Abolfazl Asudeh
Fall 2020
9/22/2020

InDeX Lab
Innovative
Data
eXploration
Laboratory

CS 594: 4- Bias and other related notions

References:

1. S. Barocas and A. D. Selbst. Big data's disparate impact. Calif. L. Rev., 104:671, 2016
2. Drosou, Marina, et al. "Diversity in big data: A review." Big data 5.2 (2017): 73-84.
3. Jon Kleinberg "Fairness, Rankings, and Behavioral Biases“, keynote talk at FAT* 2019.
4. A. Olteanu, C. Castillo, F. Diaz, and E. Kiciman. Social data: Biases, methodological pitfalls, and ethical boundaries. in Big Data, 2:13, 2019.
5. Salma Ghoneim. 5 Types of bias & how to eliminate them in your machine learning project
6. (Tutorial) A. Asudeh, H. V. Jagadish. Fairly Evaluating and Scoring Items in a Data Set. PVLDB, 13(12): 3445-3448, 2020

Bias in data

Why is data biased?

Pre-existing bias

Historical discrimination. e.g. Redlining

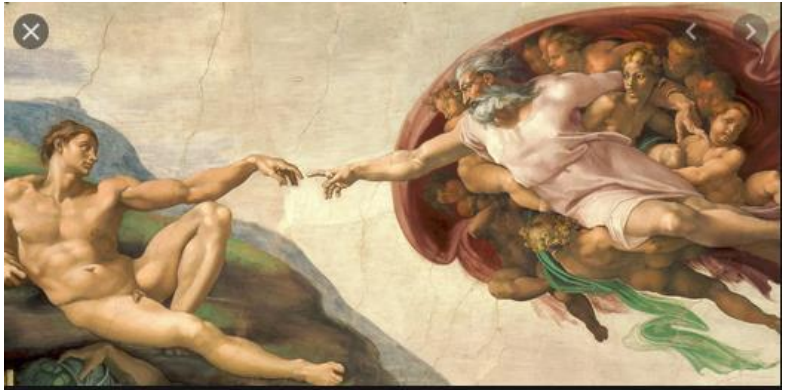
Explicit human bias

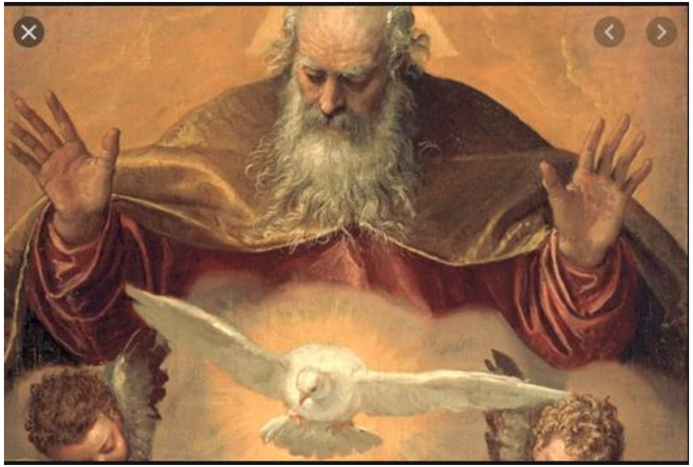
Implicit biases: Most of us, even if we try to be fair, have

