







BIG DATA

a) Definition and breakdown

The six Vs of big data

Big data is a collection of data from various sources, often characterized by what's become known as the 3Vs: *volume*, *variety* and *velocity*. Over time, other Vs have been added to descriptions of big data:

VOLUME	VARIETY	VELOCITY	VERACITY	VALUE	VARIABILITY
The amount of data from myriad sources.	The types of data: structured, semi-structured, unstructured.	The speed at which big data is generated.	The degree to which big data can be trusted.	The business value of the data collected.	The ways in which the big data can be used and formatted.
					

ICONS: ALEXANDZ/ADOBE STOCK

©2018 TECHTARGET. ALL RIGHTS RESERVED 

Source: <https://www.techtarget.com/searchdatamanagement/definition/big-data>

Big data refers to collected data sets that are so large and complex that they require new technologies, such as artificial intelligence, to process. The data comes from many different sources. Often, they are of the same type, for example, GPS data from millions of mobile phones is used to mitigate traffic jams; but it can also be a combination, such as health records and patients' app use. Technology enables this data to be collected very fast, in near real time, and get analysed to get new insights.

Big data can be produced by people: in mobile apps, on the web, including social media and commercial transactions, e-government records etc.

It can also be generated by machines and collected through sensors in objects linked to the Internet of Things, including smart cars, factories, GPS satellites and satellites collecting weather data etc.

Source: HP European Parliament

b) The potential of big data for the economy

6 big data benefits for businesses

1. Better customer insight
2. Improved operations
3. More insightful market intelligence
4. Agile supply chain management
5. Data-driven innovation
6. Smarter recommendations and targeting



ART: VISUAL GENERATION/ADOBE STOCK
©2021 TECHTARGET. ALL RIGHTS RESERVED

Companies use big data in their systems to:

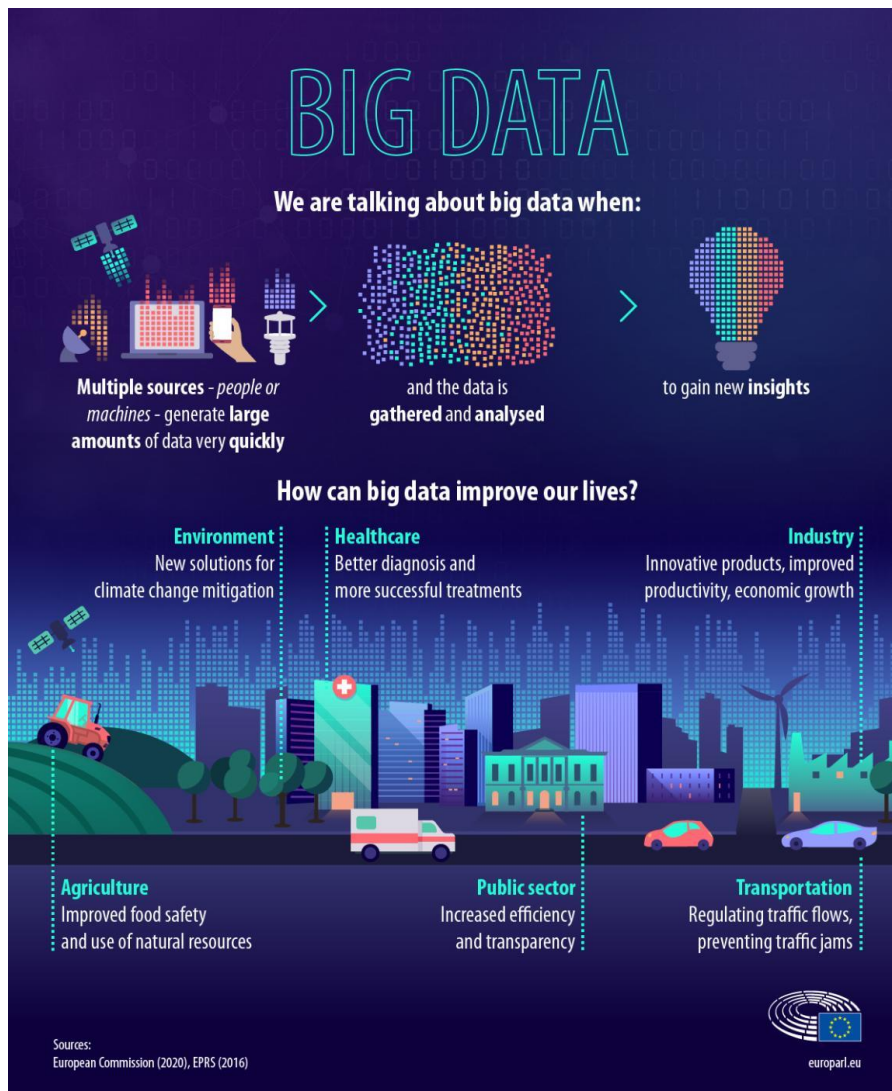
- improve the efficiency of operations
- provide better customer service
- create personalized marketing campaigns (microtargeting)
- take other actions that can increase revenue and profits

Businesses that use it effectively hold a potential competitive advantage over those that don't because they're able to make faster and more informed business decisions.

