

Imperial College
London

Addressing Racial and Ethnic Inequities in Data- driven Health Technologies

**INSTITUTE OF
GLOBAL HEALTH
INNOVATION**



Authors

O'Brien N, Van Dael J, Clarke J, Gardner C, O'Shaughnessy J, Darzi A, & Ghafur S.

Contributors

We would like to thank the following people who contributed to this document through interviews/ attendance at a roundtable and have agreed to be acknowledged. The people marked with * participated in a one-to-one interview and those marked with † took part in the roundtable discussion. The contributions of the individuals listed below have been invaluable to the development of the report, however, the report outlines a synthesis of range of insights, and opinions and as such, responsibilities for the findings and content of the report rests with the authors.

Lord Victor Adebowale *

Noa Dagan *†

Cerys Wyn Davies *†

Christine Ekechi *†

Allison Gardner †

Sennay Ghebreab *

Zubaida Haque *

Katherine Heller *

Marcus Gazette †

Nick Jennings *

Adrian Jonas *

Indra Joshi *

Rabia Khan *†

Maxine Mackintosh *†

Sultan Mahmud *†

Omar Moreea †

Johan Ordish *

Adepeju Oshisanya *

Neelam Patel *

Rafiah Patel †

Prisha Shah *

Harpreet Sood †

Michael Stern *

We would like to specifically thank Professor Aldo Faisal, Professor of AI & Neuroscience at the Department of Computing and the Department of Bioengineering, Imperial College London; Dr Ana Luisa Neves, Associate Director and Advanced Research Fellow at the NIHR Imperial Patient Safety Translational Research Centre, Imperial College London; and Melanie Leis, Director, Big Data Analytics Unit (BDAU), Institute of Global Health Innovation, Imperial College London for their reviews of the report.

This report was funded by the Health Foundation. The views expressed are those of the authors and not necessarily those of the Health Foundation.

Contents

5	Foreword
6	Introduction
7	Context
13	Methodology
14	The development and deployment of data-driven health technologies
15	Prioritisation in the research agenda: Innovation on an uneven playing field
18	Sources of bias in data generation and collection: AI systems are only as good as the data they are built upon
24	Algorithm development: A multidisciplinary approach can improve performance
28	Deployment, utilisation and patient and clinical outcomes: Implementation in the NHS
32	Overarching considerations
32	a. Developing appropriate governance systems, legislation and regulation
36	b. Improving the product and its impact through diversity in IT, tech and healthcare
40	c. Patient and public involvement and engagement throughout the development and deployment of data-driven health technologies
42	Addressing key knowledge and practice gaps
44	References



Foreword

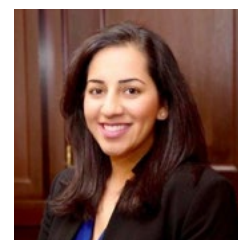
With the use of artificial intelligence (AI) and machine learning for healthcare delivery increasing, there are numerous examples of more accurate, tailored, and personalised opportunities for the management of patients across healthcare systems in the UK. However, the use of AI and machine learning is not without challenges, particularly in striving towards safe and high-quality healthcare for all. The NHS has relied on digital technologies, including data driven technologies to facilitate care throughout the pandemic and has seen significant and tangible benefits from being able to access and analyse vast amounts of health care data. The COVID-19 pandemic has highlighted that there are widespread healthcare inequities, with poorer access, treatment, and health progression for some minority ethnic groups. Several recently published reports have highlighted various aspects of these inequities. It is vital to ensure that developments in AI and machine learning do not exacerbate or create further challenges to equity in healthcare.

There are a range of challenges to enabling data-driven health technologies and AI for minority ethnic groups. Funding disparities, bias in data and datasets, under-theorised or un-justified use of ethnicity data, biased decision-making, and unfair or harmful deployment, amongst other challenges, can all result in negative outcomes for minority ethnic groups. Although AI and data-driven technology may have the potential to improve the health of minority ethnic groups, it is clear that further understanding is required to address these risks to ensure they can be mitigated across the innovation pipeline.

Despite a growing number of cases demonstrating racial and ethnic bias in AI and data-driven technologies, the research and evidence base on this topic remains nascent. The aim of this review is to help synthesize the current evidence in academia and policy and help to identify high-priority areas of inquiry that need to be addressed, both in the UK and globally.

The report highlights core considerations to improve equity through the development and deployment of AI and machine learning technology, offering key research and practice priorities to be considered at each step. However, we acknowledge that there are a significant range of nuances and complexities that we may not fully address in this report and that there are many questions that need further study. We hope that this paper acts as a catalyst and framework for a much-needed national conversation on how the UK's health data infrastructure and advances in data-driven technologies, including AI, can be best used to improve healthcare for all.

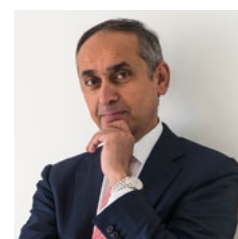
Dr Saira Ghafur - Lead for Digital Health at the Institute of Global Health Innovation, Imperial College London and Honorary Consultant at the Imperial College Healthcare NHS Trust



Lord James O'Shaughnessy - Visiting Professor and Chair of the Data, Health and Wealth Programme at the Institute of Global Health Innovation



Lord Ara Darzi - Co-Director of the Institute of Global Health Innovation and Chair of the Accelerated Access Collaborative



Introduction

Inequity in healthcare has been well documented in the UK. The COVID-19 pandemic has starkly highlighted the disproportionate impact that the disease is having on some population groups, with the risk of dying among those diagnosed with COVID-19 significantly higher in Black and Asian groups than in White ethnic groups¹. The reasons for health disparities are complex and highlight existing issues around unequal access to healthcare, worse experiences of healthcare, and wider inequities in society. Despite strong legal and constitutional backing, and national policy guidance and frameworks, limited progress has been made on this critical agenda².

Data-driven technologies offer great potential for the delivery of health and care in general and may have a specific part to play in helping to reduce disparities in health. The UK holds some of the

best health data sets globally and has the ability to bring together comprehensive and longitudinal data for over 65 million people³. It is imperative that the benefits of this data are maximised for public benefit. The Government has invested capital and driven initiatives to improve these data sets³. However, at the same time, there has been little exploration of whether and how AI and data-driven technologies could be utilised to help improve the health of minority ethnic groups and as such, a better evidence base is required for this. The World Health Organization published ethical principles for AI in health in June 2021⁴, highlighting the need to ensure inclusiveness and equity. As these digital technologies become integrated into almost every aspect of healthcare in the UK, it is crucial that they do not create or perpetuate any existing inequities for minority ethnic groups.

Context

Race and ethnicity: terminology and use in health policy and research

Ethnicity has been described as “a form of collective identity that draws on notions of shared ancestry, cultural commonality, geographical origins and shared biological features”². According to the Office of National Statistics (ONS) Annual Population Survey 2018, 85.6% of the total population of the UK was White. People from Asian or Asian British ethnic groups made up the second largest percentage of the population (7.5%), followed by Black, African, Caribbean or Black British ethnic groups (3.4%), Mixed/Multiple ethnic groups (1.7%) and Other ethnic groups (incl. Arab ethnic groups) (1.8%)⁵. England represents the most diverse of the four nations, with minority groups accounting for 16.1% of the population, while Northern Ireland is the least diverse with minority groups representing 2.2% of the population.

Despite significant differences between the terms race and ethnicity, they are often used interchangeably and synonymously when it comes to health and social care research and policy. A review of the definition and operationalisation of race or ethnicity in biomedical research found widespread under-theorisation and unspecified use of the two concepts and concluded that such inconsistency compromises an adequate understanding of the causes for differential health outcomes⁶. In medicine and biomedical research, race has historically (and often problematically) been used to refer to groups defined by shared ancestry and/or physical attributes and understood to play a genetical role in driving health outcomes⁷. More recent research challenges the extent to which observed differences in outcomes relate to genetic or biological factors,

providing evidence that a range of factors such as unequal treatment or access to health, socio-economic conditions, and environmental factors play a significant role (e.g. ^{8–10}).

Minority ethnic groups are also often inappropriately grouped together when there are significant differences between them in terms of their health outcomes¹¹. This is reflected in the common use of the term ‘BAME’ in the UK, which stands for Black, Asian and Minority Ethnic. Although this term was born out of the need to discuss the shared racialisation of minority ethnic groups relative to the majority White population¹², it obscures significant heterogeneity in lived experiences that exists both within and between minority ethnic groups and negates the variation in the unequal power structures in which different groups are situated. The term also excludes White ethnic minorities such as Gypsy, Roma and Traveller of Irish Heritage groups, which are some of the most marginalised and disadvantaged communities^{11,12}. In the recent years, the term has been widely criticised, including in the 2021 Commission on Race and Ethnic Disparities report¹³. This has important wider lessons in terms of thinking about the classification or labelling of diverse ethnic minorities under single categories¹⁴, and their representation in health data sets used to train AI – as discussed in the findings of the report.

When considering whether and how to include ethnicity in a given model, it is further critical to consider differences based on interactions with wider social and demographic factors (e.g. age, class, gender). The increased recognition that social categories intersect poses both a significant opportunity and challenge for population health and data science research. Historically, such

approaches have tended to use either unitary or multiple approaches to factor social positioning¹⁵. An intersectional approach acknowledges that using single or additive categories to model socio-demographic variation risks obscuring the unique variance that stems from different interactions of such categories. As health outcomes are themselves a result of the interaction between social and individual circumstances, biological factors, and structural disparities in health policy and delivery¹⁶, the potential benefits of considering these intersections in data science research are increasingly recognised for their enormous potential to support more tailored, patient-centric models of healthcare and policy^{17–20}. We hope that some of the insights in this report will aid this effort. As such, when diversity is discussed in this report, we do not speak exclusively of ethnicity, but across a wide range of social categories such as age, class, gender, ability status, and sexual orientation.

Health disparities, the social determinants of health, and inequity in healthcare in the UK

Ethnicity-based disparities exist widely within health and healthcare in the UK, as seen most recently in the disproportionate toll of COVID-19 among minority ethnic individuals and groups^{21,22}. Black women are four times more likely to die of COVID-19 than white women, with drivers of inequity extending beyond disparity in clinical predisposition²³. The risk of psychosis in Black Caribbean groups is estimated to be significantly higher (almost seven times) than in the White population²⁴. In children, there are higher levels of obesity among those aged 10-11 in most minority ethnic groups as compared to a White population of the same age²⁵. Gypsies and Travellers experience poorer health than the average population, for example bronchitis (40% vs 10%), anxiety (38% vs 13%), with Gypsy and Traveller women twice as likely to be anxious as men, and asthma (65% vs 40%)²⁶. They also have an average life expectancy 10 years lower than the UK national average²⁷.

In many cases, the causes for these disparities include non-biological factors, and (at least partly) result from unequal healthcare access and treatment. The 2014 Adult Psychiatric Morbidity Survey showed that Black men had the highest mean score for severity of mental health symptoms but were the least likely to receive treatment for their illness²⁸.

In 2019-20 the rate of detentions under the Mental Health Act in England among Black people was four times higher than for White people²⁹, prompting the development of reforms to the Mental Health Act in 2021³⁰. Surveys of maternity care experiences in the NHS indicate that minority ethnic women, and in particular Black women, more often feel that they are not treated with dignity and respect, with common reports of staff dismissing their pain or concerns^{31–32}. These findings have been echoed in the recent report by the NHS Race and Health Observatory³³.

Other research has indicated minority ethnic groups may be less likely to receive appropriate preventative care (e.g., screening, vaccinations); are more likely to have undiagnosed comorbidities (e.g., heart disease, chronic obstructive pulmonary disease)⁸; and are more at risk of patient safety incidents³⁴. Similarly, inequities can exist in the functionality of medical devices. For example, a 2020 study on pulse oximetry showed the technology's racial bias, with hypoxaemia (low blood oxygen) identified less readily in Black patients than White patients risking inaccurate measurements³⁵. More generally, a 2017 survey found that respondents who identified as Black reported lower levels of satisfaction with the NHS (44%) than respondents who identified as White (58%)²¹.

At national level, more is being done to tackle some of these inequities in healthcare. In February 2022, the Government launched landmark reviews to tackle health disparities, including a review into the health impact of potential bias in medical devices³⁶. NHS England and NHS Improvement announced the launch of the Core20PLUS5 programme in December 2021. This programme will focus on the most deprived 20% of the population and will focus on the following five areas: maternity care, mental illness, respiratory diseases, cancer, and cardiovascular disease³⁷. The programme also provides the opportunity for local areas to focus on their population's needs, through the role of integrated care systems (ICSs)³⁸.

In addition to unequal access and treatment, factors explaining disparities include wider inequities in society, often referred to as 'social determinants'. Public Health England describes the social determinants of health as "the broad social and economic circumstances that together influence health throughout the life course" and similarly recognises the role of these circumstances as contributors to health outcomes³⁹. For example, access to education reduces the likelihood of suffering from long-term diseases and mental health

issues such as depression and anxiety⁴⁰. The 2010 Fair Society, Healthy Lives: The Marmot Review found that the lower one's social and economic status, itself shaped by education, occupation, income, gender, ethnicity and race, the poorer one's health is likely to be⁴¹. Although this report initially fell short of highlighting ethnicity as a factor relating to inequity and disadvantage⁴², Marmot has since been vocal in advancing the association between ethnicity and inequity, both in Health Equity in England: The Marmot Review 10 Years On⁴⁴ and Build Back Fairer: The COVID-19 Marmot Review⁴⁵, noting that structural racism is a fundamental driver for social determinants, and should be considered in any efforts to tackle social determinants behind disparities in health.

Despite clear evidence of the health inequities affecting minority ethnic groups, the report of the Commission on Race and Ethnic Disparities¹³ published in March 2021 (the CRED report), rejected “the common view that ethnic minorities have universally worse health outcomes compared with White people”, arguing that inequities in healthcare are driven by “deprivation, geography and differential exposure to key risk factors” without acknowledging the role of institutional racism in driving some of these inequities. The conclusions of the report have faced criticism from a range of groups and organisations across the UK⁴⁶. Among the medical community, the report has been criticised for its failure to adequately consider the evidence of racial disparities in health, nor the relationship between structural racism and health.

The Royal College of Psychiatrists responded by saying:

“It is deeply disappointing that the report failed to take notice of the compelling evidence that racial disparities in health, and particularly in mental health, are driven in large part by social factors which are structurally determined.”⁴⁷

The British Medical Association questioned the value of continuing to seek evidence of ethnic disparities in health and across the UK, which is already widely documented, encouraging a greater focus on supporting and monitoring action to address inequity:

“It is essential that poverty is recognised as a key driver of racial and ethnic inequalities in the UK. The fact that Black and Asian populations experience poverty at higher rates, and also have worse health, is no coincidence. There is a wealth of evidence to support this, and alleviating poverty must be an objective of any strategy to reduce race and ethnic disparities and health inequalities.”^{48(p.1)}

With regards to data-driven technologies, although the report acknowledges some technical aspects of AI bias, such as those resulting from missing data on ethnicity, it has been criticised for failing to acknowledge the wider social and historical context in which data is created, managed, and applied in practice. Drawing from their recent report on the ‘data divide’⁴⁹, The Ada Lovelace Institute noted:

“Unless we acknowledge that structural racism is a fundamental driver and cause of the data divide – which affects who can benefit equitably and fairly from AI and algorithmic systems – we cannot evidence and understand inequalities caused by data.”⁵⁰

Notwithstanding the limitations identified above, the CRED report usefully highlighted the breadth of factors beyond ethnicity that contribute to health outcomes including age, gender, disability, poverty, housing, rurality¹³. These can directly drive outcomes and are also interlinked with ethnicity in complex ways, and all need to be considered within any approach to reducing inequity in health and care.

The COVID-19 pandemic

The COVID-19 pandemic has exacerbated and emphasised the widespread disparities in healthcare within the minority ethnic groups in the UK. Despite making up 13% of the population, data up to January 2021 in England and Wales shows that 22% of patients requiring intensive care support were of Black or Asian background⁵¹. The latest ONS release

shows that the rate of deaths involving COVID-19 between December 2020 and June 2021 was 5 times higher for Bangladeshi men than White men, and 4.5 times higher for Bangladeshi women compared to White women⁵². Similarly, all Census-defined minority ethnic groups other than Chinese had a higher rate of deaths involving COVID-19 than the White ethnic population for both men and women. The initial COVID-19 data response was slow to detect these issues, with estimates based on simple aggregates of the overall population (e.g. based on age and comorbidity)⁵³, thereby masking disproportionately poor outcomes for minority subsets in the data. Other issues rest with poor data quality and comprehensiveness more generally. A Nuffield Trust report published in June 2021 found that mis-coding in hospital data challenges the reliability of assessing infection rates, hospitalisations and deaths by ethnicity, and the ability to fully understand the role of ethnic inequities through the pandemic and in healthcare more widely⁵⁴.

