Intelligent Interactive Information Access Hub

DIPARTIMENTO
DI INGEGNERIA
DELL'INFORMAZIONE

Bias and Fairness in AI: New Challenges with Open Data?

Giorgio Maria Di Nunzio
Dept. of Information Engineering
University of Padua

Aura Network Workshop
Artificial Intelligence and Archives: What comes next?
16/03/2021

Outline

- Problems
- Definitions
- Possible Solutions
- Open Data in a "FAIR" Ecosystem

Background

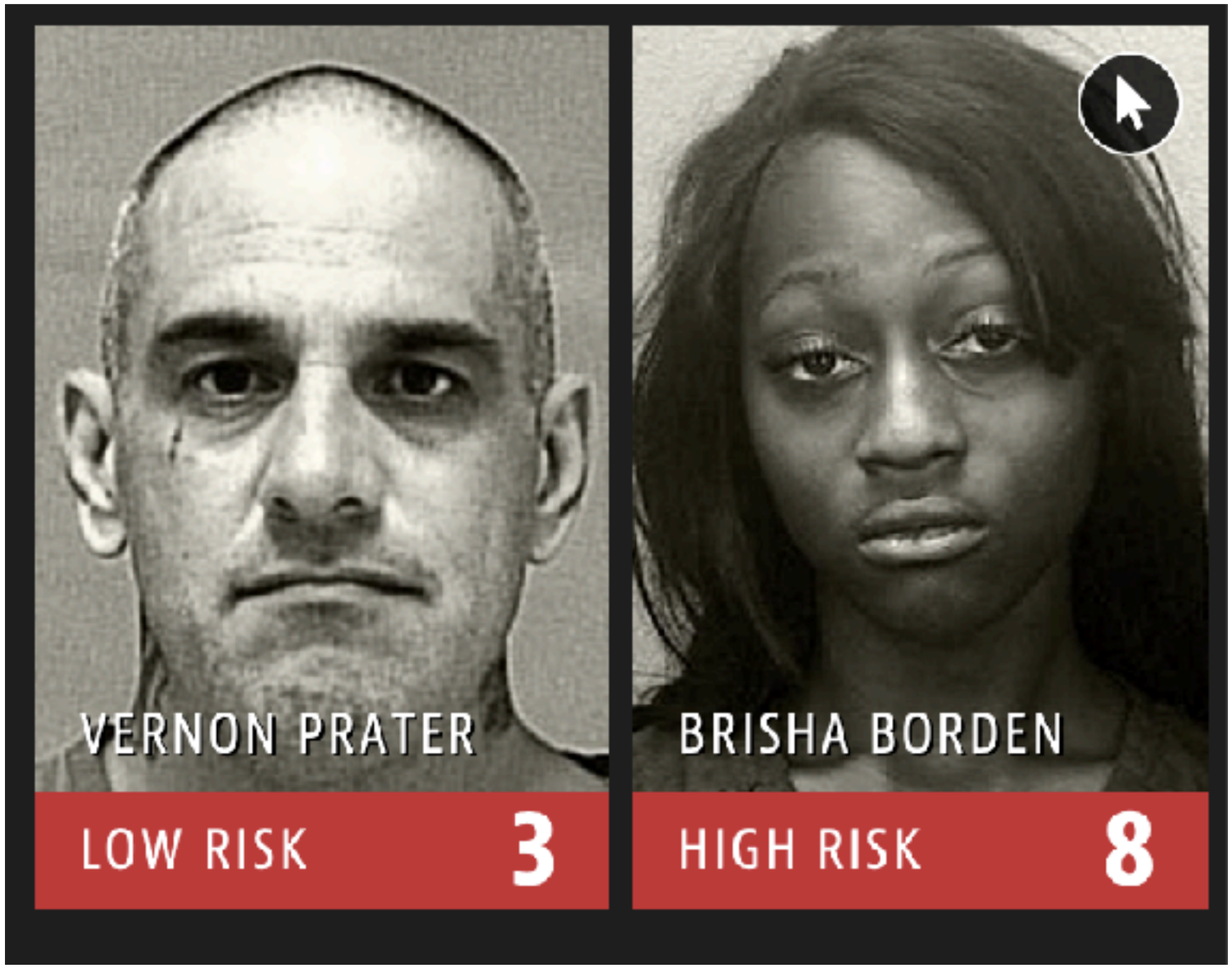
Explainability

Explainability

Explainability

Explainability

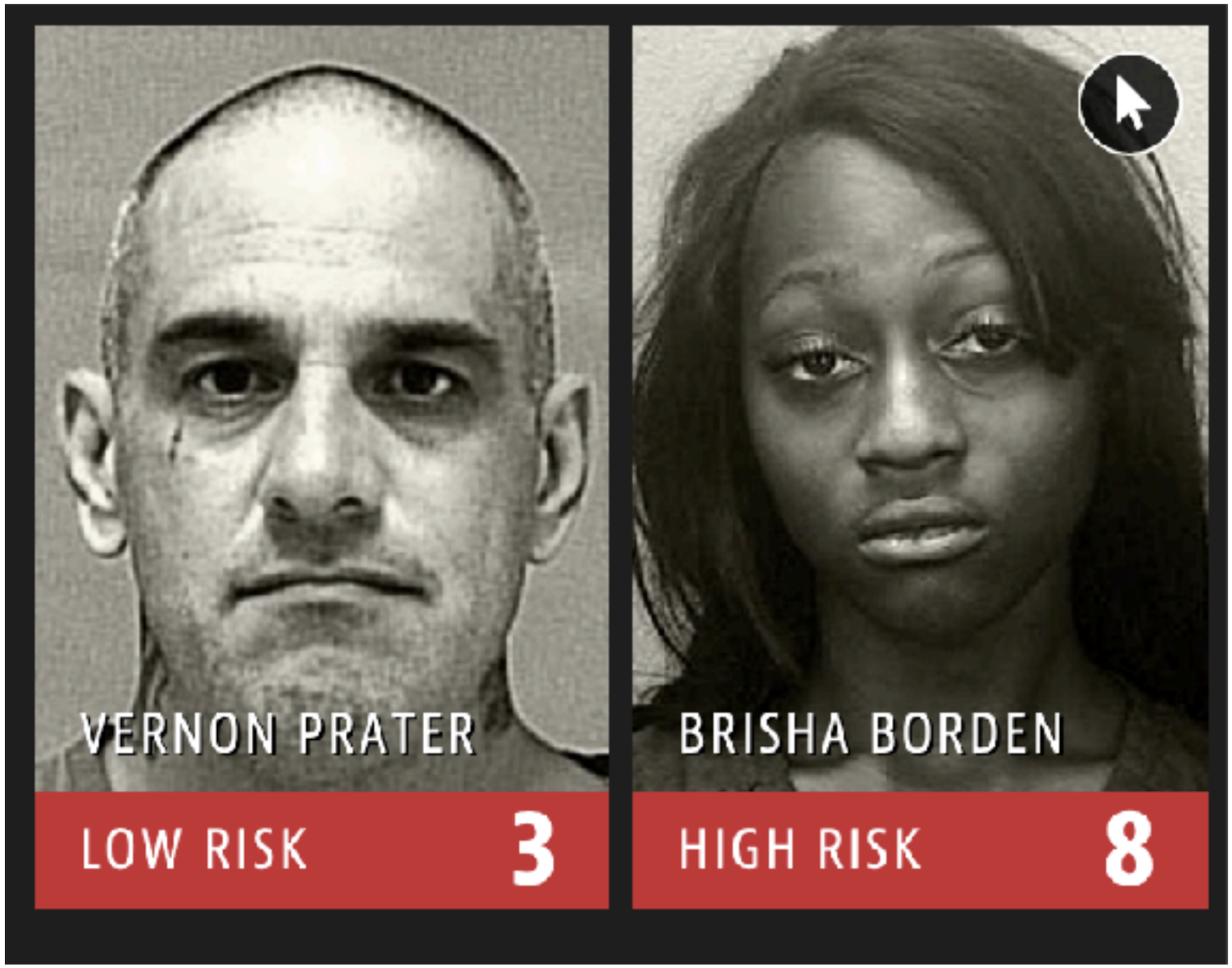
