ML for Finance

MAFN Spring 2023

Columbia University

1st Week

- Class Overview
- Class Motivation
- Anaconda
- Jupyter
- Data Exploration
- Taste of ML

Logistics

- Class Meets
  - Saturdays 2:40-4:30pm
  - 312 Mathematics Building (via Zoom for emergencies only)
  - Mathematics GR5430 section 001
- Instructor
  - Renzo Silva
  - email: rs333@columbia.edu
  - Office hours: Fridays 4-6pm by appointment and via Zoom
- TAs
  - Xiaoyang Chi xc2599@columbia.edu
  - Weining “Winnie” Qu wq2155@columbia.edu
  - Office hours: TBD

Assignments & Exams

- 5 homeworks, 40% of grade
  - Grade best 4 out of 5
- 2 “projects”, 20% of grade
  - 1 individual project
  - 1 team project, random team memberships
- 2 exams, 40% of grade
  - Closed book, online, via CourseWorks/Proctorio
  - Some questions will require coding in python

Schedule

- Modules 1-2, weeks 1-3
  - Data Exploration
  - Data Visualization
  - Data Collection, prep, pipeline
- Modules 3-4, weeks 4-11
  - ML Algos
- Module 5, weeks 12 - 13
  - Computing, including cloud

Textbooks

All optional

1. Machine Learning in Finance, Dixon, Halperin, Bilokon (MLF)
2. Python for Finance: Mastering Data-Driven Finance 2nd Edition (PF)
3. The Elements of Statistical Learning: Data Mining, Inference, and Prediction, Second Edition by Hastie, Tibshirani, Friedman (ESL)
4. Python for Data Analysis by Wes McKinney (PDA)
5. Machine Learning by Tom M. Mitchell 1st Edition (ML)
6. An Elementary Introduction to Mathematical Finance 3rd Edition by Sheldon M. Ross (EIMF)

What are the goals?

- Make your work/future company more efficient
- New ways to do financial analysis or new investment research ideas
- Use the existing tools, don’t recreate them
  - Developing new tools is fine, if it improves the art
- Become the bridge between tech and finance

Class Motivation

- Why ML for Finance? Who cares?
- Better career-wise?
  - "Skate to where the puck is going, not where it has been." - W. Gretzky
  - What if everyone goes there?
- All models are wrong, some are useful*
  - Some are harmful
- What role does technology play in economics?

* statistician George Box (1976)

Solow-Swan Model

- Economic growth rate is made up of 3 components
  - Financial growth rate
  - People growth rate
  - Technological growth rate

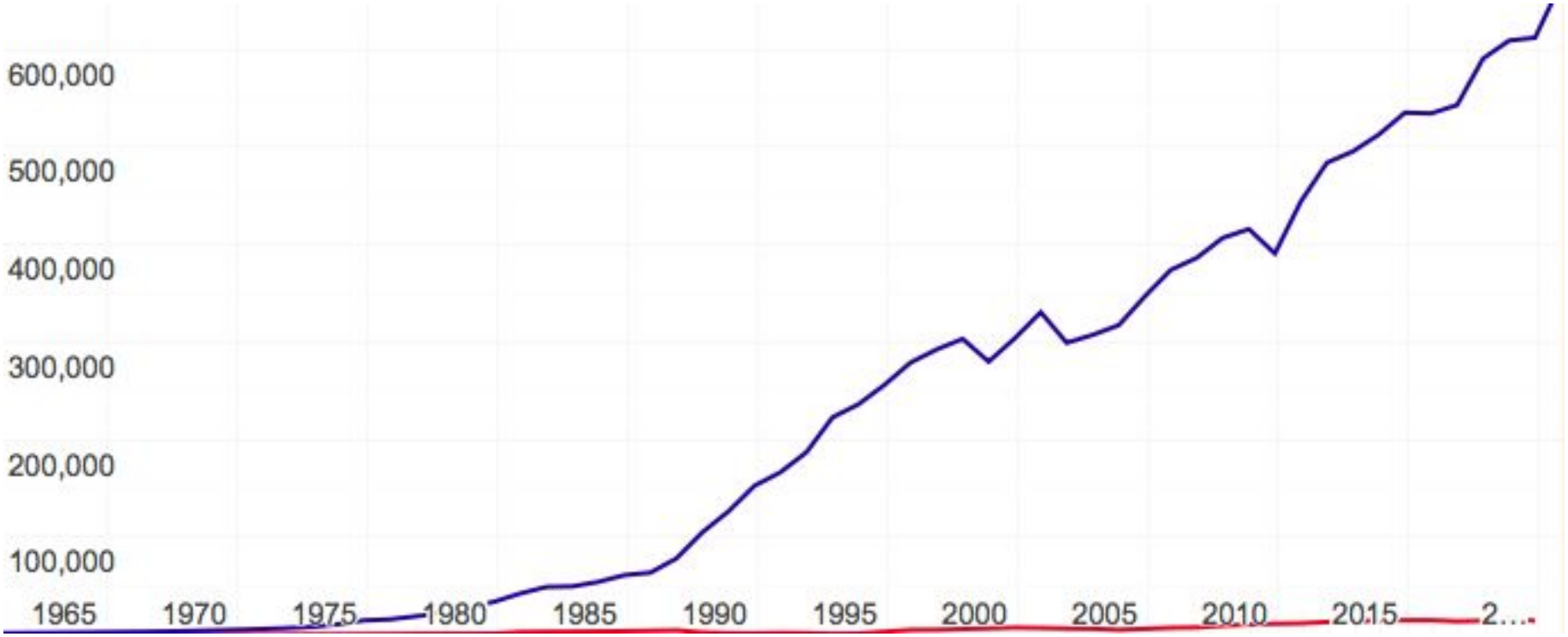
$$Y(t) = K(t)^{\alpha}(A(t)L(t))^{1-\alpha}$$

Y = total production
L = labor (or working population)
K = capital
A = labor-augmenting **technology**, or "knowledge"
$0 < \alpha < 1$ is the elasticity of output with respect to capital

Source: Wikipedia

Comparing 2 Countries

GDP of 2 countries over 50 years



- In 1950s, two countries were comparable. Similar GDP, population, technological state.
- Over the next 50 years, one country invested heavily in technology. Population and education level also grew. (could be interesting project with future projections)

