



HEIDELBERG
FACULTY OF
MEDICINE



Who Benefits? Justice, discrimination, and better health outcomes in digital healthcare

PD Dr. Nadia Primc

Institute for the History and Ethics of Medicine

Primc@uni-heidelberg.de

- I. Justice in healthcare: forms and foundations
- II. AI and the reproduction of injustice
- III. Summary



**Co-funded by
the European Union**

Mind the gap!

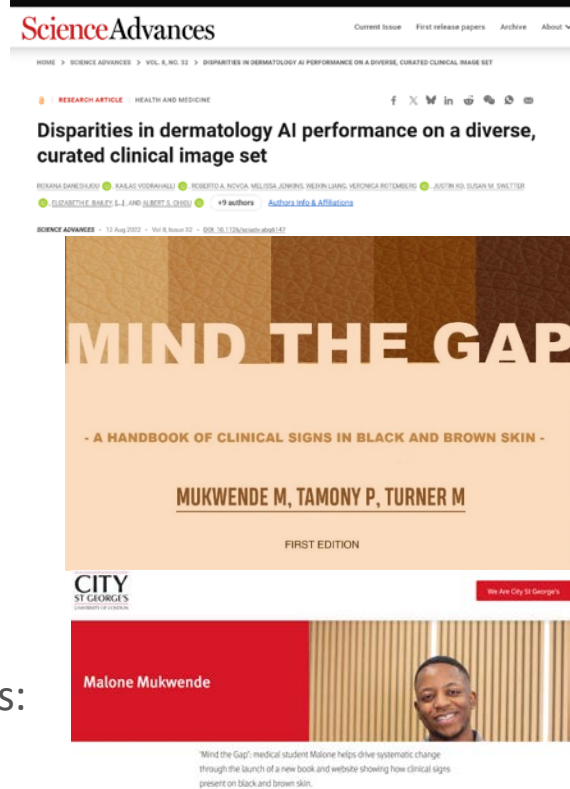
Malone Mukwende, “Mind the gap” (2020)

- Medical textbooks show only white skin
- Clinical signs look different on darker skin
- Decades of medical training did not teach this
- *"The reference body of medicine is white"*

Dermatology AI (Daneshjou et al., 2022)

- ISIC dataset: >70% light skin, <8% dark skin
- AI performance drops dramatically on darker skin
- Higher melanoma mortality in darker-skinned patients: worsened by delayed diagnosis
- Risk of automation bias: AI perceived as objective

The gap predates AI. AI reproduces it: automatically, at scale, and with a claim to objectivity.



The gender data gap: #DeathByChatbot

The Babylon Test (Dr. Murphy / Dr. Watkins, 2019):
Patient: 59-year-old smoker. Sudden chest pain and nausea.

	FEMALE	MALE
Diagnosis suggested	Panic attack or depression	Possibility of serious heart problems
Recommendation	Stay at home	Go to emergency room, call ambulance

Women are already 50% more likely than men to be misdiagnosed after a heart attack. The app learned from decades of clinical research focused primarily on male patients: the gender data gap (e.g. gendermedicine)

“The reference body of medicine is not only white — it is male.”



Does an App a Day Keep the Doctor Away? AI Symptom Checker Applications, Entrenched Bias, and Professional Responsibility

Moin H Zawalil, Michael Lang



Dr Murphy (aka David Watkins)
 @DrMurphy11

Afficher la traduction

The @babylonhealth Chatbot has descended to a whole new level of incompetence, with #DeathByChatbot #GenderBias.

Classic #HeartAttack symptoms in a FEMALE, results in a diagnosis of #PanicAttack or #Depression.

The Chatbot ONLY suggests the possibility of a #HeartAttack in MEN!

Some of the symptoms you reported might need to be checked out by a GP within the next 6 hours.

Based on the information you gave, some possible causes are listed below.

If symptoms persist, worsen or you are concerned, seek further medical advice.

Panic attack
 A sudden period of intense fear and anxiety.
 This can usually be treated at home.

Depression
 A persistent low mood/loss of interest.

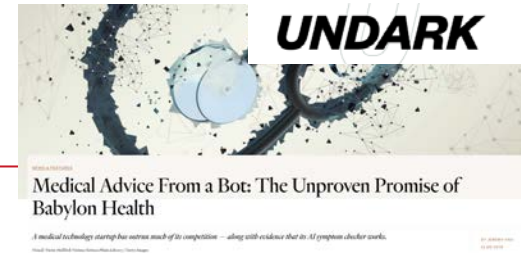
eMed UK et 2 autres personnes

9:02 PM - 8 sept. 2019

Pericarditis
 An inflammation of the thin membrane surrounding the heart.
 This is usually treated at the emergency department.

This usually requires seeing a GP urgently.

Panic attack
 A sudden period of intense fear and anxiety.
 This can usually be treated at home.