The COVID-19 vaccine roll-out further highlighted incomplete data generation and collection, particularly for minority ethnic populations. Prior to the pandemic, anecdotal evidence suggests ethnicity was only thought to be recorded in 60-70%

of GP records⁵⁵. Since the COVID-19 vaccine drive started, there was significant concern with the lack of data collection on patient ethnicity at the point of vaccination, thereby obstructing efforts to ensure good uptake in different minority ethnic groups, and perpetuating misconceptions about poor uptake⁵⁵. NHS England now collects routine data on ethnicity at the point of vaccination, which has helped identify communities with poorer vaccine uptake.

Levelling Up

The UK Government has affirmed its commitment to the “levelling up” agenda and through the recently published a white paper has outlined a 12 key missions to help reduce regional inequalities by 2030⁵⁶. These include health, housing, education and digital connectivity and emphasise the government’s commitment to address disparities in key services and improve outcomes and metrics such as healthy life expectancy which are critical to addressing the underlying issues discussed above. Although welcome, there is criticism that tackling health inequalities will require billions of pounds more in funding than is currently available to meet these ambitious targets⁵⁷.

What are data-driven technologies in health and care?

In the context of healthcare, the term “data-driven technology” describes technologies that work by collecting, using and analysing data, including patient health and care data, to support the care of individuals and patients, the functioning and improvement of health services and public health, and the development of medical research and innovation⁵⁸. Artificial intelligence (AI) is one type of data-driven technology that is increasingly being used in healthcare. Unlike traditional use of information communication technology (ICT) where a computer performs a task based on the decisions of its user, AI performs tasks by learning rules from the data it has access to. Used together with big data, AI in healthcare can produce powerful models to automate prevention, detection, diagnosis and treatment, enabling more accurate, tailored and effective management of patients⁵⁹⁻⁶¹. Machine learning describes the application of AI, which gives systems the

ability to automatically learn and improve from experience without being explicitly programmed.

The UK Government published a National AI Strategy in September 2021 as part of efforts to ensure the UK keeps pace with innovation and can unlock the power of AI and data-driven technology⁶². Within the document, the Government re-affirmed its commitment to taking a sector-led approach to regulating AI. An early manifestation of that goal is to produce a Draft National Strategy for AI-driven technologies in Health and Social Care⁶³ and a further White Paper on governing and regulating AI is also expected in 2022.

In healthcare, a broad range of techniques are used to create data-driven technologies and AI that can carry out or assist in medical tasks traditionally done by professional healthcare practitioners⁶⁴. **Figure 1** outlines the high-level areas in which such technologies can be applied in the healthcare context.

The application of AI in clinical settings is in its infancy, and where applied is often limited to medical image analysis, whether in radiology, histopathology, dermatology or ophthalmology^{60,65–67}. In each case, structured image data are readily available and are relatively consistent between patients. Similarly, many of the models so far developed aim to classify within a small number of diagnostic classes, for example the diagnosis of retinal pathology⁶⁶. Clinical decision support tools are already being used in healthcare provision, assisting clinicians in their decision-making and planning. Similarly, automated efficiency tools are increasingly used to support back-office tasks in healthcare such as scheduling, including scheduling staff time, predicting patient no-shows, and optimising slots within clinics to reduce patients' waiting times⁶⁰.

As with any technology, potential risks and harms need to be addressed, particularly in the application of data-driven technology in healthcare. Common issues related to AI and machine learning include values and biases reflected in AI systems, the explainability of AI systems, the potential for bias across the development and deployment process, transferability, concerns regarding data, legal, moral and professional responsibility, and critically, patient/end user experience and outcomes⁶⁸. Despite the growing number of ethical guidelines and standards for AI and machine learning, including the development of QUADAS-AI, a specific framework to aid researchers and policymakers in evaluating the risk of bias and applicability when conducting reviews that evaluate AI diagnostic test accuracy⁶⁹, there is a notable absence of insights specifically related to the risks and challenges with regards to ethnicity-based inequities.

This is critical given the wide acknowledgement of racial bias in AI, including in academic literature and mainstream media. For example, facial-recognition systems from companies like IBM have been shown to be less accurate in recognising darker-skinned individuals: 11-19% less accurate in recognising images of Black men and women, increasing to 34% less accurate when shown images of Black women only⁷⁰. Another well-known example is skin cancer melanoma, where AI algorithms have been predominantly developed based on data from White patients, resulting in fewer data points to develop a full picture of melanoma in ethnic minorities⁷¹.

The purpose of this report is to identify key areas of inquiry in addressing risks and challenges with applying data-driven technologies to help improve the health of minority ethnic groups in the UK. The report highlights core considerations to improve equity through the development and deployment of AI and machine learning technology, offering key research and practice priorities to be considered at each step.

In this report we have reviewed existing literature and we have augmented this with a range of examples and case studies. Given the paucity of UK research, and subsequent examples in this area, we present a range of international examples and case studies, as well as examples from both health and social care. We have had the benefit of the involvement of experts from varying academic disciplines and professional practices and provide a series of high-level recommendations for future research to address key risks and challenges in developing and deploying data-driven technologies in health and care for minority ethnic groups.

“

Diagnostics	<ul style="list-style-type: none"> • <i>Image recognition</i> • <i>System checkers and decision support</i> • <i>Risk stratification</i>
Knowledge Generation	<ul style="list-style-type: none"> • <i>Drug discovery</i> • <i>Pattern recognition</i> • <i>Greater knowledge of rare diseases</i> • <i>Greater understanding of causality</i>
Public Health	<ul style="list-style-type: none"> • <i>Digital epidemiology</i> • <i>National screening programmes</i>
System Efficiency	<ul style="list-style-type: none"> • <i>Optimisation of care pathways</i> • <i>Prediction of Do Not Attends</i> • <i>Identification of staffing requirements</i>
P4 Medicine	<ul style="list-style-type: none"> • <i>Prediction of deterioration</i> • <i>Personalised treatments</i> • <i>Preventative advice</i>

”

Fig. 1: Areas of care in which such automated tasks could make a difference described in NHS X, Artificial Intelligence: How to get it right (2019) ⁶⁰

Methodology

The aim of this report is to map existing evidence on current risks and challenges with using AI and data-driven technologies for minority ethnic groups in the UK and to identify high-priority questions that need to be addressed in future research and practice. The findings in this report are based on a review and synthesis of existing academic literature, policy evidence, and expert involvement. Involving experts, both national and international and across the spectrum from academia to patient-facing organisations, has allowed us to better understand knowledge gaps in this nascent field of research and identify future priorities.

A series of informant interviews (n = 20) were conducted on a one-to-one basis with experts identified as having a strong expertise in this topic (i.e., race equality, health policy and machine learning experts). Interviewees were representatives from industry (technology and life sciences), NHS policy and practice, legal and regulatory bodies, academic institutions, patient groups and charities. A combination of convenience and snowball sampling was used to identify the participants. The spread of expertise ensured that we covered the challenges and opportunities when exploring this topic at a research, practice, and policy-level. A semi-structured topic guide was developed based on a review of academic and policy literature and piloted with a digital health expert (not included in the final interviews). Interview findings formed the basis of a

roundtable discussion (n = 14) where potential gaps and contested issues were discussed to try and reach consensus on research priorities where possible for the final framework and findings. This event included experts who were not interviewed prior.

Interviews and the roundtable discussion were transcribed verbatim and analysed using the framework analysis method⁷². The analysis was conducted in an iterative manner using a combination of theory-deductive and data-inductive analysis. Expert insights were mapped onto each stage outlined in [Chen et. al's ethical pipeline for health care model development](#)⁷³ (**Figure 2**) and used to complement to and contextualise existing knowledge. The following sections present a synthesis of the literature review and interview and roundtable analysis, offering key research priorities at each stage.

Limitations

This report is intended to be a scoping review to understand the gaps in knowledge that we have when understanding this topic specifically in the NHS context. Although this includes existing evidence, this is not intended to be an exhaustive review of the literature or existing research. Due to the COVID-19 pandemic the expert roundtable was hosted online and to ensure maximum utility of the remote session we limited the number of participants.

The development and deployment of data-driven health technologies

There are several areas that need to be explored and addressed as the health, technology and academic sectors work towards ensuring AI and machine learning offer health and healthcare benefits across society. Inherent bias exists within AI and machine learning technologies themselves, due to such bias being prevalent within UK society and the digital technology industry more widely. As such, understanding inequity and bias at each stage of development and deployment is an essential starting point ahead of developing future research and priority setting.

A seminal paper in this field by Chen et al. (2021) outlines five steps in what they call an “ethical pipeline for healthcare model development” (Figure 2)⁷³. The paper offers a clear and concise outline of where challenges may arise through the process of development and deployment, and as such we use an adapted version of this pipeline as a framework to guide the structuring of our research findings and review of literature. Based on findings from the expert interviews (most of whom did not explicitly discuss “outcome definition”) we map our findings to the main stages represented in Chen’s pipeline (excluding the ‘outcome definition’, e.g. mapping the end-goal, stage which has been incorporated under the broad umbrella of ‘algorithm development’).

“

Problem Selection

Disparities in funding and problem selection priorities are an ethical violation of principles of justice.



Data Collection

Focus on convenience samples can exacerbate existing disparities in marginalised and under-served populations, violating do-no-harm principles.



Outcome Definition

Biased clinical knowledge, implicit power differentials, and social disparities of the healthcare system encode bias in outcomes that violate justice principles.



Algorithm Development

Default practices, like evaluating performance on large populations, violate beneficence and justice principles when algorithms do not work for sub-populations.



Post-Deployment Considerations

Targeted, spot-check audits and lack of model documentation ignore systematic shifts in populations and risks patient safety, furthering risk to under-served groups.

”

Fig. 2: Chen et al’s five steps in the ethical pipeline for health care model development. Republished with permission of Annual Review of Biomedical Data Science, from Ethical Machine Learning in healthcare, Chen et al., 2021;4; permission conveyed through Copyright Clearance Centre, Inc.⁷³

Prioritisation in the research agenda: Innovation on an uneven playing field

Summary of our research findings

- A lack of diversity at a strategic level (e.g., in the NHS and in government), has implications for the problems and needs of minority ethnic populations being met in the development of AI and data-driven technologies, as minority ethnic voices are not adequately heard in the design and delivery of health services.
- Given the risk of commercial opportunity prioritisation as opposed to societal need driving prioritisation in research, there is a role for government and public funding bodies, such as NIHR, Wellcome Trust, and MRC, to fund needs-based innovation focused on improving health for minority ethnic groups.
- Some progress has been made in recent UK grant funding calling for research to explore the disproportionate health challenges and the health needs of minority ethnic groups in the UK, but more government and public funding bodies must fund needs-based innovation.
- In addressing disparities in research prioritisation and funding, qualitative research and patient and public involvement and engagement initiatives, such as patient panels and citizen juries, may help challenge or complement priorities set at a top-level.

Inequities in AI development and deployment start with which problems are prioritised in the first place, as reflected in research funding and policy priorities. A recently published paper outlining racism in the UK health research landscape provides an overview of the challenges across academia, funding bodies, and peer reviewed journals⁷⁴ in the UK; according to a 2019 report by AdvanceHE, there were 19,285 professors in UK universities in total, including 12,795 White men, 4,560 White women, 90 Black men, and 35 Black women⁷⁵. Similarly, there is also evidence of limited diversity at a strategic level within the NHS⁷⁶. The 2020 NHS Workforce Race Equality Standard

reported that 10% of board members from NHS Trusts were from ethnic minority backgrounds. While this figure is an increase on previous years, it still falls short of being representative of the 21% of NHS staff from identifying as being from a minority ethnic group⁷⁷. One expert interviewee noted:

“There is some real research needed and some action around the kind of the representation at the very top of the NHS. It is sorely lacking. Quite a lot of people who are from Black, Asian and Minority Ethnic backgrounds, who aspire to senior positions, will leave the NHS after they get to a certain point because there is a perception that you are bouncing your head against the wall for so long and then think, ‘Well, why don’t I go to the private sector and it pays me a lot more?’. The problem with that is that actually it’s those very voices that we need to kind of change the way we deliver and design services.”

- NHS healthcare leader

Beyond funding disparities in the NHS, the experts expressed concern with a lack of diversity within research teams, and more generally in terms of the types of research institutions that receive funding. This is not a problem unique to the UK, but as a result, essential minority ethnic voices may not be heard in discussions of what research must be prioritised within academic departments, or what subsequent innovations can be developed from existing research. The experts we interviewed also noted a lack of diversity, particularly in terms of ethnicity and class, at a strategic level (e.g., in the NHS and in government), within research teams, and more generally in terms of the types of research institutions that receive funding.

Considering the impacts of funding disparities in healthcare

Chen et al.⁷³ focus on the commonly used example to highlight disparities in research funding demonstrated by the funding allocated to sickle cell disease (predominantly affecting Black children) and cystic fibrosis (predominantly affecting White children) research^{78,79}, where cystic fibrosis receives 3.5 times more National Institutes of Health funding⁸⁰ than SCD, despite significantly more children being affected by SCD. Such funding disparities ultimately challenge the treatment of conditions. However, potential solutions have also been noted: the development of a research community determined to ensure health research avoids reinforcing and exacerbating inequities⁷⁴. Working towards this must in part involve improved representation of minority ethnic groups in research commissioning leadership roles, key committees and panels, and in patient and public communities involved in research.

We found that there were concerns among experts that the prioritisation of healthcare challenges in the UK is based on commercial opportunity as opposed to societal need. Some industry experts we interviewed highlighted the greater need for government and public funding bodies such as the National Institute for Health Research (NIHR), Wellcome Trust, and Medical and Research Council (MRC) to fund needs-based innovation focused on improving health for minority ethnic groups.

“There is a good reason why governments drive research programmes, because I guess if you leave it to the private sector, they will only focus on what’s profitable for them. So, at the end of the day, research companies are a business, they will go for diseases they feel they can be successful in, regardless of the population. But the role that institutions like the NIHR play is very important, because they fund research regardless of whether

there is a commercial revenue stream down the line or not. It’s really about the population and how are we serving them.”
- AI and healthcare expert

Disparities in light of the COVID-19 pandemic

The COVID-19 pandemic has shed light on the importance of investing in research projects that reflect health healthcare priorities in minority ethnic groups. A promising start was made by NIHR and UK Research and Innovation (UKRI) who launched funding calls worth £4.3 million for further research to understand the disproportionate death rate from COVID-19 in minority ethnic groups since evidence of these disparities emerged^{81–84}. A paper published in April 2021, reporting research findings from the recipients of a portion of the funding at the University of Leicester, found that the risk of testing positive for COVID-19, developing severe disease, and mortality from COVID-19 in South Asian and black communities in the UK could all be substantially reduced with population level policies targeting deprivation⁸⁵. Similarly, the Health Foundation and NHS AI Lab have launched a much-needed funding call for research into data-driven technologies and AI that aims to meet the health needs of minority ethnic groups in the UK. They have recently awarded £1.4m to four projects to better understand these issues⁸⁶.

The importance of bottom-up initiatives

In terms of addressing disparities in prioritisation in the research agenda and funding, experts across disciplines commented on the importance of bottom-up approaches, for example through qualitative research or patient and public involvement initiatives, to help challenge or complement priorities set at higher level. Two experts (AI and health policy) we interviewed provided the example of involving the target population in project development and grant reviews, such as through patient panels and citizens’ juries, to help estimate whether proposed solutions are in fact beneficial to clinicians (e.g., in terms of providing better care to minority patient populations) or patients (e.g., in terms of using or receiving data-driven health solutions). Yet several

of the AI experts we interviewed were hesitant about such solutions, raising concerns that they may not sufficiently represent the breadth and scale of the large data sets used in AI. This concern about homogenous representation is supported by literature that suggests that while involvement of patients and members of the public in the planning, implementation analysis and evaluation of academic research has increased in recent years, minority ethnic populations have been underrepresented⁷⁵. Further, involvement is often restricted to the planning of research, with contributions from patients rarely occurring during the analysis or dissemination phases of research. Therefore, crucial opportunities for minority ethnic patients and members of the public to guide the evolution of innovation are often missed.