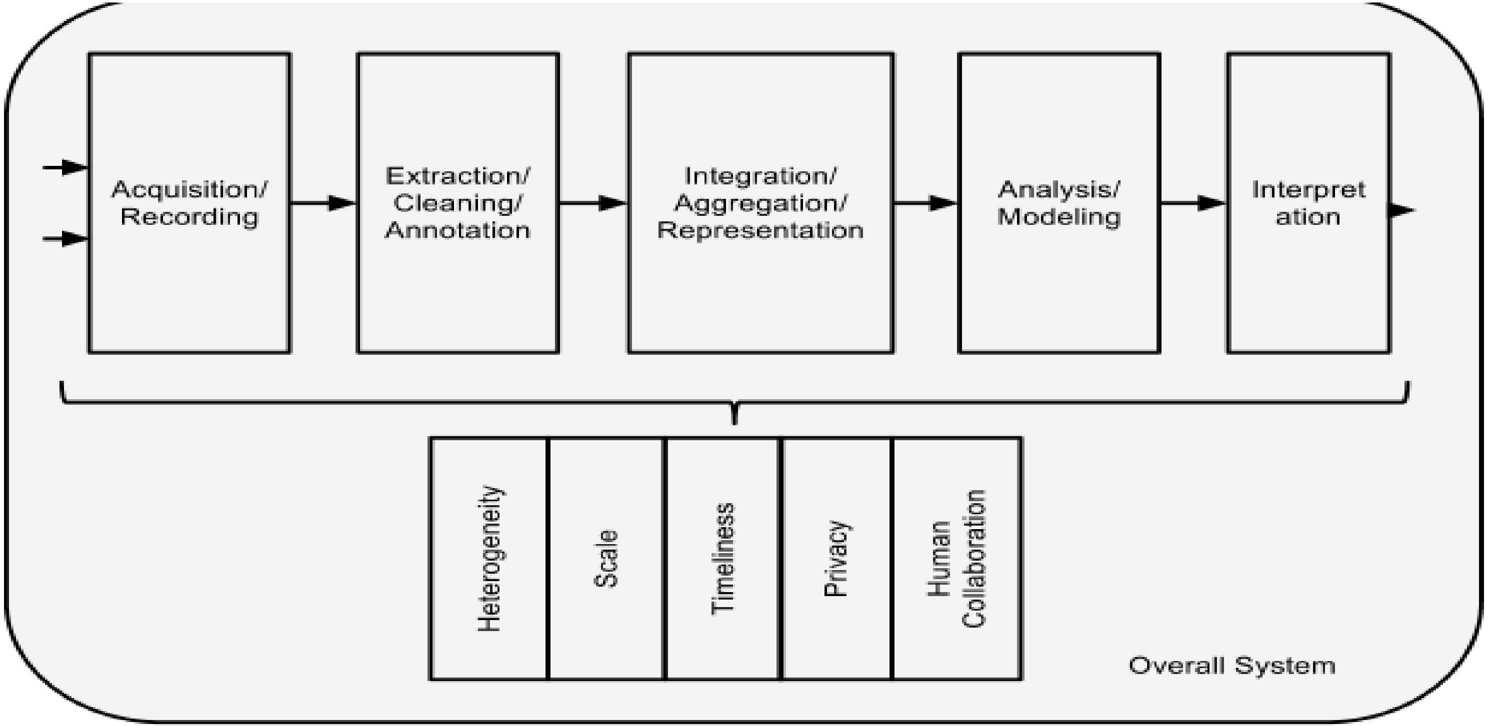
There is a long history of discrimination against people based on sex, religion, race, ethnicity, sexual orientation, and so on. These biases get reflected in the training data we see.

False stereotypes

- When you imagine a “doctor” you probably picture a man, but for a nurse you picture a woman.
- How about god?







The Big Data Pipeline (CACM 2014)

Bias in Data acquisition (selection bias, aka sampling bias)

collected data doesn't accurately represent the environment the program is expected to run into.

