Lifting as we all rise: Addressing challenges to AI bias in healthcare

— Human–AI Collaboration in Healthcare workshop at CSCW 2019 —
Austin, Texas, USA

Identifying Challenges and Opportunities in Human–AI Collaboration in Healthcare

CSCW 2019 Workshop (Saturday, Nov. 9th, Austin, TX)

~ Charles R. Senteio, Kaitlin L. Costello, Vivek Singh ~

Lifting as we all rise: Addressing challenges to AI bias in healthcare

Charles R. Senteio
Rutgers University
charles.senteio@rutgers.edu

Kaitlin L. Costello
Rutgers University
k.costello@rutgers.edu

Vivek Singh
Rutgers University
v.singh@rutgers.edu

Abstract
Artificial Intelligence (AI) is influencing healthcare in multiple ways. However, known human and algorithmic bias in both diagnosis and treatment present persistent challenges to human-AI collaboration in healthcare. We discuss two key scenarios (prostate cancer and mental health assessment) where equity of diagnosis and treatment is critically important, but the requirements for AI are quite different due to the different stages of technological development. While addressing human bias in the clinical encounter is a critical initial step for prostate cancer, reducing bias in currently adopted automated algorithms by refining the operationalization of constructs is a critical next step in mental health assessment. We assert the imperative to utilize current understanding of human and algorithmic bias to inform development of human-AI collaboration which identifies and abates inequities in healthcare, rather than magnifying them.

Author Keywords
AI bias; healthcare; human bias; algorithmic bias

CSS Concepts
• Applied computing ~ Health care information systems

Introduction
Artificial intelligence (AI)–the capability of a machine to perform cognitive tasks to complete a specific goal using provided data–continues to revolutionize healthcare. Computational power, pattern recognition algorithms, and accessibility to 'big data' enables machines to recognize patterns within unstructured bits of information with increasing precision [1]. For example, systems using machine learning show promise to transform healthcare delivery by enhancing complex cognitive tasks which are currently performed by human specialists, such as assessment, diagnosis, and treatment selection.

Alongside these advancements in developing algorithmic solutions to these vexing and enduringly imprecise cognitive tasks, we must also attend to the persistent challenge of recognizing and addressing the negative effects of inequity and bias. In order to develop trustworthy AI, we must attend to equity: AI applications in other domains are susceptible to bias and reproduce and reinforce existing social inequalities. For example, face detection algorithms do not work well for women and non-binary people, and automated parole decisions are systematically biased against people of color [2]. In light of these persistent challenges concerning AI bias, we discuss two healthcare application domains, or key scenarios, which

CSCW 2019

RUTGERS
School of Communication and Information

Alongside great promise is known human and algorithmic bias in Healthcare applications

- AI supporting **complex** cognitive tasks currently performed by medical specialists
  - o Assessment, diagnosis and treatment

- AI **bias in other domains** are well described, which reinforce existing social inequity
  - o Facial recognition for non-binary individuals
  - o Automated parole decisions for people of color (Buolamwini & Gebru, 2018)

- Selected **two areas** of healthcare applications – illustrative examples how bias can manifest in AI
  1. Treatment – prostate cancer
  2. Diagnosis – mental health

To develop *trustworthy* AI we must attend to Equity

Prostate cancer treatment selection is ambiguous and stressful … and wrought with bias

- Persistent **racial disparities** in prevalence and mortality – African American Men (AAM) show 3.0-fold higher mortality risk (Di Pietro, Chornokur, Kumar, Davis, & Park, 2016)

- Multi-faceted **drivers** of disparities, differences in treatment (Division of Cancer Control and Population Sciences (DCCPS), 2019)(Hoffman et al., 2001)

  o AAM receive less decisive treatment (e.g., radiation therapy, radical prostatectomy, or combination therapy) than White men (52.0% vs. 64.8%, P<0.001) (Mahal et al., 2014)(Moses, Orom, Brasel, Gaddy, & Underwood, 2016)

- Treatment selection **strongly associated** with physician recommendations (Aning, Wassersug, & Goldenberg, 2012)(Ramsay et al., 2011)(Saigal, Lambrechts, Seenu Srinivasan, & Dahan, 2017)(van Tol-Geerdink et al., 2013)

- Pertinent bias in **clinical encounters**

  o Physicians perceive that African American patients are less effective communicators than White patients (Street, Gordon, & Haidet, 2007)

  o Physicians are more contentions and more verbally dominant with African American patients (Johnson, Saha, Arbelaez, Beach, & Cooper, 2004)

Underrepresentation in research presents a persistent risk of bias

- AAM **underrepresented** in cancer research, from the bench, to the bedside, to communities (Ahaghotu, Tyler, & Sartor, 2016)(Byrne, Tannenbaum, Glück, Hurley, & Antoni, 2013)

- Lack of representation is enduring public health **concern** – presents potential barriers to racially equitable development of strategies for prevention and treatment (Reifenstein & Asare, 2018)

  o Machine learning being used to identify gene expression to predict prostate cancer outcomes. Bolsters promise of personalized medicine by determining "optimal" treatment (Goldenberg, Nir, & Salcudean, 2019)

- Socio-technical barriers (e.g., trust) present **barriers** to research and use of IT-enabled tools

To develop *trustworthy* AI we must attend to Equity

Mental health assessment and access must attend to bias, especially pertaining to gender

- Gender **inequities** in mental health assessment, which include conceiving gender as binary (Spiel, Keyes, & Barlas, 2019)

- "Disorder" is **socially constructed** and mental illness highly **stigmatized**

- Patient **consent** required for storing and diagnostic profiling

- Practitioners should understand **risk** of bias and use of AI as **adjunctive** method to inform diagnosis, and treatment selection

Must identify and target appropriate diagnostics & treatment selection

- AI can **address** inequity in outcomes, but can also **exacerbate** them

- Current understanding of bias can **inform taxonomy** of illness types which address social determinants of health, inequities to access, and acuteness and chronicity of various health conditions

- Taxonomy can aid in **focus** of future research

  - <u>Prostate cancer</u> – **toolkits** and **workflows** which help prompt physicians to understand source of bias in treatment selection and gaps in current understanding of prognosis considering patient preferences and values

  - <u>Mental health</u> – careful **operationalizing** of gender, mental health assessment, and patient consent

Fulfilling AI potential must require identifying bias & mitigating its deleterious effects – doing so will translate to equitable gains