

# ETHICAL DATA SCIENCE

## Systems Analysis

Author: Eng. Carlos Andrés Sierra, M.Sc.  
[cavirguezs@udistrital.edu.co](mailto:cavirguezs@udistrital.edu.co)

Lecturer  
Computer Engineer  
School of Engineering  
Universidad Distrital Francisco José de Caldas

2024-III



# Outline

1 What is ethics?

2 Study Cases

3 Large Language Models



# Outline

1 What is ethics?

2 Study Cases

3 Large Language Models



# Ethical Behaviours

Ethical behaviors are hard to define and deal. Sometimes it is hard just to explain why some actions could be not ethical.

$\rightarrow \frac{1}{2}$  dead  $\rightarrow \frac{1}{2} (1)$

$\rightarrow$  all  $\Rightarrow \frac{1}{2}$



Figure: Thanos



# Bad Use of Data

**Bad use of data** is a problem, ignore details for example. Even worse, use **unbalanced data** to make decisions.

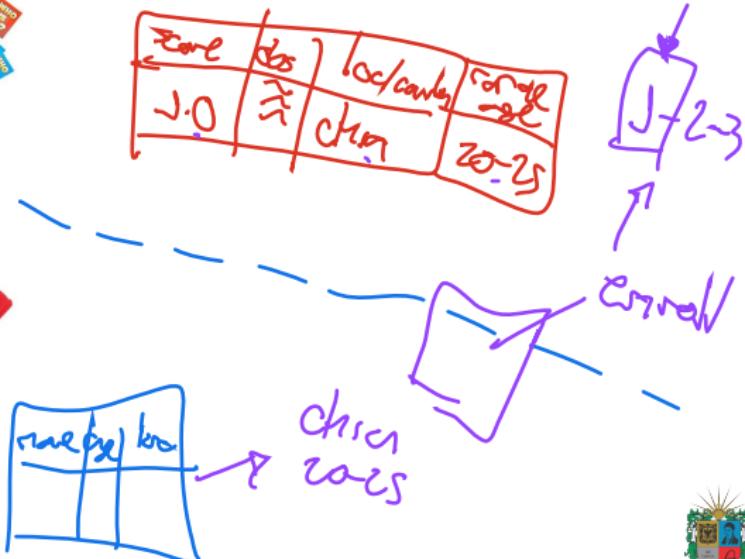


# Sharing Data

The share information is another problem right now. Even if you try to convert some part of the information into anonymous, someone could de-identified just making the right questions over the data.



Figure: Guess Who?



# Privacy in Modern Times

Privacy of the data is also a big problem. Internet and social networks are some sources of data filtration, you could be under attack at any moment.

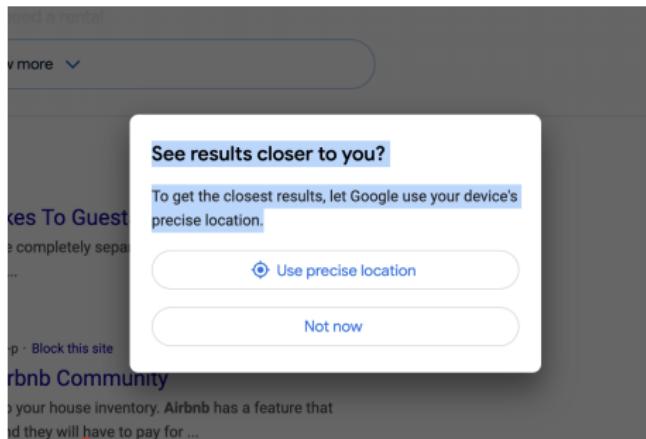
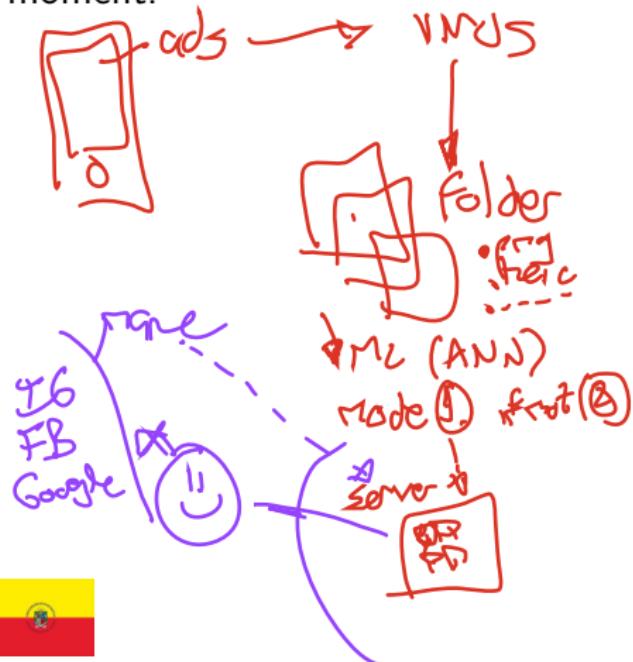


