PATIENT-CENTERED OUTCOMES RESEARCH INSTITUTE

WWW.PCORI.ORG | INFO@PCORI.ORG | FOLLOW US @PCORI

ARTIFICIAL INTELLIGENCE IN THE COVID-19 RESPONSE

VOLUME 2

Strategies to Improve the Impact of AI on Health Equity

Artificial Intelligence in the COVID-19 Response

Volume 2

Strategies to Improve the Impact of AI on Health Equity

Sean Mann, Lawrence Baker, Federico Girosi, Osonde Osoba, and Carl T. Berdahl

RAND Health

February 2023

Prepared for Patient-Centered Outcomes Research Institute (PCORI)

Suggested citation: Sean Mann, Lawrence Baker, Federico Girosi, Osonde Osoba, and Carl T. Berdahl. Artificial Intelligence in Health Care: Strategies to Improve the Impact of AI on Health Equity. Patient-Centered Outcomes Research Institute; February 2023. Prepared by RAND under Contract No. IDIQ-TO#22-RAND-ENG-AOSEPP-04-01-2020.

Table of Contents

Figures **6**

Tables **6**

Summary **7**

Problem Statement 7

Approach 7

Key Findings 7

1. Introduction **9**

Objectives 10

Organization of the Report 11

2. Methods **12**

Stakeholder Interviews 12

Document Sources, Searches, and Screening 15

Literature Searches 15

Document Screening 16

Full Text Review of Documents 16

Identifying Equity-Related Issues and Strategies 17

3. Issues Raised by the Use of AI in Health Care **18**

Background Context 19

Biased or Nonrepresentative Developers 19

Diminished Accountability 20

AI-Enabled Discrimination 20

Data Characteristics 21

Limited or Poor Information on Population Characteristics 21

Unrepresentative Data or Small Sample Sizes 22

Bias Ingrained in Data 22

Inclusion of Sensitive Variables 23

Exclusion of Sensitive Variables 24

Limited Reporting of Information on Protected Groups 25

Model Design 25

Algorithms Are Not Interpretable 25

Optimizing Algorithm Accuracy and Fairness May Conflict 26

Ambiguity in and Conflict Among Conceptions of Equity ... 26
Deployment Practices ... **26**
Proprietary Algorithms or Data Unavailable for Evaluation ... 27
Overreliance on AI Apps ... 27
Underreliance on AI Apps ... 27
Repurposing Existing AI Apps Outside Original Scope ... 28
Rushed Development or Implementation ... 28
Unequal Access to AI ... 29
Summary ... **29**
4. Strategies Proposed to Address AI and Equity Issues ... ***30***
Background Context ... **32**
Foster Diversity ... 32
Train Developers and Users ... 33
Engage the Broader Community ... 33
Improve Governance ... 35
Data Characteristics ... **36**
Improve Diversity, Quality, or Quantity of Data ... 36
Exclude Sensitive Variables to Correct for Bias ... 38
Include Sensitive Variables to Correct for Bias ... 39
Model Design ... **40**
Enforce Fairness Goals ... 40
Improve Interpretability or Explainability of Algorithm ... 41
Evaluate Disparities in Model Performance ... 42
Use Equity-Focused Checklists, Guidelines, and Similar Tools ... 43
Deployment Practices ... **44**
Increase Model Reporting and Transparency ... 44
Seek or Provide Restitution for Those Negatively Affected by AI ... 45
Avoid or Reduce Use of AI ... 45
Provide Resources to Those With Less Access to AI ... 45
Use of AI to Address Preexisting Inequities ... **46**
Develop and Field AI That Is Less Biased Than Standard Decision-Making ... 47
Use AI to Examine Disparities ... 47
Use AI to Identify People Facing Disparities ... 48
Use AI to Reduce Disparities ... 48
Summary ... **49**
5. Linking Issues and Strategies ... ***50***
Summary ... **54**
6. Discussion ... ***55***

The Literature Focuses on a Small Set of Issues 55
Strategies Are Multipurpose 56
Stakeholders Can Focus on a Small Set of Strategies 56
Study Limitations 57
Potential Future Areas of Research 58
Conclusion 59
Appendix A. Documents Proposing Strategies to Address Equity-Related Issues Raised by the Use of AI in Health Care 60
Appendix B. Stakeholder Interview Guide 70
Study Description 70
Informed Consent Protocol 71
Interview Guide 71
Appendix C. Literature Searches and Screening 74
Literature Search 76
Document Screening 77
Detailed Information on Search Queries 78
Abbreviations 98
References 99

Figures

Figure 5.1. Mapping the Connections Between AI Health Equity Issues and Strategies Proposed to Address Them 53

Figure C.1. PRISMA Literature Flow Diagram 75

Tables

Table 2.1. Summary of Stakeholder Feedback on Study Approach and Health Equity in AI 14

Table 3.1. Issues Related to AI and Equity From the Literature 18

Table 4.1. Strategies to Address AI and Equity Issues From the Literature 30

Table 4.2. Strategies That Use AI to Address Preexisting Health Disparities 31

Table 5.1. The Most Common Strategies Mentioned in the Literature for Each Health Equity Issue 50

Table A.1. Documents and the Strategies They Propose to Address Equity-Related Issues Raised by the Use of AI in Health Care 60

Table C.1. Document Types and Search Results, by Database and Subject 76

Summary

This study was prepared for the Patient-Centered Outcomes Research Institute (PCORI) Emerging Technology and Therapeutics Report series. This study, on strategies to improve the impact of artificial intelligence (AI) on health equity, is the second of 2 reports examining AI in health care.

Problem Statement

The use of AI in clinical care, public health, and health system administration has expanded rapidly in recent years. AI applications in health care have the potential to improve accuracy, personalization, and fairness, but may also introduce new biases or perpetuate existing inequalities as a result of data limitations and other challenges. At the same time, the COVID-19 pandemic has underscored the persistence of health disparities in the United States and abroad, with disadvantaged populations facing high rates of infection, hospitalization, and death. The pandemic has provided further evidence that AI developers, users, and policy makers will increasingly need strategies to mitigate negative impacts and enhance positive impacts of AI on health equity.

Approach

We conducted a scoping review to identify strategies used to address equity issues posed by AI in health care. We conducted interviews with stakeholders to inform our research questions and guide our study design. We searched, screened, and reviewed a wide range of academic and gray literature as part of our study.

Key Findings

- We identified 18 equity-related issues that are raised by the use of AI in health care. Concerns about unrepresentative and biased data are most commonly mentioned in the literature. Other prominent issues include balancing any potential trade-offs between model accuracy and fairness, biased or nonrepresentative AI developers, and limited information on population characteristics.
- We identified 15 strategies proposed to address these equity-related issues posed by the use of AI in health care. The most commonly proposed strategies in the literature were evaluating disparities in model performance, improving data inputs, engaging the broader community in AI development, and improving governance. Most issues are complex and are likely best addressed through multiple complementary strategies.

- In some cases, AI can be used to address long-standing problems of health inequity, whose causes are rooted in societal issues that are independent of the use of AI in health care. This includes implementing AI models that are less biased than current decision-making practices, using AI to better understand the extent and cause of health disparities, using AI to target health service delivery to those who need it most, and fielding AI that directly improves marginalized communities' access to care.
- Efforts to improve the impact of AI on health equity could benefit from further research on which strategies have proved the most effective in real-world settings as well as on best practices for strategy implementation.

1. Introduction

The use of artificial intelligence (AI) applications in health has the potential to improve patient outcomes, but modern AI can also, in conjunction with social and historical inequities, exacerbate disparities in health. [1-9] As AI becomes more integrated into decision-making, it is imperative that stakeholders take an active role in examining and intervening to achieve health equity. Equity considerations have become particularly pertinent given the demonstrably biased performance of some AI applications in health[10] and the unequal risks faced by historically disadvantaged groups during the COVID-19 pandemic. [11,12] Disparities in health care access, treatment, and outcomes across race, ethnicity, gender, sexual orientation, disability status, and other characteristics persist throughout the United States and around the world. [13-15] In this context, AI developers, users, and policy makers will increasingly need strategies to mitigate negative impacts and enhance positive impacts of AI on health equity.

Many different sources of potential bias exist in AI algorithms. The use of inappropriate proxy variables, biased data, small sample sizes, uninterpretable model features, and unrepresentative study populations can all lead to unequal performance across different groups. [1-7,16] Deployment practices, user training, regulatory processes, and continuous evaluation and monitoring mechanisms also influence how AI applications impact health disparities. [17] Box 1.1 illustrates a real-world example of health equity impacts caused by AI algorithms. Many strategies have been proposed to ensure that AI applications maintain equity, although these efforts remain in their early stages. [18-20] AI may even have the potential to reduce long-standing inequities if it is applied strategically to detect and mitigate problems affecting vulnerable groups. [2,3,21-24]

Although there is a broad literature on health equity in AI, there is no comprehensive guidance about which mitigation strategies are best suited to address certain issues. One objective of this report is to provide an overview of which strategies have been proposed to address specific health equity issues to support future efforts to produce guidance that could fill this gap.

Box 1.1. A Real-World Example of Health Equity Impacts Caused by AI Algorithms

A recent study by Obermeyer et al examined an algorithm deployed by health systems and payers to allocate resources to patients with greater health care needs. [10] They discovered that the algorithm used spending as a proxy for health, without considering differential access by race. Because Black patients have historically suffered from lack of access to resources and therefore experienced lower levels of health care utilization, the algorithm mistook lower health spending levels for better health. As a consequence, for the same level of need, fewer resources were allocated to Black patients.

Our research effort focused on 2 related but separate topics: (1) the use of AI applications in the COVID-19 response, and (2) strategies to mitigate negative impacts and enhance positive impacts of AI on health equity. The first of these topics is covered in a companion report. This report examines the second of these topics. It focuses on AI and health equity considerations relevant to all health conditions and not just to COVID-19.

The Patient-Centered Outcomes Research Institute (PCORI) and the RAND study team collaborated in the development of the study topics and research approach. We also consulted a range of health care stakeholders early in the study process to understand their perspectives and to ensure that our research questions examined the aspects of AI and its potential impacts on health equity that they thought were most important for further study. This report is intended to inform the efforts of patient advocates, clinicians, policy makers, software developers, and others involved in supporting the responsible use of AI in health care.

This report draws from the content of a previously published journal article on strategies to improve the impact of AI on health equity. [25] This report expands upon that article by including full descriptions of each of the equity-related issues raised by the use of AI in health care as well as the strategies proposed to address those issues that we identified in our review.

Objectives

The goal of this report is to identify strategies to improve the impact of AI on health equity. To accomplish this goal, we reviewed a wide range of academic and gray literature to answer the following 2 research questions:

1) What equity-related issues are raised by the use of AI in health care?
2) What strategies are proposed to address these issues and thus improve the impact of AI on health equity?

We consider equity concerns around gender, race/ethnicity, disability, sexual orientation, and any other relevant demographic characteristics that might be associated with health disparities. Many of the examples we provide in this report concern the social construct of race because that is a major focus of the literature we examined in our review. [26]

Organization of the Report

The report is organized as follows. Chapter 2 details the methods used to elicit stakeholder input, review published literature, and identify issues and strategies surrounding the impact of AI on health equity. Chapter 3 describes the issues raised in the literature we reviewed, followed by Chapter 4, which describes the strategies proposed to address them. Chapter 5 presents our findings on the connections made between specific strategies and issues in the literature. Chapter 6 concludes with a discussion of our results and directions for future research.

Appendix A contains a list of the documents we reviewed that propose specific strategies to address particular equity-related issues raised by the use of AI in health care. The remaining appendices contain additional details on the documents we reviewed, stakeholder interviews, and literature review methods.

2. Methods

This chapter describes the methods that we used in our study. Our study effort consisted of research into 2 topics: the use of AI in the COVID-19 response and strategies to address potential impacts of AI on health equity. We began our research by conducting stakeholder interviews, literature searches, and screening of documents that discussed either of these 2 topics. The methods we used to undertake these shared research steps are also described in a separate journal article[27] and companion report on the use of AI in the COVID-19 response. [28] The methods we used to examine each topic then diverged as we carried out subsequent research steps, including extracting information during full text review of documents and analysis of this information.

In this chapter, we first describe the interviews that we conducted with stakeholders at the outset of our study to inform our development of research aims and study design. We then provide a summary of the search strategy that we used to identify documents to include in our review and the approach we used to screen these documents for inclusion in our full text review. The chapter closes with a description of the methods we used in our full text reviews of documents to extract information about issues and strategies in AI health equity.

Stakeholder Interviews

The project team conducted interviews with 9 stakeholder representatives early in our study effort. These interviews focused on understanding stakeholder perspectives on the impact of AI on health equity, potential strategies, and any concerns for further study. This section describes our approach to conducting these interviews, together with a summary of stakeholder feedback on the impact of AI on health equity. Interviewees were also asked about the use of AI in COVID-19; the findings from this part of the interviews are summarized in another report. [28]

We first identified relevant stakeholder groups, drawing on PCORI's recommended list of stakeholders, [29] that could provide a broad range of perspectives on the impact of AI on health equity. These groups included patients and patient advocates, clinicians, hospitals and health systems, payers and insurers, public policy makers, public health officials, researchers, and industry representatives. We then identified stakeholder representatives from each of these groups to contact for a potential interview.

We emailed a total of 19 individuals to ask for their participation in an interview. Nine individuals responded and agreed to participate, including 1 patient advocate, 2 clinicians, 1 health system manager, 1 insurer representative, 1 public policy maker, 1 public health official, 1 industry representative, and 1 researcher.

RAND's Human Subjects Protection Committee approved our interview approach as exempt from further review on October 29, 2021. Interviews were conducted between November 2021 and January 2022. Interviews were approximately 1 hour in length and were conducted by phone or videoconference. Interviews were not recorded.

We sent interviewees an informed consent protocol and description of the study via email before the interview. We also drafted an interview guide, including a list of potential questions, for internal use by researchers conducting the interviews. All team members were involved in drafting the interview guide and questions, which were then pilot tested in 2 practice interviews with RAND colleagues before they were finalized. In carrying out stakeholder interviews, we adapted our questions to the perspective and experience of each individual interviewee.

All interviews were conducted by 2 members of the study team, 1 serving as lead interviewer and the other serving as lead note taker. The lead note taker recorded near-verbatim typed notes to the extent possible, focusing on statements the interviewee made. The lead interviewer also recorded notes on overall themes arising from the discussion. Following the interview, these sets of notes were reviewed for accuracy by both members of the interview team. A PCORI representative also attended some of these interviews. In these cases, following introductions, the PCORI representative participated solely as an observer. The interview guide sent to interviewees, which includes a study description and informed consent protocol, can be found in Appendix B.

We used the interviews to ensure that stakeholder priorities and perspectives were reflected in our review scope, research questions, and study design. To do this, 2 team members reviewed all notes taken during each interview to organize responses according to specific topics. Topics were determined inductively based on a first reading of interview notes and were selected based on their potential use in informing revisions to our study design or providing background context. We also analyzed interview notes to gather information on health equity in AI. A summary of interviewee responses about the impact of AI on health equity is found in Table 2.1.

