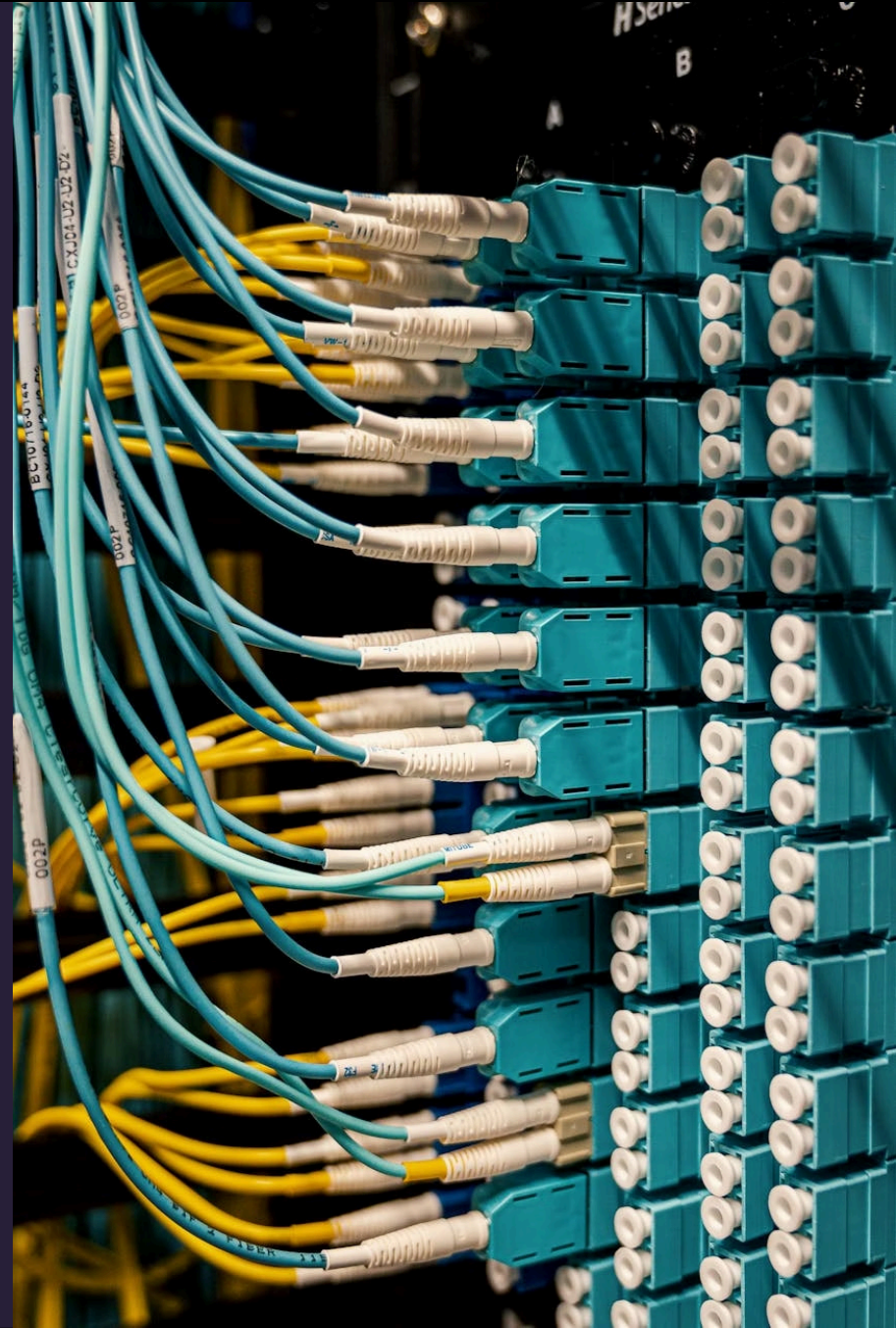
Educational Module

Behind Digital Platforms

Thematic Unit 2 - Big Data: what they are and what they are used for, how they operate, how they are produced



What are big data?

The term *big data* refers to datasets so vast and complex that they require new technologies, such as artificial intelligence, to be processed.

Data come from various sources. They often have internal uniformity, such as GPS data from millions of mobile phones used to ease traffic congestion. Big data can also result from a combination of sources, for example medical records and smartphone applications intended for patients.

How are big data generated?

Not all big data are naturally generated directly by our phones and apps.

