



Trustworthy AI: Fairness



Destiny Pounds

M.S. Student, Biomedical Data Science

Advised by Vibhuti Gupta, Ph.D.

Assistant Professor, Computer Science and Data Science

School of Applied Computational Sciences

Meharry Medical College

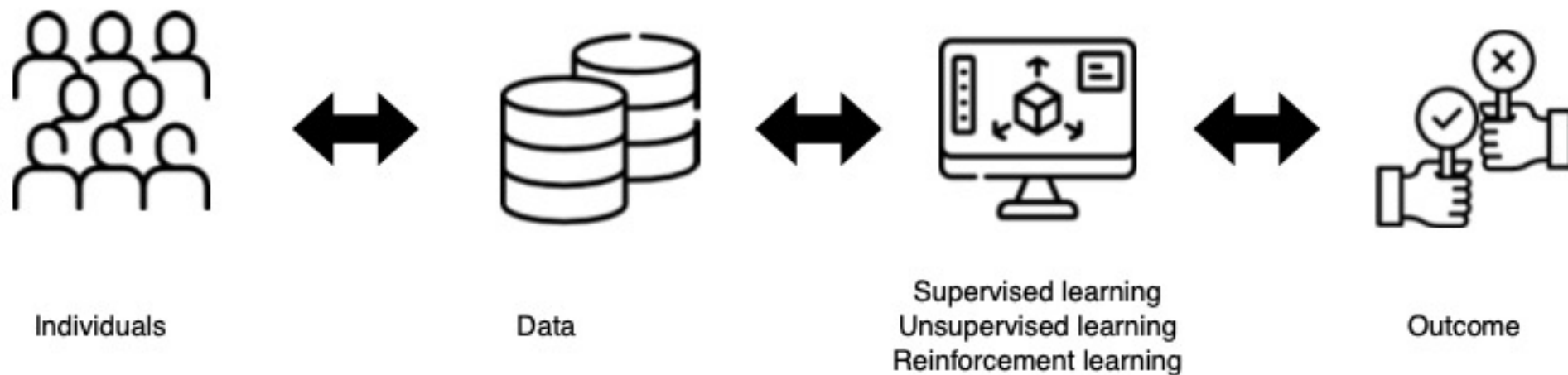


Overview


- Algorithmic Bias/Discrimination
- Categories of Bias
 - Data Bias
 - Interpretation Bias
- Examples of Bias
- Addressing Fairness in ML



Algorithmic discrimination is the result of bias.



bias ▶



Biases introduced during the machine learning process come in different forms.

Data Bias

Algorithms are trained using biased data.

Examples:

- Selection bias
- Sampling bias
- Reporting bias
- Participation bias
- Non-response bias
- Coverage bias

