

COMP4434 Big Data Analytics

Lecture 1 Introduction to Big Data Analytics

HUANG Xiao

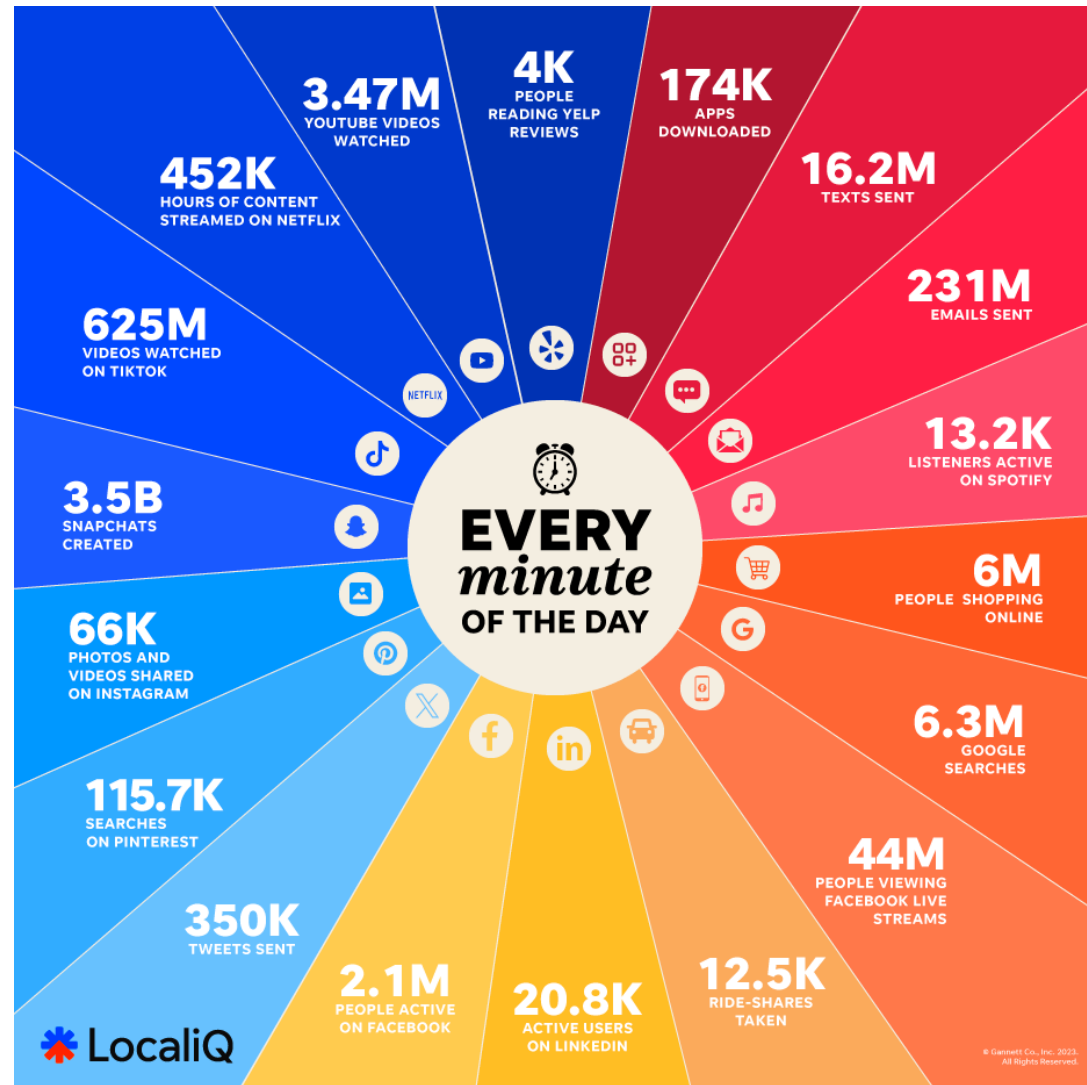
xiaohuang@comp.polyu.edu.hk

Arrangement

- Prerequisites:
 - Basic statistics, probability, linear algebra
 - Basic Computer Science fundamentals
 - programming (Python)
 - data structures
 - algorithms
 - database systems
- 26 hours Lectures + 13 hours Lab
 - Four lab sessions

What is big data?

- Posts on social media sites
- Purchase transaction records
- Digital pictures and videos
- Software logs
- Microphone & camera recordings
- Cell phone GPS signals
- Sensing data
- Scans of government documents
- Traffic data
- ...