Car telemetry, navigation systems, airplane engines, and banking transactions are examples of big data sources that are collected, analyzed, processed, and reused.

We are not always the direct generators of these data. Although, on closer inspection, even when not directly, our activities are often behind the production of these big data.

Big data can be:

Structured

Already collected in processable databases, organized from the moment of production (e.g., when we enter name and surname for a payment).

Semi-structured

Mixed between structured and unstructured; mainly used to transmit data between a server and a web application, e.g., e-mails, XML files.

Unstructured

Without a predefined model; cannot be organized into rows and columns and are processed when used, e.g., images, photos, videos, posts, IoT (Internet of Things).

They have changed our territories and our lives.

Thanks to algorithms and big data, platforms offer goods and services that we will not even have to search for — they will find us!

But in what sense do "products find me"?

Creating value, expanding markets: Internet of Things and big data

What do you mostly do when you are on your smartphone?

I send DMs to my bros and sistas

I take photos

I scroll social media

I work/study

Obviously, the correct answer is: I work.

"Why would I be working?"

Because many online products are free: Instagram Google Search Google Meet Glovo ...and many others.

Apparently different, they share one common characteristic.

> When we use them, we authorize their owners or developers to collect our personal data.

Every time we browse the internet from our phone/tablet/PC/smartwatch, interact, or carry out operations of any kind, we "work," because we produce an enormous amount of data and information.

Why collect all this data?

We provide data:

Directly

Upon explicit request (gender, age, place of residence, etc.).

Indirectly

By generating data through web searches, orders, document requests, payments, etc.

> **Why is someone interested in collecting, cataloguing, processing, and reusing this immense mass of data?**