Future considerations

As minority ethnic groups are less likely to receive appropriate preventative care, are more likely to have undiagnosed comorbidities⁸, and are more at risk of patient safety incidents³⁴, there are several high-opportunity research areas that could be impactful:

- **Identifying high-opportunity areas for data-driven diagnostic or therapeutic solutions (e.g., using existing cardiovascular data sets)**
- **Determining high-opportunity areas for data-driven operational applications and service use in the NHS (e.g., attendances to emergency departments)**

Sources of bias in data generation and collection: AI systems are only as good as the data they are built upon

Summary of our research findings

- There are a variety of ways that a lack, or imbalance in the collection, of diverse data occurs within research and healthcare data sets in the UK (e.g., the Hospital Episode Statistics (HES) and Clinical Research Practice Datalink (CRPD) databases)
- Three common types of data generation and collection problems may result in bias were highlighted by experts we interviewed: use of data that is itself unrepresentative or developed based on representative data but subsequently applied to an unrepresentative population; use of data that includes socio-historical bias in how it was entered and collected; and use of data that does not account for or misrepresents social categories and determinants of the intended outcome
- Efforts should be made to ensure the data on which algorithms are based is representative of the populations that they will be deployed in and with sufficient breadth and depth to capture the multitude of clinically important associations between ethnicity, demographic, social and clinical features that may exist
- The appropriateness and implications of including race or ethnicity as features in AI models must be critically appraised

As noted previously, the use of data-driven technology and AI to enable healthcare delivery has expanded in recent years, backed by the availability of large quantities of data and advancements in machine learning models, as well as increased funding for AI development⁸⁷. This has been evidenced in the creation of NHSX and the NHSX AI lab in 2019 to ensure the NHS and social care system can benefit from the advances in AI⁸⁸. NHSX has now integrated with the Transformation Directorate at NHS England as of February 2022, with the existing work programmes expected to continue. In October 2021, the UK Chancellor further announced a £6 billion package of funding for the NHS, including £2.1 billion over the SR21 period for “innovative use of digital technology so hospitals and other care organisations are as connected and efficient as possible, freeing up valuable NHS staff time and ensuring the best care for patients wherever they are.”⁸⁹ (p.94).

In the NHS, where the combination of routine data collection, longitudinal data collection, a unique patient identifier and universal coverage has resulted in large-scale data being collected and stored across the health system³. However, AI systems are only able to make decisions based on the data with which

they are trained, and that data may be lacking in critical information or imbalanced in the proportion of minority ethnic data points available. A recent study found discordance between self-reported ethnicity and NHS’ routine hospital data, the Hospital Episode Statistics (HES) database, for 20-30% of minority ethnic patients (compared to only 2.2% for White British patients)⁹⁰. Similarly, a study that compared ethnicity data coding in HES to the Clinical Research Practice Datalink (CRPD) found discordance in ethnicity coding between the datasets for over 50% of ethnic minority patients, with a significant proportion of missing data (in particular in the HES database)⁹¹.

The AI experts (academic; industry) we interviewed commented on three common types of data generation and collection problems that may result in bias. First, use of data that is either itself unrepresentative or developed based on representative data but subsequently applied to an unrepresentative implementation population (e.g., the First Derm example⁹² outlined in **Table 1**). Second, use of data that includes socio-historical bias in terms of how it was entered and collected (e.g., the use of racialised labels in the London Metropolitan

Police's Gangs Matrix, which resulted in Black people being disproportionately represented and retained in the database⁹³). The challenge of socio-historic bias in datasets used for AI development was explained by an AI researcher we interviewed:

“For some conditions, you really do not want to incorporate variables like ethnicity or socioeconomic status in your model because you may take on a historical bias. And if you train a model and use such a variable, then you may be integrating that bias into the solutions that you are creating. In these cases, maybe you want to take the extra step of not just leaving the variable out of the model that you are creating, but actually making sure that the model predictions are statistically independent of that variable.”

- AI researcher

The third problem highlighted by the experts we interviewed is the use of data that does not account for or misrepresents social categories and determinants of the intended outcome. For instance, when data on social determinants is not collected or is misrepresented, predictions may be less accurate and underestimate risk for populations who, for example, do not have access to healthcare, do not have financial resources to seek health support, or are employed in high-risk industries. Examples of this are commonplace in everyday clinical practice, including in predictive models that predate AI⁹⁴. A race equality expert we interviewed noted:

“The kind of data that is being collected is dependent on the kinds of questions that you have asked. I think one of my biggest concerns is that, like in a lot of the COVID-19 measures of vulnerability among black and ethnic

minorities, AI applications in health are very biologically driven as opposed to taking into account the context of which people live their lives. So the social and economic circumstances, the housing circumstances, the types of jobs that they do, I think that is the sort of information that I would want to see into any data-driven technology that measures health access, health risk, health services, and so on.”

- Race equality expert

Our literature review and expert interviews highlighted that, in addition to improving the representativeness of minority ethnic groups in health data sets, the appropriateness and implications of including race or ethnicity as features in AI models must be critically appraised. The use of race or ethnicity in clinical risk prediction models is both widespread and a major source of bias in clinical practice. Some clinical risk prediction tools in use today incorporate race as an input variable and in doing so may perpetuate pre-existing healthcare discrimination⁹⁴. For example, a controversial race-based correction factor is still widely used to estimate glomerular filtration rate (eGFR), a test to assess kidney function in patients with chronic kidney disease, despite evidence indicating that using the factor may lead to underdiagnosis and delayed access to treatment for Black populations⁹⁵.

The finding of significant association between race and clinical outcomes in a data set may not itself be sufficient justification for the inclusion of race as an input variable into predictive models that influence clinical practice. In a recent editorial in the *New England Journal of Medicine*, the prevalence and harms of predictive models based on race were clearly elucidated⁹⁴:

“...the racial differences found in large data sets most likely often reflect effects of racism... In such cases, race adjustment would do nothing

to address the cause of the disparity. Instead, if adjustments deter clinicians from offering clinical services to certain patients, they risk baking inequity into the system”⁹⁴ (p.289)

The developers of AI models must therefore carefully appraise the inclusion or exclusion of race or ethnicity as features in the data on which their models are produced. Where included, appropriate justification should be given and measures taken to ensure such inclusion. Vyas et al. (2020)⁹⁴ propose three questions that should be asked of clinicians and algorithm developers:

1. Is the need for race correction based on robust evidence and statistical analyses?
2. Is there a plausible causal mechanism for the racial difference that justifies the race correction?
3. Would implementing this race correction relieve or exacerbate health inequities?

Matching training data to the implementation population

To develop data-driven technology and AI without inherent bias arising from the data it has access to, it is crucial that efforts are made to ensure the data on which algorithms are based are representative of the population in which they will be deployed. There are a variety of ways that a lack or imbalance in the collection of diverse data occurs within research and healthcare (see **Table 1** which outlines examples from health systems around the world). While it may not always be possible to match training data to the implementation population, additional measures may be applied to resample available data in such a way as to better match the patient population in which an algorithm may be applied⁹⁶.

It is crucial therefore to draw a distinction between better data generation and collection of race or ethnicity itself within healthcare data sets, and better collection of data pertaining to a wide range of social and clinically relevant factors across the entire population. Recognising the ubiquity of biases in predictive models in healthcare today, many of which have been in use for decades, is an important step

to ensuring AI models in healthcare do not replicate these same mistakes¹⁰⁰.

Alternative data sources and methods

Many machine learning methods rely upon a relatively even distribution of the number of cases falling into each class within a classification. In medical data sets, imbalance in the distribution of cases is common, particularly when attempting to classify rare events. In such cases, models may fail to adequately predict rare outcomes effectively while providing good overall performance, largely driven by accurately predicting the common outcome. A range of techniques exist to address these issues of imbalance in source data. In its simplest form, over-sampling involves duplication of less common outcomes within a data set, while under-sampling involves removal of more common outcomes to improve balance in the number of cases across classes. A range of more complex methods exist including the widely used Synthetic Minority Oversampling Technique (SMOTE) to over-sample less common outcomes by generating synthetic data based on existing data. In doing so, these methods are only able to work with information already contained within the dataset, which may fail to encompass variation arising as a result of intersections based on age, gender, deprivation or health status that may not adequately be featured in the initial data¹⁰¹. As noted by one of the academic experts we interviewed, synthetic data is often generated based on the underlying data set and would not necessarily correct for original bias within that data set. Considering this in a real-world example, elderly minority ethnic patients in the UK are known to have substantially worse self-reported health status than White people of a similar age, yet 2004 was the last year that the Health Survey for England specifically oversampled minority ethnic patients to establish these findings¹⁰². According to a recent editorial in the British Medical Journal:

“The exclusion of older ethnic minority people from population-based studies is a form of institutional racism and leads to a worrying and socially unjust dearth of knowledge about the health and social conditions of an

already disadvantaged part of the UK population.”¹⁰² (p.2)

For AI to effectively benefit minority ethnic groups, it is crucial that not only are data collected from across ethnic groups but collected with sufficient breadth and depth to capture the multitude of clinically important associations between ethnicity and demographic, social and clinical features that may exist.

Methods to improve impartiality

Similarly, bias in the process of data generation and collection may impair the effectiveness of algorithms for minority populations in the UK. A few options to mitigate this bias have already been proposed in academic literature from health and social care settings around the world. As AI bias is related to the data-generating process, one of the simpler ways of improving impartiality could be to preferentially use data sources that better represent the population to which a model will be applied, or that extend beyond routinely collected clinical and demographic data alone. There are several examples of routine non-invasive monitoring systems that collect substantial amounts of data on all patients which are relatively unbiased, such as recorded vital signs during surgical operations, and triage data collected from the first hour after presentation at the accident and emergency department, which avoid the potential

bias of clinical judgements¹⁰³. While examples of efforts to reduce bias in AI in health and social care do exist, results have been mixed, as indicated in **Table 2**, and overt efforts to minimise bias with respect to ethnicity are not commonplace.

The House of Lords Committee on Artificial Intelligence report (2017-2019) further outlines various areas where the Government could play a role in addressing bias in AI¹⁰⁸. One key action could be for the Government to release open data sets representative of the population in order to address the inequities associated with the less demographically diverse privately held datasets. The report also proposed the creation of guidance for the development of new data sets to help to minimise bias. In October 2021, the UK Medicines and Healthcare products Regulatory Agency (MHRA) released the ‘Good Machine Learning Practice for Medical Device Development: Guiding Principles’¹⁰⁹ alongside the US Food and Drug Administration (FDA) and Health Canada. They have jointly agreed ten guiding principles to inform the development of good machine learning practice to help promote safe, effective, and high-quality medical devices using AI. Principle 3 particularly states that good practice including ensuring “Clinical Study Participants and Data Sets Are Representative of the Intended Patient Population” to manage any bias, promote appropriate and generalizable performance across the intended patient population, and better identify where the model may be underperforming.

Future considerations

As there are several challenges associated with data generation and collection and equitable AI and machine learning technologies there are a number of research areas that could be impactful. Fundamental starting points for improvement in this area include:

- **Developing and refining catalogues of publicly available data sets with detailed descriptions of their make-up, data dictionaries, and consideration of whether repurposing may impact on inequity**
- **Working towards improving the breadth and quality of NHS routine data across the following three areas:**

1. Improving front-line entry by clinical staff through staff education, adequate training and simplified data entry systems
2. Identifying factors required to increase data sharing by patients such as through enhanced information and transparency (e.g., on usage, perceived benefits and risks)
3. Enabling wider data linkage of healthcare data sets to more fully understand the clinical interactions of patients

What was done?	Whose data was used?	Where was the bias?	What was the impact?
<p>AI-powered dermatological algorithm called Skin Image Search⁹², by First Derm was tested on Fitzpatrick 6 skin type (dark skin) and dermatological conditions relevant in the Ugandan setting</p>	<p>Dermatological images selected from a total of 173 images from the electronic database of The Medical Concierge Group, a Ugandan telehealth company</p>	<p>The poor accuracy of the AI app suggests the First Derm application was mainly trained based on predominantly Fitzpatrick 1/2 skin types (Caucasian skin) and less of Fitzpatrick 5/6 skin types (dark coloured skin). Additionally, the poor accuracy reflects that the app was trained on dermatitis data sets rather than fungal images, likely based on the prevalence of dermatitis among North American and European settings</p>	<p>Overall diagnostic accuracy of the AI app was low at 17%, with varying accuracy of predictability levels. The AI app made the correct diagnosis in up to 80% of dermatitis cases but did not make a correct diagnosis for fungal diseases</p>
<p>The American Heart Association’s Get with the Guidelines–Heart Failure intended to predict in-hospital mortality in patients with acute heart failure⁹⁷</p>	<p>Clinical records of a cohort of 39,783 patients admitted to hospital with a diagnosis of heart failure</p>	<p>A risk score was developed in which non-Black patients are assigned higher risk score than Black patients, which therefore regards Black patients as lower risk, thereby raising the threshold for clinical intervention. However, the American Heart Association does not provide a rationale for this adjustment</p>	<p>The algorithm remains in use and readily available as a risk prediction resource for clinicians. Black patients may be less likely to receive cardiac care</p>
<p>Algorithm developed by Optum^{98,99} used to identify patients with complex medical needs to develop a “risk score”</p>	<p>Health records for nearly 50,000 patients of an unnamed US medical centre: 44,000 White patients and 6,000 Black patients</p>	<p>The algorithm was trained to identify patients with higher anticipated future healthcare costs as higher risk. However, the algorithm prioritised patients with higher incomes and health insurance plans that covered more doctor visits and more expensive prescriptions, while ignoring that many low-income patients (disproportionally African American) are more likely to seek medical care once their symptoms are severe</p>	<p>African Americans comprised 18% of the medical centre’s high-risk group but, looking at the actual patient health data, should have represented 47% of the high-risk group. The disparity resulted in White patients being granted access to resources ahead of African American patients in poorer health</p>

Table 1: Examples of bias arising from predictive models in healthcare and the population affected

What was done?	Whose data was used?	How was it done/ how does it work?	What was the impact?
<p>Machine learning models used to predict eight outcomes for individual cases in children’s social care¹⁰⁴, while determining whether they do this equally well for different groups</p>	<p>Data from between c.700 and c.24,000 children’s social care cases provided by four local authorities in England</p>	<p>The research team completed the analysis for each local authority separately and prepared the data sets for the model, validated the data and created new ways of summarising or categorising the data. They also extracted from the text ‘vulnerabilities’ (e.g., food poverty, parental substance abuse) the child was experiencing to paint a richer picture than the structured data alone could</p>	<p>The models built missed most children at risk of the outcome, which led to the model discouraging a social worker to support a child or young person, potentially resulting in harm to them. The overall transparency of reporting and considerate nature of data curation exposes the weaknesses of using such data for decision-making in social care. Such reporting of negative findings is helpful to inform future research and practice</p>
<p>AI and machine learning technology used to understand individual health information needs from internet searches^{105,106}</p>	<p>Individuals across Africa with regard to their needs as related to HIV/AIDS, malaria and tuberculosis</p>	<p>An analysis of Bing searches related to HIV/AIDS, malaria, and tuberculosis from 54 African nations automatically derived a set of common topics for each disease, revealing a widespread interest in various types of information, including disease symptoms, drugs, as well as stigma, beliefs in natural cures</p>	<p>The project created a new data set useful for improving the understanding of health needs in Africa and can be further used to improve access to care and allocate resources</p>
<p>Smartphones and wearable devices used to detect diabetes^{98,107}</p>	<p>Individuals initially enrolled in an online heart study, as well as African Americans and Asian Americans who participated in extra rounds of clinical trials designed for their explicit inclusion</p>	<p>The optical technique photoplethysmography (PPG) is used to allow the algorithm to extract features such as blood pressure and heart rate. Users are required turn on the phone’s light, place their fingertip over the phone’s camera lens to provide data to the algorithm</p>	<p>The project demonstrates that a PPG signal alone can be used in deep learning to detect diabetes. To ensure the algorithm is as effective on darker complexions, extra rounds of clinical trials on African Americans and Asian Americans (both higher risk groups) are being conducted</p>

Table 2: *Examples of attempts to reduce AI bias in healthcare and the population affected*

Algorithm development: A multidisciplinary approach can improve performance

Summary of our research findings

- Defining the end goal is a crucial first step to develop a technology that will benefit minority ethnic groups once deployed that several of the AI experts interviewed felt is often overlooked. This should involve critical reflection on what the algorithm will directly be used for, who it will be affecting, and what this means in terms of data requirements and outcome variables used
- There is consensus in academic literature that developers should build in sensitivity checks to reduce bias. Such checks may include simulated data sets and running counterfactual simulations, during the design and subsequent phases of AI
- Several of the experts we interviewed noted that algorithm development should not only draw from mathematical notions but from multi-disciplinary knowledge on disease prevalence and potential causes for differential outcomes
- There are several statistical methods to prevent lower performance of algorithms for minority ethnic groups, such as data stratification or sub-population weighting, but these are still not standard practice in the development and testing of algorithms.
- A challenging aspect of AI development is the problem of explainability or interpretability – often known as the “black box” problem. In the context of healthcare, it is crucial that training and support for clinicians is provided to clearly explain healthcare AI decisions and reflects the breadth of circumstances in which AI may be applied

Outputs versus inputs: who is it for and who will it affect?

Defining the end goal should involve critical reflection on what the algorithm will directly be used for, who it will be affecting, and what this means in terms of data requirements and outcome variables used. There was consensus among the AI experts (academic; industry) we interviewed that defining the end goal is arguably one of the most essential – but commonly overlooked – steps in the pipeline with regards to ensuring equity for minority ethnic groups.

