

# Introduction: The 3 Pillars

Algorithms, Data and Security  
A.Y. 2024/25

**Valeria Cardellini**

Global Governance, 3rd year  
Major in Science and Technology

## The 3 course pillars

---

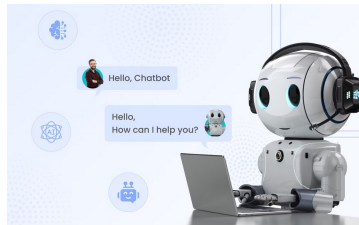
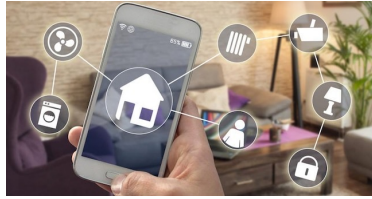
### Algorithms, Data and Security

- Why should we care?
- Technology (digital innovation) is changing and shaping our world

# The IT revolution

---

- Smart living
- Smart cities
- AI and automation
- Future of work
- New experiences



Valeria Cardellini - ADS 2024/25

2

## Why IT sector is so important?

---

- Largest phone companies?
- Most valuable retailer?
- Most popular media companies?
- Largest movie house?
- Largest software vendors?
- Largest taxi company?
- Largest accommodation provider?

Valeria Cardellini - ADS 2024/25

3

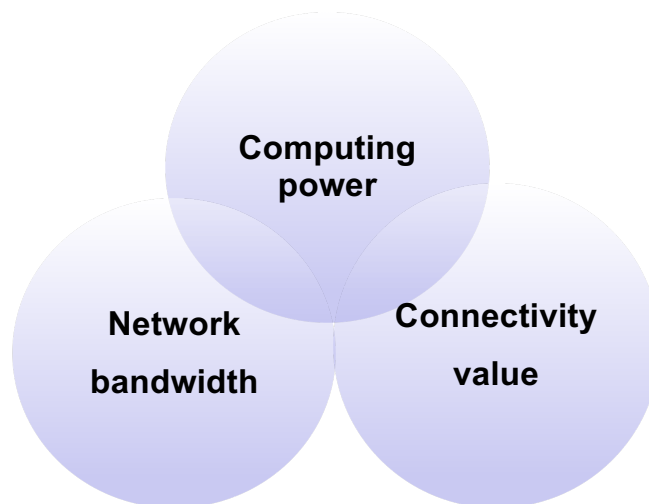
# Why IT sector is so important?

---

- Largest phone companies? *No telco infrastructure (Skype, Telegram, WeChat, WhatsApp)*
- Most valuable retailer? *Carries no stock (Alibaba)*
- Most popular media companies? *Create no content (Facebook, Instagram, TikTok, Youtube)*
- Largest movie house? *Owns no cinema (Netflix)*
- Largest software vendors? *Write no app (Apple, Google)*
- Largest taxi company? *Owns no car (Uber)*
- Largest accommodation provider? *Owns no property (Airbnb)*

# Why is it possible?

---



- Moore's law

*The number of transistors on a computer chip double approximately every 18 months*

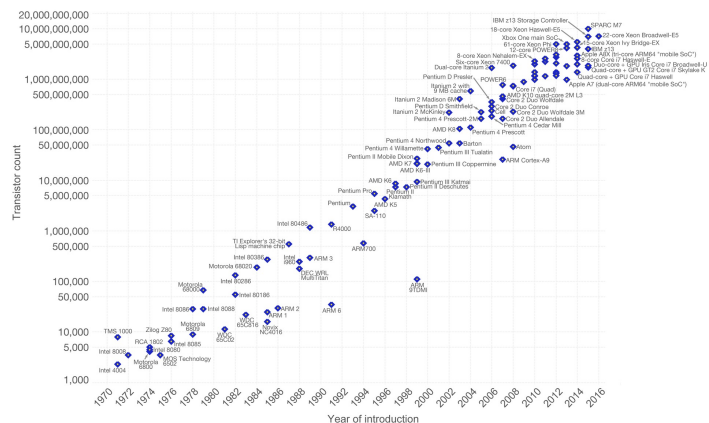
- Began to show signs of slowing in 2015-2017
- But we have multicore computing architectures

**Moore's Law** – The number of transistors on integrated circuit chips (1971-2016)

This advancement is important as other aspects of technological progress – such as processing speed or the price of electronic products – are strongly linked to Moore's law.

