ITINERIS

Data mining and machine learning

Technical challenges and limitations of AI

Vittoria Mascellaro

IR0000032 – ITINERIS, Italian Integrated Environmental Research Infrastructures System
(D.D. n. 130/2022 - CUP B53C22002150006) Funded by EU - Next Generation EU PNRR-Mission 4 "Education and Research" - Component 2: "From research to business" - Investment 3.1: "Fund for the realisation of an integrated system of research and innovation infrastructures"

Finanziato dall'Unione europea NextGenerationEU

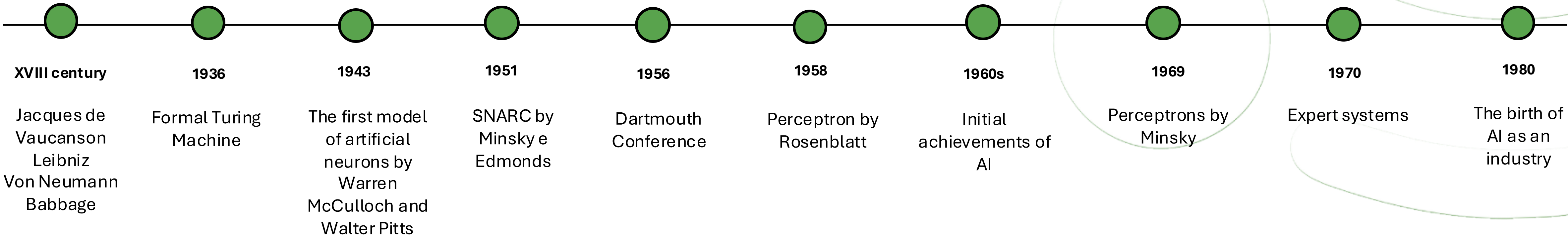
Ministero dell'Università e della Ricerca

Italiadomani

Time	Duration	Training Module - Topic	Speaker
09:00 – 10:45	1h45m	Artificial Intelligence and data	Vittoria Mascellaro
10:45 – 11:00		Coffee Break	
11:00 – 13:00	2h00m	Data ethics, AI ethics (Real-world cases of misure or controversy)	Vittoria Mascellaro
13:00 – 14:00		Lunch Break	
14:00 – 15:30	1h30m	Bias	Vittoria Mascellaro
15:30-15:45		Coffee Break	
15:45 – 16:30	45m	Group activity	Vittoria Mascellaro

Module 1: AI and Data

Historical perspective



XVIII century
Jacques de Vaucanson
Leibniz
Von Neumann
Babbage

1936
Formal Turing Machine

1943
The first model of artificial neurons by Warren McCulloch and Walter Pitts

1951
SNARC by Minsky e Edmonds

1956
Dartmouth Conference

1958
Perceptron by Rosenblatt

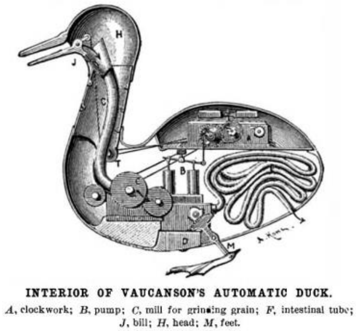
1960s
Initial achievements of AI

1969
Perceptrons by Minsky

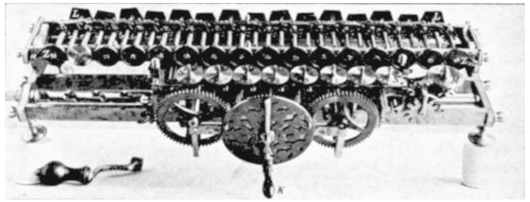
1970
Expert systems

1980
The birth of AI as an industry

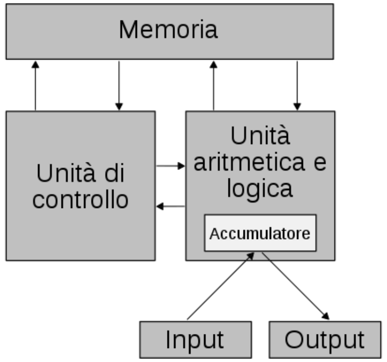
Historical perspective


INTERIOR OF VAUCANSON'S AUTOMATIC DUCK.
A, clockwork; B, pump; C, mill for grinding grain; F, intestinal tube; J, bill; H, head; M, feet.

The *Digesting Duck* by Jacques de Vaucanson (1738)

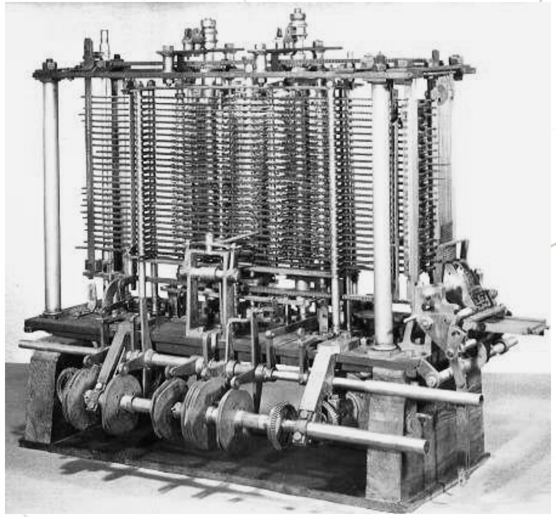


Calculus ratiocinator By Leibniz



Von Neumann architecture

Charles Babbage's Difference Engine



Charles Babbage's Analytical Engine

- These figures allow us to refer to the tradition of formalist research.
- This tradition helps us understand how artificial performance is part of human practice.
- In particular, through the projects of mathematician Charles Babbage, we see the human tendency to self-imitate using machines.
- Mathematician Ada Lovelace, in 1840, recognized the potential of Babbage’s Analytical Engine.
- Lovelace was interested in the machine’s ability to process symbols that could represent all objects.
  She was the first to foresee the advent of a form of Artificial Intelligence.
- Artificial Intelligence was possible, but it was not yet clear how to achieve it.

ARTIFICIAL INTELLIGENCE

=

THE SCIENCE THAT ADDRESSES THE PROBLEM OF HOW TO REPRESENT AND BUILD KNOWLEDGE

ARTIFICIAL INTELLIGENCE

refers to

«Systems that exhibit intelligent behavior by analyzing their environment and taking actions, with a certain degree of autonomy, aimed at achieving specific objectives.»