COMPAS System



Correctional Offender Management Profiling for Alternative Sanctions

COMPAS System

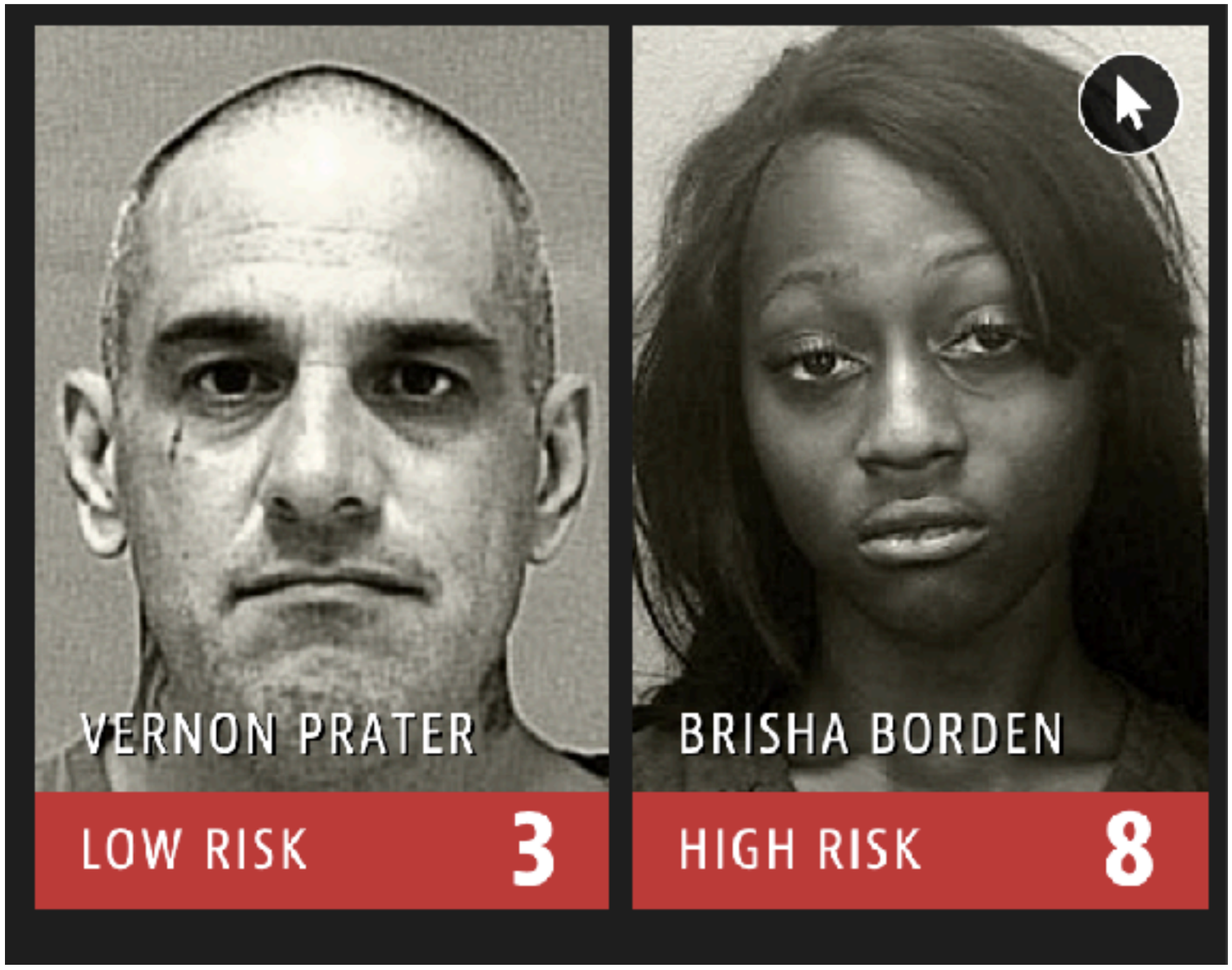
Who



Correctional Offender Management Profiling for Alternative Sanctions

COMPAS System

Who

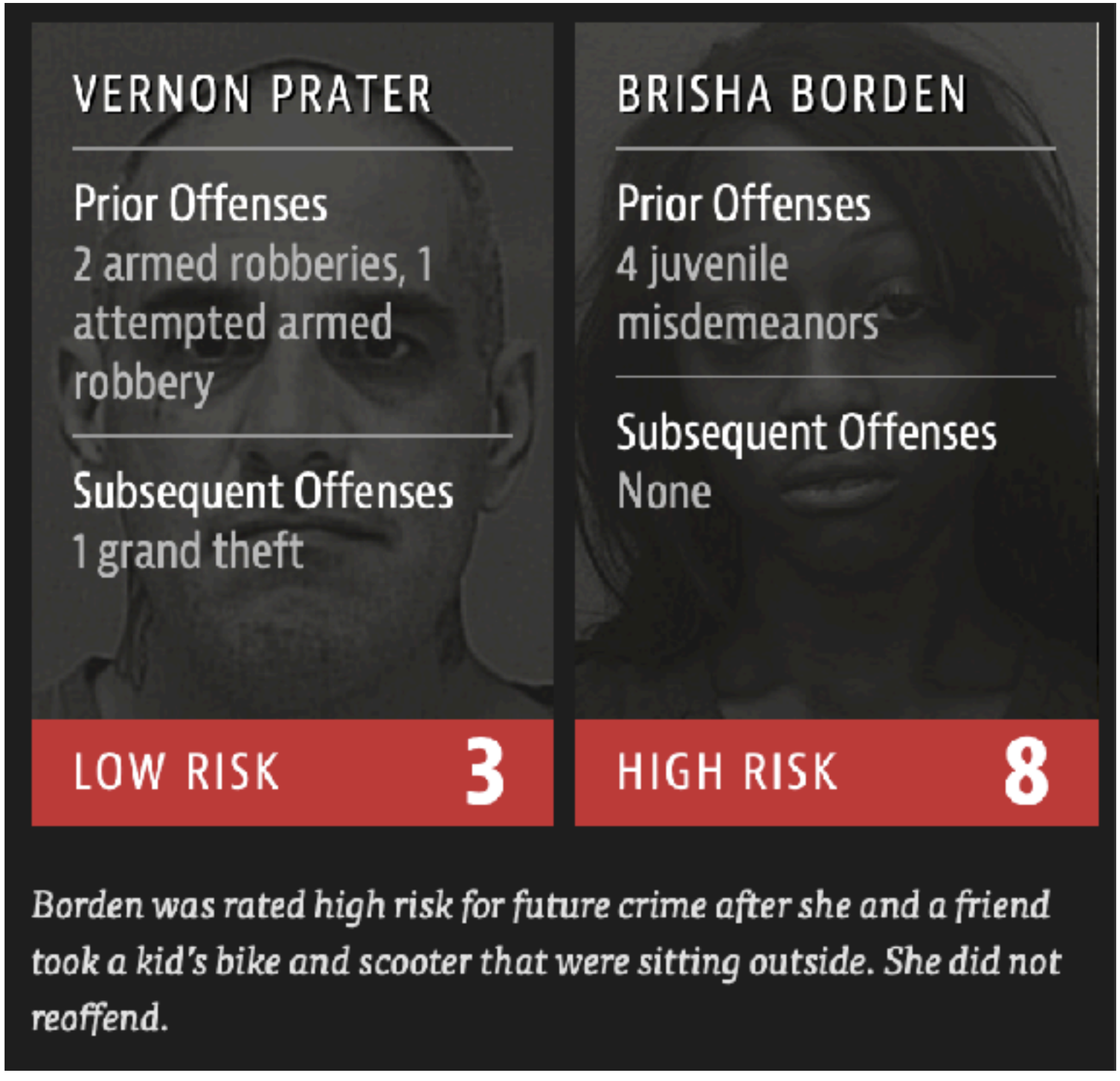


Why

Correctional Offender Management Profiling for Alternative Sanctions

COMPAS System

Who



Why

Problems

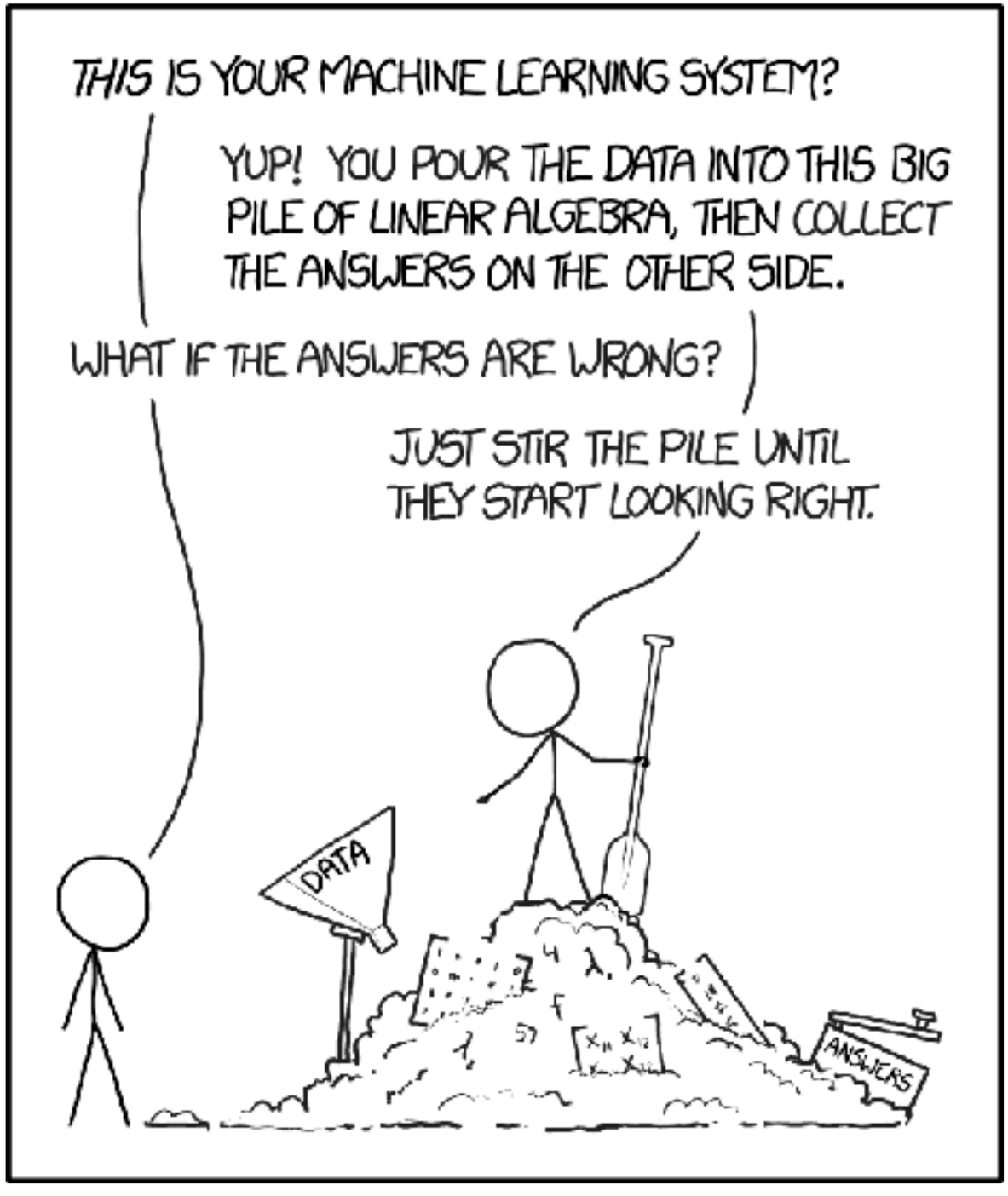
- Common distortions that can produce “unfairness”:
  - Bias encoded in data
  - Minimising overall error
  - Need to explore new data

