



THE UNIVERSITY *of* EDINBURGH

Edinburgh Research Explorer

Does "AI" stand for augmenting inequality in the era of covid-19 healthcare?

Citation for published version:

Leslie, D, Mazumder, A, Peppin, A, Wolters, MK & Hagerty, A 2021, 'Does "AI" stand for augmenting inequality in the era of covid-19 healthcare?', *BMJ (Clinical research ed.)*, vol. 372, n304.
<https://doi.org/10.1136/bmj.n304>

Digital Object Identifier (DOI):

[10.1136/bmj.n304](https://doi.org/10.1136/bmj.n304)

Link:

[Link to publication record in Edinburgh Research Explorer](#)

Document Version:

Publisher's PDF, also known as Version of record

Published In:

BMJ (Clinical research ed.)

General rights

Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy

The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.



Does “AI” stand for augmenting inequality in the era of covid-19 healthcare?

Artificial intelligence can help tackle the covid-19 pandemic, but bias and discrimination in its design and deployment risk exacerbating existing health inequity argue **David Leslie and colleagues**

Among the most damaging characteristics of the covid-19 pandemic has been its disproportionate effect on disadvantaged communities. As the outbreak has spread globally, factors such as systemic racism, marginalisation, and structural inequality have created path dependencies that have led to poor health outcomes. These social determinants of infectious disease and vulnerability to disaster have converged to affect already disadvantaged communities with higher levels of economic instability, disease exposure, infection severity, and death. Artificial intelligence (AI) technologies—quantitative models that make statistical inferences from large datasets—are an important part of the health informatics toolkit used to

fight contagious disease. AI is well known, however, to be susceptible to algorithmic biases that can entrench and augment existing inequality. Uncritically deploying AI in the fight against covid-19 thus risks amplifying the pandemic’s adverse effects on vulnerable groups, exacerbating health inequity.

Cascading risks and harms

Interacting factors of health inequality include widespread disparities in living and working conditions; differential access to, and quality of, healthcare; systemic racism; and other deep-seated patterns of discrimination. These factors create disproportionate vulnerability to disease for disadvantaged communities, as a result of overcrowding, compelled work, “weathering” (that is, the condition of premature ageing and health deterioration due to continual stress), chronic disease, and compromised immune function.¹⁻³ This greater vulnerability manifests as increased risks for exposure to covid-19, susceptibility to infection, severity of infection, and death.⁴⁻⁶ The evidence for these outcomes is rapidly increasing: mortality rates for covid-19 are more than double for those living in more deprived areas⁷; black, Asian, and minority ethnic Britons are up to twice as likely to die if they contract covid-19 in comparison with white Britons.^{8,9} When controlling for age, black men and women are more than four times more likely to die than white men and women.¹⁰

Although AI systems hold promise for improved diagnostic and prognostic decision support, epidemiological monitoring and prediction, and vaccine discovery,^{11,12} much research has reported that these systems can discriminate between, and create unequal outcomes in, different sociodemographic groups.¹³ The combination of the disproportionate impact of covid-19 on vulnerable communities and the sociotechnical determinants of algorithmic bias and discrimination might deliver a brutal triple punch. Firstly, the use of biased AI models might be disproportionately harmful to vulnerable

groups who are not properly represented in training datasets, and who are already subject to widespread health inequality. Secondly, the use of safety critical AI tools for decision assistance in high stakes clinical environments might be more harmful to members of these groups owing to their life and death impacts on them. Lastly, discriminatory AI tools might compound the disproportionate damage inflicted on disadvantaged communities by the SARS-CoV-2 virus.

Despite their promise, AI systems are uniquely positioned to exacerbate health inequalities during the covid-19 pandemic if not responsibly designed and deployed. In this article, we show how the cascading effects of inequality and discrimination manifest in design and use of an AI system (fig 1). To mitigate these effects, we call for inclusive and responsible practices that ensure fair use of medical and public AI systems in times of crisis and normalcy alike.

