Fairness via AI:

Bias Reduction in Medical Information

Shiri Dori-Hacohen (PI); Sherry Pagoto, Scott A. Hale (Co-PIs)

Joint work with Roberto E. Montenegro, Fabricio Murai, Adimika Arthur, Charmain Jackman, Keen Sung, Michela Blain, Jennifer Edwards-Johnson

An aspirational gap

Most AI Fairness work is inherently limited due to the "Bias In, Bias Out" principle (Mayson, 2018)

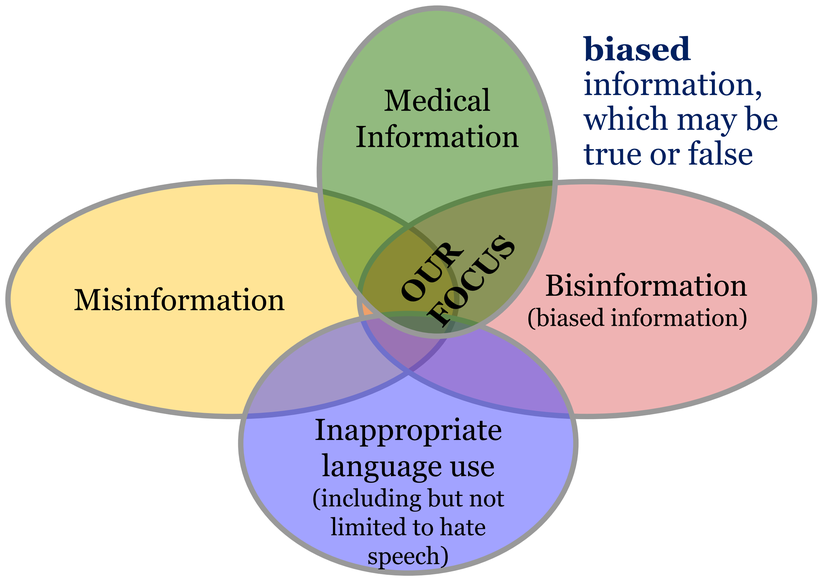
Upper bound is true "fairness", i.e., inequities, in real world

Fairness **via** AI

- Rooted in insights from medical education, sociology, and antiracism
- A broader lens on fairness: greater aspiration
  - Rather than Fairness **in** or **of** AI,
  - Using AI to study & remedy inherently unequal situations in society
  - Accelerate progress
  - Debiasing the world, rather than just debiasing AI

Bisinformation

(\ˌbis-ˌin-fər-ˈmā-shən\)



Venn diagram illustrating the partial overlaps between these issues along with BRIMI's focus

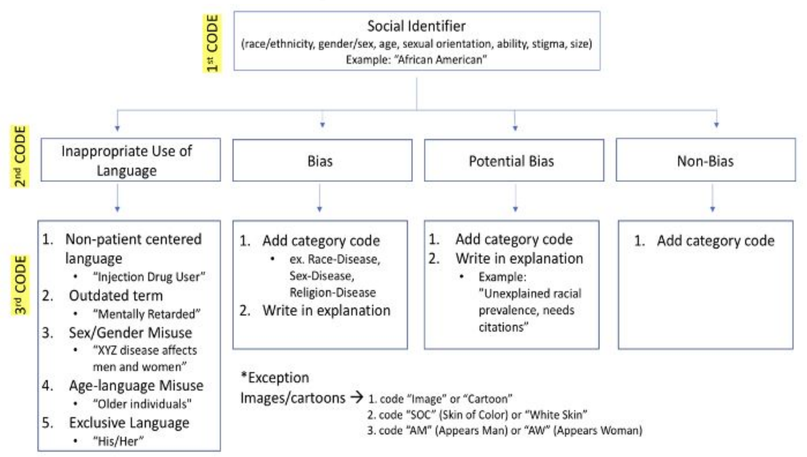
Case Study: Bias in Medical Information

- Bias and language misuse can have detrimental impact on patient outcomes
- Inappropriate usage of social identifiers in a medical context can be harmful, even if strictly true
  - e.g. prevalence of illness in racial category
  - w/o accounting for Social or Structural Determinants of Health
- Similarly for inappropriate language use, e.g.:
  - conflating gender with sex
  - using racial identity as a stand-in for phenotype
- Even more harmful when the information is actually false
  - Example: Long-debunked Salt Gene Hypothesis (still taught)

Reducing bisinformation in medical information

- Bias reduction is critical for accurate evidence-based medicine
- Reducing bisinformation in medical content is a manual and time consuming effort
  - Yet, immense downstream impact
- SOTA AI and NLP can scale up these significantly –
  - But must be undertaken with great care

Medical Experts' Annotation Guidelines



Data collection: complex domain knowledge and extensive training

Prior Results and Work Underway

- Montenegro et al. developed and piloted bias reduction process in med schools
  - Yale, Seattle Children's Hospital, Michigan State University
- Dataset with 10,594 pages of instructional material, with 3,500 annotated examples of bias and related constructs, which will serve as seed data for BRIMI (more constantly added)
- Concurrent grant from NBME's Stemmler Fund with a focus on medical curricula and education - underway
  - Machine learning models to detect bias - under construction by AuCoDe
- NSF FAI grant focuses on online & social media information
  - Data collection - underway

Deliverables

- Traditional research outputs (publications, datasets), plus…:
- Triage guidelines for bis- and misinformation disseminated to public health officials and journalists
- Bias reduction tools integrated in medical schools
- Twitter bot that flags biased or false medical claims

Outreach and Broader Impacts

- BRIMI will embed our research into practice through:
  - outreach to patients
  - integration into existing systems
  - disseminating reports and tools to practitioners
- BRIMI offers outsized promise for improved equity in health for minoritized communities and patients
  - Better patient outcomes
  - Also improves DEI in medical schools!
- Research and Education are integrated
  - by leveraging advances in reducing bias in medical education and applying them to the online information ecosystem.