JOHNS HOPKINS
WHITING SCHOOL
of ENGINEERING

Origins and Data

9/7/23

Logistics

- Change response deadline to 9am the day of class?
- Finalize discussion leaders for the next few weeks

Course Project

1. Literature Review **[DUE THURSDAY SEPTEMBER 28]**
2. Proposal
   [Short in-class proposal presentations]
3. Final paper
   [In-class presentations]

Discussion Time

- Should crowdsourcing require IRB approval? (Xiaowen, Muhammad, Ammar, Kristen, Owen, Abe)
- What are the potential harms and benefits of seeking lower-cost labor?
  - At what point does it constitute "ethics dumping" or exploitation? (Krithika, Ammar)
  - Does it also harm data quality? (Bolun, Muhammad, Haonan, Pulkit)
  - Thoughts on Sama's marketing as an ethical AI company and differences between reality and what's publicized? (Kristen, Taryn)
  - What is "fair" pay? (Nikhil, Krithika, Abe, Kevin, Elisée)
- What kind of protections could be in place to reduce power imbalances between companies and crowd-workers / protect rights of workers? (Yaohan, Bolun, Jiahui, Chi)
  - What challenges and opportunities do crowdworkers have for collective action? (Zhiqing)
- What are potential harms of crowd-sourcing that neither article fully addresses?
- What is the role of investigative journalism regarding accountability?
- How should we prevent models from outputting toxic content? Is there a better solution than crowd-sourcing labels? Is the issue in how crowd-workers are treated?

Next Topic: Fairness, Bias, and Stereotypes

- Tuesday September 12: Fairness metrics
- Thursday September 14: Classification/Prediction
- Tuesday September 19: Generation

A little background: ProPublica COMPAS Report (2016)

- "The formula was particularly likely to falsely flag black defendants as future criminals, wrongly labeling them this way at almost twice the rate as white defendants."

- "White defendants were mislabeled as low risk more often than black defendants."

https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing

A little background: risk assessment

- U.S. courtrooms have employed various forms of **risk assessment**
- Given someone that has been arrested for a crime, models are typically trained to predict if they will be arrested again in the future
  - "Recidivism" if a convicted person will "reoffend"
- Used at all stages of criminal justice system:
  - When determining if someone can be released while they are awaiting trial
  - When determining programs while they are incarcerated
  - Level of supervision (home confinement, electronic monitoring) when they are released

https://bja.ojp.gov/program/psrac/basics/what-is-risk-assessment#:~:text=As%20a%20cornerstone%20of%20this,and%20identify%20areas%20for%20intervention.

A little background: risk assessment

- Consistency
  - "can be viewed as more defensible and credible than more subjective and less transparent decision-making processes"
- Efficiency
  - "help practitioners make more efficient use of limited justice resources"
- Effectiveness
  - "help practitioners more effectively improve criminal justice outcomes (e.g., reduce reoffending, improve compliance)"

https://bja.ojp.gov/program/psrac/basics/what-is-risk-assessment#:~:text=As%20a%20cornerstone%20of%20this,and%20identify%20areas%20for%20intervention.

Fairness metrics

Group Fairness

- Equalized odds:
  - Protected and unprotected groups should have equal rates for true positives and false positives
  - Example: COMPAS
- Demographic (statistical parity)
  - Likelihood of a positive outcome should be the same regardless of whether or not the person is in the protected group
  - Example: men and women people should be equally able to get loans
- [etc]

Individual Fairness

Similar individuals should be treated similarly

Tuesday's Readings

1. Dwork, Cynthia, et al. "Fairness through awareness." Proceedings of the 3rd innovations in theoretical computer science conference. 2012.
2. Chouldechova, Alexandra. "Fair prediction with disparate impact: A study of bias in recidivism prediction instruments", Big Data, Special issue on Social and Technical Trade-Offs. 2017.
3. (optional) Corbe-Davies, Sam et al. "Algorithmic Decision Making and the Cost of Fairness", KDD. 2017.
4. (optional) Mehrabi, Ninareh et al. "A Survey on Bias and Fairness in Machine Learning", ACM Computing Surveys. 2021.