- I. Justice in healthcare: forms and foundations**
- II. AI and the reproduction of injustice
- III. Summary



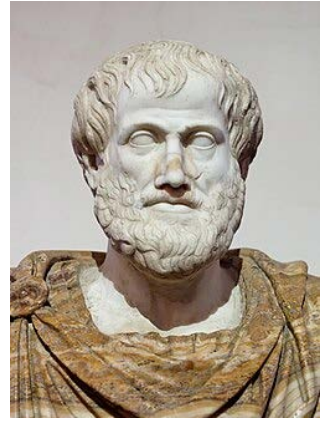
**Co-funded by
the European Union**

Forms of justice: Aristotle

Distributive Justice = Fair allocation of goods, resources, and burdens

Corrective Justice = Fair terms in transactions and exchanges

- **Voluntary transaction:** fair prices, fair salaries, fair working conditions
- **Involuntary transaction:** compensation for harm done



Distributive justice as the most fundamental form of justice in the healthcare system!

Who receives what? Who gets access to treatment, diagnostics, medical innovation?

Distributive justice: formal and material principles

The Formal Principle: Equals must be treated equally, unequals unequally in proportion to relevant differences.

But: who counts as equal? What counts as a relevant difference?

Material Principles:

- Ability to pay: Free-market exchange (e.g. the US system)
- Merit or effort: Health behaviour, contributions, insurance payments
- Need: The more care you need, the more you receive ✓
- Benefit: Prospects of success (Will the intervention actually help?) ✓

When need and benefit conflict: *Need + benefit must be balanced, especially in case of scarce resources (e.g. donor organs)*

The limits of distributive justice: social determinants of health

The problem: *Need-based distribution asks “How do we distribute care fairly once someone is sick?”, but not “Why are certain people sicker in the first place?”*

WHO and the social determinants of health

- The conditions in which people are born, grow, live, work, and age (e.g. income, housing, education, employment, food security, environment)
- Distributed unequally along lines of race, gender, and socioeconomic status
- Health is not only a biological fact, but also a social one (social justice)

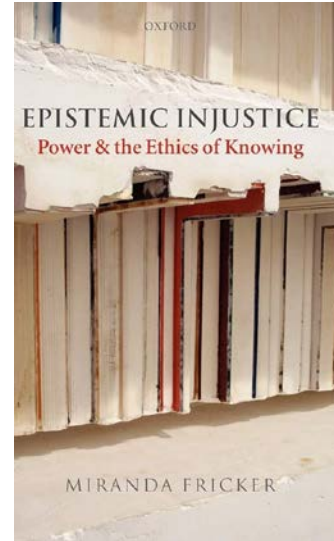
An AI system trained on health data has not only learned from biology.

→ It has learned from the outputs of a social system that already distributes health and illness unequally!

Beyond distribution: epistemic justice (Fricker 2007)

People can be wronged not only in what they receive, but also in how they are known, heard, and understood!

- **Testimonial Injustice:** Being given less credibility than deserved because of who you are (*e.g. the patient whose pain is not believed, who is told it is anxiety, who is sent home*)
→ Rooted in prejudice related to race, gender, class
- **Hermeneutical Injustice:** A gap in shared concepts and categories leaves experience unintelligible (*e.g. sexual harassment; example from healthcare: ME/CFS - Myalgic Encephalomyelitis/Chronic Fatigue Syndrom*)



Both forms of injustice leave traces in clinical records and in the data that AI systems learn from.

- I. Justice in healthcare: forms and foundations
- II. AI and the reproduction of injustice**
- III. Towards ethical AI in healthcare



**Co-funded by
the European Union**

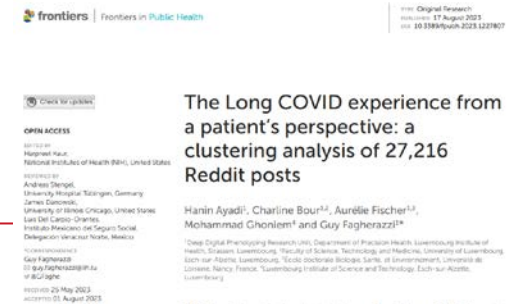
Biases in training data: layers of injustice



The core problem: AI learns from historical data that was generated by a healthcare system that already embodies the inequalities we have discussed.

Layer	Form of injustice
Underrepresentation and structural inequalities	Distributive injustice and social determinants of health
Credibility bias in documentation	Testimonial injustice
Misclassification	Hermeneutical injustice

But AI can also help: By analysing large volumes of data and patient-reported experiences, AI can identify patterns that fall outside existing clinical categories!



The digital divide and the digital determinants of health

The core problem: Even a perfectly unbiased AI system produces unequal outcomes if access to it is unequal!

- Different levels of the digital divide, e.g.
 - infrastructure (*e.g. no smartphone, or reliable internet*)
 - literacy (*e.g. digital and health literacy unevenly distributed*)
 - language (*tools designed in dominant languages, BUT: can also help to translate*)
 - trust/willingness to use tools (*e.g. worsens representation gaps in training data*)

Digital Determinants of Health

- Access to digital technology is itself a determinant of health outcomes (e.g. income). The digital divide is not a technical problem, but a structural health equity issue. It deepens an existing problem.