European Parliament:

“Big data enables companies to innovate, either through better analysing people's needs and wants or by offering entirely new products. While personal data is central for the operation of apps and platforms that have become an important part of our lives and economy, better exploitation of industrial data could bring a new wave of innovation in the EU. Data can also boost productivity and help cut costs, for example by predicting sales or maintenance in smart factories.”

c) The potential of big data for society



Environment

Satellite data can improve research and help the EU reduce greenhouse emissions and help in preventing and responding to natural disasters such as wildfires. Improving production efficiency in the industry should lower emissions and waste.

Healthcare

The analysis of large clinical data sets - for example: anonymised health records or data entered by patients in apps - can enable better diagnostics, treatment and development of medicine, while reducing costs.

Agriculture

Farmers can use data from satellites and sensors to better use resources such as water or sunlight and adapt crops to changing circumstances.

Public sector

Data and advanced analytics can increase the efficiency and effectiveness of public services, to improve transparency and offer better-tailored services.

Transportation

Big data collected from GPS and social media can help mitigate traffic jams. Better regulation of traffic flows also contributes to savings in time and fuel and lowers CO2 emissions.

d) The value of data

Consider the following facts:

- There is a market for data, just like there is a market for fish or mobile phones.
- Data is considered to be a new form of currency in our digital age.
- Some people claim that data has surpassed oil in value.

As mentioned before, big data can only be effective and efficient if large amounts of data are collected quickly. In other words, whoever relies on data for their business or political goals, needs lots of it and fast. And this creates several problems.

e) The problems with big data

Big data certainly has the potential to benefit our society. However, there is also a downside to this new data market. Here are some of the most pressing issues.

- **Data security**

Who is allowed to collect your data? When? How?

Is there legislation which regulates the collection of data?

Are people aware that their data is being collected? If not, can they sue?

What about the protection of our basic rights to privacy? Are ethical concerns justified?

Where is your data? Who has it?

- **Cybercrime**

Cybercriminals can steal your data

Cybercrime has consequences on real people and their real lives

Your own personal data can be used against you.

- **Quality of data**

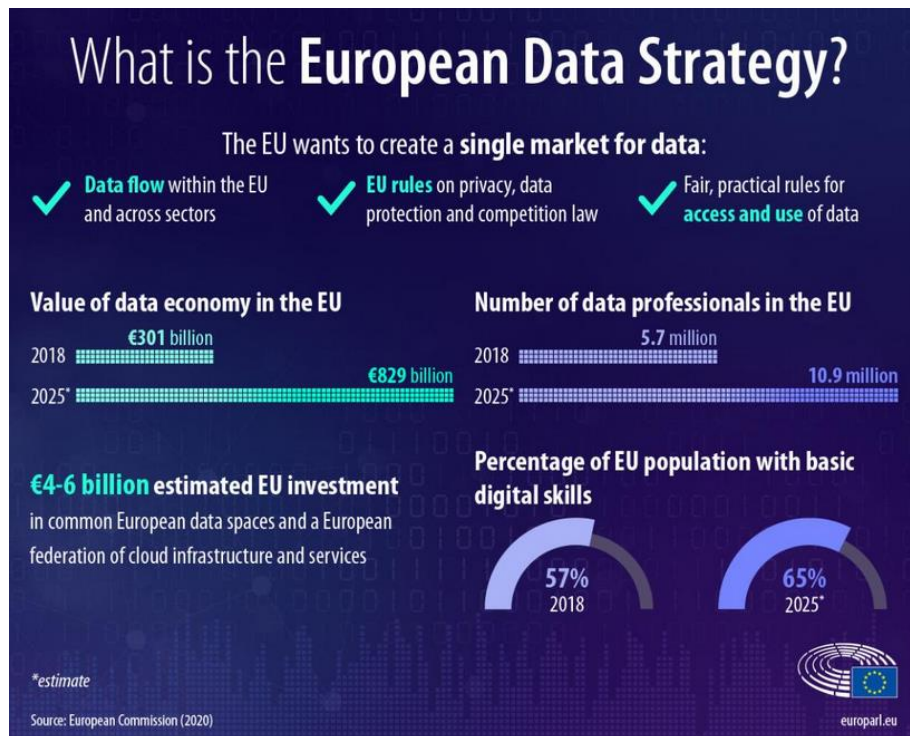
How reliable and relevant is the data that was gathered?

