

How AI Can Improve Your Data Strategy

Hugo Bowne-Anderson
[@hugobowne](https://twitter.com/hugobowne)



DataCamp

→ Hugo Bowne-Anderson, data scientist at DataCamp

- ◆ Undergrad in sciences/humanities (double math major)
- ◆ PhD in Pure Mathematics (UNSW, Sydney)
- ◆ Applied math research in cell biology (Yale University,
Max Planck Institute)
- ◆ Python curriculum engineer at DataCamp
- ◆ Host of [DataFramed, the DataCamp podcast](#)
- ◆ Data & AI evangelist, strategy consultant



1. Data science
2. Machine learning
3. Artificial intelligence



- Defining data strategy
- What is data science?
- What is machine/deep learning?
- What is AI?
- Foundations for your data strategy

- Defining data strategy
- What is data science?
- What is machine/deep learning?
- What is AI?
- Foundations for your data strategy

- “Data strategy refers to an organization’s ability to deploy advanced analytics at every point of interaction – human as well as machine – to continuously improve decision-making quality and accuracy.”



- “A real strategy involves a clear set of choices that define what the firm is going to do and what it’s not going to do.”

STRATEGY EXECUTION

Many Strategies Fail Because They're Not Actually Strategies

by [Freek Vermeulen](#)

**Harvard
Business
Review**

STRATEGY

What's Your Data Strategy?

by [Leandro DalleMule](#) and [Thomas H. Davenport](#)

From the May-June 2017 Issue

The Elements of Data Strategy

	DEFENSE	OFFENSE
KEY OBJECTIVES	Ensure data security, privacy, integrity, quality, regulatory compliance, and governance	Improve competitive position and profitability
CORE ACTIVITIES	Optimize data extraction, standardization, storage, and access	Optimize data analytics, modeling, visualization, transformation, and enrichment
DATA-MANAGEMENT ORIENTATION	Control	Flexibility
ENABLING ARCHITECTURE	SSOT (Single source of truth)	MVOTs (Multiple versions of the truth)

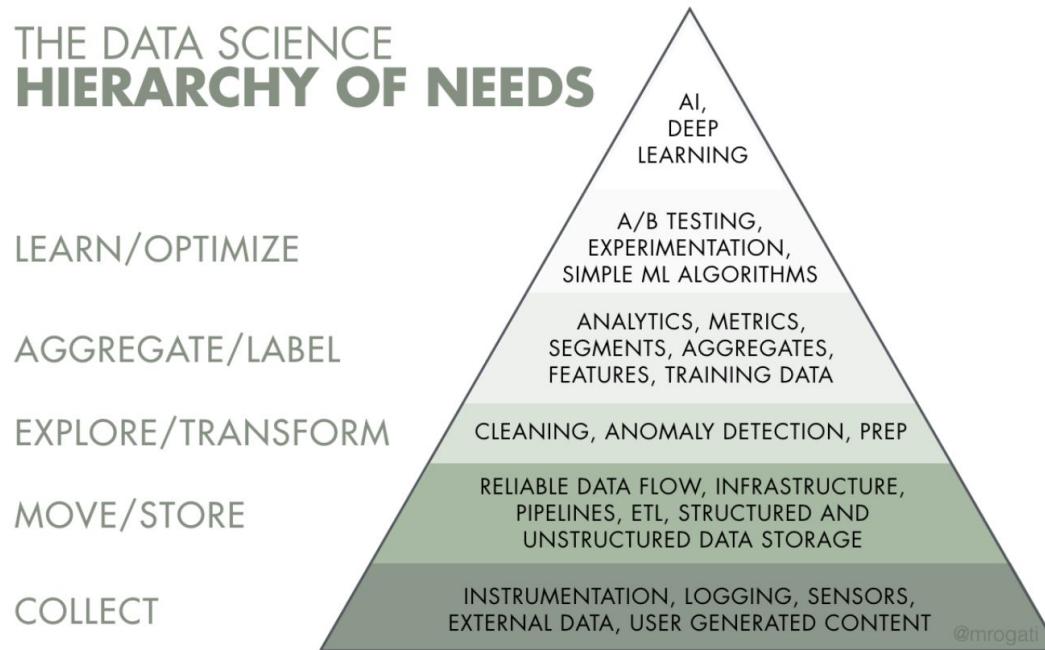
From "WHAT'S YOUR DATA STRATEGY?" BY LEANDRO DALLEMULE AND THOMAS H. DAVENPORT, MAY-JUNE 2017

© HBR.ORG

Harvard
Business
Review

- Most often, data is a key tool to serve business decisions & strategy
- Literally, data translates as “things given”
- We need to rebrand data to mean “things collected”





You need a solid foundation for your data before being effective with AI and machine learning.

Monica Rogati, [The AI Hierarchy of Needs](#)

Robert Chang (Airbnb) identified 3 data stages for tech startups: ([Episode 7, DataFramed](#)):

1. Data foundation
2. Growth and online experimentation
3. Machine learning

```
total_poker <- sum(poker_vector)
11 # Total winnings with roulette
12 total_roulette <-
13
14 # Total winnings overall
15 total_week <-
16
17 # Print out total_week
18
```

Instructions

- Calculate the total amount of money that you have won/lost with roulette and assign to the variable `total_roulette`.
- Now that you have the totals for roulette and poker, you can easily calculate `total_week` (which is the sum of all gains and losses of the week).
- Print out `total_week`.

 Take Hint (-30%)

 Submit Answer

AI? Data Science?

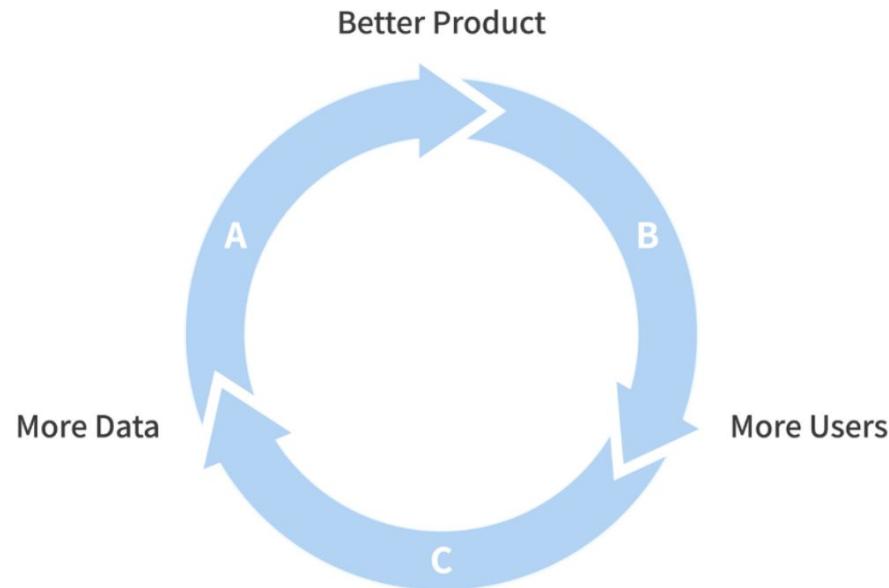
- **Data science** produces **insights**
- **Machine learning** produces **predictions**
- **Artificial intelligence** produces **actions**

David Robinson, [What's the difference between data science, machine learning, and artificial intelligence?](#)

“The most impactful and effective AI strategies will stand on the shoulders of robust data science capabilities.”

