Department of Statistical Science

Critical Decisions

Managing AI Bias in Intensive Care Settings

Ali Akbar Septiandri — 24.02.2025

CHIMERA

Engineering and Physical Sciences Research Council

AI uses in ICUs

- Monitoring vital signs
- Mortality prediction
- Understanding medical documents
- Bed allocation

Photo by Anna Shvets

Photo by National Cancer Institute

“Among the top 20 occupations impacted by AI, seven are from the healthcare domains.” – Septiandri et al. (2024)

"Don't be a radiologist, be a plumber instead. It's much harder for AI to take your job!"

Biases in data acquisition

Photo by Google DeepMind

Photo by cottonbro studio

Photo by David McElwee

“Statistical bias refers to an algorithm that produces a result that differs from the true underlying estimate.”

(Parikh et al., 2019)

Highest-risk groups

Bias in the sampling process

- Breast cancer in women (Arnould et al., 2006; Giordano, 2018)
- Cardiovascular disease in men (Vogel et al., 2021)
- Skin cancers in whiter skins (Gloster Jr and Neal, 2006)

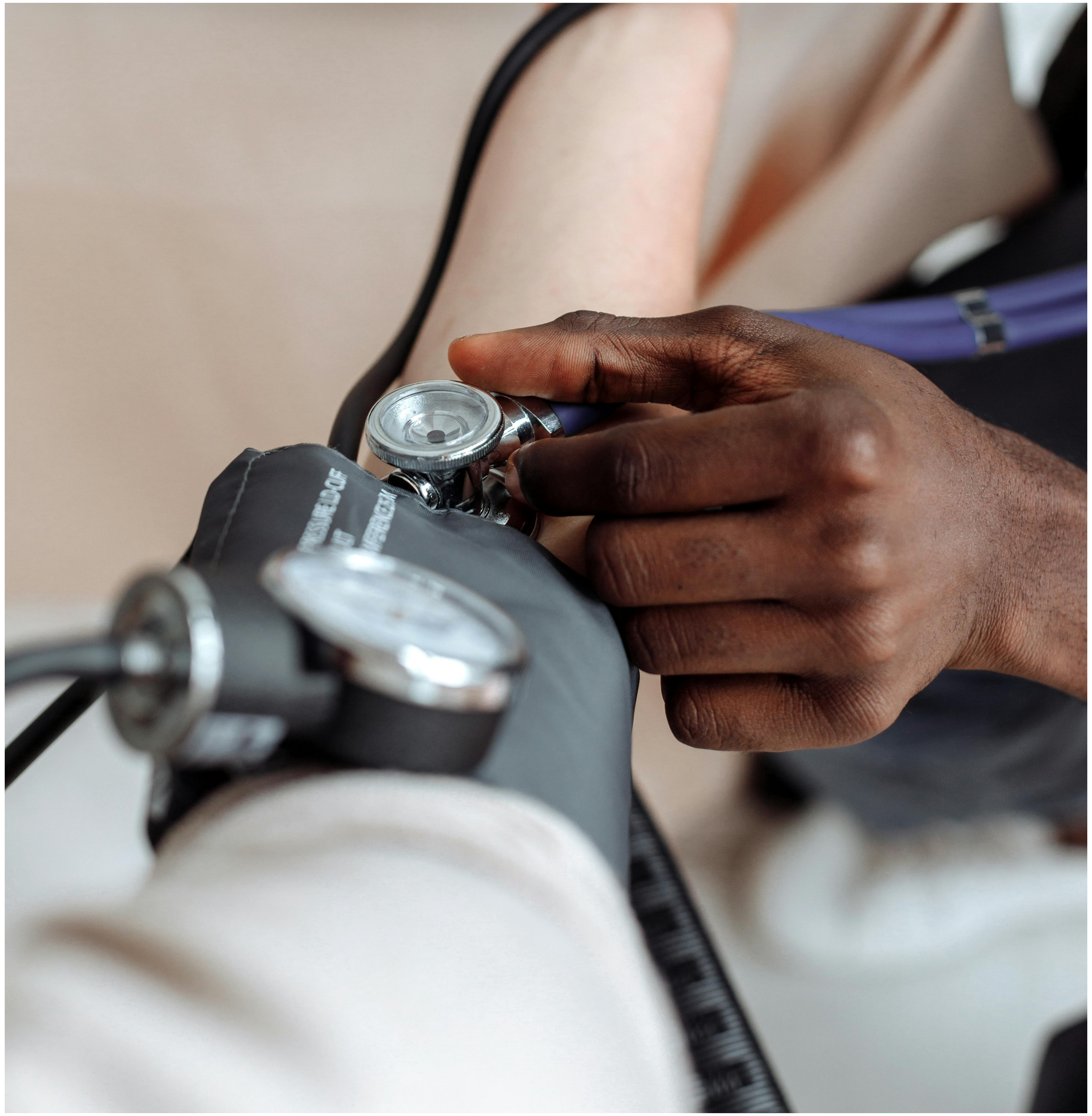
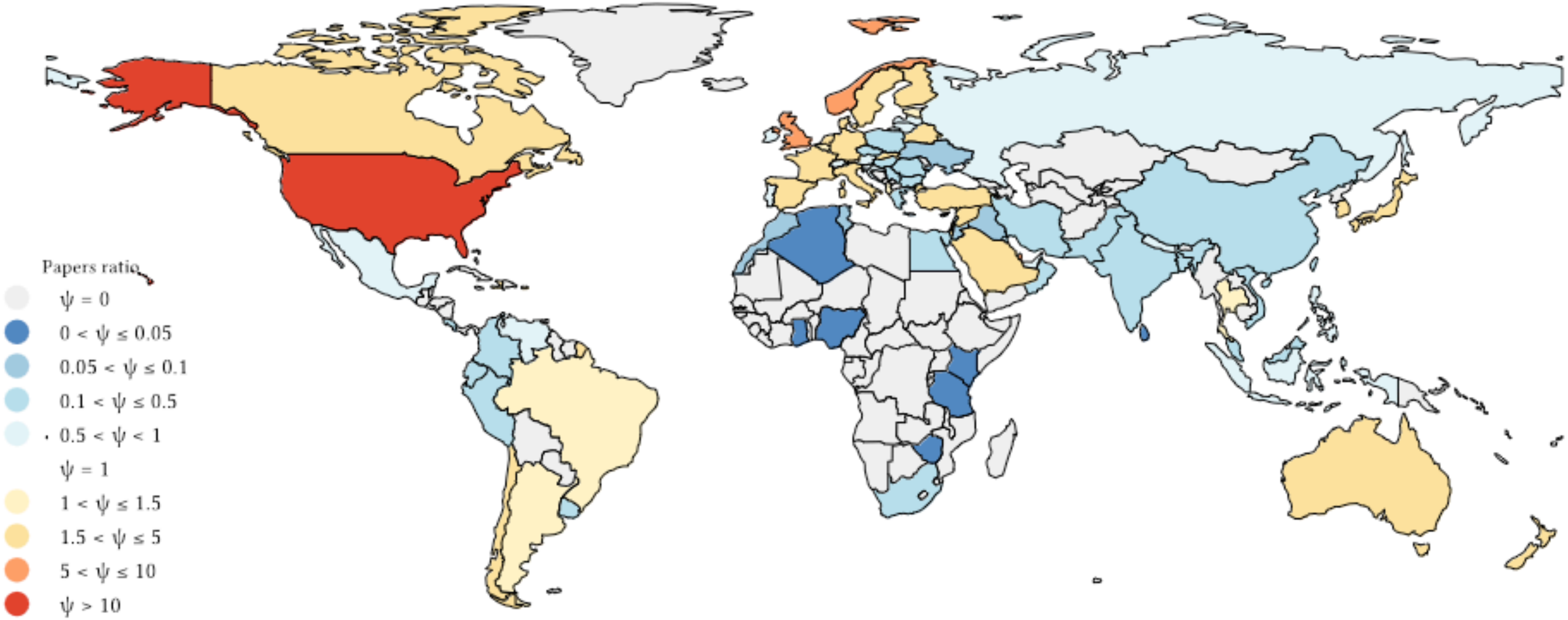


Photo by Thirdman



ACM FAccT authors are mainly
using off-the-shelf datasets from the U.S. (Septiandri et al., 2023)

"Social bias in health care refers to inequity in care delivery that systematically leads to suboptimal outcomes for a particular group."