Embedding inequality in AI systems

Patterns of health inequality permeate AI systems when bias and discrimination become entrenched in the conception, design, and use of these systems across three planes. Discriminatory structures become ingrained in the datasets used to train systems (eg, data from underserved communities are excluded owing to their lack of access to healthcare); deficiencies arise in data representativeness (eg, under-sampling of vulnerable populations); and biases crop up across the development and implementation lifecycle (eg, failure to include clinically relevant demographic variables in the model leads to disparate performance for vulnerable subgroups).¹⁴

Health discrimination in datasets

AI technologies rely on large datasets. When biases from existing practices and institutional policies and norms affect those datasets, the algorithmic models they generate will reproduce inequities. In clinical and public health settings, biased judgment and decision making, as well as

KEY MESSAGES

- The impact of covid-19 has fallen disproportionately on disadvantaged and vulnerable communities, and the use of artificial intelligence (AI) technologies to combat the pandemic risks compounding these inequities
- AI systems can introduce or reflect bias and discrimination in three ways: in patterns of health discrimination that become entrenched in datasets, in data representativeness, and in human choices made during the design, development, and deployment of these systems
- The use of AI threatens to exacerbate the disparate effect of covid-19 on marginalised, under-represented, and vulnerable groups, particularly black, Asian, and other minoritised ethnic people, older populations, and those of lower socioeconomic status
- To mitigate the compounding effects of AI on inequalities associated with covid-19, decision makers, technology developers, and health officials must account for the potential biases and inequities at all stages of the AI process

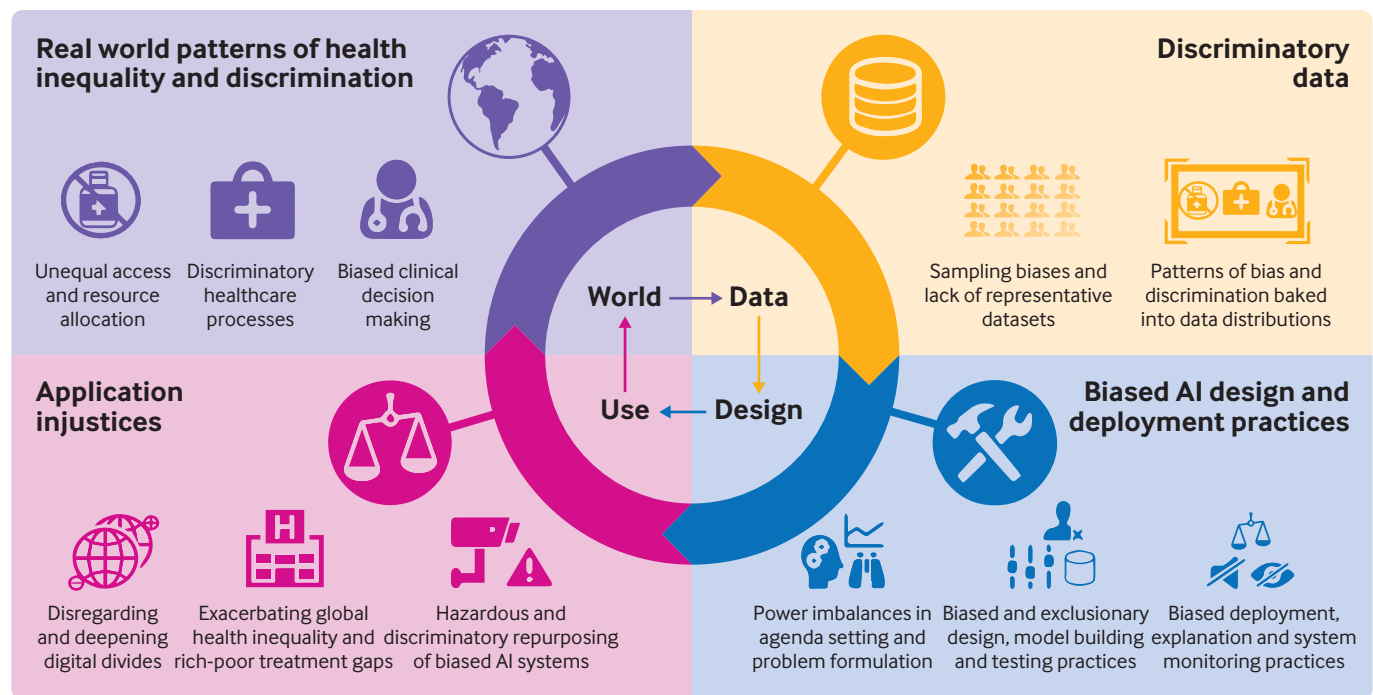


Fig 1 | Cascading effects of health inequality and discrimination manifest in the design and use of artificial intelligence (AI) systems

discriminatory healthcare processes, policies, and governance regimens can affect electronic health records, case notes, training curriculums, clinical trials, academic studies, and public health monitoring records. During clinical decision making, for example, well established biases against members of marginalised groups, such as African American^{15 16} and LGBT^{17 18} patients, can enter the clinical notes taken by healthcare workers during and after examination or treatment. If these free text notes are then used by natural language processing technologies to pick up symptom profiles or phenotypic characteristics, the real world biases that inform them will be silently tracked as well.