Ethics Guidelines for Trustworthy Artificial Intelligence, 2019

ARTIFICIAL INTELLIGENCE

refers to

«Systems that exhibit intelligent behavior by analyzing their environment and taking actions, with a certain degree of autonomy, aimed at achieving specific objectives.»

Ethics Guidelines for Trustworthy Artificial Intelligence, 2019

ARTIFICIAL INTELLIGENCE

refers to

«Systems that exhibit intelligent behavior by analyzing their environment and taking actions, with a certain degree of autonomy, aimed at achieving specific objectives.»

Ethics Guidelines for Trustworthy Artificial Intelligence, 2019

Data as the foundation

- Artificial Intelligence relies on large volumes of data.
- Data fuels machine learning models: **more data → better accuracy**.
- **Raw data → Information → Knowledge → Automated decisions**
- AI doesn't just analyze data — it transforms it into **intelligent actions** (e.g., predictions, recommendations, classifications).

Definition of data

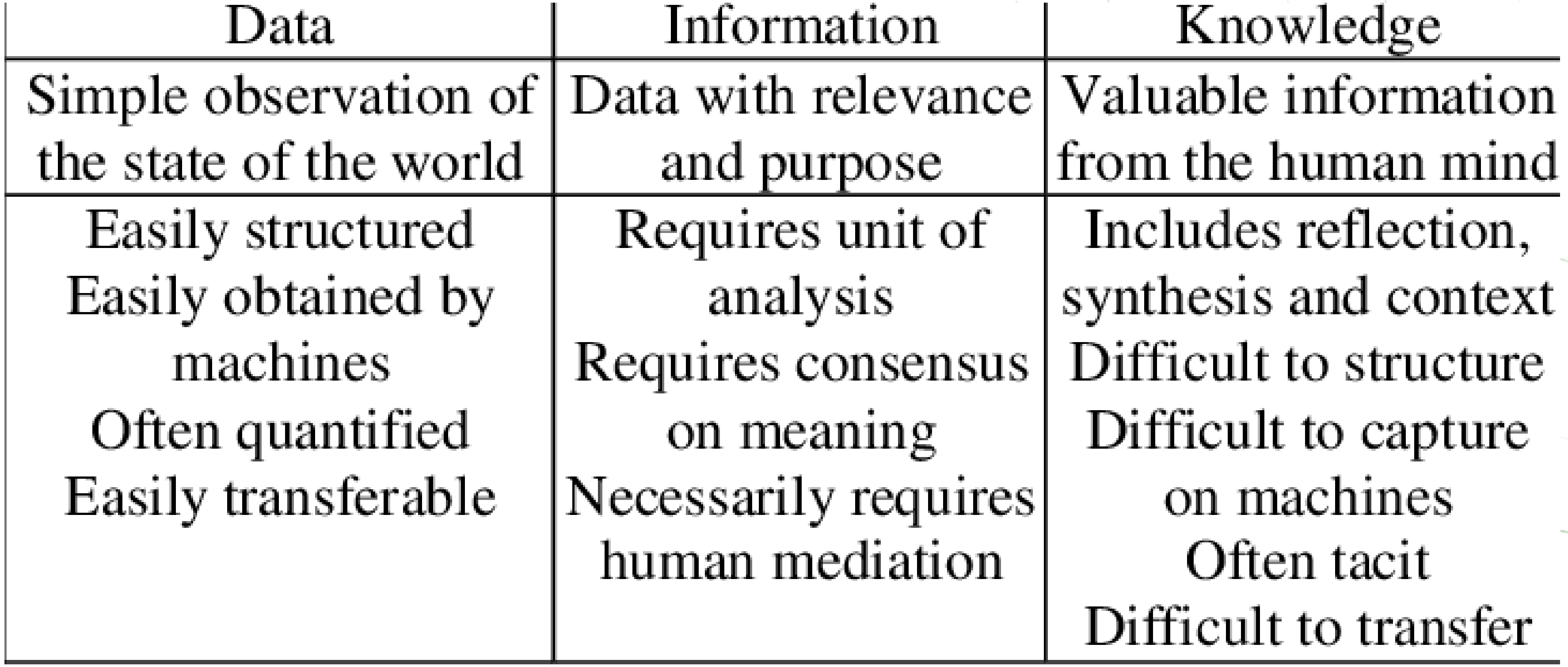
Data are **original representations** — that is, not yet interpreted — **of a phenomenon, event, or fact**, conveyed through symbols, combinations of symbols, or any other expressive form associated with a medium

Definition of data

Data are **original representations** — that is, not yet interpreted — **of a phenomenon, event, or fact**, conveyed through symbols, combinations of symbols, or any other expressive form associated with a medium

⇩

Data are representations of events or facts:
– **Not interpreted (original)**
– **Expressed through symbols (or combinations of symbols)**
– **Stored or conveyed on some medium (expressive form)**



Data	Information	Knowledge
Simple observation of the state of the world	Data with relevance and purpose	Valuable information from the human mind
Easily structured Easily obtained by machines Often quantified Easily transferable	Requires unit of analysis Requires consensus on meaning Necessarily requires human mediation	Includes reflection, synthesis and context Difficult to structure Difficult to capture on machines Often tacit Difficult to transfer

Source: [8].

Structured data vs Unstructered data

Structured data refers to data that follows a predefined and expected format → as a table in a database, with columns for name, date, temperature — each entry follows a set structure

VS

Unstructured data lacks a predefined format (e.g.Podcast, video files…)

Quality of data

HOW GOOD IS THIS DATA?

What makes data "High quality"?

- **Accuracy**
- **Consistency**
- **Timeliness**
- **Completeness**
- **Spatial and temporal resolution**
- **Metadata and documentation**

Data Quality and validation according to ISTAT

According to ISTAT, the final output of a statistical survey can be broken down into **three levels of information**:

- **Microdata** = individual data points
- **Macrodata** = statistical summaries
- **Metadata** = documentation about the data

Together represent the **statistical information** produced by a survey. That’s why ISTAT refers not just to **data quality**, but more broadly to the **quality of information >** we must define what “quality” means at **each of the three levels** — individual data, aggregated results, and metadata.