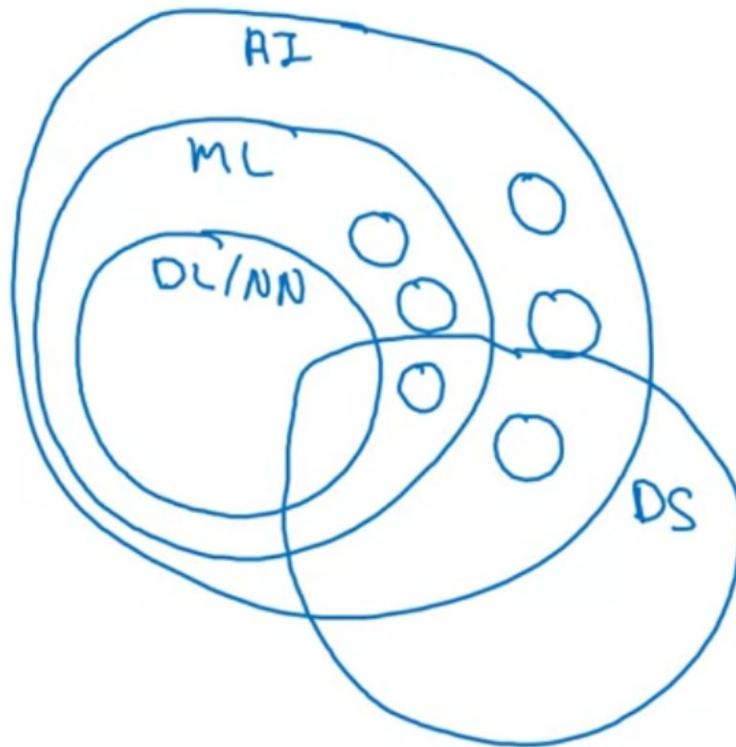
Me

The Virtuous Cycle of AI



Andrew Ng, [The AI Transformation Playbook](#)

Yet another data science Venn diagram



Andrew Ng, [AI for Everyone](#)

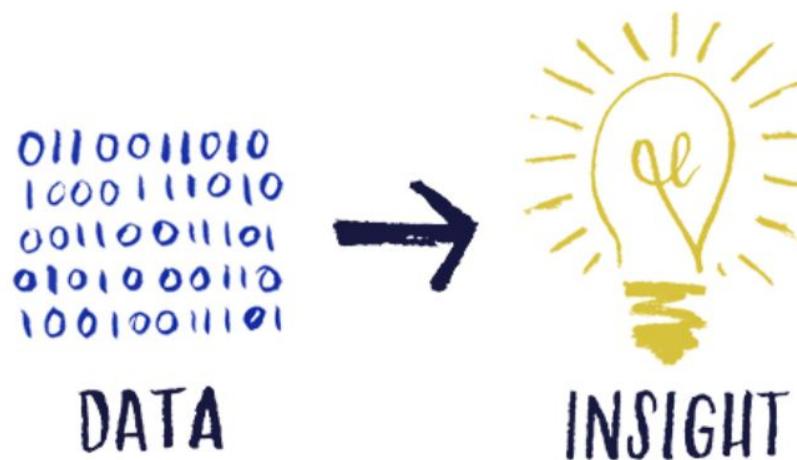
- Data science will, yes.
- Machine learning will if you need to make predictions.
- AI will if you build data products (products that make decisions based on data in real-time).
- Serious warning: people use the term AI to mean any of the above.



- Working definitions of data strategy
- Data strategy will *always* be a function of business strategy
- The Data Science/AI Hierarchy of Needs
- All data strategy starts with data foundations
- Teaser: DS, ML, AI, and data strategy

- Defining data strategy
- What is data science?
- What is machine/deep learning?
- What is AI?
- Foundations for your data strategy

- Data Science (DS) is about making discoveries and creating insights from data and communicating these insights and discoveries to non-technical stakeholders.



Today's data scientist...

- Data Collection and Cleaning
- Building dashboards and reports
- Data visualization
- Building models (statistical inference & machine learning)
- Communicate results to stakeholders
- Business decisions are then made!

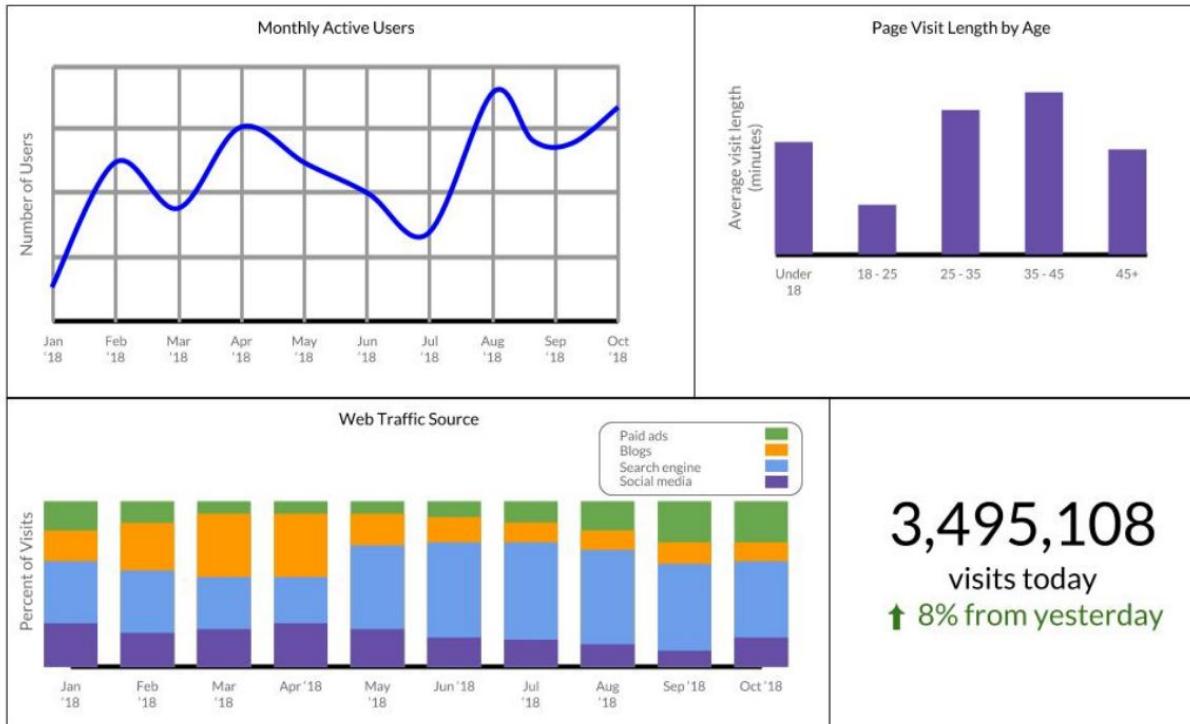
Jacqueline Nolis breaks data science into 3 components ([Episode 28, DataFramed](#)):

