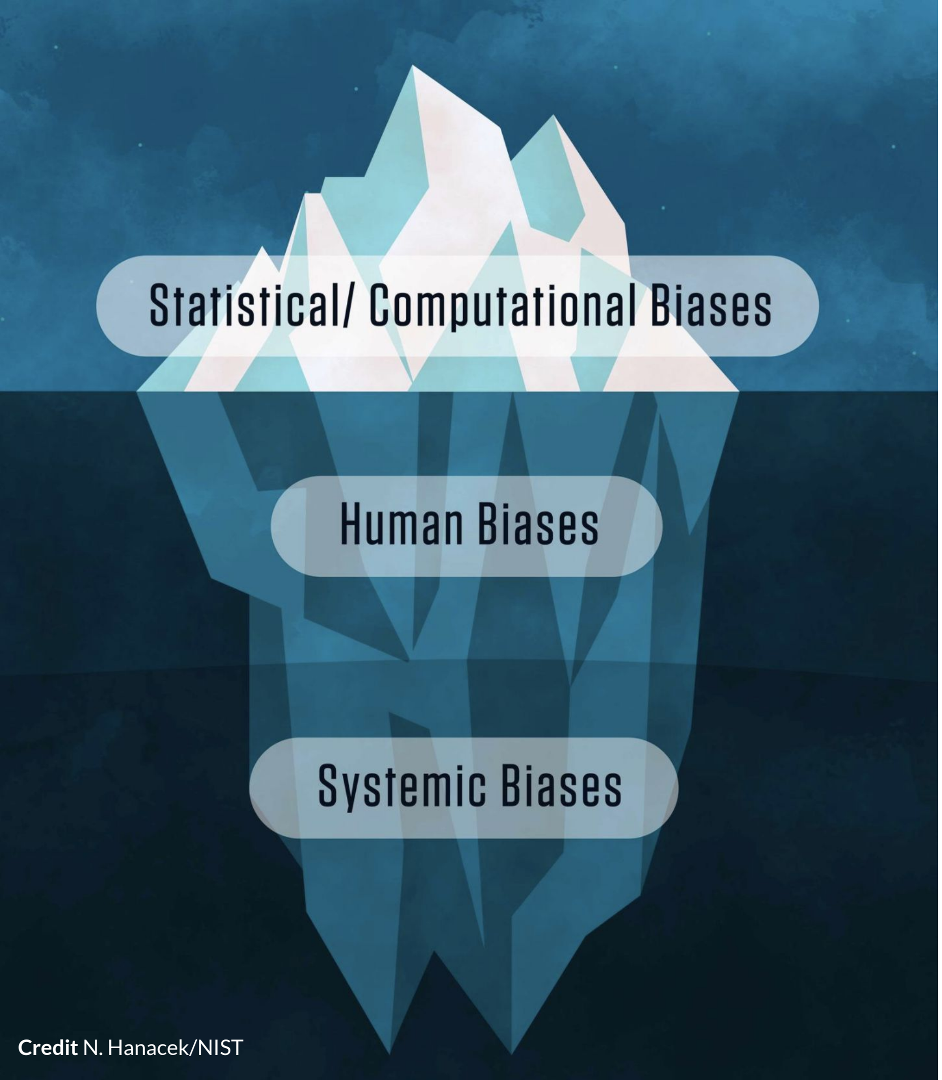
STANDING Together

Developing STANdards for data Diversity, INclusivity and Generalisability

Professor Alastair Denniston

Dr Xiao Liu



Credit N. Hanacek/NIST

*"**Missing data matters:** it can exacerbate inequalities on a societal scale.*

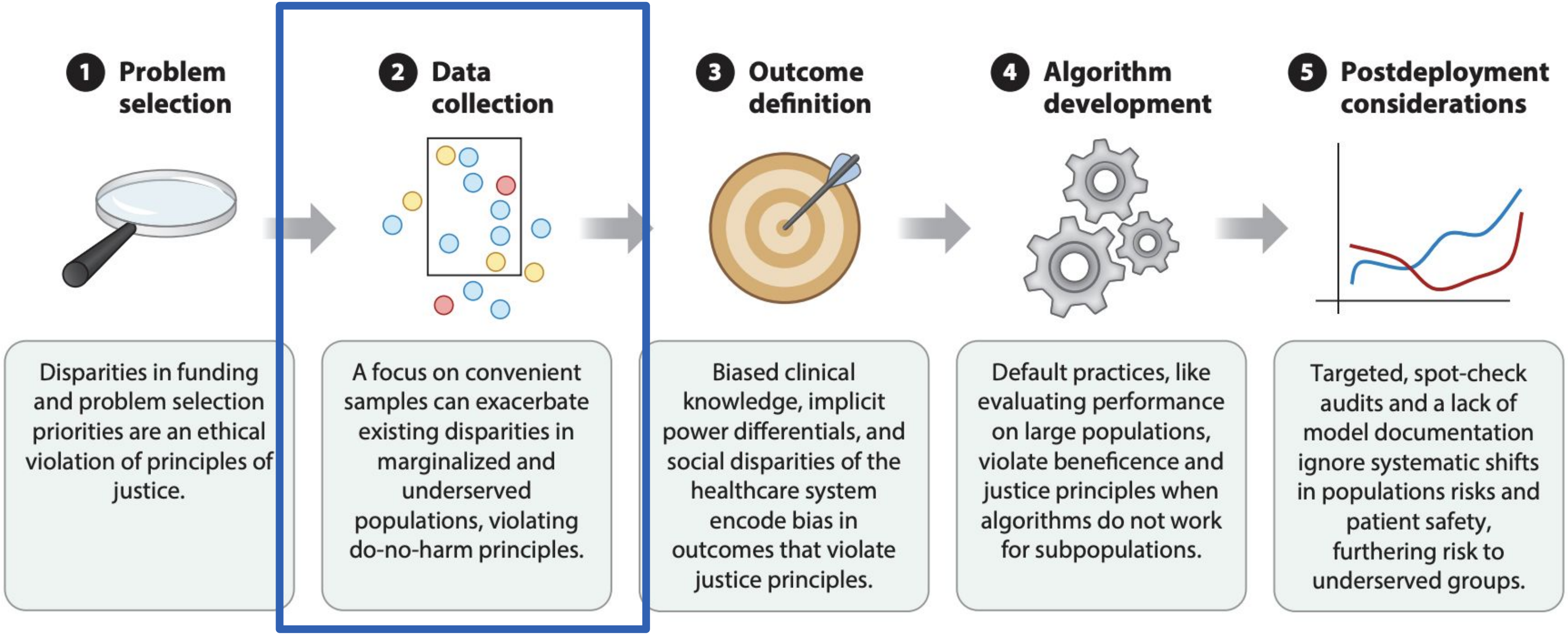
When that data is operationalised into algorithmic decision-making systems and AI, the social processes that produce racial inequality—mechanisms of power, economics, knowledge, culture and language—can be written into technologies with huge societal impacts."