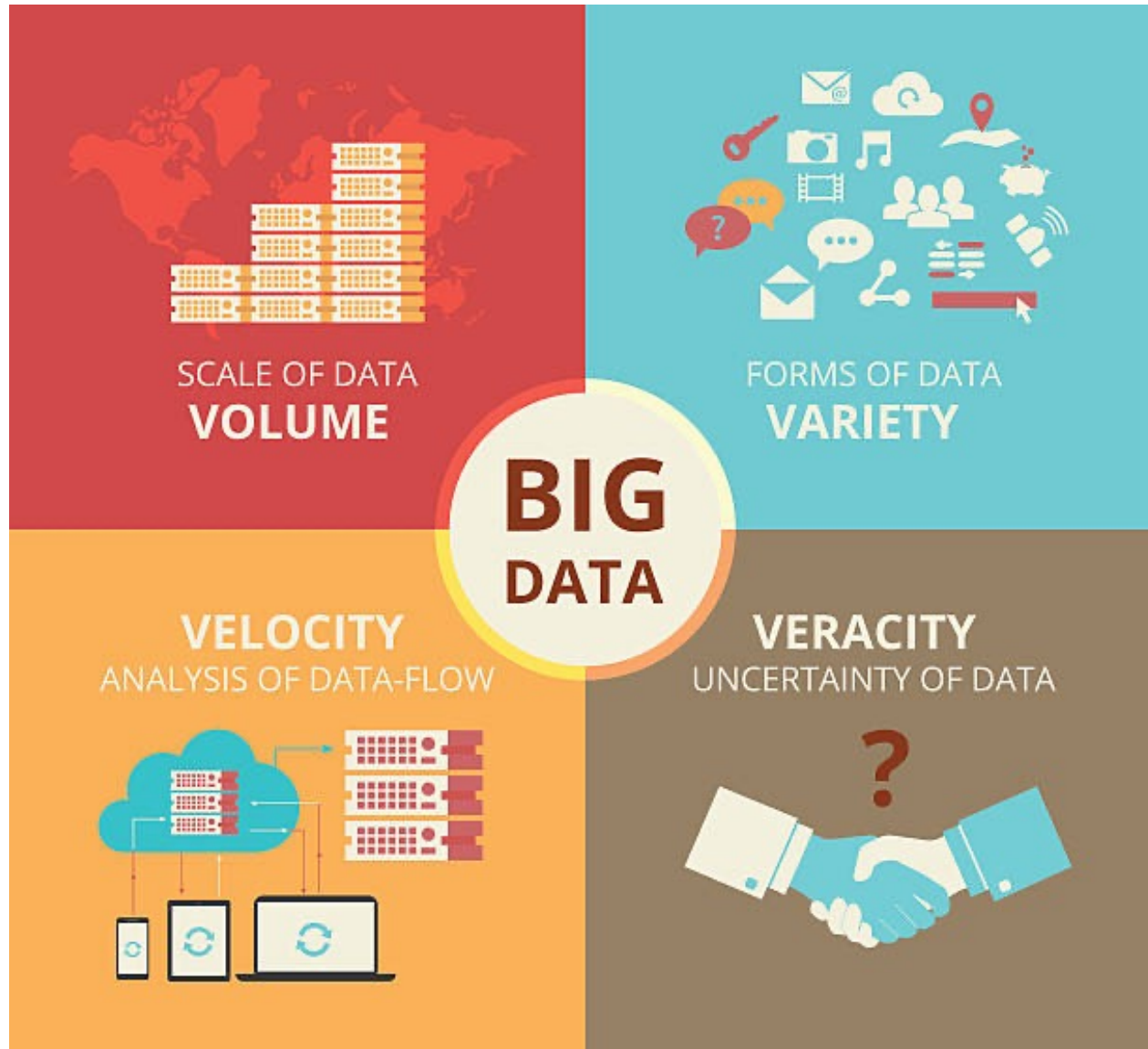


Definition of big data

- Definition 1: a combination of structured, semi-structured and unstructured data collected by organizations that can be mined for information and used in machine learning projects, predictive modeling and other advanced analytics applications.
- Definition 2: “Big data is high-**volume**, high-**velocity** and high-**variety** information assets that demand **cost-effective**, **innovative** forms of information processing for **enhanced insight and decision making**.” -- Gartner
- **4Vs** Characteristics of big data
 - **Volume, Velocity, Variety, Veracity**

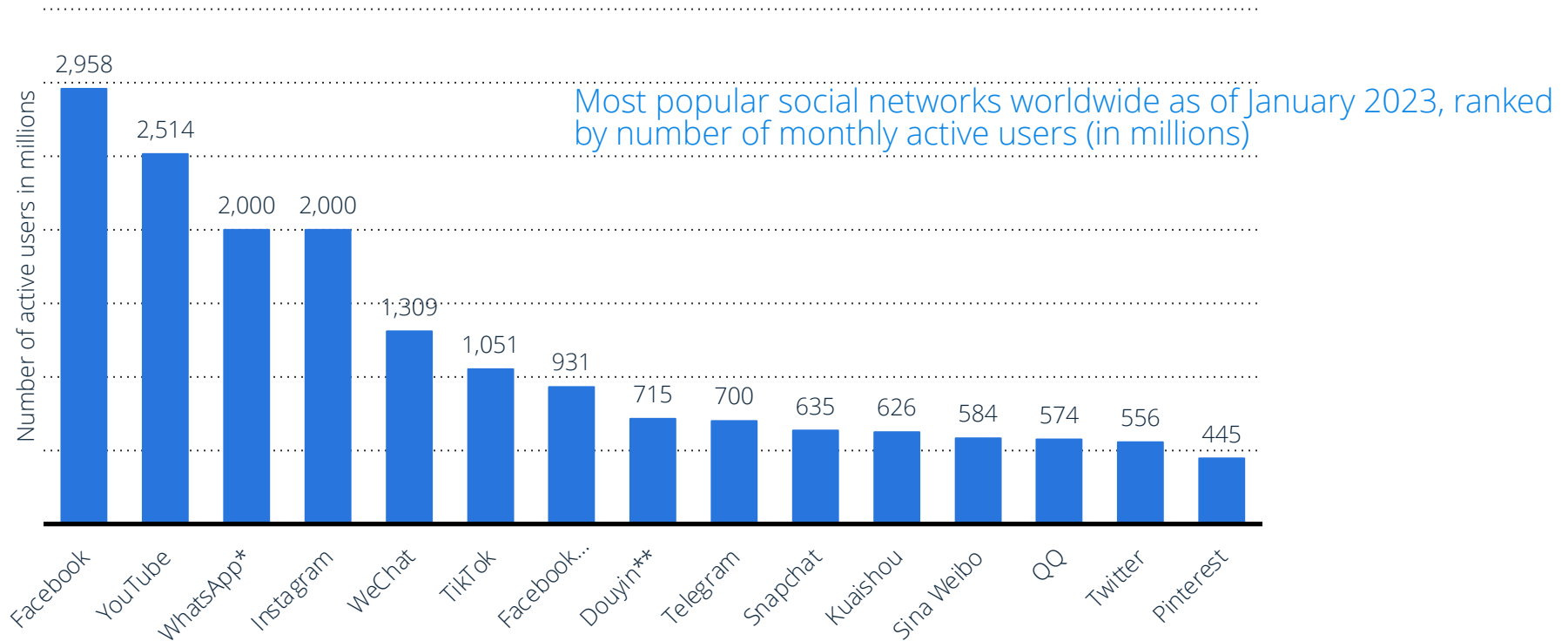
4Vs Characteristics of big data

- Volume, Velocity, Variety, Veracity



Volume: scale of data

- Data volume is increasing exponentially
- Generated by huge number of devices and sensors
 - The number of smartphone mobile network subscriptions worldwide reached almost 6.4 billion in 2022



Velocity: speed of data generation

- Data is generated fast
- Data need to be processed fast
 - Online Data Analytics: late decisions mean missing opportunities
 - e.g., 1: Based on your current location and your purchase history, send promotions right now for store next to you
 - e.g., 2: Sensors monitoring your activities and body, notify you if there are abnormal measurements

	Amount per minute
Emails sent	231400000
Cryptocurrency purchased (USD)	90200000
Texts sent	16000000
Searches conducted on Google	5900000
Snaps shared on Snapchat	2430000
Pieces of content shared on Facebook	1700000
Swipes on Tinder	1100000
Hours streamed	1000000
USD spent on Amazon	443000
USD sent on Venmo	437600
Tweets shared on Twitter	347200
Hours spent in Zoom meetings	104600
USD spent on DoorDash	76400

Media
usage in an
internet
minute as of
April 2022

Variety: data in many forms

- A single application may generate/collect many types of data, e.g., types of data are stored in emails
 - Tabular data: attributes like subject, to, from
 - Text (in email body)
 - Image (in attachment)
 - Hyperlinks
- Types of data
 - Relational Data (e.g., Tables)
 - Text Data (e.g., comments)
 - Semi-structured Data (e.g., XML)
 - Graph Data (e.g., social network)
 - What else?

Data in Many Forms

John likes to watch movies.
Mary likes movies too.

Text

Tuple	x_1
1	John also likes to watch football games.
2	John likes to watch movies. Mary likes movies too.



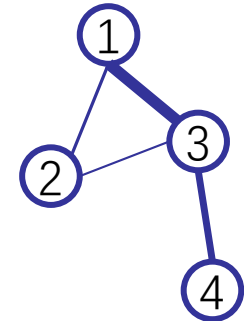
**Signal
(Voice, Audio)**

	signal1	signal2
	13.58	7.24
	12.11	12.50
	13.49	8.66
	11.25	10.98
	14.57	13.75
	13.22	9.02



Image

12	20	22
5	0	5
10	15	25
5	0	5
15	75	17
		5



Graph

A	1	2	3	4
1	0	1.2	4.3	0
2	1.2	0	0.8	0
3	4.3	0.8	0	2.6
4	0	0	2.6	0

E	Node1	Node2	Weight
1	1	2	1.2
2	1	3	4.3
3	2	3	0.8
4	3	4	2.6

Veracity: uncertainty of data

- Is the data accurate?
 - Measurement error
 - Human errors like typos in names/addresses
- Does the data come from a reliable source?
- What if data from different sources are not consistent?