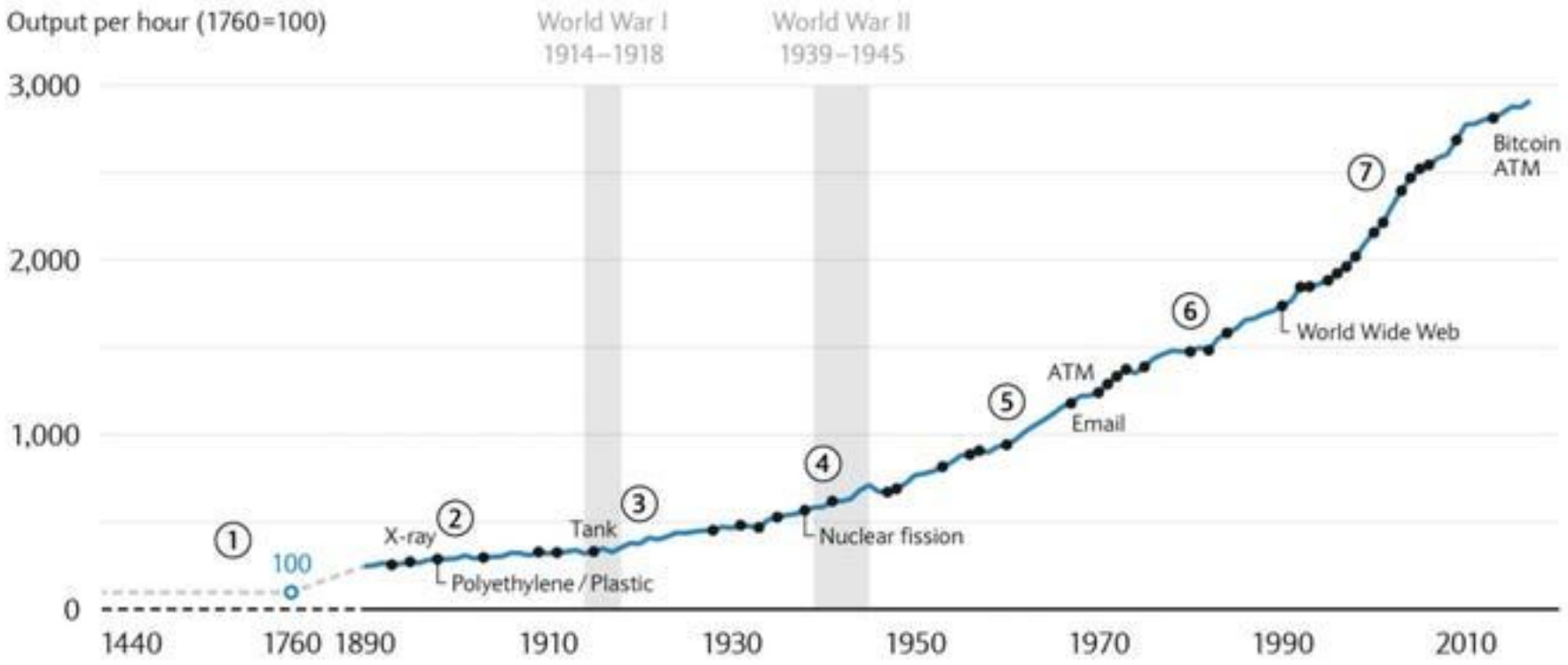
Tech and Productivity

- Since the 1760, how much has productivity increased on earth?
  - 30x, per person
  - In general, what took 30 hours in 1760, takes 1 hour today
- Largely due to technological innovations like electricity, telephones, and the internet.

Tech and Productivity

FIGURE 1

From the printing press to the global internet, technology has evolved, and human societies with it



Source: Barclays

Tech and Productivity

① 1890 and earlier

1440 Gutenberg Printing Press
1480 Sea astrolabe
1589 Mechanical knitting machine
1608 Telescope
1630 Slide rule
1765 Watt's steam engine
1790 Sewing machine
1816 Telegraph
1867 Dynamite
1879 Light bulb

② 1891–1910

1893 Diesel Engine
1895 X-ray
1898 Polyethylene / Plastic
1903 Gas turbine
1909 Television broadcast

③ 1911–1930

1911 Cloud chamber
1915 Tank
1928 Penicillin

④ 1931–1950

1931 Electron microscope
1933 FM radio
1935 Nylon
1938 Nuclear fission
1941 Polyester
1947 Hydraulic fracturing
Transistor
1948 Atomic clock

⑤ 1951–1970

1953 Video tape recorder
1956 Hard disk drive
1957 IBM 610
Sputnik 1
1960 Laser
1967 ATM (Barclays)
1970 Pocket calculator

⑥ 1971–1990

1971 Email
Intel 4004
1972 Magnavox Odyssey
1973 Capacitive touchscreen (CERN)
1975 Altair 8800 (Microcomputer revolution and internet protocol suite)
1980 Flash memory
1982 CD-ROM
1984 Cell phone
1990 World Wide Web
Hubble Space Telescope

⑦ 1991–present

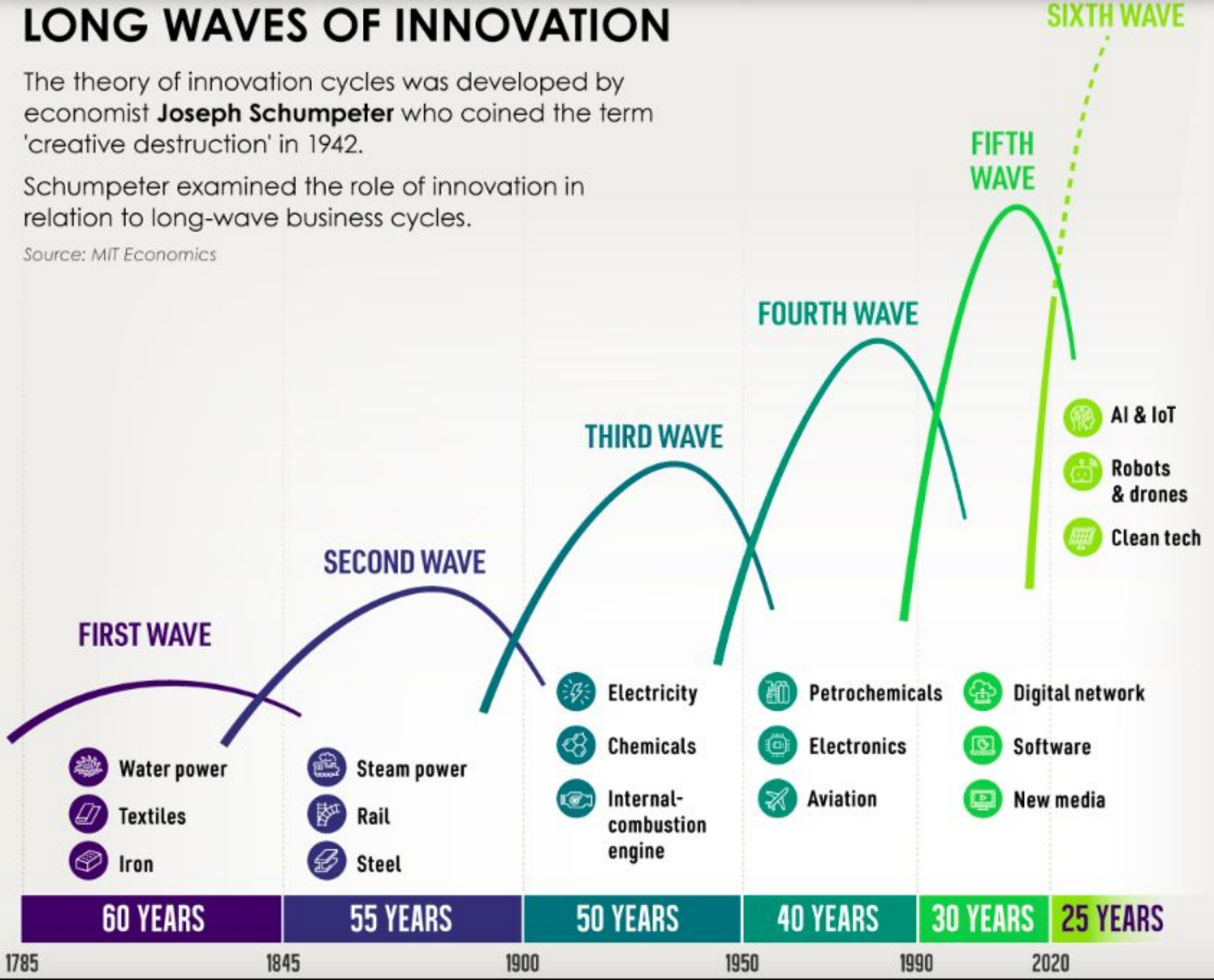
1992 Text messaging
1993 Apple Newton
Mosaic (Web browser)
1995 DVD
Windows 95
1996 USB ports
1997 Netflix
1998 Google
2000 Bluetooth
2001 iPod
2003 iTunes Music Store
2004 Facebook
2005 YouTube
2006 Twitter
2009 Bitcoin
2013 Bitcoin ATM

LONG WAVES OF INNOVATION

The theory of innovation cycles was developed by economist **Joseph Schumpeter** who coined the term 'creative destruction' in 1942.

Schumpeter examined the role of innovation in relation to long-wave business cycles.

Source: MIT Economics



What does it mean for you?

- Every industry is impacted by tech
- Every job is impacted by tech
  - What is the most common tool/tech in finance in the last 30 years?
  - What is the most disruptive tech in finance (and the world) in the next 30 years?
- Will AI replace all of us?
  - No, but some jobs will be
  - Electricity example in the 1900s

New tech, all good?