Table 2.1. Summary of Stakeholder Feedback on Study Approach and Health Equity in AI

Topic	Summary of Stakeholder Feedback
Overall research topic	• None of the interviewees suggested changes to the study topics proposed for our review
Data sets and health equity	• Six interviewees were interested in understanding how often subpopulation characteristics are reported for the data sets used to train, validate, or evaluate AI applications, as well as how these subpopulations were defined • Four interviewees mentioned the challenges involved in collecting data related to health disparities, given that disparities are often driven by social determinants of health that are difficult to measure, and that categories used to characterize people may be contested or poorly defined • Three interviewees mentioned that there can be a tension between ensuring that data used in AI applications are inclusive of underrepresented groups on the one hand, which may require asking for and storing sensitive information such as patient-reported race and ethnicity, and ensuring that patient privacy and trust are maintained on the other hand
Variables relevant to health equity	• Four interviewees mentioned the use of geography as an important and readily available determinant of health, although some cautioned that it can be an imperfect proxy or can render disadvantaged groups invisible if they are dispersed over a wide area, rather than geographically concentrated in a particular ZIP code or neighborhood • Two interviewees mentioned that AI algorithms can identify (and thus take into account) race, ethnicity, or other sensitive personal attributes even from seemingly unrelated information, such as medical images • When asked about the personal characteristics most relevant to understanding health equity, interviewees mentioned geography/location, education, housing status, immigration status, employment status, insurance status, incarceration status, access to transport, and family structure, in addition to characteristics such as race, ethnicity, gender, socioeconomic status, and disability
Evaluating equity-related outcomes	• Five interviewees were interested in the topic of AI model acceptability to different types of patients, including whether patients are comfortable sharing potentially sensitive information that may be needed as data inputs • Four interviewees mentioned the importance of AI impacts on cost of care, access to care, or quality of care received as important from an equity perspective, in addition to impacts on measurable health outcomes
Availability of equity-related information on AI algorithms	• Four interviewees mentioned that AI may be used internally by an organization (such as a health system, government agency, insurance company, or technology company), but little information may be publicly available on these algorithms. This was variously attributed to concerns over publishing proprietary information, lags in academic publishing timelines, and in 1 case, the fear that publishing information on algorithm use in health care could lead to bad publicity if that algorithm was determined to be biased

Topic	Summary of Stakeholder Feedback
Recommended documents and data sources	• Interviewees recommended several documents to screen for inclusion in our review, including peer-reviewed papers and gray literature documents • Interviewees also recommended more general data sources to examine in our review, including paper authors, software developers, and types of articles, such as conference proceedings, as well as viewpoints published in academic journals and commentaries in the news media and gray literature

Document Sources, Searches, and Screening

Following stakeholder interviews and refinement of our study questions, we conducted a scoping review of published literature on the impact of AI on health equity. This effort was guided by established methods for narrative reviews, scoping studies, and evidence maps. [30,31]

To conduct the review, the project team and a RAND research librarian first performed a comprehensive literature search. Search results were then screened to determine whether they fit within the scope of our review. Documents that passed initial screening were then reviewed to identify specific health equity issues and strategies. They were also reviewed to identify documents examining the use of AI in the COVID-19 response, which is the subject of the companion report. [28] Each of these steps is discussed below.

Literature Searches

Full documentation of our search approach, databases, dates, and search terms used is in Appendix C.

We performed 3 sets of searches: 1 targeting AI in the COVID-19 response, 1 targeting AI health equity, and 1 targeting AI health equity in the COVID-19 response. For each search, we selected search terms corresponding to each of the concepts that defined the scope of our study. Results from all searches were combined into a single set of documents that were then screened for relevance to either of our 2 topics.

We tailored search queries to specific databases, including PubMed, Web of Science, the Institute of Electrical and Electronics Engineers (IEEE) Xplore Digital Library, the U.S. Food and Drug Administration (FDA) Center for Devices and Radiological Health (CDRH) Document Library and Emergency Use Authorization (EUA) website, and ClinicalTrials.gov. The search included both peer-reviewed and gray literature. The latter was captured in the previously mentioned databases (eg, conference proceedings from IEEE and ClinicalTrials.gov study records), as well as in targeted Google searches to identify documents from government, nonprofit, and industry sources, following methods used in similar gray literature reviews. [32] In cases in which searches of academic databases would otherwise return several thousand results because of the amount of literature available on our study topics, we limited our search to just review articles or articles flagged in the database as highly cited.

We also conducted a search of media on ProQuest US Newsstream and Academic Search Complete to capture news articles, commentaries, and other documents.

Our searches resulted in a total of 2244 unique documents following removal of duplicate records. Of these unique documents, 1897 were accompanied by abstracts that we used to conduct initial screening. The remaining 347 documents were not accompanied by abstracts and thus proceeded directly to full text review.

We also reviewed other documents in addition to those resulting from our systematic search and screening process. This includes relevant documents identified via citations using a snowball method and documents recommended by stakeholder interviewees.

Document Screening

Following the literature search, the 1897 search results for which we had abstracts underwent screening based on their title and abstracts to determine whether to include them in our full-text review. We included English-language documents of 2 types in our review: documents that discussed AI applications used in the clinical and public health response to COVID-19 and documents that discussed the potential impacts of AI application usage on health equity.

Of these 1897 documents, 313 were selected for inclusion in our full text review during this screening process. This included 277 from PubMed and Web of Science, 11 from IEEE Xplore, and 25 from ClinicalTrials.gov.

Appendix C provides a Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) literature flow diagram of the search and screening results (Figure C.1).

Full Text Review of Documents

Each document screened as eligible for full-text review was read by 1 of 3 members of the study team. Relevant citations from this initial set of documents were also reviewed, as were documents recommended by stakeholders.

For a document to be included in our review, it needed to meet the following 2 criteria: (1) the document discussed AI, and (2) the document discussed potential impacts on health equity.

Criterion 1: Discussion of AI

We included documents that discussed AI, “machine learning,” or “deep learning” algorithms. Documents did not need to use these exact terms; we also included documents that discussed algorithms that used specific machine learning (ML) methods, including advanced ML (eg, neural networks, support vector machines, and complex decision trees), traditional ML (eg, logistic or linear regression prediction models), and unsupervised ML (eg, clustering, principal components analysis).

Criterion 2: Discussion of Potential Impacts on Health Equity

We included documents that discussed the potential impacts of AI use on health equity. In assessing this, we used the definition provided by the World Health Organization (WHO): "Health equity is defined as the absence of unfair and avoidable or remediable differences in health among population groups defined socially, economically, demographically or geographically.… Pursuing health equity means … giving special attention to the needs of those at greatest risk of poor health, based on social conditions." [33] The WHO definition is very similar to other prominent definitions of health equity, such as that from the Robert Wood Johnson Foundation, which states, "Health equity means that everyone has a fair and just opportunity to be as healthy as possible." [34]

Identifying Equity-Related Issues and Strategies

Our unit of analysis was an issue-strategy pair, defined as a discussion of a particular strategy proposed to address a specific equity issue raised by the use of AI in health care. We extracted information on each document using a 3-item data collection form, consisting of the reference for each document, a specific issue that the document discussed, and a particular strategy that the document proposed could be used to address that issue. Each document could include multiple issue-strategy pairs, including instances in which a particular issue was addressed by multiple proposed strategies, as well as instances in which 1 strategy was proposed to address multiple issues. These data on issue-strategy pairs were then used in our analyses of how frequently each issue or strategy was mentioned in the literature and which strategies were commonly proposed to address which issues, as presented in Chapter 5.

We created our set of issues and strategies inductively from our review of the literature: whenever an equity issue or strategy discussed in a document was not adequately described by the current set, we created a new entry. Definitions were refined in group meetings among 3 members of the study team.

Several reviews have organized issues and strategies in AI health equity using development pipeline models, although there is no consensus on a single framework. [3,17,35] We adopted a version of the organizing framework proposed by Chen et al[36] with two modifications. The original framework proposed by Chen et al was targeted at developers. To make it applicable to a broader set of stakeholders, we expanded the "problem selection" category to include other aspects of the "background context" of AI development and use. In addition, we grouped variable choices (a separate category in the framework by Chen et al) with characteristics of the data because data characteristics define whether variable choices are appropriate or problematic. The 4 categories we used were:

- **Background context:** Systemic and structural elements
- **Data characteristics:** Quality and quantity of the data
- **Model design:** Choice of model, variables, and objective function
- **Deployment practices:** Model evaluation, use, and maintenance

The next chapter presents the findings of our literature review on the equity-related issues raised by the use of AI in health care, organized according to these 4 categories.

3. Issues Raised by the Use of AI in Health Care

In this chapter, we describe the equity-related issues raised by the use of AI in health care that we identified in the literature. We identified a total of 18 issues and grouped them into the 4 categories corresponding to different aspects of the AI development pipeline discussed in Chapter 2. These 18 issues are summarized in Table 3.1. We then present a summary of each of these issues based on their description in the academic and gray literature we reviewed.

Table 3.1. Issues Related to AI and Equity From the Literature

	Issue	Description
Background context	Biased or nonrepresentative developers	Development teams may be biased or not representative of the public, patients, or users, leading to mismatched priorities
	Diminished accountability	Lack of accountability makes it more difficult to obtain compensation for AI harms and enables malicious actors to target certain groups
	Enabling discrimination	Developers may use AI algorithms to purposely discriminate, either out of malice or for economic gain
Data characteristics	Limited or poor information on population characteristics	Insufficiently granular data on population characteristics may lead to inappropriately aggregating dissimilar groups, such as classifying race into White and Non-White
	Unrepresentative data or small sample sizes	Inadequate representation of groups in training data can lead to worse model performance in these groups; this often occurs when training and deployment populations do not match
	Bias ingrained in data	When data reflect past disparities or discrimination, algorithms may learn and perpetuate these patterns
	Inclusion of sensitive variables	Inclusion of sensitive information, such as race or income, may cause algorithms to inappropriately discriminate on these factors
	Exclusion of sensitive variables	Exclusion of sensitive information may reduce algorithm accuracy and lead to systematic bias due to a lack of explanatory power
	Limited reporting of information on protected groups	Lack of reporting on the composition of training data or model performance by group makes it difficult to know where to appropriately use models and whether they have disparate impacts

	Issue	Description
Model design	Algorithms are not interpretable	When we do not understand why models make decisions, it is difficult to evaluate whether the decision-making approach is fair or equitable
	Optimizing algorithm accuracy and fairness may conflict	Optimizing models for fairness introduces a trade-off between model accuracy and the strictness of the fairness constraint, so equity comes at the expense of decreased accuracy
	Ambiguity in and conflict among conceptions of equity	There are many conceptions of fairness and equity, which may be mutually exclusive or require sensitive data to evaluate
Deployment	Proprietary algorithms or data unavailable for evaluation	When training data, model design, or the outputs of algorithms are proprietary, it is difficult to independently evaluate them for bias or disparate impact
	Overreliance on AI apps	People may blindly trust algorithmic outputs, implementing decisions despite contrary evidence, perpetuating biases if the algorithm is discriminatory
	Underreliance on AI apps	People may be dismissive of algorithm outputs that challenge their own biases, perpetuating discrimination
	Repurposing existing AI apps outside original scope	Models may be repurposed for use with new populations or to perform new functions without sufficient evaluation, bypassing safeguards on appropriate use
	Application development or implementation is rushed	Time constraints may exacerbate equity issues if they push developers to inappropriately repurpose existing models, use low-quality data, or skip validation checks
	Unequal access to AI	AI applications tend to focus on high-income areas, potentially amplifying preexisting disparities

Background Context

Applications within the same functional category tended to share similar data inputs, predicted variables, users, settings, and proposed benefits, as discussed here.

Biased or Nonrepresentative Developers

Developers of AI applications tend to be less diverse in their sociodemographic characteristics compared with the general population. [37-41] For example, only 22% of AI professionals worldwide are female, [42] and Black or Hispanic employees represent only a small portion of the workforce at major firms involved in AI research and development. [43-45] As a result, conscious and unconscious developer biases may affect various steps of the development process, including the decision to undertake development, choice of key variables, model training, testing, evaluation, interpretation, and dissemination. [3,6,46-51] Moreover, the lack of representation for members of minority groups in AI development has consequences for the acceptance of technology among minority populations because of the fear that AI may be misused.[38]

Diminished Accountability

The decisions made by AI algorithms are a function of data, design decisions, and use, so it is difficult to attribute harm to a single entity. If an algorithm provides a physician with an incorrect treatment recommendation that harms the patient, it is unclear whether the developer, physician, health system, state, or no one is at fault.[52] The use of AI means that the risk of discrimination may now be "inherent" in the procedure, in the same way that there are risks inherent in heart surgery.[53]

Existing legal frameworks are ill-equipped to deal with algorithmic decision-making. Antidiscrimination laws rely on the plaintiff showing the conscious intent to discriminate, an impossible standard for algorithms.[18] Even proving a lower standard, such as disparate outcomes, is difficult because this requires knowledge of system performance across population groups. Large outcome data samples are likely only available to developers or large clients, both of which may have limited incentives to address equity issues. In contrast, users may only have access to information on their own outcomes.[54] Without clear liability or readily available evidence, it is more difficult to hold actors accountable and to seek restitution for harm. This diminished accountability is an equity issue because some groups may disproportionately experience algorithmic errors.

In some cases, there may also be economic incentives to use algorithms to avoid regulation because variations in the enforceability of antidiscrimination laws create opportunities for regulatory arbitrage. Actors can outsource decisions that might be discriminatory to algorithms, avoiding liability while hiding behind the supposed neutrality of their AI systems.[55]

We found little discussion of such accountability issues specifically in the context of health equity, but several authors pointed to hiring algorithms potentially allowing firms to bypass fair hiring laws or legislation for protecting specific groups, like the Americans with Disabilities Act.[54] The lack of specific examples is unsurprising, as firms may be unwilling to indicate that they are attempting to bypass regulation.

AI-Enabled Discrimination

Although most equity issues are unintentional, as in the case of developers using unrepresentative data or inappropriate proxies as outcome variables, some may use AI algorithms specifically to discriminate. ML is excellent at discrimination, because it can flexibly target any set of characteristics that defines a labeled group.

Some applications that use AI to discriminate require a bad actor, such as an authoritarian government using facial recognition to identify and surveil a disfavored minority group. However, other applications merely require that rational economic actors follow profit incentives, including in health care. For instance, several companies offer "propensity to pay" algorithms for health care providers. These algorithms are marketed as making debt collection efforts more efficient but could potentially also be used to inform health system decisions on what care to provide to which patients. Other potential opportunities include "cream skimming" (seeking to enroll only the most profitable patients) in health insurance markets. [56] ML algorithms could effectively identify these profitable patients even after risk adjustment, leading to differential access to care and exacerbating health disparities.

Data Characteristics

Modern AI models are, by nature, statistical models. They learn relevant patterns from data samples and use these patterns to make predictions. Therefore, if the data are problematic—for example, because they reflect existing biases, are not representative of the general population, or do not contain an appropriate set of variables—the predictions may be biased and unfairly affect vulnerable populations. Consequently, some of the equity concerns associated with the use of AI in health care settings stem from properties of the data.

Limited or Poor Information on Population Characteristics

The utility of AI algorithms is often limited because they use data sets that do not account for or misrepresent important sociodemographic characteristics of populations and subpopulations. [39,57, 40] Person-generated data from smartphones is particularly problematic because even basic demographic characteristics such as age and sex may be unknown, making it challenging to do meaningful analyses across subgroups. [19]

Sometimes information is available about the populations studied during training or validation, but the information may not contain variables that are important to key stakeholders. For example, during the COVID-19 pandemic, efforts to identify vulnerable communities were undermined by the lack of granular data describing population characteristics, such as the prevalence of comorbidities among patients in a given geographic area. [58] The use of race and ethnicity in AI algorithms can be particularly problematic because of issues with "reliability, completeness, and lack of comparability across data sources." [59,60] In reaction to this problem, investigators commonly categorize patients into 5 races and 2 ethnicities, which is a reductive approach that likely fails to capture clinically relevant differences. [61-63] Categorization of sex is often limited to male and female, which inhibits efforts to study marginalized populations such as those who identify as nonbinary or transgender. [3]

Combining data sets from different sources also causes similar problems because harmonizing demographic variables across multiple data sets introduces additional sources of error and may magnify problems caused by missing or inaccurate data. This problem, dubbed "aggregation bias," causes problems with validity because a model may not produce adequate results when applied to a population with differently coded demographic characteristics. [64] The end result can be failure to recognize meaningful differences across subpopulations because of data limitations, [51] which is especially consequential for cases in which intersectionality of multiple traits is important. [19]

Unrepresentative Data or Small Sample Sizes

As statistical models, the certainty with which an AI model identifies specific patterns is a function of how many samples exhibit the pattern in the data set. Data sets that have unbalanced representation of subjects in the general population will tend to produce models that are less certain when making predictions regarding members of unrepresented populations. [65] This presents as unequal sensitivity (true positive rates) in models based on such data. [3,6,18] This issue can be especially dangerous when there are subpopulations that do not engage with or participate in health services with as much frequency or in the same manner as the majority. [18,66] Models based on data sets drawn from such unrepresentative contexts will perform less accurately on minority populations and, in some cases, may even lead to adverse health outcomes. [60,67,68]

Concerns around data representativeness carry a second dimension: the portability of AI models across contexts. Context here refers to both the population context as well as the decision context. Justified AI model deployment requires the statistical distributions of the data during training and deployment to match (an assumption of *statistical regularity*). [59,69] The use of a model on a distinct statistical population makes standard checks for internal model validity (eg, model accuracy measures) less reliable. So, for example, models developed using data from specific countries or patient populations are not guaranteed to work well for other populations without significant modeling and validation effort. [70,71] Besides the population context, the decision context also matters. The use of AI models outside the scope of the decision question they are designed to address can lead to decision failures. We see this in examples such as the model described in the Obermeyer et al analysis[10] in which a model addressing the prediction question of health cost is applied to the related but separate prediction question of illness severity.