Data Quality and validation according to ISTAT

ISTAT adopts a definition of quality originally proposed by **O. Arkhipoff** in 1986:

"The quality of a product is its ability to meet the guarantees provided by the producer."

These guarantess includes both the **design characteristics and tolerance**

Data Quality and validation according to ISTAT

Design guarantess

1. Timeliness
2. Theoretical relevance
3. Effective relevance
4. Transparency
5. Tolerence

Tolerance guarantess

1. Sampling precision
2. Non-sampling precision

Dimensions of data quality

Dimension	Definition	Defined by
1. Relevance	The extent to which statistics meet the real needs of users.	Eurostat
2. Accuracy	The closeness between statistical estimates and the true values.	Eurostat
3. Timeliness	The delay between the reference period and the availability of data.	Eurostat
4. Punctuality	The degree to which data is released according to the planned schedule.	Eurostat
5. Accessibility	The ease with which users can access the data.	Eurostat
6. Clarity (Transparency)	The clarity of presentation and documentation, enabling users to understand and interpret data.	Eurostat
7. Comparability	The possibility of comparing data across time, regions, or countries.	Eurostat
8. Coherence	The internal consistency of data and its compatibility with other datasets.	Eurostat
9. Completeness	The extent to which required data are available without gaps.	Eurostat
10. Confidentiality Protection	Ensuring the privacy of respondents and secure handling of individual data.	**ISTAT** (added)

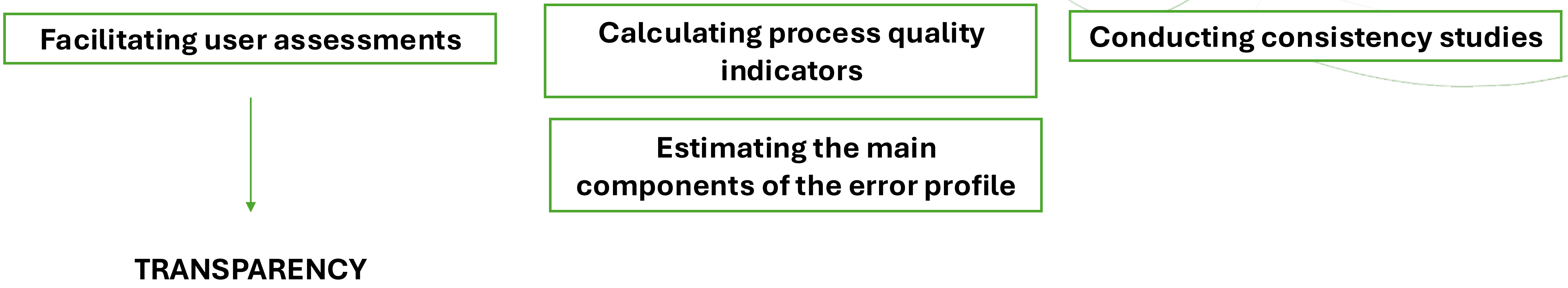
Data validation

Data validation involves examining all the characteristics that define the **dimensions of data quality**, and it has two main objectives:

a) To assess whether the **quality of the data is sufficient** for public dissemination.

b) To identify the **most significant sources of error**, and to introduce changes in the production process in order to reduce errors in future surveys.

Four key validation measures:



Facilitating user assessments

Calculating process quality indicators

Conducting consistency studies

Estimating the main components of the error profile

TRANSPARENCY

Data validation

According to a definition provided by **Marescotti (1985)**, **environmental information** has three fundamental characteristics:

1. **Complexity**
2. **Uncertainty**
3. **Conflict**

Data abundance vs data scarcity

- **Data Abundance:**
  When there is a large volume of data available, often from multiple sources, sometimes even overwhelming in size.
  Example: Social media data, satellite imagery, sensor networks producing continuous streams of information.

- **Data Scarcity:**
  When data is limited, either in quantity, quality, or both. This can happen due to cost, accessibility, or rarity of events.
  Example: Rare disease cases, remote environmental measurements, early-stage research data.

Challenges of Big Data:

STORAGE

PROCESSING

NOISE

Challenges of Small Data:

OVERFITTING

LACK OF REPRESENTATIVENESS

Group activity: Exploring data quantity challenges

Group 1

Small Sample Size

- **Scenario:** A city wants to model traffic flow but only has traffic count data from 3 days in a year.
- **Challenge Questions:**
  - What problems might arise using such a small sample?
  - How might this affect the model's reliability and predictions?
  - What strategies could improve data quantity or address this issue?

Group 2

Missing Data

- **Scenario:** A weather dataset has temperature readings for every day, but 20% of the data is missing randomly.
- **Challenge Questions:**
  - How could missing data impact analysis?
  - What are possible risks when building models with this dataset?
  - What are common techniques to handle missing data?

