**LECTURE 7A 🡪 Algorithmic Bias (part 1)**

Bias is intentional but not conscious. It is an inclination or prejudice for or against one person or group, especially in a way considered to be unfair.

To understand algorithmic bias 🡪 understand human bias (point of this lecture)

In humans, biases are a trade-off between accuracy, speed and efficiency; the mind will make a decision without giving all facts due consideration

Biases are grouped into 4 major types:

* ACTION ORIENTED
* STABILITY
* PATTERN-RECOGNITION
* INTEREST BIAS (including SOCIAL BIAS)

Action Oriented Bias

Drive us to **speedy** actions by focusing attention and deflecting procrastination due to **self-doubt**

E.g. if someone does this, I think people like him will do that

**Overoptimism** and **overconfidence** are examples of that bias

Stability Bias

**Minimise** cognitive and physical efforts by gluing us to the status quo. (**efficiency)**

**Anchoring** is a specific bias that can compromise estimates by rooting them in completely random or seriously flawed reference points

E.g. write last 2 digits of phone number and make estimate of how much X costs. You tend to say a higher number if your last digits are high

Pattern Recognition bias

Leads to **flawed predictions** by either forming rules from random patterns or by applying rules inappropriately

E.g. if I wear a yellow tie on my tests and score high, it is a ‘lucky tie’

Also includes the **priming effect**, which is the ‘eche eche que beben las vacas?’

**Confirmation bias** is a particular pattern-recognition bias that compromises the data we consider when developing pattern-based rules. E.g. liking the journals that support your political views

Interest Bias

Comprise objective judgement by blending it with our own **self-interest**. They answer to the question: ‘What do I want’

**Social bias** are a particular type that make us agree with our social environment to fit better. E.g. agreeing with your boss

How do algorithms debias decisions I?