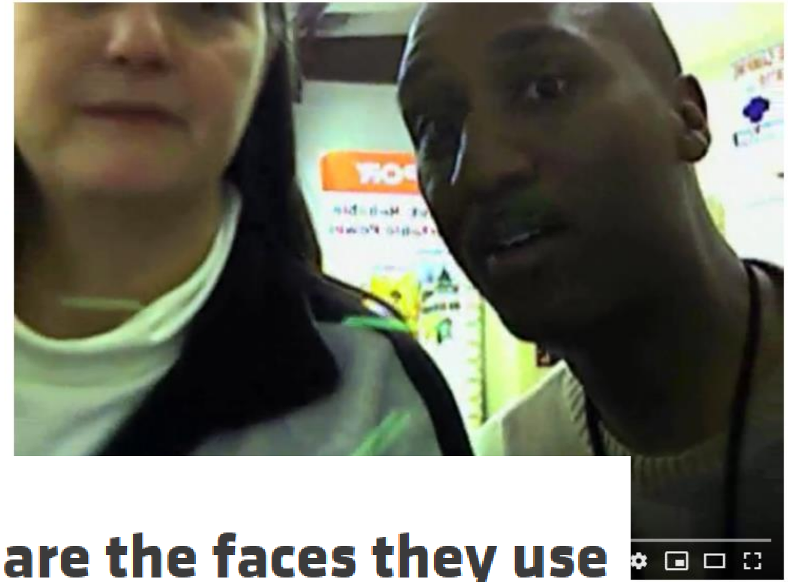
Often we cannot get the data we want, so we use the data we can get.

We need opinions for all citizens, we use Twitter as a proxy, knowing that not everyone tweets, and this is a biased sample, which skews younger, more tech-savvy, better off, and so on.

We want to know the number of crimes committed, which is really unknown. So we use the number of crimes recorded by the police, which is not a random sample of the crimes committed.

Selection bias

- Google Gorilla
- Nikon camera’s open eyes detection
- The face tracking feature of the HP web cams

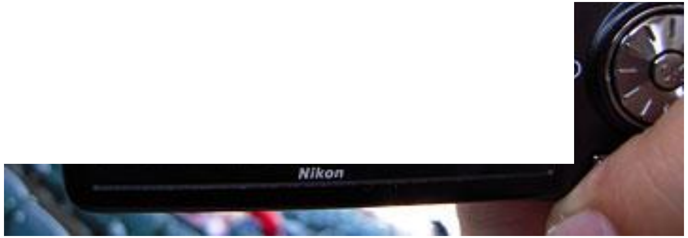


ARTIFICIAL INTELLIGENCE | DIVERSITY

Most engineers are white — and so are the faces they use to train software

A black researcher had to wear a white mask to test her own project.

By Tess Townsend | Jan 18, 2017, 11:45am EST



Selection bias

- **Medical drug trials**

Understanding ANIMAL RESEARCH

Why we need female mice in drug trials

Mia Rozenbaum

Male mice can outnumber female mice five to one

The under representation of female mice in biomedical research is based on the assumption that females are intrinsically more variable than males. This idea is mostly due to their oestrous cycle where hormones vary periodically - every 4 days in mice. Female mice are generally tested at each of the four stages of oestrus to generate reliable data.

eLife

eLife. 2016; 5: e13615.
Published online 2016 Mar 3. doi: 10.7554/eLife.13615
PMCID: PMC4821800
PMID: 26939790

Bias in the reporting of sex and age in biomedical research on mouse models

Oscar Flórez-Vargas,[1,*] Andy Brass,[1,*] George Karystianis,[2] Michael Bramhall,[1] Robert Stevens,[1]
Sheena Cruickshank,[3] and Goran Nenadic[2,4]

Bias in feature extraction

- How to digitize an object can introduce bias
  - Resume
  - GPA
- Feature definition: If exactly two values are allowed for gender, then we so not have the ability to represent other genders.

Bias in data cleaning -- Exclusion bias

- excluding some feature(s) from our dataset usually under the umbrella of cleaning our data

- ***Deleting a record from the majority is not very harmful***
- ***Deleting a record from minorities (under-represented group) can significantly increase bias***

Bias in data annotation – observer bias

- The tendency to see what we expect to see, or what we want to see. (passing human bias into data)
  - It may significantly increase bias, as it is the opinion of (probably) a single individual
- Example: Is Intelligence influenced by status? — The Burt Affair
  - Burt’s approach to intelligence testing allegedly proved that children from the working classes were in general, less intelligent.
  - This led to the creation of a two-tier educational system in England in 1960s