- Ada Lovelace Institute

Annual Review of Biomedical Data Science

Ethical Machine Learning in Healthcare

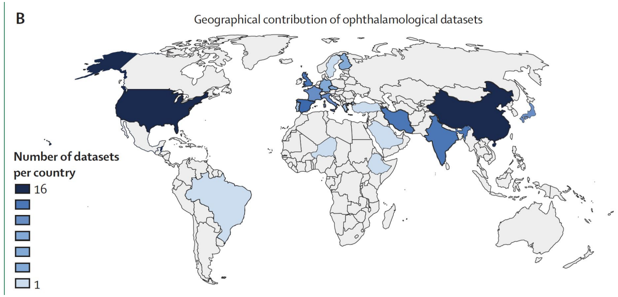
Irene Y. Chen,[1] Emma Pierson,[2] Sherri Rose,[3]
Shalmali Joshi,[4] Kadija Ferryman,[5]
and Marzyeh Ghassemi[1,6]



THE LANCET
Digital Health

A global review of publicly available datasets for ophthalmological imaging: barriers to access, usability, and generalisability

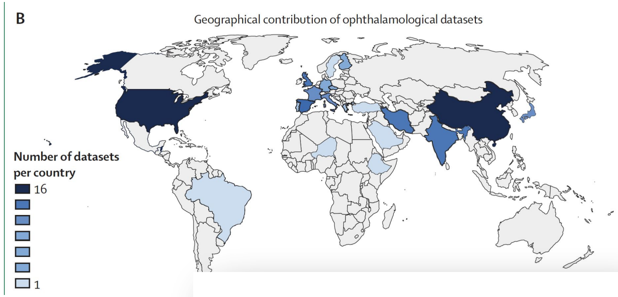
Saad M Khan, Xiaoxuan Liu*, Siddharth Nath, Edward Korot, Livia Faes, Siegfried K Wagner, Pearse A Keane, Neil J Sebire, Matthew J Burton, Alastair K Denniston*

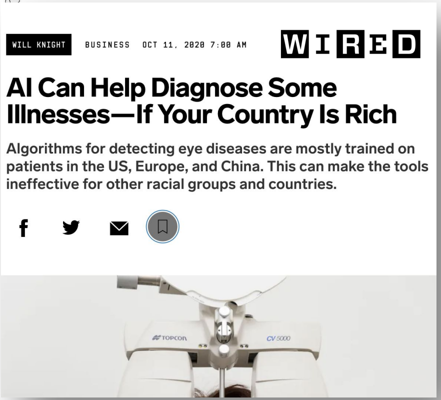


THE LANCET
Digital Health

A global review of publicly available datasets for ophthalmological imaging: barriers to access, usability, and generalisability

Saad M Khan*, Xiaoxuan Liu*, Siddharth Nath, Edward Korot, Livia Faes, Siegfried K Wagner, Pearse A Keane, Neil J Sebire, Matthew J Burton, Alastair K Denniston




WILL KNIGHT BUSINESS OCT 11, 2020 7:00 AM

WIRED

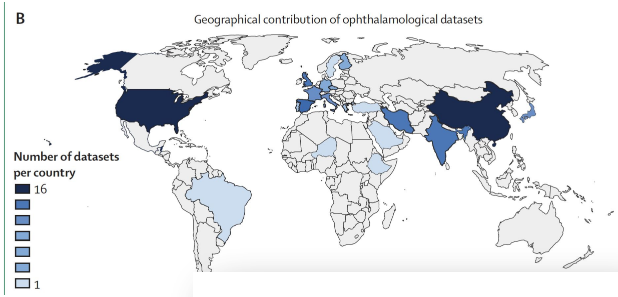
AI Can Help Diagnose Some Illnesses—If Your Country Is Rich

Algorithms for detecting eye diseases are mostly trained on patients in the US, Europe, and China. This can make the tools ineffective for other racial groups and countries.