“A good portion of our research actually focuses less on, what are the inputs to the algorithm, and instead focusing on, what are we trying to predict? And is that what we’re actually trying to predict?”

And also, particularly in the context of, what does this mean? I do not think there is any one way to automate this. I think it needs to be very context-specific and say, what is this algorithm directly being used for? Who is it affecting?”

- AI researcher

The Obermeyer et al. study (2019)⁹⁹ (see **Table 1**, the model which used predicted future health needs based on prior use of healthcare services) was mentioned by many of the AI experts we interviewed as an example of why it is crucial to understand how and where a model will be

implemented, who it will be affecting, and what this means in terms of data requirements and outcome variables used. In the Obermayer et al.⁹⁹ example, differences in the availability of healthcare services were not considered as a potential influence on the relationship between healthcare needs and healthcare utilisation, meaning that the algorithm systematically underestimated health needs of Black patients who have less access to healthcare due to disparities in US health system spending. It shows how insufficient consideration of potential application outcomes can result in creating new inequities in the wider healthcare system, with Black people being less likely to be referred to health services than White people who were equally sick.

Developing AI with bias in mind

Most AI models implemented in clinical practice seek to predict an outcome based on what is known about a patient and their clinical context. In doing so, the objective of models has been to predict outcomes correctly as frequently as possible. In such optimisation, it is likely that the predictive accuracy of a model will vary across patient characteristics including ethnicity, age, gender and health status. Producing the most effective predictive model across the whole population may therefore be at odds with a need to ensure the model works similarly well for all members of that population. The extent to which model accuracy across the whole population should be reduced to ensure similar accuracy across patient groups remains a largely subjective and context dependent question for those developing and implementing AI models in healthcare.

A proposed way forward is to work with developers to build in sensitivity checks, including creating simulated data sets with high numbers of omitted variables to determine whether models would produce similar results if features of the underlying data on which they are trained are added or removed¹⁰³. As described in Obermeyer et al. (2019), without the techniques to readily scrutinise and critique the performance of predictive models, it is possible that bias based on ethnicity “can arise from reasonable choices” and that by using the overall predictive accuracy of a model “cost seemed to be an effective proxy for health yet still produced large biases.”⁹⁹

Methodological innovations

In the rapidly evolving arena of AI in healthcare, methodological innovations have generally proceeded ahead of the means by which to ensure these models are equitable. Well-intentioned innovations with high predictive accuracy meeting an important clinical need may unknowingly introduce biases that are only identified in the post-implementation period. To address this, a range of methods have recently been developed to identify the potential for AI models to systematically introduce bias or inequity in their implementation, however their application is not yet widespread^{110–113}. Instead of seeking to achieve the highest predictive accuracy in the overall population, models may instead seek to minimise the occurrence of a ‘worst case scenario’ of harmful misclassification¹¹⁴ or incorporate other measures of fairness directly into the objective function of the model¹¹⁵. Improving the fairness of such models by as much as 98% was associated with only a 4% reduction in overall predictive accuracy¹¹⁵, indicating that in some cases the cost of building fairness into modelling considerations may be small.

Appropriate modelling of social determinants of health

Several of the experts we interviewed, across disciplines (academic; industry; healthcare; charities), highlighted that equitable modelling requires a robust understanding of the issue at hand, and the involvement of clinicians and public health experts may be essential to fully understand and reflect the intersection between the relevant biological, social, and genetic factors and how they impact on the outcome variable. As such, algorithm development should not only draw from mathematical notions but draw from intersectionality theory and multi-disciplinary knowledge on disease prevalence and potential causes for differential outcomes.

In line with data generation and collection considerations highlighted above, the appropriateness of including or controlling for social factors should further be considered in the context of how the algorithm will be used in practice. For example, one academic expert we interviewed commented that an algorithm that uses ethnicity as a proxy to predict the prevalence of a medical condition by association might be acceptable when

used for low-risk preventative health interventions but becomes highly unethical when aiming to deny insurance or treatment. Yet, others we interviewed (healthcare; academic) were concerned that the use of race and ethnicity as associative variables risks perpetuating incorrect assumptions about the mechanism of causality in health disparities (e.g., such as assuming a biological foundation where it is in fact explained by social determinants). This point is echoed in articles found through our literature review on the implications of undertheorised operationalisation of ethnicity and race in biomedical research^{6,116}, such as noted in Williams (2002):

“It is always necessary to indicate why race or ethnicity is being utilised, the limitations of racial and ethnic data, and how findings should be interpreted. Given the strong association between racial/ethnic status and socioeconomic status (SES), the presentation of racial differences should routinely stratify them from SES across racial groups. Failure to attend to these issues of presentation and interpretation may mis specify complex health risks and even perpetuate negative racial stereotypes.”¹¹⁶
(p. 4837)

Tools and resources for equity-sensitive analytics

Performance testing is often measured based on aggregate outcomes of the overall population, thereby obscuring the potentially poorer performance of an algorithm for minority sub-sets. In the context of challenges for equitable practice in the AI development stage, several of the AI experts in

academic research and industry we interviewed noted that statistical methods to prevent lower performance of algorithms for minority ethnic groups, such as data stratification or sub-population weighting, are still not standard practice in the development and testing of algorithms.

“We usually focus on the accuracy of an algorithm across the whole population. We assume that the performance is the same for everyone within the test data set. But what people are not doing is picking apart the sub-populations within the test data set and then looking at the performance on each of them separately. And that is just a subgroup analysis, which is often done in other areas of research, but not enough yet in AI.”
- AI researcher

The early COVID-19 data response provided a key example of the potential impacts of overly simplistic data science approaches, where insufficient breakdown of data based on gender and ethnicity masked the disproportionate impact of the pandemic on Black and minority ethnic women. Experts we interviewed highlighted that statistical methods to prevent this are not particularly complex and could be promoted through widely accessible tools and resources that guide developers through equity-sensitive development and testing. For example, testing and validating the accuracy of a model on population subgroups can provide an indication of how well the algorithm performs across different groups, but is currently not often conducted prior to wider trialling or implementation. Another important but commonly overlooked step highlighted by the interviewees is out-of-sample testing to validate algorithm performance on data other than the training data used to develop the algorithm. Without such validation, the algorithm risks performing worse on populations with different demographics than the training data.

Explainability/interpretability of problems

Another challenging aspect of AI development is the problem of explainability or interpretability, often known as the “black box” problem⁶⁴. An explainable algorithm is one in which humans can know how the algorithm is doing what it is doing. However, currently, less explainable algorithms seem to be more accurate. As such, future healthcare AI and data-driven technology may recommend healthcare decisions affecting the safety of patients that cannot be straightforwardly explained to patients and clinicians. While resources are emerging to aid the explanation of AI model findings in everyday practice, it is crucial that training and support for clinicians to

clearly explain healthcare AI decisions reflects the breadth of circumstances in which AI may be applied.

However, such a relationship between the technology and humans raises both technical and moral concerns. As AI is increasingly integrated into clinical practice, robust quality assurance processes for technologies and systems, in order to ensure understanding, review and maintenance of consistency in results, will be required⁶⁴. Deprived of the ability to explain predictions within models, clinicians and policymakers may be unable through “black box” models alone to understand the relationships that may exist between groups of patients and outcomes. To prevent this, transparency and explainability of AI models is a central component of the Code of Conduct for Data-Driven Health and Care Technology^{60, 117}. It is also necessary that clear expectations regarding explainability are developed and understood by clinicians, patients and other relevant stakeholders before, during and after deployment.

Key definitions

Data stratification: The sorting of data into smaller, more defined groups or layers (strata) based on a predetermined set of criteria

Out-of-sample testing: Testing the performance of a model or algorithm using new data on which it was not trained. This may include testing using data provided by entities outside of the team developing the original model

Resampling: Selecting a subsample of a data set in order for a model to better represent its intended implementation population

Sensitivity checks/modelling: A range of processes examining changes in the outcome of a model following removal or modification of different input variables

Sub-population weighting: Altering the proportion of individuals from subpopulations in a data set to enhance representation of subpopulations otherwise underrepresented in the data set

Over-sampling: The duplication of under-represented outcomes within a data set

Synthetic data: Data that is artificially created rather than being generated organically

Future considerations

There is still substantial work to do to ensure the end goal is well defined, and develop statistical methods to improve algorithm performance in the AI space. Some initial high-opportunity areas that could be impactful include:

- **Developing and extending existing research guidelines (e.g., extension of CONSORT AI), including principles for transparent reporting of sub-population performance and characteristics of training data**
- **Developing and evaluating computational approaches to prevent and identify bias in modelling across the potential scope of implementation of AI in healthcare**

Deployment, utilisation and patient and clinical outcomes: Implementation in the NHS

Summary of our research findings

- Efforts to enhance the availability of data for minority ethnic populations are important at the when generating and collecting data, but post-deployment, any underrepresentation may result in the development and roll-out of products that are less likely to support the needs of certain groups or may exacerbate existing inequities
- The Office for National Statistics (ONS) reported that a significant proportion of the UK population does not have internet access. As evidence suggests a digital divide with regard to access and utilisation of technology across geographic location, age, ethnicity, gender, state of health and socioeconomic status it is important to promote initiatives to bridge this divide.
- Planning and continuous audit after roll-out can illuminate potential unequal impacts of technologies on different population groups, specifically minority ethnic groups, identifying potential bias and other negative outcomes, which can then be addressed in a timely manner
- Engaging minority ethnic patients on their experience of the technology can provide a nuanced understanding of patient outcomes, such as patient experience or perceived quality of care
- Clear information for clinicians on the purpose and use of data-driven technologies and AI is a critical tool in guiding best practice, optimising clinical buy in, and aiding precise application of the technology

The experts we interviewed from NHS practice highlighted that the adoption of AI solutions by digitally mature care providers may leave behind other centres and their patient populations, thereby further widening existing inequities. As discussed by both the charity and industry experts we interviewed, it is often unclear what the intended outcome from a deployed algorithm is or should be. For example, an algorithm may improve economic effectiveness or diagnostic accuracy, but reduce outcomes in terms of healthcare quality or how patients experience care, and this may differ across patient or user groups. As such, there was consensus between those we interviewed, across disciplines, that there is an urgent need to explicate how to define and measure deployment, utilisation and patient and clinical outcomes of AI relating to ethnic equity, including those related to distributional impact in terms of:

1. Disparities in deployment, access, and usage;
2. Unequal patient outcomes, such as systematic differences in clinical impact of an AI intervention but also the impact of

innovation on, for example, perceived quality of care or healthcare utilisation;

3. The cultural and socio-technical impact of AI on clinician performance and the wider health infrastructure.

Challenges in equitable deployment, access and utilisation of AI technologies

As previously noted, efforts to enhance the availability of data for minority ethnic populations are important. Post-deployment, such underrepresentation in data may result in the development and roll-out of products that are less likely to support their needs or may exacerbate existing inequities. As such, the collection of data and evidence on the deployment, utilisation, and health and wider social impacts of the technology among minority ethnic groups is also essential in striving towards equity. As noted by one academic expert we interviewed, disparities in tech access and usage can result in “data poverty”⁽¹¹⁸⁾ becoming a social determinant of health in and of itself.

Acknowledging the digital divide

While we have thus far outlined the challenges associated with technology primarily used by clinicians, a range of technologies and apps are designed for use by patients themselves. Such technologies raise further challenges around data poverty and the digital divide.

Evidence from international literature suggests a digital divide with regard to access and utilisation of technology among the public across geographic location, age, ethnicity, gender, state of health and socioeconomic status^{119–122}, although it remains somewhat unclear if similar patterns exist in the UK. In 2020, the Office for National Statistics (ONS) reported that a significant proportion of the UK population (4%) did not have internet access, and approximately 6% of the UK population did not use the Internet^{123,124}. In healthcare, research conducted by the Ada Lovelace Institute found that individuals from minority ethnic groups were less likely to have used health and wellbeing apps, personal fitness apps, COVID-19 contact tracing apps, and symptom checking apps than individuals from White ethnic groups⁴⁹.

The diffusion of innovation theory¹²⁵ argues that novel technologies are more frequently adopted by individuals or countries of higher social or economic status. In some English-speaking countries, it has further been suggested that cultural and language barriers may account for lower use of technology amongst some populations^{119,126}. Like their impact on the wider social determinants of health⁴⁴, social factors such as lack of housing or geographic location have also impacted patients' accessibility to technology¹²⁷.

The Government's recent "Levelling up" white paper outlines enhanced digital connectivity as a core mission to ensure there is nationwide 4G coverage with the ambition of 5G coverage for the majority of the population⁵⁶. There are some concerns that it may be difficult to achieve this where it is not commercially viable to do so, and the full coverage goal (to be realised by 2030) does not cover the hardest to reach areas, including over 100,000 households¹²⁸.

In addition to widening internet access, it is important to support health and technology literacy to bridge the digital divide. An example of an initiative designed to bridge the digital gap in healthcare was NHS England's Widening Digital

Participation programme, which by 2016 had trained over 220,000 people to use digital health resources, including those most likely to be socially or digitally excluded¹²⁹.

The need for patient-driven evaluation and outcomes

As noted, the voices of patients, particularly those from minority groups rarely feature in defining the clinical challenges to which AI is applied. Our interviewees mentioned the need for harder-to-measure outcomes, such as more qualitative insights on patient experiences or the perceived quality of care. For example, one charity expert we interviewed suggested the enhanced use of cross-sectional surveys or qualitative studies to understand the real impacts of these technologies on patients' lives across different intersections of society. Other experts mentioned the need for patient engagement in their interviews, such as those proposed by the AI Now Institute in New York which developed a framework for local communities to enable them to review algorithms prior to implementation by a public agency. Recommendations for meaningfully involving diverse patients and public members in varying stages of the pipeline can be found in the section "Patient and public involvement and engagement" below.

Cultural and socio-technical impacts of data-driven technologies

Some experts we interviewed (charity; healthcare; academic) discussed the somewhat underexplored issue of the cultural and socio-technical impact of data-driven technologies on health delivery systems and the patient-clinician relationship. This may be particularly relevant to the use of decision support algorithms (e.g., rather than AI used for administrative health purposes) where data-driven systems may directly inform and shape the clinician's treatment decisions or advice to the patient. One interviewee, an academic expert in AI and behavioural science, suggested the potential of decision support algorithms to limit implicit clinician bias based on race or ethnicity^{130,131}, arguing that this would support enhanced focus on physiological input (i.e., "not filtered through the mindset and assumptions of the clinician"). Yet, other experts we interviewed (charity; healthcare) had concerns about the limited understanding of how decision

support tools and clinical decision-making interact, particularly in the context of an already unequal power structure between patients and clinicians.