The datasets which are the basis of data driven AI and machine learning models thus reflect complex and historically situated practices, norms, and attitudes. This means that inferences drawn from such medical data by AI models to be used for diagnosis or prognosis might incorporate the biases of previous inequitable practices, and the use of models trained on these datasets could reinforce or amplify discriminatory structures. Risks of this kind of discrimination creep pose special challenges during the covid-19 pandemic. For instance, hospital systems are already using natural language processing technologies to extract diagnostic information from radiology and pathology reports and clinical notes.¹⁹⁻²¹ As these capacities

are shifted on to tasks for identifying clinically significant symptoms of SARS-CoV-2 infection,²² hazards of embedding inequality will also increase. Where human biases are recorded in clinical notes, these discriminatory patterns will probably infiltrate the natural language processing supported AI models that draw on them. Similarly, if such models are also trained using unrepresentative or incomplete data from electronic health records that reflect disparities in healthcare access and quality, the resulting AI systems will probably reflect, repeat, and compound pre-existing structural discrimination.

Data representativeness

The datasets used to train, test, and validate AI models are too often insufficiently representative of the general public. For instance, datasets composed of electronic health records, genome databases, and biobanks often undersample those who have irregular or limited access to the healthcare system, such as minoritised ethnicities, immigrants, and socioeconomically disadvantaged groups.²³⁻²⁵ The increased use of digital technologies, like smartphones, for health monitoring (eg, through symptom tracking apps) also creates potential for biased datasets. In the UK, more than 20% of the population aged 15 or older lack essential digital skills and up to 10% of some population subgroups do not own smartphones.²⁶ Datasets from pervasive sensing, mobile technologies,

and social media can under-represent or exclude those without digital access. Whether originating from medical data research facilities or everyday technologies, biased datasets that are linked—such as in biomedical applications that combine pervasive sensing data with electronic health records²⁷—will only exacerbate unrepresentativeness.

The prevalence and incidence of diseases and their risk factors often vary by population group. If datasets do not adequately cover populations at particular risk, trained prediction models that are used in clinical AI decision support might have lower sensitivity (true positive rates) for these populations and systematically underdetect the target condition.²⁸ Every time a prediction model which has been tailored to the members of a dominant group is applied in a “one-size-fits-all” manner to a disadvantaged group, the model might yield suboptimal results and be harmful for disadvantaged people.²⁹

The data flows emerging from the covid-19 outbreak present a set of problems that could jeopardise attempts to attain balanced and representative datasets. Tendencies to produce health data silos create a channelling effect where usable electronic health records from patients who have contracted covid-19 overly reflect subpopulations who non-randomly have access to particular hospitals in certain, well-off neighbourhoods. This problem arises because resources needed

to ensure satisfactory dataset quality and integrity might be limited to digitally mature hospitals that disproportionately serve a privileged segment of a population to the exclusion of others. Where data from electronic health records resulting from these contexts contribute to the composition of AI training data, problems surrounding discriminatory effects arise. If such dataset imbalances are not dealt with, and if thorough analyses are not performed to determine the limitations of models trained on these data, they will probably not be sufficiently generalisable and transportable. The models will simply underfit members of vulnerable groups whose data were under-represented in the training set, and will perform less well for them.

Biases in the choices made for AI design and use

Lack of representativeness and patterns of discrimination are not the only sources of bias in AI systems. Legacies of institutional racism and the implicit—often unconscious—biases of AI developers and users might influence choices made in the design and deployment of AI, leading to the integration of discrimination and prejudice into both innovation processes and products.³⁰

At the most basic level, the power to undertake health related AI innovation projects is vested with differential privileges and interests that might exacerbate existing health inequities. The sociodemographic composition (that is, class, race, sex, age) of those who set research and innovation agendas often does not reflect that of the communities most affected by the resulting projects.^{31 32} This disparity lays the foundation for unequal outcomes from AI innovation. Decisions in setting the agenda include which clinical questions should be reformulated as statistical problems, and which kinds of data centric technologies should be developed. During the covid-19 pandemic this is of particular concern, as the urgency to find solutions and the institutional hierarchies in decision making are at cross purposes with consensus building mechanisms and with the diligence needed to ensure oversight and involvement of the community in setting the agenda.