- Not all tech impact is good
  - **For every new tech, there is new set of problems**
    - E.g. airplanes introduced plane crashes and challenge of how to avoid them
    - AI is no different, it introduces new set of problems
- Keep the potential AI/ML problems in mind
  - Bad data
  - Biases
  - Cost
  - Bad problem
  - Bad solution
  - Trust, ethics
  - Computation limits
  - Safety
  - Job losses

“Simple” Tech Example

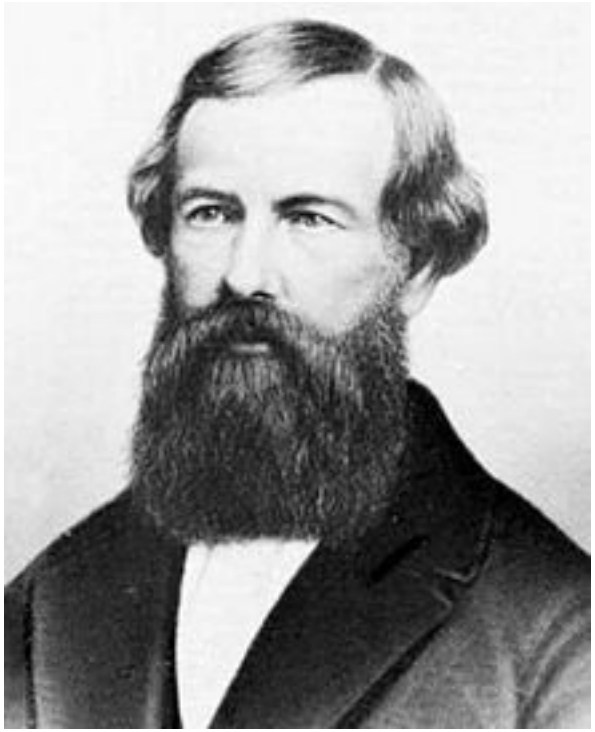
- When were canned goods introduced?
  - ~1810-1811 in London by Girard/Durand/Donkin based on glass jar process by Appert, who developed it for a 12,000 franc prize awarded by the french government for Napoleonic wars
- Benefits
  - Long term food storage without refrigeration
  - Easier to transport
- What new challenges/problems did they introduce?
  - Botulism, especially from home-canning
  - Package reuse/recycling
  - How to open them easily?
    - Ezra Warner from CT patented the can opener in **1858**, ~ 50 years after tin canning began. Hammers and chisels were used previously!
    - Pull tab for cans was only patented in 1963!

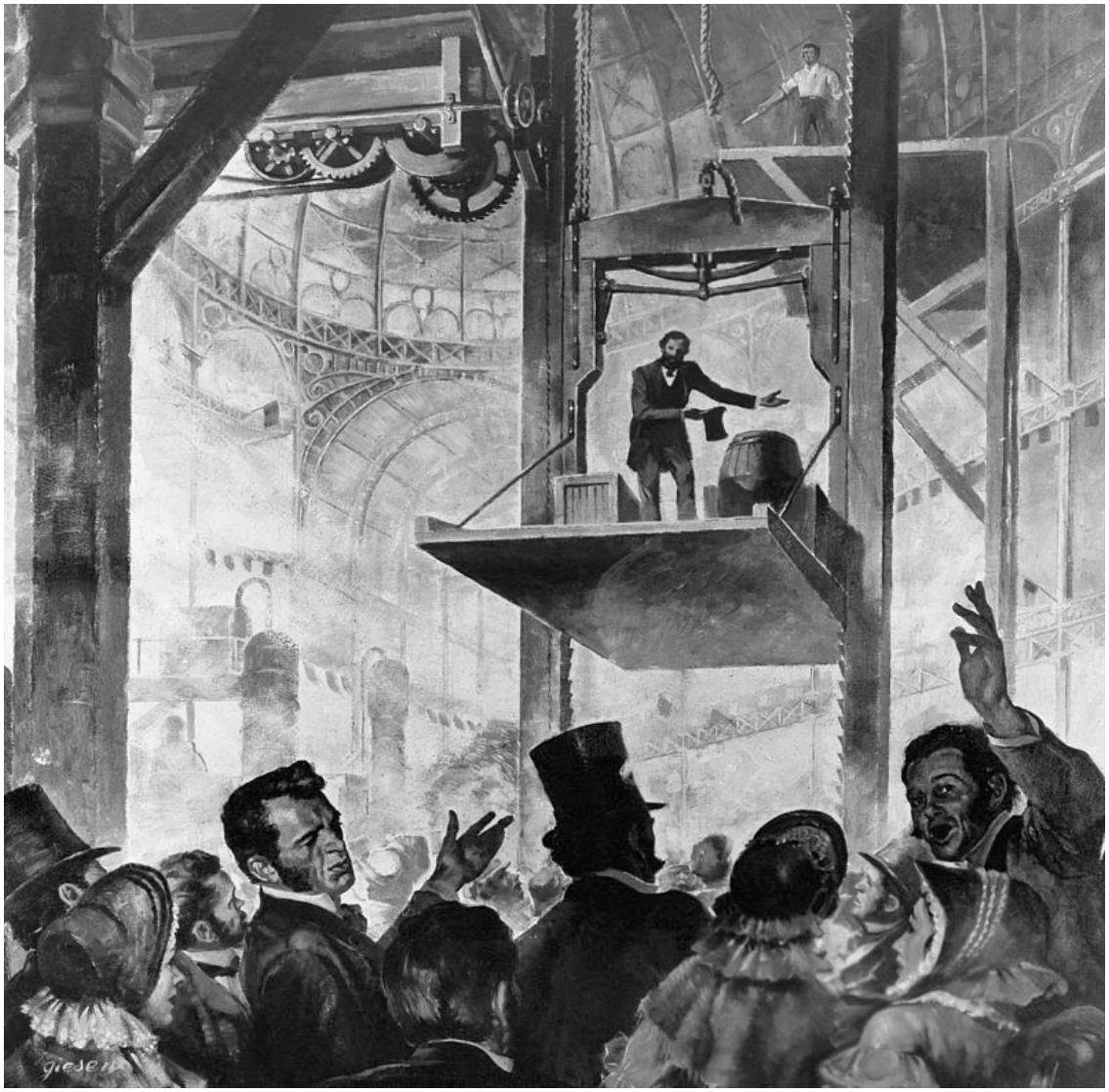
https://en.wikipedia.org/wiki/Can_opener

Example: Otis Elevator Company

- When were elevators invented?
  - Hoists date back to 3rd century BC
  - Human, animal powered, some times water-driven mechanisms
- Modern elevators developed in the 1800s
  - What new problem did they introduce?
    - Cable cuts were common, with passengers plunging to their unfortunate demise
  - Until 1852, Elisha Graves Otis introduced a special mechanism to lock the elevator car in place should the hoisting ropes fail.
  - 1880, first electric elevator. Revolutionized modern cities.
  - Tech can take a long time to stabilize and be useful/safe.

Otis' "Safety Elevator"





1853 New York
World's
Fair

"Quantum Computers Could Solve Countless Problems—And Create a Lot of New Ones"