The need to further understand the socio-technical aspects of deployment, utilisation and patient and clinical outcomes in healthcare was discussed by the experts we interviewed at two levels. First, in terms of the degree to which clinicians are influenced by decision support tools and, in the case of biased algorithms, trust the predicted outcome or are able to correct for it. For instance, in some cases, ability to question or contest results from such tools may be complicated by the aforementioned explainability/interpretability problem¹³². Second, it was discussed by the experts in terms of how data-driven decision support tools may impact the clinician-patient relationship, shared decision-making, and subsequent healthcare quality. Beyond existing power asymmetries between patients and clinicians, there is evidence of racial disparities in clinician-patient interactions, such as with regard to how patients are engaged in treatment decisions^{133,134} and feel listened to (for example, in the case of Black women in maternity care)¹³⁵, which may now be further overlaid by the additional complexity of a data-driven system that influences, guides, or shapes how clinical decisions are made. One expert explained:

Future considerations

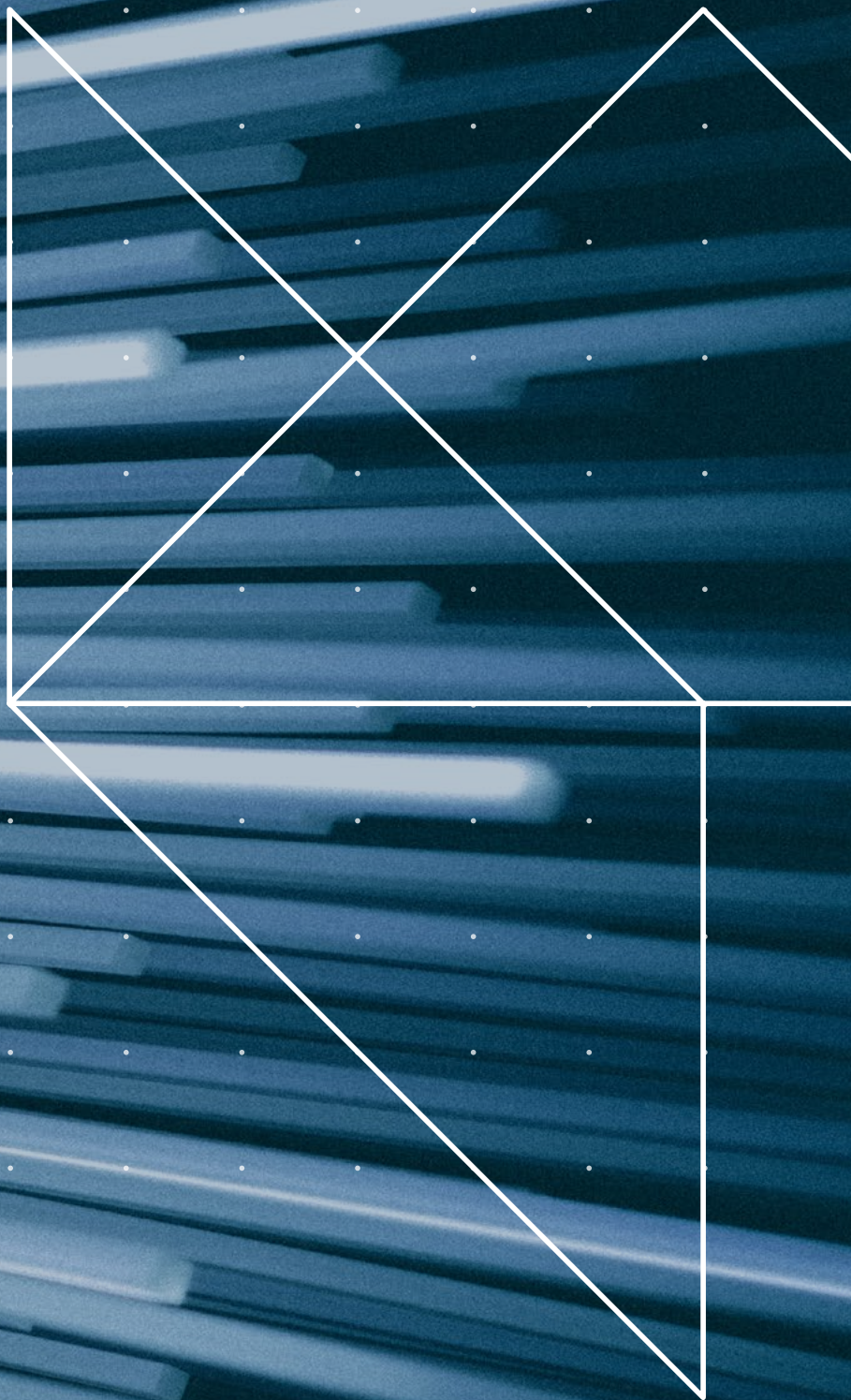
Given that the post-deployment phase of the pipeline has a direct impact on patient access, utilisation and outcomes, there are some key areas for consideration such as:

- **Design and trialling of toolkits to evaluate distributional impact including on 1) sub-population performance; 2) potential disparities in deployment, access and usage; 3) patient outcomes**
- **Developing participatory design initiatives to support codesign of AI applications with local clinician and patient populations, and the public**
- **Improving understanding of the socio-technical and cultural impact of AI technologies on the doctor-patient relationship and wider healthcare infrastructure**

“When you are looking at the patient-healthcare provider relationship, it is not an equitable relationship, not a fair relationship. And then on top of that you overlay additional complexity which is the notion that a data-driven system is in some ways informing, influencing, or guiding the practitioner’s advice to the patient. So, what effectively is happening is that a relationship that was initially about the medical practitioner’s expertise and the patient’s willingness to understand a little bit more about what is going on and be guided, but make informed choice, is now being additionally complicated by the use of a data-driven technology system.”

- Equality expert

So far, such impacts have not been adequately investigated and require further research.



Overarching considerations

a. Developing appropriate governance systems, legislation and regulation

Summary of our research findings

- As AI becomes more embedded in healthcare systems, data protection guidelines will be increasingly important to protect citizen rights and ensure public trust and transparency
- Recent changes to the legal framework for the use of AI in the EU ushered in a new era of AI regulation, but there is uncertainty about what Brexit will mean for the UK legal framework in the long term.
- An industry developer we interviewed explained that guidance from regulators on translating legal requirements into practice can provide assurance to deployers, end-users, and patients around data, bias, governance, transparency and the intended outcomes. However, others we interviewed noted the challenges of developing regulation in the context of AI as the different aspects of AI development, intended outcome, and use are highly context-dependent
- Alongside legislation, experts we interviewed highlighted the value add of high-level ethical principles expected of AI systems, such as explainability, transparency and accountability, which are now coming into existence across several governmental, non-governmental and international organisations

At present, elements of the AI and machine learning development and deployment pipeline are covered by relevant UK and European Union (EU) legislation. For example, the UK equalities legislation bans discrimination on the basis of nine characteristics including ethnicity, sex and age. Special category data, including health-related data which is subject to additional protection under the EU General Data Protection Regulation (GDPR), partly overlaps with some of the protected characteristics of equalities legislation. Following Brexit, GDPR has been incorporated into UK data protection law (as UK GDPR) with little change to the core data protection principles, rights and obligations¹³⁶.

As AI and machine learning technologies become more embedded in our healthcare systems, data protection guidelines will be increasingly important to protect citizen rights and ensure public trust and transparency regarding data use. The risk of not having a clear procedure in place to scrutinise data used by governmental bodies or public services is illustrated in a recent non-health example in the Netherlands, which led to the Dutch government's resignation in January 2021. In this case, the governmental tax authority incorporated data on (dual) nationality in the development of risk profiles for tax evasion, resulting in a large number of individuals with minority ethnic backgrounds being wrongly accused¹³⁷.

Legal framework

Changes to the legal framework for the use of AI is expected in the EU and will usher in a new era of AI regulation. In 2020, the European Commission published a white paper on AI and an inception impact assessment and opened public consultations on an overall AI regulatory framework, seeking to position the EU as an ecosystem of 'excellence' and 'trust' for AI¹³⁸. As such, the Commission released its proposal for AI regulation in May 2021¹³⁸. There is

a great deal of uncertainty on what legal framework and subsequent regulation will look like in the UK following Brexit. However, the Medicines and Medical Devices Act 2021¹³⁹, which received Royal Assent in Spring 2021¹³⁹, seeks to address safety, the pace of innovation and the regulation of medical devices and medicines for the future in post-Brexit UK. The bill is aimed at extending the UK’s global position in personalised medicine and AI in health by allowing regulators to develop regulation in this space.

The need for improved regulation

As noted, the UK government published a National AI Strategy in September 2021 as part of efforts to signal the UK’s intention to build a world-leading pro-innovation regulatory environment⁶², with a further White Paper on governing and regulating AI expected in 2022.

AI experts in both academia and industry noted the need for improved scrutiny from external regulators, such as the Medicines and Healthcare products Regulatory Agency (MHRA) in the case of the UK. Although ethnic equity of new developments may be scrutinised through academic “gatekeeping”, such as through peer reviewers, journal editors, and ethical review boards, and supported by research reporting guidelines such as CONSORT AI (a reporting guideline for randomised trials evaluating interventions with an AI component), or potentially also quality assessment tools for AI such as QUADAS AI⁶⁹, non-academic AI innovation does not necessarily undergo the same level of scrutiny. While MHRA requires all manufacturers show their device (regardless of class) be evidenced by data that is “representative of the entire intended purpose with all patient populations and all claims foreseen for the device under evaluation”¹⁴⁰, where ‘representative’ includes (where relevant) ethnicity, race, country of origin, etc., such requirements are not expressly defined for AI/machine learning. However, the release of the Good Machine Learning Practice for Medical Device Development: Guiding Principles in October 2021 highlight the MHRA’s commitment to encouraging good practice, including the reduction of bias, in medical device development using AI¹⁰⁹. One industry developer we interviewed noted the importance of regulation in encouraging best practice among developers:

“It comes back to regulation. If the FDA says to me, ‘you are about to go build an algorithm

that you are going to want to test to predict sepsis, I am going to want to see the validity of your algorithm across these ethnic groups’, then as a developer I would probably start thinking about that day one because I know that is already a regulatory requirement.”

- Industry developer

Hence, there was consensus among interviewed experts that additional review is required at the regulatory approval stage by MHRA, approved (notified) bodies, and standard setting organisations, such as the British Standards Institute, particularly around scrutiny of the technical elements of AI, where regulation is still “catching up” with rapidly growing knowledge and innovation. For example, one industry expert we interviewed emphasised the combined role of regulatory bodies in government and industry bodies like Google and felt that it was concerning that UK and EU regulation in some circumstances are less stringent than what Google itself produces.

Robust guidance from regulators, specifically in relation to AI and machine learning technologies, on translating legal requirements into practice can provide assurance to deployers, end-users, and patients around data, bias, governance, transparency and the intended outcomes. That said, experts we interviewed noted the challenges of developing regulation in the context of AI, in terms of data requirements, outcome selection, and modelling, given they are highly context-dependent and will vary significantly depending on the intended outcome and use of the algorithm. Across the expert interviews we conducted, the following considerations were highlighted as requiring scrutiny through external regulators and public bodies (i.e., the MHRA and/or The National Institute for Health and Care Excellence (NICE) in the case of the UK):

- A description of the data set and its representativeness of different minority ethnic populations
- What measures were taken to prevent and

address bias across different minority groups in:

1. data used,
 2. defined outcome, and
 3. modelling
- Whether the performance of the algorithm has been tested and validated on different minority ethnic subgroups
 - Ethical considerations regarding how the algorithm will be used once deployed and whether there are risks of this creating or perpetuating ethnic disparities in health and healthcare

Several experts in AI we interviewed, including in academia and industry, believed that the presence of such regulation would motivate developers to increasingly consider equity (i.e., over overall performance) as an essential output of a project, and thereby support critical awareness of data requirements, sensitivity modelling, and pre-deployment validation in the data generation and collection and algorithm development stages of the pipeline. Experts we interviewed from legal and healthcare practice felt that such regulation would be an enabler, rather than a barrier, to innovation, clarifying risks, liability and potential outcomes of deployment.

Establishing ethical principles

Alongside legislation, high-level ethical principles expected of AI systems, such as explainability, transparency and accountability, are now coming into existence across several governmental, non-governmental and international organisations. For example, the UK Information Commissioner's Office (ICO) and the Alan Turing Institute have collaborated on detailed guidance to help organisations explain processes, services and decisions delivered by AI systems and data protection. Similarly, NHS X developed 'Artificial Intelligence: How to get it right', a document which offers an entry point to understand and navigate regulation to put policy into practice, and the NHS AI Lab was established in part to respond to some of the regulatory challenges identified in that report. The Lab released guidelines, such as 'A Buyer's Guide to AI in Health and Care' to support buyers making purchasing decisions for is working with key regulators and public bodies

to improve regulatory assurance for AI^{60,141}. As an example, the AI Lab's regulatory programme will fund a joined up regulatory advice and approval service and enhancements to the MHRA's Yellow Card Scheme to strengthen post-market surveillance¹⁴².

Safeguarding effectiveness in the clinical context

Furthermore, in the UK clinical context, the National Institute for Health and Care Excellence (NICE) is in the process of setting up a pilot digital health technologies (DHT) evaluation programme to establish a robust process for the national evaluation of AI and machine learning technologies⁶⁰. NICE is a public body with a longstanding role in developing guidance, advice, quality standards and information services for health, public health and social care. The establishment of the DHT programme will enable NICE to issue recommendations to the NHS and care system about which technologies should be deployed at scale. Largely, the NICE evaluation will be based on the established medical technologies guidance development process and methods, but in its pilot, it was also supplemented with a technical assessment, including examining the extent of the use of AI⁶⁰. A key aspect of the pilot focused on the role of NICE in clearly specifying the data that is required to address uncertainties in the evidence as early as possible, and to include this information in the NICE standards framework as it is further developed⁶⁰. Following the completion of the pilot, NICE announced in January 2021 that the organisation is committed to working with NHS X (now as part of NHS England) and the MHRA on developing a streamlined regulatory to access pathway in England for digital health technologies¹⁴³. One academic expert in AI quality assurance we interviewed added that, alongside NICE's role, individual NHS trusts which procure AI technologies could be better supported to review questions around ensuring safety, fairness, and equity when considering how to implement new AI technologies, such as through guidance akin to the NHS AI Lab's 'Buyer's Guide'.

Future considerations from interview findings

Governance systems, legislation and regulation, if correctly developed and implemented offer assurances that will encourage greater innovation in the AI space. As such, some key research areas for consideration include:

- **Development of an overarching framework to scrutinise equity-sensitivity which should include as a minimum:**
 - A description of the data set and its representativeness of different minority ethnic populations
 - What measures were taken to prevent and address bias across different minority groups in 1) data used, 2) defined outcome, and 3) modelling
 - Whether the performance of the algorithm has been tested and validated on different minority ethnic subgroups

Ethical considerations of how the algorithm will be used once deployed and whether there are risks of this creating or perpetuating ethnic disparities in health and healthcare

- **Development of a guide for NHS deployment settings to support informed decision-making and audit with regard to inequity risks of AI**

b. Improving the product and its impact through diversity in IT, tech and healthcare

Summary of our research findings

- The precise role of developers in producing biased algorithms is contested, but several prominent researchers working in the AI industry note that AI bias is not only a result of biased data
- The experts we interviewed highlighted that a more diverse workforce may be better equipped to predict, identify, and mitigate ‘blind spot’ problems with bias throughout development and deployment that stem from an overly homogeneous workforce
- Several experts stressed that a diverse workforce does not end at recruitment. Diversity hiring needs to be matched with the promotion of a culture of diversity and inclusion that enables equal participation in review and feedback processes and provides equal opportunities for upward progression. The unequal treatment of staff within the NHS is an example of where a culture of diversity and inclusion can be improved
- Improving diversity in the workforce should start early: tackled at the primary and secondary school level, as well as at the graduate and postgraduate level

The conscious or unconscious preferences of individual programmers, the lack of diversity within the AI developer workforce, the AI and tech industry culture, as well as the culture within individual organisations, pose a crucial challenge to the equity of the AI and machine learning technologies they develop. While the precise role of developers in producing biased algorithms is contested, several prominent researchers working in the AI industry, such as Dr Timnit Gebru and Emily Denton, as well as Kate Crawford and colleagues at the AI Now Institute at New York University, note that AI bias is not solely a result of biased data, but rather a product of unequal workforce and societies¹⁴⁴.

When considering prioritisation in the research agenda/deployment, the experts we interviewed as part of this project questioned what issues developers are designing solutions for and how these are impacted by teams that are insufficiently diverse in terms of socio-cultural background, gender, or education. The experts we interviewed noted that a lack of diversity in developer teams can lead to solutions that may not be representative of or appropriate for the intended end-users, or perpetuate stereotypes (e.g., default use of female-gendered voice assistants such as Alexa and Siri¹⁴⁵). Further, as outlined in this report, the data generation and collection, defining the end goal, and AI development stages require a significant degree of human decision-making, which will to some extent be influenced by developers’ wider assumptions about the world, and thus differ depending on developers’ educational background/discipline, socio-cultural background, and personal experiences. When development teams are too homogenous in terms of such factors, problems with bias during these processes may be more easily overlooked¹⁴⁶. For example, this was exemplified in the algorithm that misidentified two African-American women as “gorillas”, following which the respective Google developer admitted his lack of anticipation of potential faulty performance of the algorithm for darker-skinned faces^{146,147}.

Impact on clinician performance and wider hierarchies and discrimination in healthcare

Currently, the regular application of AI in healthcare is limited to a small number of clinical contexts including, but not limited to, radiology, ophthalmology and bedside clinical decision

support, as well as system efficiency and operations, where predominantly classification-based tasks are performed to assist clinicians and administrative staff. These projects arose in part through the ready availability of appropriate data, and well-defined applications in a collaboration between clinicians, academic institutions and industry. However, it is important to consider the best way to ensure clinical staff are utilising data-driven technologies while prioritising clinical and patient safety.