Bias Encoded in Data

- Pedro Domingos (20 Dec. 2020):
  - “[…] machine-learning algorithms […] are essentially just complex mathematical formulas […], can’t be racist or sexist any more than the formula $y = a\,\boldsymbol{x} + b$ can.”
- Andrew Gelman (21 Dec. 2020):
  - “The point is that if **x** is biased, then $a\,\boldsymbol{x} + b$ will be biased too.”

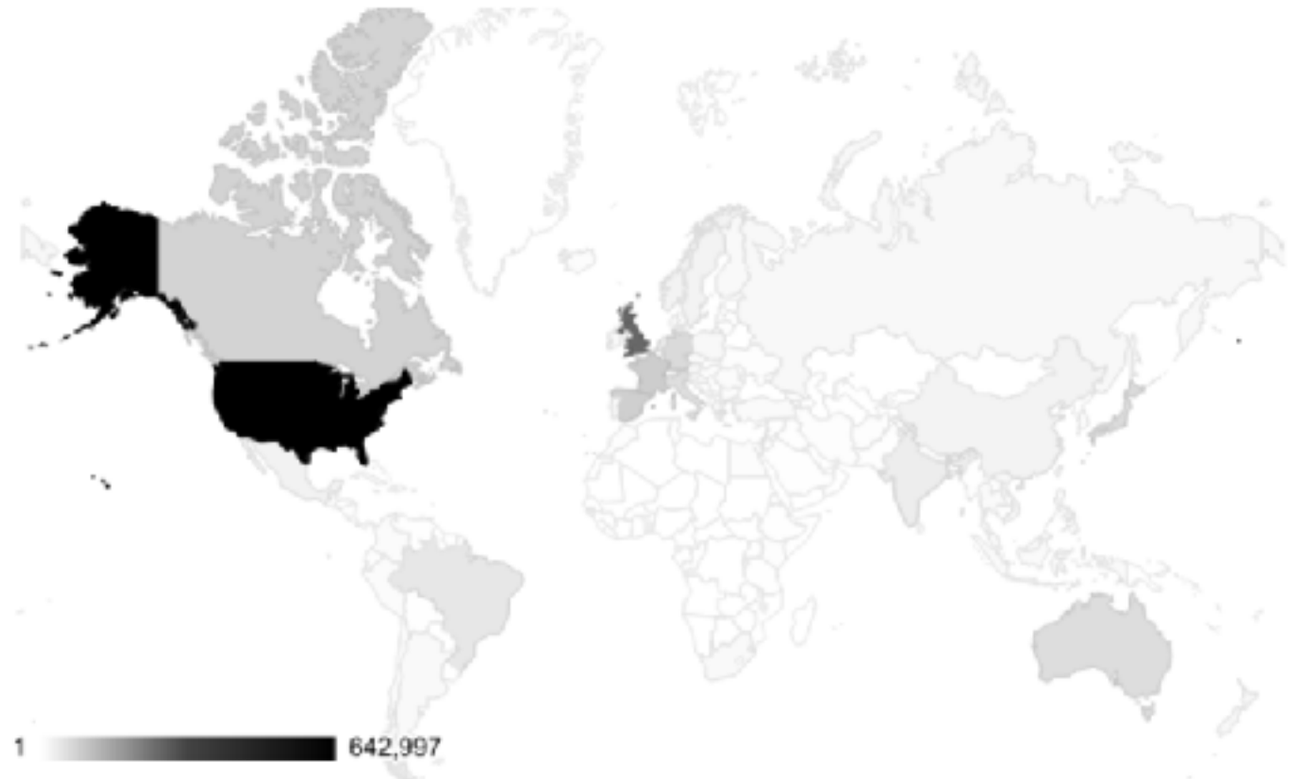
Bias Encoded in Data

- Pedro Domingos (20 Dec. 2020):
  - “[…] machine-learning algorithms […] are essentially just complex mathematical formulas […], can’t be racist or sexist any more than the formula $y = a\,\boldsymbol{x} + b$ can.”
- Andrew Gelman (21 Dec. 2020):
  - “The point is that if **x** is biased, then $a\,\boldsymbol{x} + b$ will be biased too.”



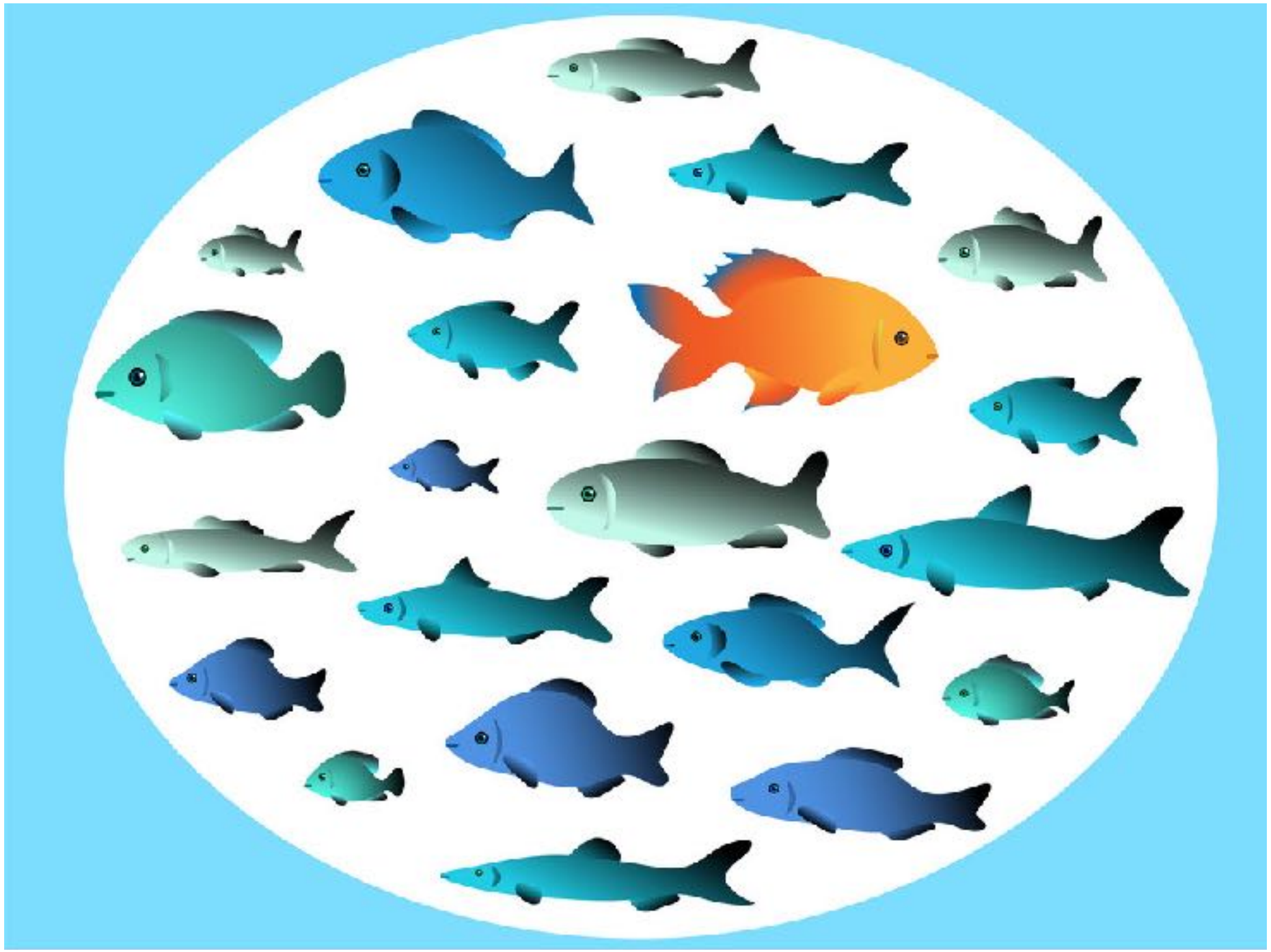
Bias Encoded in Data

- Historical Bias
- Representation Bias
- Measurement Bias
- ….