(Parikh et al., 2019)

Disparity in access

- Socioeconomic status
- Race and ethnicity
- Geographic location
- Insurance coverage
- Language barriers

Photo by Leeloo The First

AI-assisted decision making

Photo by Tima Miroshnichenko

Intrinsic uncertainty in medicine

Cabitza et al. (2017)

Photo by Pixabay

Black box AI models

Cabitza et al. (2017)

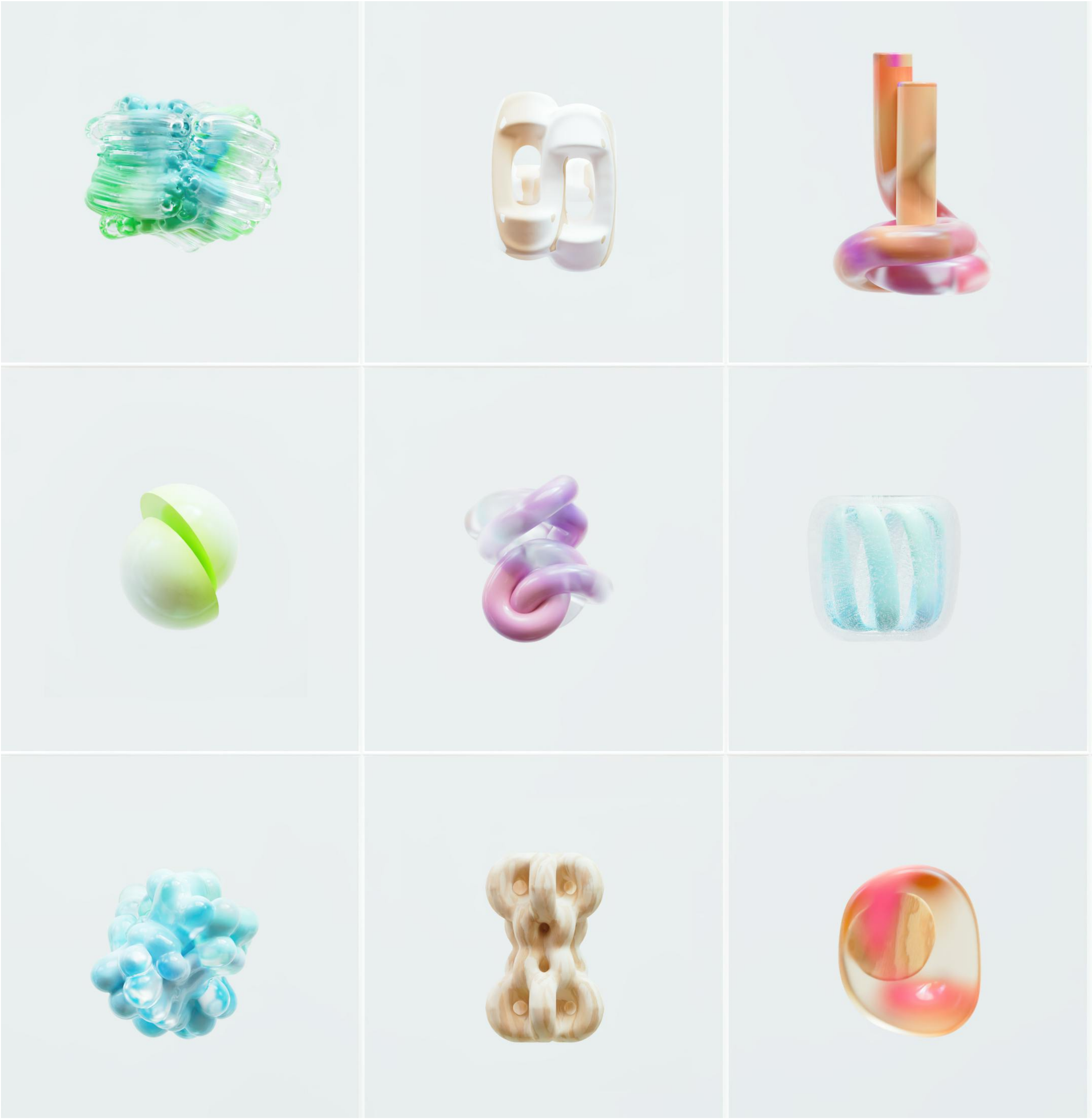


Photo by Google DeepMind

Privacy-utility trade-off

Photo by Pixabay

Publicly available critical care databases are mainly from the Global North

MIMIC-III (The Medical Information Mart for Intensive Care III)