Once an AI innovation project is under way, choices must be made about how to define target variables and their quantifiable proxies. At this stage of problem formulation, any latent biases of designers, developers, and researchers

might allow structural health inequalities and injustices to be introduced in the model via label determinations (that is, choices made in the specification of target variables) that fail to capture underlying complexities of the social contexts of discrimination.³³ This bias was seen in a recent study, which showed that the label choice made by the producers of a commercial insurance risk prediction tool discriminated against millions of African Americans, whose level of chronic illness was systematically mismeasured because healthcare costs were used as a proxy for ill health.³⁴

At the stages of extraction, collection, and wrangling of data, measurement errors and faulty data consolidation practices could lead to additional discrimination against disadvantaged communities. For example, if data on skin colour are not collected together with pulse oximetry data, it is almost impossible for AI models to correct for the effect of skin tone on oximetry readings.³⁵

Similar discriminatory patterns can pass into design-time processes at the data preprocessing and model construction stages. The decisions made about inclusion of personal data such as age, ethnicity, sex, or socioeconomic status, will affect the way the model performs for vulnerable subgroups. When features such as ethnicity are integrated into models without careful consideration of potential confounders, those models risk identifying as biological, characteristics that have socioeconomic or environmental origins. As a result, structural racism might be integrated into the automated tools that support clinical practice. A well known example is the flawed “race correction” mechanism in commercial spirometer software.³⁶

Lastly, AI systems might introduce unequal health outcomes during testing, implementation, and continuing use. For instance, in the implementation phase, clinicians who over-rely on AI decision support systems might take their recommendations at face value, even when these models might be faulty. On the other hand, clinicians who distrust AI decision support systems might discount their recommendations, even if they offer corrections to discrimination. For example, when a decision support model provides pulse oximetry values that have been correctly adjusted for skin tone, the results might conflict with a clinician’s own preconceptions about the validity of raw oximetry data. These results might lead the clinician to dismiss the model’s

recommendation based upon their own potentially biased professional judgment.

Equity under pressure

During the covid-19 pandemic, demand for rapid response technological interventions might hinder responsible AI design and use.^{37 38} In a living systematic review of over 100 covid-19 prediction models for diagnosis and prognosis, Wynants et al have found that owing to the pressure of rushed research, the proposed systems were at high risk of statistical bias, poorly reported, and overoptimistic. Up to this point, the authors have recommended that none of the models be used in medical practice.³⁹

To make matters worse, some hospitals are hurriedly repurposing AI systems (which were developed for use, and trained on data, in situations other than the pandemic) for sensitive tasks like predicting the deterioration of infected patients who might need intensive care or mechanical ventilation.⁴⁰ These models run considerable risks of insufficient validation, inconsistent reliability, and poor generalisability due to unrepresentative samples and a mismatch between the population represented in the training data and those who are disparately affected by the outbreak.⁴¹

AI systems are similarly being swiftly repurposed in non-clinical domains, with tangible consequences for public health. In an attempt to curb the spread of covid-19, the United States prison system, for example, has used an algorithmic tool developed for measuring the risk of recidivism to determine which inmates will be released to home confinement. This tool has been shown to exhibit racial biases, and so repurposing it for the management of health risks makes black inmates more likely to remain confined and, consequently, subjected to increased exposure to covid-19 infection and disease related death.⁴² At the beginning of the second US wave of the pandemic in June, such repurposing took place while the five largest known clusters of covid-19 in the US were at correctional institutions,⁴³ and against a backdrop of mass incarceration based on historic and systemic racism.⁴⁴

Conclusion

AI could make a valuable contribution to clinical, research, and public health tools in the fight against covid-19. The widespread sense of urgency to innovate, however, should be tempered by the need

to consider existing health inequalities, disproportionate pandemic vulnerability, sociotechnical determinants of algorithmic discrimination, and the serious consequences of clinical and epidemiological AI applications. Without this consideration, patterns of systemic health inequity and bias will enter AI systems dedicated to tackling the pandemic, amplifying inequality, and subjecting disadvantaged communities to increasingly disproportionate harm.

