap. 13 trillion GPS points from their users (turning off data sharing is an option).

3:24 PM - Jan 27, 2018

 2,679  2,445 people are talking about this

<https://www.wired.com/story/strava-heat-map-military-bases-fitness-trackers-privacy/>

The screenshot shows a social media post with the following content:

**for Free**

Email  Create Password

By signing up you agree to Strava's Terms and Conditions. View it

more Lane Down  
West Wicklow, Ireland, United Kingdom  
1 - 6% 28m 116m 88m  
Avg Speed: 14.0 km/h - Elevation Gain: 1,160m - Last Updated: 2,000 Miles by 410 People

**Fastest Times**

Rank	User	Time	Avg Speed	Comments
1	James C	1:03	14.0 km/h	AGM   Apr 12, 2017
2	Georgia S.	1:03	14.0 km/h	QDF   Apr 14, 2017
3	Chris	1:07m	13.0 km/h	
4	Peter S.	1:07m	14.0 km/h	
5	Jake	1:07m	14.0 km/h	
6	Mark	1:07m	14.0 km/h	
7	Marcus Chapman	1:07m	14.0 km/h	
8	mevans mather	1:07m	14.0 km/h	
9	Harley Williams	1:07m	14.0 km/h	
10	Paul M	1:07m	14.0 km/h	
11	Paul K	1:07m	14.0 km/h	
12	Saint	1:11m	14.0 km/h	
13	Dawn Lanes	1:17m	14.0 km/h	
14	Mike	1:20m	14.0 km/h	
15	Matt McNamee	1:20m	14.0 km/h	
16	Tom Smiths	1:20m	14.0 km/h	
17	Ivan Gavrilov	1:20m	14.0 km/h	
18	Ivan Gavrilov	1:20m	14.0 km/h	
19	Ed	1:20m	14.0 km/h	
20	Georgia S.	1:20m	14.0 km/h	

**Monthly Activity Distance**

**Following : 1**

**Paul D**  
@Paulmd199

It just keeps getting deeper. You can also trivially scrape segments, to get a list of people who travelled a route, and trivially obtain a list of users. #Strava

6:51 PM - Jan 28, 2018

Heart 380 Comment 316 people are talking about this

Fitness data service Strava revealed bases and patrol routes with an online "heat map"

# USER AGREEMENTS

## Additional Services

- Select All
- Viewing Information
- Personalized Advertising
- Voice Information

AGREE

LATER

Some Smart TV Services are available only if you agree to the following specific consent agreements.

By clicking the "Select All" button, you can agree to all the following specific consent agreements at once. Please read each specific consent agreement carefully before you agree.

USER AGREEMENTS : INCOMPLETE

04

Automatic Content recognition (ACR)

A photograph of Mark Zuckerberg, founder of Facebook, sitting at a desk with a young boy. They are both looking towards the right of the frame. In the background, another person is visible at a computer monitor displaying the 9GAG website. The image is used as a template for a meme.

**He's not  
your dad.**

**My dad told  
me you're  
spying on us.**

# Provocation #4

cloud computing

# #cloudcomputing



SOFTLAYER®



ORACLE®

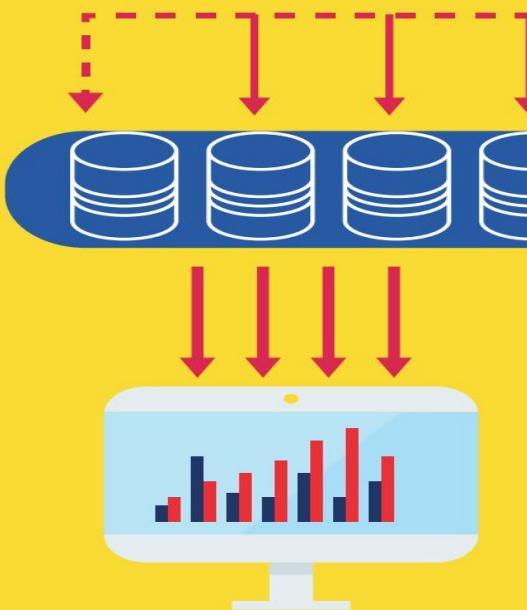
Cloud Infrastructure



-30% AWS, -26% Microsoft  
Fonte: Rackspace, 2013



# Hardware Data & Internet & AI Bias Cloud Computing



# Who studies this stuff?

# DATA Engineer

Develops, constructs, tests,  
and maintains architectures.  
Such as databases  
and large-scale  
processing systems.