Edit

Introduced by Johnson et al. in MIMIC-III, a freely accessible critical care database

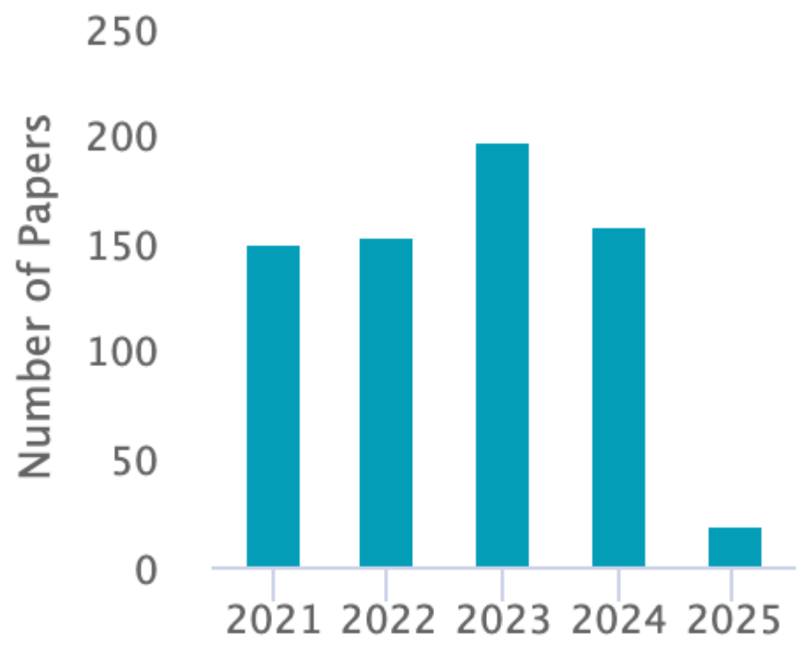
The Medical Information Mart for Intensive Care III (**MIMIC-III**) dataset is a large, de-identified and publicly-available collection of medical records. Each record in the dataset includes ICD-9 codes, which identify diagnoses and procedures performed. Each code is partitioned into sub-codes, which often include specific circumstantial details. The dataset consists of 112,000 clinical reports records (average length 709.3 tokens) and 1,159 top-level ICD-9 codes. Each report is assigned to 7.6 codes, on average. Data includes vital signs, medications, laboratory measurements, observations and notes charted by care providers, fluid balance, procedure codes, diagnostic codes, imaging reports, hospital length of stay, survival data, and more.

The database supports applications including academic and industrial research, quality improvement initiatives, and higher education coursework.

Source: MIT Laboratory for Computational Biology

Homepage

Usage



Federated learning

Collaboratively training a model while keeping their data decentralised

(Mondrejevski et al., 2022)

