Accountability in Algorithmic Decision Making

Decision Making Process

- Prioritization
- Classification
- Association
- Filtering

Prioritization

- brings attention to certain things at the expense of others
- Search engine
  - quality of schools and hospitals
  - the riskiness of illegal immigrants on watch list
- criteria used in a ranking, how they are defined and datafied, and their weighting are essential design decisions that deserve careful consideration and scrutiny

Classification

- Classification decisions mark a particular entity as belonging to a given class by considering key characteristics of that entity
- Bias in the training data
- The accuracy of classification algorithms
  - False Negative / False Positive
- The consequences of these errors

Association

- Association decisions revolve around creating relationships between entities
- The consequences of mis-association
- Collaborative filtering is a popular class of algorithm that defines an association neighborhood (a cluster, really) around an entity and uses those close ties to suggest or recommend other items

Filtering

- Filtering decisions involve including or excluding information according to various rules or criteria
- Moderation and filtering are crucial elements when publishing on social media
- Online comments are sometimes filtered algorithmically
- Danger of going into censorship

Filtering

ACM Code of Ethics for software engineering

- act in the public interest:
- to be accountable and responsible for their work
- to moderate private interests with public good
- to ensure safety and privacy, to avoid deception
- to consider the disadvantaged

An Algorithmic Transparency Standard

- Human involvement
  - explaining the goal, purpose, and intent of the algorithm, including editorial goals and the human editorial process or social context crucible from which the algorithm was cast
- Data
  - What features or variables are used in the algorithm?
  - Of-ten those features are weighted: What are those weights?
  - What were the tools used to do the modeling?
- The Model
  - Benchmarking against standard data- sets and with standard measures of accuracy to disclose some key statistics
  - What are the assumptions (statistical or otherwise) behind the model, and where did those assumptions arise?
  - And if some aspect of the model was not ex- posed in the front end, why was that?

An Algorithmic Transparency Standard

- Inferencing
  - What is the margin of error?
  - What is the accuracy rate, and how many false positives versus false negatives are there?
  - What kinds of steps are taken to remediate known errors?
- Algorithmic presence
  - disclose if and when an algorithm is being employed at all