FAIRNESS AND ABSTRACTION IN SOCIOTECHNICAL SYSTEMS

Introduction

- Abstraction: Defined as black boxes precisely by inputs, outputs and relationship between them.
- Fair-ML researchers miss the broader picture by abstracting the social context away.
- Five failure modes(traps) due to this abstraction error.
- Science and Technology Studies (STS) – sociotechnical systems (a combination of technical and social components)
  - Key- Shift from solution-oriented approach to process-oriented.
  - Draws the boundary of abstraction to include social actors, institutions, and interaction.

The Abstraction Trap1- Framing Trap

- Failure to model the entire system over which a social criterion, such as fairness, will be enforced.
- Algorithmic Frame: representations of data & labeling of outcomes
- Data Frame: algorithms as well as its inputs and outputs.
- Sociotechnical Frame: ML model is part of a sociotechnical system, other components of the system needed to be modeled.

An STS Lens on Framing Trap

- Adopt a “heterogeneous engineering” approach
- Example: Cell phones
  - Satellites, wireless protocols, batteries, electrical outlets to companies like Apple, regulatory agencies like the FCC, standards setting organizations like the IEEE
  - Categorical Mistake: conceptually separating ML from the social context = company that designs a cell phone without knowledge of data plans, satellites, regulators and so on

The Abstraction Trap2- Portability Trap

- Failure to understand how repurposing algorithmic solutions designed for one social context may be misleading, inaccurate, or otherwise do harm when applied to a different context.
- Portability is equally important in machine learning.

An STS Lens on Portability Trap

- Contextualizing user "scripts"
  - Example: studies on how light bulbs and generators, developed in France as part of a development project, failed once imported to West Africa.
- Scripts demonstrate, that concepts such as "fairness" are not tied to specific objects but to specific social contexts.
  - Attaching the label "fair" to the code will erroneously encourage to appropriate this code without understanding how the script changes or is disrupted with a shift in social context.

The Abstraction Trap3: Formalism Trap

- Failure to account for the full meaning of social concepts such as fairness, which can be procedural, contextual, and contestable, and cannot be resolved through mathematical formalisms.
- Limiting the question to a mathematical formulation gives rise to two distinct problems in practice.
- First, there is no way to arbitrate between irreconcilably conflicting definitions using purely mathematical means
- Second, no definition might be a valid way of describing fairness.
    - Procedurality
    - Contextuality
    - Contestability

An STS Lens on Formalism Trap

- Identifying “interpretive flexibility”, “relevant social groups”, and “closure”
- Social Construction of Technology program (SCOT)
- Social groups have the power to shape technological development
- Rhetorical Closure

The Abstraction Trap4: Ripple Effect Trap

- Failure to understand how the insertion of technology into an existing social system changes the behaviors and embedded values of the pre-existing system
- unintended consequences are the ways in which people and organizations in the system will respond to the intervention.
- Technologies can also alter the underlying social values and incentives embedded in the social system

An STS Lens on Ripple Effect

- Avoiding "reinforcement politics" and "reactivity"
- Awareness of several common changes avoids common pitfalls that may negatively affect the fairness of their proposed systems
- Reinforcement politics
- Reactivity behaviors
- Heterogenous Engineers:

Abstraction Trap5: Solutionism Trap

- Failure to recognize the possibility that the best solution to a problem may not involve technology
- By starting from the technology and working outwards, there is never an opportunity to evaluate whether the technology should be built in the first place
- Fairness definitions can be politically contested or shifting, a model may not be able to capture how it moves
- The modeling required could be so complex as to be computationally intractable

An STS Lens on Solutionism Trap

- Considering when to design
- Careful consideration of the complex sociotechnical system at play
- Cooperation between fair-ML researchers and domain experts
- Not all problems can or should be solved with technology

Conclusion

- When considering designing a new fair-ML solution, this would mean determining if a technical solution:
  - requires a nuanced understanding of the relevant social context and its politics (Solutionism).
  - remains unchanged after the introduction of the technology (Ripple Effect).
  - can handle robust understandings of social requirements (Formalism).
  - has appropriately modeled the social and technical requirements of the actual context in which it will be deployed (Portability).
  - is heterogeneously framed to include the data and social actors relevant to the localized question of fairness (Framing).

Thank You