With these dynamics in mind, it is essential to think not just of risks but also of remedies (fig 2). On the latter view, developing and deploying AI systems safely and responsibly in medicine and public health to combat covid-19 requires the following:

In technological development—Incorporation of diligent, deliberate, and end-to-end bias detection and mitigation protocols. Clinical expertise, inclusive community involvement, interdisciplinary knowledge, and ethical reflexivity must be embedded in AI project teams and innovation processes to help identify and remedy any discriminatory factors. Similarly, awareness of the social determinants of disparate vulnerability to covid-19 must be integrated into data gathering practices so that data on socioeconomic status can be combined with other race, ethnicity, and sensitive data to allow for scrutiny of subgroup differences in processing results.^{45 46}

In medical and public health practices— Interpretation of the outputs of AI systems with careful consideration of potential algorithmic biases, and with understanding of the strengths and limitations of statistical reasoning and generalisation. Stakeholders in healthcare must use tools available in public health, epidemiology, evidence based medicine, and applied ethics to evaluate whether specific uses of the quantitative modelling of health data are appropriate, responsible, equitable, and safe.

In policy making—Benefits, limitations, and unintended consequences of AI systems must be considered carefully when setting innovation agendas, without discrimination. Policies will need to be formulated in processes that are open to all stakeholders and prioritise individual and community consent in determining the purpose and path of AI innovation projects.

Finally, as a society, we must deal effectively with systemic racism, wealth disparities, and other structural inequities, which are the root causes of discrimination and health inequalities and evident in algorithmic bias. If we do so, AI can help counter exacerbations of inequalities, instead of contributing to them.

Contributors and sources: The authors have academic and policy backgrounds from social science, computer science, statistics, and linguistics, with expertise in technology, ethics, social impacts, and human rights. DL is the ethics theme lead and ethics fellow at the Alan Turing Institute. AM is AI and justice and human rights theme lead at the

Alan Turing Institute. AP is a researcher of AI and society at the Ada Lovelace Institute, an independent research institute looking at data and AI. MKW is a reader in design informatics, School of Informatics, and academic associate of the School of Philosophy, Psychology, and Language Sciences at the University of Edinburgh, and a faculty fellow at the Alan Turing Institute. AH is a research associate at the Centre for the Study of Existential Risk and associate fellow at the Leverhulme Centre for the Future of Intelligence, University of Cambridge. All authors contributed to discussing, writing, and editing the article. DL is the guarantor.

Competing interests: We have read and understood BMJ policy on declaration of interests and declare that we have no competing interests.

Provenance and peer review: Commissioned; externally peer reviewed.

This collection of articles was proposed by the WHO Department of Digital Health and Innovation and commissioned by *The BMJ*. *The BMJ* retained full editorial control over external peer review, editing, and publication of these articles. Open access fees were funded by WHO.

David Leslie, ethics theme lead and ethics fellow¹
Anjali Mazumder, AI and justice and human rights theme lead¹

Aidan Peppin, researcher of AI and society²

Maria K Wolters, reader in design informatics³

Alexa Hagerty, research associate⁴

¹Alan Turing Institute, London, UK

²Ada Lovelace Institute, London, UK

³School of Informatics, University of Edinburgh, UK

⁴Centre for the Study of Existential Risk and Leverhulme Centre for the Future of Intelligence, University of Cambridge, Cambridge, UK