THE LANCET
Digital Health

A global review of publicly available datasets for ophthalmological imaging: barriers to access, usability, and generalisability

Saad M Khan*, Xiaoxuan Liu*, Siddharth Nath, Edward Korot, Livia Faes, Siegfried K Wagner, Pearse A Keane, Neil J Sebire, Matthew J Burt
Alastair K Denniston

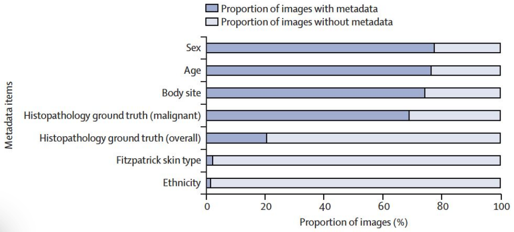


THE LANCET
Digital Health

Characteristics of publicly available skin cancer image datasets: a systematic review

David Wen, Saad M Khan, Antonio Ji Xu, Hussein Ibrahim, Luke Smith, Jose Caballero, Luis Zepeda, Carlos de Blas Perez, Alastair K Denniston,
Xiaoxuan Liu*, Rubeta N Matin*

Publicly available skin image datasets are increasingly used to develop machine learning algorithms for skin cancer diagnosis. However, the total number of datasets and their respective content is currently unclear. This systematic review aimed to identify and evaluate all publicly available skin image datasets used for skin cancer diagnosis by exploring their characteristics, data access requirements, and associated image metadata. A combined MEDLINE,



WILL KNIGHT BUSINESS OCT 11, 2020 7:00 AM WIRED

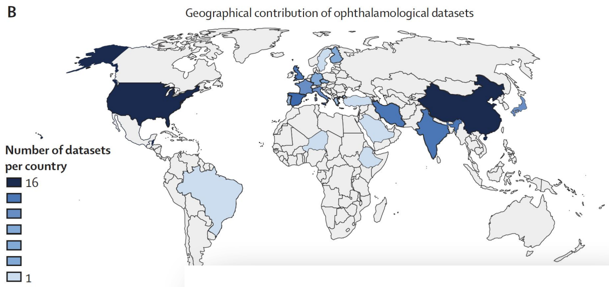
AI Can Help Diagnose Some Illnesses—If Your Country Is Rich

Algorithms for detecting eye diseases are mostly trained on patients in the US, Europe, and China. This can make the tools ineffective for other racial groups and countries.

THE LANCET
Digital Health

A global review of publicly available datasets for ophthalmological imaging: barriers to access, usability, and generalisability

Saad M Khan*, Xiaoxuan Liu*, Siddharth Nath, Edward Korot, Livia Faes, Siegfried K Wagner, Pearse A Keane, Neil J Sebire, Matthew J Burt
Alastair K Denniston

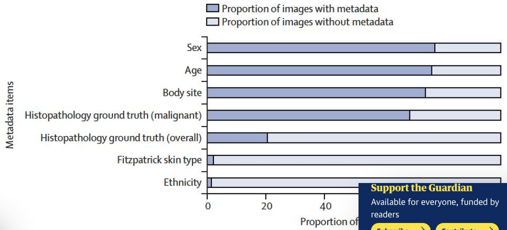


THE LANCET
Digital Health

Characteristics of publicly available skin cancer image datasets: a systematic review

David Wen, Saad M Khan, Antonio Ji Xu, Hussein Ibrahim, Luke Smith, Jose Caballero, Luis Zepeda, Carlos de Blas Perez, Alastair K Denniston, Xiaoxuan Liu*, Rubeta N Matin*

Publicly available skin image datasets are increasingly used to develop machine learning algorithms for skin cancer diagnosis. However, the total number of datasets and their respective content is currently unclear. This systematic review aimed to identify and evaluate all publicly available skin image datasets used for skin cancer diagnosis by exploring their characteristics, data access requirements, and associated image metadata. A combined MEDLINE,



WILL KNIGHT BUSINESS OCT 11, 2020 7:00 AM WIRED

AI Can Help Diagnose Some Illnesses—If Your Country Is Rich

Algorithms for detecting eye diseases are mostly trained on patients in the US, Europe, and China. This can make the tools ineffective for other racial groups and countries.

Support the Guardian
Available for everyone, funded by readers
Subscribe Contribute
Sign in
The Guardian
News website of the year

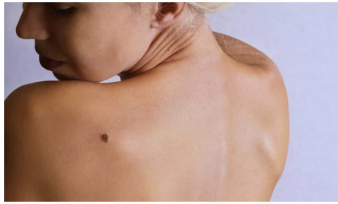
News Opinion Sport Culture Lifestyle

UK World Coronavirus Climate crisis Football Business Environment UK politics More

Skin cancer

AI skin cancer diagnoses risk being less accurate for dark skin - study

Research finds few image databases available to develop technology contain details on ethnicity or skin type

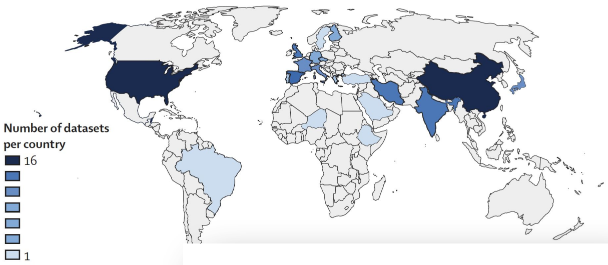


Studies suggest image recognition technology can classify skin cancers as successfully as humans. Posed by model. Photograph: ChesiireCat/Getty Images/iStockphoto