Interpretation Bias

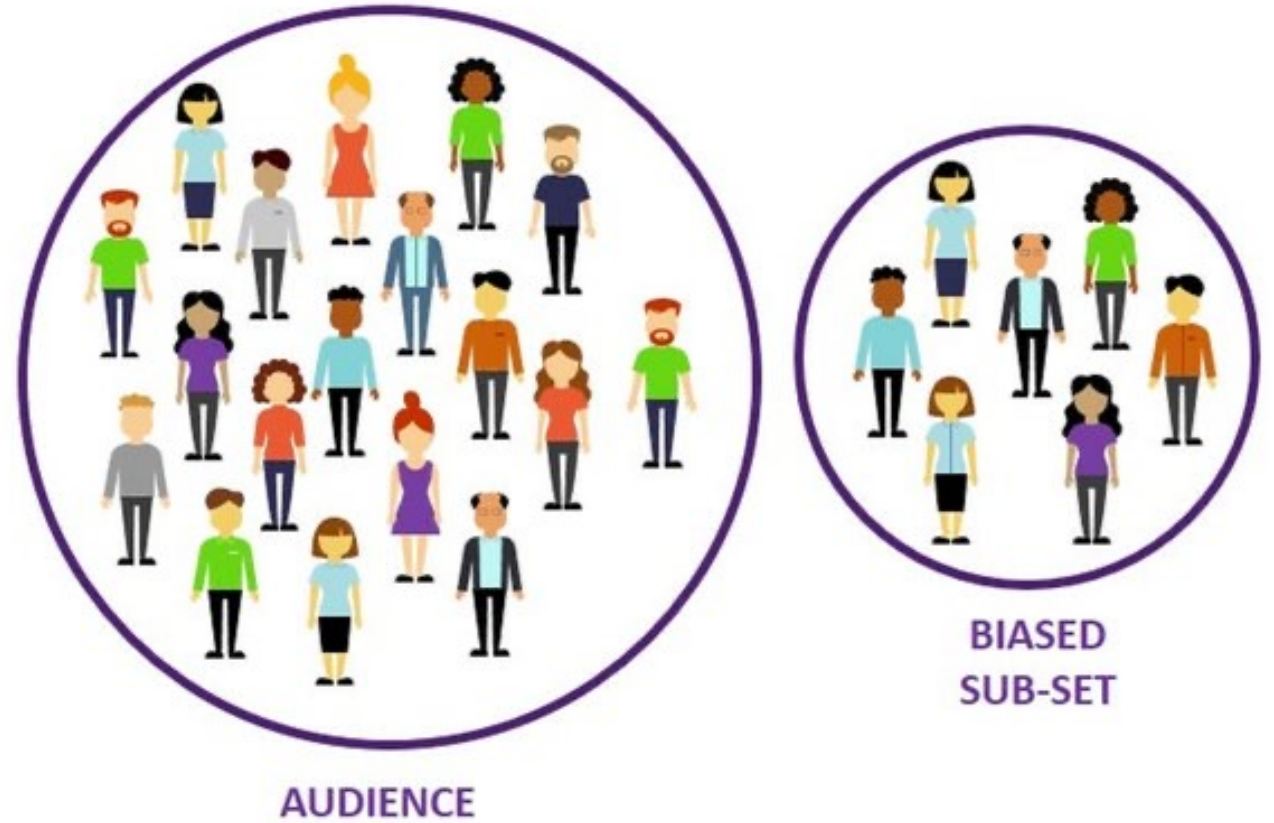
Human assumptions perpetuate a skewed interpretation of results.

Examples:

- Automation bias
- Overgeneralization
- Confirmation bias
- Experimenter's bias
- Group attribution bias
- Implicit bias
- Correlation fallacy