Bias in data aggregation

- Wrong Entity Resolution for underrepresented groups

Bias in representation (embeddings)

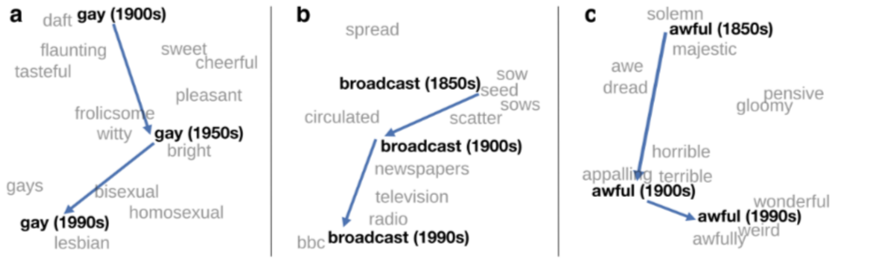
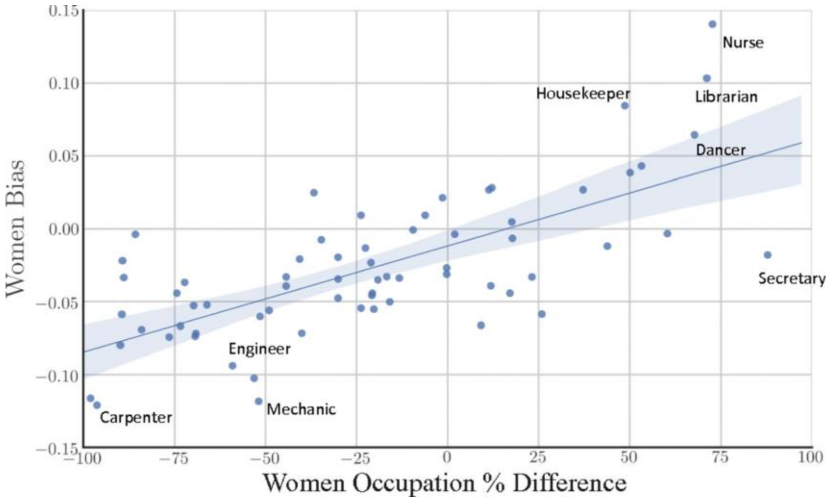


Figure 1: Two-dimensional visualization of semantic change in English using SGNS vectors.[2] **a**, The word *gay* shifted from meaning "cheerful" or "frolicsome" to referring to homosexuality. **b**, In the early 20th century *broadcast* referred to "casting out seeds"; with the rise of television and radio its meaning shifted to "transmitting signals". **c**, *Awful* underwent a process of pejoration, as it shifted from meaning "full of awe" to meaning "terrible or appalling" (Simpson et al., 1989).



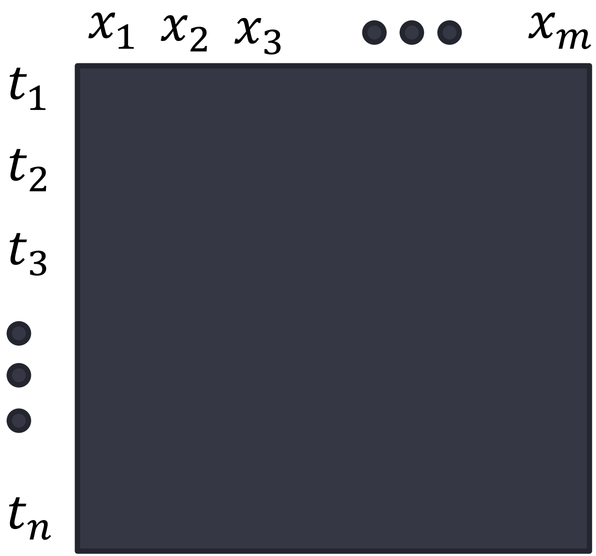
Bias in Modeling

If we use a standardized test as a proxy for intellectual ability, we are selecting particular aspects of intelligence to focus on, and we are ignoring the impact of test-taking skills and test preparation.