Bias Ingrained in Data

Data are often the products of social processes. Members interacting on a social media platform produce language data. Citizens and police officers interacting produce criminal justice records. And patients and doctors interacting in settings produce health records. Social processes are inflected by pervasive biases that are then recorded as data for AI models to learn from. As statistical pattern recognition tools, AI models, trained on such socially generated data, are bound to learn these biases ingrained in the data and transmit these biases into future decisions if no meaningful oversight is applied. [65,72] The effect of this phenomenon does not end with a one-time transmission of bias to future AI-enabled decisions. Such biased AI-enabled decision outcomes, in turn, produce biased data that further feeds AI training pipelines directly or indirectly. This sketches out a *runaway feedback loop* of adverse outcomes on disadvantaged subpopulations that may compound over time, barring meaningful interventions. [3,18,72,73]

AI models for natural language models have historically furnished the most salient examples of both the social construction of data (language use is a social act) and the bias ingrained in data. A 2017 study found that standard biases or implicit associations in language use (eg, “female names are more associated with family than career words, compared with male names”). [74] The criminal justice system provides another heavily social data-generating context, with examples such as reports of how sociohistorical biases led to disproportionate targeting of Black people by police interested in drug and gang activity. [40,65]

In the context, the process of ingraining bias into data streams can take myriad forms. For example, language barriers, health literacy, the kinds of accessible care (eg, teaching vs nonteaching hospitals), and other socioeconomic factors may lead to differences or disparities in care received and thus affect relevant health data used to train models. [75] Another example of biases in delivery involves the level of pain management in postpartum or postoperative care. [76,77] Recent research highlights racial disparities in pain management “that cannot be explained by less perceived pain.”[75]

The case discussed by Obermeyer et al [10] is also a relevant example here. The authors reviewed an algorithm deployed by insurance companies to allocate resources to patients. They discovered that the algorithm used spending as a proxy for health without considering differential access by race. As a consequence, for the same level of need, fewer resources were allocated to Black customers. [10] This example illustrates bias ingrained in the data: bias presents itself in the form of differences in how White and Black patients are reported to engage with health services relative to their needs. This example also illustrates how biased models create feedback loops that compound disadvantages on vulnerable groups. Left unchecked, that model would have continued to make recommendations that reduced care levels for Black patients. [10]

Ingrained data biases can be difficult to mitigate. AI-deploying institutions will need to continually monitor their models, not just for internal measures of quality (eg, held-out set accuracy), but also for external measures of quality. Such external measures would include deep survey of diverse user experiences with AI-equipped health decision-making use cases. It will also be important for health AI designers and deployers to be aware of the societal biases that are relevant to their data sets and come up with a plan to mitigate these societal biases.

Inclusion of Sensitive Variables

Sensitive information, such as an individual’s race or income level, is sometimes included in the data used to train AI applications. Views differ on whether this leads AI applications to discriminate in ways that exacerbate health disparities, and it may depend on the specific circumstances of an application’s use. This section presents a summary of the literature discussing this issue as a contributor to health inequities. Conversely, the next section summarizes literature that supports an opposing view, which holds that excluding sensitive variables from AI applications can contribute to health inequities, at least in some circumstances.

Data on patient race, for example, have been used in regression-based AI algorithms widely adopted in US clinical practice to predict kidney function, [78] heart surgery risk, [78] and successful vaginal birth. [79] These algorithms produce systematically different predictions for Black or other non-White patients as compared with others. Their use may have exacerbated racial health inequities through delayed diagnosis of kidney disease, recommendations against heart surgery, and recommendations for caesarean section rather than vaginal birth for non-White patients. [61] In fact, these concerns led the race variable to be dropped from the equations used to predict kidney function. [80-82] Moreover, some researchers have worried that when people see a sensitive variable included in an AI model, that by itself may "[perpetuate] incorrect assumptions about the mechanism of causality in health disparities," by, for example, "assuming a biological foundation" for disparities that are "in fact explained by social determinants." [40]

Including sensitive information in AI algorithms may be particularly problematic under certain conditions. Some researchers have argued that sensitive information, such as race or gender, should not be included in algorithms when they have no known biological effect on the health outcome of interest. [19,36,61,70,83] Furthermore, although sensitive information may not be appropriate for training clinical decision support algorithms, it might be allowable when used to evaluate disparities in algorithm performance or as part of algorithms that examine health disparities and their contributors. [40,61] Another key distinction may be whether an AI application uses sensitive information in ways that disproportionately harm, rather than benefit, a marginalized group. [40,61] In the health care context, this latter distinction may not always be clear; however, given that many AI-informed interventions (such as heart surgery or diagnosis of kidney disease), while intended to benefit patients, also carry some potential for harm (from postsurgery mortality, unnecessary dialysis, or withheld antibiotics in the case of overdiagnosed kidney disease). [84]

Exclusion of Sensitive Variables

Sensitive information is often excluded from the data used to train AI applications, whether for reasons of data availability, legal restrictions, or the concerns over algorithmic discrimination described previously. When sensitive variables serve as determinants of health, excluding them may reduce algorithm accuracy and can potentially lead to systematic bias. For example, information that might be highly sensitive, such as skin color, may be needed to evaluate and correct for bias in the case of AI-enabled smartwatches, pulse oximeters, or other wearable sensors that measure light transmission through a patient's skin to estimate heart rate or blood oxygen levels. [6,85,86]

Exclusion of sensitive variables may exacerbate disparities even when the sensitive variable serves as a social, rather than biological, determinant of health. One study found that an AI algorithm used in California to allocate additional COVID-19 vaccines to disadvantaged communities was less accurate in identifying neighborhoods facing disparities in life expectancy when it excluded race as an input variable. [87] Another study compared other COVID-19 vaccine allocation algorithms used elsewhere in the United States and found that the AI algorithm that excluded race as an input "yielded lower benefits to minority groups" than a different algorithm that included race. [88,89]

Limited Reporting of Information on Protected Groups

Lack of reporting and outcome standards poses a significant challenge to improving AI equity. Relatively few studies report even basic demographic information about their training data. A review of mechanical ventilation algorithms found that only 19% of studies reported race and ethnicity for training data. [90] Similarly, outcome data by subgroup are also often not considered. A review of 130 FDA-approved AI-based medical devices found that only 13% reported that demographic subgroup performance was considered in evaluations. [91] Without these data, it is impossible to know which populations and contexts algorithms generalize to or whether they will have disparate impacts. [91]

Model Design

Data characteristics are not solely responsible for AI equity concerns. [92] Choices in algorithm design and deployment have a role to play in mediating fair outcomes in the use of AI models. The following subsections highlight some model design choices and constraints that can influence model fairness. Fairness issues around model design, while complex, are often more directly addressable because the stakeholders responsible for model design choices are easiest to identify: the model designers.

Algorithms Are Not Interpretable

Some ML algorithms are not interpretable, meaning that it is difficult to examine the exact relationships between data input and algorithm outputs. This issue, often referred to as the problem of "black-box" AI, applies primarily to more complex algorithms, such as neural networks that rely on multiple layers of data analysis and abstraction. This can hinder efforts to evaluate whether an algorithm's decision-making approach is fair, potentially undermining user trust[55,93] or the ability of those affected by algorithms to seek legal redress. [18]

Algorithm opacity is sometimes mentioned as potentially contributing to AI applications having a negative impact on equity. [53,59,72,94] However, none of the documents we reviewed described a specific case in which this has occurred. Furthermore, a prominent researcher has argued that "humans are inscrutable in a way that algorithms are not" when it comes to racial bias, [95] giving the example of his own successful detection of bias in an algorithm, despite not having visibility into its inner workings.

Nevertheless, algorithm opacity precludes straightforward evaluation of how an AI application handles race or other sensitive information when those are present in input data. One such evaluation may have contributed to the removal of race as an input variable from a widely adopted AI-based vaginal birth after cesarean risk calculator. [96] This evaluation relied on the presence of interpretable variable coefficients to critique how a previous version of the calculator used a race variable to predict that Black and Hispanic patients had a lower chance of successful vaginal birth. [61] Had this been a more complex and uninterpretable algorithm, rather than a simple regression-based ML model, it would have been more difficult to examine how race affected algorithm predictions.

Optimizing Algorithm Accuracy and Fairness May Conflict

Typical adaptations for making AI models fair(er) aim to optimize both the model's accuracy as well as its performance in achieving a particular measure of fairness. [97-99] For example, a model might be developed to optimize both its accuracy in identifying individuals who could benefit from mental health services within a population as a whole, while also seeking to ensure that these services are accurately and equitably provided to specific population subgroups. [100]

This type of joint optimization effort will often feature a trade-off between the accuracy of the model and the strictness of the fairness constraint. The degree to which optimizing a model for fairness may result in lower accuracy varies depending on the AI model. However, recent work shows that this trade-off may also be negligible in certain applications and for carefully selected measures of model accuracy and model fairness. [60]

Ambiguity in and Conflict Among Conceptions of Equity

There are innumerable ways of conceiving of what it means to be fair or equitable, with examples including statements such as "model allocations of opportunity should reflect the population distribution," "false positive rates should be equally distributed across subgroups," or "model outcomes must be independent of demographic attributes." [18,65,97,101]

The choice of what definition of fairness to incorporate into the design of an AI model is not always clear and depends on broader consideration of health care goals and conceptions of equity. Suggested frameworks exist for navigating this choice. [102] But, in general, the choice requires detailed domain knowledge and consultation with relevant stakeholders because the legitimacy of a fairness norm will ultimately depend on the buy-in from affected stakeholders.

Different equity norms can conflict with each other. The recent literature on fair ML articulates a set of statistical impossibility theorems that show that it is often not possible to jointly satisfy collections of fairness definitions. [65] Selecting a fairness norm will typically preclude the satisfaction of other kinds of fairness. This is especially true if the model is predicting the occurrence of a condition that varies in prevalence across relevant subpopulations.

Deployment Practices

It is useful to frame the use of AI in health care as an intervention or experiment to improve decision-making. This emphasizes the basic fact that practices in AI application deployment can significantly influence the fairness of even the most carefully developed AI systems. The following discussion highlights several issues concerning how deployment practices surrounding the use of AI in health care can affect equity.

Proprietary Algorithms or Data Unavailable for Evaluation

Independent evaluation of the fairness and generalizability of AI applications is a powerful tool to prevent poor equity outcomes. [103] However, external evaluations can only take place if the data, preprocessing, and algorithms used by the tool are widely available. [73] Information on preprocessing and variable selection are particularly important steps because developers have much subjective choice in these decisions. Often, AI application developers do not release algorithm development information, especially if the algorithm is being used for commercial purposes. This issue can also compound problems related to lack of reporting on population characteristics described earlier in this chapter.

One example is the algorithms used by health insurance firms, as demonstrated by the algorithm used for resource allocation studied by Obermeyer et al. [10] This algorithm was not public, and, had it not been made available to researchers by a health system user, racial biases might never have been uncovered. This raises the question of what undiscovered biases exist in the proprietary algorithms used by other health insurance firms and hospital systems.

Overreliance on AI Apps

Some authors caution that use of AI may lead to unintended consequences if algorithmic outputs are relied upon too heavily or accepted at face value without appropriate validation or inquiry. [6,48,55] A phenomenon known as "automation bias" is defined as the tendency for humans to accept machine-generated decisions over conflicting human decisions, despite the existence of contrary information. [53,104] Overreliance on algorithmic outputs can lead to errors of omission when humans fail to discern failure of an AI tool. For example, when a high volume of decisions is made automatically (eg, using AI to interpret chest x-rays at a large medical center), there may be no human available to validate computer-generated results. Conversely, overreliance on AI can lead to errors of commission when humans implement an algorithm's decision despite evidence to the contrary. [104] Automation bias may perpetuate structural biases for vulnerable populations, leading to discriminatory outcomes because of overreliance on AI applications. [3,4,104,105]

Underreliance on AI Apps

Underreliance on AI applications may also lead to negative impacts on health equity under certain circumstances. For instance, an algorithm may produce a result that a health care professional dismisses without adequate investigation, even if the algorithm result is in fact less biased than the professional's own judgment. Murphy et al provide an example where the outputs of IBM's Watson Oncology, an AI-powered diagnostic decision support system, were rejected by physicians for reasons that were poorly justified. [48] This example highlights that underreliance may be a more salient problem when intended users mistrust the process or have inadequate understanding of how algorithms operate, especially if algorithm outputs conflict with user judgment. [6,106] Unfounded dismissal of algorithm results might be more likely to harm vulnerable groups because of existing structural biases. [3]

Repurposing Existing AI Apps Outside Original Scope

Ideally, applications using AI are trained, validated, and then deployed for use with a defined, preplanned purpose. However, during a time of crisis, an AI application might be deployed for use cases outside its intended scope. [6,18,64] For example, during the COVID-19 public health emergency, a widely available algorithm designed to predict clinical deterioration among patients before the pandemic was repurposed to predict deterioration among patients with COVID-19. Not surprisingly, the performance of the algorithm was found to be poorer than other algorithms that were trained and validated among patients with COVID-19. [107,108] This case illustrates that, although repurposing AI may be faster than designing a new application, the application's output may be "unpredictable, unexpected, or biased" because of out-of-scope use. [109] Additionally, repurposing AI likely causes problems related to equity that go undiscovered and unaddressed, especially if repurposing is done under crisis conditions without appropriate safeguards for vulnerable populations. [110] This may have occurred during the early stages of the pandemic, when algorithms trained to predict spread of disease in Europe performed less accurately in African countries. [51, 51, 111]

Rushed Development or Implementation

Rushing the development of applications compounds equity issues. During the COVID-19 pandemic, there was significant pressure to reduce the time required to develop AI applications because the virus was moving so fast, and there was fierce competition to be the among the first applications released. Many of the strategies designed to mitigate problems in AI equity take time, such as consulting community members, improving data quality, or evaluating model outputs. In high-pressure situations, these aspects of development may be ignored.

AI developers or users may also attempt to rapidly repurpose algorithms in contexts outside their intended scope. For example, during the COVID-19 pandemic, hospitals used triage applications trained on prepandemic data to determine who might need intensive care or mechanical ventilation, despite the fact that COVID-19 patients were not in the training set and therefore no validation had been performed on this group of patients. [6] More generally, facing major time constraints, developers may also use data from one population, such as those initially affected by the pandemic, and extrapolate to other populations, without addressing representativeness. [4] Many of the other equity issues discussed in this chapter are more likely to appear in this context.

