

# ETHICAL DATA SCIENCE

## Systems Analysis

Author: Eng. Carlos Andrés Sierra, M.Sc.  
[carlos.andres.sierra.v@gmail.com](mailto:carlos.andres.sierra.v@gmail.com)

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

2024-I



# Outline

- 1 What is ethics?
- 2 Study Cases
- 3 Large Language Models



## Outline

- ## 1 What is ethics?

## 2 Study Cases

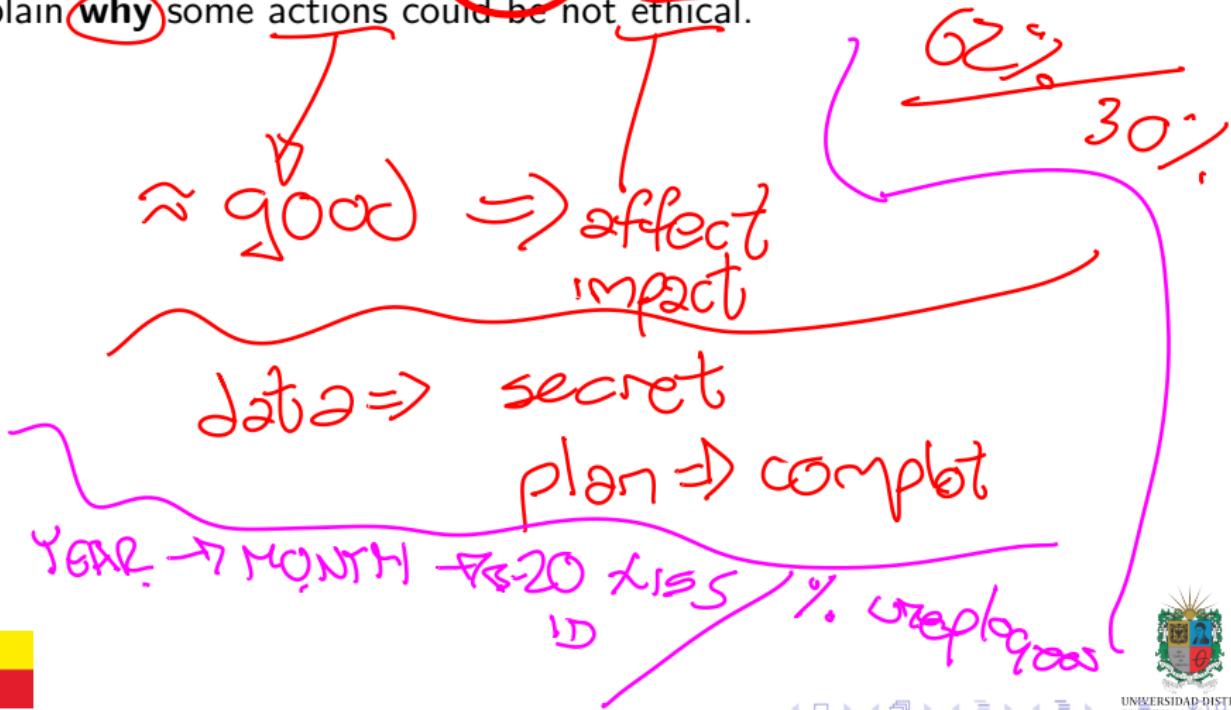
## ③ Large Language Models



MSc. C.A. Sierra (UD FJC)

# Ethical Behaviours

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



# Bad Use of Data

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

best target?

W-2

M-JT

1000000 → 2.7 clicks  
M.T=F DANG → COSES

5 4 3 2 100 100



# Sharing Data

The share information is a 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.