Roll over image to zoom in

Lumia 650 Case, CoverON [HexaGuard Series]
Slim Hybrid Hard Phone Cover Case for
Microsoft Lumia 650 - Black & Black

Visit the CoverON Store
★★★★☆ 5 ratings

Currently unavailable.
We don't know when or if this item will be back in stock.

Color: Black & Black

Material Silicone, Rubber, Polycarbonate

Brand CoverON

Color Black & Black

Compatible Phone Microsoft Lumia 650

Models

About this item

- CoverON [HexaGuard Series] protective hard hybrid case for the Microsoft Lumia 650
- Hard polycarbonate outer layer combined with a soft flexible rubber silicone inner layer
- Enhanced "basketball" grip on the back
- Shock absorbing and protective yet thin and lightweight
- 90 day manufacturer warranty

★★★★★ This case is really good, it looks very nice and doesn't add bulk ...

By [redacted] on January 16, 2016

Color: Black & Black

The case is lightweight and slick. This case is really good, it looks very nice and doesn't add bulk or excessive size to my phone. It is so easy to get in and out of my pocket and the color is nice. It fits snugly and provides good protection. It's so pretty and it seems durable. I receive this product to give honest review.

2 comments 0 of 8 people found this helpful. Was this review helpful to you?

★★★★★ It protects well and feels good on my phone

By [redacted] on January 20, 2016

Color: Black & Black

This case fits my phone well and is very durable. It protects well and feels good on my phone. I received this for free for doing this review.

2 comments 0 of 8 people found this helpful. Was this review helpful to you?

★★★★☆ Full phone protection.

By [redacted] on January 17, 2016

Color: Black & Black

I receive this case yesterday and give this CoverON HexaGuard series hybrid case for the Microsoft Lumia 650 for a family member. This case really protects the phone and this case is an excellent choice for those who want the protection of a 2-layer hybrid case, however it does add some bulkiness to the phone. The title said its slim, but for me it's a little thick but to whom I've given it to, he said it's great. This case has a flexible rubber silicone skin as inner shell with thicker extra padding on the corners. Then a tough polycarbonate outer shell. I got this at a discounted price.

Fake, Paid-For Reviews in Amazon

Applications: Artificial Intelligence with big data

- Artificial Intelligence: the theory and development of computer systems able to perform tasks normally requiring human intelligence

- Before the Age of “big data”

- ELIZA is an early natural language processing computer program created from 1964 to 1967 at MIT
- On May 11, 1997, an IBM computer called IBM Deep Blue beat the world chess champion after a six-game match

```
Welcome to

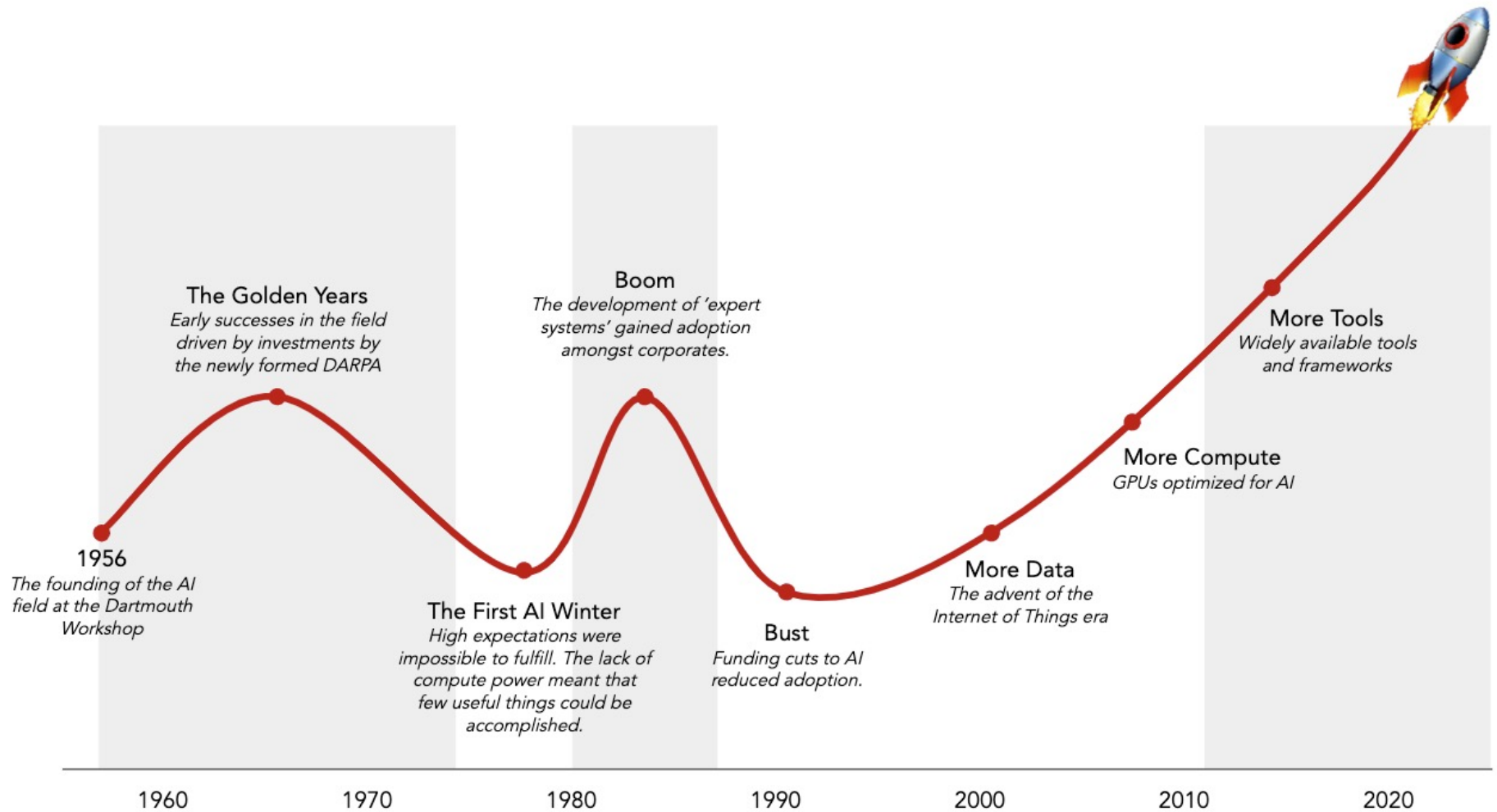
EEEEEE LL      IIII  ZZZZZZ  AAAAA
EE      LL      II    ZZ    AA  AA
EEEEEE LL      II    ZZ    AAAAAA
EE      LL      II    ZZ    AA  AA
EEEEEE LLLLLL  IIII  ZZZZZZ  AA  AA

Eliza is a mock Rogerian psychotherapist.
The original program was described by Joseph Weizenbaum in 1966.
This implementation by Norbert Landsteiner 2005.

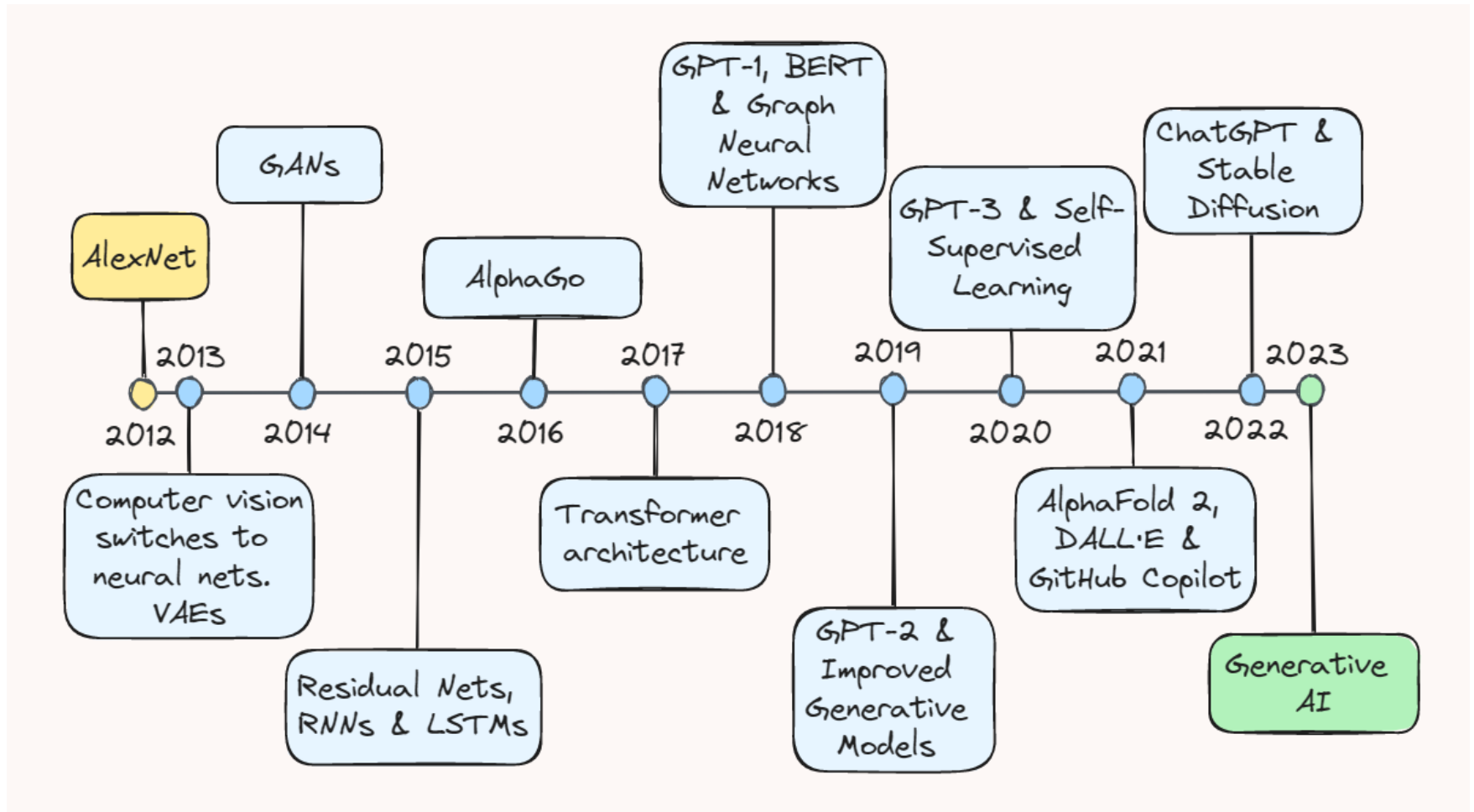
ELIZA: Is something troubling you ?
YOU:   Men are all alike.
ELIZA: What is the connection, do you suppose ?
YOU:   They're always bugging us about something or other.
ELIZA: Can you think of a specific example ?
YOU:   Well, my boyfriend made me come here.
ELIZA: Is it important to you that your boyfriend made you come here ?
YOU:   He says I'm depressed much of the time.
ELIZA: I am sorry to hear that you are depressed.
YOU:   It's true. I am unhappy.
ELIZA: Can you explain what made you unhappy ?
YOU:
```