Unequal Access to AI

Research and implementation of AI applications tend to be focused on areas that are relatively high income. [112] This disparity in access to AI exists not only between high- and low-income countries but also within individual countries. [113-115] As a result, use of AI may amplify preexisting disparities for vulnerable subpopulations. [3,116]

Disparities in internet access is one reason for inequity in the distribution of AI. For example, although an estimated 4 billion people worldwide accessed the internet in 2019, only 28% of people in Africa had internet access compared with 82% in Europe. [117] Thus, any advances in health because of internet-based AI applications likely improved health for those in Europe more than those in Africa. Additionally, within most countries, vulnerable groups are less likely to have access to technologies necessary to use AI applications, such as broadband internet or smartphones. [118] Taken together, inequities in access to AI can lead to accrual of benefits “to the highly educated and wealthier segment of the population, while displacing the less educated workforce, both by automation and by the absence of educational or retraining systems capable of imparting skills and knowledge needed to work productively alongside [AI applications].” [119] In the case of unequal access to AI, the subpopulations most likely to experience disproportionate benefits are those that are “tech savvy, highly health literate, self-directed, information seeking, English fluent, health focused, and well insured.” [120]

Summary

We identified 18 equity-related issues raised by the use of AI in health care that are discussed in the literature. These include issues related to the background context that shape AI development, the characteristics of data used as model inputs, the design of AI models, and the ways that AI-based health interventions are deployed in practice.

The next chapter describes the strategies that were proposed in the literature to address these issues and ensure that the use of AI improves, rather than undermines, health equity.

4. Strategies Proposed to Address AI and Equity Issues

This chapter describes the 15 strategies we identified that were proposed in the literature to address the 18 equity issues raised by the use of AI in health care. As in the previous chapter, we organize discussion of these strategies according to 4 categories that correspond to different aspects of the AI development pipeline: background context, data characteristics, model design, and deployment practices.

These 15 strategies are listed in Table 4.1. This is followed by a more detailed summary of each strategy in the rest of this chapter, based on their description in the academic and gray literature we reviewed.

Table 4.1. Strategies to Address AI and Equity Issues From the Literature

	Strategy	Description
Background context	Foster diversity	Create teams with diverse characteristics, experiences, and roles to increase consideration of equity throughout development
	Train developers and users	Train AI developers and users in equity considerations and the ethical implications of AI, which may be missing from their formal education
	Engage the broader community	Foster community involvement throughout development, from conception to postdeployment, to increase the likelihood that developers prioritize equity concerns
	Improve governance	Enact robust regulation and industry standards to align AI applications with social norms, including equity, safety, and transparency
Data characteristics	Improve diversity, quality, or quantity of data	Train models with large diverse samples that are representative of the target population for the application and contain all relevant features
	Exclude sensitive variables to correct for bias	Exclude sensitive variables or replace them with variables that are more directly relevant to patient health outcomes to prevent models from discriminating directly on these characteristics (although they may still discriminate on latent approximations of these characteristics)
	Include sensitive variables to correct for bias	Include sensitive variables to improve model accuracy, increase explanatory power, and enable easier testing for disparate impact

	Strategy	Description
Model design	Enforce fairness goals	Formulate an equity norm and enforce it on the model by editing the input data, objective function, or model outputs
	Improve interpretability or explainability of algorithm	Choose models that are inherently explainable (such as decision trees), build secondary explainable models, or explore explainable local approximations to model decision-making
	Evaluate disparities in model performance	Evaluate model performance on a wide range of metrics across subgroups, particularly groups that might face disparate impact, then report and act upon the results
	Use equity-focused checklists, guidelines, and similar tools	Incorporate equity-focused checklists into workflows for developers, reviewers of AI models, health care providers using an application, or patients who want to understand algorithm outputs
Deployment practices	Increase model reporting and transparency	Provide more information on AI equity issues, including publishing standardized equity-related information on models, increasing independent model reviews, and requiring equity discussion in academic journals
	Seek or provide restitution for those negatively affected by AI	Proactively provide restitution to those harmed by AI or create legal frameworks so they can seek restitution
	Avoid or reduce use of AI	If equity issues are severe, or improvements have been fruitless, consider discontinuing model use
	Provide resources to those with less access to AI	Improve access to AI for disadvantaged groups and low-income countries by subsidizing infrastructure, creating education programs, or hosting AI conferences in these locations

We also identified 4 additional strategies proposed in the literature we reviewed that use AI to proactively enhance equity by addressing preexisting health disparities, meaning disparities that are not themselves caused by the use of AI. These 4 strategies are summarized in Table 4.2 and are described in greater detail at the end of this chapter.

Table 4.2. Strategies That Use AI to Address Preexisting Health Disparities

Issue	Description
Develop and field AI that is less biased than standard decision-making	Accurate and well-designed AI-based health interventions may help question biases inherent in health-professionals' standard decision-making practices
Use AI to examine disparities	AI can improve researchers' and health professionals' understanding of the extent and source of health disparities

Issue	Description
Use AI to identify people facing disparities	AI can be used to identify individuals or communities that face health disparities as part of efforts to target service delivery to those who need it most

Background Context

Equity considerations start at the very beginning of the development process. Decisions made on what applications are valuable, how to conceptualize problems, and whose views are important set the context for the rest of the AI development life cycle. The equity strategies in this section focus on who has input into decisions in AI development, how well they understand equity issues, and how regulation can channel and constrain their actions.

Foster Diversity

The benefits of increasing diversity in the workplace have grown clear over the last several decades. For example, a recent article in the *Harvard Business Review* highlighted the business case for improving diversity, concluding that diverse teams "are more effective than homogeneous teams." [121] Having a diverse team of developers for AI applications is particularly important in the health sector, because a lack of insight into ingrained biases in the design process may lead to adverse health outcomes for members of vulnerable groups. Diversity at all levels of an organization, including among those individuals who are involved in setting AI development priorities, can lead to improvements in equity for communities affected by an application, including through the implementation of inclusive community involvement in design processes. [6]

Ensuring representation for vulnerable subpopulations in AI developer workplaces is challenging, so pursuing multiple approaches may be necessary to advance equity. [122] First, hiring practices can be adjusted to improve diversity at all levels of organizations involved in AI development and implementation. [123] A second approach could be to recruit diverse experts, not as employees, but rather as advisors for particular steps of the design process. [49] A third approach is to improve the workplace climate for underrepresented groups. This may require altering job advertising practices, modifying position prerequisites that discourage diverse applicants from applying, and changing day-to-day culture to improve acceptance of people with underrepresented backgrounds (eg, sharing of preferred pronouns in email signatures to encourage acceptance of gender-nonconforming individuals). [122]

In addition to improving sociodemographic diversity, gathering interdisciplinary perspectives is also important for ensuring equity in AI applications. [48,75] A team of technical developers tends to benefit from having other experts involved in the design of AI applications related to health, because understanding the societal, ethical, and economic impacts of an application on equity is not a straightforward process. [124] Adding team members with other areas of expertise, such as attorneys and AI ethicists, has been recommended. For applications with broad impact, interdisciplinary advisory boards could also be considered so that issues related to equity can be identified early and addressed. [109]

Train Developers and Users

Several publications in our review noted that training related to equity is more common in health care programs than in the science, technology, engineering, and math (STEM) education for AI developers. [46,113,122] Lack of training leads to less awareness and fewer skills to address gaps in equity. [113] To advance equity, STEM training programs can design courses that inform students about equity-relevant concerns for vulnerable population subgroups and also train students more generally about the ethical implications of AI. [38,124,125] In addition to training in formal educational programs such as degree programs, workplaces can also institute training for current employees. [126]

In clinical contexts, clinicians and patients benefit from awareness of the risks posed by AI algorithms. Understanding the risks and benefits of AI tools can help physicians fulfill their responsibilities to minimize harm to patients. [127] This includes informing patients of the potential benefits and dangers posed by the use of clinical algorithms, such as the risk that the algorithm may be biased against disadvantaged groups. [53] Patients and providers may benefit from more general AI education, [48] both so that they can take advantage of advances in AI[38] and so they can adopt a healthy skepticism of algorithms. [16] Missing either of these perspectives could exacerbate inequities, as groups either lose out on benefits because of a lack of adoption or are exposed to discriminatory outcomes because of an overreliance on algorithmic decision-making.

Engage the Broader Community

Many of the sources we reviewed recommended engaging the broader community to identify and address concerns related to equity in AI applications. Several authors advocated for community involvement from start to finish of the AI design process, including the decision to build an application, goal-setting for the model, defining fairness, interpreting data output for protected groups, and guarding against unintended consequences after deployment. [3,38,101,128-130] One comprehensive approach for engaging community stakeholders is through a community-based participatory process. The basic premise of community-based participatory activities is that researchers or developers partner with community organizations to elevate diverse voices, establish a respectful dialogue throughout a project, and make joint decisions that take into account all stakeholder perspectives. [131] Through this process, community perspectives can be embedded into the design process to allow communities to own how AI design "is constructed, conducted, and disseminated, thus reversing the power imbalance currently present" in most development activities. [132]

Depending on the context, different stakeholders may need to be included. For AI applications in health care, community stakeholders might include patients, service end-users, physicians, and nonprofessional caregivers, [133] as well as members of historically underrepresented groups who have experienced oppression. [70] These groups may represent people with disabilities, older adults, racial and ethnic minority groups, lower income individuals, sexual and gender minorities, and individuals with intersecting identities. [134] One author recommended that AI applications, such as those assisting with infectious disease surveillance, should "consider the rights of people from diverse regions and communities." [135] A stakeholder analysis of multiple communities in the United States conducted during the COVID-19 pandemic underscored the importance of engaging historically underrepresented groups in AI development. Community members of diverse racial and ethnic backgrounds shared stories about feeling invisible, and they advised that there was an "increased need for inclusiveness and involving community voices in the development and deployment of AI within health care settings." [38]

Additional steps developers can take to engage communities include using consensus-building mechanisms, encouraging oversight by and involvement of community members in agenda setting, [6] and creating formalized channels to bring community voices into critical parts of the design or auditing processes. [3,18] For example, the National COVID Cohort Collaborative held a Tribal Consultation to "decide how to appropriately make [American Indian and Alaska Native] data available to tribal researchers and the broader scientific research community." The relevant data will not be released until the National Center for Advancing Translational Sciences releases a report on the Tribal Consultation and outlines the terms of data sharing. [136]

Communities can also be engaged in ongoing data governance decisions. For example, a data governance panel could be established to periodically review data sets used for training AI applications. Such a panel could "work to achieve a clearly articulated data collection and utilization strategy that will guide documentation, workflow, a review of influencing factors and monitoring standards." This panel would be comprised of diverse interdisciplinary members that have experience with AI, including clinicians, managers, patient group representatives, and technical and ethics experts. [137]

Another approach to community engagement involves having organizations create recommendations for how developers should approach equity—so-called soft governance—in partnership with community groups. Soft governance mechanisms may enable diverse opinions to be voiced, but they lack mechanisms to ensure compliance. At least 2 examples of soft governance frameworks can be applied for AI applications in response to the pandemic. [18] The Toronto Declaration is a framework of international human rights law to promote rights to equality and nondiscrimination in ML[138]; and the Asilomar AI Principles are a judicial transparency framework that has implications for addressing algorithmic equity. [139] These frameworks were created and promulgated by AI researchers and civil society organizations, including IEEE, the Association for the Advancement of Artificial Intelligence, and AI Now.

Finally, governments often engage communities when making policies governing AI applications, and governments may sometimes encourage or require oversight of AI applications through formalized community engagement programs. [140,141] The Organization for Economic Co-operation and

Development (OECD) provides a list of AI-related policies and strategies by country on their OECD.AI website.[142] One government initiative, the "Chilean Participation Process on AI," lists 4 policy instruments related to community engagement: public awareness campaigns and civic participation activities; regional roundtables to discuss AI challenges and policy opportunities; self-organized roundtables to discuss national AI policy; and opportunities for the public to contribute written feedback on national AI policy. [143] Readers interested in local models for community engagement in AI that have been initiated by governments should consider visiting the OECD.AI website to browse programs in their respective countries. [142]

Improve Governance

Robust governance of AI can go a long way to ensure the alignment of AI applications with social norms, including equity, safety, and transparency. There is broad interest in ensuring the equitable use of AI across a broad spectrum of domains, [94,101,144,145] including in health applications. [3,130,146,147] Two concerns are implicated in the push for effective regulation: the translation of desirable norms into clear regulatory guidelines and the enforceability of any selected regulations. We address each below.

Much AI regulatory interest is expressed in the form of guiding principles for AI use, [145,146] including mitigating disparate model performance and how to prevent AI applications from being used beyond their original scope. [18,130] However, little is said about the translation of such principles into development practice. This practical translation is important given the trade-offs and tensions inherent in enforcing consideration of equity concerns in AI development and use. Implementation forces familiarity with choices such as which definitions of equity are relevant, how equity will be measured, and how much of a trade-off in model accuracy is acceptable. These practical details are important for regulation.

AI regulation is difficult because AI is a *broad-spectrum* technology in that it can be applied in many different applications. AI models find use cases in disease surveillance, drug discovery, medical diagnoses, vitals monitoring, and automatic control of medical devices. It is hard to accommodate all regulatory authority for these applications in a single agency. This suggests that, although consensus may exist on the goals of AI regulation, practical regulation may need to be multifaceted and multiagency. Preexisting regulatory authorities can and do already assert some level of authority over different health-relevant uses of AI, such as FDA authority over AI use in some types of medical devices. [91,148-150] However, this mandate does not extend to all health-relevant uses of AI, such as disease-forecasting models and other public health applications.

Other modes of regulation are available besides *hard* regulation of the sort that the FDA typically wields. As mentioned earlier, *soft governance* or forms of regulation not backed by hard government enforcement may be relevant. [18,151] This includes nonbinding standards setting (eg, as provided by the National Institute for Standards and Technology) and industry-organized protocols, standards, and statements of principles. Industry-based self-regulation approaches have the benefit of relying on the expertise of the community of AI practitioners. Self-regulation has a history of success in technology domains that rely on interoperability (eg, communications technology).[152,153] However, industry self-regulation can also be more susceptible to capture by commercial interests.

Hard AI regulation must contend with several limitations. The enforcement of regulations requires the regulator's ability to test and measure for compliance. This can be problematic because of a lack of measurement standards, intellectual property concerns, and the technical maturity of the regulatory workforce. Regulators benefit from sufficient technological maturity to judge the compliance of new AI-based products, which is hard to ensure in a fast-moving space such as AI. Finally, AI regulatory impulses must balance their regulatory mandates against the burden of regulation on innovation.

Data Characteristics

Data are the cause of many biases in AI algorithms, and many important strategies to enhance equity focus directly on improving the quality of the data itself. This can be challenging, however, given the complexities involved in the processes of data collection and curation. For instance, developers need to identify the most suitable proxies for unobservable outcomes, decide how much information to include about protected groups, identify which variables to include in a model, and define protected groups in a granularity appropriate to their application. Not all these data may be readily available, and so compromises or additional work must be directed toward addressing data inadequacies.

Improve Diversity, Quality, or Quantity of Data

Many developers start with their data, not their algorithms, when they look for sources of bias. One survey found that most developers considered equity at the point of data collection or curation. [49] Data can create algorithmic bias if it is not representative (or small), contains biased outcomes, lacks information on group identifiers, or is missing key explanatory variables. The solutions to these issues are simple in theory: use large diverse samples that are representative of the target population for the application and contain all relevant features. [154,155] Algorithms retrained with more representative data have fewer equity issues and perform better overall. [67,156] However, accessing these data may require modified collection methods. For prospective trials, this could be changing recruitment procedures to target traditionally underrepresented groups. [120,157] When judging the diversity of cohorts, researchers should remember that no group is a monolith and that socioeconomic, clinical, and demographic factors all interact. The sample for each group must cover the breadth of the target population across all factors of interest and be large enough for the algorithm to learn clinically important interactions between these features. [40] For example, an algorithm trained on a cohort that consisted only of men under 40 years of age and women over 65 would be unable to learn how age and gender interact.