Group 3

Uneven Sampling

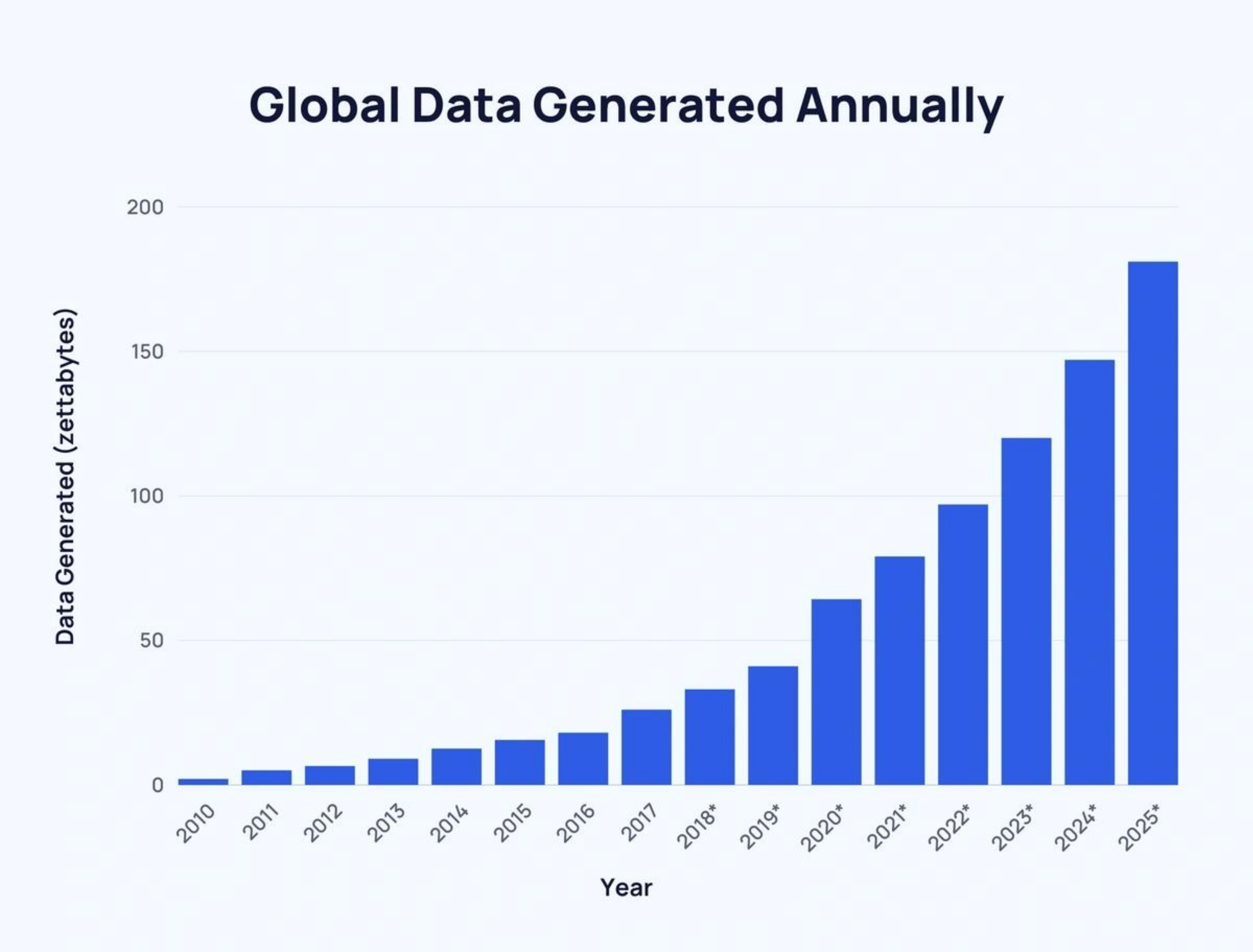
- **Scenario:** Environmental sensors are deployed in a forest, but some sensors record data hourly while others record daily.
- **Challenge Questions:**
  - What challenges could uneven sampling frequencies cause?
  - How might this bias the results or the model?
  - How could you standardize or correct this inconsistency?

Group 4

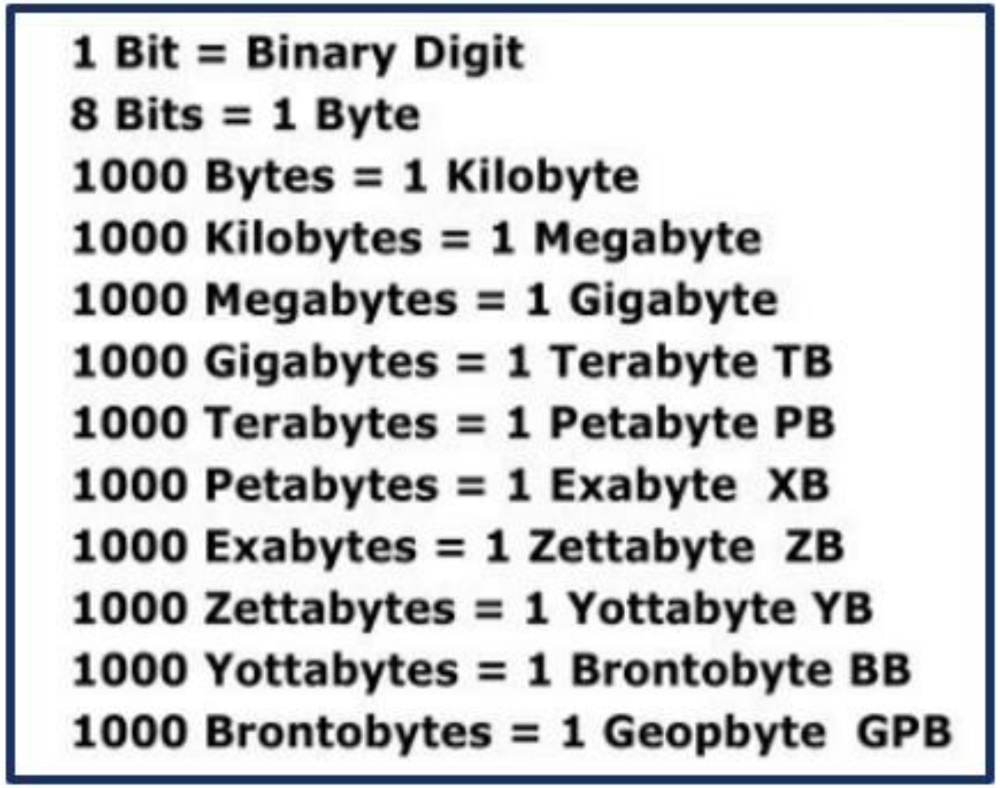
Excessive Data / Overfitting Risk

- **Scenario:** A model uses a very large dataset with thousands of features but limited observations (high dimensionality).
- **Challenge Questions:**
  - What issues can arise from having too many features relative to data points?
  - How can this affect the model's performance?
  - What approaches can reduce this risk?

The amount of data produced every day is growing exponentially



The amount of data produced every day is growing exponentially



1 Bit = Binary Digit
8 Bits = 1 Byte
1000 Bytes = 1 Kilobyte
1000 Kilobytes = 1 Megabyte
1000 Megabytes = 1 Gigabyte
1000 Gigabytes = 1 Terabyte TB
1000 Terabytes = 1 Petabyte PB
1000 Petabytes = 1 Exabyte XB
1000 Exabytes = 1 Zettabyte ZB
1000 Zettabytes = 1 Yottabyte YB
1000 Yottabytes = 1 Brontobyte BB
1000 Brontobytes = 1 Geopbyte GPB

Datafication

"Datafication is not just the quantification of information but the rendering of many aspects of the world into a data format that makes it calculable and accessible to digital algorithms.”