The list of modeling choices is very long, with personal biases getting reflected in these choices, often without even the modeler aware of it.

Bias in rows v.s. columns

- Bias in rows: Not enough representative tuples from minority (sub)groups
- Bias in columns: Features are biased (correlated) with sensitive attributes



Other notions

Diversity

Diversity constraints require that there be enough representation of even small protected groups so that they are not ignored.

Bias in rows

Diversity v.s. fairness

Stability and Robustness

- Small changes in model parameters change the output
  - Decisions based on which are questionable (not fair)
- Small changes in input data changes the algorithm output/performance
  - Noisy data (who knows what the real value was?)
  - Distribution drift

Robustness

- Robust is a characteristic describing a model's, test's, or system's ability to perform effectively while its variables or assumptions are altered. A robust concept will operate without failure and produce positive results under a variety of conditions.

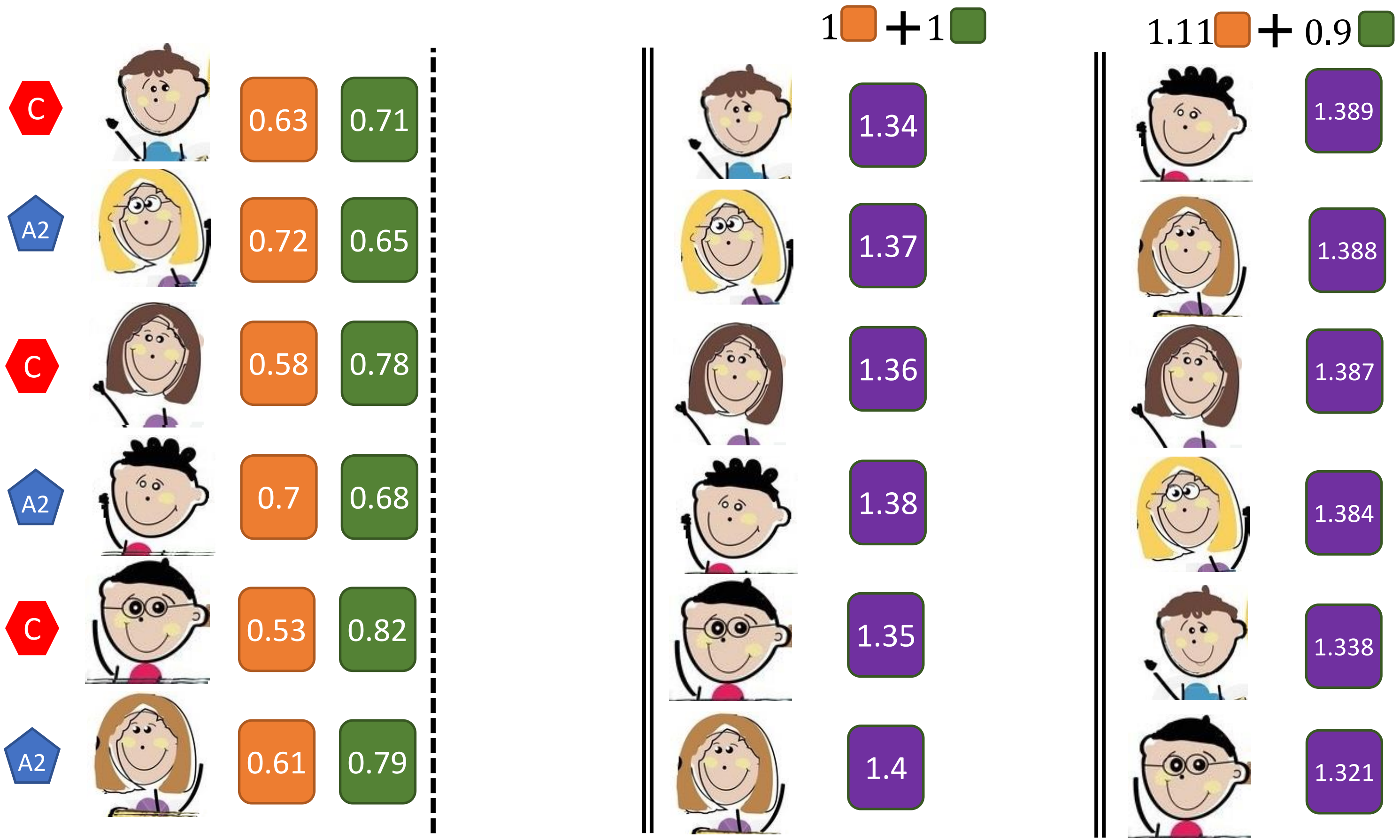
- For statistics, a test is robust if it still provides insight into a problem despite having its assumptions altered or violated.