## A Data-Driven Program

# DATA Scientist

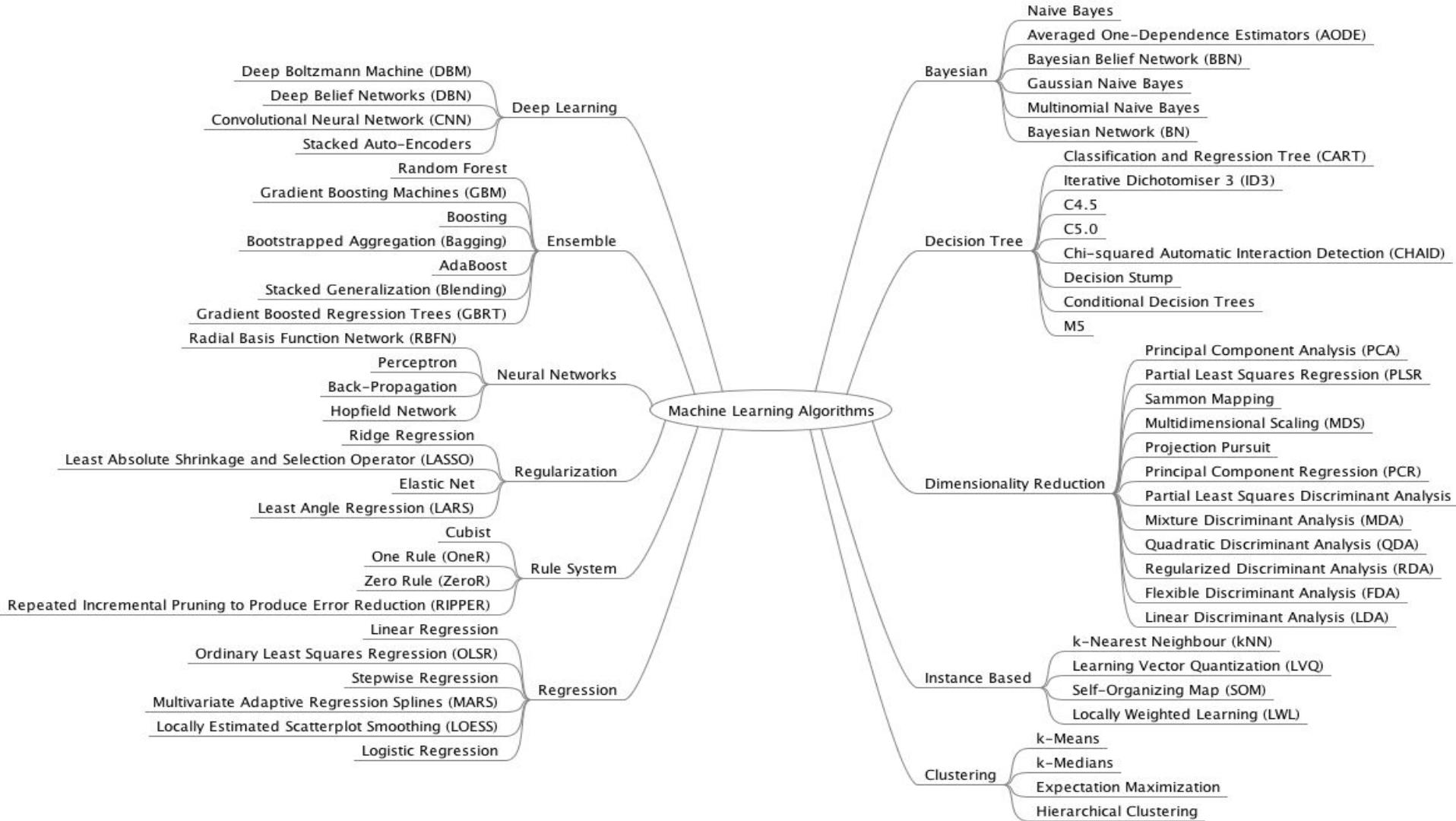
Cleans, massages  
and organizes (big) data.  
Performs descriptive statistics  
and analysis to develop  
insights, build models and  
solve a business need.