— **van Dijck, J.** (2014). *Datafication, dataism and dataveillance: Big Data between scientific paradigm and ideology*.

Data power

When we talk about «data power» by bringing together the two concepts: that of datafication on the one hand and that of power on the other hand, we mean that there forms of power, the ability to achieve one’s will against the resistance of others, which can now be achieved by means of controlling power. Data has become an instrument of power, and a condition for the exercise of power.

Lecture

BREAK

Training object development for all target categories,including future generations of Earth stewards: "Artificial Intelligence applied to enviromental monitoring ", Lecce, 3/06/2025 34

Module 2: AI ethics

Data Ethics

"The real risk of artificial intelligence is not malice, but competence."

Stephen Hawking

AI Ethics

- **Ethics** is the branch of philosophy concerned with judging whether actions are good or bad.
- **AI Ethics** is the branch of technology ethics that specifically focuses on artificially intelligent systems.
- It involves the **creation of a test capable of determining whether decisions made by AI are ethical**.

Three approaches

Ethics in AI

Ethics of AI

Ethics for AI

Ethics in AI

The one embedded in AI software.

1. Ethics in AI

The one embedded in AI software.

2. Ethics of AI

It concerns the interaction between human beings and artificial agents.

1. Ethics in AI

The one embedded in AI software.

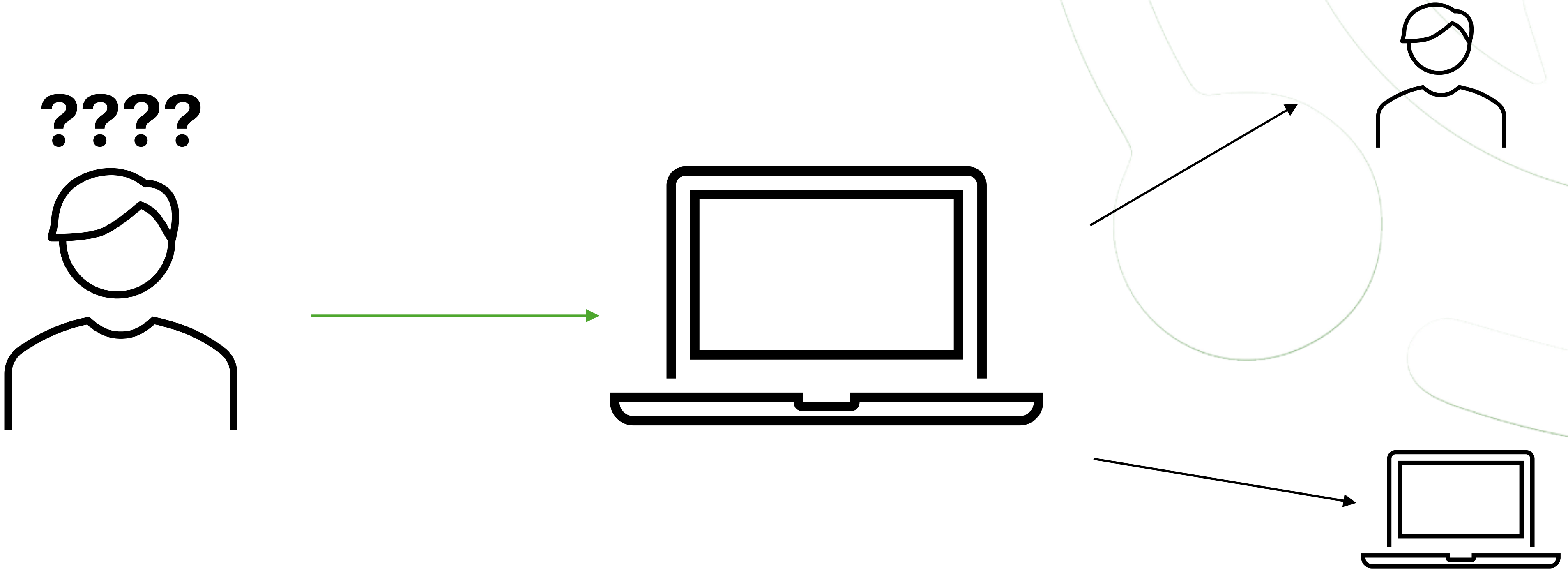
2. Ethics of AI

It concerns the interaction between human beings and artificial agents.

3. Ethics for AI

What shared and shareable ethics? The idea of a global and comprehensive ethics because AI is a global phenomenon and requires equally global responses.

The imitation game



Purposes of AI

- **Should we give AI a purpose?**
  If so, **what kind of purpose** should that be?
- **How can we define goals** for an AI system in a way it can understand and follow?
- **How can we ensure** those goals are maintained over time, especially as the system evolves or learns?
- **What are the purposes of human beings?**
  And should AI align with them, replicate them, or challenge them?

Friendly Artificial Intelligence

Eliezer Yudkowsky

Artificial intelligence **whose goals are aligned with ours**, based on the principle of **coherent extrapolated volition.**

↓

It means building an AI that does **what we would want it to do**, **if we knew more**, were **more rational**, and had **more time to think**.

↓

AI should help us fulfill our **better, wiser, long-term goals**, —not just our immediate desires or flawed preferences.

Breakdown of the problem:

- Ensuring that AI **understands** our goals
- Ensuring that AI **adopts** our goals
- Ensuring that AI **preserves** our goals

Understanding human goals: solution

Two key problems:

1. **Finding an effective way to encode arbitrary systems of goals and ethical principles** into a machine.
2. **Enabling machines to determine which specific system** of goals or values corresponds to the behavior they observe.

Understanding human goals: solution

- **Inverse Reinforcement Learning (IRL)**
  (Proposed by Stuart Russell)
- This approach expects the AI to **infer something about our goals** by observing the **decisions and actions it takes**.
- In other words, the AI learns what we want by analyzing behavior, rather than being explicitly told.

Adopting our goals: solution

Corrigibility

=

It is possible to give AI a system of goals
that **can be corrected or adjusted** by humans.

Adopting our goals: solution

But are we sure that AI’s goals won’t evolve as its intelligence evolves?