THE LANCET
Digital Health

A global review of publicly available datasets for ophthalmological imaging: barriers to access, usability, and generalisability

Saad M Khan*, Xiaoxuan Liu*, Siddharth Nath, Edward Korot, Livia Faes, Siegfried K Wagner, Pearse A Keane, Neil J Sebire, Matthew J Bur
Alastair K Denniston



AI Can Help Diagnose Some Illnesses—If Your Country Is Rich

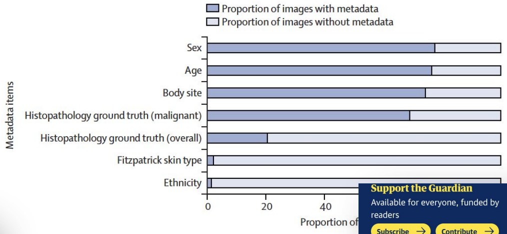
Algorithms for detecting eye diseases are mostly trained on patients in the US, Europe, and China. This can make the tools ineffective for other racial groups and countries.

THE LANCET
Digital Health

Characteristics of publicly available skin cancer image datasets: a systematic review

David Wen, Saad M Khan, Antonio Ji Xu, Hussein Ibrahim, Luke Smith, Jose Caballero, Luis Zepeda, Carlos de Blas Perez, Alastair K Denniston, Xiaoxuan Liu*, Rubeta N Matin*

Publicly available skin image datasets are increasingly used to develop machine learning algorithms for skin cancer diagnosis. However, the total number of datasets and their respective content is currently unclear. This systematic review aimed to identify and evaluate all publicly available skin image datasets used for skin cancer diagnosis by exploring their characteristics, data access requirements, and associated image metadata. A combined MEDLINE,



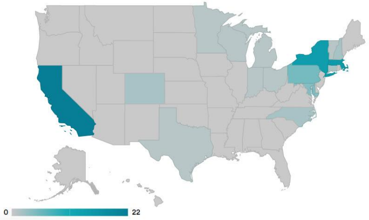
The Geographic Bias in Medical AI Tools

SHANA LYNCH September 21, 2020

Home / Blog

SHARE THIS:

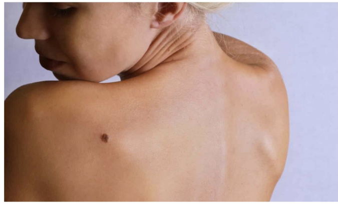
Patient data from just three states trains most AI diagnostic tools.



Skin cancer

AI skin cancer diagnoses risk being less accurate for dark skin – study

Research finds few image databases available to develop technology contain details on ethnicity or skin type

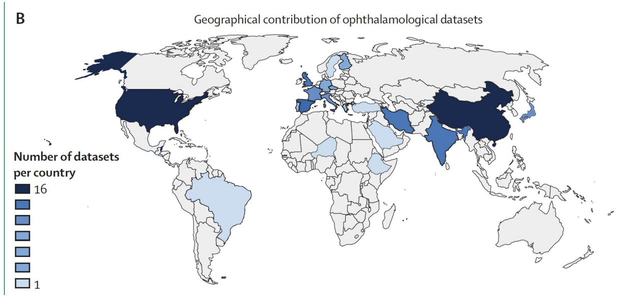


Studies suggest image recognition technology can classify skin cancers as successfully as humans. Posed by model. Photograph: ChesiireCat/Getty Images/iStockphoto

THE LANCET
Digital Health

A global review of publicly available datasets for ophthalmological imaging: barriers to access, usability, and generalisability

Saad M Khan*, Xiaoxuan Liu*, Siddharth Nath, Edward Korot, Livia Faes, Siegfried K Wagner, Pearse A Keane, Neil J Sebire, Matthew J Burt
Alastair K Denniston

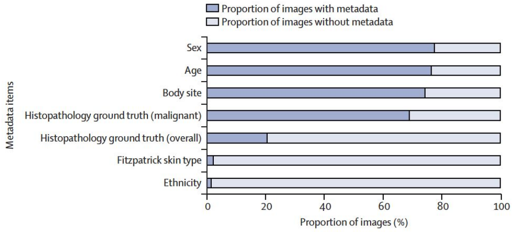


THE LANCET
Digital Health

Characteristics of publicly available skin cancer image datasets: a systematic review

David Wen, Saad M Khan, Antonio Ji Xu, Hussein Ibrahim, Luke Smith, Jose Caballero, Luis Zepeda, Carlos de Blas Perez, Alastair K Denniston, Xiaoxuan Liu*, Rubeta N Matin*

Publicly available skin image datasets are increasingly used to develop machine learning algorithms for skin cancer diagnosis. However, the total number of datasets and their respective content is currently unclear. This systematic review aimed to identify and evaluate all publicly available skin image datasets used for skin cancer diagnosis by exploring their characteristics, data access requirements, and associated image metadata. A combined MEDLINE,



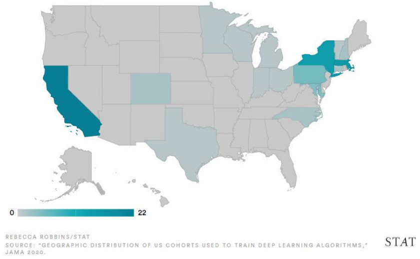
The Geographic Bias in Medical AI Tools

SHANA LYNCH September 21, 2020

Home / Blog

SHARE THIS:

Patient data from just three states trains most AI diagnostic tools.