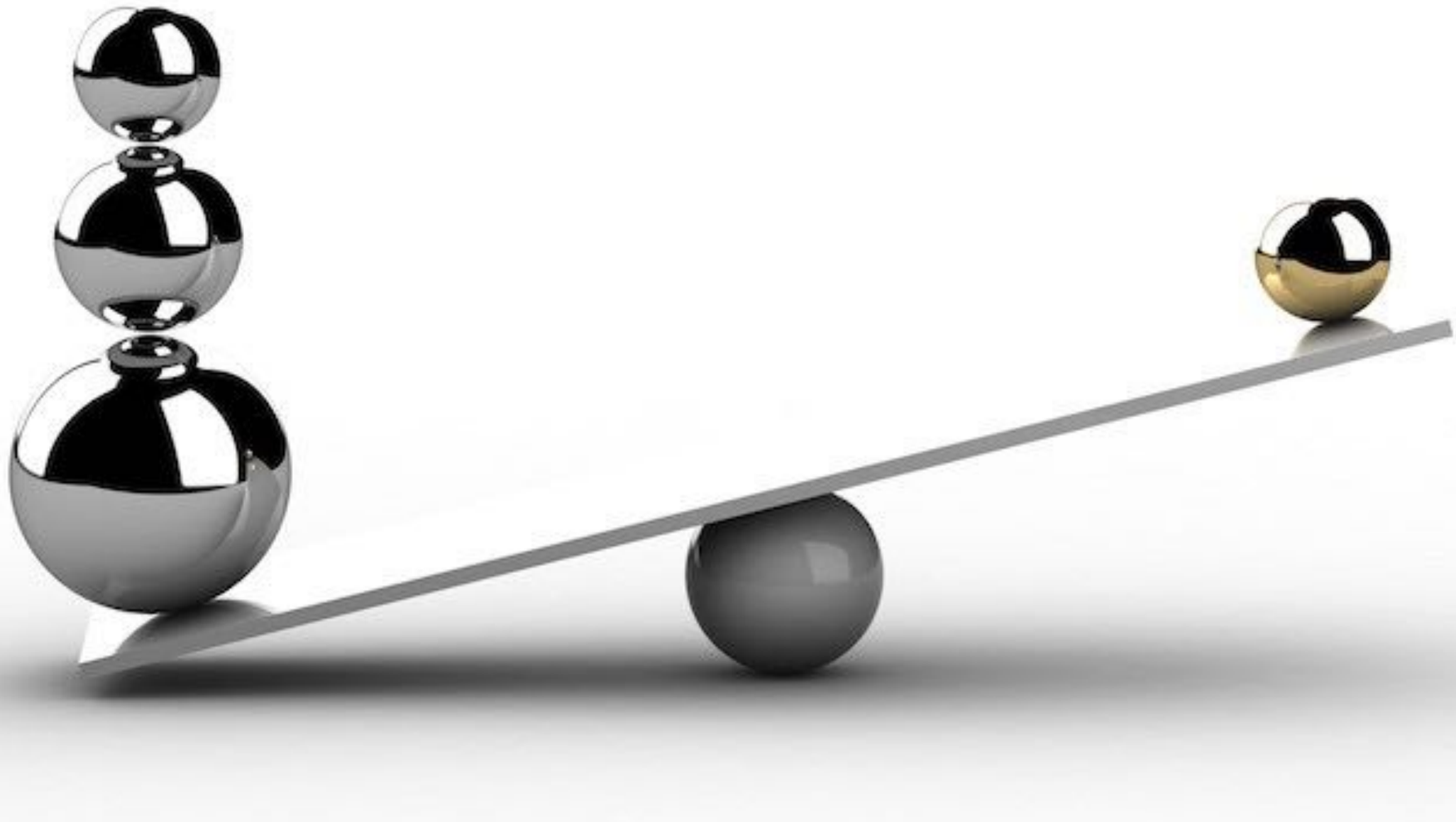
ImageNet Database

Minimising Overall Error



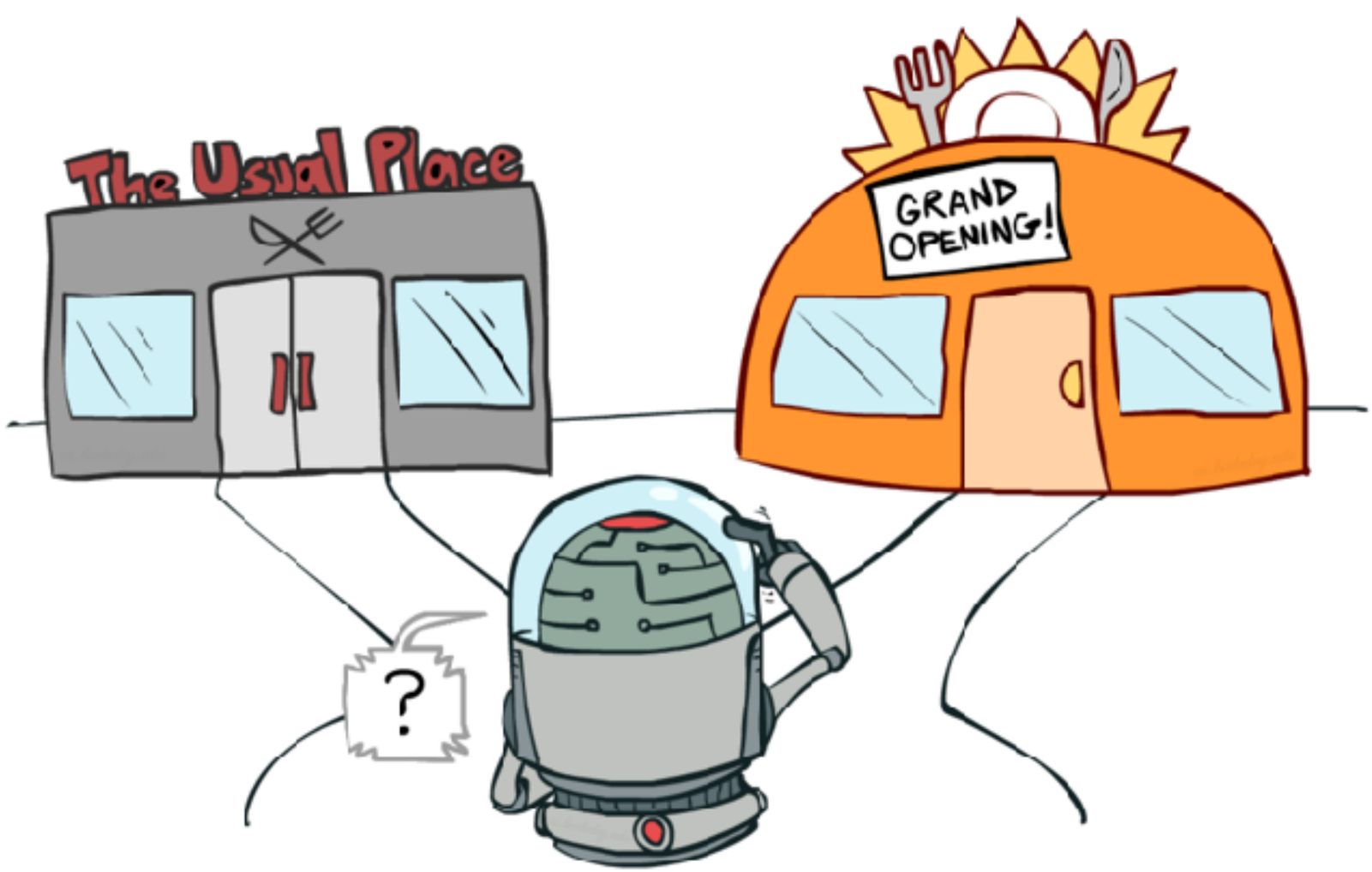
Minimising Overall Error

- A group-blind classifier trained to minimise the overall error, it will fit the majority population



Need to Explore

- In order to effectively learn, we need to explore - that is, sometimes take actions we believe to be sub-optimal in order to gather more data.