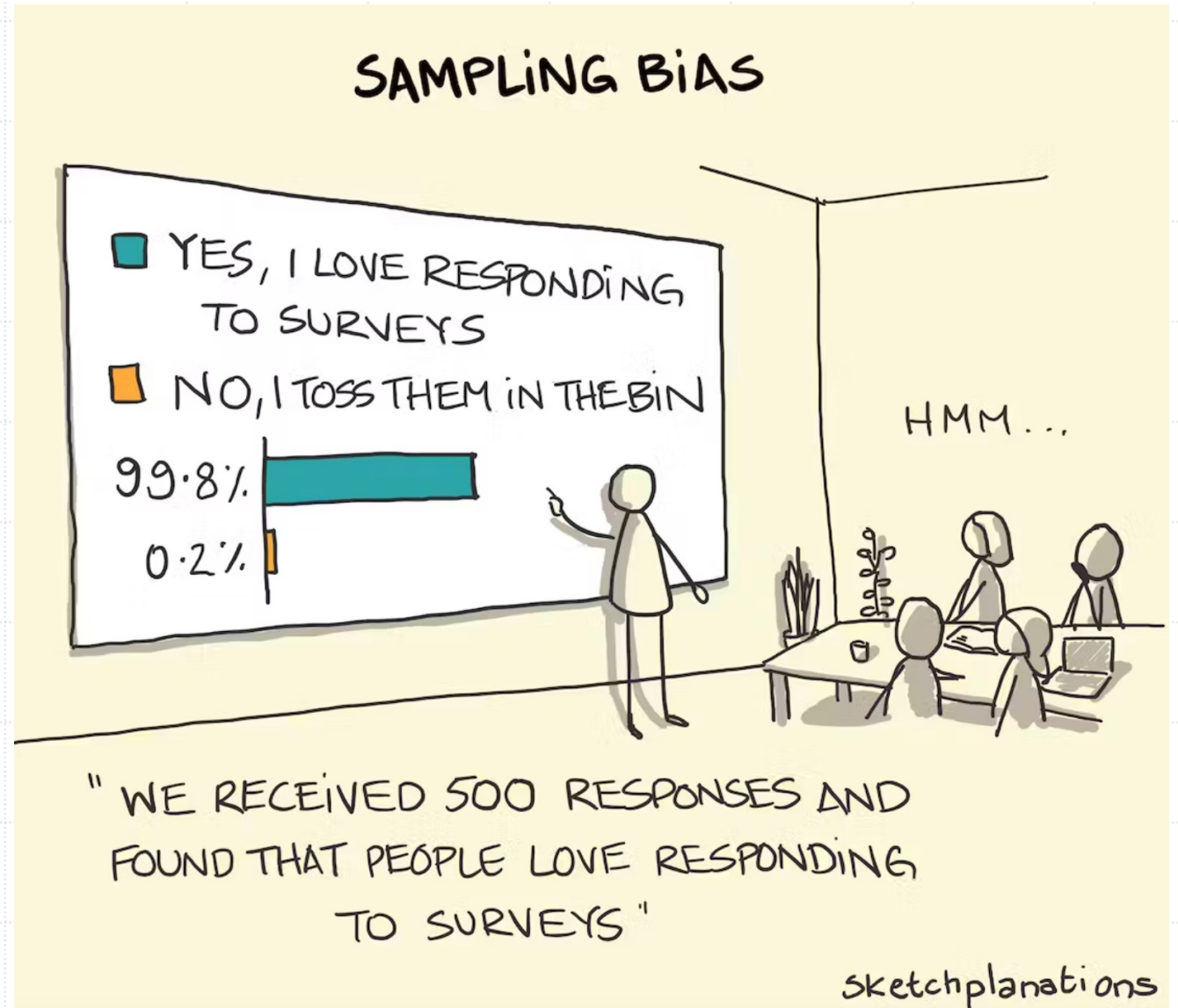
Data Bias

- Selection Bias – Data selections don't reflect randomization



Data Bias

- Sampling bias – data instances are more frequently samples