- Big data has changed AI: “AI Would Be Nothing Without Big Data”
 - “Data is the new oil”

History of AI



Recent breakthroughs in AI



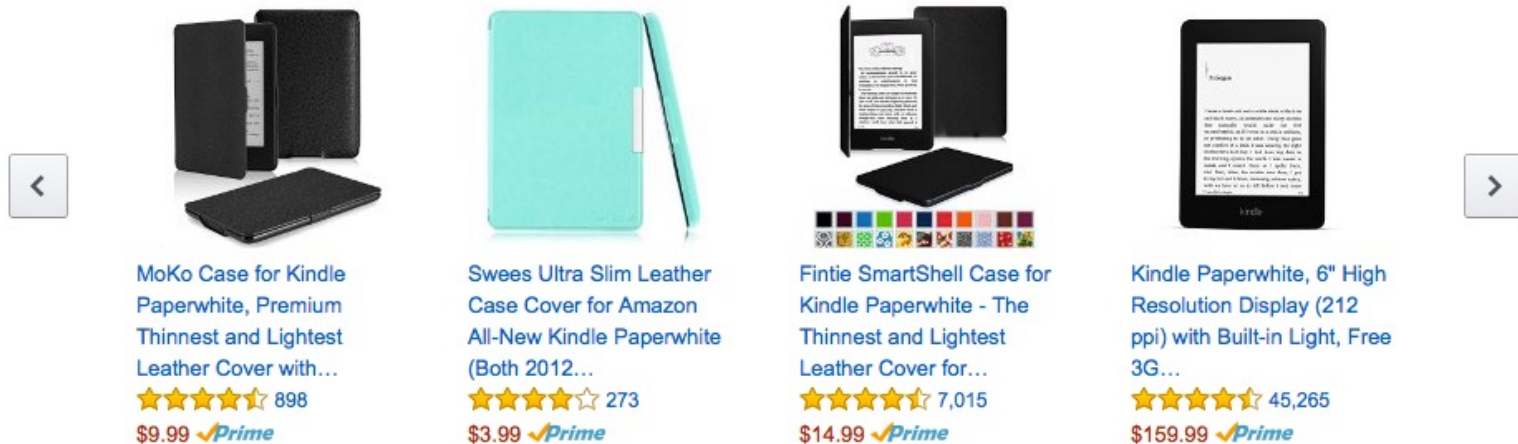
- At the 2017 Future of Go Summit, the Master version of AlphaGo beat Ke Jie, the number one ranked player in the world at the time, in a three-game match, after which AlphaGo was awarded professional 9-dan by the Chinese Weiqi Association

What Big Data Can Help

- Recommendation
 - Ex: Amazon, YouTube, Netflix,
 - What item for what people?
 - How to improve users' satisfaction?

Recommended for You Based on Kindle Paperwhite, 6" High Resolution Display w...