1. Descriptive analytics (Business Intelligence)
2. Predictive analytics (Machine Learning)
3. Prescriptive Analytics (Decision Science)

1. Business Intelligence (descriptive analytics)

- Taking data company already has
- Getting that data to the right people
- In form of dashboards, reports, emails





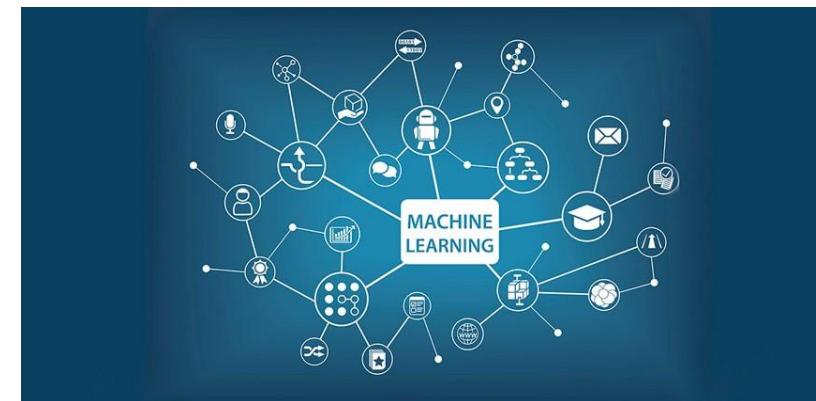


« Everyone at any level, C-level, entry level, should be looking and diving into data the same way that you were expected to start using email 20 years ago. »

TANYA CASHORALI

2. Machine Learning (predictive analytics)

- Make predictions
- Often put models continuously into production (AI)
- E.g. will a customer churn?



3. Decision Making (prescriptive analytics)

- Take the insight discovered from the data science work
- Use it to help company decision making
- E.g. what do you if your data science work tells you a particular type of customer will churn?



0110011010
1000111010
00110011101
0101000110
10010011101



DATA



INSIGHT



Data



Insight



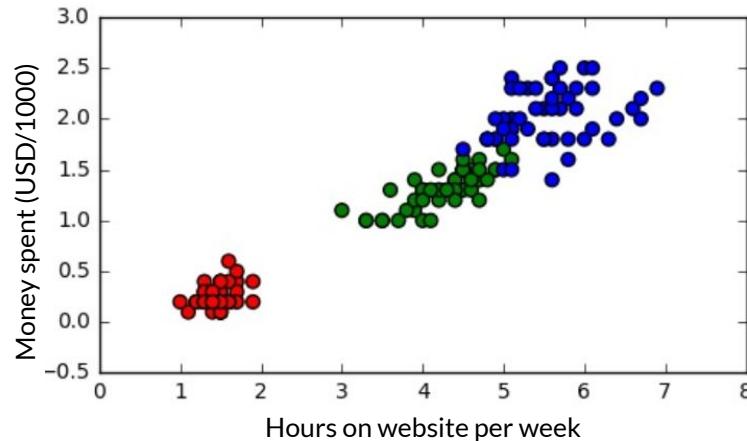
Action

- Working definitions of data science
- What data scientists actually do
- Slicing data science into *descriptive, predictive, and prescriptive* analytics
- Data science serves decision making and strategy!

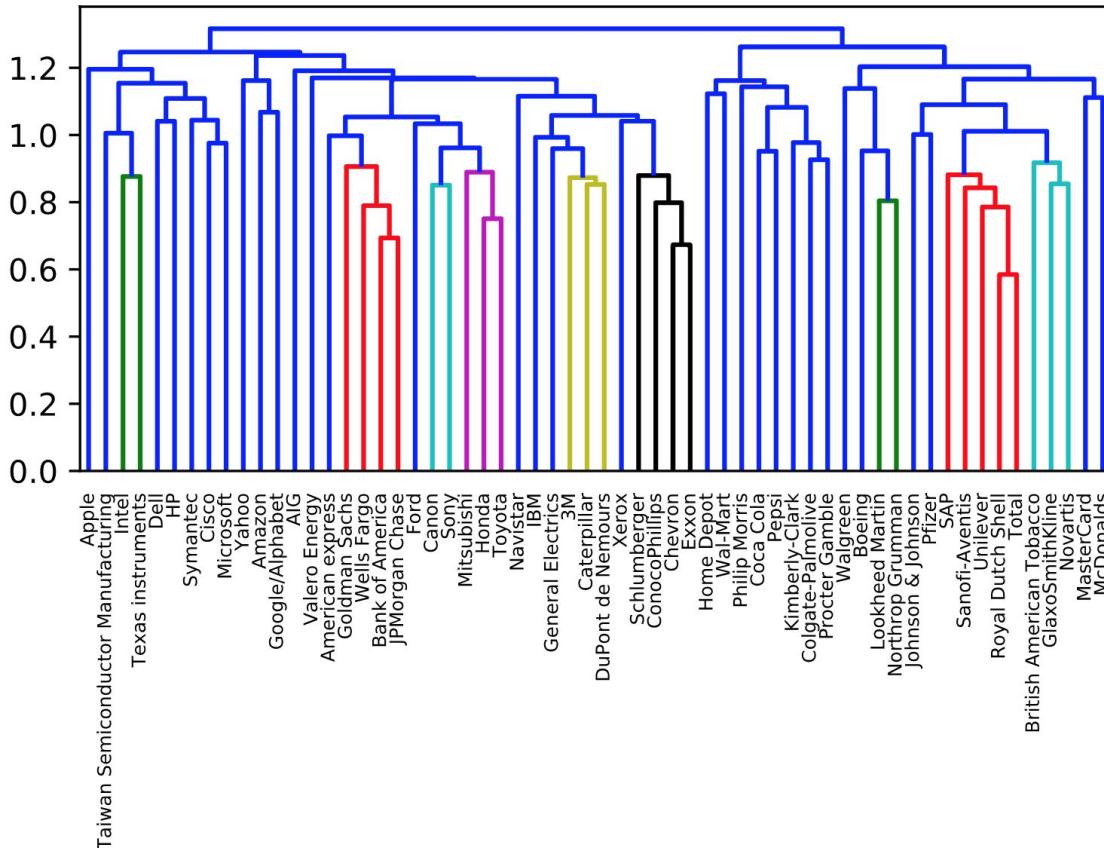
- Defining data strategy
- What is data science?
- What is machine/deep learning?
- What is AI?
- Foundations for your data strategy

- “the science and art of giving computers the ability to learn from data without being explicitly programmed.”
 - ◆ Email: spam or not (supervised)
 - ◆ Text clustering (unsupervised)

- the machine learning task of uncovering hidden patterns and structures from unlabeled data:
 - ◆ E.g. Clustering texts, customers, genes.