Accuracy versus fairness: a question of justice

The problem: Overall accuracy = average performance across the whole dataset. If the dataset is skewed, the average hides inequality (e.g. dermatology app)

Two sources of tension

- Biased training data: Fix the data → much of the tension dissolves
- Unequal disease distribution: Even with good data, choices remain: Whose health do we prioritise? What does fairness mean?

These are not technical questions. They are questions of justice and they must not be hidden inside a performance metric!

Regulatory and ethical frameworks: are they adequate?

Framework	Type	Key relevance
HLEG Guidelines (EU, 2019)	Soft law: ethical guidelines	Foundational ethical document for AI in Europe
EU AI Act (2024)	Hard law: legally binding	High-risk classification for clinical AI
European Health Data Space (2025)	Hard law: legally binding	Governs access to health data for AI training
WHO Guidelines (2021)	Soft law: aspirational	Global perspective on equity and inclusion

Main question: Do these frameworks adequately address the forms of injustice we have identified? (e.g. distributive justice, social determinants, epistemic injustice)

The Regulatory Landscape: HLEG Guidelines for Trustworthy AI (2019)

Seven requirements for trustworthy AI

- Human agency and oversight
 - Technical robustness and safety
 - Privacy and data governance
 - Transparency
 - Diversity, non-discrimination and fairness
 - Societal and environmental wellbeing
 - Accountability
- **Distributive justice:** Non-discrimination and fairness, but no specification of which material principle takes priority
 - **Social determinants:** Societal wellbeing, acknowledged but not developed
 - **Epistemic injustice:** Not addressed

These are guidelines, not law. High ethical ambition, but no enforcement, no specificity, not designed for medicine.

The Regulatory Landscape: The EU AI Act (2024)

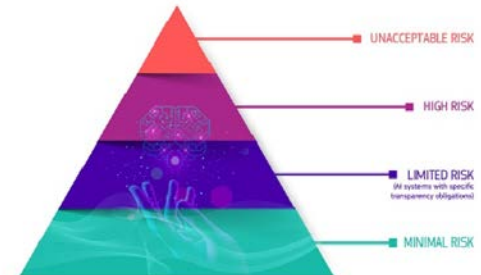
Risk classification: Unacceptable (*prohibited*) - **High risk** - Limited risk - Minimal risk
Most clinical AI falls into the high-risk category.

Requirement	What it says	Limitation
Data governance (Art. 10)	Datasets must be sufficiently representative and examined for bias	Does not specify which groups, which metrics, what counts as sufficient
Human oversight (Art. 14)	Clinicians must be able to understand, interpret and override outputs. Explicitly names automation bias	Does not specify what meaningful oversight requires in practice

A Risk-based Approach

The AI Act defines 4 levels of risk for AI systems:

Genuine steps forward, but stated at a high level of abstraction.
 The Act does not engage with the categorical problem, epistemic injustice, or the social determinants.



The Regulatory Landscape: The European Health Data Space (2025)



What it addresses: Who is allowed to use patient health data and for what purposes?

Every clinical encounter generates data. Access to that data for research and AI training has been governed inconsistently across EU member states.

What it does:

- Creates a common EU-wide framework for secondary use of health data
- Explicitly permits training and testing of AI algorithms as a secondary use
- Requires secure processing environments
- Gives patients the right to opt out

But: Governs access, not quality, representativeness, or fairness of data.

Whose data is still missing? Who benefits?

The Regulatory Landscape: WHO Ethics Guidelines for AI in Health (2021)

Six principles: Human autonomy, wellbeing and safety, transparency, accountability, inclusiveness and equity, sustainability

What sets it apart: The only framework that explicitly addresses global health inequalities and low- and middle-income countries.

But: Aspirational — no enforcement mechanism.

Tells us what a just framework would need to include. Does not deliver it.

Overall assessment: The regulatory landscape is necessary — but not sufficient.



- I. Justice in healthcare: forms and foundations
- II. AI and the reproduction of injustice
- III. Summary**



**Co-funded by
the European Union**

Closing the trilogy

Lecture 1: What do the principles of medical ethics require when AI is involved in a clinical decision?

Lecture 2: What does it mean to keep the patient in the loop?

Lecture 3: Which patients are actually in the loop?

A structural argument: The patients least well served by AI are those

- whose bodies were not the norm in medical knowledge production
- whose symptoms were not fully believed or recorded
- whose social conditions were not addressed by need-based distribution
- whose access to digital healthcare is limited by the same inequalities that make them sicker

Thank you very much for your attention!

Brief (!) Survey



Next lecture: Immersive Surgical Edu



05/05/2026 at 11:00



Prof. Filippo Sanfilippo