Health data poverty: an assailable barrier to equitable digital health care

Hussein Ibrahim, Xiaoxuan Liu, Nevine Zariffa, Andrew D Morris*, Alastair K Denniston*

The inability for individuals, groups, or populations to benefit from a discovery or innovation due to insufficient data that are representative of them

ARTICLES

https://doi.org/10.1038/s41591-021-01595-0

nature medicine

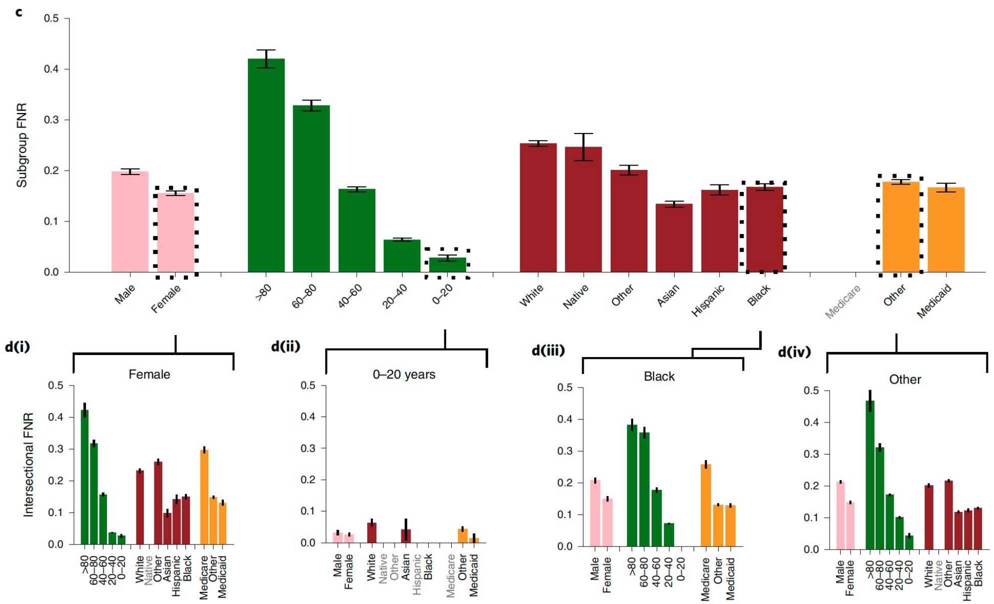
Check for updates

OPEN

Underdiagnosis bias of artificial intelligence algorithms applied to chest radiographs in under-served patient populations

Laleh Seyyed-Kalantari[1,2], Haoran Zhang[3], Matthew B. A. McDermott[3], Irene Y. Chen[3] and Marzyeh Ghassemi[2,3]

We have shown consistent underdiagnosis in three large, public datasets in the chest X-ray domain. The algorithms trained on all settings exhibit systematic underdiagnosis biases in under-served subpopulations, such as female patients, Black patients, Hispanic patients, younger patients and patients of lower socioeconomic status (with Medicaid insurance). We found that these effects persist for intersectional subgroups (for example, Black female patients)

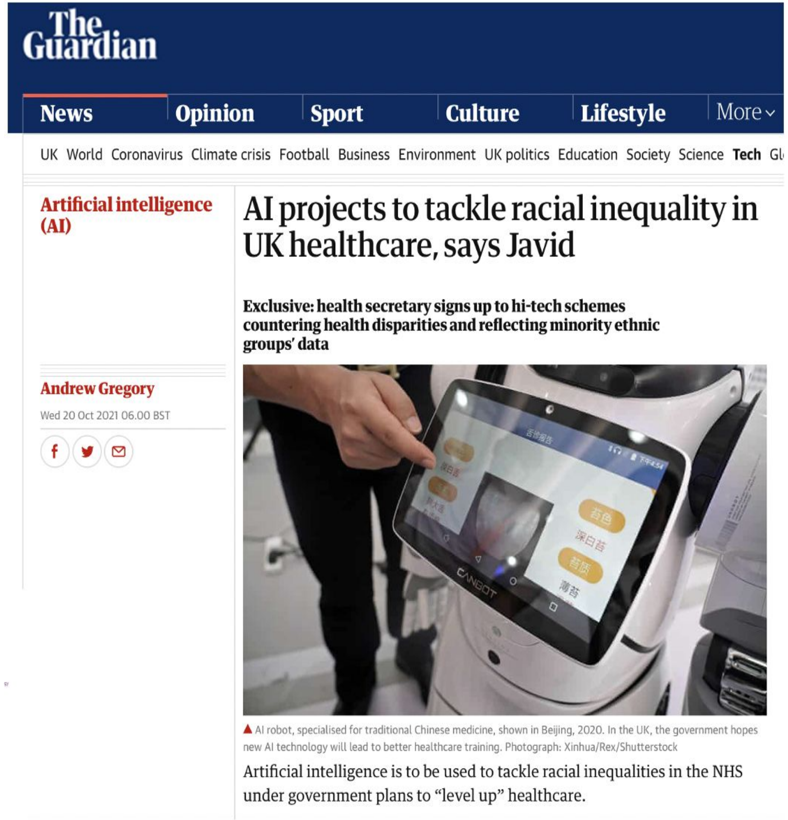


The Health Foundation

NIHR | National Institute for Health Research

1. Understanding and enabling opportunities to use AI to address health inequalities

2. Optimising datasets, and improving AI development, testing, and deployment



The Guardian

News | Opinion | Sport | Culture | Lifestyle | More

UK World Coronavirus Climate crisis Football Business Environment UK politics Education Society Science Tech Gl