- Computers get

- Smaller
- Cheaper
- Power efficient
- Faster



Valeria Cardellini - ADS 2024/25

Data source: Wikipedia ([https://en.wikipedia.org/wiki/Transistor\\_count](https://en.wikipedia.org/wiki/Transistor_count))

The data visualization is available at [OurWorldinData.org](https://ourworldindata.org). There you find more visualizations and research on this topic.

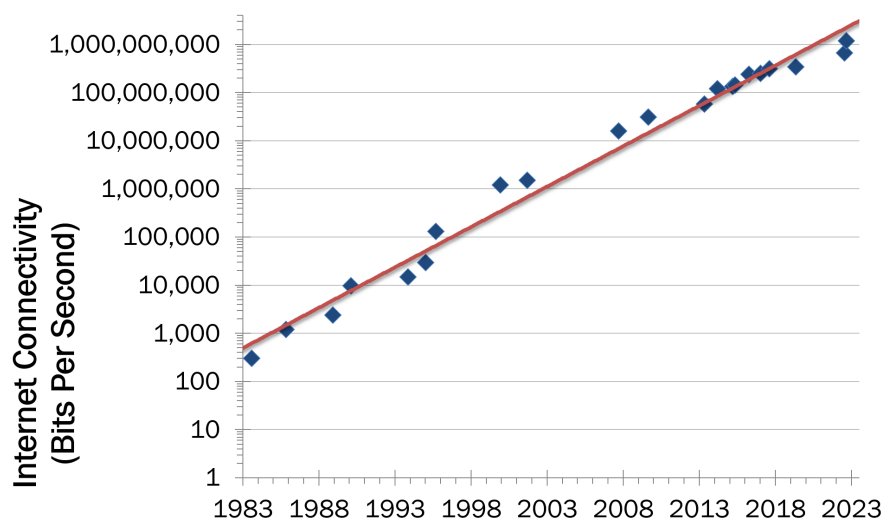
Licensed under [CC-BY-SA](#) by the author Max Roser.

6

- Nielsen's law

*Users' bandwidth grows by 50% per year*

<https://www.nngroup.com/articles/law-of-bandwidth/>



Valeria Cardellini - ADS 2024/25

7



## Network bandwidth

---

- With **5G** we have super-fast low latency Internet everywhere
  - Faster than typical home Internet connection
  - Wireless
- How fast is 5G?
  - 4G tops out at a theoretical 100 megabits per second (Mbps)
  - 5G tops out at 10 *gigabits* per second (Gbps)
    - 100x faster than current 4G (theoretical maximum speed)



Valeria Cardellini - ADS 2024/25

8

## Connectivity value

---

- Metcalfe's law

*The value of a telecommunications network is proportional to the square of the number of connected users of the system*

Networking is *socially* and *economically* interesting

facebook

Google

TikTok

Instagram

Valeria Cardellini - ADS 2024/25

9

# What is an algorithm?

---

- Informal definition
- Sequence of **instructions**, given **step by step**, that could be executed “**easily**” and so that they produce a desired result (**output**)

# What is an algorithm?

---

- How to solve a problem: compute (efficiently) output from input
- Problem: Wish to prepare a chocolate cake
- How to solve this problem? Recipe!

## Input

200g **golden caster sugar**

200g **unsalted butter**, softened plus extra for the tins

4 **large eggs**

200g **self-raising flour**

2 tbsp **cocoa powder**

1 tsp **baking powder**

½ tsp **vanilla extract**

2 tbsp **milk**

Recipe  
and you!

## Output



## A classical algorithm

---

- **Eratosthenes' sieve** (275-195 BC): output all primes  $\leq n$

1. Prepare a list of natural numbers: 2, 3, ..., n
2. Start from  $k = 2$
3. Mark all multiples of  $k$  between  $k^2$  and  $n$  (they cannot be prime!)
4. Replace  $k$  with smallest unmarked number  $> k$
5. If  $(k^2 \leq n)$  go back to step 3
6. Output all unmarked numbers