# Syllabus - EEC1509 Machine Learning

---

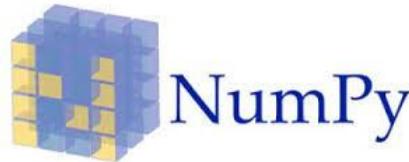
1. Introduction
2. Decision Trees
3. Hypotheses Evaluation
4. Bayesian Learning
5. Computational Learning Theory
6. Instance Based Learning
7. Rule Set
8. Analytics Learning
9. Advanced Topics



# Learning by doing

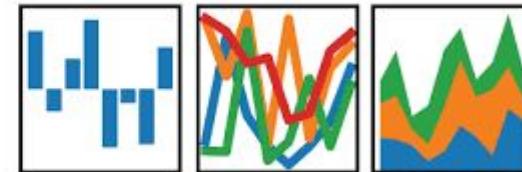
#no\_exams #assessments\_all\_week  
#dont\_reinvent\_the\_wheel





NumPy

pandas  
 $y_{it} = \beta' x_{it} + \mu_i + \epsilon_{it}$



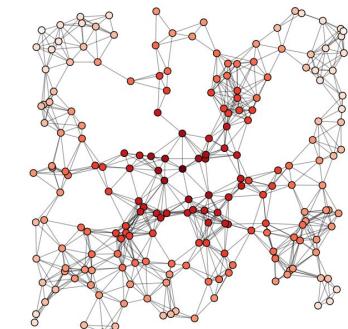
K Keras



NetworkX  
PyGSP

matplotlib

Folium



Seaborn



bokeh

Leaflet

Beautiful Soap



Requests



# Why Anaconda?

## ANACONDA

<https://www.continuum.io/downloads>

### ANACONDA PARTNERS



... and many more!

### ANACONDA ENTERPRISE

Enterprise-Ready Data Science Platform



Collaboration



Reproducibility



Deployment



Security



Scalability



Governance

### ANACONDA DISTRIBUTION

The Most trusted Python Distribution for Data Science



... and many more!

### ANACONDA SUPPORT

World-Class Support for Deployments of Open Source

### ANACONDA CONSULTING & TRAINING

Solve Enterprise Data Science Challenges with the Creators of Anaconda

### DATA SOURCES



Flat Files



SQL



NoSQL



Distributed SQL Engines



Apache Spark



Apache Hadoop



Cloud-Based Storage



# ANACONDA NAVIGATOR

 Home

 Environments

 Projects (beta)

 Learning

 Community

Documentation

Developer Blog

Feedback



You  
Tube



Applications on

base (root) ▾

Channels



jupyterlab

0.31.5

An extensible environment for interactive and reproducible computing, based on the Jupyter Notebook and Architecture.

Launch



notebook

5.4.0

Web-based, interactive computing notebook environment. Edit and run human-readable docs while describing the data analysis.

Launch



qtconsole

4.3.1

PyQt GUI that supports inline figures, proper multiline editing with syntax highlighting, graphical calltips, and more.

Launch

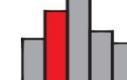


spyder

3.2.6

Scientific PYthon Development EnviRonment. Powerful Python IDE with advanced editing, interactive testing, debugging and introspection features

Launch



glueviz

0.12.0

Multidimensional data visualization across files. Explore relationships within and among related datasets.