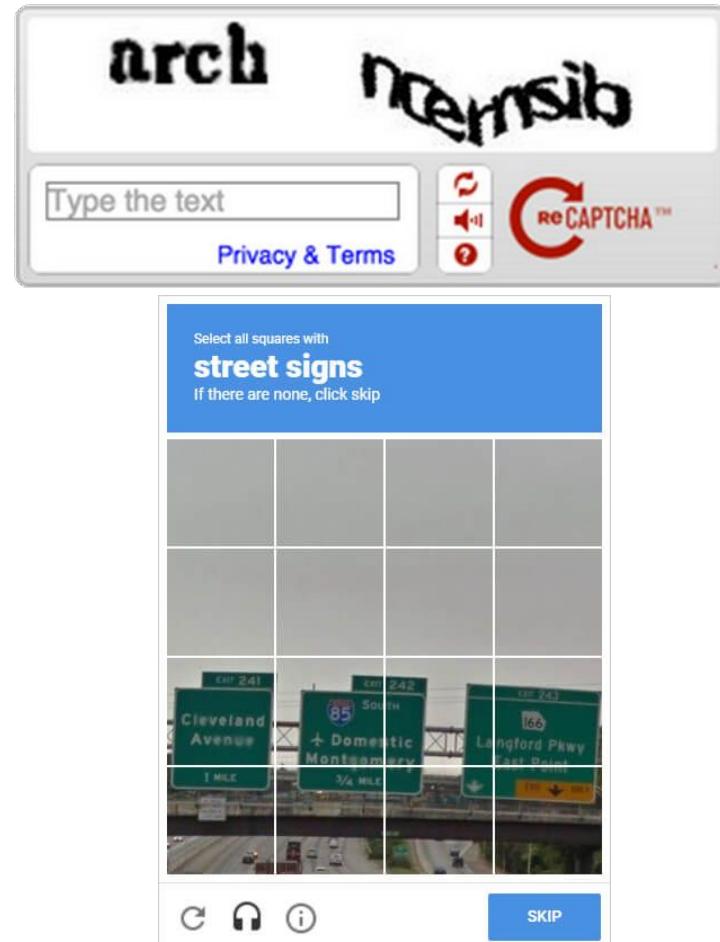
Hierarchical Clustering: Stock Movements



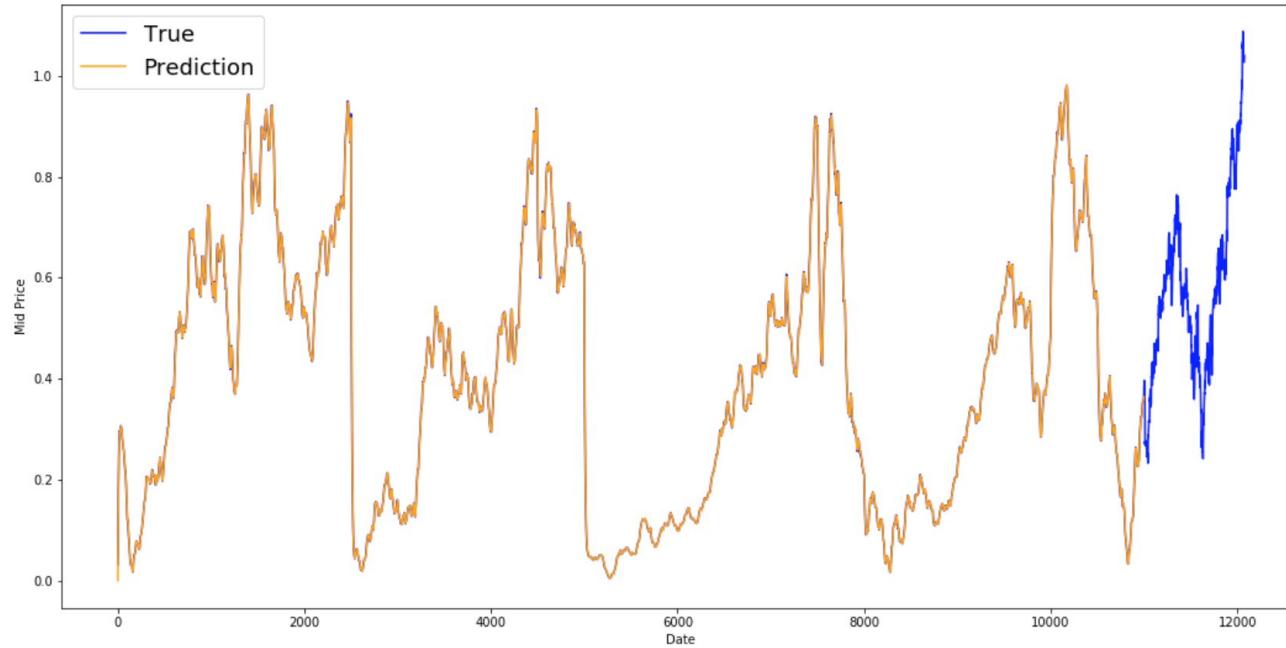
- automate a time-consuming or expensive manual task, such as a doctor's diagnosis or
- to make predictions about the future, say whether a customer will click on an ad or not.

Need for labelled data

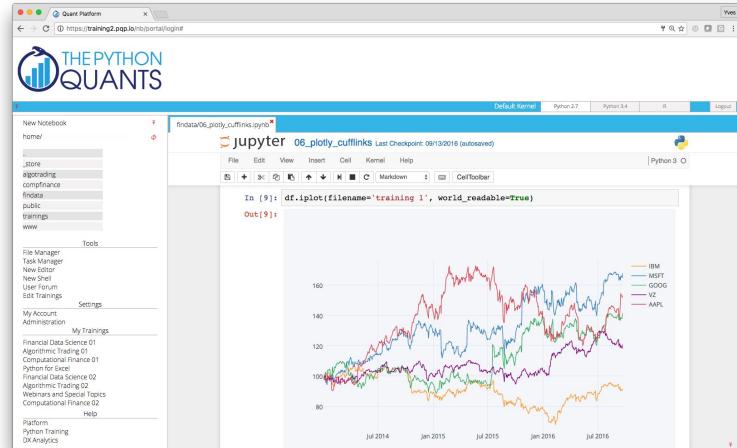
- Goal: learn from data for which the right output is known, so that we can make predictions on new data for which we don't know the output.
- Historical data
- Do an experiment
- Crowd-source (e.g. ReCaptcha, FB)