## Eratosthenes' sieve: example with $n=120$

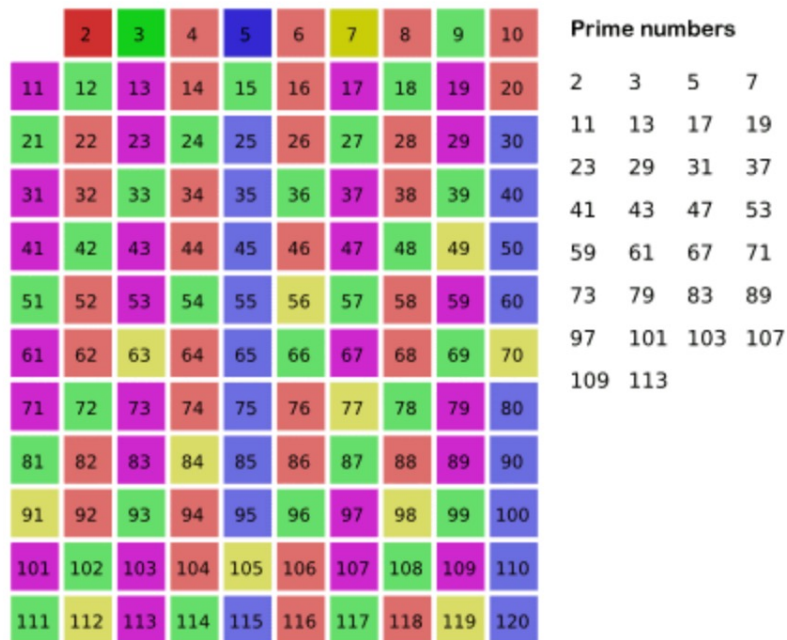
---

	2	3	4	5	6	7	8	9	10	Prime numbers
11	12	13	14	15	16	17	18	19	20	
21	22	23	24	25	26	27	28	29	30	
31	32	33	34	35	36	37	38	39	40	
41	42	43	44	45	46	47	48	49	50	
51	52	53	54	55	56	57	58	59	60	
61	62	63	64	65	66	67	68	69	70	
71	72	73	74	75	76	77	78	79	80	
81	82	83	84	85	86	87	88	89	90	
91	92	93	94	95	96	97	98	99	100	
101	102	103	104	105	106	107	108	109	110	
111	112	113	114	115	116	117	118	119	120	

# Eratosthenes' sieve: example with n=120

- See animation

[https://en.wikipedia.org/wiki/Sieve\\_of\\_Eratosthenes](https://en.wikipedia.org/wiki/Sieve_of_Eratosthenes)



Valeria Cardellini - ADS 2024/25

14

## Other examples

- Count number of objects having distinct values (e.g., b, a, b, c, g, d, d, g)
- Sort list of objects
- Search on Google
  - We will study **PageRank**, invented by Google's founders to rank web pages in search results
  - Now used in bibliometrics, social network analysis, and for link prediction and recommendation

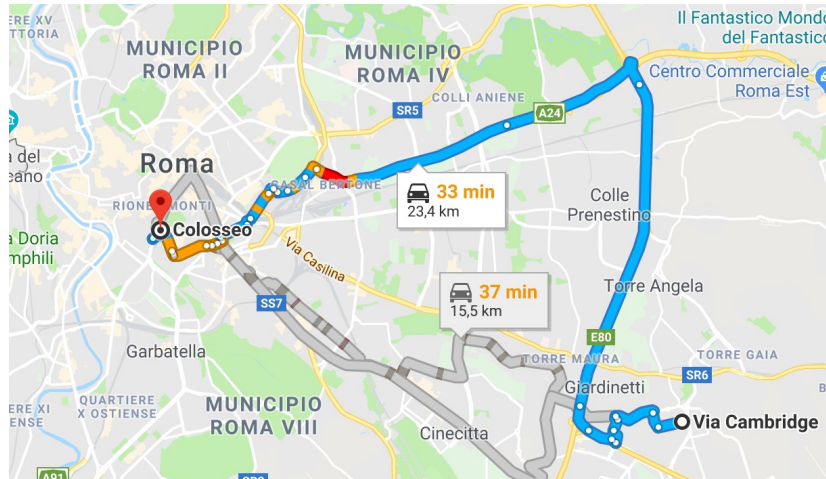


Cerca con Google

Mi sento fortunato

## Other examples

- Find the shortest distance: Google Maps uses the **shortest path algorithm** invented by Dijkstra



## Other examples

- AlphaGo** beats the world's top Go player
  - Go: sophisticated board game, more complex than chess for computers
  - AlphaGo: computer program that plays Go
    - Its algorithm finds the moves based on knowledge previously “learned” by **machine learning**, specifically by an artificial neural network



<https://www.nytimes.com/2017/05/23/business/google-deepmind-alphago-go-champion-defeat.html>

# Artificial neural network

- **Artificial neural network**: computing system inspired by biological neural networks that constitute animal brains
  - It “learns” to perform tasks by considering examples, generally without being programmed with task-specific rules
    - E.g., in image classification, a neural network might learn to identify images that contain cats or dogs by analyzing a set of images labeled as “cat” or “dog” and using the results to identify cats or dogs in new images