Researchers may lack the resources to collect their own representative data sets, especially during fast-moving crises such as COVID-19. To address this, many authors suggested that pooling publicly available data sources could support more research with representative samples that generalize to the wider population and are not only from affluent academic medical centers. [4,158] Some collection efforts are already underway, including the Google Baseline Study[159] and the National Institute of Health's All of Us Research Program that is collecting nationally representative data from 340 sites and already has more than 175 000 biospecimens and 112 000 electronic health records. [160] Public data sources could be used as benchmarks to directly compare algorithm performance across diverse patient cohorts. [90] These data sets would also help coordinate across global issues that affect many countries, such as adverse reactions to the COVID-19 vaccination. [161] Finally, more data sharing would decrease disparities in who has access to data. For instance, more than 170 countries have no available ophthalmologic images for training algorithms, which precludes algorithm development specific to patients from these regions. [158]

Technical methods can mitigate some data inadequacies and are especially helpful when data are unavailable for privacy reasons. Federated learning enables model coefficients to be calculated locally and combined without sharing the underlying data, which means models can be trained on larger and more diverse data sets, without private health information leaving a protected system. [64] Transfer learning can be used to ensure that models have a basic framework of clinical facts (such as heart disease risk increasing with age) that do not have to be learned anew in each model. This enables models to be trained with smaller data sets. [64] Similarly, synthetic data sets can be constructed to protect privacy, resample data to provide more weight to minority classes (this mitigates model overfitting to data from the majority class), [162] or incorporate data from other sources to minimize bias. [50]

Even when data are representative, they may still contain biases from the data-generation process. For instance, outcome labels may be created by biased humans[49] or may be the product of inequitable processes, such as disparities in access to care. [10] Unfortunately, as no basic statistical checks exist for label bias, developers would need to carefully examine any bias-creating mismatch between the real-world outcomes they are trying to predict and the specific variable or label their model is trained to predict. [156] Developers can also establish processes that automatically build label bias checks and data updates into the development pipeline. [55] Some researchers suggested that certain data sources or variables are inherently less likely to be biased, such as noninvasive medical monitoring devices[40] or diagnoses. [36] However, there will always be the possibility for biases to enter data. Medical devices might be more accurate or available for some groups, for example, and diagnoses rely on access to health care.

In some data sets, group identifiers may inappropriately aggregate dissimilar groups or may be missing entirely. This reduces the accuracy of results and obscures variation and may preferentially fit the majority group. [65] When the use of sensitive variables is appropriate, adopting more granular identifiers, such as additional race categories or inclusive gender, sex, and sexual orientation categories can improve accuracy and equity. [163] For instance, adding 2 additional racial or ethnic groups to an equation that originally only distinguished between Black and non-Black patients for predicting chronic kidney disease resulted in lower levels of bias for some Asian populations. [63] If individual group identifiers are not available, then using imputation or aggregated statistics may be better than omitting these variables entirely. [156] Some demographic data may be imputed from surname data using validated algorithms, for example, or from patient address, given that socioeconomic data are often available at the census block level from administrative surveys. [156]

If all relevant features are not included in a data set, then algorithms can draw spurious conclusions, as part of the explanatory power of omitted variables is assigned to other correlated variables. Several studies emphasized the importance of including social determinants of health, including employment, education, and the quality of the built environment, in the data used by AI applications. [58] Including social determinants of health improves model performance and helps to identify the root cause of disparities. [164] For instance, because race is correlated with income, if income is excluded from a model, then poor outcomes may wrongly be attributed to race alone, degrading predictive performance. Some organizations have recognized this need and begun collecting new data as a result: UnitedHealthcare has helped to develop new International Classification of Diseases codes relating to social determinants of health that can be integrated within the data inputs used in their predictive models. [165]

Exclude Sensitive Variables to Correct for Bias

The question of whether and how to incorporate sensitive variables in an AI model is highly contested in the literature. Excluding sensitive variables from AI models has been proposed as a strategy to reduce algorithmic bias; the opposite remedy—including sensitive variables in models in which they are lacking—is also proposed. Much depends on the context of AI model use and the perspective of the author.

As discussed in the previous chapter, including sensitive variables in AI applications may lead to negative impacts on health equity, at least under certain circumstances. To address this issue, researchers and others have proposed excluding race or other sensitive variables from AI algorithms, a strategy that is sometimes discussed in academic literature as "fairness through unawareness" [18,166,167] or "fairness through blindness." [168]

This strategy might be pursued in several ways. AI developers and health research institutions might voluntarily choose to exclude sensitive variables from AI applications. [61,83,169] Researchers can highlight cases of existing algorithms that use sensitive variables in ways that exacerbate disparities. [61] Clinicians or other users might advocate for removing sensitive variables from algorithms or for their health systems to stop using algorithms that include these variables. [61,170,171] As a result of such efforts, updated versions of prominent AI applications that no longer use race have been released, including calculators used to estimate kidney function[80,172] and outcomes for vaginal birth after cesarean section. [169,173]

Policy makers might also issue guidelines or establish and enforce rules barring the use of sensitive variables in AI algorithms. In some cases, existing rules forbid consideration of an individual's race or other sensitive information in resource allocation or other government decision-making, whether those decisions are made by humans or based on AI algorithms. The Healthy Places Index used in California's COVID-19 vaccination campaign, for example, deliberately excluded race as an input variable to comply with California Ballot Proposition 209 passed in 1996. [89] At least 1 recent legislative proposal in the United States has focused on regulating the use of race and other sensitive variables specifically in AI algorithms, although it has not yet become law. [174,175]

Putting this strategy into practice generally requires finding an alternative algorithm or rebuilding an existing algorithm to exclude sensitive variables. In some cases, sensitive variables might be replaced with other data that are more directly relevant to patient health. [19,83] Researchers have cautioned against simple workarounds, such as continuing to use an already built algorithm and removing the sensitive variable from the input (or similarly entering a standard answer for all patients), as this can concentrate errors for particular groups and can exacerbate disparities in care. [80,82]

The effectiveness of this strategy depends on whether excluding sensitive information will reduce an algorithm's ability to discriminate. Excluding sensitive variables does not necessarily prevent algorithms from discriminating against marginalized groups, given that other, less visible but more significant, sources of potential bias exist. [18,60,156] Algorithms can learn to discriminate using patterns associated with seemingly less sensitive variables, such as a person's address or zip code, which may nevertheless be highly correlated with race or other sensitive information, in an AI-enabled version of redlining. [168] Researchers have demonstrated that deep learning AI models can infer a patient's race even from apparently innocuous data, including images from x-rays or CT scans. [176] Similarly, other algorithms use deep learning to identify people with mental health conditions using audio recordings of their voices. [177]

Include Sensitive Variables to Correct for Bias

This strategy is the opposite of the one described previously. Each may apply in different contexts and both strategies are contested.

Yet there are also some areas of general agreement. Even some advocates of excluding sensitive variables acknowledge that there are cases in which sensitive data might be used in algorithms in ways that advance equity, such as in epidemiologic research[61] or in applications intended to support affirmative action. [175] Similarly, we found few criticisms of the use of sensitive variables in evaluating disparities in model performance, analyzing data limitations, or enforcing fairness in model design. We also found few criticisms of AI algorithms that used sensitive variables when they had a clear and accepted connection to the health condition of interest, such as skin pigmentation level when used to assess melanoma risk, or patient sex when examining risk of diseases that disproportionately affected women rather than men.

Some researchers make a much broader case in favor of including sensitive variables in AI applications in health. [59] Several studies have found that including gender or race in training data for AI can improve accuracy in a wide range of applications, including in facial recognition, [178] estimates of life expectancy, [89] and risk of contracting or dying from COVID-19. [179-181] For this strategy to be effective in reducing disparities, however, any improvements in algorithm performance must ultimately contribute to improved care for disadvantaged groups.

Model Design

This section covers steps that developers can take during the model design process to minimize potential negative effects of health AI on equity. Some strategies are about the design of the algorithm itself. AI algorithms are designed to maximize objective functions, typically predictive accuracy, but they can incorporate other metrics as well, such as fairness. Models can be modified to produce fairer outcomes, whether through data editing, different objective functions, or postprocessing. However, this requires quantitative definitions of what fairness means, a subject of much debate. [97,166,168] Model design also determines if we can understand how an algorithm reached a decision. After an algorithm has been created, developers and evaluators can test its equity impacts, whether in training data, synthetic data sets, or in real-world implementation. The results from these evaluations can be used to inform decisions on whether to retain, revise, or retire a model.

It can be complex to balance equity and accuracy, identify model limitations, and rigorously evaluate equity impacts. Checklists and other tools can provide validated frameworks for these tasks and integrate them into the development workflow.

Enforce Fairness Goals

Fairness considerations can be incorporated directly into model design, through editing data inputs, building equity into the objective function of the algorithm, or postprocessing of model outputs. [65,97] The key aspects of these approaches include identifying the relevant equity norm for the proposed application, formulating a statistical interpretation of this norm (to facilitate measurement), and enforcing the equity norm on the AI model, usually during training or via postprocessing. [18,60,102] Identifying the relevant equity norm can be difficult because there are myriad defensible forms of fairness. [97,166,168] It often requires multidisciplinary, multistakeholder consultation to settle on a definition of equity that is satisfactory to all parties and is feasible given market, legal, and resource constraints.

There are statistical translations of most unambiguous equity norms. The use of statistical measures is pivotal because it enables developers and evaluators to measure and enforce compliance with a selected equity norm. A few approaches exist for enforcing equity norms, typically split into preprocessing, in-processing, and postprocessing. [182]

Preprocessing edits the data before they are used within the model. For instance, feature values can be edited to ensure similar marginal distributions for each group while preserving rank ordering, [182,183] "fairness through awareness," which presorts subjects by a task-specific similarity metric, [168] and fair representation learning, which finds a latent representation of the data that conceals protected attributes while preserving other aspects of the data. [65] The most direct approach uses in-processing to integrate a statistical formulation of the equity norm into the learning objective of the model. [182,184] One example is prejudice removal, meaning an algorithm designed to simultaneously maximize 2 objectives: the first is the accuracy of the prediction, and the second is a term that measures the fairness of the predictions (or the lack of prejudice, hence the name). [185] Other examples of this strategy are adversarial debiasing[186] and classification without disparate treatment. [187] Postprocessing strategies adjust the predicted probabilities of an algorithm in such a way that some fairness criterion is satisfied. [188] For example, the equalized odds procedure ensures that a model's true-positive and false-positive rates are equal across population subgroups (eg, defined by race or sex). [189]

Improve Interpretability or Explainability of Algorithm

Explainable AI (often referred to as XAI) aims to expose how AI algorithms reach their decisions. [190] Model mechanisms can be global (the overall logic of the model) or local (what features were important in determining a particular decision). [137] XAI methods fall into 2 groups: interpretability and explainability. Interpretability relies on developers choosing inherently interpretable AI algorithms, such as linear regression, decision trees, or k-nearest neighbors models. These are sometimes called *glass-box* approaches to contrast with unexplainable *black-box* approaches, such as neural networks, support vector machines, or random forests. [191] The trade-off is that these algorithms may be less accurate for some tasks[192] and maintaining interpretability limits the number of available parameters and precludes use of ensemble models. [35]

In the second approach, explainability, there are no constraints on which algorithms can be used. Instead, a second model is built to interpret the initial black-box algorithm. This is done by training an interpretable model on the output of the black-box model. Global explainability is then derived from the aggregation of local explainability. [191] In its simplest form, this could be training a decision tree on the output labels from a neural net. There are also several software packages available for XAI. Shapely Additive exPlanations [193] generates feature importance that approximates the impact of excluding that variable from the model and can be linearly summed to interpret their contribution to a model prediction. [193] Local Interpretable Model-Agnostic Explanations fit local linear models by perturbing inputs to estimate feature importance for an instance. [194] Although these methods can be flexibly applied to any model, transitioning from local to global explanations can be difficult, and explainability approaches can be computationally intensive for large data sets. [195]

XAI methods fulfill several needs. First, they provide a mechanism to examine fairness because humans can review the decision-making process, instead of only the output. [196] Second, they can help identify if the model is making decisions on spurious correlations, potentially improving accuracy. [195]

For instance, 1 large ML model found that asthma was a protective factor for pneumonia, contradicting medical understanding. On examination of the model with rule-based learning, it was found that asthma was only protective because it resulted in quicker admission to the intensive care unit. [192] Finally, interpretable algorithms may increase adoption because decision-makers and users feel more comfortable using models they understand. [197]

XAI can be used together with other strategies to address equity concerns. One group of researchers examined racial disparities in knee pain and found that their algorithm could explain significantly more disparities than radiologists, potentially improving access to treatment for those who need it most. Their use of XAI methods enabled them to identify and visualize particular x-ray image features associated with knee pain that radiologists had not found an image-based explanation for on their own. [198]

Evaluate Disparities in Model Performance

Several software tools are available to explore model fairness, such as AEquitas, Fairlearn, and FairLens. [101,199] Most automate the process of stratifying evaluations by specified variables, and some adopt aspects of explainable AI to show variable contributions to an individual prediction. [200] These tools lower the barriers to exploring bias and measuring algorithmic fairness. [195]

Reporting model performance across different groups can be a direct way to understand a model's equity implications and the range of populations over which the model results can be trusted. The most common approach is stratified testing, in which model metrics are evaluated separately in each subgroup of interest. [182,201,202] These tests evaluate how well the model generalizes to different groups and highlight where bias is introduced. [35] Typically, researchers evaluate results across protected categories such as race or gender, sometimes in addition to other variables, to build a more complete picture of if and how bias enters the model. Some researchers evaluate outputs across other attributes such as procedure type, [64] whereas, others advocate for integrating socioeconomic variables into stratification to explore the interactions between poverty and group status. [6]

Raw accuracy percentages are not the useful metric to evaluate. Evaluators can also look at false-positive and false-negative rates for each group, [55] and, if possible, at downstream impacts such as clinical benefit. [64] It is important to investigate the impact of the model on equity in practice to directly measure an application's effect on patient outcomes. Model metrics are only intermediate proxies that can be confounded by unexpected behavior. For instance, an unbiased model may be applied unfairly by human operators or used to recommend a treatment course that is only effective for some groups. [60] Evaluations that do not look at model performance in implementation will miss these aspects of fairness.

Several approaches complement stratified testing to explore model fairness. Variables can be omitted to see how model predictions change in the absence of protected categories. [203] Evaluators can examine whether model outputs are statistically independent of the omitted variable to ensure that the algorithm is not reconstructing the omitted variable from other available information. [40] An alternative approach is to swap variables to explore their impact: for instance, analyzing how someone would be treated differently if only their race changed. [204] Similarly, "matched case-control" data sets involve creating test data in which potentially biased variables have the same distribution across groups; if predictive power is decreased, then this implies a risk of bias. [64] Finally, explainable AI methods (see section titled Improve Interpretability or Explainability of Algorithm) may identify whether models are learning causal mechanisms or merely extrapolating from biased correlations. [205]

Evaluations can occur during AI development, deployment, and continuously over the life cycle of an AI application because the target population, data-generating process, and implementation practices adopted by users can shift over time. [40,69,123] The performance of AI models can degrade silently if inputs shift away from their training conditions. [102] The model itself may cause changes in the data if users strategically react to it or if model outputs impact model inputs (for instance, if the model is used to allocate resources). To address this, developers can use directed graphs to think about the causal structure of their model[64] and put processes in place to mitigate the impact of feedback loops. [35] Evaluations can be made more rigorous by complementing them with a clinical trial, [3] investigating algorithm impacts across multiple implementation sites, [91] or inviting independent third parties to conduct evaluations (potentially blinding developers to protected groups during this process). [206] In some cases, external evaluators may be able to evaluate disparities without having direct access to the internal workings of an algorithm, instead relying solely on examining model inputs and outputs that are available to the user. [10] There are multiple examples of this strategy being put into practice and resulting in changes to algorithms used in health care. [46,172,207]

Use Equity-Focused Checklists, Guidelines, and Similar Tools

Checklists are a simple behavioral intervention that have been shown to improve safety outcomes in complex procedures, such as surgeries. [18] In health care, AI checklists can be designed for developers, external reviewers who are evaluating models, health care providers using an application, or patients who want a framework to understand algorithm output. [208] Checklists can either be general or specific to certain application types, such as a 5-point checklist suggested for assessing racial equity impacts of AI models used in care for patients with diabetes. [59] Among developers, widespread demand exists for equity and fairness checklists that incorporate easily into the development workflow. [49]

Some checklists are designed to prompt thinking about the wider role of AI in promoting equity. One 7-part framework is designed to incorporate equity considerations throughout the AI development life cycle, from problem conception through maintenance and monitoring. [110] Another is designed for institutions to evaluate their use of AI entirely, advising organizations to inventory all algorithms in use, screen for bias, improve or retire biased algorithms, and put processes in place to prevent future bias. [156]

Deployment Practices

Once algorithms are implemented, they can take on a life of their own by influencing data inputs, being applied in unexpected ways, or being taken into new contexts. Thorough model reporting efforts can help to ensure that models are used responsibly, with limitations in mind, in approved applications and in populations similar to training data.