Feb 2023

Source:
https://time.com/6249784/quantum-computing-revolution/

Tales of caution

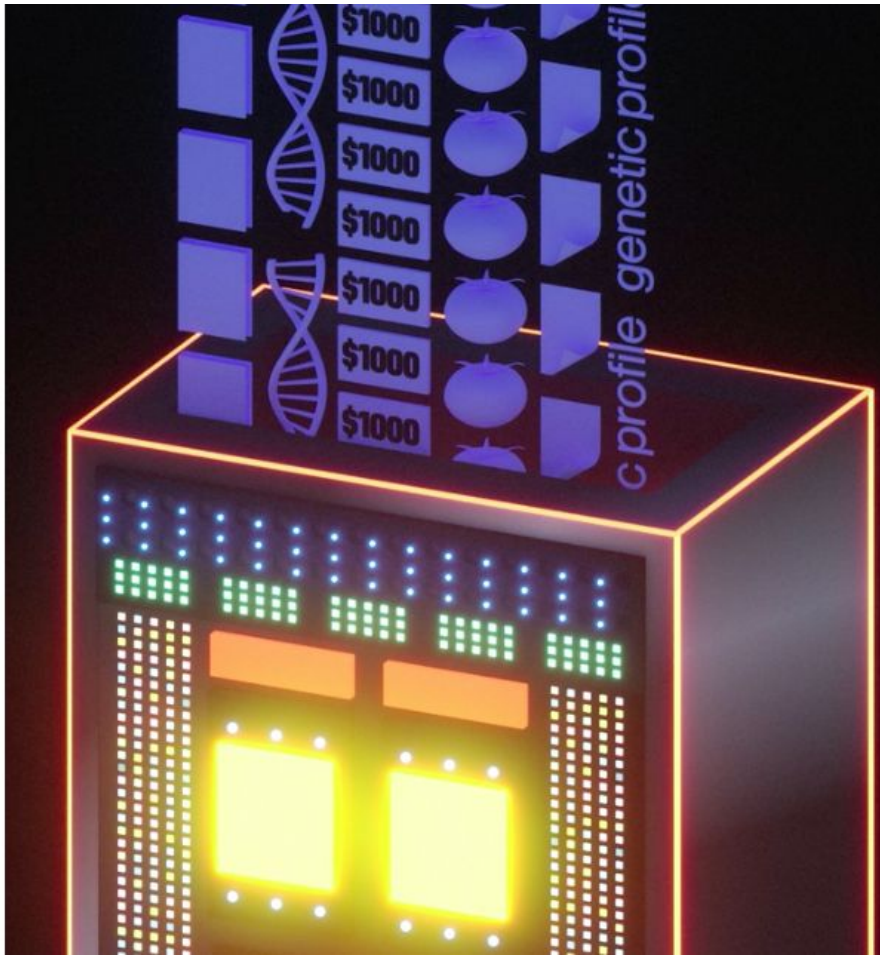
- In the US gold rush of 1849, who made the most money?
  - The gold speculators?
  - The merchants selling goods & services to the speculators
- Cisco CCIE certifications in the 1990s
- Is AI/ML our tech gold rush?

AI and ML: Hype or Reality?

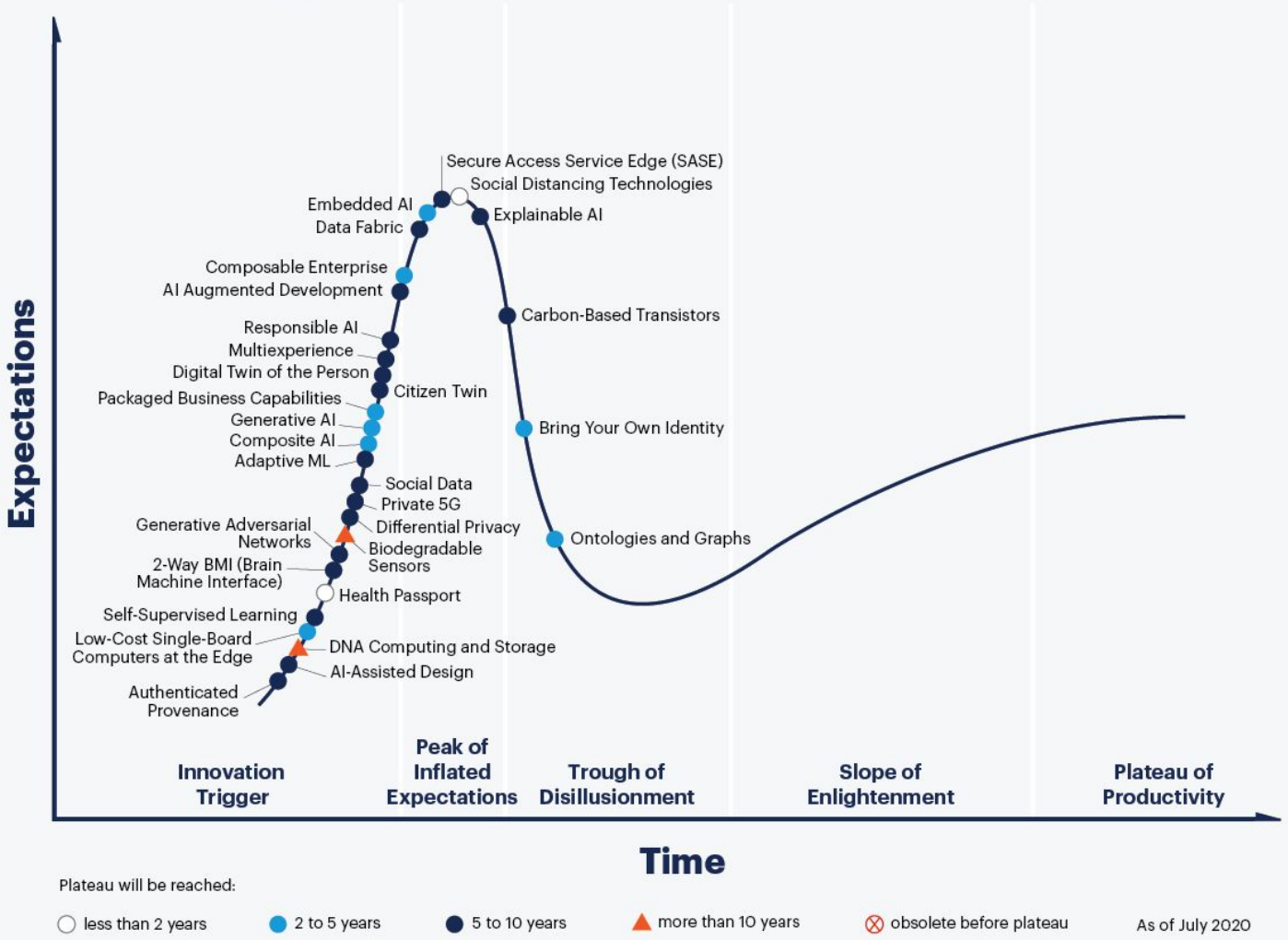
The New York Times

What Ever Happened to IBM's Watson?

IBM's artificial intelligence was supposed to transform ndustries and generate riches for the company. Neither has panned out. Now, IBM has settled on a humbler vision for Watson.



Hype Cycle for Emerging Technologies, 2020



Plateau will be reached:

○ less than 2 years ● 2 to 5 years ● 5 to 10 years ▲ more than 10 years ⊗ obsolete before plateau

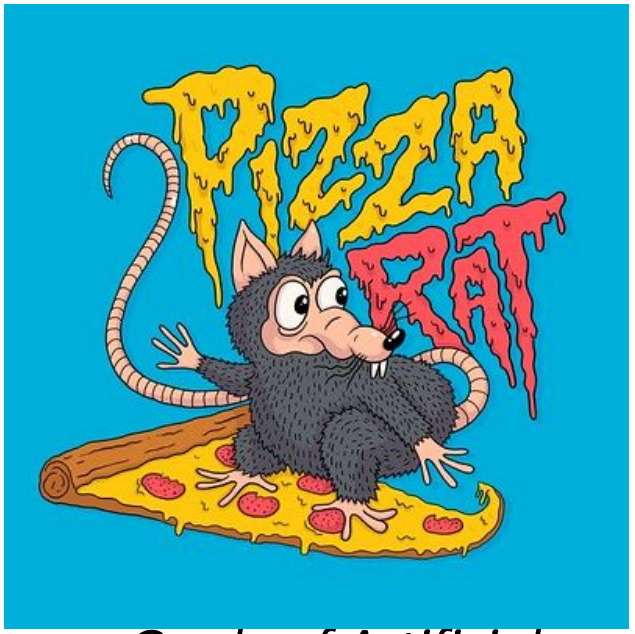
As of July 2020

gartner.com/SmarterWithGartner

Source: Gartner

Gartner

Is (current) AI smarter than …



- Adult human?
- Baby human?
- NYC pizza rat?
- Cockroach?
- Worm?
- Not yet
- Research:
  - *Intelligence Quotient and Intelligence Grade of Artificial Intelligence, 2017*
  - By Feng Liu, Yong Shi, Ying Liu
- You could say we are at the infancy of AI, “baby AI”
  - In many ways,”naïve AI” or “baby AI” is intelligence mimicry

AI vs AGI

- Artificial General Intelligence (AGI)
  - Can analyze a situation on its own an take calculative decision, without having to be programmed in advanced
  - Aka “true” or “real” or human intelligence
  - Doesn’t exist currently, but active area of research

The New York Times

Account

How Smart Are the Robots Getting?