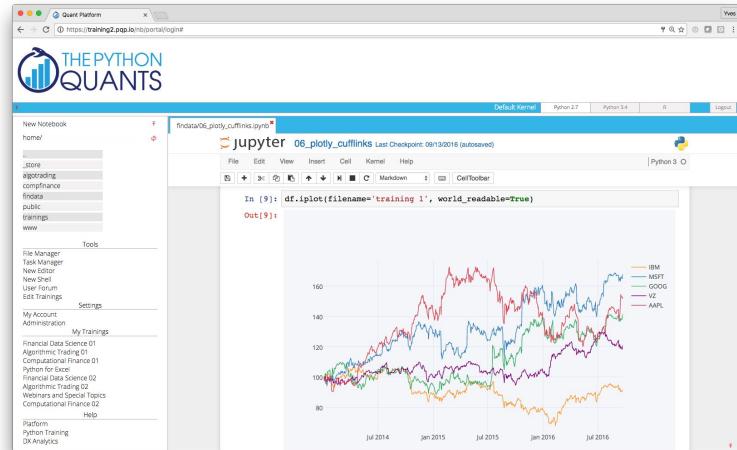
Stock market prediction



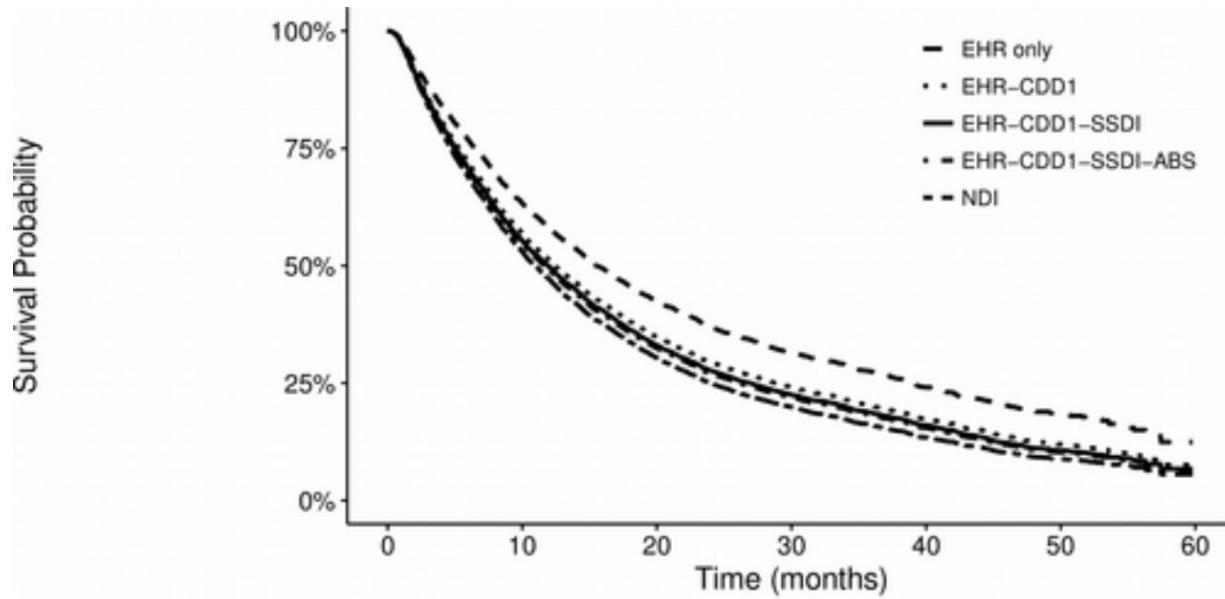
→ “Traders used to be first-class citizens of the financial world, but that’s not true anymore. Technologists are the priority now...” -- Robin Wigglesworth



→ “...with a little bit of training, you can accomplish quite a lot compared to the traditional approaches in this field.” -- Yves Hilpisch

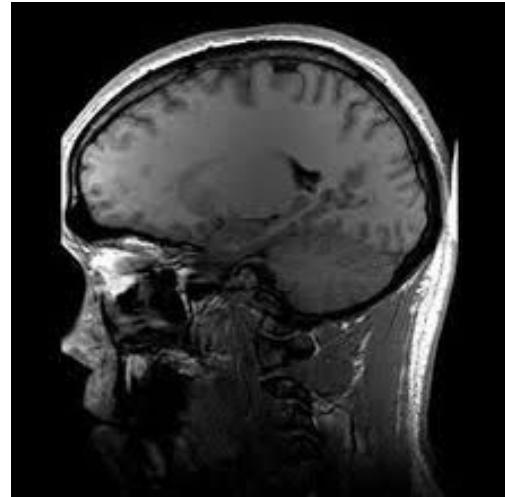


Cancer patient survival probability

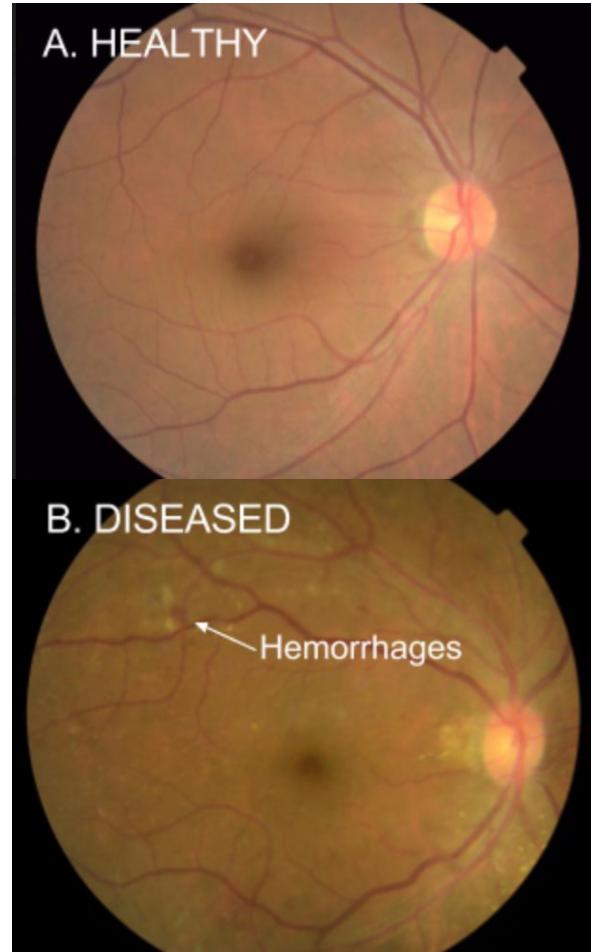


Flatiron Health

- Disease diagnosis from imaging data
(e.g. MRI, PET)



- Google AI have developed an algorithm to detect diabetic retinopathy as well as experts
- One of the leading causes of blindness globally. Around 400 million people have this disorder.