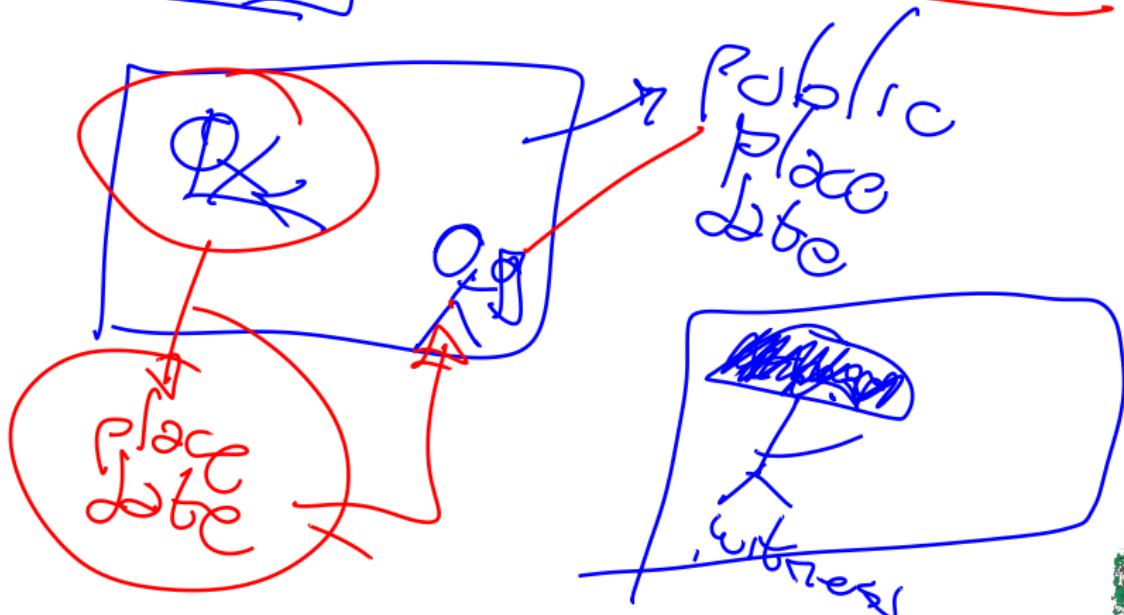
Gore,	Age long	Big Color
F	15-25	Purple

De-Identification



# Privacy in Modern Times

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



# Outline

1 What is ethics?

2 Study Cases

3 Large Language Models



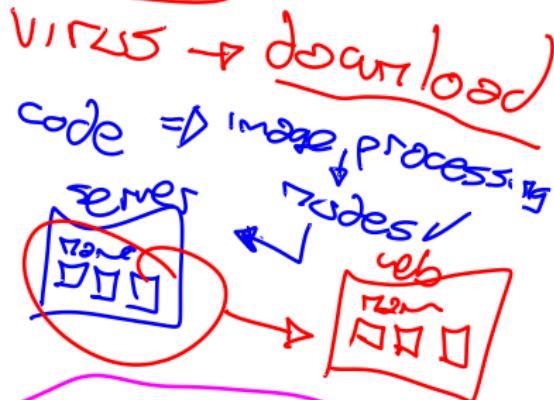
# Selfies in Social Networks

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.