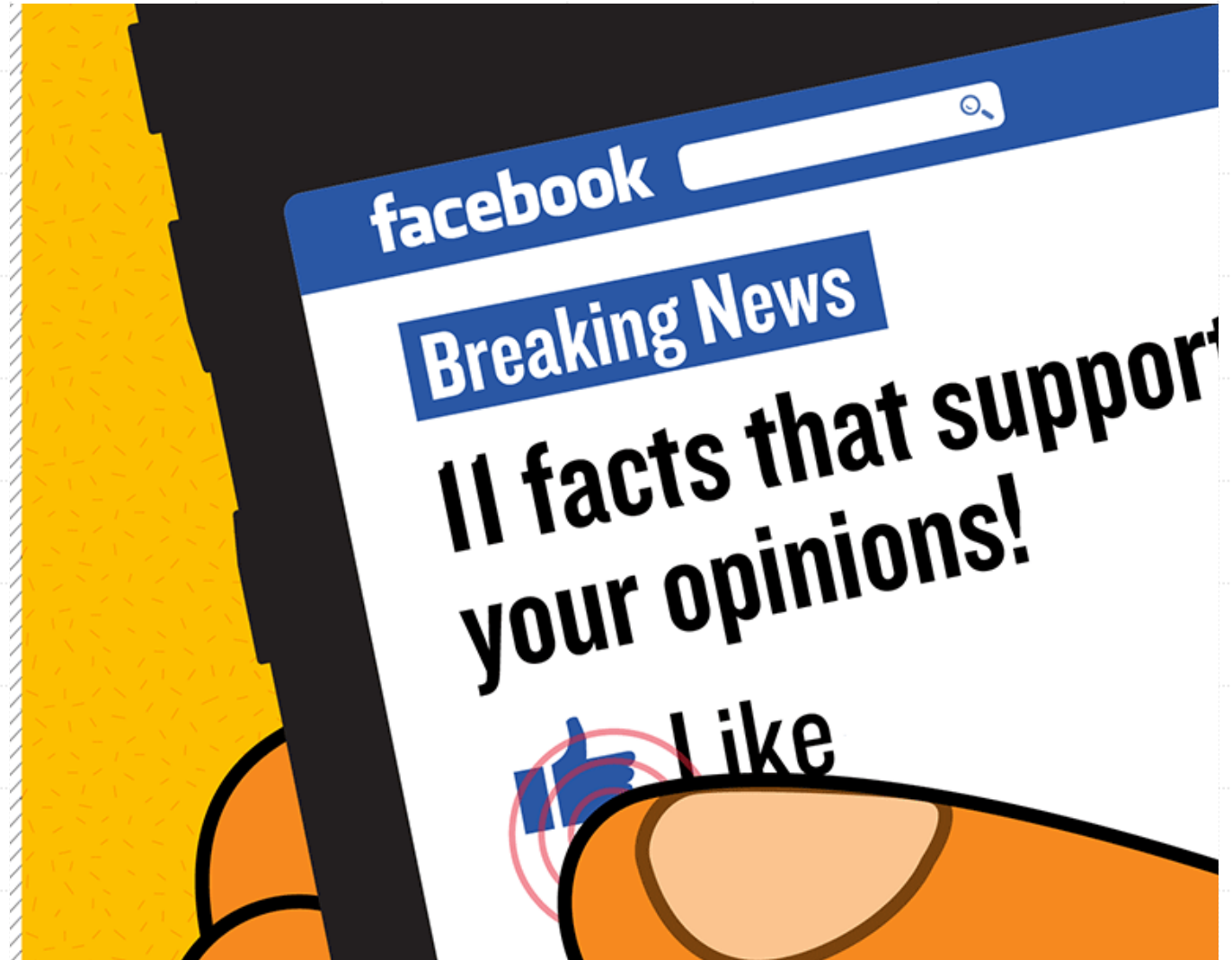
Interpretation Bias

- Overgeneralization – making more general conclusions from limited testing data

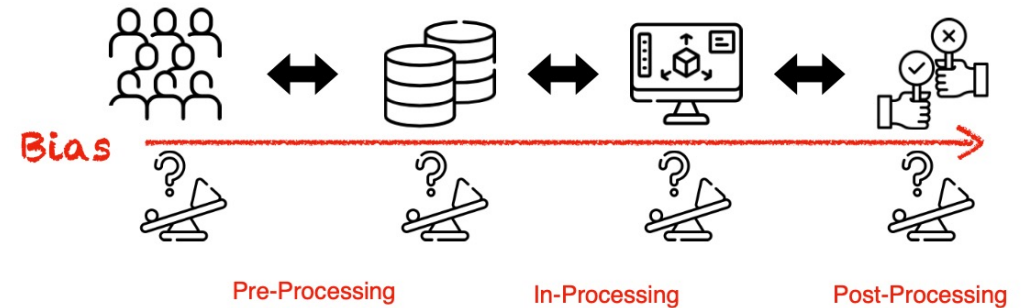
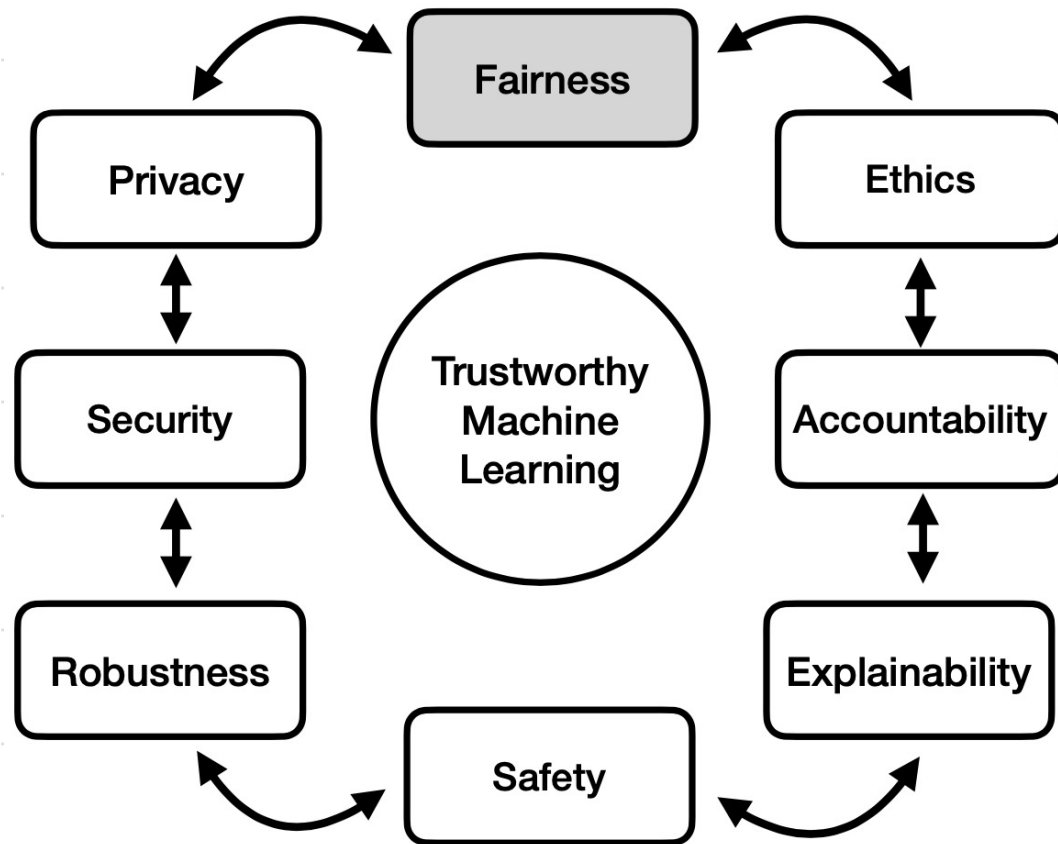