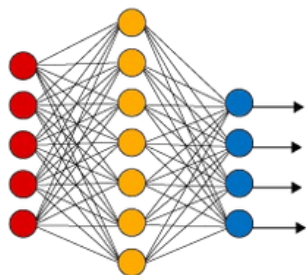
Valeria Cardellini - ADS 2024/25

18

# Artificial neural network

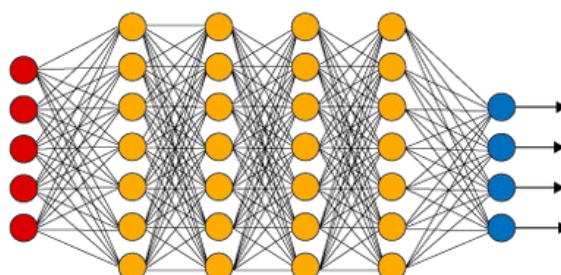
- Neural network works by making links between massive numbers of nodes
  - Hidden layers perform nonlinear transformations of inputs entered into the network

Simple Neural Network



● Input Layer

Deep Learning Neural Network



● Hidden Layer

● Output Layer

Valeria Cardellini - ADS 2024/25

19



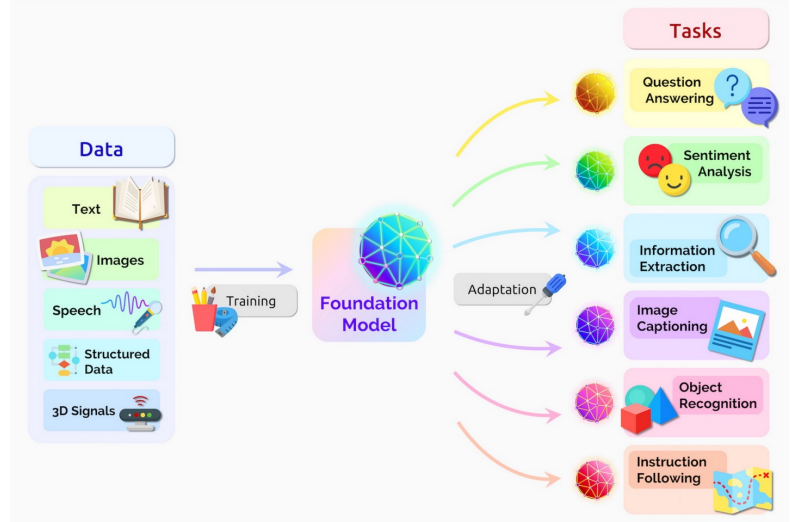
# Transformer model

- A neural network that learns context and thus meaning by tracking relationships in sequential data like the words in this sentence

Aka **foundation model**:

"The sheer scale and scope of foundation models over the last few years have stretched our imagination of what is possible"

[crfm.stanford.edu/report.html](https://crfm.stanford.edu/report.html)



# Algorithms and Big Data

- Algorithms more important today because of Big Data
  - The Economist, 2010: [The Data Deluge](#)
  - NYT in 2012: [The Age of Big Data](#)



# Why Big Data?

How much data is created every single minute of the day?



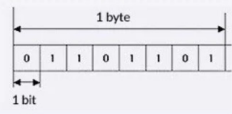
Global Internet population in Jan. 2025: 5.56 billion (67.9% of world population)  
1 billion in 2005



Source: <https://www.domo.com/data-never-sleeps>

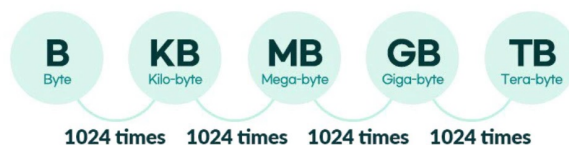
## Demystifying data units

- Bit and byte

Bit (binary digit, bit)	Byte
Measurement unit that can only have two values, 0 and 1	Unit that indicates the amount of data, consisting of 8 bits
 <b>0 OFF FALSE</b>  <b>1 ON TRUE</b>	

- Each byte unit is 1,024 times the previous

= Data Unit =



Unit	Definition	Storage space size
Bit	0 or 1	Yes/No
1 Byte	8 bit	Alphabets and one number
1 kilobyte (KB)	1,024 Byte	A few paragraphs
1 megabyte (MB)	1,024 KB	One minute-long MP3 song
1 gigabyte (GB)	1,024 MB	30 minute-long HD movie
1 terabyte (TB)	1,024 GB	About 200 FHD movies