Artificial intelligence (AI)

AI projects to tackle racial inequality in UK healthcare, says Javid

Exclusive: health secretary signs up to hi-tech schemes countering health disparities and reflecting minority ethnic groups' data

Andrew Gregory

Wed 20 Oct 2021 06.00 BST

▲ AI robot, specialised for traditional Chinese medicine, shown in Beijing, 2020. In the UK, the government hopes new AI technology will lead to better healthcare training. Photograph: Xinhua/Rex/Shutterstock

Artificial intelligence is to be used to tackle racial inequalities in the NHS under government plans to "level up" healthcare.


STANDING Together
Developing STANdards for data Diversity, INclusivity and Generalisability

To build AI healthcare technologies which benefit all patients, we need datasets which represent the diverse range of people they are intended to be used in.

Unfortunately, health datasets often do not adequately represent minority populations.

FDA U.S. FOOD & DRUG ADMINISTRATION

Health Canada Santé Canada

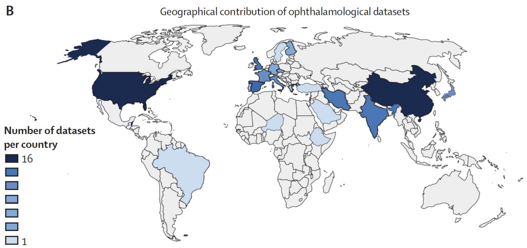
Medicines & Healthcare products Regulatory Agency

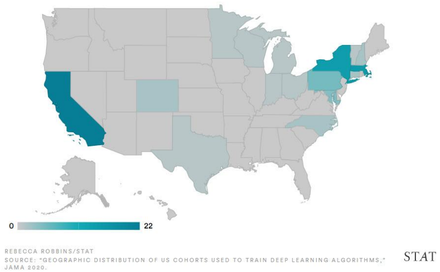
Good Machine Learning Practice for Medical Device Development: Guiding Principles

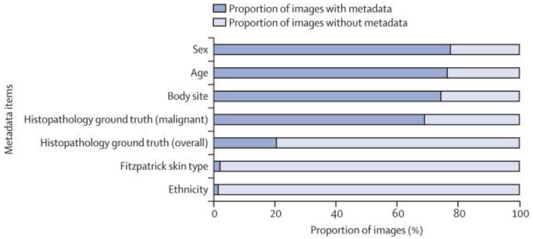
Guiding Principles

1. **Multi-Disciplinary Expertise Is Leveraged Throughout the Total Product Life Cycle:** In-depth understanding of a model's intended integration into clinical workflow, and the desired benefits and associated patient risks, can help ensure that ML-enabled medical devices are safe and effective and address clinically meaningful needs over the lifecycle of the device.
2. **Good Software Engineering and Security Practices Are Implemented:** Model design is implemented with attention the "fundamentals": good software engineering practices, data quality assurance, data management, and robust cybersecurity practices. These practices include methodical risk management and design process that can appropriately capture and communicate design, implementation, and risk management decisions and rationale, as well as ensure data authenticity and integrity.
3. **Clinical Study Participants and Data Sets Are Representative of the Intended Patient Population:** Data collection protocols should ensure that the relevant characteristics of the intended patient population (for example, in terms of age, gender, sex, race, and ethnicity), use, and measurement inputs are sufficiently represented in a sample of adequate size in the clinical study and training and test datasets, so that results can be reasonably generalized to the population of interest. This is important to manage any bias, promote appropriate and generalizable performance across the intended patient population, assess usability, and identify circumstances where the model may underperform.
4. **Training Data Sets Are Independent of Test Sets:** Training and test datasets are selected and maintained to be appropriately independent of one another. All potential sources of dependence, including patient, data acquisition, and site factors, are considered and addressed to assure independence.
5. **Selected Reference Datasets Are Based Upon Best Available Methods:** Accepted, best available methods for developing a reference dataset (that is, a reference standard) ensure that clinically relevant and well characterized data are collected and the limitations of the reference are understood. If available, accepted reference datasets in model development and testing that promote and demonstrate model robustness and generalizability across the intended patient population are used.
6. **Model Design Is Tailored to the Available Data and Reflects the Intended Use of the Device:** Model design is suited to the available data and supports the active mitigation of known risks, like overfitting, performance degradation, and security risks. The clinical benefits and risks related to the product are well understood, used to derive clinically meaningful performance goals for testing, and support that the product can safely and effectively achieve its intended use. Considerations include the impact of both global and local performance and uncertainty/variability in the device inputs, outputs, intended patient populations, and clinical use conditions.
7. **Focus Is Placed on the Performance of the Human-AI Team:** Where the model has a "human in the loop," human factors considerations and the human interpretability of the model outputs are addressed with emphasis on the performance of the Human-AI team, rather than just the performance of the model in isolation.
8. **Testing Demonstrates Device Performance During Clinically Relevant Conditions:** Statistically sound test plans are developed and executed to generate clinically relevant device performance information independently of the training data set. Considerations include the intended patient population, important subgroups, clinical environment and use by the Human-AI team, measurement inputs, and potential confounding factors.
9. **Users Are Provided Clear, Essential Information:** Users are provided ready access to clear, contextually relevant information that is appropriate for the intended audience (such as health care providers or patients) including: the product's intended use and indications for use, performance of the model for appropriate subgroups, characteristics of the data used to train and test the model, acceptable inputs, known limitations, user interface interpretation, and clinical workflow integration of the model. Users are also made aware of device modifications and updates from real-world performance monitoring, the basis for decision-making when available, and a means to communicate product concerns to the developer.
10. **Deployed Models Are Monitored for Performance and Re-training Risks Are Managed:** Deployed models have the capability to be monitored in "real world" use with a focus on maintained or improved safety and performance. Additionally, when models are periodically or continually trained after deployment, there are appropriate controls in place to manage risks of overfitting, unintended bias, or degradation of the model (for example, dataset drift) that may impact the safety and performance of the model as it is used by the Human-AI team.