The Turing test used to be the gold standard for proving machine intelligence. This generation of bots is racing past it.

- A best case
- A somewhat scarier case

Dall E 2

TEXT DESCRIPTION

An astronaut Teddy bears A bowl of soup

mixing sparkling chemicals as mad scientists shopping for groceries working on new AI research

as kids' crayon art on the moon in the 1980s underwater with 1990s technology

→

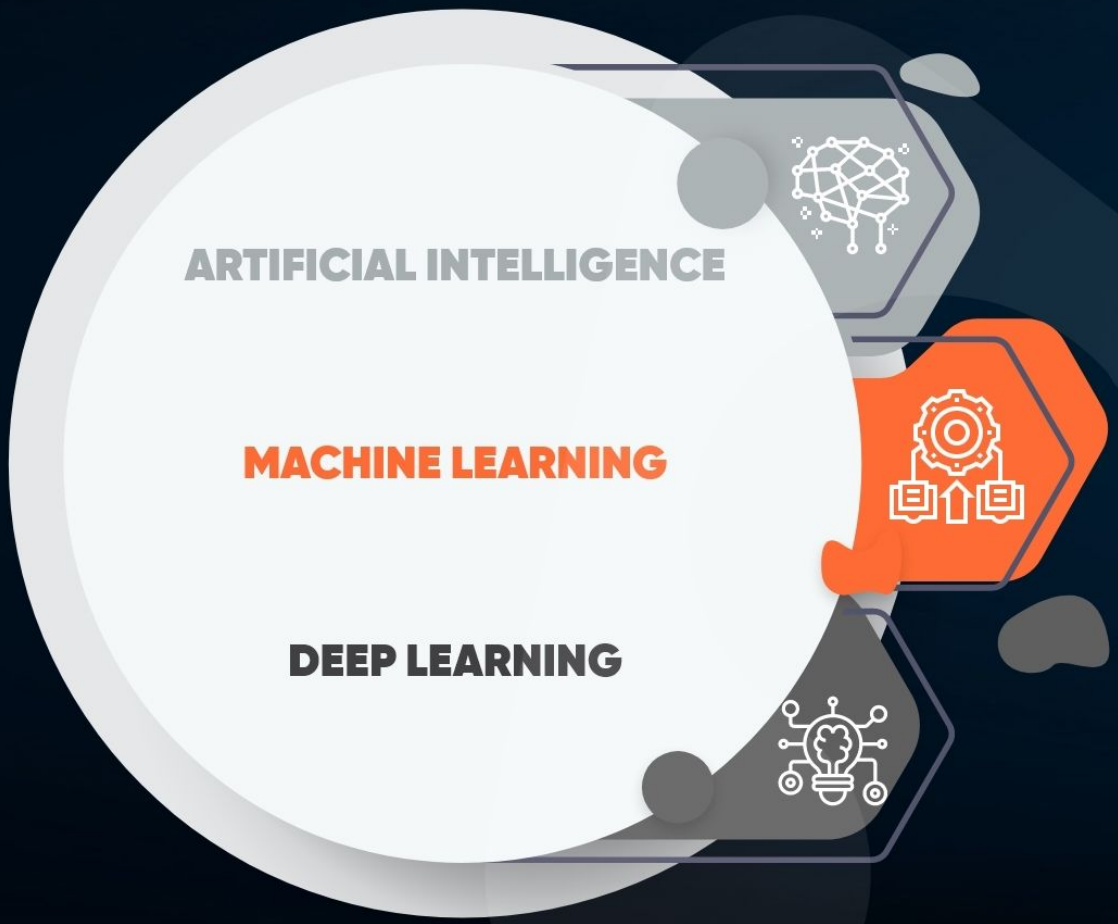
DALL·E 2

https://openai.com/dall-e-2/

How soon before we can ask AI systems to do more complicated things?

Comparing human intelligence and artificial intelligence

Sequence	Human View	Machine View
Event occur in the enviroment	Human have general intelligence and respond to many stimuli, but only some events are observed and are revelant to our senses.	Machines are custom-built to react to specific events, but can be programmed to sense things humans cannot (e.g., ultrasound).
Sensors measure data about the event	Human use our natural senses and the system of the body to capture information and send it to the brain through the nervous system	Machine sensors can be built for specific data and placed within objects to collect "Big Data". Digital Twin projects from industrial companies render physical things in virtual worlds as a digital nervous system.
Information is sent for Prossessing	Different system in the organism are used to measure and react to stimuli-- from conscious and unconcious thought, to emotions and hormonal systems.	Machines must structure the data they collect and find ways to store and process massive amounts of information, which only now has become possible with cheap computing power.
Processing check new data against a mental model		
Reaction is trigged in response to event	The sensory experience has impacted a system, which has made a determination for action (e.g., a person will withdraw a hand from the flame after proccesing a pain signal through the spinal cord brain).	Upon receiving the data, a model will be updated, perhaps even changing t he model and an action will be caused (e.g., a self-driving car may change course on perceiving an obstacle).



ARTIFICIAL INTELLIGENCE

Technique enabling machines to mimic human behavior)

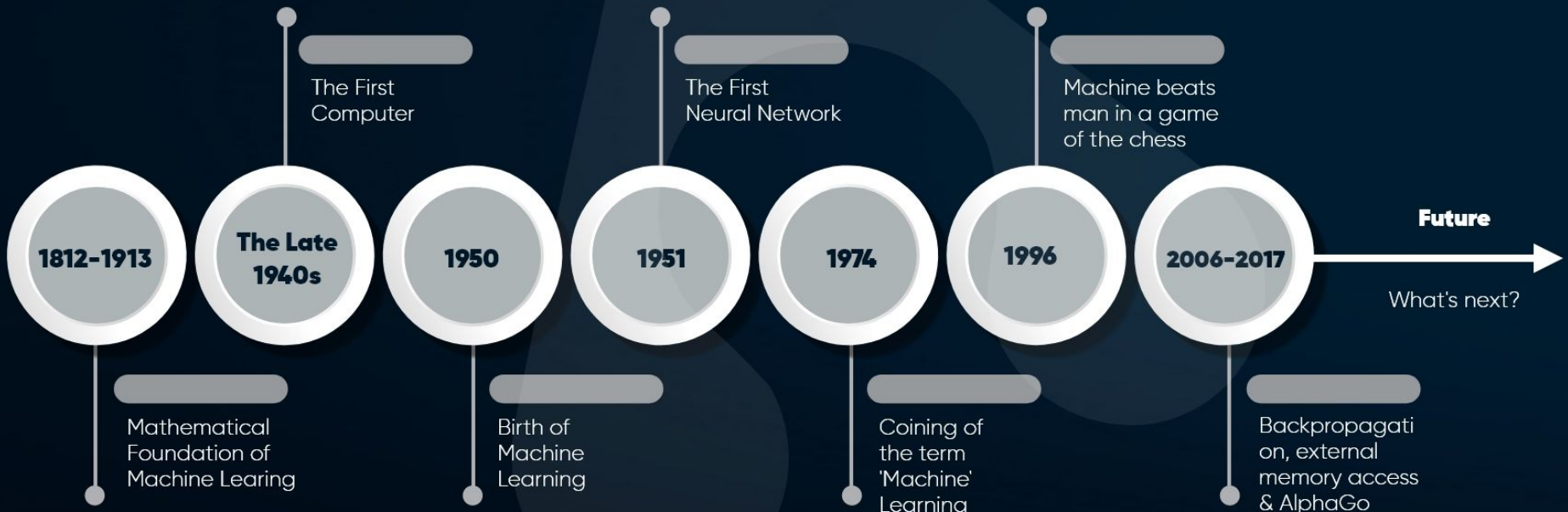
MACHINE LEARNING

A subset of AI technique that uses statistical methods for enabling the machine to improve with experience

DEEP LEARNING

A subset of AI meant for making the computation of multi-layer neural networks feasible