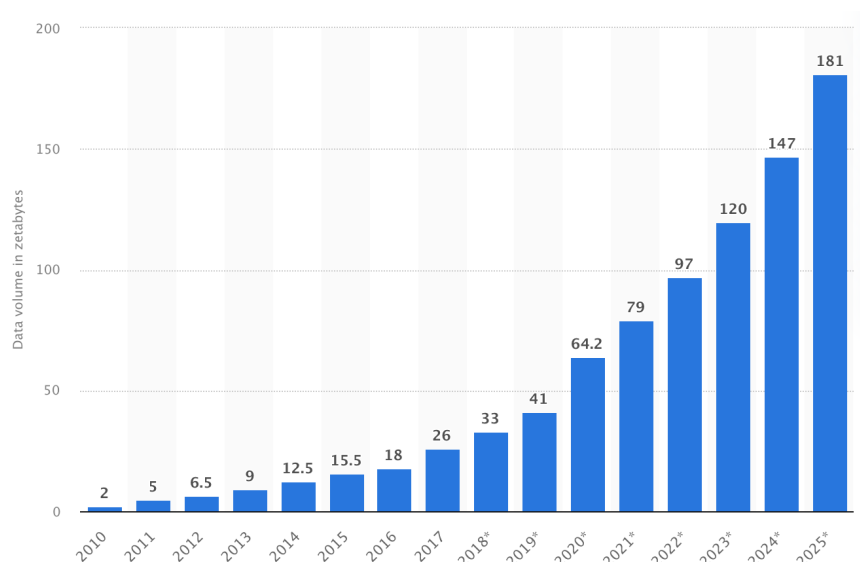


# How much data?

- Big data volume: from Terabytes to Zettabytes
  - How big is a Zettabyte?
  - **1 ZB** =  $2^{70}$  B =  $(2^{10})^7$  B  $\approx (10^3)^7$  B =  **$10^{21}$  B**
    - Remember that  $2^{10} = 1024 \approx 10^3$
    - Also, **1 ZB** =  $2^{70}$  B =  $2^{40} 2^{30}$  B =  **$2^{40}$  GB**, being  $2^{30}$  B = 1 GB
- 147 Zettabytes of data generated by 2024
  - 147 Zettabytes ( $147 \times 2^{70} \approx 147 \times 10^{21}$ ) ...
  - $\approx 147,000$  Exabytes ( $147,000 \times 10^{18}$ ) ...
  - $\approx 147,000,000$  Petabytes ( $147,000,000 \times 10^{15}$ ) ...
  - $\approx 147,000,000,000$  Terabytes ( $147,000,000,000 \times 10^{12}$ ) ...
  - $\approx 147,000,000,000,000$  Gigabytes ( $147,000,000,000,000 \times 10^9$ ) ...
  - $\approx 147,000,000,000,000,000,000,000$  bytes!

# How much data?

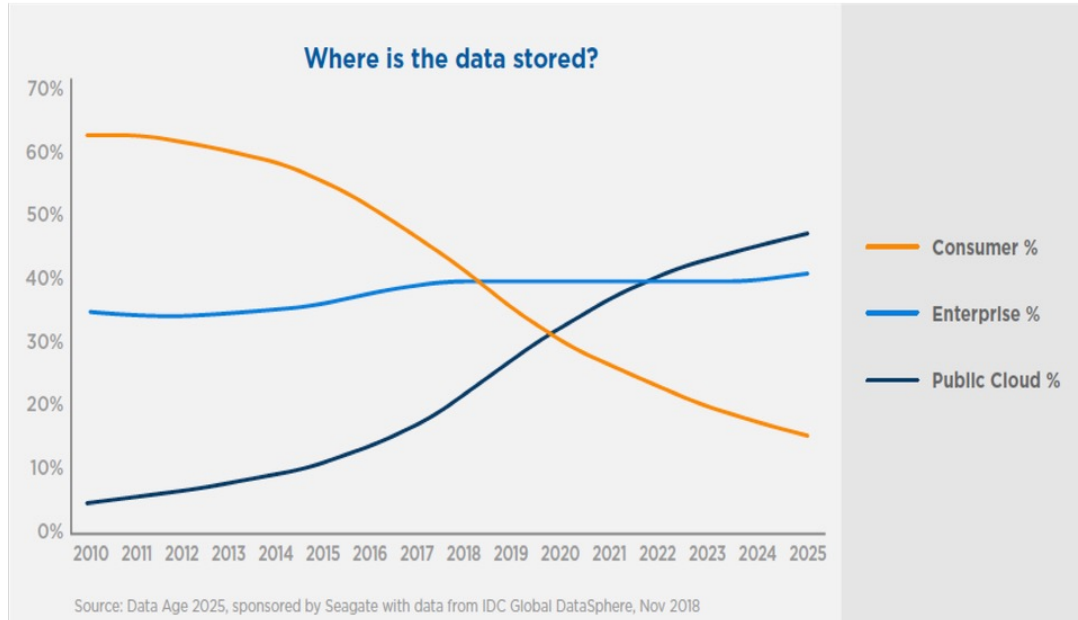
- Recent explosion in data volume
  - In 2013: 90% of all the data in the world was generated over the last two years
  - 74x growth from 2010 to 2024