In the UK healthcare context, hierarchies among healthcare staff magnify discrimination within the NHS workforce, which includes widespread reports of bullying, harassment, abuse of minority ethnic staff by their colleagues and seniors, and unequal access to promotion¹⁴⁸. In 2022, interim findings of the British Medical Association's 'Racism in Medicine' survey reported that, 75.6% of the doctors who responded, reported that they had experienced racism at least once in the last two years, with 17.4% experiencing incidents on a regular basis⁴⁸. Additionally, the results of the survey found large numbers of minority respondents saw racism as a barrier to career progression as compared to 4.2% of White British respondents¹⁴⁹. A recent example of ethnicity-based discrimination, is the first successful prosecution of a race discrimination case against the General Medical Council (GMC)¹⁵⁰. The GMC themselves commissioned a report in 2019 which showed, minority ethnic physicians are twice as likely as their White counterparts to be referred to local disciplinary bodies¹⁵¹. Given the substantial challenges faced by some minority ethnic staff within the NHS, it is important to recognise the complex relationship between ethnicity and hierarchies within the NHS and develop policies to encourage staff to speak up if they are concerned with an aspect of the technology or its application, without fear of repercussions

Diversity in the NHS

Given the implications of a complex relationship between ethnicity, hierarchies, and unequal treatment of staff within the NHS, with potential impact on the equity and success of AI at the post deployment stage, it is an imperative that greater attention is paid to improving diversity within the NHS. Data from 18 NHS acute trusts in London in 2020-21 showed that White doctors were six times more likely to be offered jobs in London compared with Black doctors¹⁵². One example from Kingston

Hospital NHS Foundation showed that despite 418 Black candidates (as compared to 317 White candidates) applying for medical positions in 2020-21, no Black doctor was offered a position during that period¹⁵². As such, much more must be done at all levels to ensure diversity at all levels of the NHS, as well as fairness, inclusion, and respect towards staff.

Promoting a culture of diversity in tech

CognitionX, a global network of AI experts, previously explained "one of the reliable ways we know we can mitigate [the problem of bias and discrimination] is to have more diverse development teams in terms of specialisms, identities and experience", yet such a statement simplifies the challenges associated with minority ethnic representation in the workforce¹⁰⁸. Even when people with minority backgrounds are employed, subsequent workplace engagement and participation can be restricted by a non-inclusive workplace culture that penalises employees for raising problems with diversity and bias. The Chief Executive of Google's parent company, Alphabet, was forced to apologise for the departure of a prominent minority ethnic AI researcher, whose exit raised questions of Google's commitment to diversity and the responsible development of AI technology¹⁵³. Dr Timnit Gebru was allegedly fired in December 2020, after sending an email criticising the company's lack of progress in hiring women and minorities¹⁵⁴. This example highlights the importance of a culture of diversity and inclusion within the workforce. Such culture should avoid tokenistic inclusion of diversity policies but rather strive for meaningful and open discussion of the challenges among staff and executives.

The House of Lords Committee on Artificial Intelligence rightly recognised work towards improving diversity should start at a much earlier stage: tackled at the primary and secondary school level, as well as at the graduate and postgraduate level. One key recommendation of the report was that the Alan Turing Institute should establish mechanisms to encourage AI PhD applications from female and minority ethnic candidates to help address diversity challenges within the sector¹⁰⁸. However, engagement with minority ethnic groups to identify intersecting challenges is a fundamental starting point to facilitate greater access into the technology workforce. As noted above, open and meaningful discussion within the technology industry

and within individual organisations must also be facilitated in a non-tokenistic way to improve the experiences for minority ethnic individuals in the workplace, where clear challenges are set at all levels of organisations to remove some of these barriers.

Addressing bias in the workforce

Researchers and developers working on healthcare technologies at each of the pre-deployment stages need a more developed understanding of health inequities and the issue of bias within the technology industry, and within technology itself. Without a clear understanding, researchers and developers may not be able to understand, or be aware of the need to consider biases embedded in the algorithms themselves. The UK House of Lords Committee report explained this idea succinctly:

“Human developers set the parameters for machine learning algorithms, and the choices they make will intrinsically reflect the developers’ beliefs, assumptions and prejudices. The main ways to address these kinds of biases are to ensure that developers are drawn from diverse gender, ethnic and socio-economic backgrounds, and are aware of, and adhere to, ethical codes of conduct.”¹⁰⁸



c. Patient and public involvement and engagement throughout the development and deployment of data-driven health technologies

Summary of our research findings

Experts we interviewed stressed that the needs of patients should be incorporated into the design, funding, implementation and evaluation of new AI interventions to ensure innovation does not exacerbate existing inequities. Such inclusion can dispel misconceptions and instil trust with regards to the use of AI and machine learning

- Approaches for recruiting and involving patients should be tailored to the problem being addressed and should consider multiple socio-cultural factors including ethnicity, age, geographical location, sexual orientation, ability status, etc. NIHR offers

published practical guidelines on how to ensure patient and public voices are heard within research

- One expert noted the value of integrating lay partners as part of the team to help achieve meaningful contributions, but added that this should be combined with wider outreach initiatives to not lose the diversity of perspectives
- The use of participatory design or “co-design” methods including those focused on lived experience may offer a more integrative approach to collaboration

Developing algorithms which are the most immediately achievable or marketable may limit the opportunity for AI to reduce rather than exacerbate healthcare inequities and meet the priorities of patients. It is crucial therefore that the needs of patients are incorporated into the design, funding, implementation and evaluation of new AI interventions so they are in part designed with input from patients. It is particularly important to involve patients whose experience with digital technology and healthcare is relatively immature. Engaging patients and the public would also be useful in dispelling misconceptions and instilling trust regarding the use of AI and machine learning.

Ensuring patient and public involvement and engagement in research

The NIHR defines patient and public involvement and engagement (PPIE) in research as research being carried out ‘with’ or ‘by’ members of the public rather than ‘to’, ‘about’ or ‘for’ them¹⁵⁵. PPIE initiatives were identified as an essential component of AI development by many experts across disciplines. In general, providing the means to engage the wider public as well as underrepresented groups in research and innovation were considered as an

opportunity to educate the public in data science, unpack ‘good’ and ‘bad’ examples of AI, and increase the diversity of perspective in working groups who are defining, developing and implementing AI solutions. Examples of PPIE in AI research discussed by our interviewees included patient and public panels in data access applications (e.g., for the Whole Systems Integrated Care data set), grant panels and ethical review boards. However, it was repeatedly noted that such involvement initiatives do not often move beyond tokenism, do not tend to include a sufficiently diverse sample of representatives, and are difficult to do in a meaningful way due to the complexity and scale of AI.

Achieving meaningful contribution

Ensuring appropriate representation will depend on the make-up of the intended implementation population and should consider factors beyond ethnicity, such as age, geographical status (e.g., rural vs. urban), sexual orientation, ability status, and socio-economic status. Approaches for recruiting and involving patients should be tailored to the problem being addressed. One patient and public involvement expert we interviewed highlighted that integrating lay partners as part of the team can be particularly

helpful to achieve meaningful contributions but should be combined with wider outreach initiatives to not lose the diversity of perspectives when defining key questions, intended outputs, and strategies for dissemination. To achieve this, it is important to consider that equity of access and opportunity for citizens to be involved relies on recruitment strategies (e.g., in terms of online or physical outlets, language used, contacted organisations) and support provided to participate (e.g., travel, training, a levelled participation environment). Most AI projects will be highly complex and require substantial induction training and information provision to support and empower lay partners to contribute. Compensation for lay partners’ time and investment should go beyond financial reimbursement and include appropriate dissemination of the research outputs and efforts to ensure the project will benefit patients and the public in practice.

Moving towards participatory methods

Beyond PPIE, the use of participatory design or “co-design” methods, including those focused on lived experience (e.g., Experience-Based Co-Design)¹⁵⁶, have been proposed as a more integrative approach to collaboration. With a substantial amount of technological innovation in healthcare stemming from the US and Europe¹⁵⁷, methods such as participatory action research and community-based participatory research have been recommended to empower end-users, as well as enable researchers to gain a cultural understanding of end-users which can feed into the design and implementation of novel solutions¹⁵⁸. As noted by the experts in AI and healthcare we interviewed, such participatory methods may be particularly important for the following stages of the AI development process:

1. Shaping research questions: to identify neglected problem areas that need to be addressed and understand the nuances and intricacies of the problem being addressed based on lived experience.
2. Optimising implementation and useability: to consider how to design the front-end or user interface to ensure solutions are usable, accessible, and perceived as beneficial to local clinician and patient populations at the deployment setting.

Although good examples of PPIE and co-design in AI may still be somewhat limited, efforts to involve

patients in health research are growing at a fast pace and offer critical learning lessons towards designing more inclusive and meaningful methods for integrating patient and public perspectives in AI. Organisations such as the NIHR have published practical guidelines on how to ensure patient and public voices are heard within research, not just in the context of AI but all healthcare-related research and innovation¹⁵⁶. Additionally, the Ada Lovelace Institute has recently published a framework for involving patients and public in the collection, analysis, and storage of data¹⁵⁹.

Future considerations

As PPIE is an overarching consideration closely linked with the development and deployment process, key areas for future work related to this topic were previously listed in earlier sections of the report:

- **Identifying high-opportunity areas for data-driven solutions to respond to health needs of different minority ethnic groups (see Prioritisation in the research agenda section)**
- **Participatory design initiatives to support co-design of AI applications with local clinician and patient populations, and the public (see Deployment, utilisation and patient and clinical outcomes section)**

Addressing key knowledge and practice gaps

There is no one silver bullet to address bias and utilise AI and machine learning technology to reduce health inequities faced by minority ethnic groups. The level of dynamism required for such an undertaking requires awareness and willingness to act at several levels. Meaningful impact can only be achieved by addressing several of the factors that have been identified as part of this review. Working across government, healthcare and the technology sectors is critical to encourage positive results. While the research priorities highlighted in this report are by no means exhaustive, they offer insight into some of the most urgent areas in each stage of developing data-driven health technologies.

To develop AI and data-driven technologies that do not create or exacerbate inequities and healthcare for minority ethnic groups, it is vital that potential risks and challenges are considered at each step. This report has provided an overview of critical considerations that need to be scrutinised by AI developers, researchers, policymakers, healthcare providers and tech companies when developing, reviewing, or deploying a novel AI or machine learning approach in health and care. The report includes recommendations for addressing a wide range of issues across the development and deployment of data-driven health technologies, including but not limited to: funding disparities, bias in data and datasets, under-theorised or un-justified use of ethnicity data, biased decision-making, and unsafe or harmful deployment. As AI and data science will continue to develop at a rapid pace, it is likely that new challenges will come to light. We hope that this report will function as an initial catalyst to collectively recognise, address, and prioritise initiatives to enable data-driven technologies in health and care for the benefit of all.



References

1. Public Health England. Beyond the Data: Understanding the Impact of COVID-19 on BAME Communities [Internet]. Public Health England; 2020 Jun [cited 2020 Dec 14]. Available from: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/892376/COVID_stakeholder_engagement_synthesis_beyond_the_data.pdf
2. Salway S, Carter L, Powell K, Turner D, Mir G, George T. Race equality and health inequalities: Towards more integrated policy and practice [Internet]. Race Equality Foundation; 2014 Apr [cited 2020 Dec 14]. Available from: <https://raceequalityfoundation.org.uk/health-care/race-equality-and-health-inequalities-towards-more-integrated-policy-and-practice/>
3. Ghafur S, Fontana G, Halligan J, O'Shaughnessy J, Darzi A. NHS data: Maximising its impact on the health and wealth of the United Kingdom [Internet]. 2020 Feb [cited 2020 Dec 14]. Available from: <http://spiral.imperial.ac.uk/handle/10044/1/76409>
4. World Health Organization (WHO). Ethics and governance of artificial intelligence for health [Internet]. 2021 Jun [cited 2021 Jun 28]. Available from: <https://www.who.int/publications/i/item/9789240029200>
5. Office for National Statistics (ONS). Tables produced using the Annual Population Survey and the Labour Force Survey: ethnicity and religion broken down by country [Internet]. 2018 Jan [cited 2021 Jan 5]. Available from: <https://www.ons.gov.uk/employmentandlabourmarket/peopleinwork/labourproductivity/adhocs/007975tablesproducedusingtheannualpopulationsurveyandthelabourforcesurveyethnicityandreligionbrokendownbycountry>
6. Lee C. 'Race' and 'ethnicity' in biomedical research: how do scientists construct and explain differences in health? *Soc Sci Med*. 2009 Mar;68(6):1183–90.
7. Afshari R, Bhopal RS. Ethnicity has overtaken race in medical science: MEDLINE-based comparison of trends in the USA and the rest of the world, 1965–2005. *International Journal of Epidemiology*. 2010 Dec 1;39(6):1682–3.
8. Institute of Medicine (US) Committee on Understanding and Eliminating Racial and Ethnic Disparities in Health Care. Unequal Treatment: Confronting Racial and Ethnic Disparities in Health Care [Internet]. Smedley BD, Stith AY, Nelson AR, editors. Washington (DC): National Academies Press (US); 2003 [cited 2020 Dec 1]. Available from: <http://www.ncbi.nlm.nih.gov/books/NBK220358/>
9. McClure ES, Vasudevan P, Bailey Z, Patel S, Robinson WR. Racial Capitalism Within Public Health-How Occupational Settings Drive COVID-19 Disparities. *Am J Epidemiol*. 2020 Nov 2;189(11):1244–53.
10. Thomas B. Health and health care disparities: the effect of social and environmental factors on individual and population health. *Int J Environ Res Public Health*. 2014 Jul 21;11(7):7492–507.
11. Khunti K, Routen A, Pareek M, Treweek S, Platt L. The language of ethnicity. *BMJ* [Internet]. 2020 Nov 23 [cited 2020 Dec 1];371. Available from: <https://www.bmj.com/content/371/bmj.m4493>
12. Bunglawala Z. Please, don't call me BAME or BME! [Internet]. Civil Service. 2019 [cited 2020 Dec 1]. Available from: <https://civilservice.blog.gov.uk/2019/07/08/please-dont-call-me-bame-or-bme/>
13. Commission on Race and Ethnic Disparities. Commission on Race and Ethnic Disparities: The Report [Internet]. 2021 Mar [cited 2021 Mar 18]. Available from: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/974507/20210331_-_CRED_Report_-_FINAL_-_Web_Accessible.pdf
14. Delgado JL. Beyond Diversity—Time for New Models of Health. *New England Journal of Medicine*. 2022;386:503–5.
15. Bauer GR. Incorporating intersectionality theory into population health research methodology: Challenges and the potential to advance health equity. *Social Science & Medicine*. 2014 Jun 1;110:10–7.
16. Hankivsky O, Doyal L, Einstein G, Kelly U, Shim J, Weber L, et al. The odd couple: using biomedical

- and intersectional approaches to address health inequities. *Glob Health Action*. 2017 Dec;10(-sup2):1326686.
17. Hankivsky O, de Leeuw S, Lee J-A, Vissandjé B, Khanlou N. *Health Inequalities in Canada: Intersectional Frameworks and Practices*. Vancouver: UBC Press; 2011.
 18. Kelly UA. Integrating intersectionality and biomedicine in health disparities research. *ANS Adv Nurs Sci*. 2009 Jun;32(2):E42-56.
 19. Lapalme J, Haines-Saah R, Frohlich K. More than a buzzword: how intersectionality can advance social inequalities in health research. *Critical Public Health*. 2019 Feb 20;30:1-7.
 20. Verdonk P, Muntinga M, Leyerzapf H, Abma T. From gender sensitivity to an intersectionality and participatory approach in health research and public policy in the Netherlands. In: *The Palgrave Handbook of Intersectionality in Public Policy*. Springer; 2019. p. 413-32.
 21. The King's Fund. *Tackling racial inequalities in health and care* [Internet]. 2020 [cited 2020 Dec 1]. Available from: <https://www.kingsfund.org.uk/press/press-releases/tackling-racial-inequalities-health-care>
 22. Public Health England. *Disparities in risk and outcomes of COVID-19* [Internet]. Public Health England; 2020 Jun [cited 2020 Jun 6]. Available from: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/908434/Disparities_in_the_risk_and_outcomes_of_COVID_August_2020_update.pdf
 23. Knight M, Bunch K, Tuffnell D, Patel R, Shakespeare J, Kotnis R, Kenyon S, Kurinczuk JJ (editors) on behalf of MBRRACE-UK. *Saving Lives, Improving Mothers' Care - Lessons learned to inform maternity care from the UK and Ireland Confidential Enquiries into Maternal Deaths and Morbidity 2017-19*. Oxford: National Perinatal Epidemiology Unit, University of Oxford; 2021.
 24. Bhui K, Stansfeld S, Hull S, Priebe S, Mole F, Feder G. Ethnic variations in pathways to and use of specialist mental health services in the UK. *Systematic review*. *Br J Psychiatry*. 2003 Feb;182:105-16.
 25. Toleikyte L, Salway S. Local action on health inequalities: Understanding and reducing ethnic inequalities in health. *Public Health England*; 2018 Aug p. 81.
 26. Bhui KS, McKenzie K. Rates and risk factors by ethnic group for suicides within a year of contact with mental health services in England and Wales. *Psychiatr Serv*. 2008 Apr;59(4):414-20.
 27. Parry G, Van Cleemput P, Peters J, Walters S, Thomas K, Cooper C. Health status of Gypsies and Travellers in England. *J Epidemiol Community Health*. 2007 Mar;61(3):198-204.
 28. McManus S, Bebbington PE, Jenkins R, Brugha T, editors. *Mental health and wellbeing in England: the adult psychiatric morbidity survey 2014*. Leeds: NHS Digital; 2016.
 29. NHS Digital. *New figures on detentions under the Mental Health Act in 2019-20 published today* [Internet]. 2020 Oct [cited 2021 Jan 13]. Available from: <https://digital.nhs.uk/data-and-information/publications/statistical/mental-health-act-statistics-annual-figures/2019-20-annual-figures#>
 30. Department of Health and Social Care. *Landmark reform of mental health laws* [Internet]. 2021 Jan [cited 2021 Jan 13]. Available from: <https://www.gov.uk/government/news/landmark-reform-of-mental-health-laws>
 31. Henderson J, Gao H, Redshaw M. Experiencing maternity care: the care received and perceptions of women from different ethnic groups. *BMC pregnancy and childbirth*. 2013;13(1):1-14.
 32. Raleigh V, Hussey D, Seccombe I, Hallt K. Ethnic and social inequalities in women's experience of maternity care in England: results of a national survey. *Journal of the Royal Society of Medicine*. 2010;103(5):188-98.
 33. Kapadia D, Zhang J, Salway S, Nazroo J, Booth A, Villarroel-Williams N, et al. *Ethnic Inequalities in Healthcare: A Rapid Evidence Review* [Internet]. NHS Race & Health Observatory; 2022 Feb [cited 2022 Feb 16]. Available from: https://www.nhs-rho.org/wp-content/uploads/2022/02/RHO-Rapid-Review-Final-Report_v.7.pdf
 34. Metersky ML, Hunt DR, Kliman R, Wang Y, Curry M, Verzier N, et al. Racial disparities in the frequency of patient safety events: results from