It is inevitable that some algorithms will cause harm. When this happens, especially when harms and benefits are inequitably distributed, it has been suggested that algorithm use should be reviewed and accountability mechanisms put in place. Health care stakeholders can play a role in monitoring who is able to use beneficial AI applications and take steps toward improving access.

Increase Model Reporting and Transparency

Many strategies to address equity issues require information about model structure, training data, and implementation. However, several reviews found that many model developers do not disclose basic information relevant to health equity, such as the demographic composition of the training data. [90,209] Ensuring that models are only used in appropriate populations and evaluating their risk of bias is effectively impossible without basic data reporting. [35]

One commonly advocated solution to this problem is to establish reporting standards. [4,94,210] MINIMAR (MINimum Information for Medical Al Reporting) defines the minimum set of information necessary to interpret model output and understand limitations. MINIMAR includes the study population and setting, demographic data (including protected characteristics), the model architecture, validation, and evaluation. [211] Several different preexisting standards, such as CONSORT (Consolidated Standards of Reporting Trials), SPIRIT (Standard Protocol Items: Recommendations and Intervention Trials), and TRIPOD (Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis) are extending their frameworks to specifically include reporting guidelines for AI models. [211] Other frameworks, such as PROGRESS (Place of residence, Race/ethnicity/culture/language, Occupation, Gender/sex, Religion, Education, Socioeconomic status, and Social capital), have been designed to organize reporting on population subgroups relevant to equity. [212] Some of these frameworks also encourage reporting of errors, which could include high error rates in certain groups. [68]

Model information need not be burdensome to report or complex to interpret. Data sheets and model cards have been proposed to aid in the dissemination and interpretation of model information. Data sheets characterize model training data and therefore indicate to which populations the model can generalize. Model cards are short, structured documents that report information on an AI algorithm for potential users. In addition to reporting basic model and data information, they can also record benchmarked evaluations across different population subgroups, a list of contexts in which model use is intended, and known limitations of model performance. [162]

Seek or Provide Restitution for Those Negatively Affected by AI

It is inevitable that as AI applications become more widespread, some will cause harm to users. Some authors have considered the possibility of allowing users to pursue restitution for any harms to which the use of AI models in health care contribute. [213-215] This strategy is proposed to serve 2 purposes. First, it helps compensate users for any harms they suffered. Second, it creates an incentive structure that punishes AI developers or users who are responsible for harms, hopefully leading to more accurate and equitable AI development. [213]

Torts were the framework most advocated in the literature as a means for providing restitution, in which users would be able to sue developers if they were subject to foreseeable and avoidable harm. This framework may not easily transfer to AI because harms may not be foreseeable by developers, and physicians may not understand or control AI systems. [215] Potential remedies include making AI algorithms "persons," with a similar status to physicians, distributing liability among all actors, or modifying the standard of care so that physicians who use these algorithms would have a duty to validate them. [215] Ultimately, some authors argued "companies need to bear responsibility for the AI products they create, even when those products evolve in ways not specifically desired or foreseeable by their manufacturers." [213]

Avoid or Reduce Use of AI

AI applications may be inappropriate in some settings or for some uses, although whether and when this might be the case was an area of disagreement in the literature. Before development takes place, risk assessments can be conducted to help determine if an AI application is warranted. If there are significant risks, developers may choose to refrain from designing or fielding an AI system. [123] Similarly, keeping humans in the loop[216] or discontinuing algorithm use can always be an option to respond to equity concerns. Avoiding or reducing use of AI may be necessary if an algorithm is found to exacerbate health disparities and attempts at improving it have been fruitless. [156]

Some have argued that the problems with ML applications are so severe that their use should be prohibited for critical decisions such as treatment allocation. As one author has argued: "We must not allow certain key decisions to be left solely to the predictive output of artificially intelligent algorithms… there must be a statutory prohibition on utilizing it for any type of predictive or substantive decision-making." [217] Other researchers have speculated that the public is only just becoming aware of problems with AI decision-making and that they will increasingly advocate for diminished or at least cautious use of AI. [16,218]

Provide Resources to Those With Less Access to AI

Several articles argued that resources must be devoted to addressing inequities in access to AI and digital health technologies. Often, these proposals took the form of ensuring that the infrastructure critical to accessing AI technologies, such as broadband internet, was available and affordable for disadvantaged groups. [40,118] Paying for local digital infrastructure, such as low-cost phone plans or free wireless internet, might require subsidies from governments, multinational corporations, or international organizations.[118] In high-income countries, access might be enhanced by regulating that health insurers provide coverage for AI-based personal health technologies, such as smartwatches, if these technologies were found to have significant benefits.[219]

There are also gaps in digital health literacy, with lower levels among those who have little formal education and no employment. This may hinder individuals' ability to fully use and benefit from AI-powered health technologies.[220,221] Some authors have called for digital literacy education programs, so that socioeconomically disadvantaged individuals are better able to benefit from health AI applications.[40,222]

Another approach is to change the designs of AI products themselves to better suit low-resource contexts. Mobile apps might be designed so that they do not require internet access, for example.[118] Similarly, AI apps could be developed for devices already available in low-income country contexts, with interfaces tailored to these users.[119]

Several authors highlighted the importance of supporting knowledge development and transfer. The use of open data, publicly available code,[113] and supporting collaboration between developers in higher and lower income countries can contribute to effective knowledge sharing.[119] Some programs designed to help develop and use local expertise are already running. The Global South AI4COVID Program is providing grants to 8 organizations in lower and medium-income countries to design AI algorithms that contribute to the local response to the COVID-19 pandemic.[223]

Use of AI to Address Preexisting Inequities

AI applications can also be used to address long-standing problems of health inequity whose causes are rooted in societal issues that are independent of the use of AI in health care. These long-standing challenges include:

- The unequal incidence of disease, risk factors, and other health disparities
- Disparities in health-related beliefs and behaviors
- Unequal access to health-related services
- Biased human decision-making

Even when imperfect, AI may provide opportunities to reduce preexisting inequities, whether through making less biased decisions than standard alternatives, exploring where disparities exist, identifying disadvantaged individuals, or enabling new functionality to address disparities. These strategies are discussed as follows.

Develop and Field AI That Is Less Biased Than Standard Decision-Making

Much of the AI equity literature represents a reaction to naive views from early AI developers that algorithmic decision-making would eliminate discrimination. However, some researchers point out that although our understanding of the limitations of AI has evolved, we may at times be holding algorithms to unreasonably high standards, in which any discrimination is deemed unacceptable. [224] To reduce discrimination, AI systems do not have to be perfect; they just need to be better than current flawed systems, whether those are based on human decision-making or deterministic, non-AI algorithms. [60] Surveys show that many members of the public believe that AI decisions could be fairer than human decisions. [225,226]

AI algorithms have 3 potential advantages over other decision-making methods. First, with proper curation, they may be able to avoid some of the biases and preconceptions often present in human decision-making. [197] Second, algorithmic predictions may be more accurate than other methods because they can flexibly approximate complex functional forms. Increased accuracy is often beneficial for marginalized groups who are the least likely to be accurately assessed by other methods. [227] Finally, it may be easier to tell if an algorithm is biased than if a human is biased. [95] People may not understand or be willing to share why they made certain decisions, and it can be difficult to construct controlled settings where only protected status varies. In contrast, algorithmic performance can be measured and replicated with synthetic data. [95]

Use AI to Examine Disparities

AI can improve our understanding of disparities by analyzing large data sets to identify which factors are associated with disparities in health outcomes. For example, we found a great deal of research that applied ML algorithms to predict COVID-19 outcomes at the individual or community level using equity-related predictors such as the built environment, social determinants of health, and environmental pollution. [228-235] Some of these analyses found that socioeconomic and structural factors, such as education level, distance from the nearest hospital, and hospital funding model, were more predictive than biological factors such as comorbidities. [236] Other research focused on the secondary impacts of the pandemic, identifying which groups saw the greatest mental health impacts, [237] or examining the underlying reasons for vaccine hesitancy. [238] Although these studies are largely not causal, they are useful because they highlight factors that indicate higher need for COVID-19 care interventions and provide future research direction for causal studies.

AI algorithms can also be used to generate new data sets or representations of equity-related data that would otherwise be difficult to analyze. For instance, papers that investigated the impact of the built environment on COVID-19 outcomes used computer vision techniques, either on Google Street View images[239] or on satellite imagery, [58] to categorize the built environment. Similarly, natural language processing (NLP) techniques enabled the analysis of physician notes to identify differences in COVID-19 symptoms and prognosis by gender. [240] NLP was also used to help directly model disparate treatment of health researchers: responses to grant applications for male and female applicants were analyzed to show that reviewer comments were biased along gender lines and could not be explained by subsequent productivity. [241]

Causal ML techniques, although less commonly applied, can estimate heterogeneous treatment effects, ie, the treatment effect for different subpopulations. This is particularly useful when resources are scarce because they can be assigned to the subpopulations that benefit most. Heterogeneous effect estimates can also be used to promote equity by identifying when interventions that are beneficial on average might widen disparities because they are most beneficial for privileged groups. [125]

Use AI to Identify People Facing Disparities

Some groups can be hard to reach and may not receive all the services they are entitled to or they may have specific needs that are overlooked by standard processes. [4] AI algorithms can flexibly incorporate many features to aid with identification and service delivery for these individuals. During the COVID-19 pandemic, for example, the state of California used AI-based analysis to identify the zip codes of residents facing health disparities, which were then prioritized for vaccine allocation. [89,242] Researchers have developed AI algorithms to help identify homeless people, [243] people with psychiatric or physical disabilities, [244] women at risk of domestic violence, [244] and children at risk of lead poisoning. [245]

In some cases, improvements over existing assessment tools were dramatic. An AI-based prediction tool for adverse birth risk in pregnant women outperformed the existing paper-based method by 36%. Many of the gains were preserved by translating the AI model calculations into a new paper-based method. The new assessment still showed a 22% improvement, without the need for maintaining a complex AI system. [246]

Private companies are also using ML for the identification of disadvantaged individuals. For example, UnitedHealthcare, a large insurance company, is using AI algorithms on claims data to identify who may need support with necessities such as housing or food. [247] Whether algorithms such as these improve equity depends on whether positive impacts from service linkages outweigh any biases present in who is identified for preferential resource allocation.

Use AI to Reduce Disparities

AI may reduce disparities when it is applied directly to mitigate inequities in access to care, when it is disproportionately suited to marginalized populations or low-resource contexts, or when it reduces biases in existing decision-making.

AI can also solve complex many-to-many mapping problems, some of which are useful in directly addressing equity by improving access to health services, such as translation applications or chatbots. One application, RadTranslate, was developed by researchers who wanted to improve diagnoses and patient experience among the disproportionately large group of COVID-19 patients with limited

English proficiency. Physicians created examination instruction scripts that were translated by AI text-to-speech algorithms to provide standardized instructions in a patient's language of choice. In a field trial, the intervention was found to reduce the variance in length of appointment time between English- and Spanish-speaking patients. [248] Similarly, AI can be trained as a bridging tool between different dialects, helping patients and physicians to understand each other more quickly. [38] In another example, an AI chatbot was developed to help patients with asthma who had recently lost their jobs to navigate the process of regaining insurance coverage. [249]

AI may provide a particularly large benefit in low-resource settings. [250] For instance, when few specialist staff are available, AI algorithms might be used to support diagnostic decision-making, lowering the burden on the health care system and reducing unnecessary hospital visits. Potential applications included arrhythmia detection, [251] cancer screening, [252] and monitoring of Parkinson disease. [253] Many applications of AI in low-resource settings are still theoretical, and it is questionable whether the infrastructure requirements for effective care are lower or merely different in these settings. Although AI diagnosis and monitoring algorithms may increase the efficiency of health care professionals and enhance their abilities, they require much effort and expertise to train, validate, operate on data from remote populations, troubleshoot problems, and educate users on their limitations.

Summary

We identified 15 strategies proposed in the literature to improve the impact of AI on health equity. These strategies are proposed to address 1 or more of the equity-related issues discussed in the previous chapter. We also identified 4 proposed strategies that use AI to proactively address preexisting health disparities.

The next chapter provides additional detail on how the strategies discussed previously are linked to the issues presented in Chapter 3.

5. Linking Issues and Strategies

In this chapter, we explore how issues and strategies are linked. Some strategies are much more prominent in the literature than others; the strategies most frequently linked to each issue are shown in Table 5.1. This table is provided as a useful starting point for stakeholders to identify appropriate strategies that might be used to address particular issues raised by the use of AI in health care. A full list of reviewed documents that propose use of a particular strategy to address a specific issue is provided in Appendix A.

A small number of issues make up the majority of mentions in the literature: the top 4 issues constitute over half of all issue-strategy pairs. These issues tend to have several well-developed strategies, usually focused on improving the quality of data or evaluating bias in model decision-making. In contrast, some issues are mentioned infrequently and do not have well-developed strategies. When only 1 strategy was linked to an issue in the literature we reviewed, the second column in Table 4.1. is left blank. We include an issue frequency column as a rough measure of how often issues were addressed in the literature.

