

# Addressing the Need to integrate **Health Equity** and **Racial Justice** into the Artificial Intelligence Development Lifecycle

Brett R. South, MS, PhD

Irene Dankwa-Mullen, MD, MPH

IBM Watson Health

November 17, 2021

DAISY 2021 Workshop, University of Florida



# COI Statement

IBM's statements regarding its plans, directions and intent are subject to change or withdrawal without notice at IBM's sole discretion

*Information regarding potential future products is intended to outline our general product direction and it should not be relied on in making a purchasing decision.*

The information mentioned regarding potential future products is not a commitment, promise, or legal obligation to deliver any material, code or functionality. Information about potential future products may not be incorporated into any contract. The development, release, and timing of any future features or functionality described for our products remains at our sole discretion.

- Brett South and Irene Dankwa-Mullan are full-time employees of IBM Watson Health, a maker of AI products and regulated medical devices.
- Brett South holds a small amount of equity in ©IBM Corporation.

# Defining Health Disparities\*

## **Social disparities in healthcare:**

Disparities patterned by socioeconomic status (includes race) or social disadvantage

## **Racial and/or ethnic disparities in healthcare:**

Divergent health outcomes patterned by racism, discrimination, and exclusion

## **Biological factors and healthcare disparities:**

Differing distributions of risk in certain populations that create a unique susceptibility or protection

\*Payne P. The Journey of Health Equity data. Presented at: Watson Health Expert Series, Nov 9, 2021.



# Putting Health Disparities in Context\*

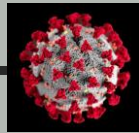
## Covid-19 Pandemic

### Healthcare

- Disparities
- Variations
- Differences

### Inequities

### Inequalities



### Physical/Personal socio-demographics

- Age
- Gender
- Sexual orientation
- Race, ethnicity, country of origin
- Religious affiliation
- Disability status

### Other socio- demographics

- Employment
- Housing status
- Income/Wealth
- Insurance Status
- Education
- Geography

\*Payne P. The Journey of Health Equity data. Presented at: Watson Health Expert Series, Nov 9, 2021.

# Integrate Equity and Justice Principles into Ethical AI

There are several ways in which AI systems, including the data and evidence on which they are trained can cause harm, each with ethical, social and equity implications

## Ethical AI Dimensions

- Accountability
- Impact of Algorithms
- Data Responsibility
- Design equity
- Discrimination and Bias
- Empathy
- Explainability
- Fairness
- Human Oversight

- Human Autonomy
- Inclusion
- Social Cohesion
- Inclusive Technology
- Moral Agency
- Privacy Protection
- Robustness, Safety
- Transparency and Trust
- Value Alignment

# IBM Commitment to Health Equity



Building **inclusive technology** and promoting inclusive language



**Code for Justice** & Call for Code Emb(race) challenge



Mitigating **artificial intelligence bias**



Building **open and transparent AI**



**Hospital equity metric:** Incorporating community health measures with a focus on equity for quality performance



**TechQuity initiative:** Integrating health equity, racial justice and ethical AI



Building trustworthy AI to ensure that it **does not further disadvantage groups**