Value of homes on Airbnb

<code>id_listing</code>	<code>host_location_city</code>	<code>avg-nightly-price</code>	<code>availability_next_180_days</code>	<code>1_year_revenue</code>
1	London	\$120	50 nights	\$
2	San Francisco	NULL	150 nights	\$\$
3	Tokyo	\$55	NULL	\$\$\$
4	New York	\$100	90 nights	\$\$\$\$

Robert Chang, airbnb

Insurance Payouts

- Given images of an accident, can you tell me the how much will be paid out?



```
total_poker <- sum(poker_vector)

11 # Total winnings with roulette
12 total_roulette <-
13
14 # Total winnings overall
15 total_week <-
16
17 # Print out total_week
18
```

Instructions

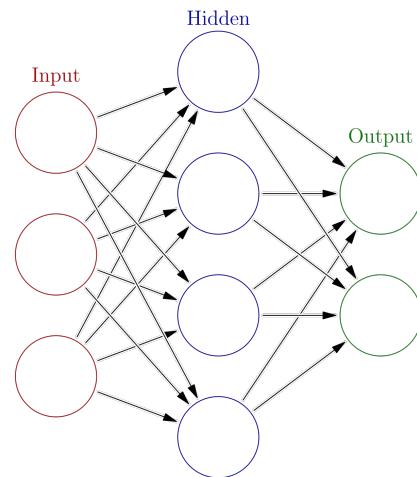
- Calculate the total amount of money that you have won/lost with roulette and assign to the variable `total_roulette`.
- Now that you have the totals for roulette and poker, you can easily calculate `total_week` (which is the sum of all gains and losses of the week).
- Print out `total_week`.

 Take Hint (-300g)

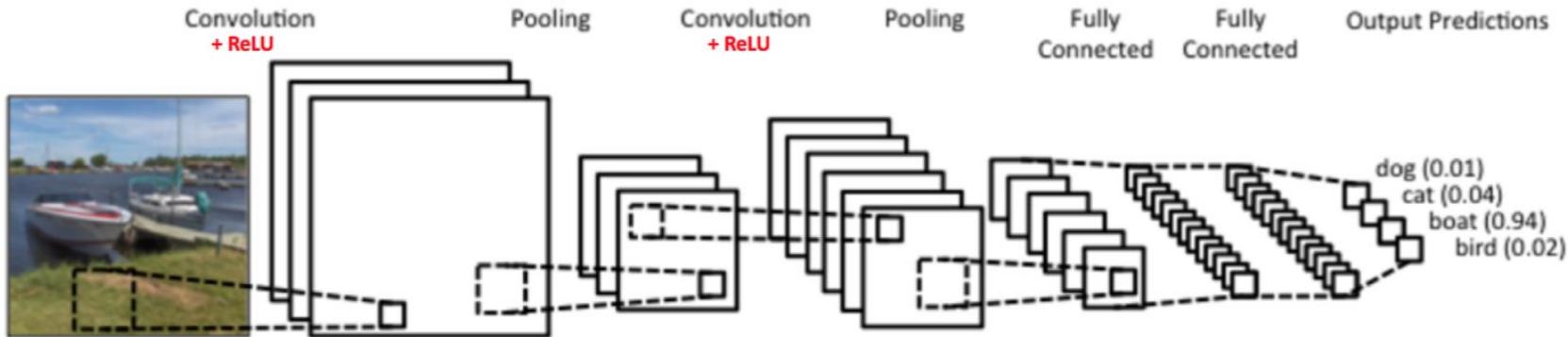
 Submit Answer

Deep Learning

- Deep learning (neural networks) is generally a particular form of supervised learning, i.e. prediction
- Used for image classification and natural language processing/textual analysis



- Deep learning (neural networks) is generally a particular form of supervised learning, i.e. prediction
- Used for image classification and natural language processing/textual analysis



- Accurate and fast at specific tasks
- Can be automated and put live

Google Translate



Text

Documents

ENGLISH - DETECTED

ENGLISH

SPANISH

FRENCH



SCOTS GAELIC

ENGLISH

SPANISH



This is deep learning



Is e ionnsachadh domhainn a tha seo



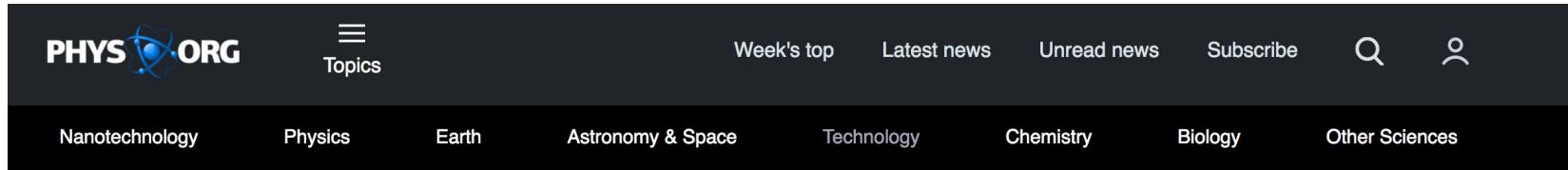
21/5000



Send feedback

- Needs lots of labelled data (1,000s or 10s of 1,000s)
 - ◆ Although: transfer learning
- Not great at complex tasks such as
 - ◆ report generation
 - ◆ intention classification in images





The screenshot shows the homepage of Phys.org. At the top, there's a navigation bar with links for "Topics", "Week's top", "Latest news", "Unread news", "Subscribe", a search icon, and a user profile icon. Below the navigation bar is a secondary menu with categories: Nanotechnology, Physics, Earth, Astronomy & Space, Technology, Chemistry, Biology, and Other Sciences. The main content area features a news article titled "Using deep-learning techniques to locate potential human activities in videos". The article is dated August 30, 2018, and is attributed to the Agency for Science, Technology and Research (A*STAR), Singapore. It includes a thumbnail image of people walking on a street with red boxes highlighting them, and a larger image of a person working on a laptop. There are social sharing icons for Facebook, Twitter, LinkedIn, and Email.

Home / Technology / Computer Sciences



(J) AUGUST 30, 2018

Using deep-learning techniques to locate potential human activities in videos

by Agency for Science, Technology and Research (A*STAR), Singapore



The 'YoTube' detector helps makes AI more human-centered. Credit: iStock



Software Engineer Salaries

[Learn more](#)

[LinkedIn](#)

Featured

Last Comments

Popular

Study suggests much more water on the moon than thought

(J) 1 HOUR AGO

0

Bird embryos respond to adult warning calls inside their shells

- Working definitions of machine learning
- ML, unsupervised learning, and supervised learning
- Examples of machine learning from finance, health, tech, and more
- Deep learning primer

- Defining data strategy
- What is data science?
- What is machine/deep learning?
- What is AI?
- Foundations for your data strategy

1. J.A.R.V.I.S.
2. A type of machine learning
3. Conscious machines
4. Term referring to systems capable of making intelligent decisions

An AI god will emerge by 2042 and write its own
bible. Will you worship it?

JOHN BRANDON @JMBrandonBB OCTOBER 2, 2017 2:10 PM

- A catch-all that refers to generally creating systems capable of making intelligent decisions
- Artificial General Intelligence (AGI) vs narrow AI

Google Translate



Text

Documents

ENGLISH - DETECTED

ENGLISH

SPANISH

FRENCH



SCOTS GAEIC

ENGLISH

SPANISH



This is deep learning



Is e ionnsachadh domhainn a tha seo



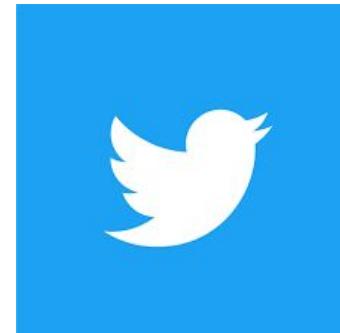
21/5000



Send feedback

What is artificial intelligence?