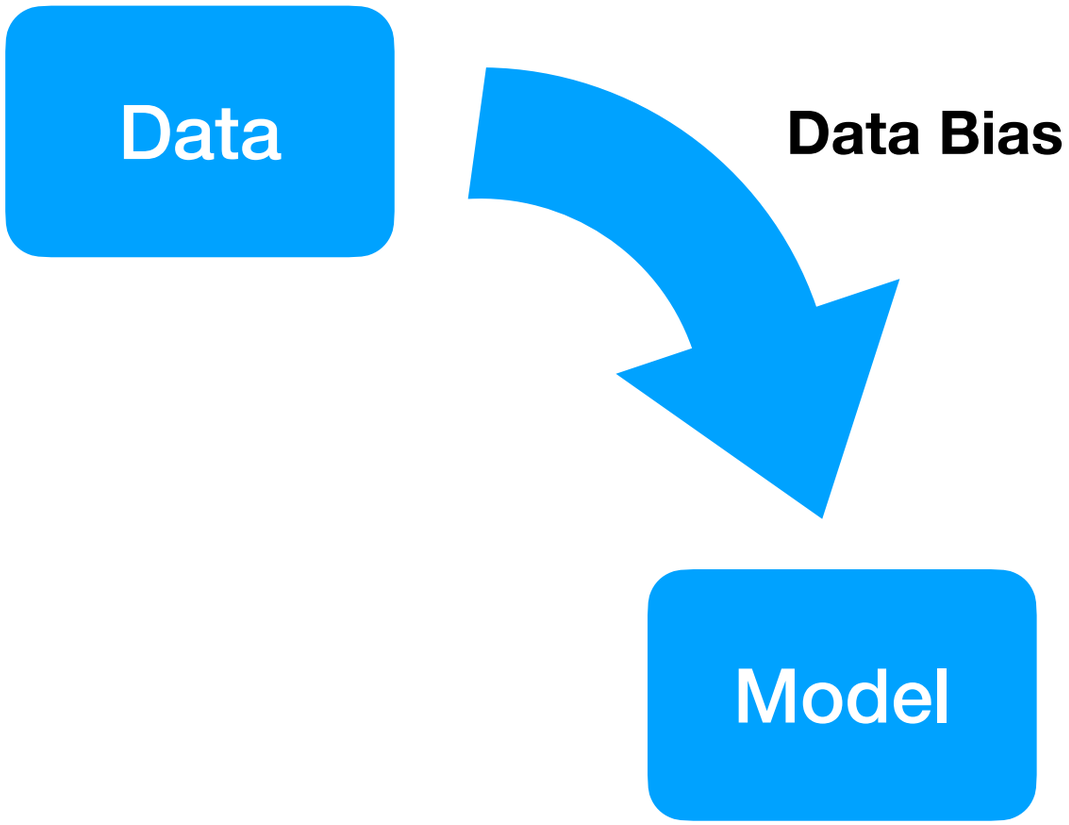
Need to Explore

- At least two (ethical?) questions:
  - Are the costs of exploring a certain sub-population worth it?
  - Are some actions “immoral”? If so, how much does this slow learning, and does this lead to other sorts of unfairness?