Page 1 of 5

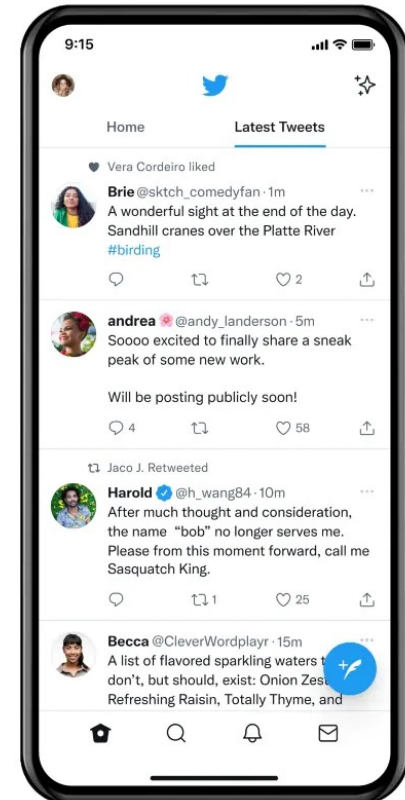
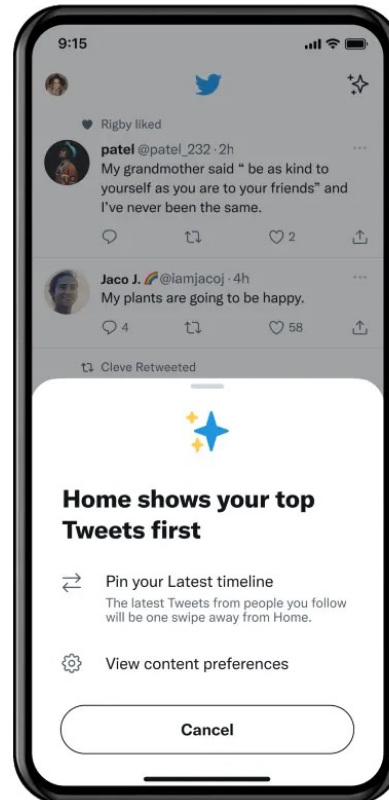
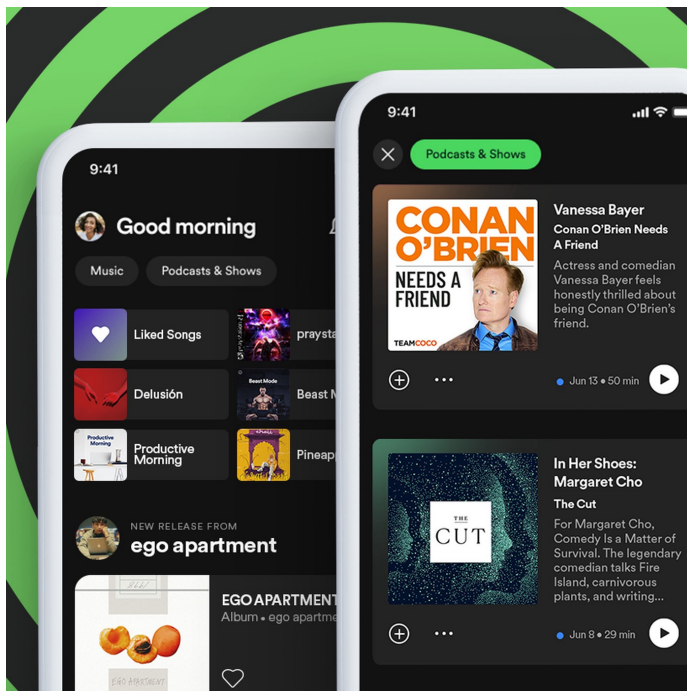


A carousel of product recommendations for Kindle Paperwhite. It features four items: three different cases (MoKo, Swees, and Fintie) and the Kindle Paperwhite device itself. Each item has a title, a star rating, and a price. Navigation arrows are visible on the left and right sides of the carousel.

Product	Rating	Price
MoKo Case for Kindle Paperwhite, Premium Thinnest and Lightest Leather Cover with...	★★★★☆ 898	\$9.99 Prime
Swees Ultra Slim Leather Case Cover for Amazon All-New Kindle Paperwhite (Both 2012...	★★★★☆ 273	\$3.99 Prime
Fintie SmartShell Case for Kindle Paperwhite - The Thinnest and Lightest Leather Cover for...	★★★★☆ 7,015	\$14.99 Prime
Kindle Paperwhite, 6" High Resolution Display (212 ppi) with Built-in Light, Free 3G...	★★★★☆ 45,265	\$159.99 Prime

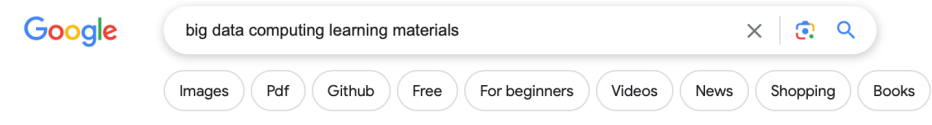
What Big Data Can Help

- More recommendations:
 - News feed
 - Music feed
 - Twitter feed



What Big Data Can Help

- Web search, image search
- Chatbot
- Virtual assistant
- High-frequency trading



About 1,400,000,000 results (0.52 seconds)

 Teach Computing
<https://teachcomputing.org> > key-stage-4 > online-safety

Lesson 3 Big data

In this lesson, learners will be introduced to the concept of **big data** and learn how it is used and why. They will be tasked with identifying what data they ...

People also ask

- What do I need to learn for big data?
- What are the 4 computing resources of big data?
- What are the 3 major components of big data?
- What are the tools and languages used in big data?

Feedback

 Shiksha
<https://www.shiksha.com> > ... > Articles > Technology

Best Online Resources to Learn Big Data in 2023 (Courses ...

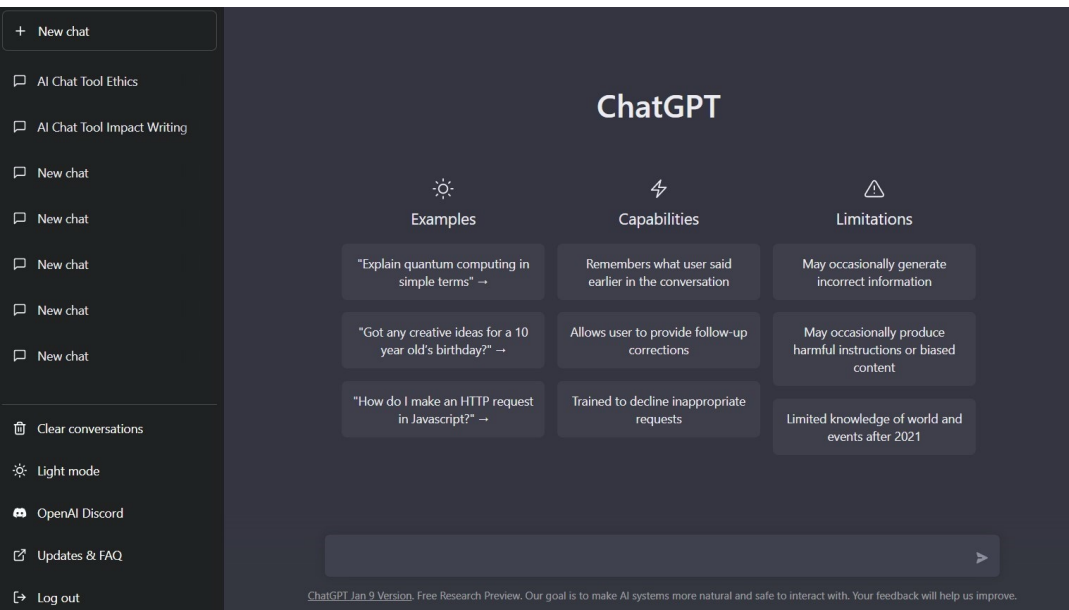
22 Nov 2022 — Best Resources to Learn Big Data · Courses · Podcasts · Books · Big Data Blogs · YouTube Tutorials · GitHub Repository – Awesome Big Data.