Figure: GPS all-the-time  
google

# Outline

1 What is ethics?

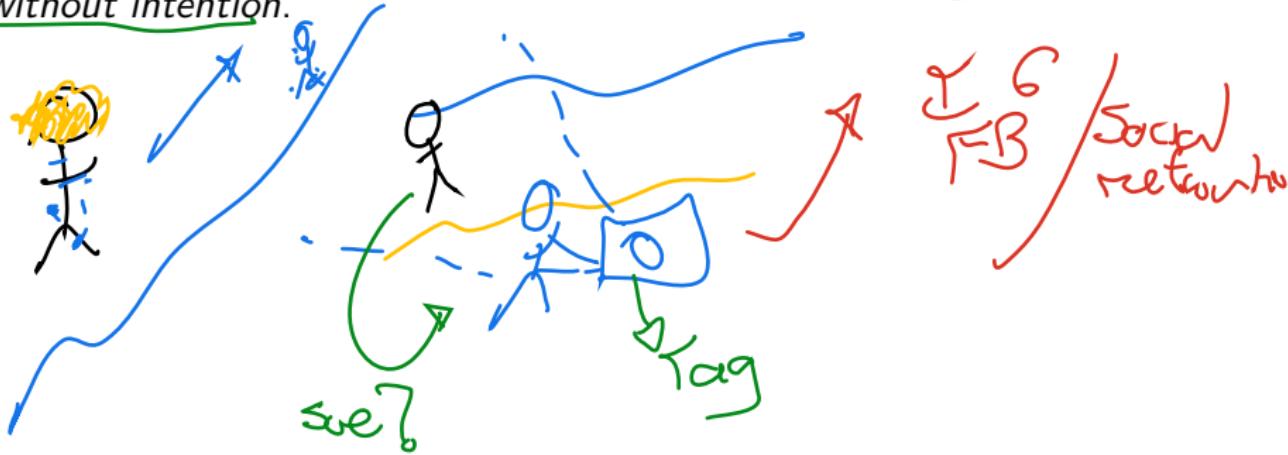
2 Study Cases

3 Large Language Models



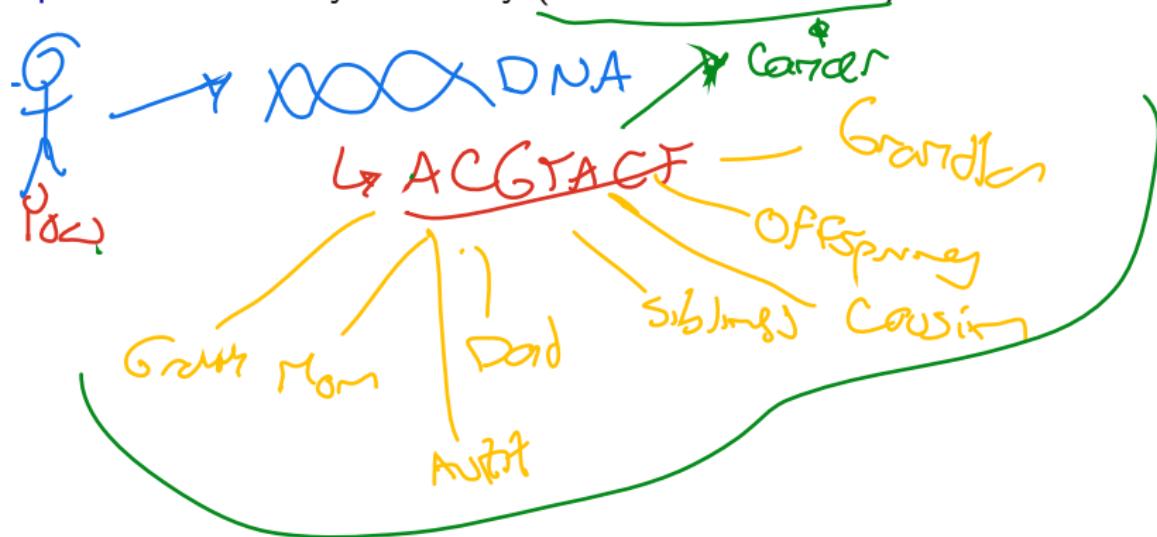
# Selfies in Social Networks

Nude photos even with distorted faces are not enough. Or a photo of you at the beach where more people appear, you are **exposing** more people without intention.



# DNA for a Good Cause

If you share your **DNA** for a *research project*, you are also **exposing** part of DNA information of **all your family**. Maybe it is not ethical unless you have **permission of all** your family (even the *dead ones*).



# Deepfakes, DeepVoices, a Deep Problem

Deepfakes, or any other kind of **supplantation** is a current problem, in **special to spread misinformation** or change people's mind in a bad way. Indeed, what is the *bad way*?