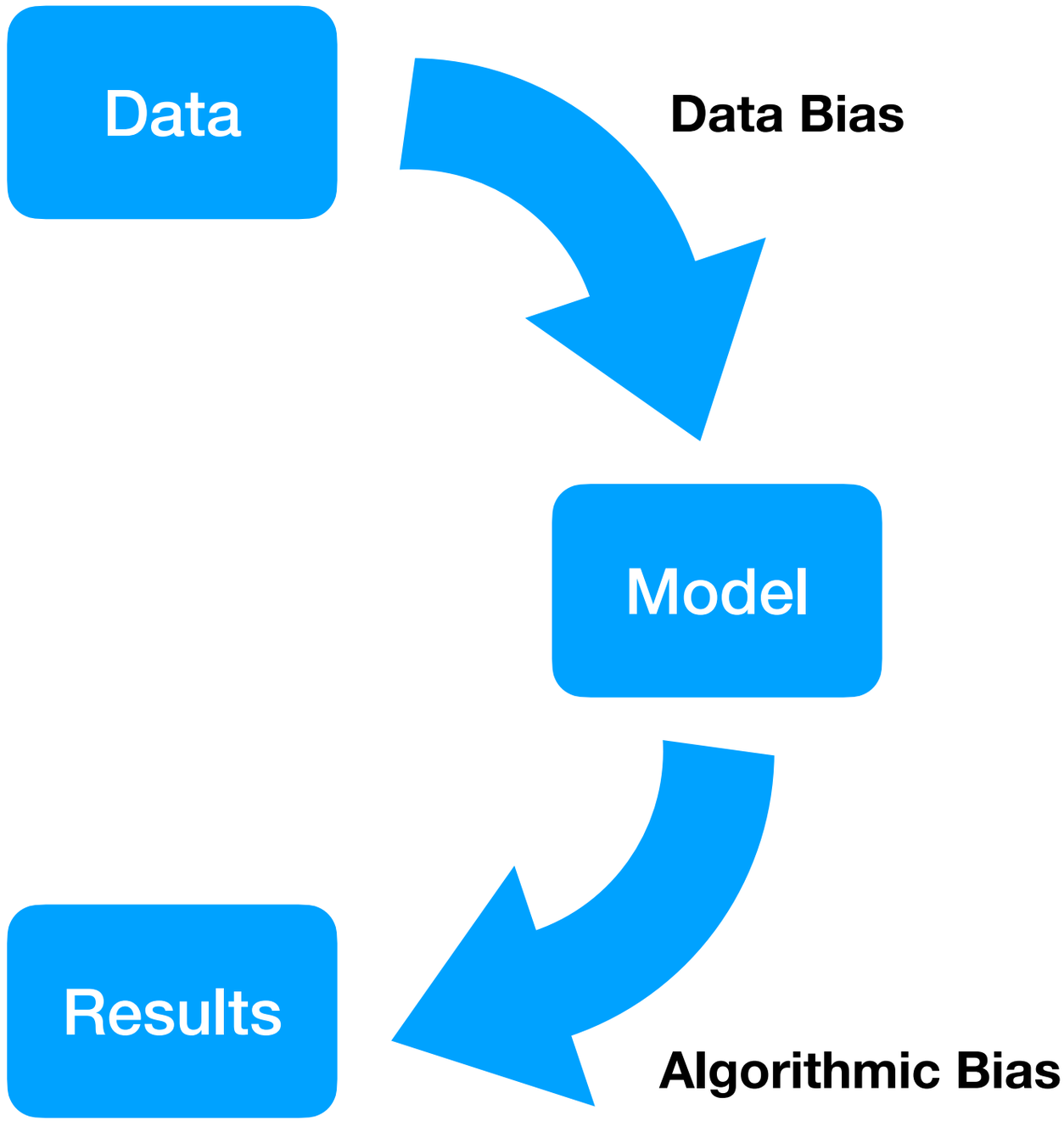
(Vicious) Cycle of Bias

Data

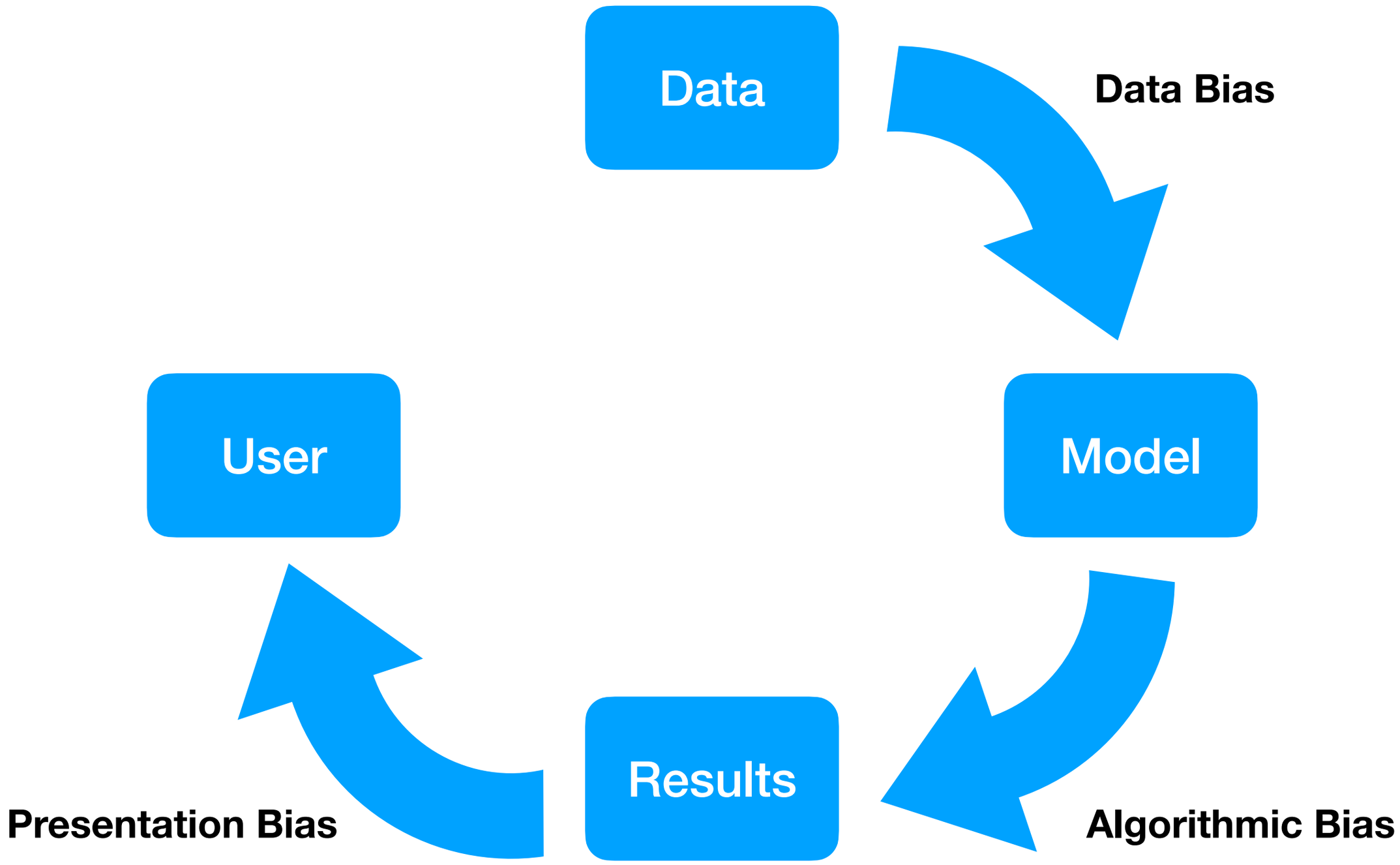
(Vicious) Cycle of Bias



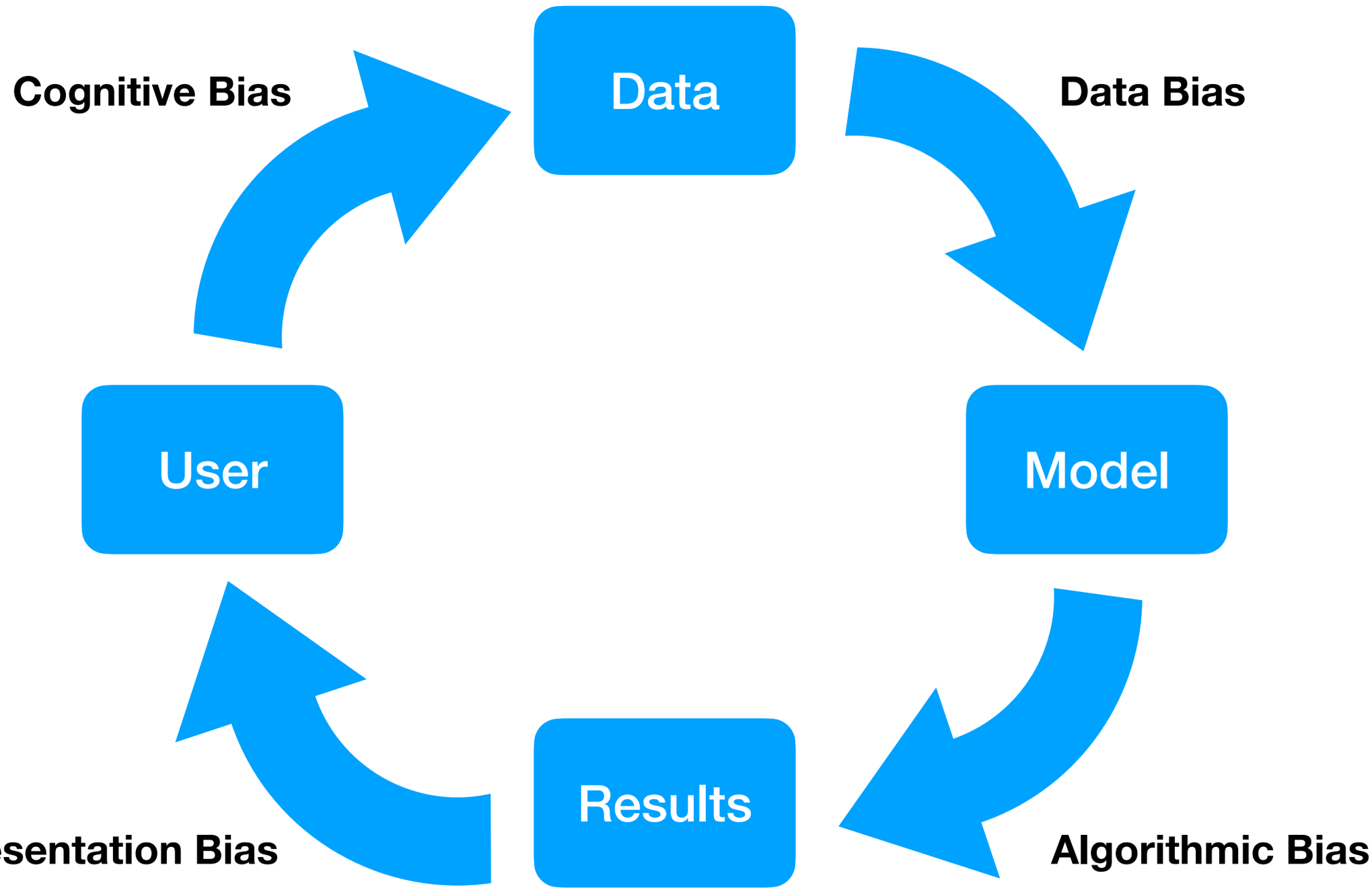
(Vicious) Cycle of Bias



(Vicious) Cycle of Bias



(Vicious) Cycle of Bias



Definitions of Fairness

- "21 fairness definition and their politics", ACM FAccT 2018
    - Conference on Fairness, Accountability and Transparency

Definitions of Fairness

- Statistical notion: Fix a small number of protected demographic groups G (such as ethnic groups), and then ask for (approximate) parity of some statistical measure across all of these groups.

- Individual notion: Ask for constraints that bind on specific pairs of individuals, rather than on a quantity that is averaged over groups.

Possible Solutions

- Best of both worlds?
- Constraints that are practically implementable without the need for making strong assumptions on the data or the knowledge of the algorithm designer?

- Fairness feedback is given by human beings who may not be responding in a way that is consistent with any metric?

Fairness Tree

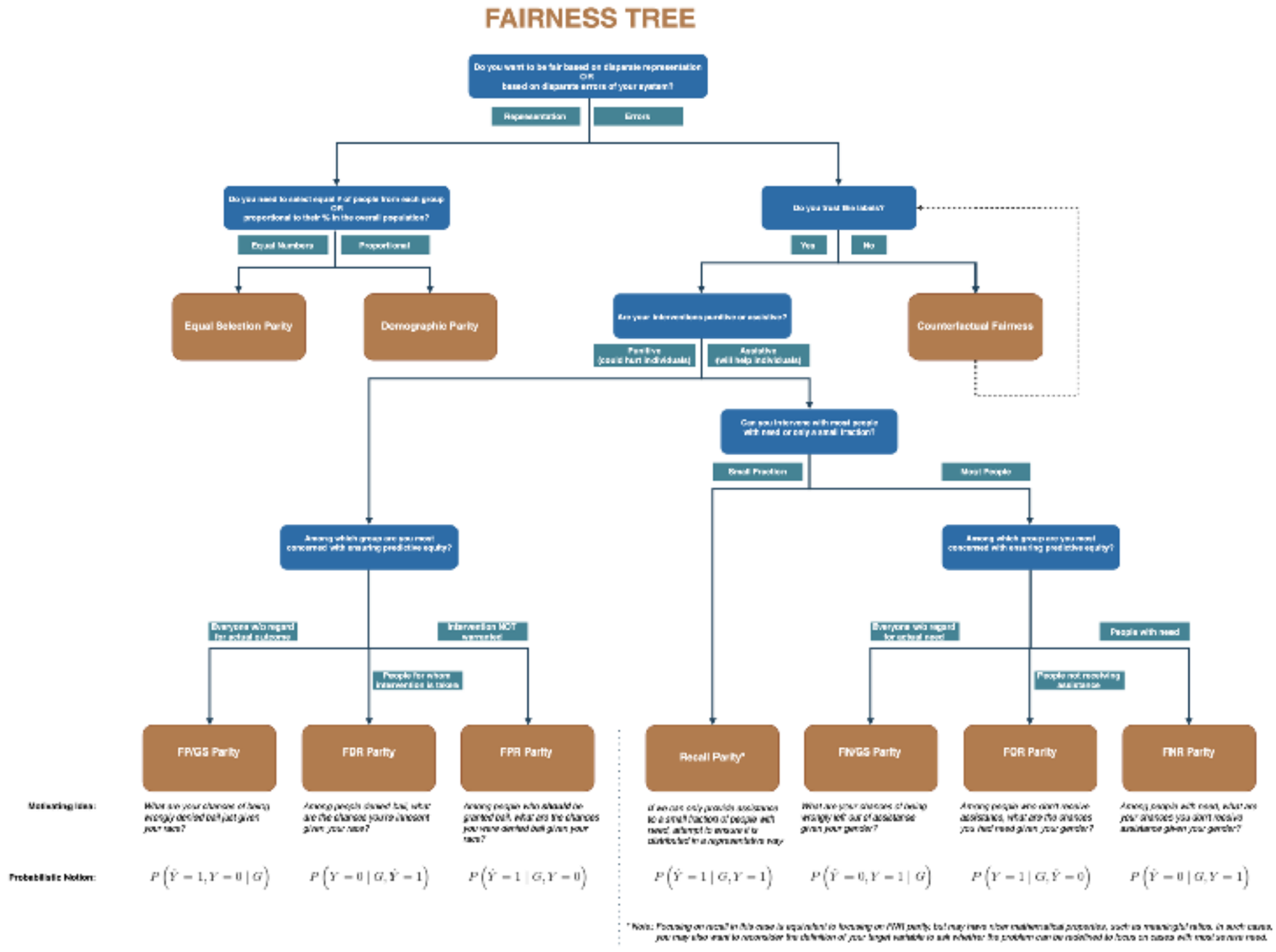
http://www.datasciencepublicpolicy.org/projects/aequitas/