- Artificial Intelligence (AI) is a “huge set of tools for making computers behave intelligently,” including voice assistants, recommendation systems, and self-driving cars.





Process Automation



Security



Underwriting and credit scoring



Algorithmic trading



Robo-advisory



BNY MELLON

Integrating robotic process automation into their banking system



Identifying suspicious account behavior

ondeck

Underwriting model evaluates thousands of historical loans to improve future credit quality of portfolio



Analyzing many data sources simultaneously to make better trading decisions



Optimizing portfolio management and personalizing investment recommendations

Don't fear automation

```
total_poker <- sum(poker_vector)

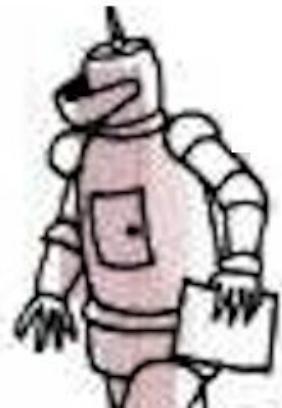
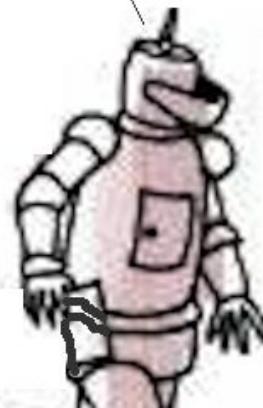
11 # Total winnings with roulette
12 total_roulette <-
13
14 # Total winnings overall
15 total_week <-
16
17 # Print out total_week
18
```



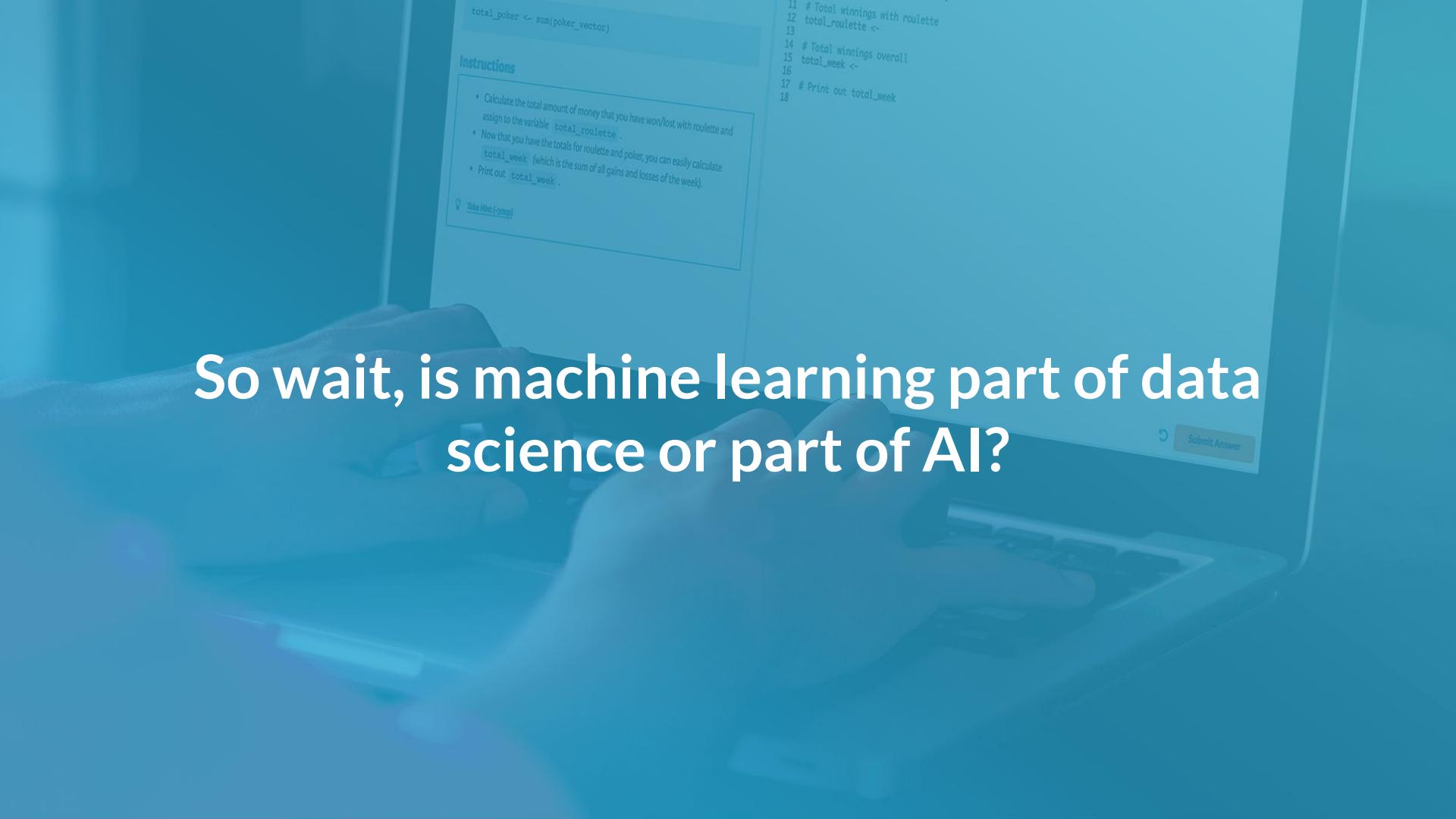
I GOT FIRED FROM
WORK TODAY

OH REALLY?
WHAT HAPPENED?

THEY REPLACED ME
WITH A HUMAN



So wait, is machine learning part of data science or part of AI?