Sun, 19 Nov [Big Data Engineer](#)

 Analytics Vidhya
<https://www.analyticsvidhya.com> > resources-big-data

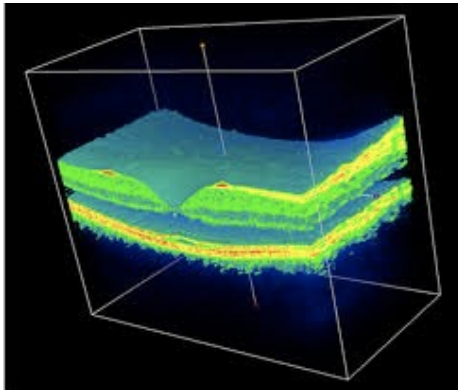
Best Resources to learn Big Data

A comprehensive list of all **Big Data Resources** for data science professionals.



What Big Data Can Help

- Disease Treatment: Joint research between Google DeepMind and Moorfields Eye Hospital
 - Eyecare professionals diagnose eye conditions by using optical coherence tomography (OCT) scans (over 1,000 a day at Moorfields alone)
 - Achieving expert error rate 5.5% comparably to the two best retina specialists (6.7% and 6.8% error rate)

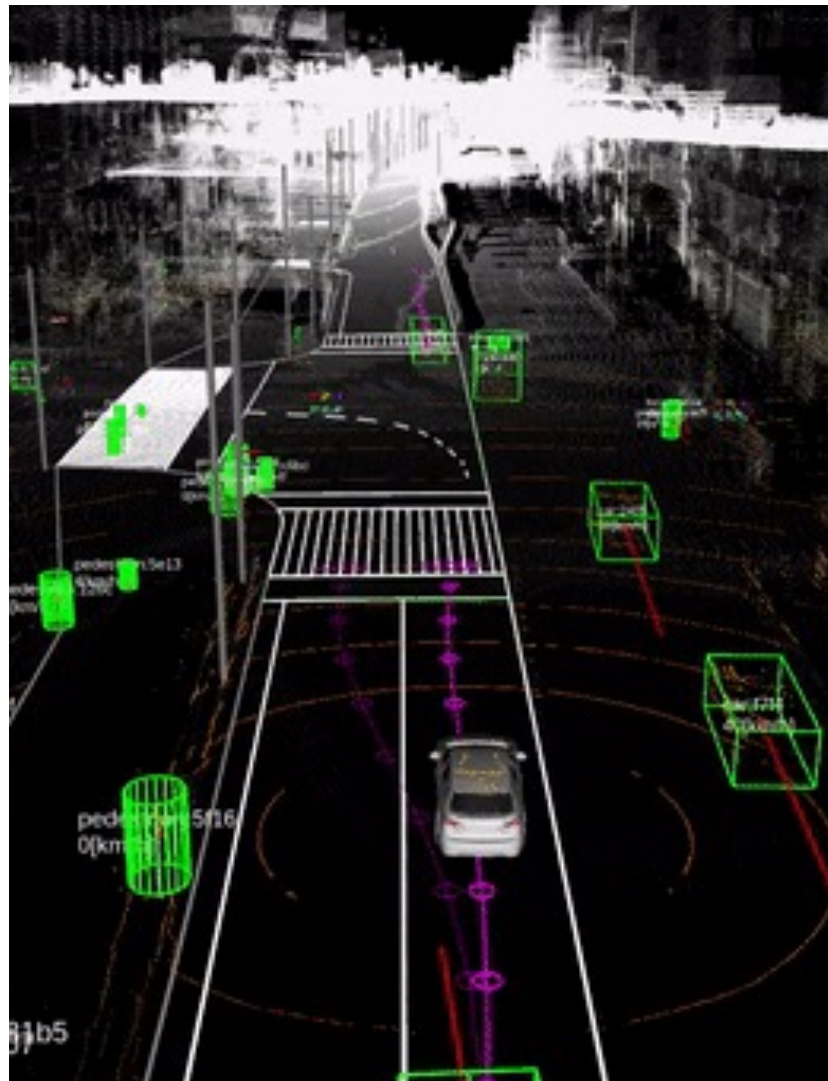


Demo 1: ChatGPT

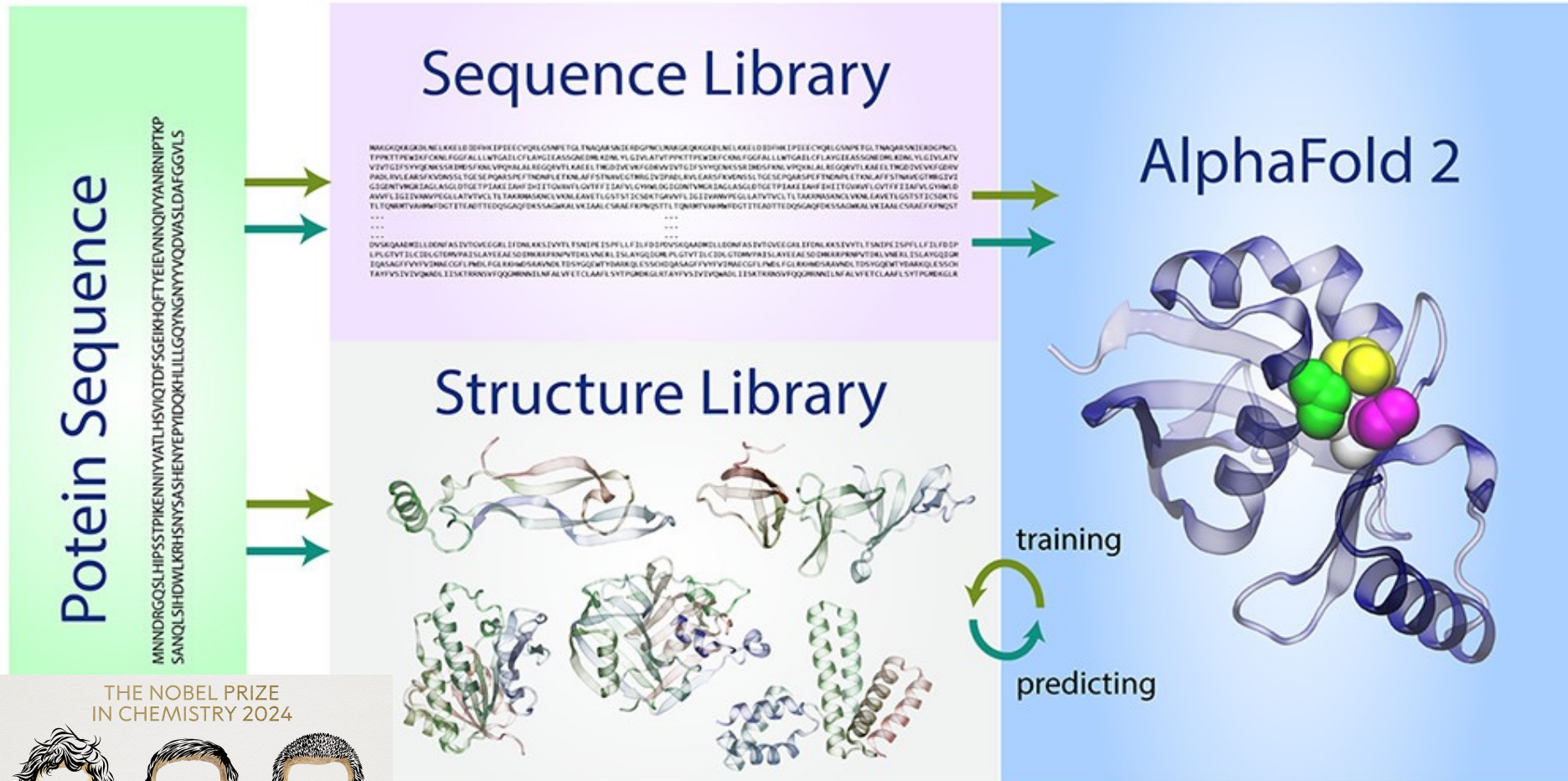
- Write essays:
<https://youtu.be/oLjZva6JvLI>
- Programming:
<https://youtu.be/TIDA6pvjEE0>
<https://youtu.be/B3yuK2XHmvM>
- Conversation:
<https://youtu.be/GYeJC31JcM0>