# Where is data stored?

---

- Data is increasingly stored in Cloud servers
- Where are Cloud servers located?



Valeria Cardellini - ADS 2024/25

26

## Big data statistics and economic impact

---

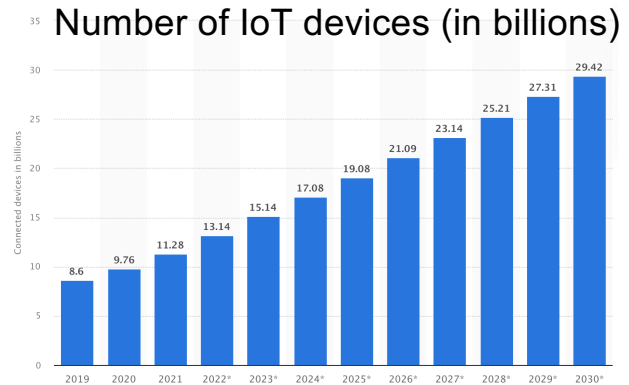
- Every person generated roughly 16 TB daily
- On average, people spend 6.5 hours on Internet every day
- Big data and analytics software and cloud services market reached \$307 billion in 2023 and is expected to reach \$745 billion by 2031
- 91% of organizations are investing in Big Data and AI
- Using Big Data and machine learning algorithms, Netflix saved \$1 billion per year on customer retention

Valeria Cardellini - ADS 2024/25

27

# Big data driving factors

- Big Data is growing fast
  - Smartphones
  - Social networks
  - Internet of Things (IoT)

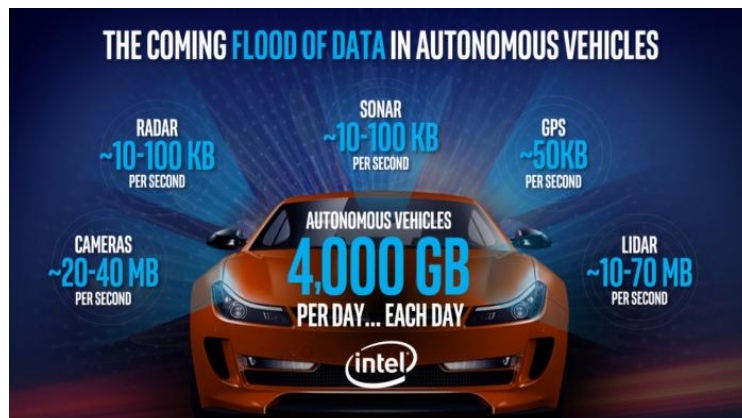


- Exponential growth in unstructured data



## How Big? IoT impact

- IoT is everywhere and largely contributes to increase Big Data challenges
  - Proliferation of data sources: by 2023 over 15 billion IoT devices installed worldwide
- Example: self-driving cars
  - Just one autonomous car can generate up to 4 TB of data/day