- the National Medicare Patient Safety Monitoring System. *Med Care*. 2011 May;49(5):504–10.
35. Sjoding MW, Dickson RP, Iwashyna TJ, Gay SE, Valley TS. Racial bias in pulse oximetry measurement. *New England Journal of Medicine*. 2020;383(25):2477–8.
 36. Department of Health & Social Care. Government launches landmark reviews to tackle health disparities [Internet]. GOV.UK. 2022 [cited 2022 Feb 16]. Available from: <https://www.gov.uk/government/news/government-launches-landmark-reviews-to-tackle-health-disparities>
 37. NHS England. Core20PLUS5 – An approach to reducing health inequalities [Internet]. [cited 2022 Feb 14]. Available from: <https://www.england.nhs.uk/about/equality/equality-hub/core2oplus5/>
 38. The King’s Fund. NHS England and tackling inequalities [Internet]. The King’s Fund. 2021 [cited 2022 Feb 14]. Available from: <https://www.kingsfund.org.uk/blog/2021/12/nhs-england-and-tackling-inequalities>
 39. Public Health England. Chapter 6: social determinants of health [Internet]. Public Health England; 2017 Jul [cited 2020 Dec 1]. Available from: <https://www.gov.uk/government/publications/health-profile-for-england/chapter-6-social-determinants-of-health>
 40. Dahlgren G, Whitehead M. Policies and strategies to promote social equity in health. Background document to WHO - Strategy paper for Europe. Institute for Futures Studies, Arbetsrapport. 1991 Jan 1;14.
 41. Marmot M, Allen J, Goldblatt P, Boyce T, McNeish D, Grady M, et al. Fair Society, Health Lives. The Marmot Review [Internet]. Institute of Health Equity; [cited 2020 Dec 1]. Available from: <https://www.instituteoftheequity.org/resources-reports/fair-society-healthy-lives-the-marmot-review/fair-society-healthy-lives-full-report-pdf.pdf>
 42. Lorant V, Bhopal R. Ethnicity, socio-economic status, and health research: Insights from and implications of Charles Tilly’s theory of Durable Inequality. *Journal of epidemiology and community health*. 2011 Mar 1;65:671–5.
 43. Razai MS, Kankam HKN, Majeed A, Esmail A, Williams DR. Mitigating ethnic disparities in covid-19 and beyond. *BMJ*. 2021 Jan 15;372:m4921.
 44. Marmot M, Allen J, Boyce T, Goldblatt P, Morrison J. Health Equity in England: The Marmot Review 10 Years On [Internet]. Institute of Health Equity; 2020 Feb [cited 2020 Dec 1]. Available from: <https://www.health.org.uk/publications/reports/the-marmot-review-10-years-on>
 45. Marmot M, Allen J, Goldblatt P, Herd E, Morrison J. Build Back Fairer: The COVID-19 Marmot Review [Internet]. 2020 Dec [cited 2021 Jan 8]. Available from: <https://www.health.org.uk/publications/build-back-fairer-the-covid-19-marmot-review>
 46. Financial Times. Campaigners accuse UK racial inequality report of complacency. 2021 Mar 31 [cited 2021 Apr 5]; Available from: <https://www.ft.com/content/8df-da80c-b7e1-4083-b874-9425adfe563c>
 47. Royal College of Psychiatrists. RCPsych response to the report of the Commission on Race and Ethnic Disparities [Internet]. 2021 Apr [cited 2021 May 10]. Available from: <https://www.rcpsych.ac.uk/news-and-features/latest-news/detail/2021/04/01/rcpsych-response-to-the-report-of-the-commission-on-race-and-ethnic-disparities>
 48. British Medical Association (BMA). BMA submission to the Commission on Race and Ethnic Disparities inquiry [Internet]. 2021 Mar [cited 2021 May 10]. Available from: <https://www.bma.org.uk/media/3583/bma-submission-race-disparities-and-inequality-in-the-uk.pdf>
 49. Ada Lovelace Institute. The data divide: Public attitudes to tackling social and health inequalities in the COVID-19 pandemic and beyond [Internet]. 2021 Mar [cited 2021 Aug 4]. Available from: https://www.adalovelaceinstitute.org/wp-content/uploads/2021/03/The-data-divide_25March_final-1.pdf
 50. Patel R, Reeve O, Strait A. How does structural racism impact on data and AI? [Internet]. 2021 [cited 2021 May 10]. Available from: <https://www.adalovelaceinstitute.org/blog/structural-racism-impact-data-ai/>
 51. Intensive Care National Audit & Research Centre (ICNARC). ICNARC report on COVID-19 in critical care: England, Wales and Northern Ireland [Internet]. [cited 2021 Nov 4]. Available from: <https://>

www.icnarc.org/Our-Audit/Audits/Cmp/Reports

52. Office for National Statistics (ONS). Updating ethnic contrasts in deaths involving the coronavirus (COVID-19), England: 8 December 2020 to 1 December 2021 [Internet]. 2022 Jan [cited 2022 Feb 16]. Available from: <https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/deaths/articles/updatingethniccontrastsindeathsinvolvingthecoronaviruscovid19englandandwales/8december-2020to1december2021>
53. Pareek M, Bangash MN, Pareek N, Pan D, Sze S, Minhas JS, et al. Ethnicity and COVID-19: an urgent public health research priority. *The Lancet*. 2020 May 2;395(10234):1421–2.
54. Nuffield Trust. Ethnicity coding in English health service datasets [Internet]. 2021 Jun [cited 2021 Jun 23]. Available from: https://www.nuffieldtrust.org.uk/files/2021-06/1622731816_nuffield-trust-ethnicity-coding-web.pdf
55. Health Service Journal (HSJ). NHS England agrees to collect ethnicity data seven weeks after covid vaccinations began. 2021 Jan 26 [cited 2021 Apr 4]; Available from: <https://www.hsj.co.uk/primary-care/nhs-england-agrees-to-collect-ethnicity-data-seven-weeks-after-covid-vaccinations-began/7029378.article>
56. Department for Levelling Up, Housing and Communities. Levelling up the United Kingdom [Internet]. 2022 Feb [cited 2022 Feb 15]. Available from: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1052706/Levelling_Up_WP_HRES.pdf
57. Iacobucci G. “Levelling up” plan needs more funding and a focus on health inequalities, say experts. 2022;376:0303.
58. Dezateux C, Banner N, Boomla K, Chessell M, Michie S, Naughton J, et al. Our data-driven future in healthcare: People and partnerships at the heart of health related technologies [Internet]. *The Academy of Medical Sciences*; 2018 Nov [cited 2020 Dec 1]. Available from: <https://acmedsci.ac.uk/file-download/74634438>
59. Fogel AL, Kvedar JC. Artificial intelligence powers digital medicine. *npj Digital Medicine*. 2018 Mar 14;1(1):1–4.
60. Joshi I, Morley J. Artificial Intelligence: How to get it right [Internet]. NHSX; 2019 Jan [cited 2020 Dec 1]. Available from: <https://www.nhsx.nhs.uk/ai-lab/explore-all-resources/understand-ai/artificial-intelligence-how-get-it-right/>
61. Komorowski M, Celi LA, Badawi O, Gordon AC, Faisal AA. The Artificial Intelligence Clinician learns optimal treatment strategies for sepsis in intensive care. *Nature Medicine*. 2018 Nov;24(11):1716–20.
62. Department for Business, Energy & Industrial Strategy, Department for Digital, Culture, Media & Sport, Office for Artificial Intelligence. National AI Strategy [Internet]. GOV.UK. [cited 2021 Nov 4]. Available from: <https://www.gov.uk/government/publications/national-ai-strategy/national-ai-strategy-html-version>
63. NHSX. The National Strategy for AI in Health and Social Care [Internet]. NHSX. [cited 2021 Nov 4]. Available from: <https://www.nhsx.nhs.uk/ai-lab/ai-lab-programmes/the-national-strategy-for-ai-in-health-and-social-care/>
64. Carter SM, Rogers W, Win KT, Frazer H, Richards B, Housami N. The ethical, legal and social implications of using artificial intelligence systems in breast cancer care. *The Breast*. 2020 Feb 1;49:25–32.
65. Hosny A, Parmar C, Quackenbush J, Schwartz LH, Aerts HJWL. Artificial intelligence in radiology. *Nat Rev Cancer*. 2018 Aug;18(8):500–10.
66. Niazi MKK, Parwani AV, Gurcan MN. Digital pathology and artificial intelligence. *Lancet Oncol*. 2019;20(5):e253–61.
67. Hogarty DT, Su JC, Phan K, Attia M, Hossny M, Nahavandi S, et al. Artificial Intelligence in Dermatology-Where We Are and the Way to the Future: A Review. *Am J Clin Dermatol*. 2020 Feb;21(1):41–7.
68. Ilhan B, Lin K, Guneri P, Wilder-Smith P. Improving Oral Cancer Outcomes with Imaging and Artificial Intelligence. *J Dent Res*. 2020 Mar 1;99(3):241–8.
69. Sounderajah V, Ashrafian H, Rose S, Shah NH, Ghassemi M, Golub R, et al. A quality assessment tool for artificial intelligence-centered diagnostic test accuracy studies: QUADAS-AI. *Nature medicine*. 2021;1–2.

70. The Medical Futurist. A.I. Bias In Healthcare [Internet]. The Medical Futurist. 2019 [cited 2020 Dec 1]. Available from: <https://medicalfuturist.com/a-i-bias-in-healthcare>
71. Noor P. Can we trust AI not to further embed racial bias and prejudice? BMJ [Internet]. 2020 Feb 12 [cited 2020 Dec 1];368. Available from: <https://www.bmj.com/content/368/bmj.m363>
72. Srivastava A, Thomson S. Framework Analysis: A Qualitative Methodology for Applied Policy Research. JOAAG. 2008 Nov 30;4.
73. Chen IY, Pierson E, Rose S, Joshi S, Ferryman K, Ghassemi M. Ethical machine learning in health-care. Annual Review of Biomedical Data Science. 2021;4:123–44.
74. Powell RA, Njoku C, Elangovan R, Sathyamoorthy G, Ocloo J, Thayil S, et al. Tackling racism in UK health research. BMJ. 2022;376:e065574.
75. Advanced HE. Staff statistical report 2019 [Internet]. Advance HE; 2019 Sep [cited 2020 Dec 1]. Available from: <https://www.advance-he.ac.uk/knowledge-hub/equality-higher-education-statistical-report-2019>
76. Kline R. The snowy white peaks of the NHS: a survey of discrimination in governance and leadership and the potential impact on patient care in London and England. 2014 [cited 2021 Jun 25]; Available from: <https://www.england.nhs.uk/wp-content/uploads/2014/08/edc7-0514.pdf>
77. NHS England. Workforce Race Equality Standard: 2020 Data Analysis Report for NHS Trusts and Clinical Commissioning Groups [Internet]. 2021 Feb [cited 2021 Aug 4]. Available from: <https://www.england.nhs.uk/wp-content/uploads/2021/02/Workforce-Race-Equality-Standard-2020-report.pdf>
78. Dyson S, Atkin K. Achieve equity in access to sickle cell services [Internet]. Health Service Journal. 2013 [cited 2021 Aug 10]. Available from: <https://www.hsj.co.uk/commissioning/achieve-equity-in-access-to-sickle-cell-services/5065176.article>
79. Geneviève LD, Martani A, Shaw D, Elger BS, Wangmo T. Structural racism in precision medicine: leaving no one behind. BMC Medical Ethics. 2020 Feb 19;21(1):17.
80. Strouse JJ, Lobner K, Lanzkron S, Haywood Jr C. NIH and national foundation expenditures for sickle cell disease and cystic fibrosis are associated with PubMed publications and FDA approvals. Blood. 2013;122(21):1739.
81. UK Research and Innovation (UKRI). COVID-19 and ethnicity: funding six new projects [Internet]. 2021 [cited 2021 Jun 20]. Available from: [https://www.ukri.org/our-work/tackling-the-impact-of-covid-19/understanding-coronavirus-covid-19-and-ethnicity/](https://www.ukri.org/our-work/tackling-the-impact-of-covid-19/understanding-coronavirus-covid-19-and-epidemics/covid-19-and-ethnicity/)
82. UK Research and Innovation (UKRI). Find COVID-19 research and innovation supported by UKRI [Internet]. 2021 [cited 2021 Jun 20]. Available from: <https://www.ukri.org/find-covid-19-research-and-innovation-supported-by-ukri/>
83. NIHR. NIHR and UKRI launch call for research on COVID-19 and ethnicity [Internet]. 2020 [cited 2021 Feb 18]. Available from: <https://www.nihr.ac.uk/news/nihr-and-ukri-launch-call-for-research-on-covid-19-and-ethnicity/24658>
84. NIHR. Multimillion investment in new research projects to investigate higher COVID-19 risk among certain ethnic groups [Internet]. 2020 Jul [cited 2021 Aug 4]. Available from: <https://www.nihr.ac.uk/news/multimillion-investment-in-new-research-projects-to-investigate-higher-covid-19-risk-among-certain-ethnic-groups/25333>
85. Razieh C, Zaccardi F, Islam N, Gillies CL, V. Chudasama Y, Rowlands A, et al. Ethnic minorities and COVID-19: examining whether excess risk is mediated through deprivation. European Journal of Public Health. 2021 Jul 13;31(3):630–4.
86. The Health Foundation. Artificial Intelligence and Racial and Ethnic Inequalities [Internet]. 2021 [cited 2022 Feb 15]. Available from: <https://www.health.org.uk/funding-and-partnerships/programmes/artificial-intelligence-and-racial-and-ethnic-inequalities>
87. Reddy S, Fox J, Purohit MP. Artificial intelligence-enabled healthcare delivery. Journal of the Royal Society of Medicine. 2019;112(1):22–8.
88. NHSX. The NHS AI Lab - one year on [Internet]. NHSX. [cited 2021 Nov 4]. Available from: <https://www.nhsx.nhs.uk/blogs/nhs-ai-lab-one-year/>
89. Her Majesty's Treasury. Autumn Budget and Spending Review 2021. London: The Controller of

Her Majesty's Stationery Office; 2021.

