HEALTH IMPACTS OF IMPLICIT BIAS IN MACHINE LEARNING

SEPTEMBER 2021

Rebekah S. Rasooly, PhD (NINR)

Rebekah Feng, Ph.D. (NINR/NIDA)

Rebecca Goodwin, J.D. (NINR/NLM)

National Institute of Nursing Research

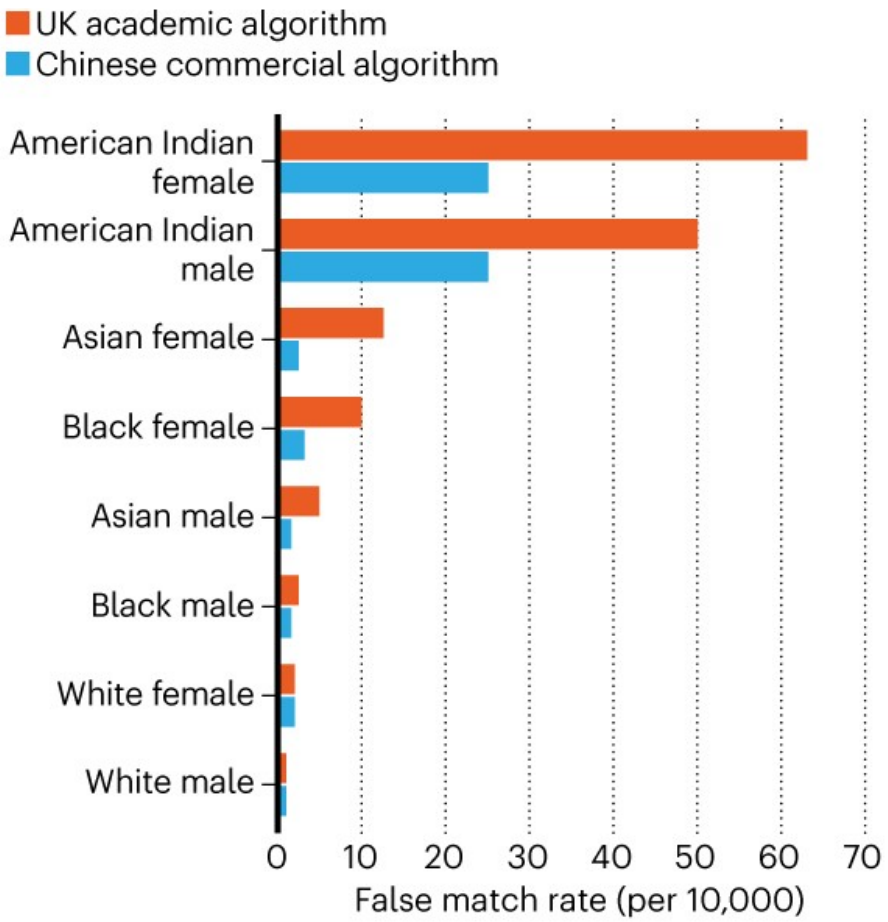
Algorithmic bias

Definition: systematic and repeatable errors in a computer system that create unfair outcomes

- For example:

  Example 1: facial recognition algorithm trained using white male data leads to higher error rates with other groups.