# Big Data definitions

---

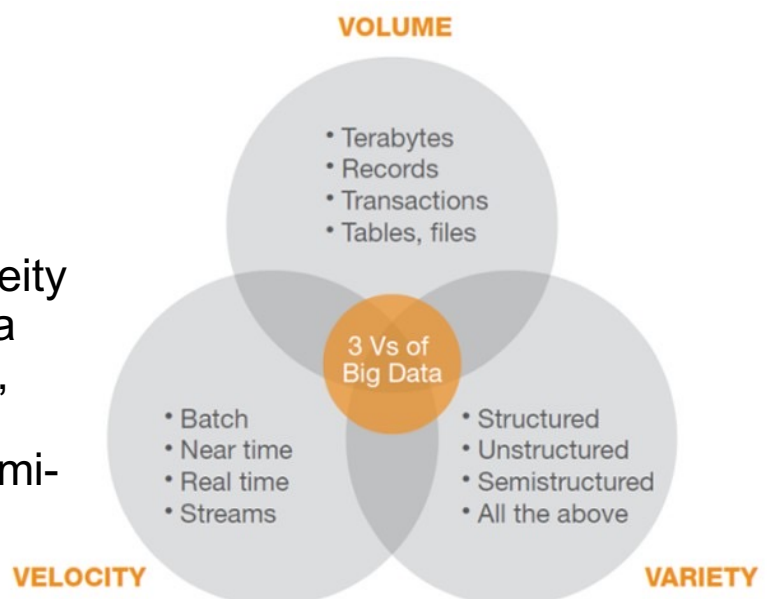
## Different definitions

- “Big data refers to data sets whose size is **beyond** the ability of typical database software tools to capture, store, manage and analyze.” *The McKinsey Global Institute, 2012*
- “Big data is a field that treats ways to analyze, systematically extract information from, or otherwise deal with data sets that are **too large** or **complex** to be dealt with by traditional data-processing application software.” *Wikipedia, 2020*
- “Big data is mostly about taking numbers and using those numbers to **make predictions about the future**. The bigger the data set you have, the more accurate the predictions about the future will be.” *Anthony Goldbloom, Kaggle’s founder*

## 3V model for Big Data

---

1. **Volume**: data size challenging to store and process (how to index, retrieve)
  2. **Variety**: data heterogeneity because of different data types (text, audio, video, record) and degree of structure (structured, semi-structured, unstructured data)
  3. **Velocity**: data generation rate and analysis rate
- Defined in 2001 by D. Laney



# The extended (3+n)V model

---

4. **Value**: Big data can generate huge competitive advantages
  - “Big data technologies describe a new generation of technologies and architectures, designed to economically extract value from very large volumes of a wide variety of data, by enabling high-velocity capture, discovery, and/or analysis.” (IDC, 2011)
  - “The bigger the data set you have, the more accurate the predictions about the future will be” (A. Goldbloom)
5. **Veracity**: uncertainty of accuracy and authenticity of data
6. **Variability**: data flows can be highly inconsistent, with periodic peaks
7. **Visualization**

## Big Data visualization

---

- Presentation of data in a pictorial and graphical format
- Why? Our brain processes images 60,000x faster than text
- Example



## The downside

---

- Every day more data to process and store
- How much energy do we consume?
  - In 2022, data centers that host popular services, including Amazon's cloud services and Google's search engine, used about 1% to 1.3% of world's current electricity use
  - Cryptocurrency mining used another 0.4%
  - By 2027 AI servers could use between 85 to 134 terawatt hours annually
    - That's similar to what Argentina, the Netherlands and Sweden each use in a year, and is about 0.5% of world electricity
- Is it sustainable?

[A.I. Could Soon Need as Much Electricity as an Entire Country](#), NYT, Oct. 2023

## Data structures

---

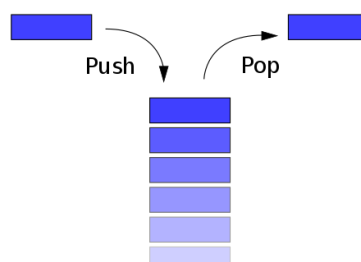
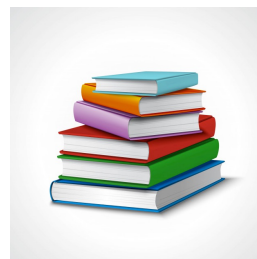
- Structure to organize data
- Support some operations
- Example: **stack**

Supports 3 basic operations:

Insert ("push")

Delete ("pop")

Display content of top item ("peek")