90. Saunders CL, Abel GA, El Turabi A, Ahmed F, Lyratzopoulos G. Accuracy of routinely recorded ethnic group information compared with self-reported ethnicity: evidence from the English Cancer Patient Experience survey. *BMJ open*. 2013;3(6).
91. Mathur R, Bhaskaran K, Chaturvedi N, Leon DA, vanStaa T, Grundy E, et al. Completeness and usability of ethnicity data in UK-based primary care and hospital databases. *Journal of public health*. 2014;36(4):684–92.
92. Kamulegeya LH, Okello M, Bwanika JM, Musin-guzi D, Lubega W, Rusoke D, et al. Using artificial intelligence on dermatology conditions in Uganda: A case for diversity in training data sets for machine learning. *bioRxiv*. 2019 Oct 31;826057.
93. Densley JA, Pyrooz DC. The Matrix in Context: Taking Stock of Police Gang Databases in London and Beyond. *Youth Justice*. 2019 Oct 25;20(1–2):11–30.
94. Vyas DA, Eisenstein LG, Jones DS. Hidden in Plain Sight - Reconsidering the Use of Race Correction in Clinical Algorithms. *N Engl J Med*. 2020 Aug 27;383(9):874–82.
95. Madhusoodanan J. Is a racially-biased algorithm delaying health care for one million Black people? *Nature*. 2020 Dec 16;588(7839):546–7.
96. Mehrabi N, Morstatter F, Saxena N, Lerman K, Galstyan A. A Survey on Bias and Fairness in Machine Learning. *arXiv:1908.09635 [cs]* [Internet]. 2019 Sep 17 [cited 2020 Dec 1]; Available from: <http://arxiv.org/abs/1908.09635>
97. Peterson PN, Rumsfeld JS, Liang L, Albert NM, Hernandez AF, Peterson ED, et al. A validated risk score for in-hospital mortality in patients with heart failure from the American Heart Association get with the guidelines program. *Circ Cardiovasc Qual Outcomes*. 2010 Jan;3(1):25–32.
98. McCullom R. Artificial Intelligence, Health Disparities, and Covid-19 [Internet]. *UNDARK*. 2020 [cited 2020 Dec 1]. Available from: <https://undark.org/2020/07/27/ai-medicine-racial-bias-covid-19/>
99. Obermeyer Z, Powers B, Vogeli C, Mullainathan S. Dissecting racial bias in an algorithm used to manage the health of populations. *Science*. 2019;366(6464):447–53.
100. O'Reilly-Shah VN, Gentry KR, Walters AM, Zivot J, Anderson CT, Tighe PJ. Bias and ethical considerations in machine learning and the automation of perioperative risk assessment. *Br J Anaesth*. 2020;125(6):843–6.
101. Gianfrancesco MA, Tamang S, Yazdany J, Schmajuk G. Potential Biases in Machine Learning Algorithms Using Electronic Health Record Data. *JAMA Intern Med*. 2018 Nov 1;178(11):1544–7.
102. Bécares L, Kapadia D, Nazroo J. Neglect of older ethnic minority people in UK research and policy. *BMJ*. 2020 Feb 11;368:m212.
103. Parikh RB, Teeple S, Navathe AS. Addressing Bias in Artificial Intelligence in Health Care. *JAMA*. 2019 Dec 24;322(24):2377–8.
104. Clayton V, Sanders M, Schoenwald E, Surkis L, Gibbons D. Machine Learning in Children's Services: Does it work? [Internet]. 2020 Sep [cited 2020 Dec 1]. Available from: <http://what-works-csc.org.uk/research-report/machine-learning-in-childrens-services-does-it-work/>
105. Abebe R, Hill S, Vaughan JW, Small PM, Schwartz HA. Using Search Queries to Understand Health Information Needs in Africa. *ICWSM*. 2019 Jul 6;13:3–14.
106. Chen IY, Joshi S, Ghassemi M. Treating health disparities with artificial intelligence. *Nature Medicine*. 2020 Jan;26(1):16–7.
107. Avram Robert, Tison Geoffrey, Kuhar Peter, Marcus Gregory, Pletcher Mark, Olgin Jeffrey E., et al. Predicting diabetes from photoplethysmography using deep learning. *Journal of the American College of Cardiology*. 2019 Mar 12;73(9_Supplement_2):16–16.
108. Select Committee on Artificial Intelligence. AI in the UK: ready, willing and able? [Internet]. 2018 Apr [cited 2020 Dec 14]. Available from: <https://publications.parliament.uk/pa/ld201719/ldselect/ldai/100/100.pdf>
109. Medicines & Healthcare Products Regulatory Agency. Good Machine Learning Practice for Medical Device Development: Guiding Principles [Internet]. 2021 Oct [cited 2021 Nov 1]. Available from: <https://www.gov.uk/government/publications>

110. Zafar MB, Valera I, Rodriguez MG, Gummadi KP. Fairness Constraints: Mechanisms for Fair Classification. arXiv: 150705259 [Internet]. 2017 Mar 23 [cited 2020 Dec 1]; Available from: <http://arxiv.org/abs/1507.05259>
111. Tramèr F, Atlidakis V, Geambasu R, Hsu D, Hubaux J-P, Humbert M, et al. FairTest: Discovering Unwarranted Associations in Data-Driven Applications. arXiv:151002377 [Internet]. 2016 Aug 16 [cited 2020 Dec 1]; Available from: <http://arxiv.org/abs/1510.02377>
112. Dwork C, Hardt M, Pitassi T, Reingold O, Zemel R. Fairness through awareness. In: Proceedings of the 3rd Innovations in Theoretical Computer Science Conference [Internet]. New York, NY, USA: Association for Computing Machinery; 2012 [cited 2020 Dec 1]. Available from: <https://doi.org/10.1145/2090236.2090255>
113. Beale N, Battey H, Davison AC, MacKay RS. An unethical optimization principle. Royal Society Open Science. 7(7):200462.
114. Sagawa S, Koh PW, Hashimoto TB, Liang P. Distributionally Robust Neural Networks for Group Shifts: On the Importance of Regularization for Worst-Case Generalization. arXiv:191108731 [Internet]. 2020 Apr 2 [cited 2020 Dec 1]; Available from: <http://arxiv.org/abs/1911.08731>
115. Zink A, Rose S. Fair Regression for Health Care Spending. Biometrics. 2020 Sep;76(3):973–82.
116. Williams, D.R. Ethnicity, race, and health. In: International Encyclopedia of the Social and Behavioral Sciences [Internet]. Oxford, UK: Elsevier; 2002. p. 4831–8. Available from: https://scholar.harvard.edu/files/davidwilliams/files/2002-ethnicity_race_and-williams.pdf
117. Department of Health & Social Care. Code of conduct for data-driven health and care technology [Internet]. GOV.UK; 2019 [cited 2020 Dec 1]. Available from: <https://www.gov.uk/government/publications/code-of-conduct-for-data-driven-health-and-care-technology>
118. Leidig M, Teeuw RM, Gibson AD. Data poverty: A global evaluation for 2009 to 2013 - implications for sustainable development and disaster risk reduction. International Journal of Applied Earth Observation and Geoinformation. 2016 Aug 1;50:1–9.
119. Mitchell UA, Chebli PG, Ruggiero L, Muramatsu N. The Digital Divide in Health-Related Technology Use: The Significance of Race/Ethnicity. Gerontologist. 2019 Jan 9;59(1):6–14.
120. Denniston A. Experts warn of risk of ‘data poverty’ if data underpinning AI health technologies is not representative and inclusive [Internet]. Health Data Research UK. 2020 [cited 2020 Dec 1]. Available from: <https://www.hdruk.ac.uk/news/experts-warn-of-risk-of-data-poverty-if-data-underpinning-ai-health-technologies-is-not-representative-and-inclusive/>
121. Impact of the digital divide in the age of COVID-19 | Journal of the American Medical Informatics Association | Oxford Academic [Internet]. [cited 2020 Dec 1]. Available from: <https://academic.oup.com/jamia/article/27/7/1147/5826352>
122. Laz TH, Berenson AB. Racial and Ethnic Disparities in Internet Use for Seeking Health Information Among Young Women. Journal of Health Communication [Internet]. 2013 Feb 1 [cited 2020 Dec 15]; Available from: <https://www.tandfonline.com/doi/full/10.1080/10810730.2012.707292>
123. Office for National Statistics. Internet users, UK: 2020 [Internet]. 2021 [cited 2022 Feb 15]. Available from: <https://www.ons.gov.uk/businessindustryandtrade/itandinternetindustry/bulletins/internetusers/2020>
124. Office for National Statistics. Internet access - households and individuals [Internet]. 2020 [cited 2022 Feb 15]. Available from: <https://www.ons.gov.uk/peoplepopulationandcommunity/householdcharacteristics/homeinternetandsocialmediausage/datasets/internetaccesshouseholdsandindividualsreferencetables>
125. Rogers EM. Diffusion of Innovations. 5th ed. New York, NY, USA: Free Press; 1983.
126. Agarwal R, Animesh A, Prasad K. Research Note—Social Interactions and the “Digital Divide”: Explaining Variations in Internet Use. Information Systems Research [Internet]. 2009 Mar 25 [cited 2020 Dec 15]; Available from: <https://pubsonline.informs.org/doi/>

[abs/10.1287/isre.1080.0194](https://doi.org/10.1287/isre.1080.0194)

127. Ramsetty A, Adams C. Impact of the digital divide in the age of COVID-19. *J Am Med Inform Assoc* [Internet]. 2020 Apr 28 [cited 2020 Dec 15]; Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7197532/>
128. Public Accounts Committee. Delivering gigabit-capable broadband [Internet]. 2022 Jan [cited 2022 Feb 15]. Available from: <https://committees.parliament.uk/publications/8513/documents/86177/default/>
129. Estacio EV, Whittle R, Protheroe J. The digital divide: Examining socio-demographic factors associated with health literacy, access and use of internet to seek health information. *J Health Psychol*. 2019;24(12):1668–75.
130. van Ryn M, Burgess DJ, Dovidio JF, Phelan SM, Saha S, Malat J, et al. The impact of racism on clinician cognition, behavior, and clinical decision making. *Du Bois Rev*. 2011 Apr 1;8(1):199–218.
131. Fincher C, Williams JE, MacLean V, Allison JJ, Kiefe CI, Canto J. Racial disparities in coronary heart disease: a sociological view of the medical literature on physician bias. *Ethn Dis*. 2004;14(3):360–71.
132. Whittaker M, Crawford K, Dobbe R, Fried G, Kaziunas E, Mathur V, et al. AI Now Report 2018 [Internet]. AI Now Institute; 2018 Dec [cited 2020 Dec 15]. Available from: https://ainowinstitute.org/AI_Now_2018_Report.pdf
133. Saha S, Arbelaez JJ, Cooper LA. Patient-physician relationships and racial disparities in the quality of health care. *Am J Public Health*. 2003 Oct;93(10):1713–9.
134. Aggarwal NK, Pieh MC, Dixon L, Guarnaccia P, Alegría M, Lewis-Fernández R. Clinician descriptions of communication strategies to improve treatment engagement by racial/ethnic minorities in mental health services: A systematic review. *Patient Education and Counseling*. 2016 Feb 1;99(2):198–209.
135. Jomeen J, Redshaw M. Ethnic minority women’s experience of maternity services in England. *Ethn Health*. 2013;18(3):280–96.
136. Information Commissioner’s Office. Data Protection after the end of the transition period [Internet]. [cited 2021 Jan 25]. Available from: <https://ico.org.uk/for-organisations/dp-at-the-end-of-the-transition-period/>
137. Childcare Allowance Parliamentary Inquiry, Committee. Unprecedented Injustice [Internet]. 2020 Dec [cited 2021 Nov 2]. Available from: https://www.houseofrepresentatives.nl/sites/default/files/atoms/files/verslag_pok_definitief-en-gb.docx.pdf
138. European Commission. White Paper on Artificial Intelligence: a European approach to excellence and trust [Internet]. 2020 Feb [cited 2020 Dec 1]. Available from: https://ec.europa.eu/info/publications/white-paper-artificial-intelligence-european-approach-excellence-and-trust_en
139. UK Parliament. Medicines and Medical Devices Bill 2019-21 [Internet]. [cited 2021 Apr 1]. Available from: <https://services.parliament.uk/bills/2019-21/medicinesandmedicaldevices.html>
140. European Commission. Guidelines on medical devices [Internet]. 2016 Jun [cited 2021 Jan 25]. Available from: <https://ec.europa.eu/docsroom/documents/17522/attachments/1/translations/en/renditions/native>
141. Joshi I, Cushnan D. A Buyer’s Guide to AI in Health and Care [Internet]. NHSX. 2020 [cited 2020 Dec 1]. Available from: <https://www.nhsx.nhs.uk/ai-lab/explore-all-resources/adopt-ai/a-buyers-guide-to-ai-in-health-and-care/>
142. NHSX. Regulating the AI ecosystem [Internet]. NHSX. [cited 2020 Dec 15]. Available from: <https://www.nhsx.nhs.uk/ai-lab/ai-lab-programmes/regulating-the-ai-ecosystem/>
143. NICE. NICE is gearing up for the evaluation of further digital health technologies as pilot project produces first result. 2021 Jan 8 [cited 2021 Jan 25]; Available from: <https://www.nice.org.uk/news/blog/nice-is-gearing-up-for-the-evaluation-of-further-digital-health-technologies-as-pilot-project-produces-first-result>
144. Crawford K, Dobbe R, Dryer T, Fried G, Green B, Kaziunas E, et al. AI Now 2019 Report [Internet]. AI Now Institute; 2019 [cited 2020 Dec 15] p. 100. Available from: <https://ainowinstitute.org/>

[AI Now 2019 Report.pdf](#)

145. UNESCO. Closing gender divides in digital skills through education [Internet]. 2019 [cited 2021 Jan 13]. Available from: https://unesdoc.unesco.org/in/document-Viewer.xhtml?v=2.1.196&id=p::usmarc-def_0000367416&file=/in/rest/annotationSVC/DownloadWatermarkedAttachment/attach_import_77988d38-b8bd-4cc1-b9b4-3cc16d631bf9%3F%3D367416eng.pdf&locale=en&multi=true&ark=/ark:/48223/pf0000367416/PDF/367416eng.pdf%5B%7B%22num%22%3A384%2C%22gen%22%3A0%7D%2C%7B%22name%22%3A%22XYZ%22%7D%2C%2C842%2Cnull%5D
146. Turner Lee N. Detecting racial bias in algorithms and machine learning. JICES. 2018 Aug 13;16(3):252–60.
147. Daniel Miller. Design biases in Silicon Valley are making the tech we use toxic, expert says. 2017 Oct 22 [cited 2021 Feb 13]; Available from: <https://www.abc.net.au/news/2017-10-23/toxic-tech-bias-and-algorithmic-racism-in-our-technology/9042288>
148. NHS Coordination Cente. NHS Staff Survey 2020 [Internet]. 2021 May [cited 2021 Jun 25]. Available from: <https://www.nhsstaffsurveys.com/Page/1056/Home/NHS-Staff-Survey-2020/>
149. British Medical Association (BMA). Doctors consider high levels of racism in the NHS to be a ‘debilitating’ problem, finds largest survey of racism in medicine [Internet]. [cited 2022 Feb 14]. Available from: <https://www.bma.org.uk/bma-media-centre/doctors-consider-high-levels-of-racism-in-the-nhs-to-be-a-debilitating-problem-finds-largest-survey-of-racism-in-medicine>
150. Dyer C. GMC is accused of complacency about discrimination in its work. BMJ. 2021 Jun 21;373:n1595.
151. Atewologun D, Kline R, Ochieng M. Fair to refer? Reducing disproportionality of in fitness to practice concerns reported to the GMC [Internet]. 2019 Jun [cited 2021 Jun 25]. Available from: https://www.gmc-uk.org/-/media/documents/fair-to-refer-report_pdf-79011677.pdf
152. Linton S. White doctors in London are six times more likely to be offered jobs than black doctors. BMJ. 2021 Oct 13;375:n2460.
153. Wakabayashi D. Google Chief Apologizes for A.I. Researcher’s Dismissal. The New York Times [Internet]. 2020 Dec 9 [cited 2020 Dec 15]; Available from: <https://www.nytimes.com/2020/12/09/technology/timnit-gebru-google-pichai.html>
154. New York Times. Google Researcher Says She Was Fired Over Paper Highlighting Bias in A.I. 2020 Dec 3 [cited 2020 Apr 12]; Available from: <https://www.nytimes.com/2020/12/03/technology/google-researcher-timnit-gebru.html#:~:text=the%20main%20story-,Google%20Researcher%20Says%20She%20Was%20Fired%20Over%20Paper%20Highlighting%20Bias,efforts%20to%20increase%20minority%20hiring.>
155. National Institute for Health Research. Engage patients to help shape your clinical research | NIHR [Internet]. [cited 2020 Dec 15]. Available from: <https://www.nihr.ac.uk/explore-nihr/industry/pecd.htm>
156. NIHR | Involve. UK Standards for Public Involvement Summary [Internet]. 2019 [cited 2020 Dec 15]. Available from: <https://www.invo.org.uk/posttypepublication/uk-standards-for-public-involvement-summary/>
157. Popic T, Schneider SM. An East–West comparison of healthcare evaluations in Europe: Do institutions matter? Journal of European Social Policy. 2018;28(5):517–34.
158. Ding H. Crowdsourcing, Social Media, and Intercultural Communication About Zika: Use Contextualized Research to Bridge the Digital Divide in Global Health Intervention. Journal of Technical Writing and Communication. 2020 Apr 1;50(2):141–66.
159. Ada Lovelace Institute. Participatory data stewardship: A framework for involving people in the use of data [Internet]. 2021 [cited 2021 Nov 1]. Available from: <https://www.adalovelaceinstitute.org/report/participatory-data-stewardship/>

**Imperial College
London**

**INSTITUTE OF
GLOBAL HEALTH
INNOVATION**