Demo 2: Autonomous driving

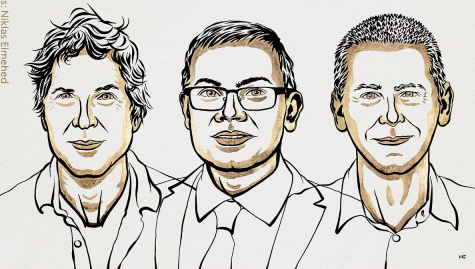
- <https://www.notateslaapp.com/news/1579/musk-live-streams-fsd-v12-and-it-s-too-human-why-that-s-a-problem-video>



Demo 3: AlphaFold



THE NOBEL PRIZE
IN CHEMISTRY 2024



David
Baker

Demis
Hassabis

John M.
Jumper

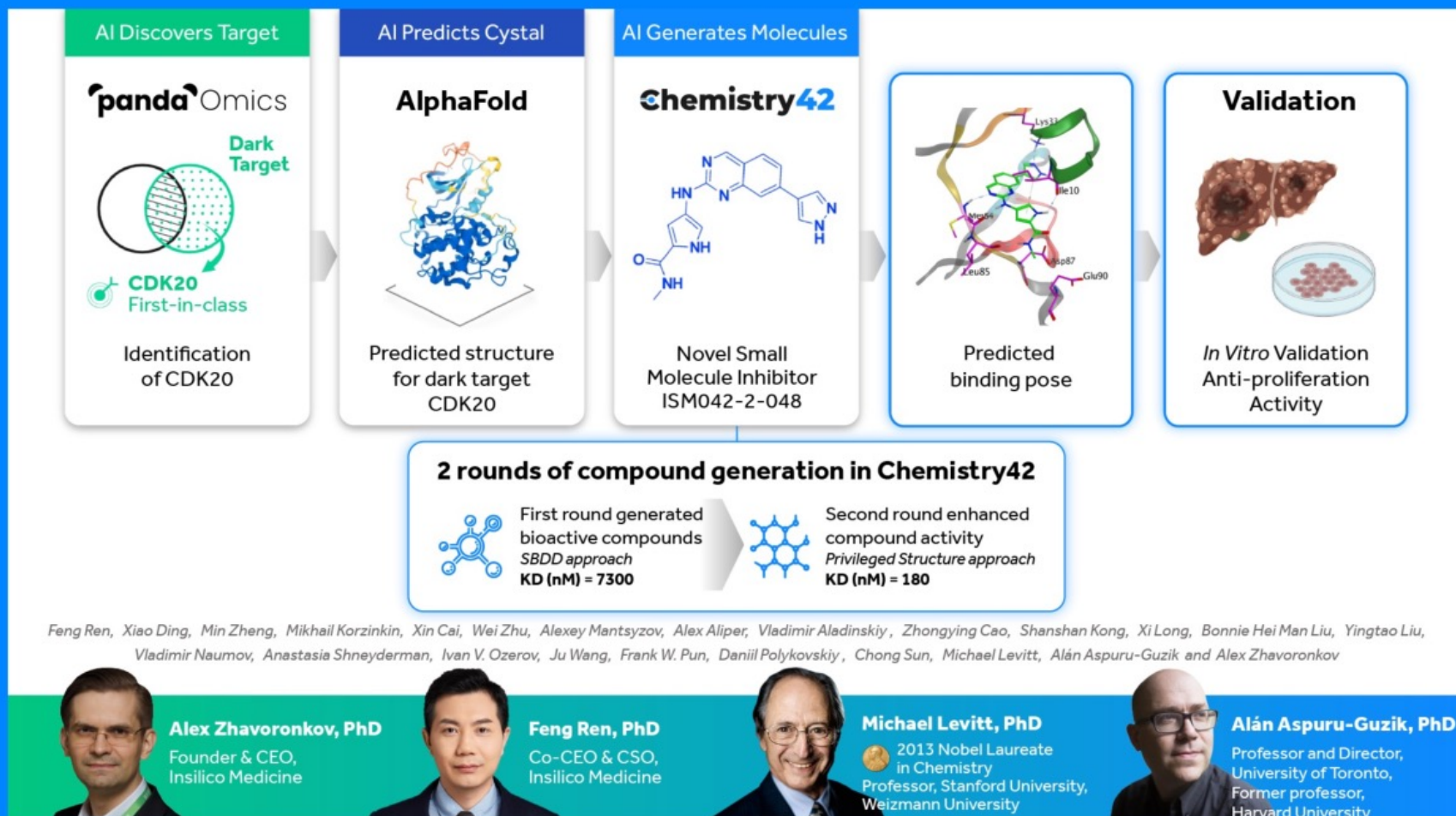
"for computational
protein design"

"for protein structure prediction"