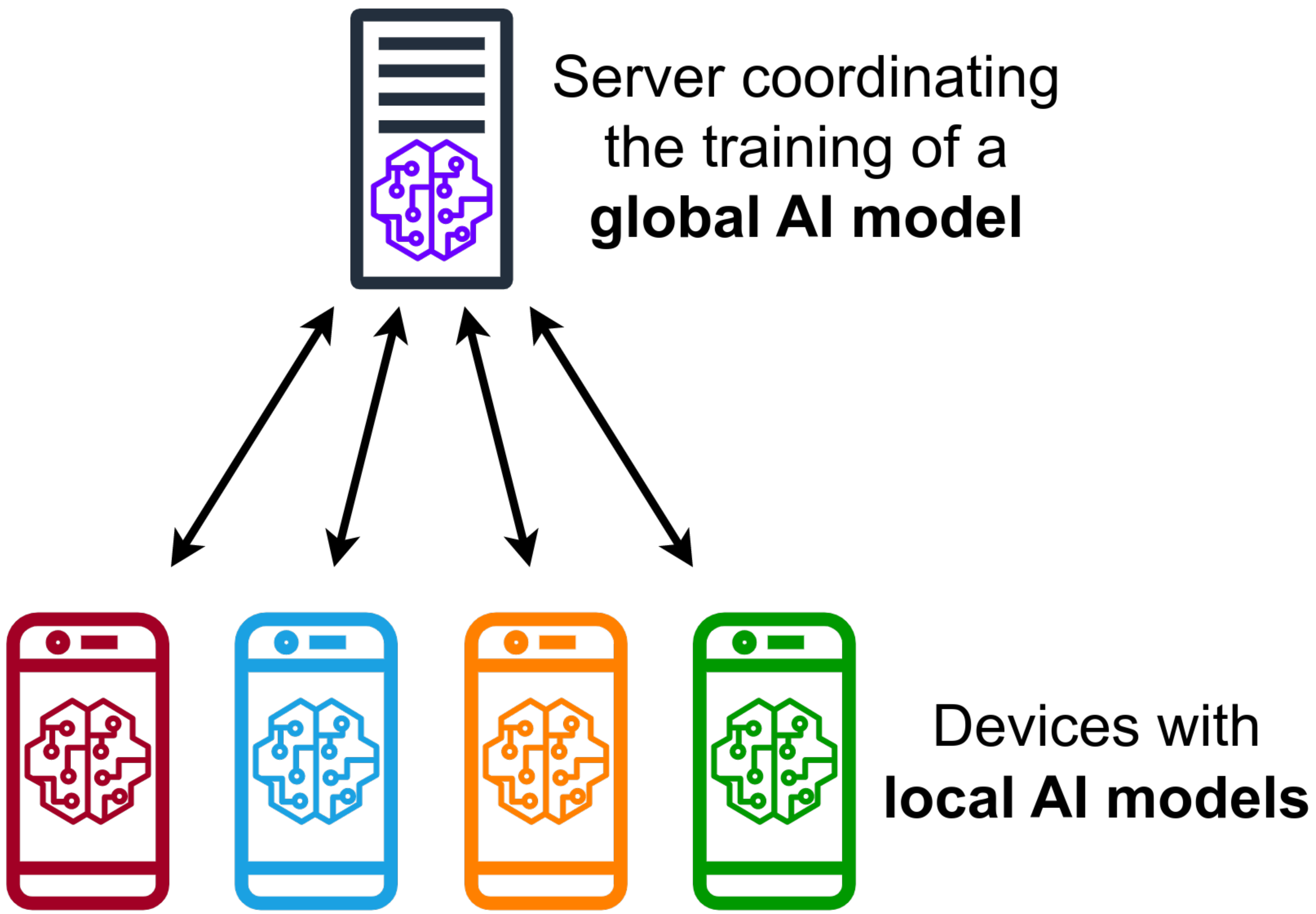


Image by MarcT0K

Mitigation strategies

Photo by Engin Akyurt

AI education for healthcare providers

Photo by Godfrey Atima

ICU dataset diversification

Photo by Google DeepMind


LREC-COLING 2024
Lingotto Conference Centre - Torino (Italia)
20-25 May, 2024
CHECK THE EVENT APP
*All Conference's updates will be communicated via the app.

Photo by Google DeepMind

Interpretable AI models

Post hoc interpretability

Contrasting factors associated with COVID-19-related ICU admission and death outcomes
(Cavallaro et al., 2021)

