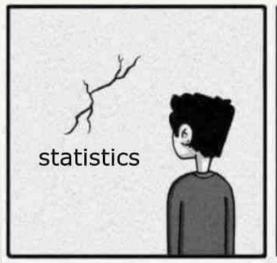
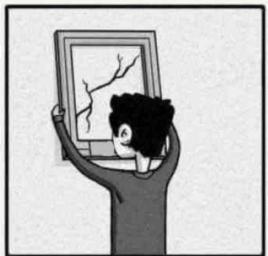
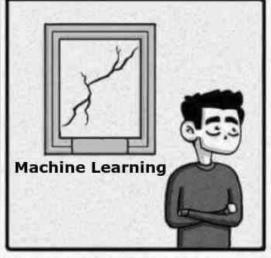
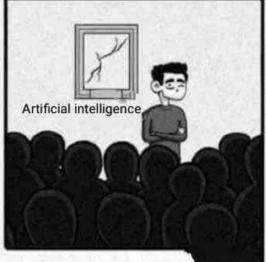
# The ethics of data collection for AI



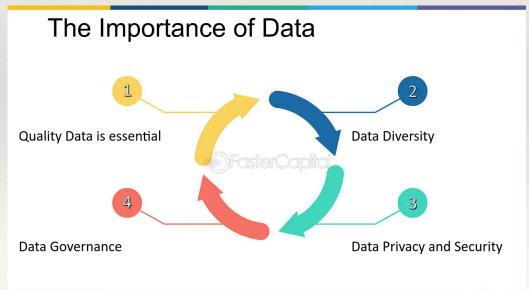






### Why data collection matters in AI

- \* Data is the foundation of AI systems, shaping their abilities and limitations.
- Collected data affects the accuracy, bias, and fairness of AI
- Ethical data collection is essential for building trustworthy and responsible AI systems.



### Privacy and consent

- \* Privacy: AI systems often require large datasets, raising concerns about user privacy.
- \* Informed Consent: Users must be aware of how their data is collected, stored, and used.
- \* Transparency: Organizations need to clarify data practices, including third-party sharing, retention, and anonymization



### Bias and Fairness

Bias in Data: AI learns from historical data, which can reflect societal biases.

**Impact on AI Decisions**: Biased data can lead to discriminatory outcomes (e.g., in hiring, law enforcement).

Mitigating Bias: Regular auditing, diverse data sources, and inclusive data practices can help reduce bias.



## Data security and protection

Security Risks: Data breaches can expose sensitive personal information

**Responsibility:** Organizations are ethically bound to protect collected data from unauthorized access.

**Compliance**: Adhering to standards (e.g., GDPR) helps ensure secure, compliant data handling practices.

### **EVERYTHING YOU NEED TO KNOW ABOUT**



7 GDPR DATA PROTECTION PRINCIPLES





### Storage limitations

Don't store personal data you do not need anymore



### **Purpose Limitations**

Only process personal data for the purpose it was intended for



### Lawfulness

Having a legal basis, being transparent and acting in the person's best interest



The data processor is responsible for complying with the GDPR

Accountability



### Data Minimization

Only gather and keep the exact amount of data that is needed



### Accuracy

Take "reasonable measures" to have the most accurate data possible



### Confidentiality

Only people who are processing the data should have access to it



### Equity and Access

**Data Representation:** Ensuring diverse, equitable data collection supports fairer AI systems.

**Digital Divide**: Recognize disparities in data access and avoid creating AI that disadvantages certain groups.

**Inclusive Data Practices**: Strive for balanced datasets to support equitable AI outcomes.

### The Nuremberg Code (1947)

- 1. The voluntary consent of the human subject is absolutely essential.
- 2. The experiment should be such as to yield fruitful results for the good of society.
- 3. The experiment should be so designed and based on the results of animal experimentation and a knowledge of the natural history of the disease.
- The experiment should be so conducted as to avoid all unnecessary physical and mental suffering and injury.
- 5. No experiment should be conducted where there is an a priori reason to believe that death or disabling injury will occur.
- 6. The degree of risk to be taken should never exceed that determined by the humanitarian importance of the problem to be solved by the experiment.
- 7. Proper preparations should be made and adequate facilities provided to protect the experimental subject against even remote possibilities of injury, disability, or death.
- 8. The experiment should be conducted only by scientifically qualified persons.
- 9. During the course of the experiment the human subject should be at liberty to bring the experiment to an end.
- 10. During the course of the experiment the scientist in charge must be prepared to terminate the experiment at any stage, if he has probable cause to believe, in the exercise of thegood faith, superior skill and careful judgment required of him that a continuation of the experiment is likely to result in injury, disability, or death to the experimental subject. (Link)

### The Belmon report (1979)