Can entire AI models and algorithms based on big data be trusted?

Should they be trusted?



f) Possible solutions



The solution to the data security challenge of big data is to implement robust security measures, such as data encryption, access controls, and data masking, throughout the data lifecycle. Additionally, organizations can use data security solutions like firewalls, intrusion detection systems, and data loss prevention tools.



Source: LinkedIn

Study Check:

SC1: Are you familiar with the saying, “If you don’t pay for a product, you are the product.” Interpret the saying in the context of big data.

SC2: Read “*The Machine that won the War*” by Asimov in the context of Big Data.

SC3: Be able to write at least 3 essay-like arguments in favour of or against Big Data.

SC4: Be able to write an entire essay on the advantages and disadvantages of Big Data.

g) Case studies

CONCEPT: divide and conquer or divide and rule

a strategy by which someone remains in power by making sure that the people under their control quarrel among themselves and so cannot unite to achieve their aims and overthrow their leader

ONE OF THE BIGGEST CONCERNS IN CURRENT SOCIETY: DIVISION

- *we constantly find ourselves in a divisive binary narrative*
- *binary means: either black or white – no nuance, nothing in between, two options only*

Examples:

Where do you stand?

- | | | |
|----------------------|-----------|----------------------|
| <i>Electric cars</i> | <i>VS</i> | <i>gas cars</i> |
| <i>vegan</i> | <i>VS</i> | <i>carnivore</i> |
| <i>pro-Israel</i> | <i>VS</i> | <i>pro-Palestine</i> |
| <i>Republicans</i> | <i>VS</i> | <i>Democrats</i> |
| <i>pro-gun</i> | <i>VS</i> | <i>anti-gun</i> |
| <i>pro-choice</i> | <i>VS</i> | <i>pro-life</i> |



Downside of binary narrative:

- *Complex discussions become reduced to two simplistic options (Israel VS Palestine)*
- *Two extreme positions destroy our ability to include many options (energy)*
- *By limiting everything to two extreme camps => division, hatred, cynicism => no solutions*

Basically:

- no progress, no evolution, no sensible discussions
- emotional response VS reasonable response

What does this have to do with big data?

The idea of “divide and conquer” has been weaponised in the digital landscape.

Weaponised? How?

Example 1: Cambridge Analytica => example => behaviour change

<https://www.youtube.com/watch?v=omc-5zi70M0>

Basically:

The emotional pulse and the dynamics of entire nations can be controlled, manipulated and triggered from a computer in an office far away.

(Brexit, Le Pen...)

If you are interested in this topic, watch "The Great Hack" on Netflix.

Example 2: Cambridge Analytica: Whistleblower => how data mining works

<https://www.youtube.com/watch?v=FXdYSQ6nu-M>

If you are interested in this topic, watch "The Great Hack" on Netflix.

Example 3: Russian Troll Factory => example => division = strategy

<https://www.youtube.com/watch?v=Vw6Jy5MeLZo>

https://www.youtube.com/watch?v=f8te_Oq93hk

Basically:

Getting news from social media platforms is a recipe for disaster. Micro-targeting works.

If you are interested in this topic, listen to the Podcast "Troll Army – Russlands Krieg im Internet".

Data Footprint of Humans



44

zettabytes

Projected volume of
global IT traffic by
2020



40

zettabytes

Volume of data
created by 2020, up
300% from 2015



2.3

zettabytes

Volume of data that
humans produce
every day

Volume Of Data Created, Captured, Copied, and Consumed Worldwide (2010 - 2020)

■ Data volume (in zettabytes)



in 2022

78% of large organizations (10,000+ employees) used Big Data



in 2022

the US had 2,701 data centers, the highest in the world.



2022-2029

the Big Data market will grow by 13.4%



by 2029

the Big Data analytics industry will be worth \$650 billion



in 2021

women's salaries in the big data sector were significantly lower than men's

The Global Analytics as a Service

market is expected to reach \$59B by 2027

in 2021

the US accounted for 51% of global spending on big data and analytics solutions