Timeline of Machine Learning



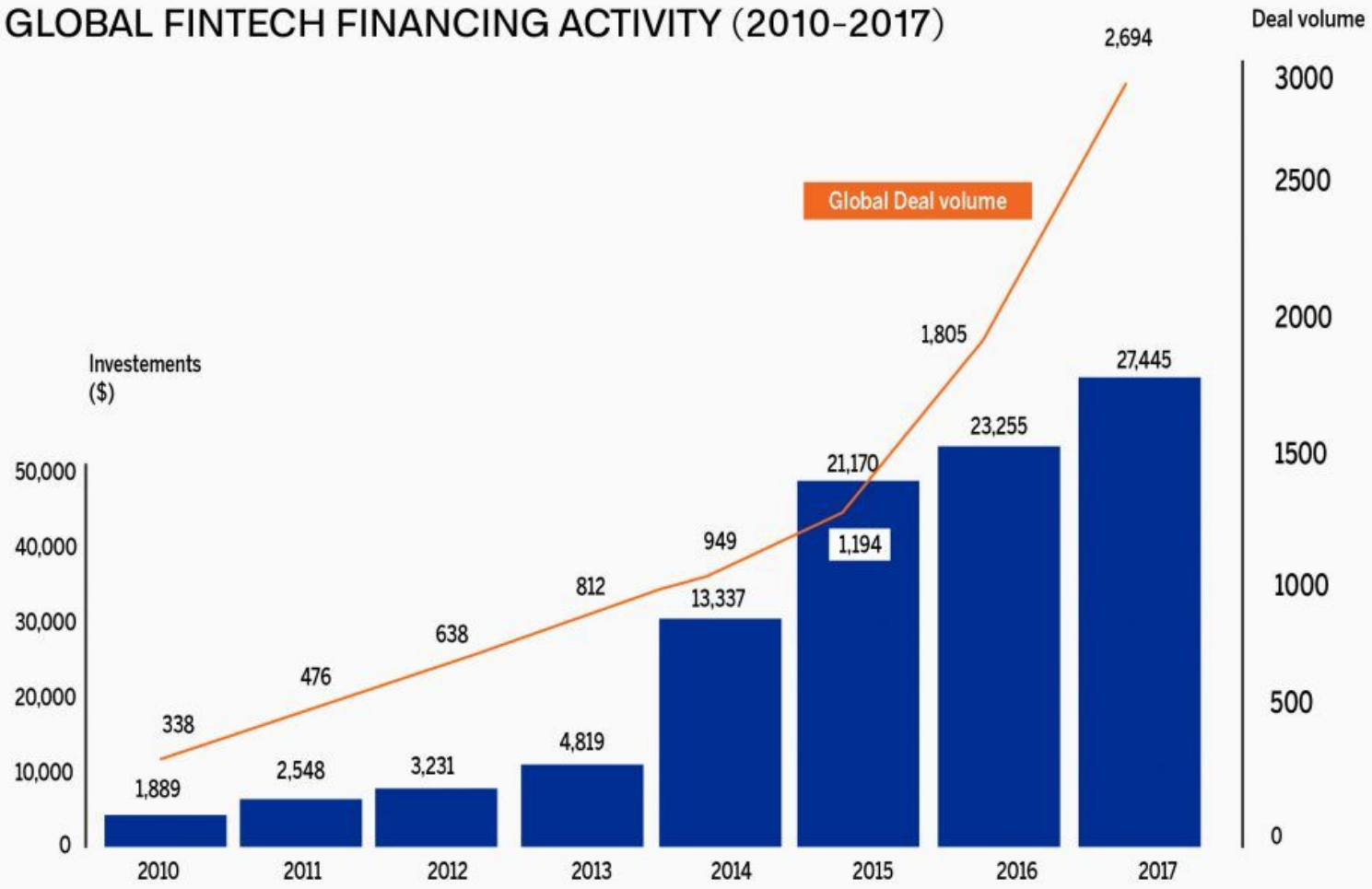
AI/ML – why is this tech different?

- Progression or step function?
- If progression, what is it progressing from?
  - Rules based programming?
- AI/ML
  - Difficult things can be very easy
  - Simple things can be very hard
  - In its infancy
  - "continuous improvement is better than delayed perfection.” –Mark Twain
- Any other potential technologies that could be a step function in change?

JPMorgan's Dimon says banks should be 'scared sh–less' about fintech challengers

Dimon said the bank faced competition from the likes of PayPal, Square, Stripe, Ant Financial, Amazon, Apple and Google

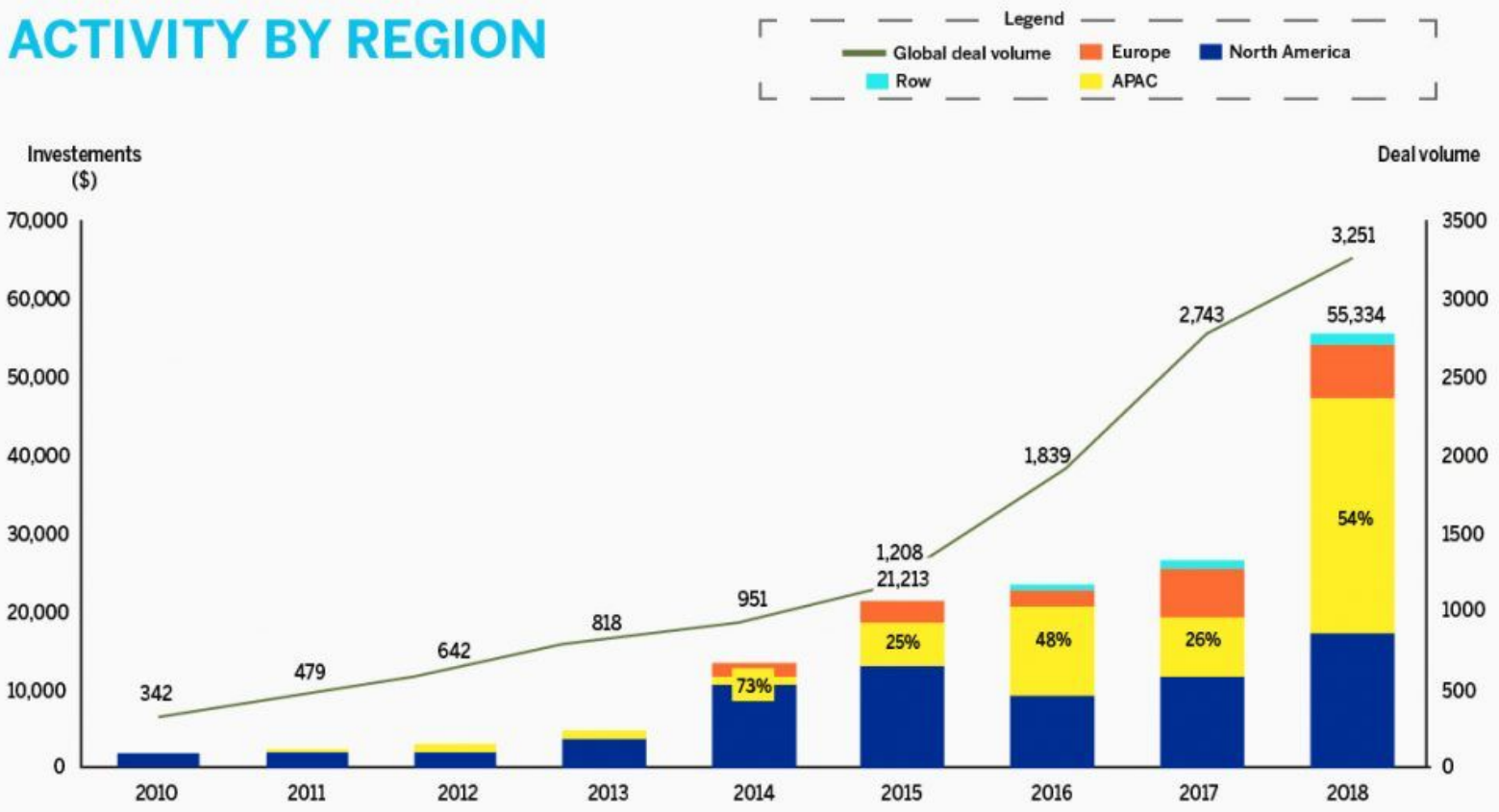
Nearly US$ 100 billion has flowed into fintech ventures since 2010