Maintaining goals: the goal preservation problem

- Steve Omohundro and Nick Bostrom argue that we can predict certain **sub-goals** of an AI regardless of its initial goals.
- If a Friendly AI self-improves, can it remain friendly?
- Therefore, it is crucial to clearly define the AI’s goals and ensure they are aligned with human values.

Goal alignment: the most important problem

- What are the goals of human beings?
- Four guiding principles:
- **Utilitarianism**
- **Diversity**
- **Autonomy**
- **Legacy**

Human Principles

UTILITARIANISM

Conscious positive experiences should be **maximized** while suffering should be **minimized**.

↓

Challenge: The problem of consciousness — how do we define and measure conscious experiences?

DIVERSITY

A varied set of positive experiences

↓

Has enabled the survival of the species

AUTONOMY

Conscious beings and societies must be free to pursue their own goals.

LEGACY

Ensures compatibility with scenarios that humans consider good.

https://www.moralmachine.net/

AI principles: can human principles align with AI principles?

Six major high-level documents:

- Asilomar AI Principles (2017)
- Montreal Declaration for Responsible AI Development (2017)
- Ethically Aligned Design: A Vision for Prioritizing Human Wellbeing with Autonomous and Intelligent Systems (IEEE, 2017)
- Statement on Artificial Intelligence, Robotics, and Autonomous Systems (EGE, 2018)
- AI in the UK: Ready, Willing and Able? (AIUK, 2017)
- AI Partnership Principles (2018)

AI principles

In 2020, the AI Ethics Guidelines Global Inventory identified **160 proposed principles**

↓

Problem: Overlap and confusion caused by so many guidelines

Top-down approach

1. **Beneficence**
2. **Non-maleficence**
3. **Autonomy**
4. **Justice**
5. **Explainability**

Beneficence

Promote well-being, preserve dignity, and support the planet

- **“The development of artificial intelligence should ultimately promote the well-being of all sentient beings.”** — Montreal Declaration for Responsible AI Development
- **“Common Good”** — Referenced in both **AIUK** and **Asilomar AI Principles**

Non-maleficence

Privacy, security, and capability caution

- It is still unclear whether the people developing these technologies should be encouraged not to do harm,
  or if it is the technology itself that should be prevented from doing harm.
- At the heart of this dilemma lies the issue of **autonomy**.

Autonomy

The power to decide to decide

Establishing a balance between the decision-making power we retain and the power we delegate to artificial agents

- "They must not compromise humans' freedom to establish their own standards and norms." — ESE
- “The autonomous power to harm, destroy, or deceive human beings should never be granted to AI.” — AIUK

Justice

Promote prosperity, preserve solidarity, and prevent inequity

Establishing a balance between the decision-making power we retain and the power we delegate to artificial agents

- “The development of AI should promote justice and strive to eliminate all forms of discrimination.”— Montreal Declaration

Are we (human beings) the patient receiving the "treatment" from AI, which presents itself as the doctor, or are we both?

Explainability

Enabling the other principles through intelligibility and accountability.

Answers the question:
HOW DOES IT WORK?

TRANSPARENCY

The five principles in the six documents

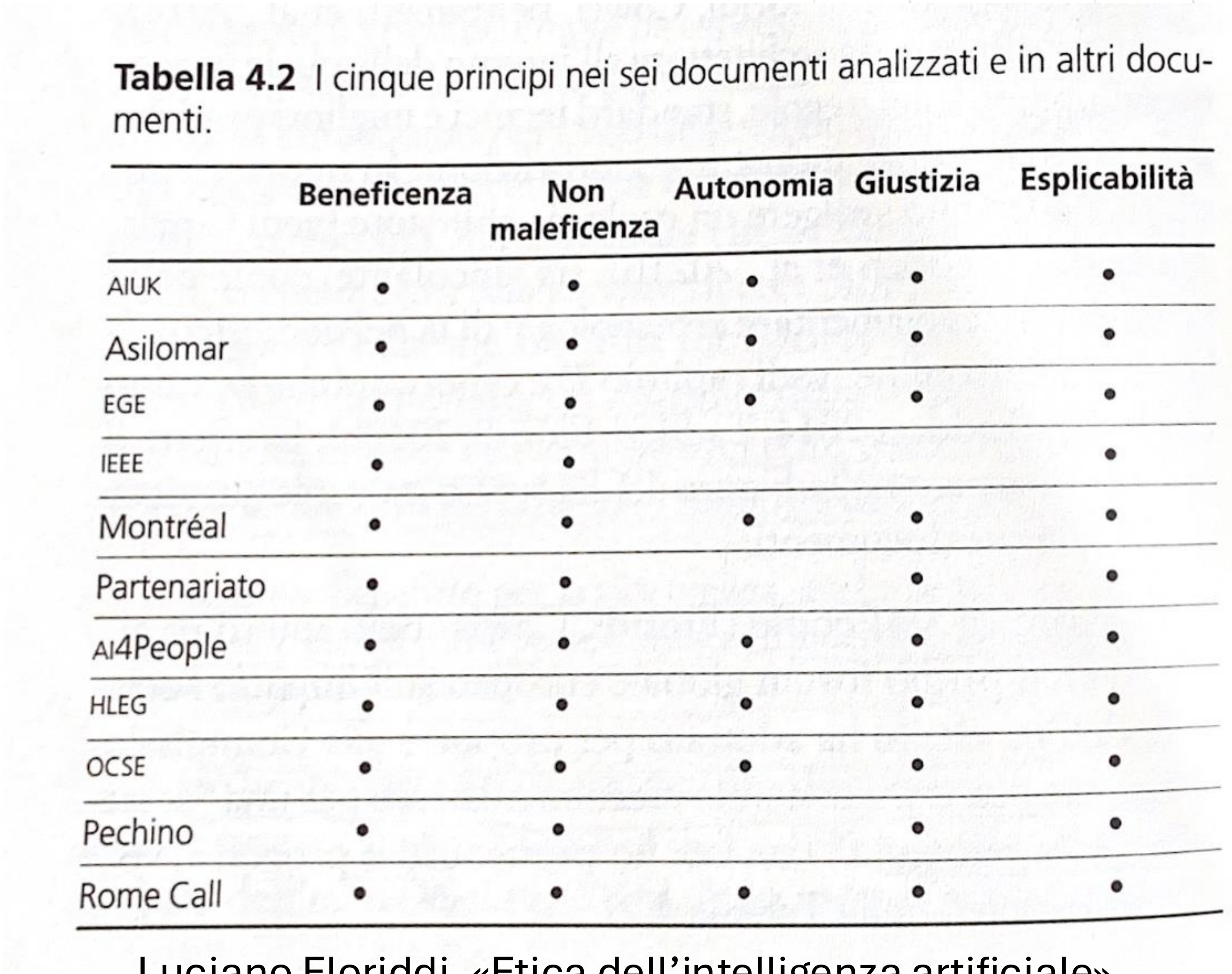


Tabella 4.2 I cinque principi nei sei documenti analizzati e in altri documenti.