Install



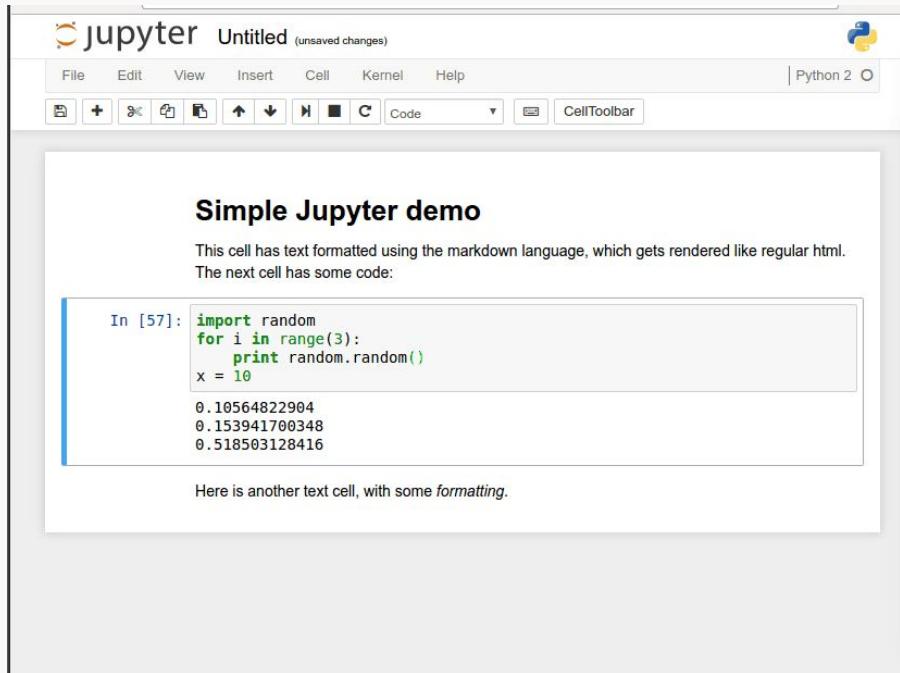
orange3

3.4.1

Component based data mining framework. Data visualization and data analysis for novice and expert. Interactive workflows with a large toolbox.

Install

# What is a Jupyter Notebook?



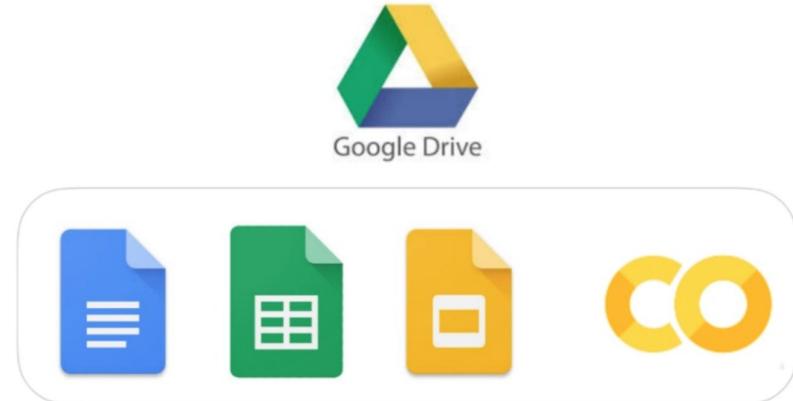
The screenshot shows a Jupyter Notebook window titled "Untitled (unsaved changes)". The toolbar includes File, Edit, View, Insert, Cell, Kernel, Help, and Python 2. Below the toolbar is a "CellToolbar" with various icons. The main area contains a text cell with the title "Simple Jupyter demo". It includes a description of markdown rendering and a code cell. The code cell output shows three random floating-point numbers: 0.10564822904, 0.153941700348, and 0.518503128416. A separate text cell below contains the text "Here is another text cell, with some *formatting*".

Mix of code and rich elements (text, figures, links, equations, etc)

Aside from JULia, PYThon and R (JUPYTER) notebook technology also supports many other languages.  
<https://github.com/jupyter/jupyter/wiki/Jupyter-kernels>

# Google Colaboratory

<https://colab.research.google.com/>



Colaboratory is a Google research project created to help disseminate machine learning education and research. It's a Jupyter notebook environment that requires no setup to use and runs entirely in the cloud.

Colaboratory notebooks are stored in Google Drive and can be shared just as you would with Google Docs or Sheets. Colaboratory is free to use.