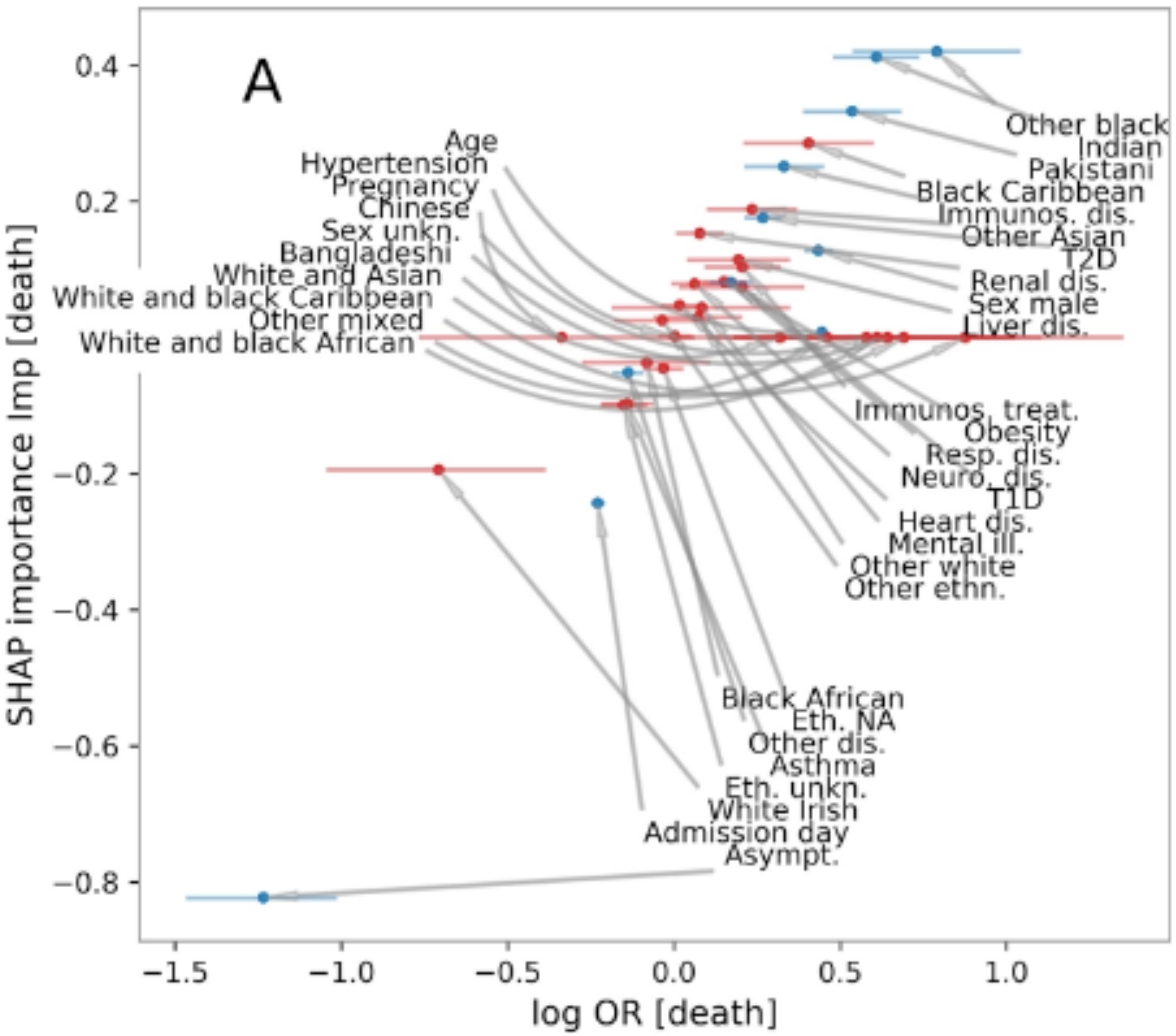
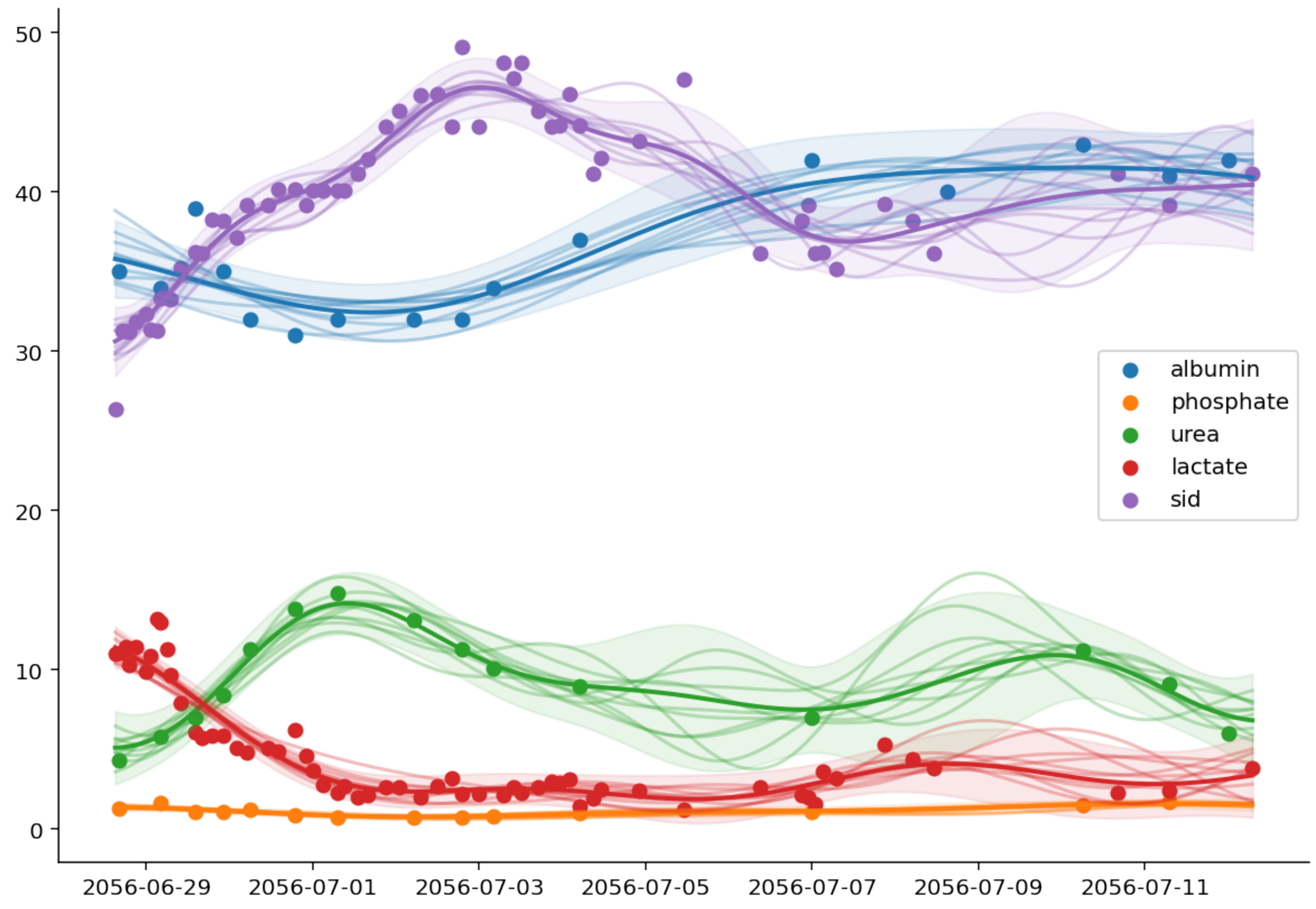