	Beneficenza	Non maleficenza	Autonomia	Giustizia	Esplicabilità
AIUK	●	●	●	●	●
Asilomar	●	●	●	●	●
EGE	●	●	●	●	●
IEEE	●	●			●
Montréal	●	●	●	●	●
Partenariato	●	●		●	●
AI4People	●	●	●	●	●
HLEG	●	●	●	●	●
OCSE	●	●	●	●	●
Pechino	●	●		●	●
Rome Call	●	●	●	●	●

Luciano Floriddi, «Etica dell’intelligenza artificiale»

Bottom-up approach

- It starts from principles
- We identify common principles starting from the analysis of the codes developed and adapted in different European countries.

Bottom-up approach

1. **Transparency**
2. **Accountability**
3. **Explainability**
4. **Respect for privacy**
5. **Justice**
6. **Security**

Bottom-up approach

1. **Transparency**
2. **Accountability**
3. **Explainability**
4. **Respect for privacy**
5. **Justice**
6. **Security**

Challenges that remain to be addressed:

- How the principles are interpreted
- The order of priority among principles
- The specific fields and actors to which they apply

Risks

- Ethical shopping
- Ethical bluewashing
- Ethical lobbying
- Ethical dumping
- Ethics evasion

Ethical shopping

“The malpractice of selecting, adapting, or revising ethical principles, guidelines, codes, frameworks, or similar standards by picking from a variety of available options, in order to give a new veneer to some pre-existing behaviors and thereby justify them retrospectively, instead of implementing or refining new behaviors by comparing them with public ethical standards.”

Ethical shopping

Risk of mixing and matching preferred ethical principles, causing incompatibility of standards

+

Risk of reduced competition, evaluation, and accountability

↓

STRATEGY:
Establish clear, shared, and publicly accepted ethical standards

Ethical guidelines for trustworthy AI

In 2021, these guidelines influenced the proposal adopted by the European Commission for an AI regulation, described as the first-ever legal framework on AI.

Ethical bluewashing

“The malpractice of making unfounded or misleading claims regarding ethical values and the benefits of processes, products, services, or other digital solutions in order to appear more ethically sound in the digital realm than one actually is.”

Ethical bluewashing

Marketing Practice

Bluewashing + Ethical shopping

=

A public or private actor acquires ethical principles and publicizes them to emphasize their ethical commitment without producing real improvements

↓

STRATEGY:

Transparency and education

(In the long term, certifications for digital products and services are also expected to be established.)

Ethical lobbying

“The malpractice of exploiting digital ethics to delay, revise, replace, or avoid appropriate and necessary legal regulation (or its enforcement) related to the design, development, and implementation of processes, products, services, or other digital solutions.”

Ethical lobbying

Undermines the foundation of ethical self-regulation

↓

And can delay the introduction of necessary regulations

↓

STRATEGY:

Good legislation and effective enforcement

Ethical dumping

“The discontent of (A) outsourcing research activities related to processes, products, services, or other digital solutions to other contexts or locations (for example, from European organizations outside the EU) in ways that would be ethically unacceptable in the original context or location; and (B) importing results of such ethically questionable research activities.”

Ethical dumping

Export of Unethical Research Practices

Involves both the export of unethical practices and the unethical import of their results

Ethical dumping may worsen in the near future due to:

1. Impact of digital technologies on healthcare, social services, defense, policing, and security
2. Ease of their deployment and use
3. Strong economic interests

Ethical dumping

STRATEGY

1. **Research ethics:** Control of public funding for research
2. **Consumer ethics:** Establishment of a certification system for products and services

Ethics evasion

“The malpractice of performing less and less ‘ethical work’ in a given context the lower the perceived return of such ethical work in that context.”

Ethics evasion

Applying double standards

STRATEGY:
Address the issue of lack of accountability

↓

More fairness, less bias, and an ethics of distributed responsibility

BREAK

Module 3: Bias

Bias

The occurrence of distorted outcomes due to human prejudices that alter the original training data or the AI algorithm itself, leading to skewed and potentially harmful outputs

Types of AI bias

- Algorithmic bias
- Cognitive bias
- Confirmation bias
- Exclusion bias
- Measurement bias
- Out-group homogeneity bias
- Prejudice bias
- Recall bias
- Sampling/Selection bias
- Stereotype bias

Algorithms that amplify biases in present data

- Algorithms can not only **absorb** those biases, but actually **amplify** them

How does this happen?

1. **Learning from biased data**
2. **Reinforcing existing trends**
3. **Feedback loops**

Why does this matter?

We risk making inequalities worse

Biases that were once hidden in society can become embedded in technology.

Case Study: Algorithmic bias – AMAZON (2014)

- Amazon used a **recruitment software** designed to analyze candidates’ résumés
- However, the **algorithm discriminated against women**
- The cause was rooted in the data used—résumés received over the previous 10 years, which were predominantly from men
- As a result, the **male profile became the dominant pattern**, leading to **bias** in the system