Correspondence to: D Leslie
dleslie@turing.ac.uk



OPEN ACCESS

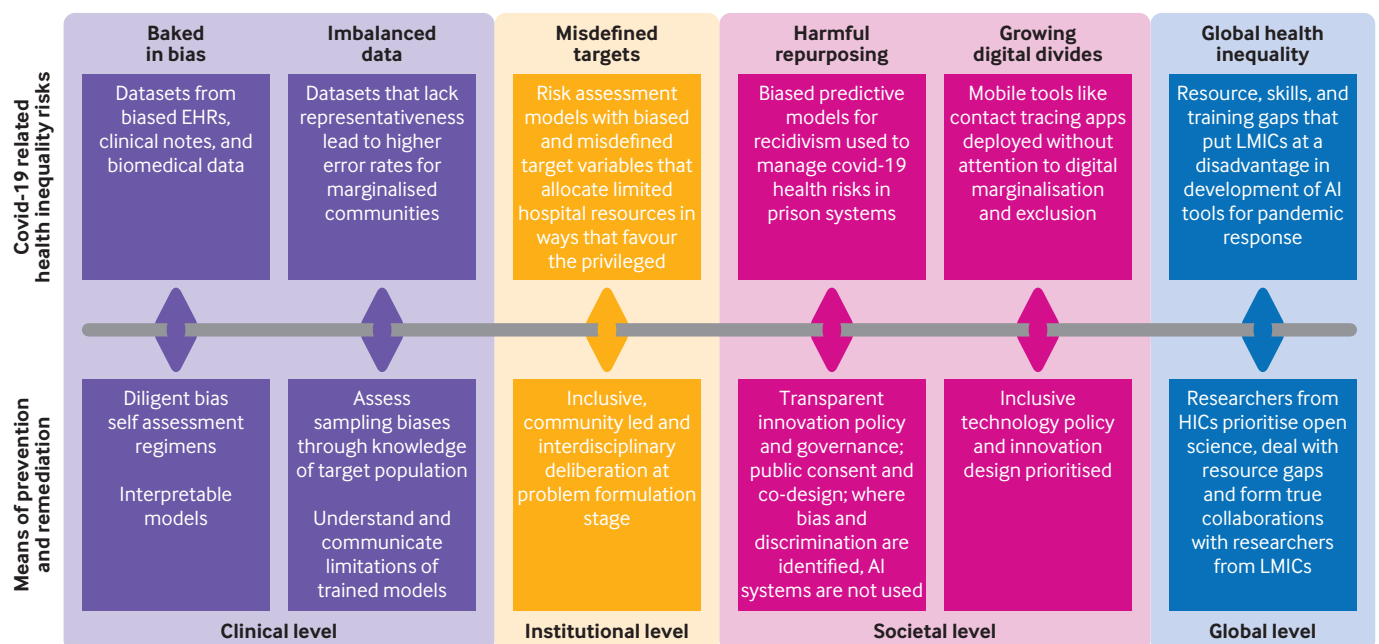


Fig 2 | Risks of, and remedies for, developing and deploying artificial intelligence (AI) systems safely. EHRs=electronic health records; HICs=high income countries; LMICs=low and middle income countries

This is an Open Access article distributed under the terms of the Creative Commons Attribution IGO License (<https://creativecommons.org/licenses/by-nc/3.0/igo/>), which permits use, distribution, and reproduction for non-commercial purposes in any medium, provided the original work is properly cited.