Interpretation bias

- Confirmation bias – the tendency to search for, interpret, favor, recall information in a way that confirms pre-existing beliefs



Fairness



- Fairness in machine learning refers to the various attempts at correcting algorithmic bias in automated decision processes.

Fairness is not a simple concept.

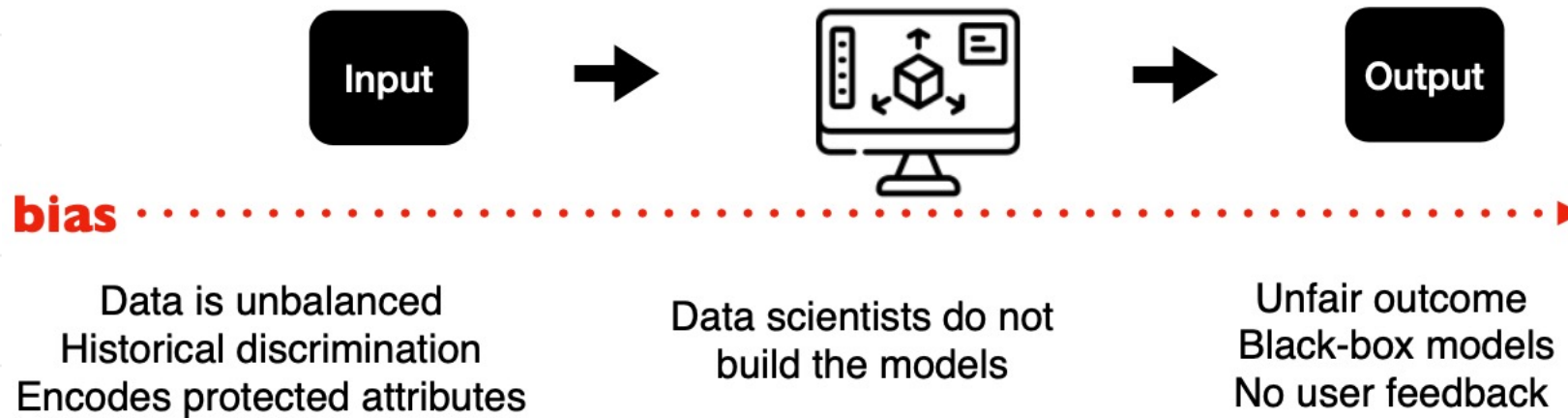
Correcting bias requires:

- Identifying the type of bias present
- Deciding what properties would indicate an unbiased process