Table 5.1. The Most Common Strategies Mentioned in the Literature for Each Health Equity Issue

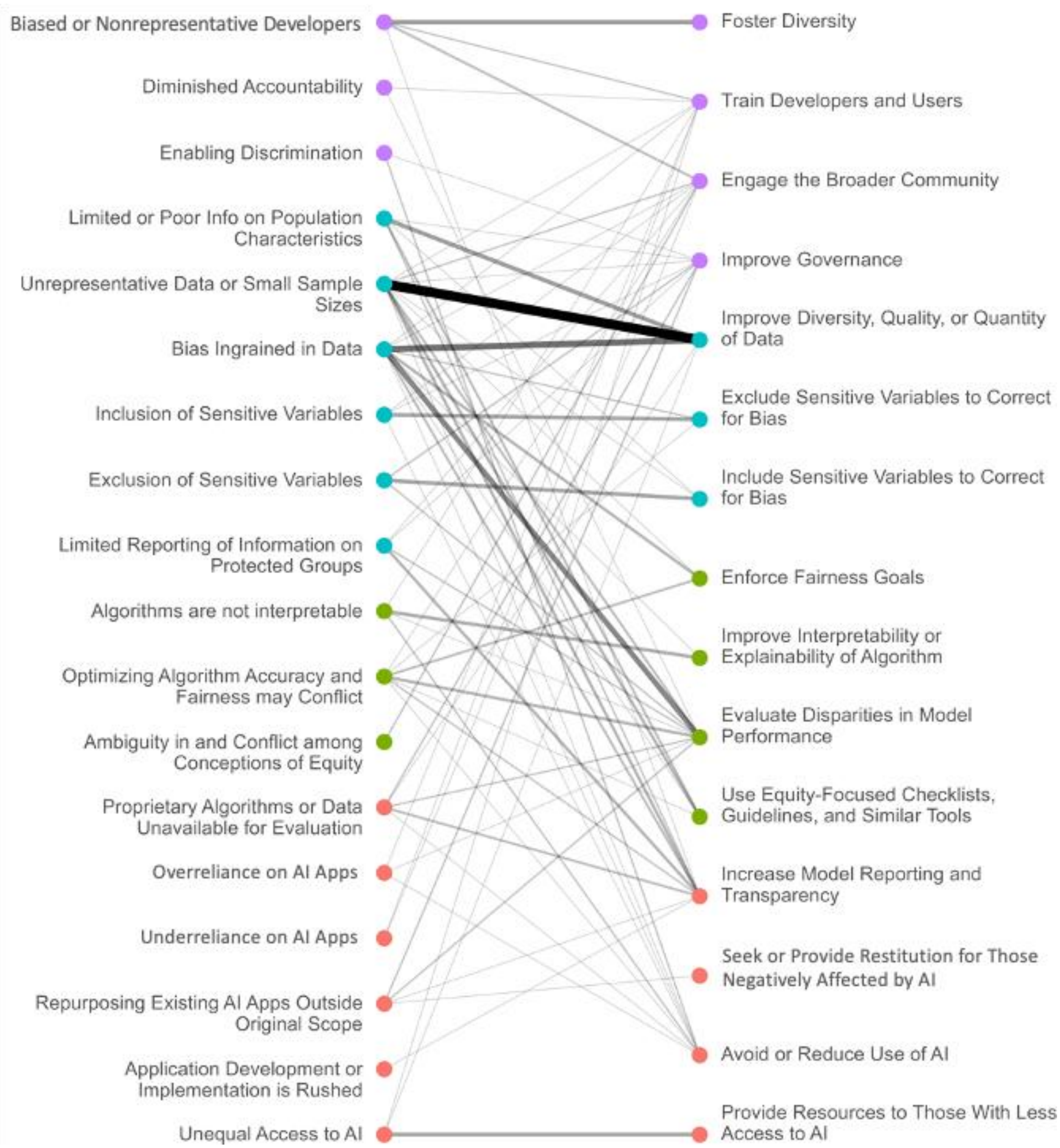
	Issue	Frequency of issue mention in literature	Most commonly linked strategy	Second most commonly linked strategy
Background context	Biased or nonrepresentative developers	13/195 = 7%	Foster diversity	Engage the broader community
	Diminished accountability	2/195 = 1%	Evaluate disparities in model performance	Train developers and users
	Enabling discrimination	3/195 = 2%	Avoid or reduce use of AI	Improve governance

	Issue	Frequency of issue mention in literature	Most commonly linked strategy	Second most commonly linked strategy
Data characteristics	Limited information on population characteristics	14/195 = 7%	Improve diversity, quality, or quantity of data	Use equity-focused checklists, guidelines, and similar tools
	Unrepresentative data or small sample sizes	46/195 = 24%	Improve diversity, quality, or quantity of data	Increase model reporting and transparency
	Bias ingrained in data	36/195 = 19%	Improve diversity, quality, or quantity of data	Evaluate disparities in model performance
	Inclusion of sensitive variables	9/195 = 5%	Exclude sensitive variables to correct for bias	Avoid or reduce use of AI
	Exclusion of sensitive variables	10/195 = 5%	Include sensitive variables to correct for bias	Evaluate disparities in model performance
	Limited reporting of information on protected groups	8/195 = 4%	Increase model reporting and transparency	Evaluate disparities in model performance
Model design	Algorithms are not interpretable	9/195 = 5%	Improve interpretability or explainability of algorithm	Avoid or reduce use of AI
	Optimizing algorithm accuracy and fairness may conflict	13/195 = 7%	Evaluate disparities in model performance	Enforce fairness goals
	Ambiguity in and conflict among conceptions of equity	2/195 = 1%	Engage the broader community	-

	Issue	Frequency of issue mention in literature	Most commonly linked strategy	Second most commonly linked strategy
Deployment practices	Proprietary algorithms or data unavailable for evaluation	9/195 = 5%	Increase model reporting and transparency	Evaluate disparities in model performance
	Overreliance on AI apps	3/195 = 2%	Avoid or reduce use of AI	Evaluate disparities in model performance
	Underreliance on AI apps	2/195 = 1%	Engage the broader community	Train developers and users
	Repurposing existing AI apps outside original scope	6/195 = 3%	Evaluate disparities in model performance	Improve governance
	Application development or implementation is rushed	1/195 = 1%	Increase model reporting and transparency	–
	Unequal access to AI	8/195 = 4%	Provide resources to those with less access to AI	Improve diversity, quality, or quantity of data

Figure 5.1 is a map of the 195 issue-strategy pairs we identified in the literature. Each paired issue and strategy mentioned in the literature is connected with a line. Bolder lines indicate strategies and issues that are more frequently linked. A comprehensive list of the documents that link a proposed strategy to a particular issue is provided in Appendix A. This figure highlights which strategies are most commonly proposed to address specific equity-related issues raised by the use of AI in health care.

Figure 5.1. Mapping the Connections Between AI Health Equity Issues and Strategies Proposed to Address Them



Note: Issues are depicted on the left of the figure, whereas strategies are listed on the right; the thickness and opacity of each line connecting an issue to a strategy are proportional to how frequently they were mentioned together. Colors represent the associated development stage: Background context (purple), data characteristics (blue), model design (green), and deployment practices (red).

Summary

In this chapter, we depict the connections between the issues described in Chapter 3 and the strategies described in Chapter 4. This provides several insights. First, although some issues have a single dominant strategy, most are linked to multiple solutions. Second, the strongest links tend to be within the same phase of the development life cycle. Third, no dominant strategy is a panacea to all equity issues. Fourth, much of the literature focuses on data issues and solutions, perhaps because these are seen as tractable problems.

The following chapter discusses the implications of these findings for health care stakeholders seeking to improve the impact of AI on health equity, as well as opportunities for future research.

6. Discussion

By analyzing the literature on AI and health disparities, we have identified 18 broad issues and 15 strategies that can be used to address them. This builds on frameworks from the existing literature, identifying specific strategies and issues associated with 4 stages of AI development and implementation. In addition, we draw 3 new insights from mapping the relationships between issues and strategies.

The Literature Focuses on a Small Set of Issues

A small set of issues dominate the literature. Much of the discourse around health AI equity focuses on data characteristics: almost two-thirds of all issue-strategy pairs we identified in the literature are related to data. These issues are complex and widely researched, and several strategies have been proposed to address them. Some strategies directly address data quality, whereas others take data limitations as a given and try to produce fair algorithm results despite poor data quality. Which approaches are most fruitful will depend on the feasibility of improving data collection.

Much of the literature on model design focuses on the trade-off between accuracy and fairness. [100,254-257] This multifaceted issue includes selecting from competing definitions of fairness and understanding the extent of any trade-offs between model accuracy and fairness. The most frequently advocated approach requires measuring disparities in model performance and revising the model if large disparities are detected. Developers can also potentially build models to optimize fairness, [55] although because definitions of fairness may conflict, developers and evaluators may need to test the impact of different constraints across a broad range of metrics, such as accuracy, false-positive rate, and false-negative rate for different population subgroups. [60]

Some issues were rarely discussed and have a limited number of associated strategies. Several issues reflect concerns about how AI is deployed, especially when AI applications are used outside their original scope, or when they are rushed through development and implemented without sufficient evaluation.

Even if an issue is not frequently discussed in the literature, it may still be important. An issue could be infrequently cited because of limited evidence of equity impact or because corresponding strategies are underdeveloped. Less discussed issues such as the repurposing of AI applications outside their original scope, the impact of insufficiently detailed population characteristics, or diminished accountability are all rich topics that would benefit from future research.

Strategies Are Multipurpose

Although some strategies, such as improving interpretability, are tailored to specific issues, most strategies are multipurpose in that they have been proposed to address several different equity-related issues raised by the use of AI in health. The top 4 most frequently mentioned strategies, which account for more than half of the issue-strategy pairs we found in the literature, are collectively linked to all 18 issues. Each of these strategies is linked to critical aspects of application development. These 4 strategies were:

- Evaluate disparities in model performance
- Improve diversity, quality, or quantity of data
- Engage the broader community
- Improve governance

Evaluating disparities in model performance is often necessary for quantifying bias across subgroups. Similarly, improving data is important across a broad range of issues because the decision-making logic of AI models flows directly from the training data. Community engagement and improved governance can increase the consideration of equity issues throughout the development pipeline. Community stakeholders may be involved at all stages of AI development, including deciding whether an application should be built, setting goals for the model, defining fairness, [101] and guarding against unintended consequences after deployment. [3,38,128,130] Improving governance is usually advocated in the form of guiding principles for AI use[145,258] or "soft governance" such as industry-organized protocols. [18,151] Regulation is less frequently advocated, although it is unclear whether this is because researchers believe it would be ineffective or because they prefer to focus on technical, rather than policy, solutions.

Stakeholders Can Focus on a Small Set of Strategies

Sometimes it is only practical to focus on a small set of strategies. For instance, in their "Algorithmic Bias Playbook," Obermeyer et al suggest that organizations identify biased algorithms and then retrain them on less biased outcome data, improve the representativeness of their data set, or consider discontinuing use. [156]

The 4 most advocated strategies (Improve Diversity, Quality, or Quantity of Data; Evaluate Disparities in Model Performance; Engage the Broader Community; Improve Governance) are collectively linked to all 18 issues. This suggests that these strategies could be considered as a useful starting point to address the impact of an AI product on health equity. However, we do not recommend that stakeholders should focus exclusively on this set. Most issues are complex and are likely best addressed through multiple complementary strategies. In addition, all 18 issues may not always be present. Issues related to biased data or the trade-off between accuracy and fairness are likely to be ubiquitous, but issues related to algorithm interpretability or handling of sensitive variables may not be present.

Once the relevant issues have been identified, stakeholders may find it helpful to refer to Table 5.1 and Figure 5.1 to select a set of strategies proposed to address those issues in the literature. The most common strategies cited previously are a good starting point because of their broad coverage of issues. However, not all these strategies may be feasible, and others may need to be complemented by additional strategies to fully cover specific issues.

Study Limitations

A primary limitation of this analysis is that we do not rate the quality of issues, strategies, or the documents in which we identified issue-strategy pairs. Some sources go into detail about health equity issues and strategies; others only make general recommendations or may represent outmoded views. The goal of this analysis was to identify which issues and strategies are highlighted in the literature. Future work could instead focus on identifying the best or most developed strategies.

In addition, some issues and strategies conflict. For example, both inclusion and exclusion of sensitive variables are discussed as having either a positive or negative influence on the impact of health AI on equity, depending on context and perspective. As a result, we include these as both issues and strategies in our study, reflecting the unsettled and context-dependent nature of debate on this topic within the literature.

The issues and strategies we identified are not entirely distinct: some are intermediaries that lead to other issues or strategies. For instance, repurposing an application is not inherently inequitable but may increase the chance that the training data are unrepresentative of the target population. Similarly, uninterpretable algorithms do not create biased outcomes but can make them more difficult to detect. The same applies to strategies. For example, using equity checklists does not directly solve problems but makes it more likely that developers identify equity issues and appropriate strategies. We included these intermediary issues and strategies because they provide a richer description of intervention points for promoting health equity.

There are other prominent concerns about AI and equity that this paper does not cover. This includes the potential for AI applications to displace human workers in ways that could increase economic disparities and the potential that AI applications could reinforce harmful stereotypes, such as via the use of female personas in AI voice assistants that perform clerical or menial tasks. [122] Although these concerns are raised in the context of economic or social disparities, we found no discussion of their impact on health equity specifically, and thus we did not include them in our study. In addition, many of the issues and strategies we identified apply to non-AI data analytics used in health care, except for some considerations that are specific to AI model design.

Potential Future Areas of Research

No single solution exists for the health equity issues introduced or exacerbated by AI. However, there are multiple avenues of research, regulation, and practice that together could push AI algorithms toward promoting, rather than degrading, health equity.

- Governance structures for guiding and regulating the use of AI in health care are in flux. Additional research might examine how guidelines, regulation, and enforcement can incentivize AI development that improves health equity and promotes intervention when algorithms deliver undesirable outcomes.
- Implementations for some strategies are underdeveloped. Statistical approaches, such as enforcing fairness constraints, often have full implementation and even supporting software packages. In contrast, *softer* strategies, such as engaging the broader community or training developers, are frequently advocated without further detail on how these approaches might be implemented.
- In this report, we identified which strategies were most frequently advocated, not which strategies are best. Further work could identify best practices for each strategy, identify which strategies are most effective, and resolve conflicts between conflicting strategies, such as whether to include or exclude sensitive variables. Ideally, this would be a living document, updated as new techniques are developed and maintained by a credibly neutral consortium that includes a broad range of stakeholders.
- More research could help illuminate when certain strategies are most appropriate. Some strategies can only be implemented by certain stakeholders (often developers), and others require specific resources, such as the ability to collect additional data. Stakeholders, particularly nontechnical groups, could benefit from concise guides on how to support efforts to improve the impact of AI on health equity.
- Whether it is best to include or exclude sensitive variables when developing and evaluating AI models is an active subject of debate. Further research could better illuminate how and when using these variables are most likely to help improve health equity, and when their use might have negative impacts on health equity and thus should be avoided.

Some strategies may always be appropriately and feasibly applied to certain types of AI models. For instance, all categorization and regression algorithms on demographic data could benefit from rigorous evaluation of outcomes across population subgroups. Research into which other strategies are most feasible and effective could support future efforts to address equity-related issues raised by the use of AI in health care.

Conclusion

Our work contributes to a growing body of AI health equity literature. We add to this literature by identifying a more granular set of strategies and issues than prior work, creating a many-to-many mapping between strategies and issues, and by reviewing the literature to identify how often each strategy is linked to each issue. This analysis is useful for a wide array of stakeholders, including AI developers, users, policy makers, and researchers.

Although no strategy can fully address the equity concerns posed by the use of AI in health care, small sets of strategies can often mitigate many of the most pressing issues. We should also recognize that existing nonalgorithmic decision-making is imperfect. By thoughtfully adopting complementary sets of strategies that cover a broad range of equity issues, AI models may offer improvements in equity over the status quo.

Appendix A. Documents Proposing Strategies to Address Equity-Related Issues Raised by the Use of AI in Health Care

This appendix lists the issue-strategy pairs we found in the literature that we used as the basis for our analysis. Table A.1 provides a list of each document we reviewed that proposes a strategy to address a particular equity-related issue raised by the use of AI in health care.