Aequitas

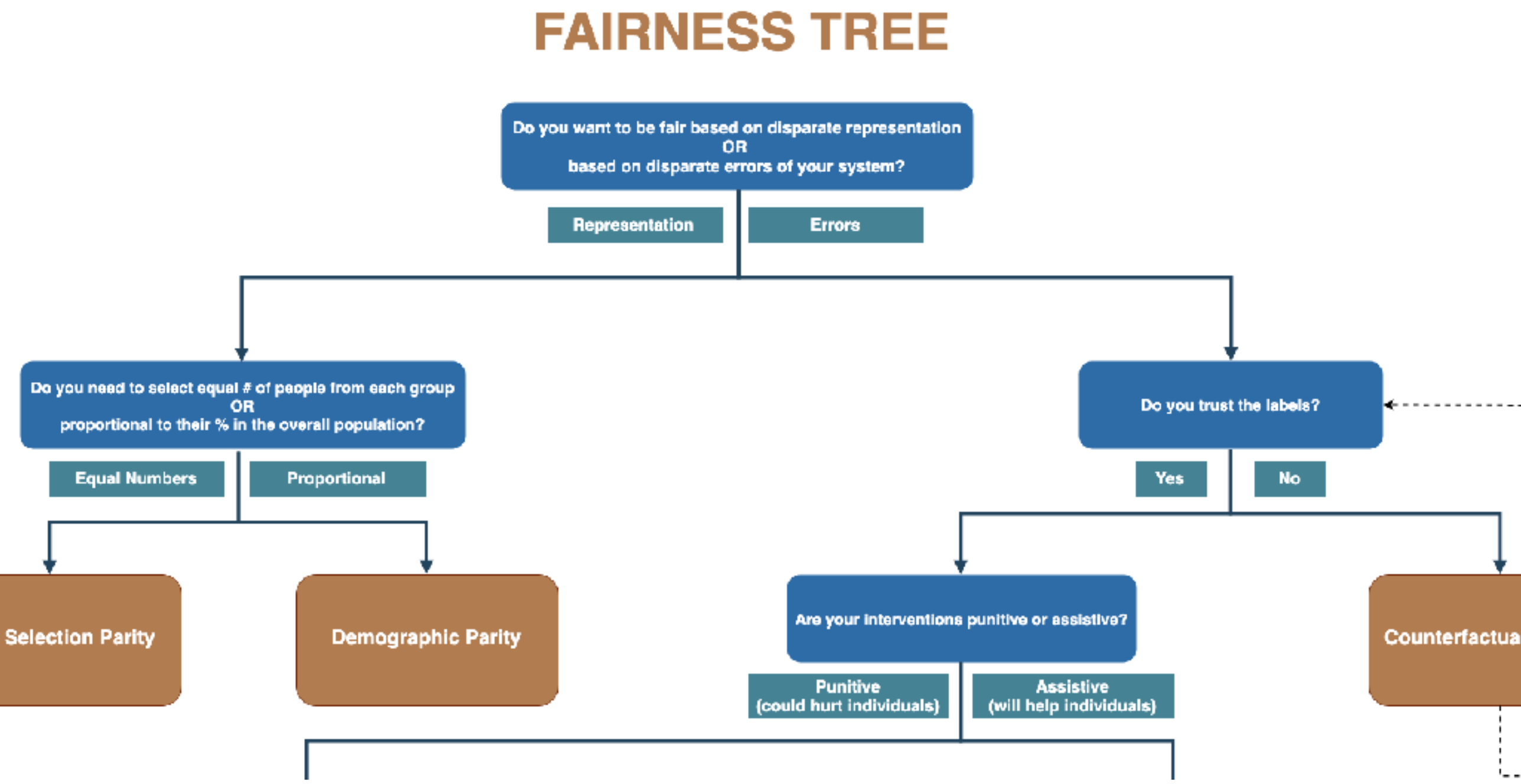
Bias & Fairness Audit

Fairness Tree

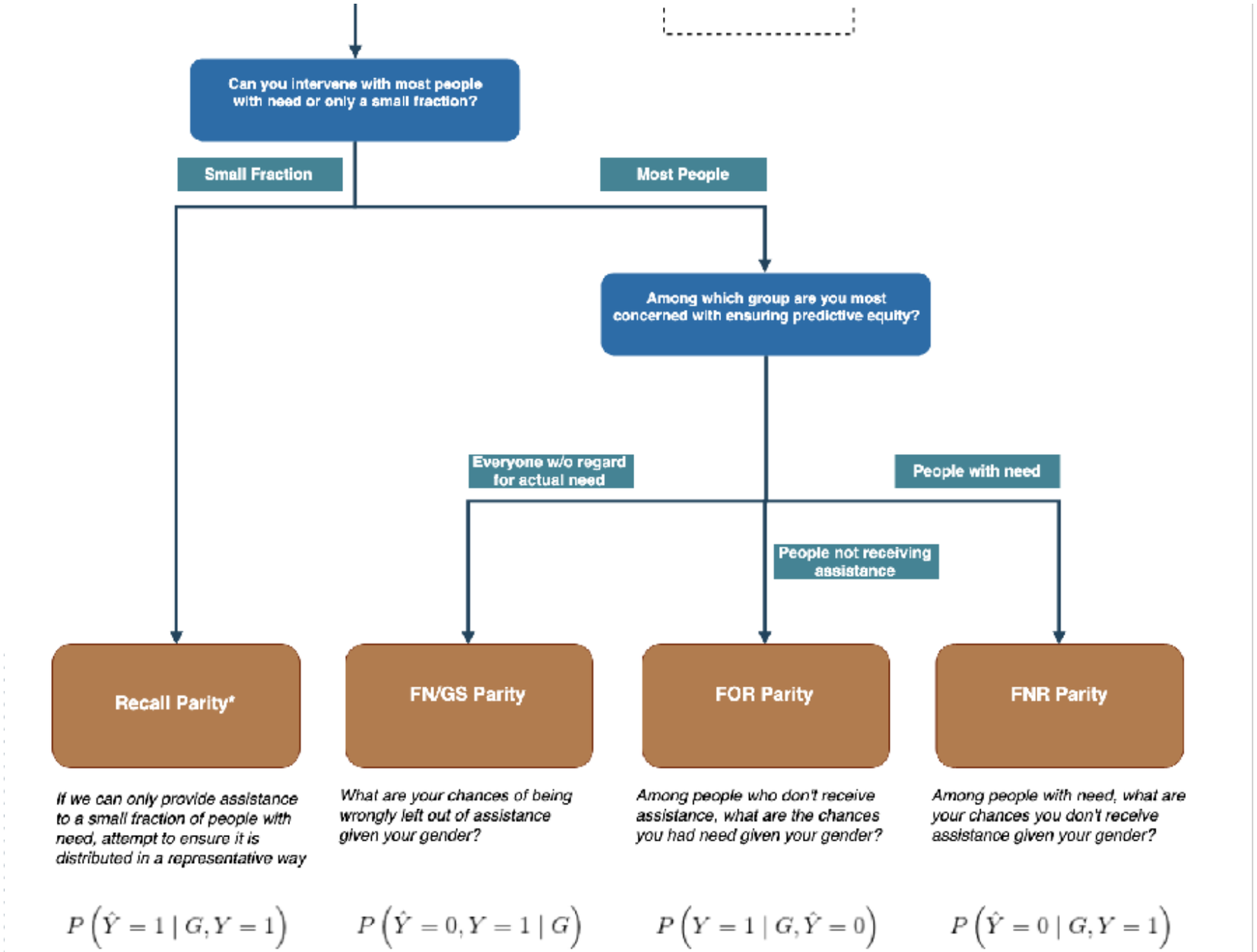
http://www.datasciencepublicpolicy.org/projects/aequitas/