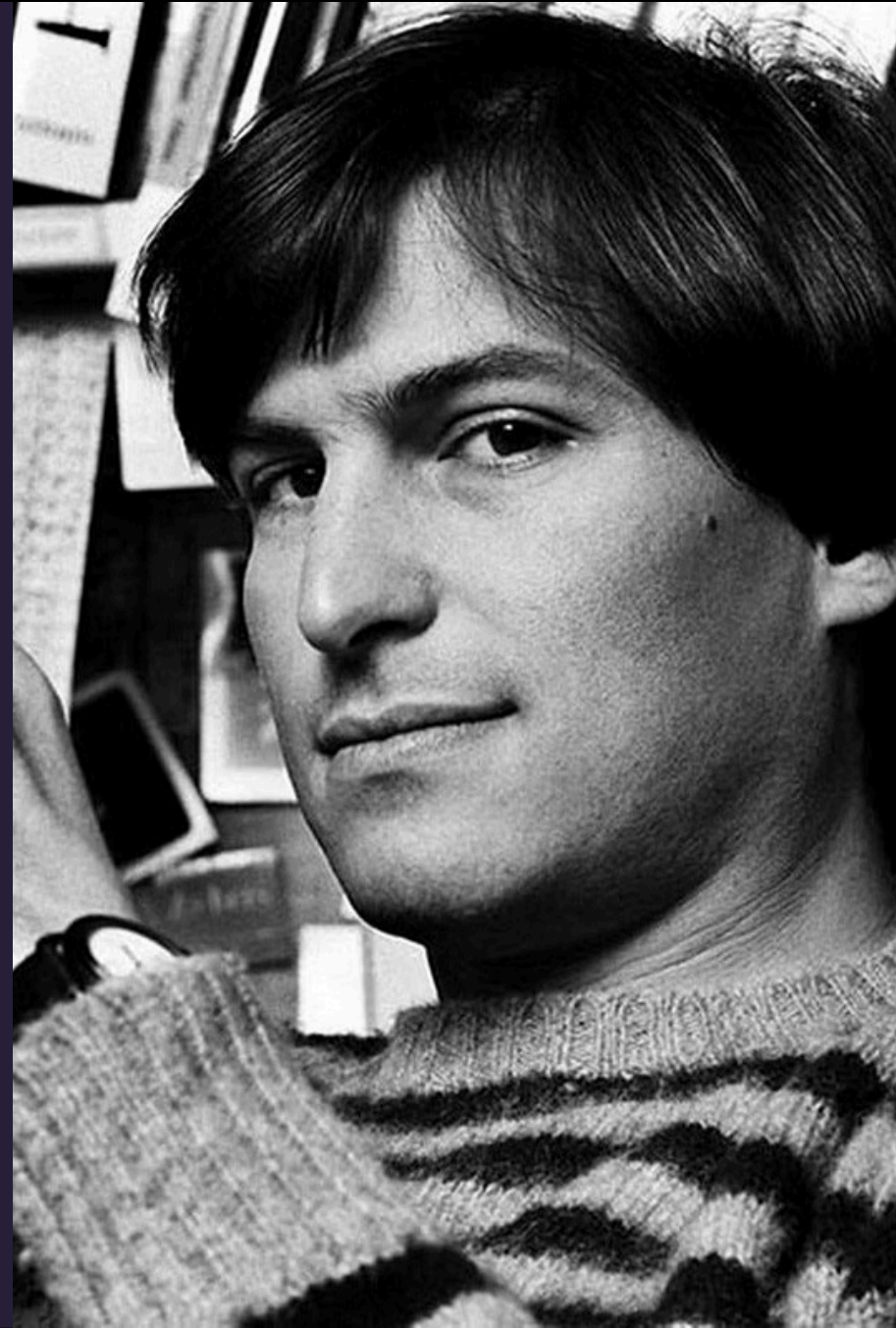
Back to the 1990s

In the 1990s, when we began connecting from our PCs, we started leaving useful trails of data.

In the technological innovation sector, it quickly became clear that collecting and analyzing the data we release while browsing the internet could become a new and fertile space for profit creation.

Everyone in the tech sector began asking: how can we extract ever greater and more usable quantities of data?

One person in particular lost sleep trying to find an answer.



The "Data Worker"

With the arrival of new technologies, the number of connected people increased dramatically, giving rise to a new profession: the **"data worker."**

> **"But the data worker doesn't work 24/7! They go for walks, to the bar, to the gym, to work, they cook... they get distracted... How can I convince them to stay actively connected to the PC from morning to evening?"**

This was the question investors and developers asked themselves.

There was no answer: there is no motivation capable of convincing people to stay at home in front of a computer all day.

The answer: "I put the internet inside the mobile phone!"

A brilliant idea.

In a very short time, smartphones became devices for continuous data production.

And they do not produce data only when we use them, but also when they lie unused in our pockets.

The apps we use are not free — we pay with our data.

App permissions

When you download an app from Google Play or the Apple Store, you receive a request for authorization.

For example: A camera app asks permission to access the camera before taking photos. That is obvious.

But it doesn't end there.

Apps request access to:

- Location
- Stored data
- Call and message logs
- Microphone
- Sensitive log data
- Cache
- Contacts
- Calendar
- Browsing history

User: "You want to listen through my microphone and read my private messages???"

Platforms: "Oh no? Then no app!!!"

User: "NO, WAIT, I WAS JOKING — TAKE THEM ALL!"

Practical consequences of big data

Let us briefly examine some practices made possible by big data:

Labor

Changes in labor contracts, increase in non-protected jobs

Environment

Increase in pollution (cloud infrastructure)

Military

Collaboration in military technologies

AI Bias

AI, although useful and efficient, is programmed and "trained" by people according to corporate, political, and economic strategies

AI, Google and Amazon

Uber Eats, Just Eat, Glovo and similar platforms are not simply food delivery services — they rent out couriers who deliver the food.

Why?

What are the consequences of algorithms and big data collection on labor and workers' economic conditions?

Decline in workers' rights
(pushed toward hyper-productivity and risk-taking)

Proliferation of digital gangmaster systems

Subordinate labor presented as "self-employed" work
Emphasizing flexibility while downplaying disadvantages

Indefinite or arbitrary schedules

Pay determined by algorithm

Payment only during active delivery time
(waiting time unpaid)

Data usage increases economic inequality
(small innovative businesses weakened)

Dark kitchens

Reduction of social spaces

Platformization of spaces and forms of citizenship

Broader impacts

The German newspaper *Bild* dismisses editorial staff due to AI.

Amazon and Google provide technology to Israel to manage attacks (calculating acceptable percentages of civilian casualties).

Therefore, we must reflect on how AI is trained, for what purposes, and to whose advantage.

Resources

Bibliographic / Web References – UT 2

- Big Data: a framework for orientation
- Big Data explained in 18 minutes | What is Big Data?
- Bild newspaper fires journalists due to AI
- Meta removes fact-checking (January 2025)
- Generative AI, Democracy and Human Rights – Harris-Shull-CIGI-2025
- Cuppini, Mezzadra, Frapporti, Pirone (2024). *Capitalism in the Platform Age*. Springer.