Clinical Study Participants and Data Sets Are Representative of the Intended Patient Population: Data collection protocols should ensure that **the relevant characteristics of the intended patient population (for example, in terms of age, gender, sex, race, and ethnicity)**, use, and measurement inputs are sufficiently represented in a sample of adequate size in the clinical study and training and test datasets, so that results can be reasonably generalized to the population of interest. **This is important to manage any bias, promote appropriate and generalizable performance across the intended patient population**, assess usability, and identify circumstances where the model may underperform.



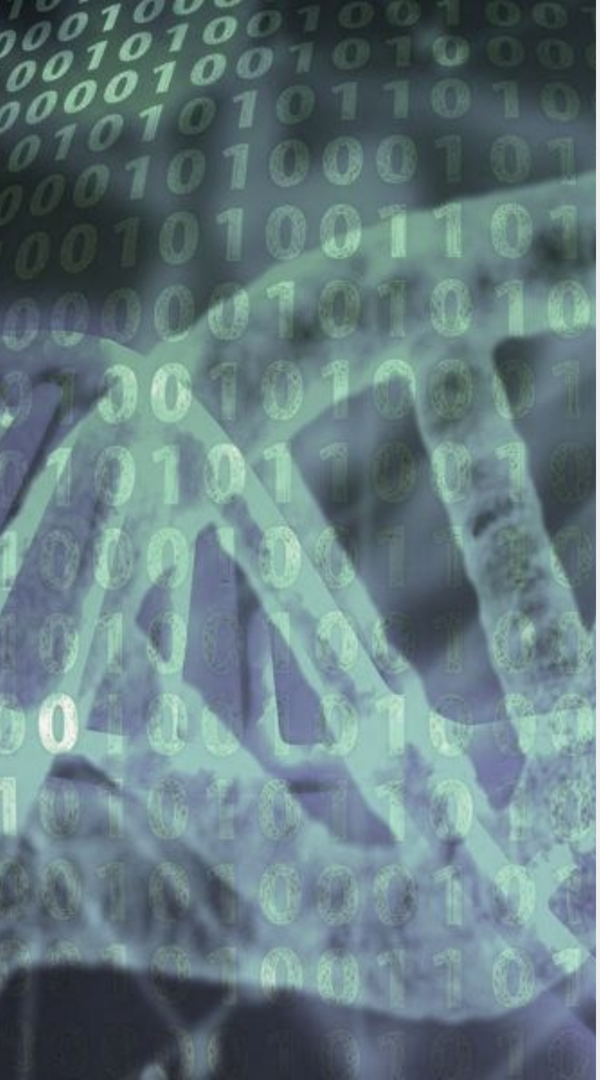




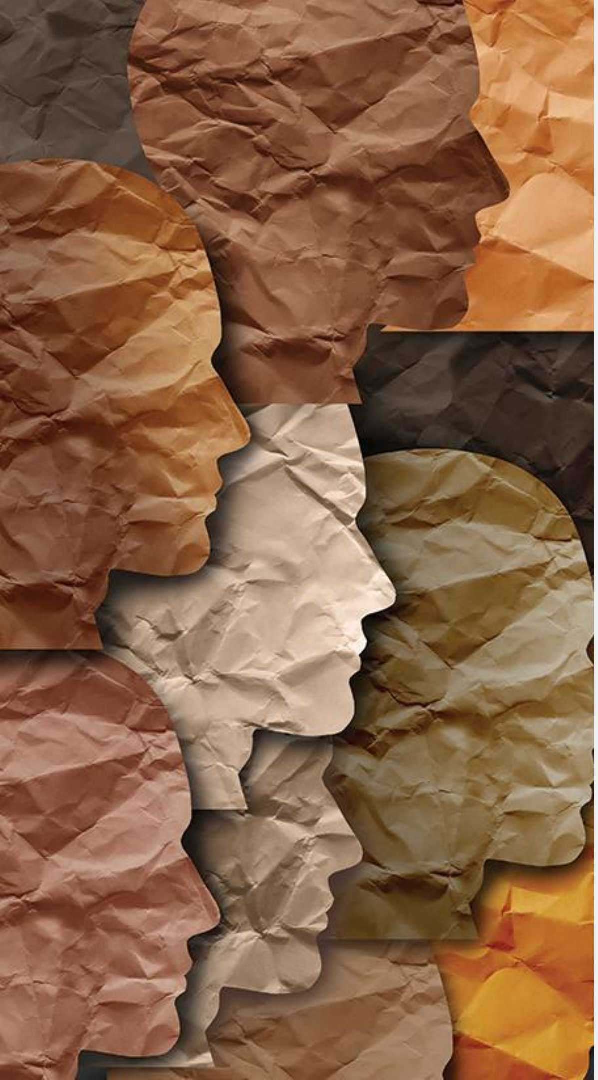
1. **What biases exist in AI health datasets?**

2. What stands in the way of reducing bias in datasets?

3. How can we ensure datasets are diverse, inclusive and promote generalisability?



1. What biases exist in AI health datasets?
2. **What stands in the way of reducing bias in datasets?**
3. How can we ensure datasets are diverse, inclusive and promote generalisability?