**IBM Tech for Justice** to devise novel and practical ways to promote fairness, racial equity and social justice



Health Insights: Building **health equity dashboards** and metrics



**Design justice** for racial equity



**Social determinants of health mapping** and precision geography with demographics



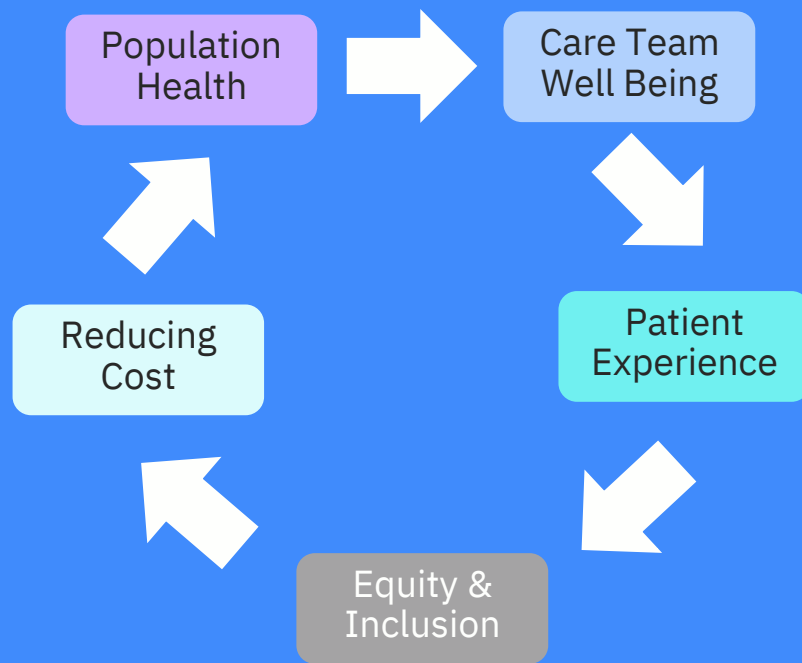
Health equity consulting: **Supporting health impact assessments**



## TechQuity initiative: Integrating health equity, racial justice and ethical AI

- **TechQuity\*** is the strategic development and deployment of technology in health care to achieve health equity and achieve racial justice and ethical AI.
- Technology can be used as a force for promoting equity in health, in ways that begin to address root causes of structural racism.

\*Rhee K. Inventing and predicting a new era of health with AI and Techquity. Presented at: Discover Brigham, Boston (MA), Nov 2019.



Quintuple Aim centered on **equity and inclusion** and founded on the principles of privacy, fairness, explainability, transparency, and trust.

# Challenges for AI Adoption in Healthcare

## Data Challenges

- Sparse, inaccurate, unbalanced, or biased data
- Missing data, misaligned in time
- Data collection biases
- Data rich, information poor?
- Interoperability issues
- Issues with data standardization / terminologies

## Security, Privacy, and Compliance Concerns

**HIPAA**



## Human Physiology

- Incomplete knowledge of human physiology
- The effects of genetics and lifestyle choices
- Health history
- Life course history and trajectory
- Labs, imaging data
- Social, environmental, and demographic data



Clinician data sources



Patient level data





# Mitigating artificial intelligence bias

- **Algorithmic bias:** refers to a situation in which the outcome is skewed by something inherent in the underlying data or the way in which it is used, leading to an inaccurate or misleading results.

## Sources of Bias:

### 1. The data used to build AI solutions:

- Different groups within our society are often underrepresented in clinical data sources, resulting in data that may be incomplete, inaccurate, or misleading.

### 2. AI modeling activities:

- In computational health, existing domain knowledge is often injected during the modeling process.



## Building open and transparent AI

- **Explainable AI:** Explainable AI must include some knowledge and understanding about different types of system training, validation, and real-world applications.
- **Trust and Transparency:** Trust is gained by providing explanations and moving away from “black box” models to “clear box” transparent applications.

When a physician (*or consumer*) wants to know if they can trust a system?

What do they need to know?

How will they decide if they should trust it?

# Differing Perspectives on Establishing Trust, Transparency, and Explainable AI\*

## Data Scientist

### Domain Knowledge:

- Speaks “AI, ML, Stats”
- Attention schemes
- Shapelets
- Model agnostic Models

### Main Focus:

- Model level performance and transparency
- Model level explanations

## Clinical Researcher

### Domain Knowledge:

- Clinical background
- Medical Informatics,
- Speaks “applied” AI, ML, Stats
- Speaks/understands medical concepts

### Main Focus:

- Model level performance and transparency
- Instance level explanations

## Clinician

### Domain Knowledge:

- Clinical expert, speaks in medical terms
- Often limited background in biomedical informatics
- May have a limited statistics background

### Main Focus:

- Patient centric focus
- Instance level explanations

Model centric

User/patient centric

\*Feldman H. Explainable AI Challenges and Future Directions. Presented at: IBM-Sponsored Roundtable Hosted by AMIA. April, 20, 2021.

# Trusting AI in Healthcare

TECHNOLOGY

## AI Can Outperform Doctors. So Why Don't Patients Trust It?

by [Chiara Longoni](#) and [Carey K. Morewedge](#)

October 30, 2019

**Harvard  
Business  
Review**

harnessing the full potential of these and other consumer-facing medical AI services will require that we first overcome patients' skepticism of having an algorithm, rather than a person, making decisions about their care.