Image by Cavallaro et al. (2021)

Uncertainty quantification
(Work in progress)



Perspective | Published: 13 May 2019

Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead

Cynthia Rudin

Nature Machine Intelligence **1**, 206–215 (2019) | Cite this article

Photo by Maksim Goncharenok

Regulatory framework

Ethics and governance of artificial intelligence for health

WHO guidance

28 June 2021 | Guideline

Download (1.9 MB)

Overview

The WHO guidance on *Ethics & Governance of Artificial Intelligence for Health* is the product of eighteen months of deliberation amongst leading experts in ethics, digital technology, law, human rights, as well as experts from Ministries of Health. While new technologies that use artificial intelligence hold great promise to improve diagnosis, treatment, health research and drug development and to support governments carrying out public health functions, including surveillance and outbreak response, such technologies, according to the report, must put ethics and human rights at the heart of its design, deployment, and use.

The report identifies the ethical challenges and risks with the use of artificial intelligence of health, six consensus principles to ensure AI works to the public benefit of all countries. It also contains a set of recommendations that can ensure the governance of artificial intelligence for health maximizes the promise of the technology and holds all stakeholders – in the public and private sector – accountable and responsive to the healthcare workers who will rely on these technologies and the communities and individuals whose health will be affected by its use.

Read more

Artificial intelligence in healthcare: Applications, risks, and ethical and societal impacts

Study – 01-06-2022

In recent years, the use of artificial intelligence (AI) in medicine and healthcare has been praised for the great promise it offers, but has also been at the centre of heated controversy. This study offers an overview of how AI can benefit future healthcare, in particular increasing the efficiency of clinicians, improving medical diagnosis and treatment, and optimising the allocation of human and technical resources. The report identifies and clarifies the main clinical, social and ethical risks posed by AI in healthcare, more specifically: potential errors and patient harm; risk of bias and increased health inequalities; lack of transparency and trust; and vulnerability to hacking and data privacy breaches. The study proposes mitigation measures and policy options to minimise these risks and maximise the benefits of medical AI, including multi-stakeholder engagement through the AI production lifetime, increased transparency and traceability, in-depth clinical validation of AI tools, and AI training and education for both clinicians and citizens.