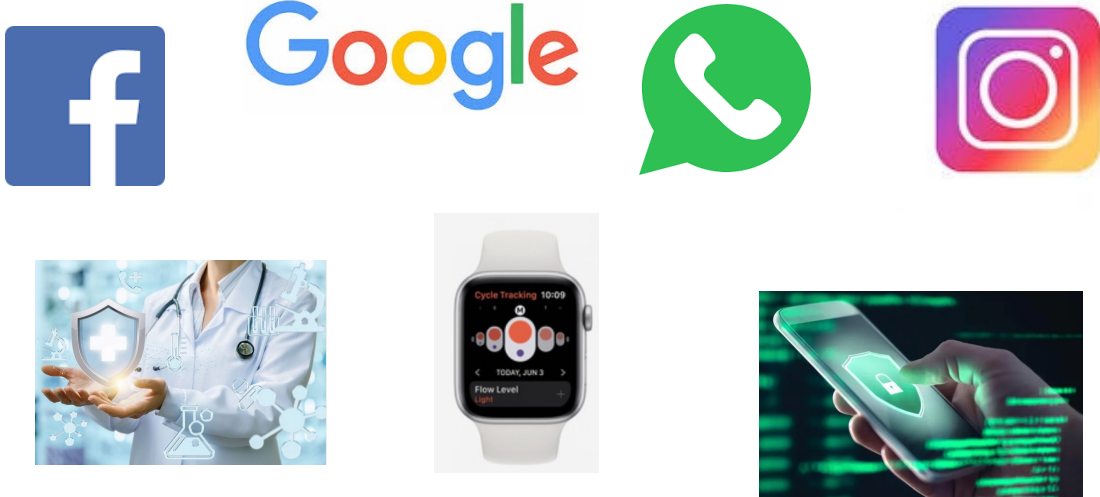




# What about security?

---

- IT and data are everywhere



## IT and data everywhere

---

- New cybersecurity threats



### Self-driving cars

### IoT wearables



# Attacking Machine Learning algorithms

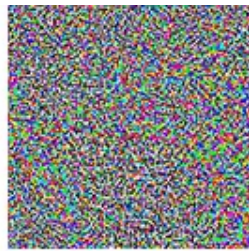
- Adversarial examples are inputs to a machine learning model (e.g., neural network) that an attacker has intentionally designed to cause the model to make a mistake
  - Like optical illusion
  - Changing only 0.04% of the pixel values in the input image (undetected by human eyes), the neural network changes its solution from the correct classification “panda” to an incorrect “gibbon” classification



“panda”

57.7% confidence

+  $\epsilon$



=



“gibbon”

99.3% confidence

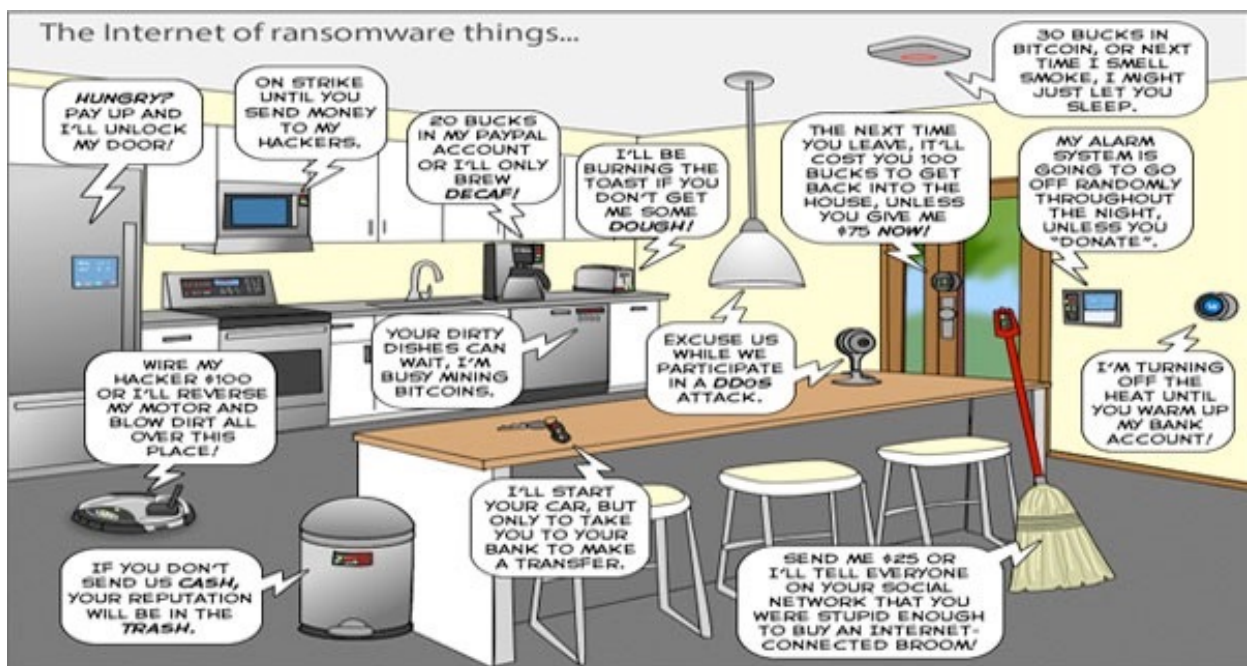


This is a gibbon!

Valeria Cardellini - ADS 2024/25

38

## Internet of Things and ransomware



**Ransomware:** kind of malware that withholds some digital assets from victims and asks for payment for the assets' release

Valeria Cardellini - ADS 2024/25

39