- 1 Cockerham WC, Hamby BW, Oates GR. The social determinants of chronic disease. *Am J Prev Med* 2017;52(1S1):S5-12. doi:10.1016/j.amepre.2016.09.010
- 2 Quinn SC, Kumar S. Health inequalities and infectious disease epidemics: a challenge for global health security. *Biosecur Bioterror* 2014;12:263-73. doi:10.1089/bsp.2014.0032
- 3 Geronimus AT, Hicken M, Keene D, Bound J. "Weathering" and age patterns of allostatic load scores among blacks and whites in the United States. *Am J Public Health* 2006;96:826-33. doi:10.2105/AJPH.2004.060749
- 4 Bolin B, Kurtz LC. Race, class, ethnicity, and disaster vulnerability. In: Rodríguez H, Donner W, Trainor JE, eds. *Handbook of disaster research 2018*. Springer, 2018:181-203. doi:10.1007/978-3-319-63254-4_10
- 5 Bavel JJV, Baicker K, Boggio PS, et al. Using social and behavioural science to support COVID-19 pandemic response. *Nat Hum Behav* 2020;4:460-71. doi:10.1038/s41562-020-0884-z
- 6 Abrams EM, Szeffler SJ. COVID-19 and the impact of social determinants of health. *Lancet Respir Med* 2020;8:659-61. doi:10.1016/S2213-2600(20)30234-4
- 7 Office for National Statistics. Deaths involving COVID-19 by local area and socioeconomic deprivation. 2020 August 28. <https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/deaths/bulletins/deathsinvolvedwithcovid19bylocalareasanddeprivation/deathsoccurringbetween1marchand31july2020>
- 8 Public Health England. Disparities in the risk and outcomes of covid-19. Jun 2020. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/892085/disparities_review.pdf
- 9 Public Health England. Beyond the data: Understanding the impact of COVID-19 on BAME groups. Jun 2020. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/892376/COVID_stakeholder_engagement_synthesis_beyond_the_data.pdf
- 10 Office of National Statistics. Coronavirus (COVID-19) related deaths by ethnic group, England and Wales: 2 March 2020 to 10 April 2020. <https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/deaths/articles/coronavirus-relateddeathsbyethnicgroupenglandandwales/2march2020to10april2020>
- 11 Islam MN, Inan TT, Rafi S, Akter SS, Sarker IH, Islam AK. A survey on the use of AI and ML for fighting the COVID-19 pandemic. arXiv:2008.07449. 2020. [Preprint.] <https://arxiv.org/pdf/2008.07449.pdf>
- 12 Syeda HB, Syed M, Sexton KW, et al. Role of machine learning techniques to tackle COVID-19 crisis: a systematic review. *JMIR Med Inform* 2021;9:e23811. doi:10.2196/23811
- 13 Ghassemi M, Naumann T, Schulam P, Beam AL, Chen IY, Ranganath R. A review of challenges and opportunities in machine learning for health. *AMIA Jt Summits Transl Sci Proc* 2020;2020:191-200.
- 14 Bickler PE, Feiner JR, Severinghaus JW. Effects of skin pigmentation on pulse oximeter accuracy at low saturation. *Anesthesiology* 2005;102:715-9. doi:10.1097/0000542-200504000-00004
- 15 Smedley BD, Stith AY, Nelson AR. Unequal TREATMENT. confronting racial and ethnic disparities in health care. National Academic Press, 2003;100. https://cdn.ymaws.com/www.aptrweb.org/resource/collection/0B76BF46-69A2-4560-BF7C-98682BBA60C1/Unequal%20Treatment_IOM%20Report.pdf
- 16 van Ryn M, Burke J. The effect of patient race and socio-economic status on physicians' perceptions of patients. *Soc Sci Med* 2000;50:813-28. doi:10.1016/S0277-9536(99)00338-X
- 17 Fallin-Bennett K. Implicit bias against sexual minorities in medicine: cycles of professional influence and the role of the hidden curriculum. *Acad Med* 2015;90:549-52. doi:10.1097/ACM.0000000000000662
- 18 Burke SE, Dovidio JF, Przeworski JM, et al. Do contact and empathy mitigate bias against gay and lesbian people among heterosexual medical students? A report from the medical student CHANGE study. *Acad Med* 2015;90:645-51. doi:10.1097/ACM.0000000000000661
- 19 Afzal N, Mallipeddi VP, Sohn S, et al. Natural language processing of clinical notes for identification of critical limb ischemia. *Int J Med Inform* 2018;111:83-9. doi:10.1016/j.ijmedinf.2017.12.024
- 20 Lopez-Jimenez F, Attia Z, Arruda-Olson AM, et al. Artificial intelligence in cardiology: present and future. *Mayo Clin Proc* 2020;95:1015-39. doi:10.1016/j.mayocp.2020.01.038
- 21 Pons E, Braun LM, Hunink MG, Kors JA. Natural language processing in radiology: a systematic review. *Radiology* 2016;279:329-43. doi:10.1148/radiol.16142770
- 22 Wang J, Anh H, Manion F, Rouhizadeh M, Zhang Y. COVID-19 SignSym—a fast adaptation of general clinical NLP tools to identify and normalize COVID-19 signs and symptoms to OMOP common data model. ArXiv. 2020 Jul 13. [Preprint.] <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7480086/>