Study PDF EN (PDF - 1 MB)

Main risks

Clinical, social, ethical

- Potential errors and patient harm
- Risk of bias and increased health inequalities
- Lack of transparency and trust
- Vulnerability to hacking and data privacy breaches

Photo by Adrien Olichon

Key takeaways

- AI can be a powerful tool in ICUs, but require thoughtful implementation
- Ensuring a responsible AI implementation involve multiple stakeholders
- There should be active participation from the Global South in the collaborative effort

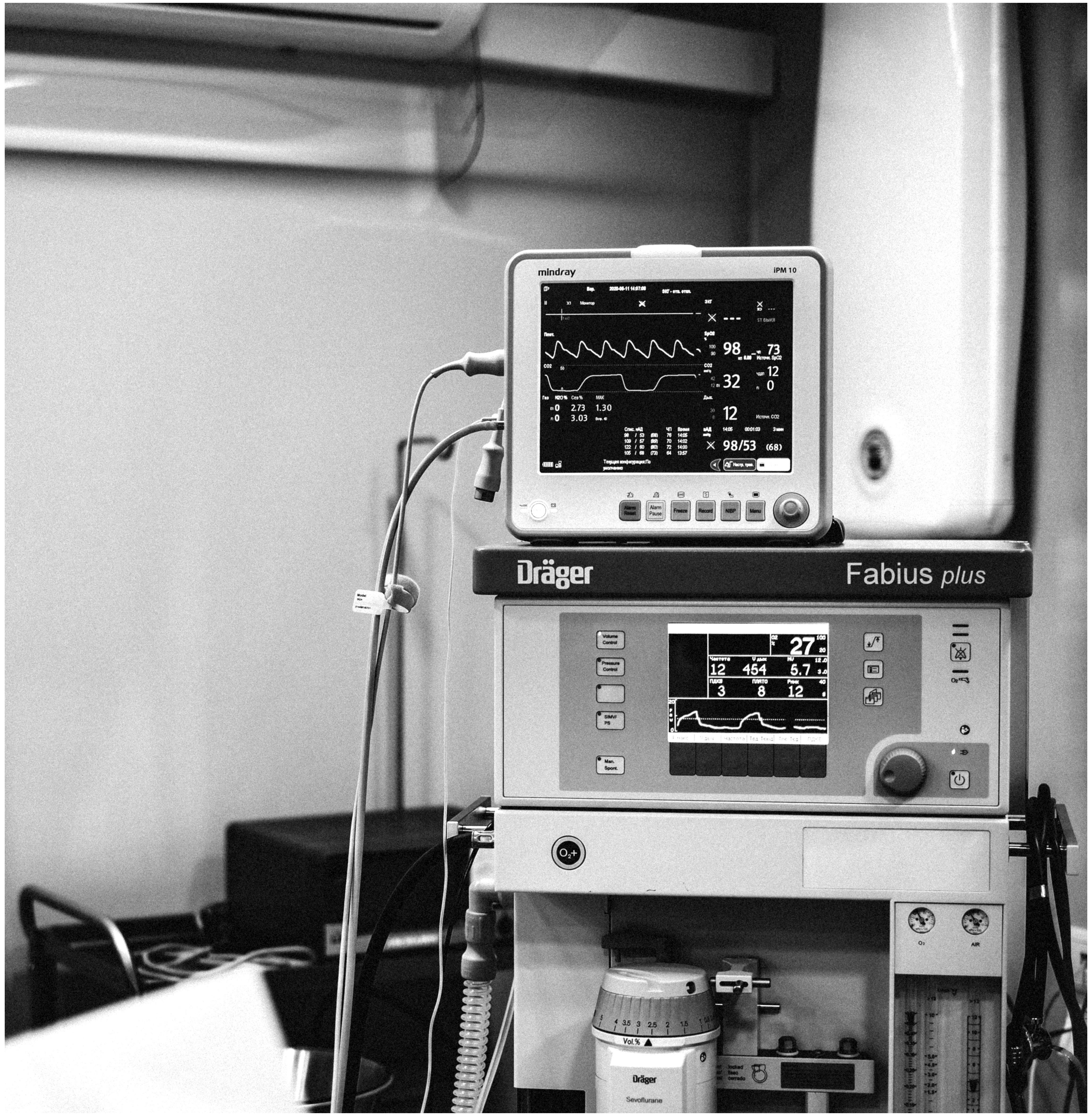


Photo by Anna Shvets

Thank you

@aliakbars.id