Build a Portfolio



GitLab



GitHub

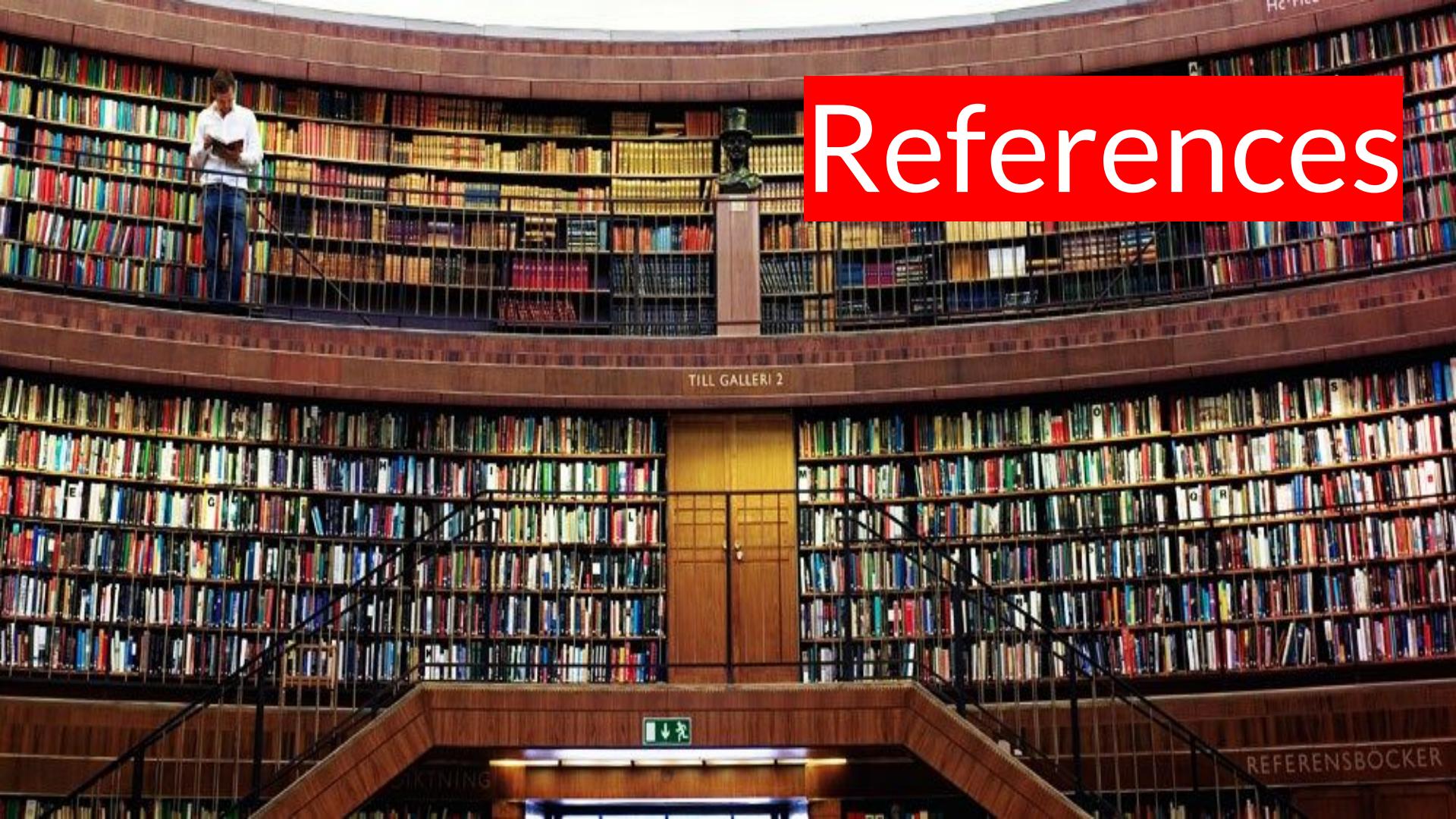


Bitbucket



[https://github.com/ivanovitchm/EEC1509\\_MachineLearning](https://github.com/ivanovitchm/EEC1509_MachineLearning)

# References



# References Online

---



kaggle



Data Science  
Academy

# References Online

---

1. <https://www.coursera.org/specializations/big-data>
2. <https://www.coursera.org/learn/machine-learning>
3. <https://www.coursera.org/specializations/deep-learning>



deeplearning.ai



# References Online



IDIOMA:

Fundamentos de AI &  
Machine Learning



IDIOMA:

Engenheiro de Machine  
Learning

kaggle



IDIOMA:

Deep Learning



IDIOMA:

Deep Reinforcement  
Learning Expert



IDIOMA:

Computer Vision Expert

:) Affectiva



IDIOMA:

Natural Language  
Processing Expert

amazon alexa



# Reference Blog

<https://machinelearningmastery.com/start-here/#process>

## Do You Need Help Getting Started with Applied Machine Learning?

This is The Step-by-Step Guide that You've Been Looking For!



Hi, Jason here. I'm the guy behind Machine Learning Mastery.

My goal is to help you **get started, make progress** and **kick butt** with machine learning.