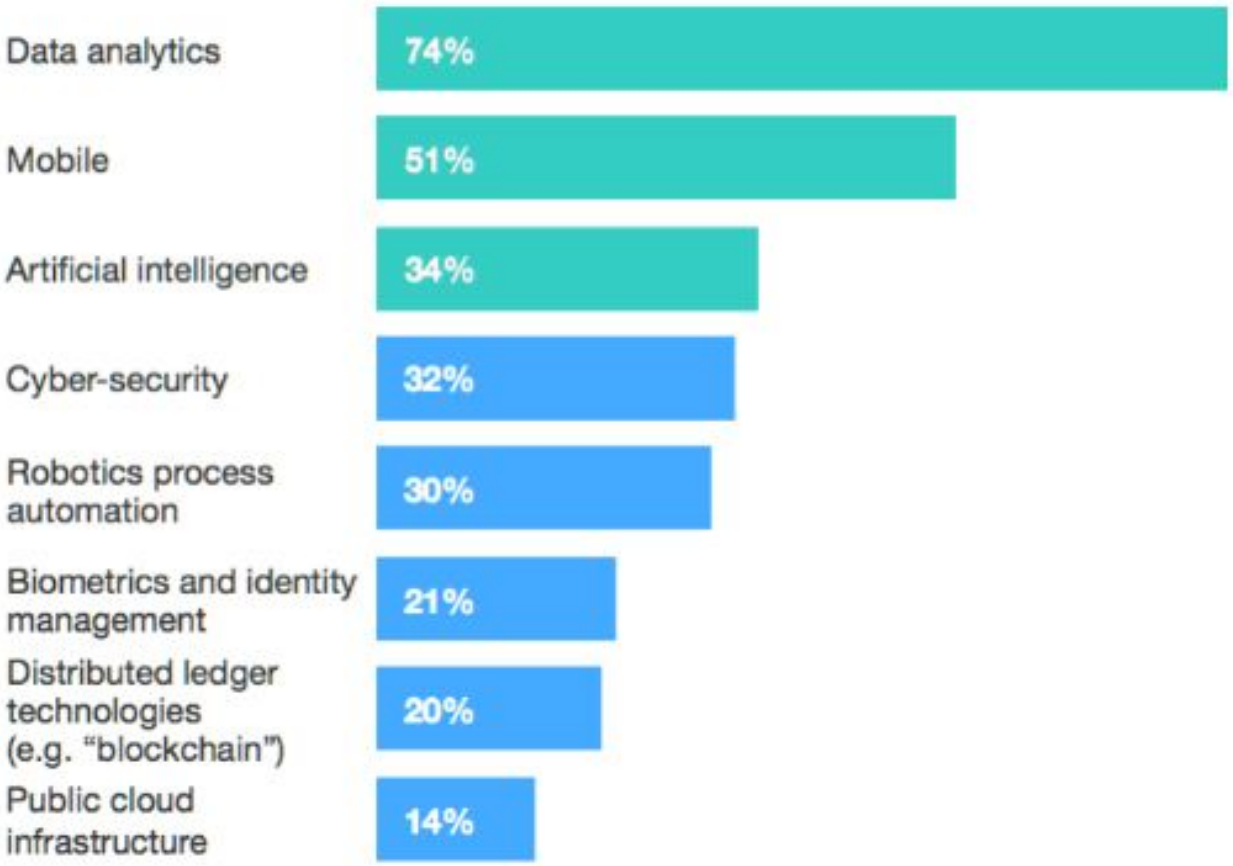
GLOBAL FINTECH FINANCING

ACTIVITY BY REGION

NO. OF DEALS GREW 18.5% YOY IN 2018. TOTAL FUNDING IN THE SAME PERIOD GREW AT 107% PRIMARILY DRIVEN BY THE $14BN ROUND IN CHINA

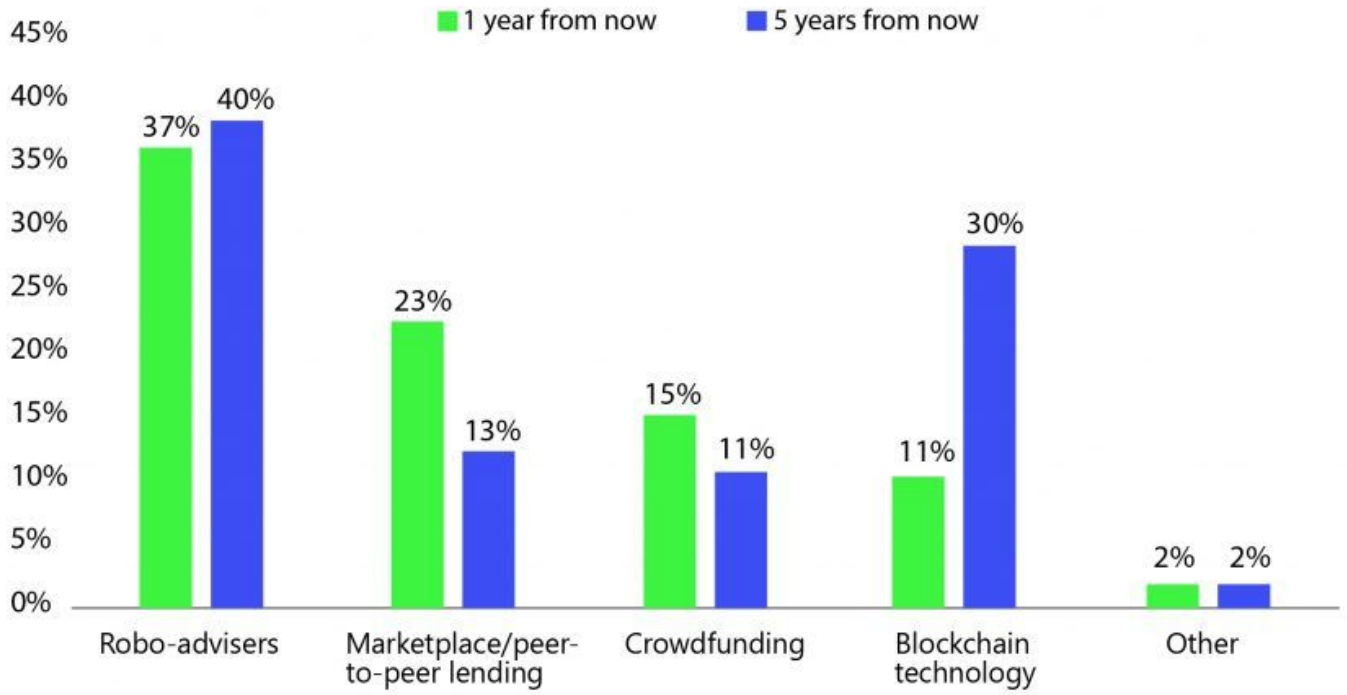


What are the most relevant technologies for your business that you plan to invest in within the next 12 months?