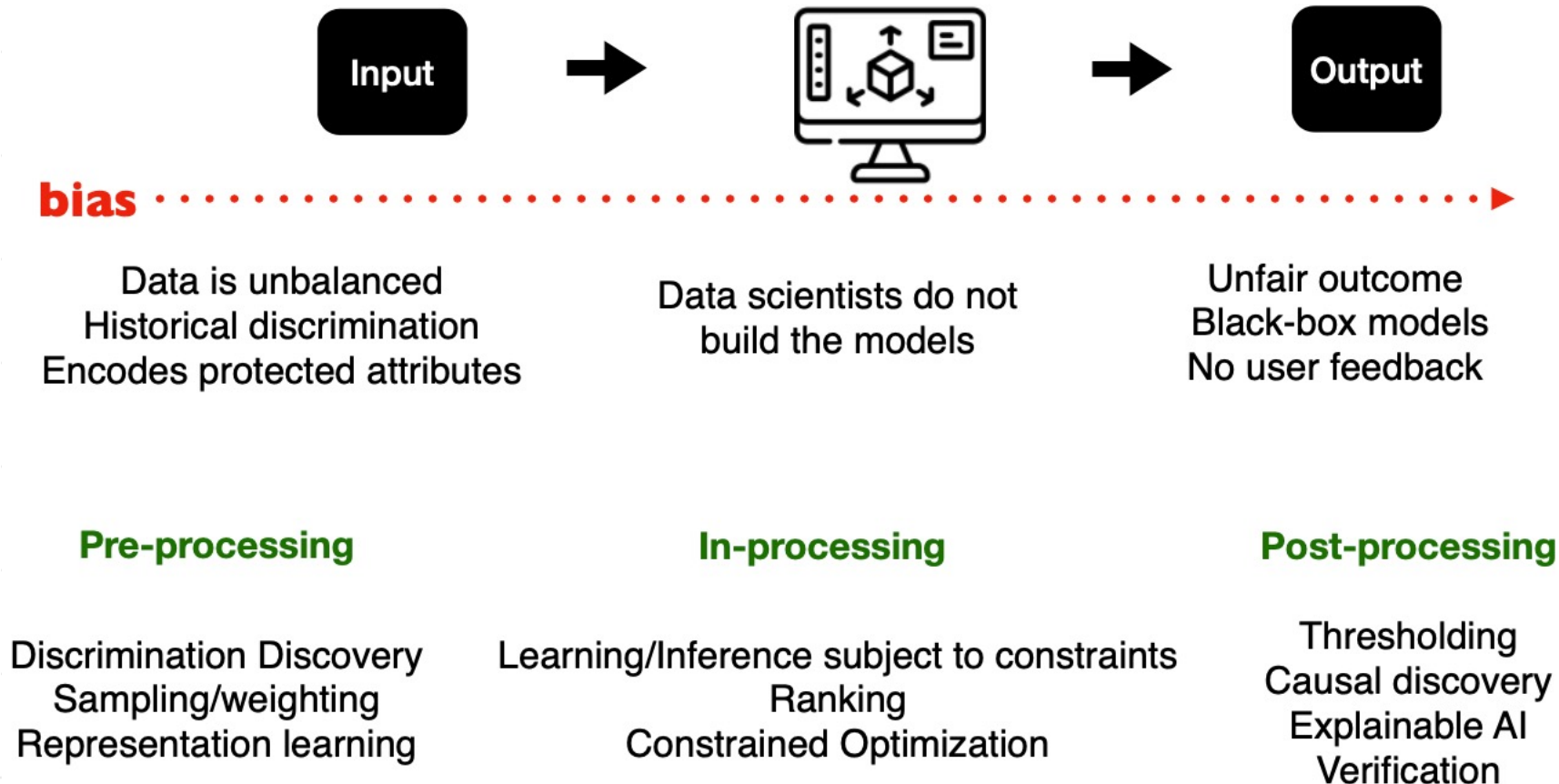
Biases are subjective, therefore, there is no single agreed-upon measure of fairness.



Addressing Fairness in Machine Learning



Addressing Fairness in Machine Learning





Thank You

Please send us your questions at:

vgupta@mmc.edu and

dpounds24@email.mmc.edu