- 23 Gianfrancesco MA, Tamang S, Yazdany J, Schmajuk G. Potential biases in machine learning algorithms using electronic health record data. *JAMA Intern Med* 2018;178:1544-7. doi:10.1001/jamainternmed.2018.3763
- 24 Popejoy AB, Ritter DL, Crooks K, et al. Clinical Genome Resource (ClinGen) Ancestry and Diversity Working Group (ADWG). The clinical imperative for inclusivity: race, ethnicity, and ancestry (REA) in genomics. *Hum Mutat* 2018;39:1713-20. doi:10.1002/humu.23644
- 25 Manrai AK, Funke BH, Rehm HL, et al. Genetic misdiagnoses and the potential for health disparities. *N Engl J Med* 2016;375:655-65. doi:10.1056/NEJMs1507092
- 26 Lloyds Bank. UK Consumer Digital Index 2020. https://www.lloydsbank.com/assets/media/pdfs/banking_with_us/whats-happening/lb-consumer-digital-index-2020-report.pdf
- 27 van der Schaar M, Humphrey J, Alaa A, et al. How artificial intelligence and machine learning can help healthcare systems respond to COVID-19 (2020). <https://www.vanderschaar-lab.com/NewWebsite/covid-19/paper.pdf>
- 28 Rajkomar A, Hardt M, Howell MD, Corrado G, Chin MH. Ensuring fairness in machine learning to advance health equity. *Ann Intern Med* 2018;169:866-72. doi:10.7326/M18-1990
- 29 Suresh H, Gutttag JV. A framework for understanding unintended consequences of machine learning. arXiv:1901.10002. 2020. [Preprint.] <https://arxiv.org/pdf/1901.10002.pdf>
- 30 Benjamin R. *Race after technology: abolitionist tools for the new jim code*. John Wiley & Sons, 2019.
- 31 West SM, Whittaker M, Crawford K. Discriminating systems: gender, race and power in AI. Al Now Institute, 2019. <https://ainowinstitute.org/discriminatingystems.html>
- 32 Nkonde M. Automated anti-blackness: facial recognition in Brooklyn, New York. *Harvard Kennedy School Journal of African American Policy*. 2019;20:30-6. <https://pacscenter.stanford.edu/wp-content/uploads/2020/12/mutalainkonde.pdf>
- 33 Passi S, Barocas S. Problem formulation and fairness. In: Proceedings of the Conference on Fairness, Accountability, and Transparency. 2019:39-48.
- 34 Obermeyer Z, Powers B, Vogeli C, Mullainathan S. Dissecting racial bias in an algorithm used to manage the health of populations. *Science* 2019;366:447-53. doi:10.1126/science.aax2342
- 35 Bickler PE, Feiner JR, Severinghaus JW. Effects of skin pigmentation on pulse oximeter accuracy at low saturation. *Anesthesiology* 2005;102:715-9. doi:10.1097/0000542-200504000-00004
- 36 Braun L. Race, ethnicity and lung function: a brief history. *Curr J Respir Ther* 2015;5:199-101.
- 37 Rössli E, Rice B, Hernandez-Boussard T. Bias at warp speed: how AI may contribute to the disparities gap in the time of COVID-19. *J Am Med Inform Assoc* 2021;28:190-2. doi:10.1093/jamia/ocaa210
- 38 Leslie D. *Tackling COVID-19 through responsible AI innovation: five steps in the right direction*. Harvard Data Science Review, 2020.
- 39 Wynants L, Van Calster B, Collins GS, et al. Prediction models for diagnosis and prognosis of covid-19 infection: systematic review and critical appraisal. *BMJ* 2020;369:m1328. doi:10.1136/bmj.m1328
- 40 Robbins R. 'Human experts will make the call': Stanford launches an accelerated test of AI to help care for Covid-19 patients. STAT News 2020. <https://www.statnews.com/2020/04/01/stanford-artificial-intelligence-coronavirus/>
- 41 Singh K, Valley TS, Tang S, et al. Validating a widely implemented deterioration index model among hospitalized COVID-19 patients. *medRxiv*. 2020. <https://www.medrxiv.org/content/10.1101/2020.04.24.20079012v1>
- 42 Fogliato R, Xiang A, Chouldechova A. Why PATTERN should not be used: the perils of using algorithmic risk assessment tools during COVID-19. Partnership on AI. 2020. <https://www.partnershiponai.org/why-pattern-should-not-be-used-the-perils-of-using-algorithmic-risk-assessment-tools-during-covid-19/>
- 43 Williams T, Seline L, Griesbach R. Coronavirus cases rise sharply in prisons even as they plateau nationwide. *New York Times* 2020 Jun 16. <https://www.nytimes.com/2020/06/16/us/coronavirus-inmates-prisons-jails.html>
- 44 Hinton E. *From the war on poverty to the war on crime: the making of mass incarceration in America*. Harvard University Press, 2016. doi:10.4159/9780674969223
- 45 Chowkwanyun M, Reed AL Jr. Racial health disparities and Covid-19—caution and context. *N Engl J Med* 2020;383:201-3. doi:10.1056/NEJMp2012910
- 46 Bailey ZD, Krieger N, Agénor M, Graves J, Linos N, Bassett MT. Structural racism and health inequities in the USA: evidence and interventions. *Lancet* 2017;389:1453-63. doi:10.1016/S0140-6736(17)30569-X

Cite this as: *BMJ* 2021;372:n304
<http://dx.doi.org/10.1136/bmj.n304>