https://www.investopedia.com/terms/r/robust.asp

Toy Example

Sale -- Normalized

Customer Satisfaction -- Normalized



$$\sum \theta_i x_i$$

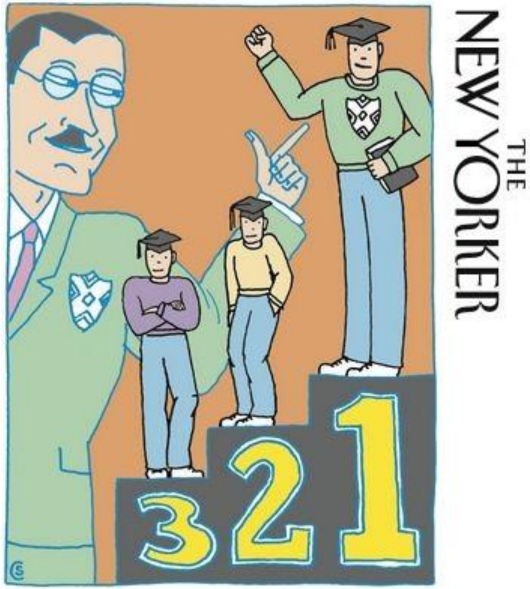
$$\theta_1 \square + \theta_2 \square$$

THE ORDER OF THINGS

What college rankings really tell us.

By Malcolm Gladwell

- "It is easy to see why the U.S. News rankings are so popular. A single score allows us to judge between entities"
- "Rankings depend on what weights we give to what variables"
- "This idea of using the rankings as a benchmark, college presidents setting a goal of 'We're going to rise in the *U.S. News* ranking' …"



Rankings depend on what weight we give to what variables.

Illustration by SEYMOUR CHWAST

Explainability

- Are you able to provide an explanation on why a specific outcome has been generated?
- To which extent the internal mechanics of a system can be explained in human terms.