Source: PwC Global FinTech Survey 2017

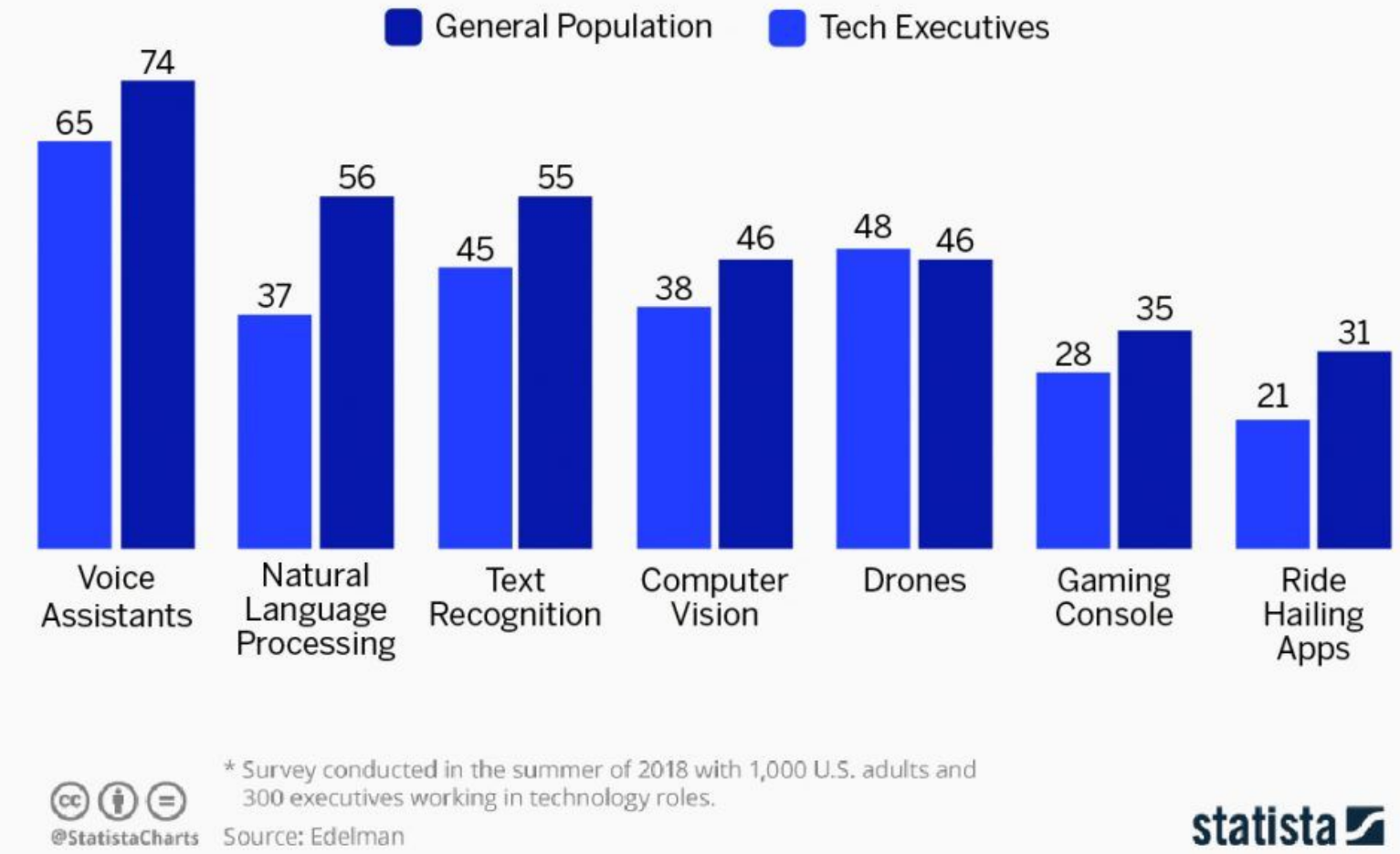
Greatest impact on financial services industry, by timeline*



Question: Which technology do you see as having the greatest impact on the financial services industry 1 year and 5 year from now?
•Excludes not sure

Bankers underestimate how their cutomers feel about Fintech services

Source: https://www.appventurez.com/blog/ai-in-fintech/



Source: https://www.appventurez.com/blog/ai-in-fintech/

Would you rather be operated on by an AI-assisted surgeon or have your finances managed by an AI-assisted bank?

Consumers uncomfortable and not confident with AI in the banking sector

Consumers who are willing to undergo AI-assisted surgery

Consumers comfortable with banks using AI to provide financial guidance

SOURCE: Autonomous © May 2018 The Financial Brand

ML and Python for Finance Professionals

Challenges & Opportunities

SHARE

CFO JOURNAL

Finance Chiefs Are Still Trying to Replace Excel With New Tools

Many companies would like to reduce their reliance on the spreadsheet application, but employees remain reluctant to give it up

Finance tech tools

- Excel still king
  - Many companies want to reduce reliance on Excel, but many employees are reluctant

"For professions that have long relied on trawling through spreadsheets, Python is especially valuable. Citigroup, an American bank, has introduced a crash course in Python for its trainee analysts." - The Economist

Benefits of Python for finance

- Python
  - Actual programming language
    - Interpreted vs compiled languages
  - State of the art (but keep an eye on Julia), especially for ML
    - Library support
  - “Production” quality, scales, automate
    - Solo excel instance vs shared program instance, with all the benefits of software control
    - Solid component of software stack/ecosystem (Excel, more difficult)
  - Common language between software engineers and quant/fundamental finance professionals
    - You as quant finance person, the bridge between the two
    - can generate excel spreadsheets if you need to
  - Pandas vs Excel
- Probably still need to be Excel power user, for now, but Python is great skill to have
  - Not even sure prototyping in excel is valid use case anymore
  - Adapt as necessary to your company needs, but push the innovation envelope

ML Use Cases

- Alternate data sets
  - Unstructured data (up to 90% of data)
- Ecommerce
- Decision making, insurance, banking
- Fraud detection
- Claim handling
- Customer support
- Wealth management
- Investing/trading
- Predictive analysis
- Risk management

This figure presents the parking lot satellite image of the Target store located at 4500 Macdonald Ave, Richmond CA 94805. The image was captured by RS Metrics on September 19, 2016 at 11:03am. The red line outlines the boundary of the parking lot associated with Target and the red dots indicate the occupied parking lot spaces. For this case, RS Metrics identifies 540 parking lot spaces with 146 of them filled.

Source: “You can make millions counting cars in parking lots from space” - Quartz

ecommerce

- CLV modeling- With the help of ML, it can check about total customer value along with learning the early indicators.
- Pricing- While analyzing the behavioral indicators, the external factors are also taken into account for dynamic pricing in real-time.
- Recommendations- Using ML, recommendations about a product or service can be made to the users in real-time. This will not just help the user but also enhance their experience to the fullest.

Roll over image to zoom in

HAGOROMO Fulltouch Color Chalk 1 Box [72 Pcs/10 Color Mix]

Visit the Hagoromo Store

4.5 out of 5 stars 820 ratings | 39 answered questions

Price: $49.90

Thank you for being a Prime Member. Get $125 off: Pay $0.00 ~~$49.90~~ upon approval for the Amazon Business Prime Card. Terms apply.

Color: **10 Color Mix**

Color	10 Color Mix
Ink Color	Blue
Brand	Hagoromo
Material	Calcium carbonate

About this item

- WELL COATED AND DUST FREE: The only thing worse than chalk breaking in half is getting dust all over your hands and clothes when writing on the board or drawing on the sidewalk. All Hagoromo chalks are well coated to prevent hands from coming into contact with chalk dust when using. Never have to worry about stained, dusty hands when using chalk ever again
- VIBRANT AND CONCENTRATED: All chalks are highly concentrated in color, are vibrant and have bright color payoff. Great visibility makes it perfect for use in large classrooms or lecture halls. Marks smoothly, creating clear and distinct lines on chalkboards and sidewalks
- NO MISTAKES: Our chalks erase easily leaving little to no residue on the chalkboard when removed. Chalk can be easily washed from most articles of clothing without leaving any

$49.90

$9.90 delivery: **Sep 21 - Oct 8**

Deliver to Renzo - New York 10036

In Stock.

Qty: 1

Add to Cart

Buy Now

Secure transaction

Ships from and sold by LoveDoki.

Return policy: Eligible for Return, Refund or Replacement.

Add to List

Share

Have one to sell?

Sell on Amazon

20% off coupon

Arteza Real Brush Pens, 48 Colors for Watercolor Painting...

AI/ML Components

- Data
- Algos
- Computing
- What is the problem and can ML be a potential solution?
  - Vs How can I apply this ML solution to a problem?
  - Framing the problem as an ML problem

The Algos

Machine learning algos are basically divided into different categories

- Learning Algorithms
  - Supervised Learning
  - Unsupervised Learning
  - Semi-Supervised Learning
- Similarity Algorithms
  - Regression Algorithm
    - Linear Regression
    - Logistic Regression
  - Decision Tree Algorithm
  - Artificial Neural Network
  - Support Vector Machine
- NLP
- Time Series
- Reinforcement Learning

Computing Providers

- Google DeepMind and GCP
- Amazon AI Platform
- Microsoft Azure AI Platform
- Custom Hardware
- Proprietary Platforms

All AI Useful? Hot dog?

- https://www.youtube.com/watch?v=ACmydtFDTGs