Fairness Tree



Fairness Tree



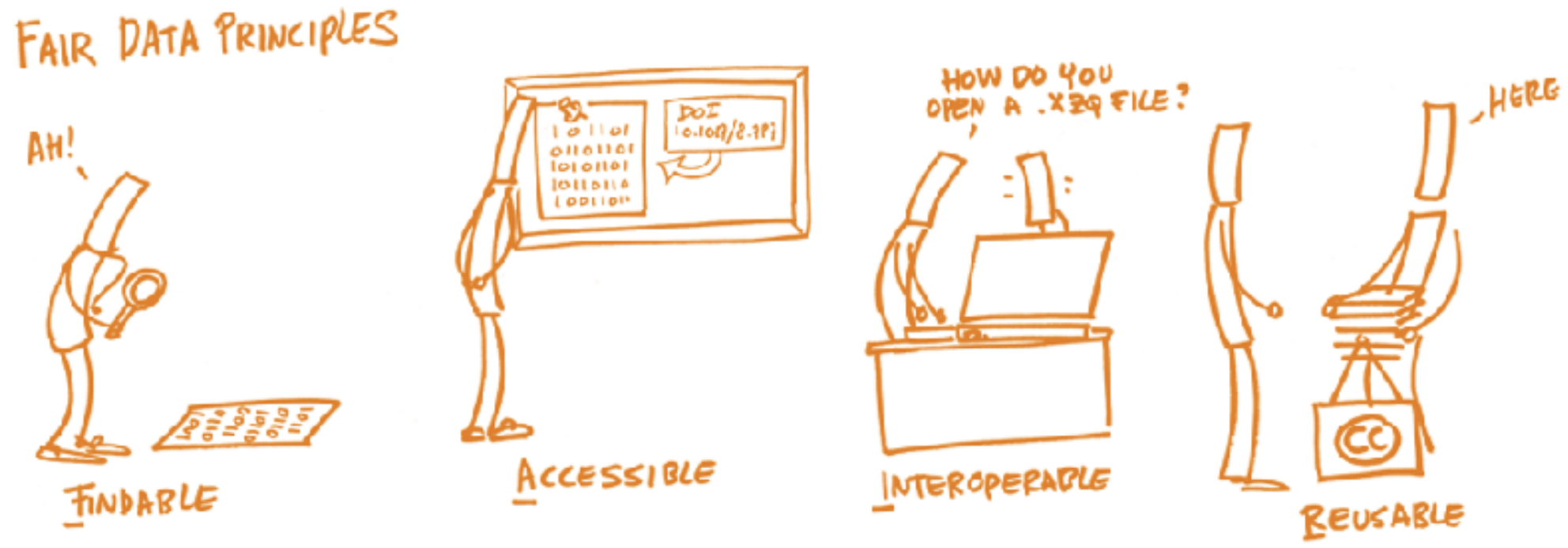
* *Note: Focusing on recall in this case is equivalent to focusing on FNR parity, but may have nicer mathematical properties, such as meaningful ratios. In such cases, you may also want to reconsider the definition of your target variable to ask whether the problem can be redefined to focus on cases with most severe need.*

What's Next with FAIRness?

EUROPEAN OPEN SCIENCE CLOUD

- Open Innovation, Open Science and Open to the World
- Implementation of a "Web of FAIR data"
  - to find, exploit, and combine linked datasets, leading to discoveries and research paradigms, and have an impact on the territory and citizens

What's Next with FAIRness?



- "The FAIR principles, although inspired by Open Science, explicitly and deliberately do not address moral and ethical issues pertaining to the openness of data. […] the degree to which any piece of data is available, or even advertised as being available (via its metadata) is entirely at the discretion of the data owner. FAIR only speaks to the need to describe a process – mechanised or manual – for accessing discovered data […]"

Conclusions

- The data itself is often a product of social and historical process that operated to the disadvantage of certain groups.

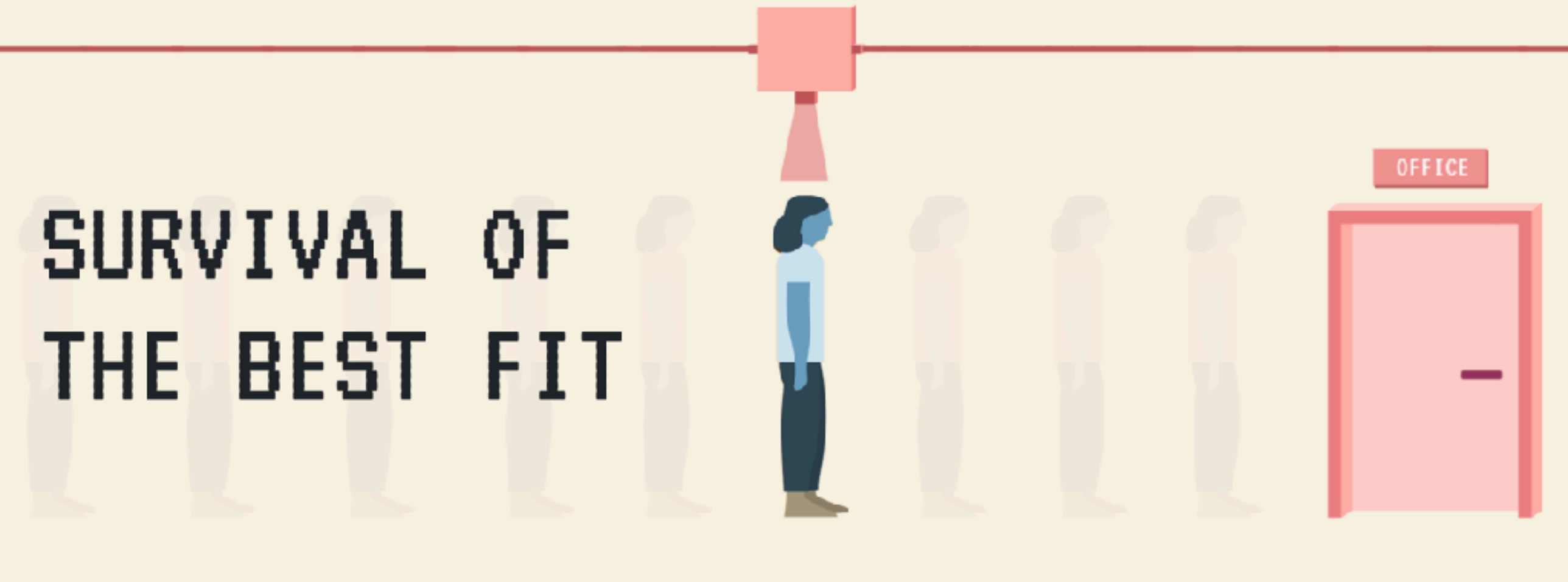
- Understanding how bias arises in the data, and how to correct for it, are fundamental challenges in the study of fairness in AI.

- Similar risks arise whenever there is potential for feedback loops.

- Correcting for data bias generally seems to require knowledge of how the measurement process is biased, or judgments about properties the data.

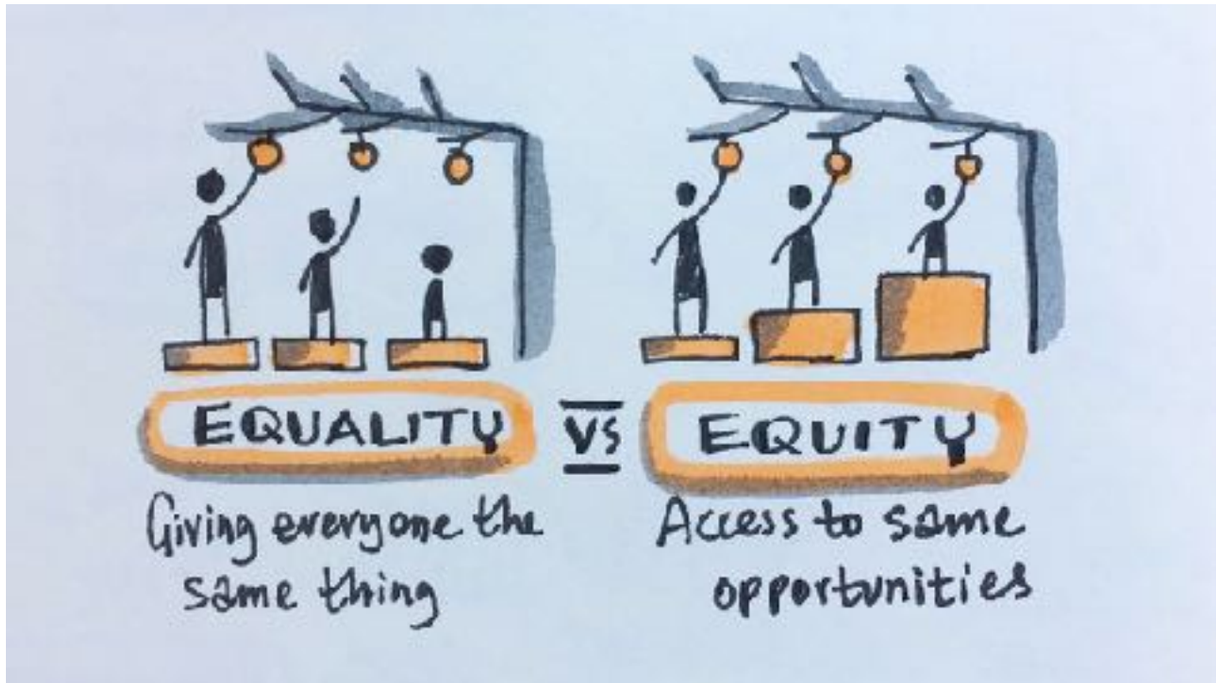
If you think you are not biased...

https://www.survivalofthebestfit.com



Bias and Fairness in AI: New Challenges with Open Data?

Thank you!



giorgiomaria.dinunzio@unipd.it
http://iiia.dei.unipd.it

Aura Network Workshop
Artificial Intelligence and Archives: What comes next?
16/03/2021