lack of laws

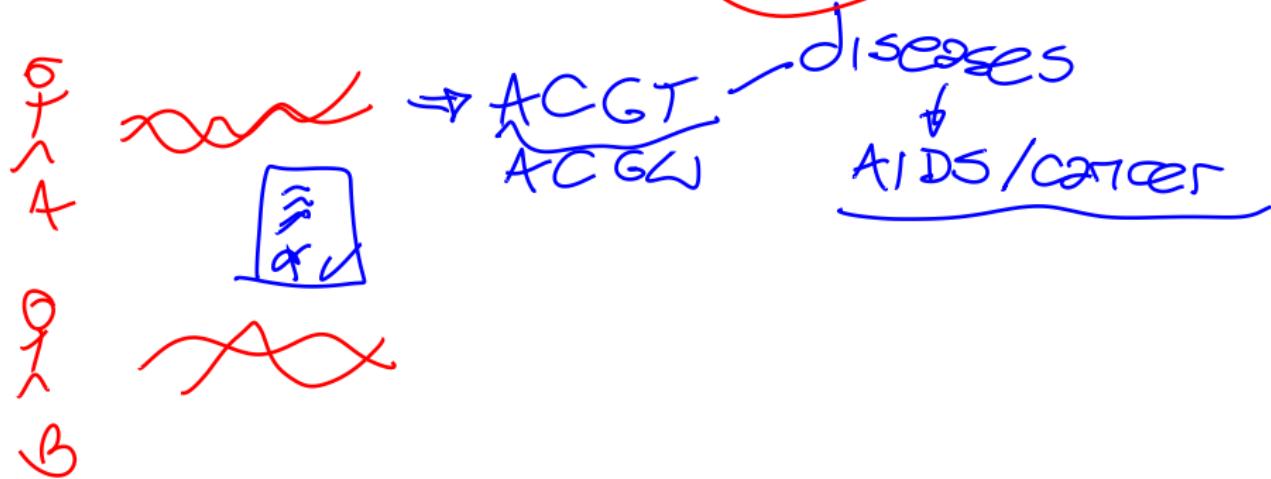


de-identification



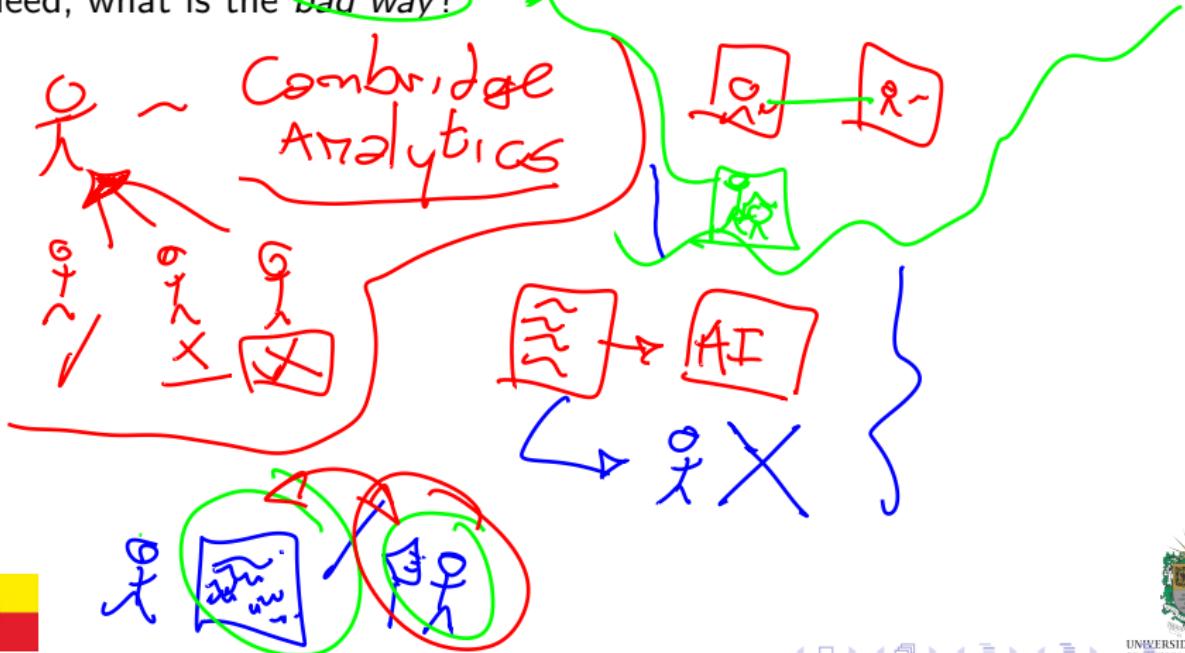
# DNA for a Good Cause

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 peoples' mind in a bad way. Indeed, what is the bad way?



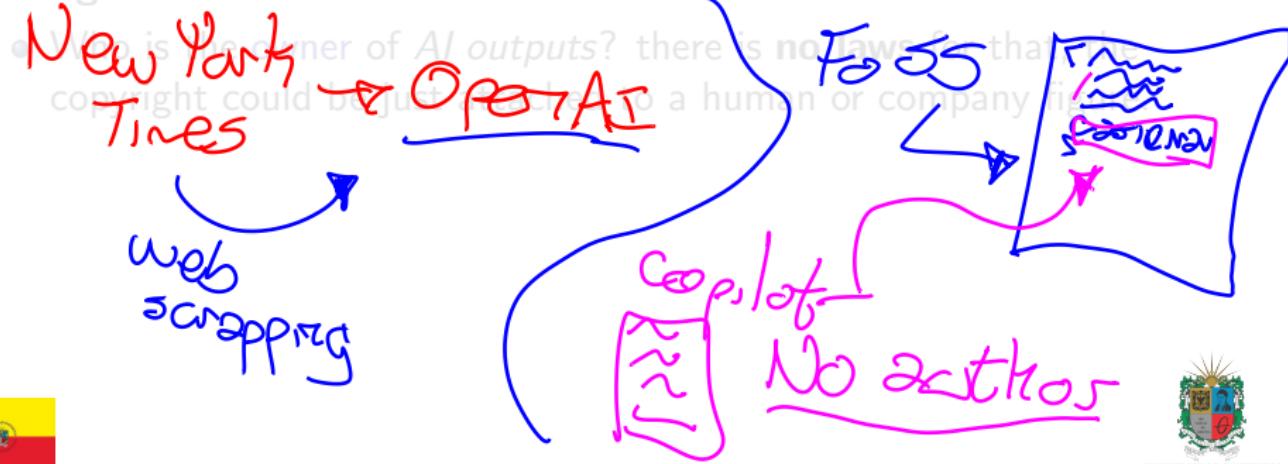
# Outline

- 1 What is ethics?
- 2 Study Cases
- 3 Large Language Models



# 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 always public, not always ethical to use. Who is the owner of legal conflicts with public and semi-public data. Who is the owner and how to respect their 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 output*? Who is responsible for that, the copyright could be just attached to a human or company figure.



# 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

- If you study Picasso a lot, and learn his style, are you a thief or need to pay money to Picassós 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>