- If ML output ---> human decision making process: data science

```
0110011010  
1000111010  
0011001101  
0101000110  
1001001101
```

DATA



INSIGHT

- If ML output ---> automated action: AI



WITH GREAT POWER COMES

GREAT RESPONSIBILITY



Amazon scraps secret AI recruiting tool that showed bias against women

Jeffrey Dastin

8 MIN READ



The Algorithmic Colonization of Africa

Startups are importing and imposing AI systems founded on individualistic and capitalist drives

Abeba Birhane

July 18, 2019

PRIVACY AND SECURITY

Google Plans Not to Renew Its Contract for Project Maven, a Controversial Pentagon Drone AI Imaging Program



Kate Conger

6/01/18 2:38pm • Filed to: PROJECT MAVEN

239.5K 161 29

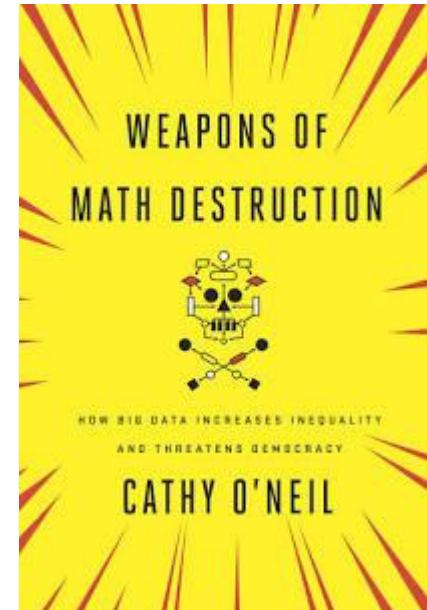


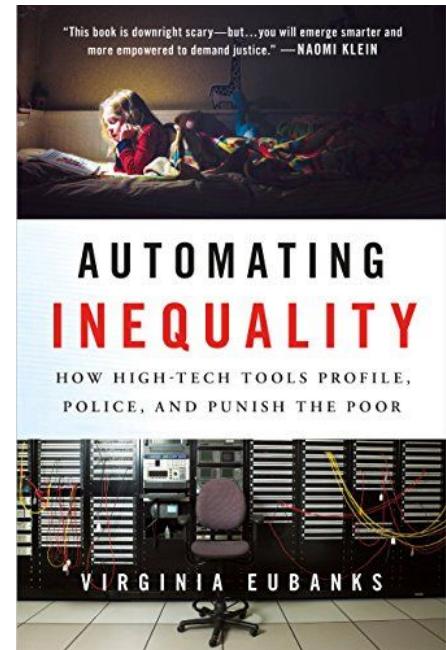
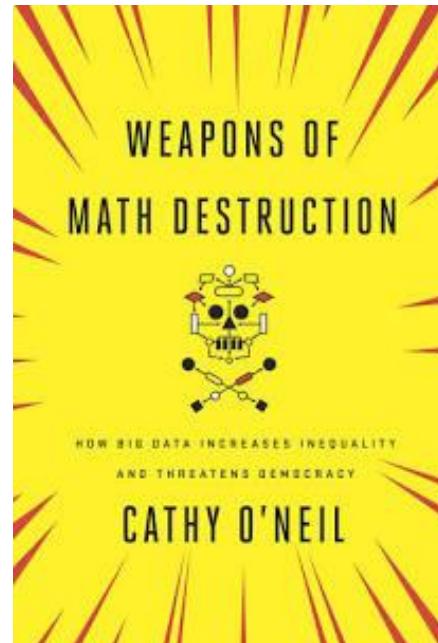
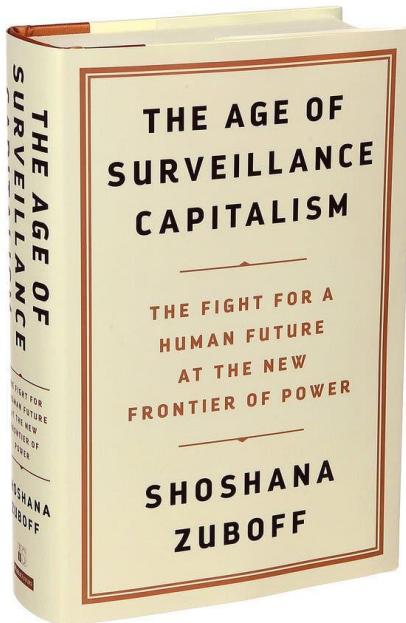
Machine Bias

There's software used across the country to predict future criminals. And it's biased against blacks.

by Julia Angwin, Jeff Larson, Surya Mattu and Lauren Kirchner, ProPublica
May 23, 2016

- Old definition of data scientist: “A data savvy, quantitatively minded coding literate problem solver.”
 - New definition: “Data science doesn't just predict the future. It causes the future.”
 - Algorithmic audits, including a sensitivity analysis
 - Ethical matrix: “rows are the stakeholders, the columns are the concerns.”





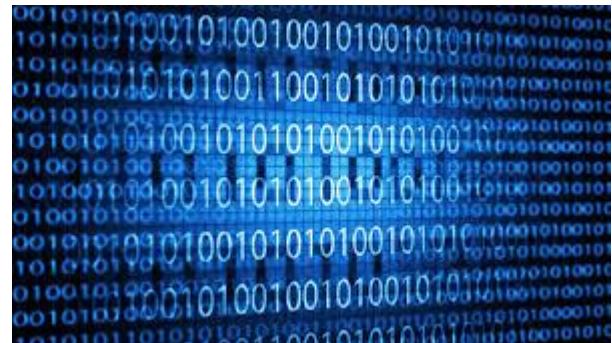
- Working definitions of AI
- The most impactful and effective AI strategies will stand on the shoulders of robust data science capabilities.
- Relationships and differences between DS, ML, and AI
- The ethics of data

- Defining data strategy
- What is data science?
- What is machine/deep learning?
- What is AI?
- Foundations for your data strategy

Taras Gorishnyy identified the necessary moving parts ([Episode 30, DataFramed](#)):

- Executive support
- Analytics vision
- Build data foundations
- Distribution of skills & establish data culture
- Establish impact of analytics early on in the process

- Data science will, yes.
- Machine learning will if you need to make predictions.
- AI will if you build data products (products that make decisions based on data in real-time).
- Serious warning: people use the term AI to mean any of the above.



Infrastructure

Set up a data lake
Enable data discovery

People

Map out roles and skills
Identify skill gaps
Personalize learning path

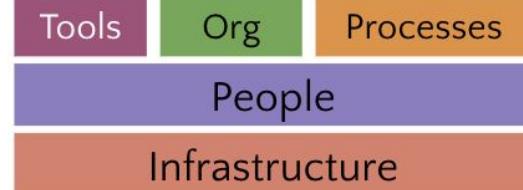
Tools

Build tools to encapsulate.
Build frameworks to automate.

Organization

Embrace a hybrid model
Build flexibility

IPTOP



Processes

Standardize project structure
Embrace version control
Embrace notebooks

- 5 foundational building blocks for your data strategy
- Will data science, machine learning, and AI impact your data strategy?
- Moving parts of a scalable data strategy (watch this space)

What's next?



```
total_poker <- sum(poker_vector)

11 # Total winnings with roulette
12 total_roulette <-
13
14 # Total winnings overall
15 total_week <-
16
17 # Print out total_week
18
```

Instructions

- * Calculate the total amount of money that you have won/lost with roulette and assign to the variable `total_roulette`.
- * Now that you have the totals for roulette and poker, you can easily calculate `total_week` (which is the sum of all gains and losses of the week).
- * Print out `total_week`.

Take Hint (-30ms)

Submit Answer

Now you know the basics, next time we'll delve deeper into key aspects of implementing data strategy.

- “5 things every business leader needs to know about data strategy”
- 11am ET, February 20 (the third Thursday of the month)



DataCamp

Thank you!



Hugo Bowne-Anderson
Data Scientist
[@hugobowne](https://twitter.com/hugobowne)