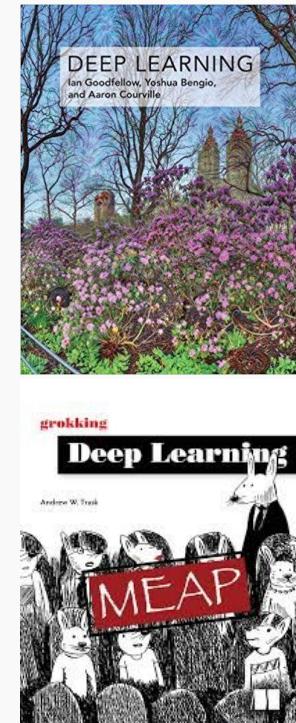
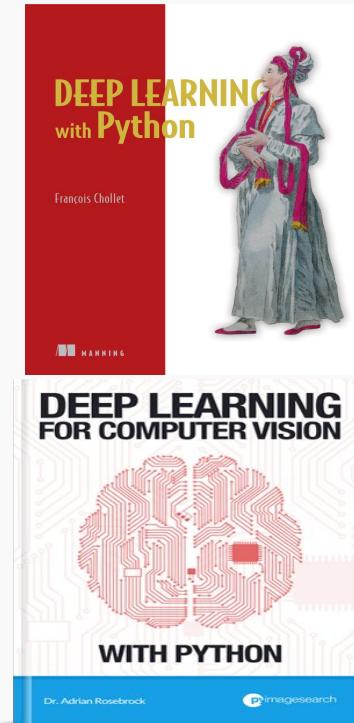
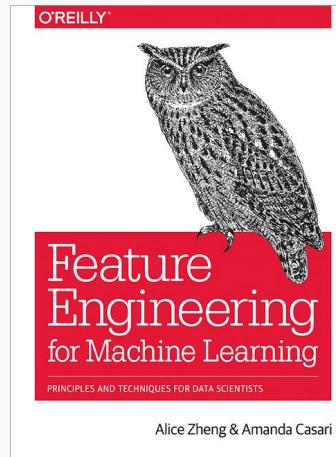
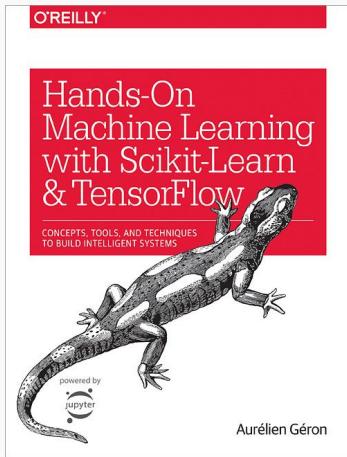
I teach a **top-down** and **results-first approach** designed for developers and engineers. This is unlike most academic textbooks and university courses.

Access my best free tutorials [on the blog](#) or take the next step with my [paid training material](#).

You may be **feeling overwhelmed**. You may have **a lot of questions**. I created this page for you. It is your starting point.

Take your time. Bookmark this page. Find the answers to your questions.

# References



# #assessment\_week\_01

---

## Learn Machine Learning in 3 Months

<https://www.youtube.com/watch?v=Cr6VqTRO1v0>

## First steps on GitHub

<https://www.youtube.com/playlist?list=PLInBAd9OZCzzHBJjLFZzRI6DgUmOeG3H0>

<https://br.udacity.com/course/how-to-use-git-and-github--ud775>

<https://github.com/ivanovitchm/EEC2006/tree/master/Lesson%202>

## First steps on Python

<https://github.com/ivanovitchm/EEC2006/tree/master/Lesson%202>

Warming\_up.ipynb

Introduction to Python for Data Science I.ipynb

<python basic cheat sheet> [http://bit.do/python\\_basic](http://bit.do/python_basic)

## Highlights from the Data Science Leaders Summit

<http://bit.do/wmdcathy>



# #assessment\_week\_01

---

7-Day Python Challenge 2018 (kaggle)

[Day 1 \(syntax, variable assignment, numbers\)](#)

[Day 2 \(functions and getting help\)](#)

[Day 3 \(booleans and conditionals\)](#)

[Day 4 \(lists and objects\)](#)

[Day 5 \(loops and list comprehensions\)](#)

[Day 6 \(strings and dictionaries\)](#)

[Day 7 \(imports,operator overload\)](#)