41,511 views | Feb 11, 2019, 12:52pm EST

## Rethinking Medical Ethics

**Insights Team** Insights Contributor

**Forbes** insights **Intel AI** **FORBES INSIGHTS** | Paid Program  
Innovation



**Forbes**

**T**he ethical guidelines laid out in the Hippocratic Oath nearly 2,500 years ago are about to collide with 21<sup>st</sup> century artificial intelligence (AI).

AI promises to be a boon to medical practice, improving diagnoses, personalizing treatment, and spotting future public-health threats. By 2024, experts predict that healthcare AI will be a nearly \$20 billion market, with tools that transcribe medical records, assist surgery, and investigate insurance claims for fraud.

Even so, the technology raises some knotty ethical questions. What happens when an AI system makes the wrong decision—and who is responsible if it does? How can clinicians verify, or even understand, what comes out of an AI “black box”? How do they make sure AI systems avoid bias and protect patient privacy?



Building trustworthy AI to ensure that it does not further disadvantage groups

**Prioritize Health Equity and Justice Principles**

Equity and justice principles are paramount and foundational in the continuum of AI design, development, and implementation



# A Proposed Framework on Integrating Health Equity and Racial Justice into the Artificial Intelligence Development Lifecycle

Irene Dankwa-Mullan, Elisabeth Lee Scheufele, Michael E. Matheny, Yuri Quintana,  
Wendy W. Chapman, Gretchen Jackson, Brett R. South



Journal of Health Care for the Poor and Underserved, Volume 32, Number 2, May 2021 Supplement, pp. 300-317 (Article)

Published by Johns Hopkins University Press  
DOI: <https://doi.org/10.1353/hpu.2021.0065>

# A Framework to Integrate Equity and Justice Principles into Ethical AI Development in Healthcare



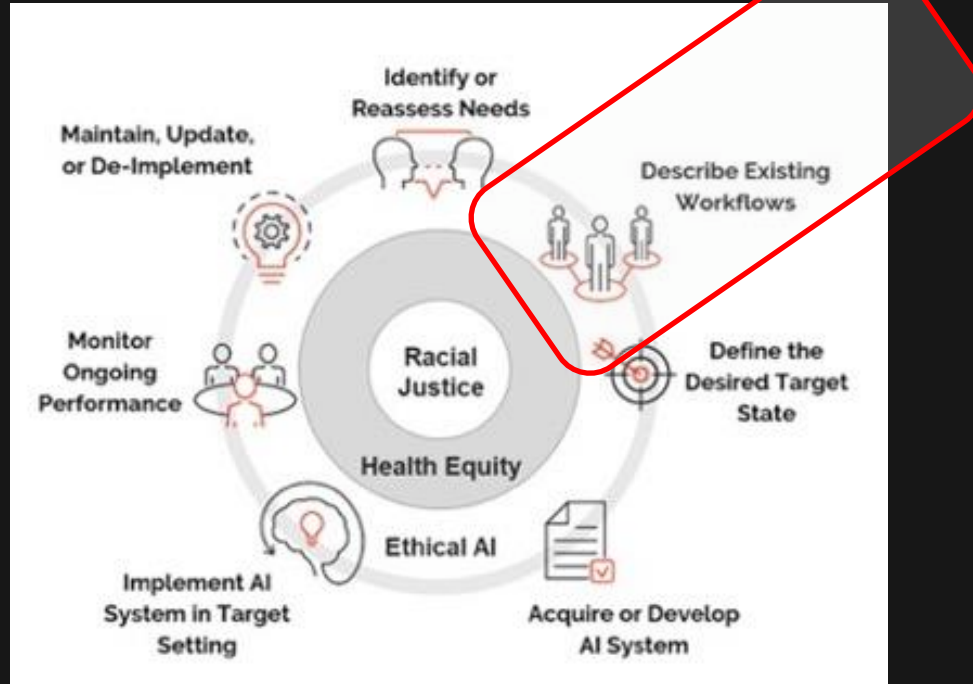
# Before AI Is Brought Onboard...