1. What biases exist in AI health datasets?
2. What stands in the way of reducing bias in datasets?
3. **How can we ensure datasets are diverse, inclusive and promote generalisability?**

We will develop standards on…

Composition (‘*who*’ *is represented*)

&

Transparency (‘*how*’ *they are represented*)

… of datasets in AI

STANDING Together

Developing STANdards for data Diversity, INclusivity and Generalisability

U.S. FOOD & DRUG ADMINISTRATION — Health Canada / Santé Canada — Medicines & Healthcare products Regulatory Agency

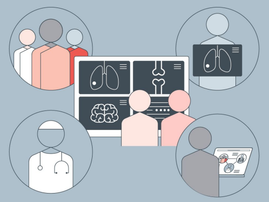
Good Machine Learning Practice for Medical Device Development: Guiding Principles

Ada Lovelace Institute

Ethics and accountability in practice

Algorithmic impact assessment: a case study in healthcare

February 2022



NIST Special Publication 1270

Towards a Standard for Identifying and Managing Bias in Artificial Intelligence

Reva Schwartz
Apostol Vassilev
Kristen Greene
Lori Perine
Andrew Burt
Patrick Hall

This publication is available free of charge from: https://doi.org/10.6028/NIST.SP.1270

NIST National Institute of Standards and Technology U.S. Department of Commerce

ISO

ICS > 35 > 35.020

ISO/IEC TR 24027:2021

Information technology — Artificial intelligence (AI) — Bias in AI systems and AI aided decision making

DOI:10.1145/3458723

Documentation to facilitate communication between dataset creators and consumers.

BY TIMNIT GEBRU, JAMIE MORGENSTERN, BRIANA VECCHIONE, JENNIFER WORTMAN VAUGHAN, HANNA WALLACH, HAL DAUMÉ III, AND KATE CRAWFORD

Datasheets for Datasets

NICE National Institute for Health and Care Excellence

Evidence standards framework for digital health technologies

CONSORT-AI

SPIRIT-AI

Diversity in Data - Ethnicity coding working group

HDRUK Health Data Research UK

DATA SCIENCE FOR HEALTH EQUITY

The medical algorithmic audit

THE LANCET
Digital Health

Xiaoxuan Liu, Ben Glocker, Melissa M McCradden, Marzyeh Ghassemi, Alastair K Denniston, Lauren Oakden-Rayner**

Scoping	Mapping	Artifact collection	Testing	Reflection	Post audit
Define audit scope	Map artificial intelligence system	**Audit checklist** • Intended use statement • Intended impact statement • FMEA clinical pathway mapping • FMEA clinical task risk analysis • FMEA risk priority number document • Datasets • Data description • Data, including explainability artifacts • Data flow diagram • The artificial intelligence model itself, if available • Model summary • Previous evaluation materials	Exploratory error analysis	Risk mitigation measures	Algorithmic audit summary report
Understand intended use	Map health-care task		Subgroup testing	Developer actions	Plan re-audit
Define intended impact	Identify personnel and resources		Adversarial testing	Clinical actions	
	Identify and prioritise risks				
	FMEA				

Adapted with permission from Raji *et al.* (2020) Closing the AI Accountability Gap: Defining an End-to-End Framework for Internal Algorithmic Auditing

Dr Joe Alderman

University of Birmingham, UK

Ms Jo Palmer

University Hospitals Birmingham NHS Foundation Trust, UK

Miss Sinduja Manohar

Health Data Research UK

Mr Cyrus Espinoza

Patient Partner

Professor Neil Sebire

Great Ormond Street Hospital for Children, London, UK

Professor Marzyeh Ghassemi

Massachusetts Institute of Technology, Massachusetts, USA

Dr Darren Treanor

University of Leeds, UK

Professor Cathie Sudlow

The University of Edinburgh & British Heart Foundation Data Science Centre, UK

Professor Melanie Calvert

University of Birmingham, Birmingham, UK

Professor Melissa McCradden

The Hospital for Sick Children (Sickkids), Toronto, Canada

Professor Elizabeth Sapey

University Hospitals Birmingham NHS Foundation Trust & University of Birmingham, UK

Dr Charlotte Summers

Cambridge University Hospitals NHS Foundation Trust & University of Cambridge, UK

Dr Stephanie Kuku

World Health Organisation & Hardian Health

Dr Rubeta Matin

Oxford University Hospitals NHS Foundation Trust, Oxford, UK

Mrs Jacqui Gath

Patient Partner

Dr Francis McKay

University of Oxford, UK

The Health Foundation | NIHR National Institute for Health Research | NHSX | HDRUK Health Data Research UK | University of Birmingham | NHS University Hospitals Birmingham NHS Foundation Trust | Medicines & Healthcare products Regulatory Agency | wellcome