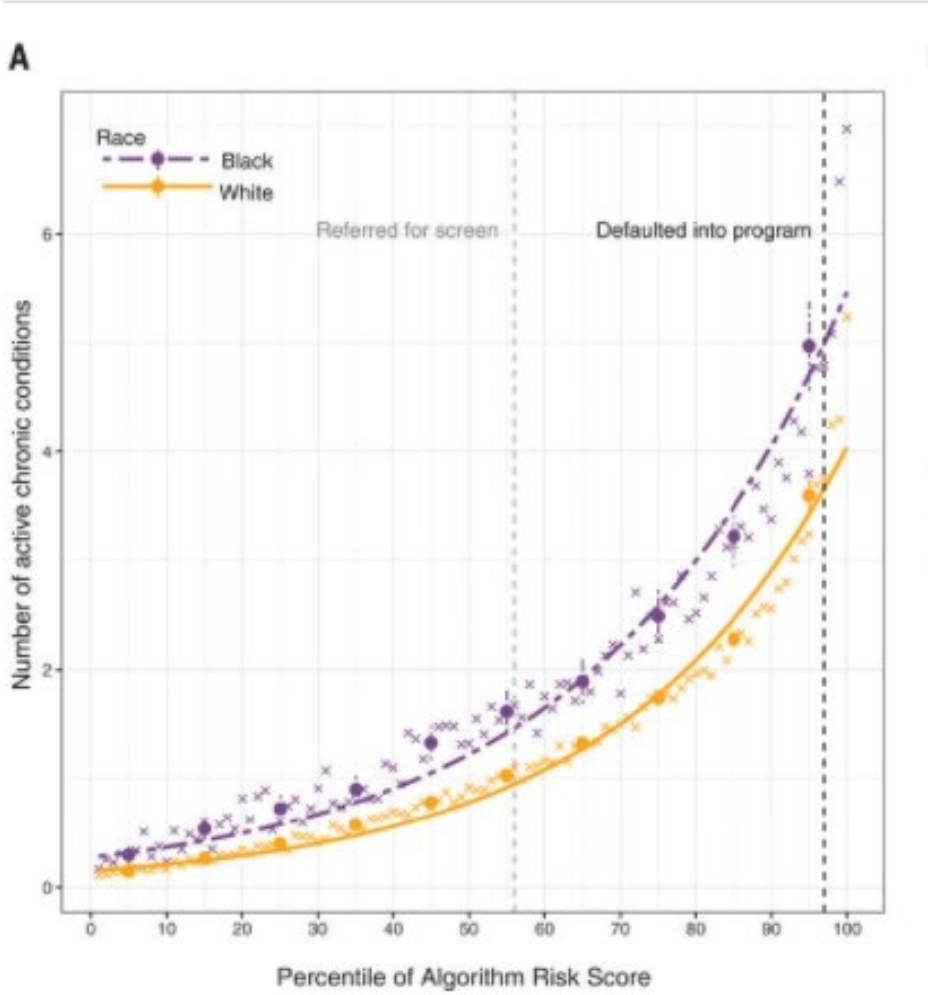
  Example 2: pulse oximeter devices that measure light absorption through the skin to measure blood [O2] perform poorly in Black patients



NISTIR 8280 Grother et al., 2019 summarized in Nature **587**, 347-349 (2020)

Algorithmic bias

- Measurement bias/label choice bias in "high-risk care management" programs:
  - Predict 'future clinical risk' based on "past healthcare costs" to identify patients for referral to specialized services.
  - "Past healthcare costs" reflects a history of more money, on average, spent on white patients than on black patients
  - But black patients actually had more chronic conditions (were 'less healthy') at a given score



Obermeyer, et al., 2019, Science 366: 447-453

Algorithmic bias - A problem that is being widely recognized:

Algorithmic bias – How do we find solutions?

"Rather, we must change the data we feed the algorithm—specifically, the labels we give it. Producing new labels requires deep understanding of the domain, the ability to identify and extract relevant data elements, and the capacity to iterate and experiment." Obermeyer, et al., 2019, Science 366: 447-453

"It's critical that we draw on nursing's holistic perspective to improve public health and advance health equity by identifying nursing practice and policy solutions across clinical and community settings that are responsive to the realities of people's lives." **Shannon Zenk**

PURPOSE and BRIEF DESCRIPTION OF WORKSHOP

- Purpose
  - Develop a workshop to bring together panelists to discuss unintended or undetected bias that results from machine learning algorithms. The goal of the workshop is to discuss ways nursing science can help to detect and prevent unintended or undetected bias that result from machine learning (ML) algorithms used in biomedical research.

- Key scientific objectives
  - To identify gaps in research and funding aimed at understanding and mitigating bias in machine learning that may be relevant to nursing research.

Workshop Outline

- Keynote speeches:
  - Representing the three perspectives: ML/AI, bias/health disparities, healthcare
- Research talks to facilitate discussion:
  - Thematic working groups to discuss specific topics
- Possible poster session/lightning talks by early-stage investigators/trainees

IMPLEMENTATION AND COLLABORATIONS – Possible Partners

- Proposed collaboration
  - Other NIH ICOs that have funded ML/AI health disparities and bias grants (e.g., NLM, NCI, NIA)
  - Other Federal agencies
    - NIST (identifying AI architecture and bias)
    - AHRQ (collecting data on clinical algorithms and bias)
  - Professional societies and other nongovernmental organizations (NGOs)
    - American Medical Informatics Association (AMIA)
    - Mobilizing Computable Biomedical Knowledge (MCBK)