THE ROYAL SWEDISH ACADEMY OF SCIENCES

- A solution to a 50-year-old grand challenge in biology

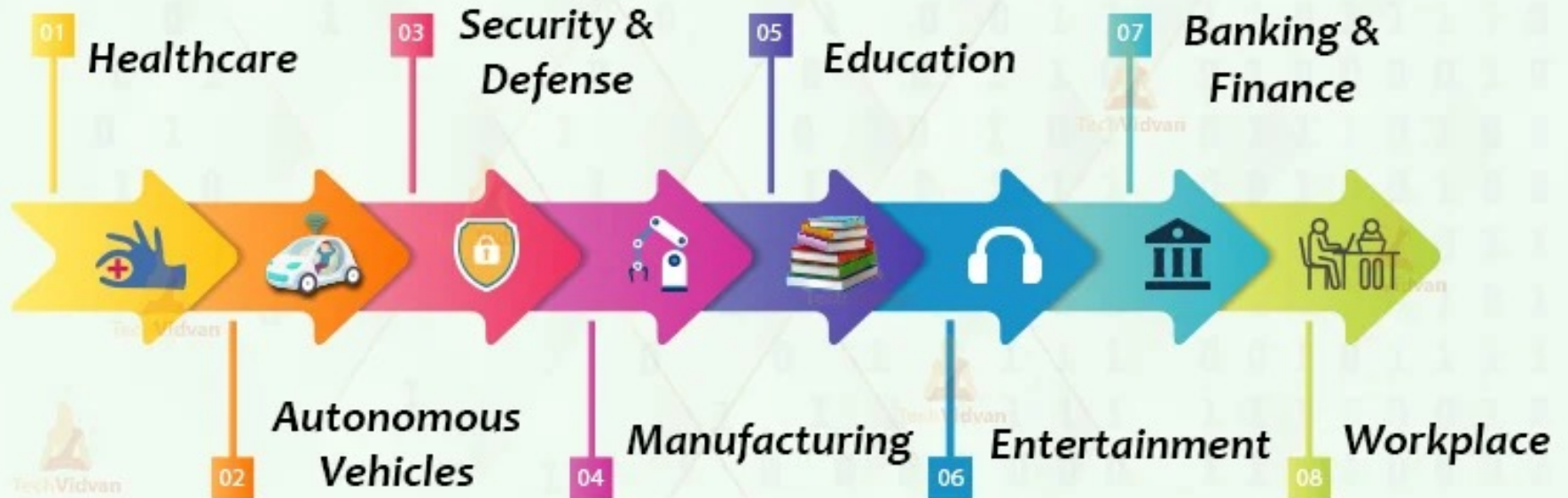
An example in drug design



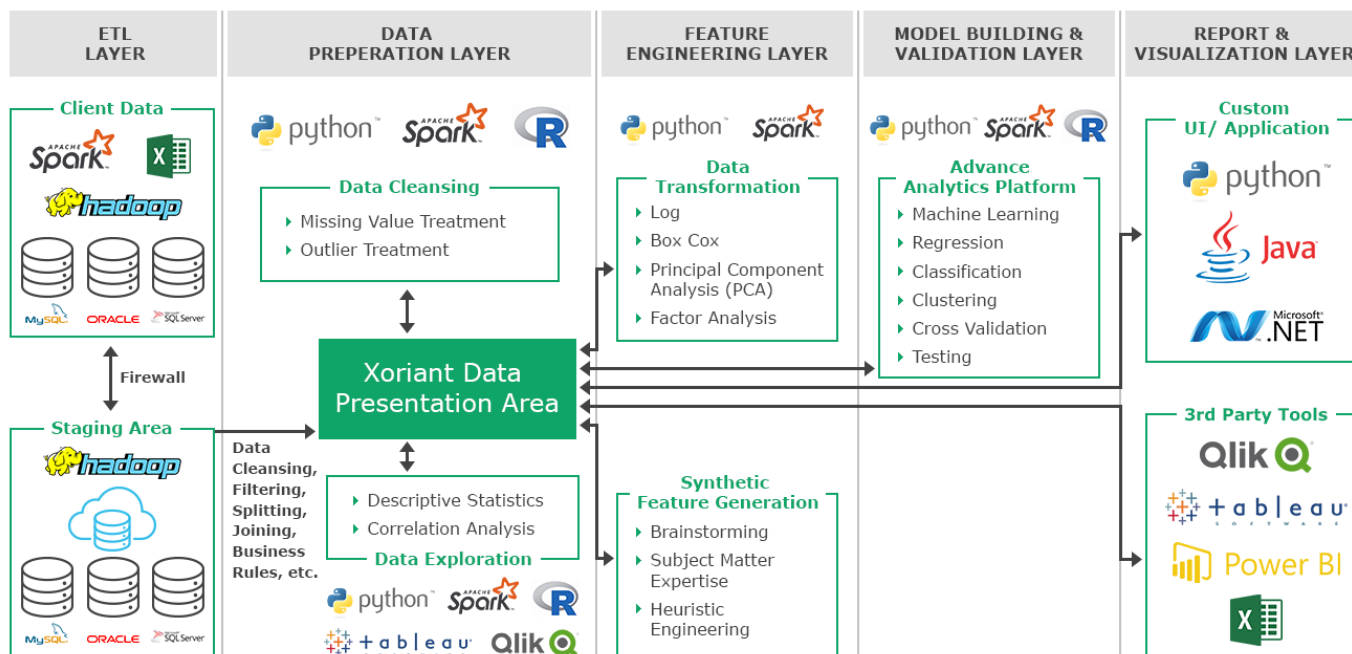
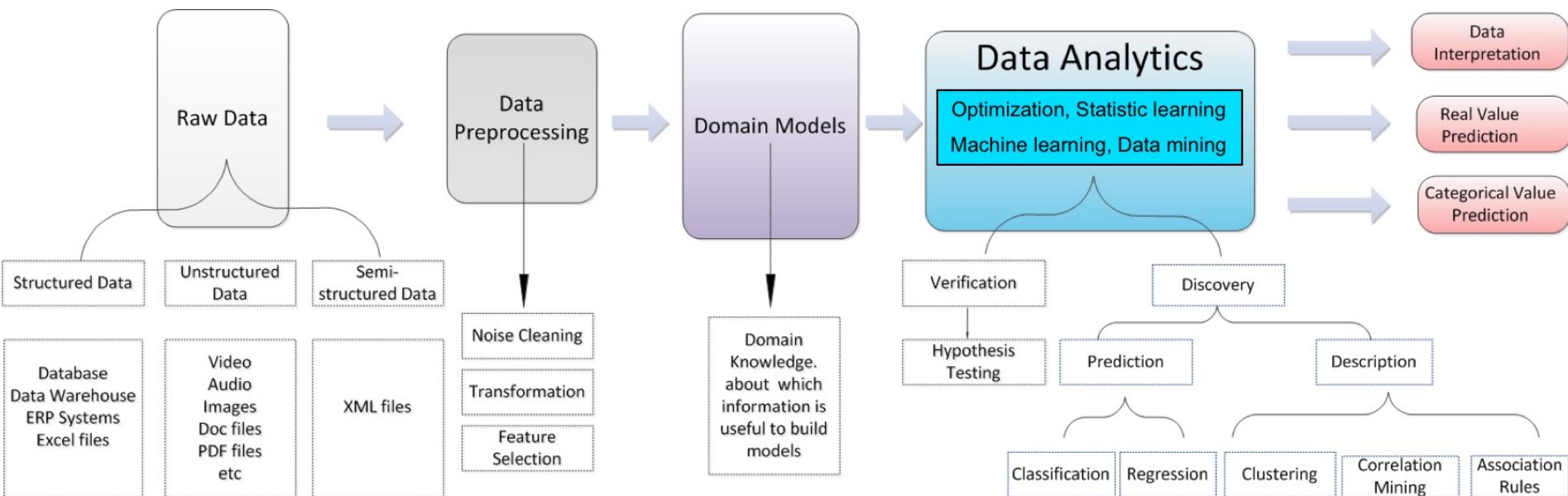
We can do more



Future of AI



Big Data Analysis Procedure



Relations among big data analytics and AI