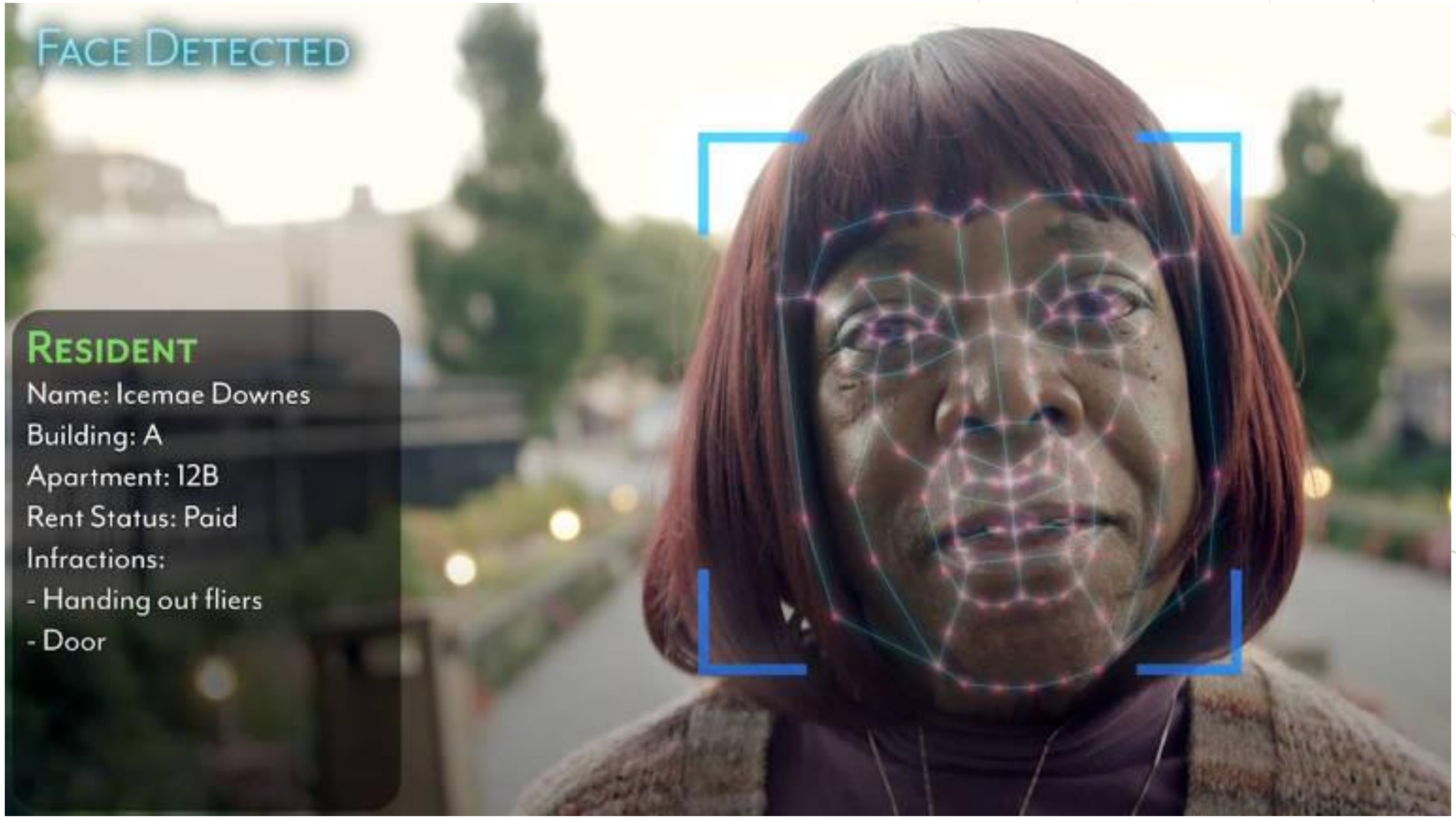
Case Study: Algorithmic bias – COMPAS

- Used by **U.S. judges** to help determine sentencing for convicted individuals
- The algorithm **embedded biases against African Americans**
- The cause was the **dataset used**, which lacked **balanced data** across different ethnic groups
- As a result, **African Americans were twice as likely** to be labeled as **high-risk**

Coded bias

- **Documentary by Shalini Kantayya about digital surveillance, released in 2020**
- It tells the story of how **data surveillance disproportionately affects groups that are still considered minorities today**
- **Joy Buolamwini**, a Ghanaian computer scientist and activist, developed **Aspire Mirror** in 2015—a device designed to display the hopes and dreams of the person reflected in the mirror
- However, Buolamwini never even had the chance to see what aspirations the device would show for her—because **the system she created failed to recognize her face**

Coded bias



FACE DETECTED

RESIDENT

Name: Icemae Downes

Building: A

Apartment: 12B

Rent Status: Paid

Infractions:

- Handing out fliers
- Door

Group activity: Invisible pollution

The problem

After a year of implementation, environmental NGOs and citizen groups noticed something strange. The model consistently reported **lower pollution levels** in certain low-income neighborhoods — despite clear evidence of heavy traffic, industrial activity, and frequent respiratory issues reported by local clinics.

Investigation findandings

- These disadvantaged areas had **fewer air quality sensors**, due to underinvestment.
- Training data was heavily weighted toward **central, wealthier zones**.
- The model assumed that areas with green spaces nearby had low pollution — but failed to account for illegal waste burning or aging heating systems common in poorer districts.
- Complaints from residents were not included in the model’s input, as they were seen as “anecdotal” data.

Discussion question

- **What types of bias are present in this case?**
- **How could these biases affect policy and public health?**
- **What changes would you suggest to make the model more fair and accurate?**
- **Can “less data” about an area be considered a form of bias in itself? Why or why not?**
- **How can local communities be involved in improving these models?**

Divide in small teams

- City government
- Data scientists
- Local community representatives
- Environmental health experts

Each group must:

1. Identify their key concern

2. Propose one concrete change to the AI model or the data collection process

How to mitigate bias

At the data level (pre-processing)

- **Data augmentation**: generate new examples for underrepresented groups
- **Resampling**: apply over/undersampling to balance classes/groups
- **Reweighing**: assign different weights to samples based on their group membership

At the model level (in-processing)

- **Fairness-aware algorithms**: models designed to incorporate fairness constraints
- **Fairness regularization**: penalize bias during training

At the output level (post-processing)

- **Equalized odds post-processing**: adjust predictions to reduce bias
- **Threshold optimization** for different groups
- **Reject option classification**: alter decisions in uncertain cases to promote fairness

THANKS!

IR0000032 – ITINERIS, Italian Integrated Environmental Research Infrastructures System
(D.D. n. 130/2022 - CUP B53C22002150006) Funded by EU - Next Generation EU PNRR-Mission 4 “Education and Research” - Component 2: “From research to business” - Investment 3.1: “Fund for the realisation of an integrated system of research and innovation infrastructures”

Finanziato dall'Unione europea NextGenerationEU

Ministero dell'Università e della Ricerca

Italiadomani