- Why x is labeled as positive?
  - Lime
  - Decision Trees

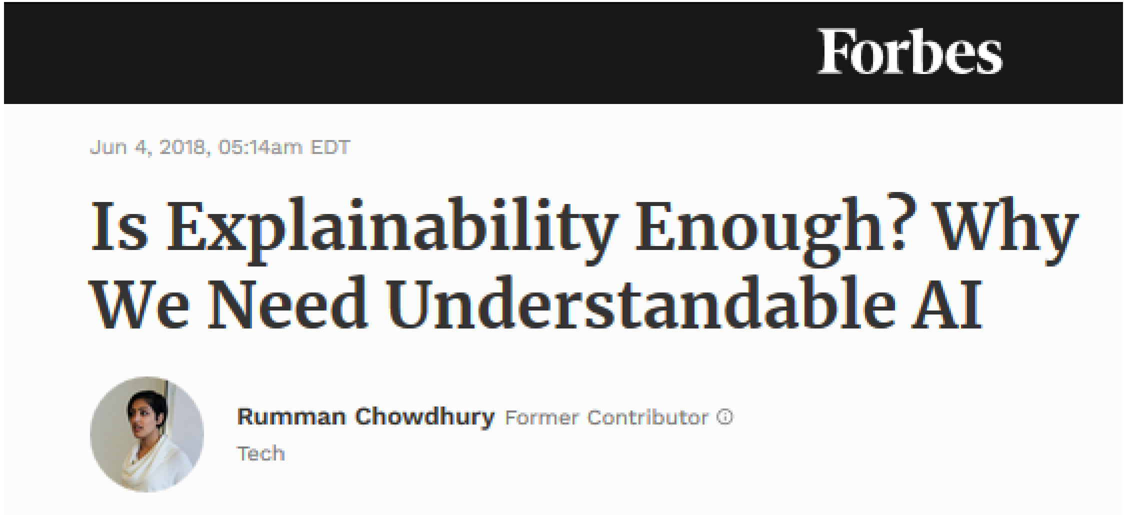
Interpretability

- Explainability and interpretability are often used interchangeably

- To which extent, a cause and effect can be observed within a system.
- To which extent, you are able to *predict* what is going to happen, given a change in input or algorithmic parameters.

Understandability

- To be understood by regular users of the system and those being impacted



Forbes

Jun 4, 2018, 05:14am EDT

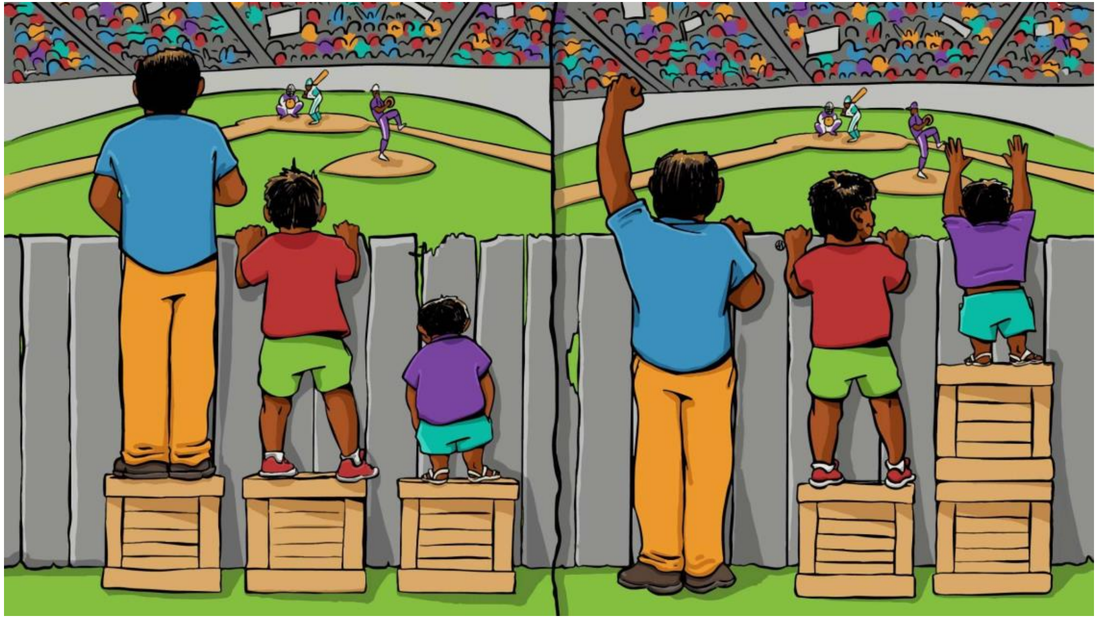
Is Explainability Enough? Why We Need Understandable AI

Rumman Chowdhury Former Contributor

Tech

Understandable AI is the domain of UI/UX designers and product developers in collaboration with AI engineers and data scientists.

Equity v.s. Equality



Trust

‘Ethics Guidelines for Trustworthy AI’ Summarised

Ben Gilburt Follow
Apr 24, 2019 · 5 min read

- To which degree one can trust the outcome of a model?