Figure: Javier Milei with GenAI



# Outline

1 What is ethics?

2 Study Cases

3 Large Language Models

Yann LeCun

L2M  $\Rightarrow$  La Ma  
↓  
AlphaGo



# Training Data, CopyRight, Laws, and AI

- Current artificial intelligence popular models, a.k.a. LLMs, are the current *attorneys preferred cases*.
- The train data used is the **first problem**: not is always public, not always is ethical to used. Also, a lot of legal conflicts with public and semi-public data. Who is the **owner** and how to respect its **creation rights**?
- Who is the owner of *AI outputs*? there is **no laws** for that, the copyright could be just attached to a human or company figure.

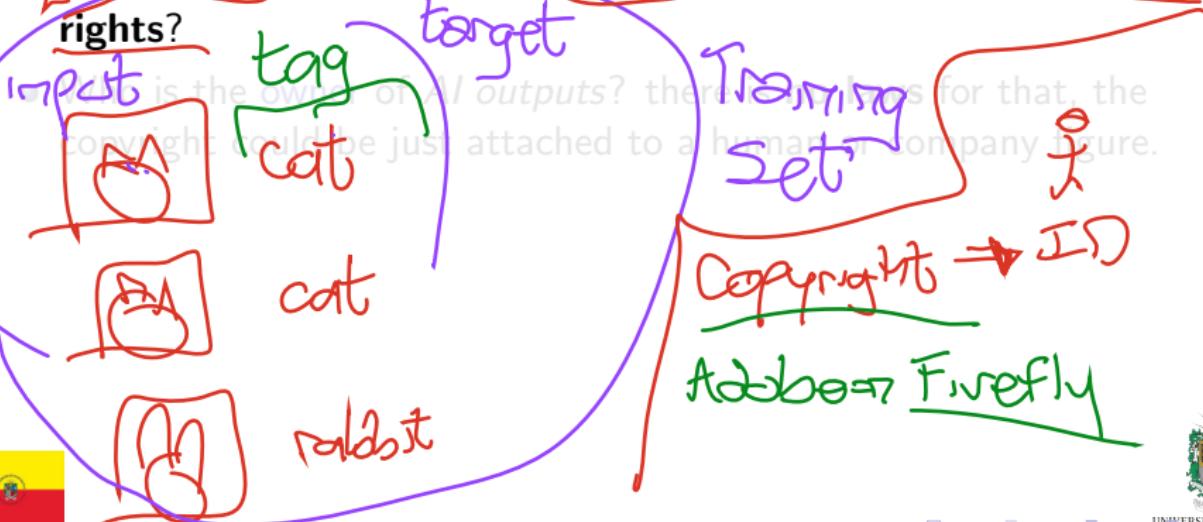


# Training Data, CopyRight, Laws, and AI

GPT-4 & 1 Trillion (20b)

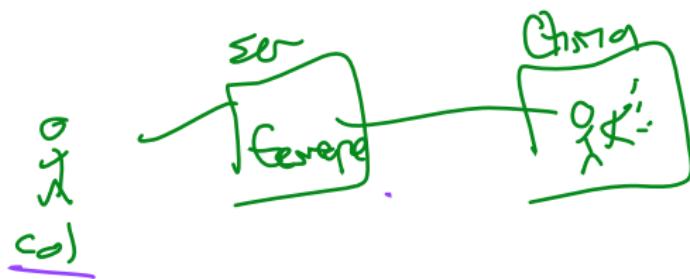
takers

- Current artificial intelligence popular models, a.k.a. LLMs, are the current *attorneys preferred cases*.
- The **train data** used is the **first problem**: not is always public, not always is ethical to used. Also, a lot of **legal conflicts** with public and semi-public data. Who is the **owner** and how to respect its **creation rights**?



# Training Data, CopyRight, Laws, and AI

- Current artificial intelligence popular models, a.k.a. LLMs, are the current *attorneys preferred cases*.
- The **train data** used is the **first problem**: not is always public, not always is ethical to used. Also, a lot of **legal conflicts** with public and semi-public data. Who is the **owner** and how to respect its **creation rights**?
- Who is the **owner** of **AI outputs**? there is **no laws** for that, the copyright could be just attached to a human or company figure.



## Picasso UD

*AD*

*Sole to OperAI* →

- If you study Picasso a lot, and learn his style, are you a **thief** or need to pay money to Picasso's family?
- Why does it is a problem in artificial intelligence?



Figure: Prompt: Draw the build of Universidad Distrital Francisco Jose de Caldas in Bogota-Colombia in a Picasso style.



# Outline

1 What is ethics?

2 Study Cases

3 Large Language Models



# Thanks!

## Questions?



Repo: <https://github.com/EngAndres/ud-public/tree/main/courses/systems-analysis>