- Put the **Need** before the **Tool**
- Clearly identify the objectives and outcomes
- Ensure resource allocation
- Identify the business owner and stakeholders
- Assess organizational readiness for change
- Examine data assets
- Obtain feedback from stakeholders
- Outline how the system will address gaps



# Before AI Is Brought Onboard...



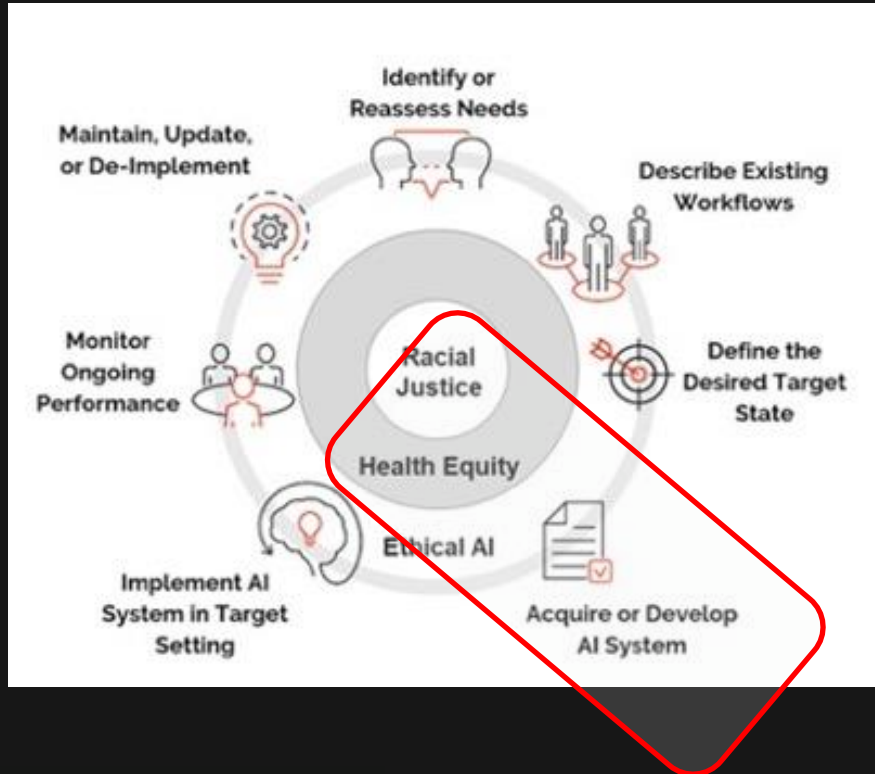
- Describe and map the existing state
- Understand the workflow and current practices and policies
- Identify potential workflows that exacerbate inequities
- Assess feasibility of workflows needs
- Understand perceived assets and barriers
- Determine necessary training and resources needed

# Before AI Is Brought Onboard...



- Define the target state
- Define measurable outcomes
- Establish equity-sensitive metrics and key performance indicator
- Define prediction-action pairs
- Promote humility in self-awareness of systematic racism
- Ensure evaluation by target end-users

# When AI is Developed/Implemented...



- Acquire/Develop AI system
- Understand relevant tools for feature engineering and model building
- Promote correction of internal algorithmic bias
- Ensure user centered design to uncover and address bias
- Define steps to integrate ethical AI and foster accountability, trust, fairness, and privacy

# When AI is Developed/Implemented...



- Implement AI system in the appropriate target setting
- Continue to ensure user centered design to uncover and address bias
- Continue to outline steps to integrate ethical AI
- Engage stakeholders in implementation processes

# When AI is Developed/Implemented...



- Monitor ongoing system performance
- Evaluate performance assessing factors that include health equity measures
- Assess how often tool is accessed and used in care delivery
- Monitor how often recommendations are accepted
- Monitor performance against historical data
- Review AI literature on bias detection, drift, algorithmic error

# When AI is Developed/Implemented...



- Maintain, apply appropriate updates, or de-implement
- Conduct routine AI model maintenance, and continuous training
- Maintain and establish trust and transparency with stakeholders
- Maintain and update policies to ensure principles of ethical AI are integrated in the system lifecycle

# Conclusions and Future Directions

Adoption of this framework provides a **strategic approach** that considers **health equity, racial justice**, and **ethical principles**

Promoting safe and effective adoption of AI and mitigating risks of exacerbating health disparities