Table A.1. Documents and the Strategies They Propose to Address Equity-Related Issues Raised by the Use of AI in Health Care

Document	Issue	Proposed strategy
Chin et al 2020[122]	Biased or nonrepresentative developers	Foster diversity
Leslie et al 2021[6]	Biased or nonrepresentative developers	Foster diversity
Vourganas et al 2020[50]	Biased or nonrepresentative developers	Foster diversity
Holstein et al 2019[49]	Biased or nonrepresentative developers	Foster diversity
O'Brien et al 2022[40]	Biased or nonrepresentative developers	Foster diversity
The Alan Turing Institute 2021[39]	Biased or nonrepresentative developers	Foster diversity
Zdawxzyk and Vallee 2021[37]	Biased or nonrepresentative developers	Foster diversity
Chin et al 2020[122]	Biased or nonrepresentative developers	Train developers and users
Ledford 2019[46]	Biased or nonrepresentative developers	Train developers and users
Leslie et al 2021[6]	Biased or nonrepresentative developers	Engage the broader community
Rajkomar et al 2018[3]	Biased or nonrepresentative developers	Engage the broader community
Zimmer et al 2021[38]	Biased or nonrepresentative developers	Engage the broader community
Holstein et al 2019[49]	Biased or nonrepresentative developers	Increase model reporting and transparency
Kiener 2020[53]	Diminished accountability	Train developers and users
Osoba et al 2019[18]	Diminished accountability	Evaluate disparities in model performance
Rose et al 2017[259]	Enabling discrimination	Improve governance
BBC News 2020[260]	Enabling discrimination	Avoid or reduce use of AI
Johnson 2022[261]	Enabling discrimination	Avoid or reduce use of AI
Ross 2021[57]	Limited or poor info on population characteristics	Improve governance
Morey et al 2022[262]	Limited or poor info on population characteristics	Include sensitive variables to correct for bias

Document	Issue	Proposed strategy
O'Neill et al 2014[212]	Limited or poor info on population characteristics	Increase model reporting and transparency
Ross 2021[57]	Limited or poor info on population characteristics	Increase model reporting and transparency
Deonarine et al 2021[58]	Limited or poor Info on population characteristics	Improve diversity, quality, or quantity of data
Kundu et al 2021[163]	Limited or poor info on population characteristics	Improve diversity, quality, or quantity of Data
Mhasawade et al 2021[19]	Limited or poor info on population characteristics	Improve diversity, quality, or quantity of data
Obermeyer et al 2021[156]	Limited or poor info on population characteristics	Improve diversity, quality, or quantity of data
Stevens et al 2011[63]	Limited or poor info on population characteristics	Improve diversity, quality, or quantity of data
The Alan Turing Institute 2021[39]	Limited or poor info on population characteristics	Improve diversity, quality, or quantity of data
UnitedHealth Group 2019[165]	Limited or poor info on population characteristics	Improve diversity, quality, or quantity of data
Mitchell 2019[162]	Limited or poor info on population characteristics	Use equity-focused checklists, guidelines, and similar tools
O'Neill et al 2014[212]	Limited or poor info on population characteristics	Use equity-focused checklists, guidelines, and similar tools
Pham et al 2021[59]	Limited or poor info on population characteristics	Use equity-focused checklists, guidelines, and similar tools
Kiener 2020[53]	Unrepresentative data or small sample sizes	Train developers and users
Morley et al 2020[128]	Unrepresentative data or small sample sizes	Engage the broader community
Reddy et al 2020[137]	Unrepresentative data or small sample sizes	Engage the broader community
US Food and Drug Administration et al 2021[150]	Unrepresentative data or small sample sizes	Improve governance
Peiffer-Smadja et al 2020[66]	Unrepresentative data or small sample sizes	Include sensitive variables to correct for bias
Vokinger et al 2021[35]	Unrepresentative data or small sample sizes	Enforce fairness goals
Peiffer-Smadja et al 2020[66]	Unrepresentative data or small sample sizes	Evaluate disparities in model performance
Rodolfa 2020[102]	Unrepresentative data or small sample sizes	Evaluate disparities in model performance

Document	Issue	Proposed strategy
Grote and Berens 2020[263]	Unrepresentative data or small sample sizes	Evaluate disparities in model performance
Park et al 2019[210]	Unrepresentative data or small sample sizes	Increase model reporting and transparency
Hernandez-Boussard et al 2020[211]	Unrepresentative data or small sample sizes	Increase model reporting and transparency
Vokinger et al 2021[35]	Unrepresentative data or small sample sizes	Increase model reporting and transparency
Wynants et al 2020[68]	Unrepresentative data or small sample sizes	Increase model reporting and transparency
Grote and Berens 2020[263]	Unrepresentative data or small sample sizes	Avoid or reduce use of AI
Arora 2020[155]	Unrepresentative data or small sample sizes	Improve diversity, quality, or quantity of Data
Chouldechova 2018[65]	Unrepresentative data or small sample sizes	Improve diversity, quality, or quantity of data
Röösli and Hernandez-Boussard 2021[4]	Unrepresentative data or small sample sizes	Improve diversity, quality, or quantity of data
Google undated[67]	Unrepresentative data or small sample sizes	Improve diversity, quality, or quantity of data
Liu et al 2019[71]	Unrepresentative data or small sample sizes	Improve diversity, quality, or quantity of data
Asiimwe et al 2021[62]	Unrepresentative data or small sample sizes	Improve diversity, quality, or quantity of data
Vourganas et al 2020[50]	Unrepresentative data or small sample sizes	Improve diversity, quality, or quantity of data
IEEE 2019[119]	Unrepresentative data or small sample sizes	Improve diversity, quality, or quantity of data
Sipior 2020[109]	Unrepresentative data or small sample sizes	Improve diversity, quality, or quantity of data
Gallifant 2021[90]	Unrepresentative data or small sample sizes	Improve diversity, quality, or quantity of data
Ferryman and Pitcan 2018[120]	Unrepresentative data or small sample sizes	Improve diversity, quality, or quantity of data
Seastedt et al 2021[264]	Unrepresentative data or small sample sizes	Improve diversity, quality, or quantity of data
Wilkinson et al 2021[265]	Unrepresentative data or small sample sizes	Improve diversity, quality, or quantity of data
Gianfrancesco et al 2018[75]	Unrepresentative data or small sample sizes	Improve diversity, quality, or quantity of data

Document	Issue	Proposed strategy
Mitchell 2019[162]	Unrepresentative data or small sample sizes	Improve diversity, quality, or quantity of data
Peiffer-Smadja et al 2020[66]	Unrepresentative data or small sample sizes	Improve diversity, quality, or quantity of data
Obermeyer et al 2021[156]	Unrepresentative data or small sample sizes	Improve diversity, quality, or quantity of data
O'Brien et al 2022[40]	Unrepresentative data or small sample sizes	Improve diversity, quality, or quantity of data
Paul and Anthony 2018[55]	Unrepresentative data or small sample sizes	Improve diversity, quality, or quantity of data
Rajkomar et al 2018[3]	Unrepresentative data or small sample sizes	Improve diversity, quality, or quantity of data
Rodolfa et al 2019[60]	Unrepresentative data or small sample sizes	Improve diversity, quality, or quantity of data
Khan et al 2021[158]	Unrepresentative data or small sample sizes	Improve diversity, quality, or quantity of data
Kings College London 2020[157]	Unrepresentative data or small sample sizes	Improve diversity, quality, or quantity of data
The All of Us Research Program Investigators 2019[160]	Unrepresentative data or small sample sizes	Improve diversity, quality, or quantity of data
Harwich and Laycock 2018[73]	Unrepresentative data or small sample sizes	Improve diversity, quality, or quantity of data
Vokinger et al 2021[35]	Unrepresentative data or small sample sizes	Improve diversity, quality, or quantity of data
Wynants et al 2020[68]	Unrepresentative data or small sample sizes	Improve diversity, quality, or quantity of data
Simonite 2022[266]	Unrepresentative data or small sample sizes	Improve diversity, quality, or quantity of data
Bozkurt et al 2020[267]	Unrepresentative data or small sample sizes	Use equity-focused checklists, guidelines, and similar tools
Leslie et al 2021[6]	Unrepresentative data or small sample sizes	Use equity-focused checklists, guidelines, and similar tools
Obermeyer et al 2021[156]	Unrepresentative data or small sample sizes	Use equity-focused checklists, guidelines, and similar tools
Osoba et al 2019[18]	Unrepresentative data or small sample sizes	Use equity-focused checklists, guidelines, and similar tools
Kiener, 2020[53]	Bias ingrained in data	Train developers and users
Kelly et al 2019[202]	Bias ingrained in data	Engage the broader community

Document	Issue	Proposed strategy
O'Neil 2016[72]	Bias ingrained in data	Exclude sensitive variables to correct for bias
Vyas 2020[61]	Bias ingrained in data	Exclude sensitive variables to correct for bias
Chouldechova 2018[65]	Bias ingrained in data	Enforce fairness goals
Chen et al 2021[36]	Bias ingrained in data	Enforce fairness goals
O'Brien et al 2022[40]	Bias ingrained in data	Enforce fairness goals
Rodolfa et al 2019[60]	Bias ingrained in data	Enforce fairness goals
Adebayo 2016[205]	Bias ingrained in data	Improve interpretability or explainability of algorithm
American Association for the Advancement of Science 2021[207]	Bias ingrained in data	Evaluate disparities in model performance
Valle-Cruz et al 2019[126]	Bias ingrained in data	Evaluate disparities in model performance
Obermeyer et al 2021[156]	Bias ingrained in data	Evaluate disparities in model performance
Osoba et al 2019[18]	Bias ingrained in data	Evaluate disparities in model performance
Parikh et al 2019[203]	Bias ingrained in data	Evaluate disparities in model performance
Rodolfa 2020[102]	Bias ingrained in data	Evaluate disparities in model performance
Park et al 2019[210]	Bias ingrained in data	Evaluate disparities in model performance
Kasturi et al 2021[201]	Bias ingrained in data	Evaluate disparities in model performance
Vokinger et al 2021[35]	Bias ingrained in data	Evaluate disparities in model performance
Obermeyer et al 2019[10]	Bias ingrained in data	Evaluate disparities in model performance
Cruz et al 2021[209]	Bias ingrained in data	Increase model reporting and transparency
Ledford 2019[46]	Bias ingrained in data	Increase model reporting and transparency
Hernandez-Boussard et al 2020[211]	Bias ingrained in data	Increase model reporting and transparency
Engler 2020[105]	Bias ingrained in data	Avoid or reduce use of AI
American Association for the Advancement of Science 2021[207]	Bias ingrained in data	Improve diversity, quality, or quantity of data

Document	Issue	Proposed strategy
Chen et al 2021[36]	Bias ingrained in data	Improve diversity, quality, or quantity of data
Google undated[67]	Bias ingrained in data	Improve diversity, quality, or quantity of data
Holstein et al 2019[49]	Bias ingrained in data	Improve diversity, quality, or quantity of data
Gianfrancesco et al 2018[75]	Bias ingrained in data	Improve diversity, quality, or quantity of data
Obermeyer et al 2021[156]	Bias ingrained in data	Improve diversity, quality, or quantity of data
O'Brien et al 2022[40]	Bias ingrained in data	Improve diversity, quality, or quantity of data
Parikh et al 2019[203]	Bias ingrained in data	Improve diversity, quality, or quantity of data
Rajkomar et al 2018[3]	Bias ingrained in data	Improve diversity, quality, or quantity of data
Rodolfa et al 2019[60]	Bias ingrained in data	Improve diversity, quality, or quantity of data
The Alan Turing Institute 2021[39]	Bias ingrained in data	Improve diversity, quality, or quantity of data
Goreke et al 2021[268]	Bias ingrained in data	Improve diversity, quality, or quantity of data
Naude and Vinuesa 2021[115]	Bias ingrained in data	Improve diversity, quality, or quantity of data
Obermeyer et al 2021[156]	Bias ingrained in data	Use equity-focused checklists, guidelines, and similar tools
Delgado et al 2022[172]	Inclusion of sensitive variables	Engage the broader community
New York City Health and Hospitals 2021[269]	Inclusion of sensitive variables	Improve governance
Vyas 2020[61]	Inclusion of sensitive variables	Exclude sensitive variables to correct for bias
Benito-León 2021[83]	Inclusion of sensitive variables	Exclude sensitive variables to correct for bias
O'Brien et al 2022[40]	Inclusion of sensitive variables	Exclude sensitive variables to correct for bias
Palmer 2021[96]	Inclusion of sensitive variables	Exclude sensitive variables to correct for bias
Zarsky 2016[270]	Inclusion of sensitive variables	Exclude sensitive variables to correct for bias

Document	Issue	Proposed strategy
Delgado et al 2022[172]	Inclusion of sensitive variables	Exclude sensitive variables to correct for bias
NEPHJC 2019[84]	Inclusion of sensitive variables	Avoid or reduce use of AI
Osoba et al 2019[18]	Exclusion of sensitive variables	Improve governance
Schmidt 2020[88]	Exclusion of sensitive variables	Improve governance
Leslie et al 2021[6]	Exclusion of sensitive variables	Include sensitive variables to correct for bias
Veale and Binns 2017[206]	Exclusion of sensitive variables	Include sensitive variables to correct for bias
Obermeyer et al 2021[156]	Exclusion of sensitive variables	Include sensitive variables to correct for bias
O'Brien et al 2022[40]	Exclusion of sensitive variables	Include sensitive variables to correct for bias
Osoba et al 2019[18]	Exclusion of sensitive variables	Include sensitive variables to correct for bias
Schmidt 2020[88]	Exclusion of sensitive variables	Include sensitive variables to correct for bias
Osoba et al 2019[18]	Exclusion of sensitive variables	Evaluate disparities in model performance
Rodolfa et al 2019[60]	Exclusion of sensitive variables	Evaluate disparities in model performance
Mitchell 2019[162]	Limited reporting of information on protected groups	Engage the broader community
Wu et al 2021[91]	Limited reporting of information on protected groups	Improve governance
Gallifant 2021[90]	Limited reporting of information on protected groups	Evaluate disparities in model performance
Wu et al 2021[91]	Limited reporting of information on protected groups	Evaluate disparities in model performance
Mörch et al 2021[113]	Limited reporting of information on protected groups	Increase model reporting and transparency
Gallifant 2021[90]	Limited reporting of information on protected groups	Increase model reporting and transparency
Mitchell 2019[162]	Limited reporting of information on protected groups	Increase model reporting and transparency
Wu et al 2021[91]	Limited reporting of information on protected groups	Increase model reporting and transparency
Kiener 2020[53]	Algorithms are not interpretable	Train developers and users

Document	Issue	Proposed strategy
Antoniadi et al 2021[191]	Algorithms are not interpretable	Improve interpretability or explainability of algorithm
Arrieta et al 2020[271]	Algorithms are not interpretable	Improve interpretability or explainability of algorithm
Osoba et al 2019[18]	Algorithms are not interpretable	Improve interpretability or explainability of algorithm
Reddy et al 2020[137]	Algorithms are not interpretable	Improve interpretability or explainability of algorithm
Vokinger et al 2021[35]	Algorithms are not interpretable	Improve interpretability or explainability of algorithm
Mullainathan 2019[95]	Algorithms are not interpretable	Evaluate disparities in model performance
O'Neil 2016[72]	Algorithms are not interpretable	Avoid or reduce use of AI
Vokinger et al 2021[35]	Algorithms are not interpretable	Avoid or reduce use of AI
Miller 2020[216]	Optimizing algorithm accuracy and fairness may conflict	Engage the broader community
O'Neil 2016[72]	Optimizing algorithm accuracy and fairness may conflict	Exclude sensitive variables to correct for bias
McCradden et al 2020[127]	Optimizing algorithm accuracy and fairness may conflict	Enforce fairness goals
Rodolfa 2021[100]	Optimizing algorithm accuracy and fairness may conflict	Enforce fairness goals
Taylor et al 2018[124]	Optimizing algorithm accuracy and fairness may conflict	Enforce fairness goals
Chouldechova 2018[65]	Optimizing algorithm accuracy and fairness may conflict	Evaluate disparities in model performance
McCradden et al 2020[127]	Optimizing algorithm accuracy and fairness may conflict	Evaluate disparities in model performance
Miller 2020[216]	Optimizing algorithm accuracy and fairness may conflict	Evaluate disparities in model performance
Osoba et al 2019[18]	Optimizing algorithm accuracy and fairness may conflict	Evaluate disparities in model performance
McCradden et al 2020[127]	Optimizing algorithm accuracy and fairness may conflict	Increase model reporting and transparency
Osoba et al 2019[18]	Optimizing algorithm accuracy and fairness may conflict	Increase model reporting and transparency
Miller 2020[216]	Optimizing algorithm accuracy and fairness may conflict	Avoid or reduce use of AI

Document	Issue	Proposed strategy
Holstein et al 2019[49]	Optimizing algorithm accuracy and fairness may conflict	Use equity-focused checklists, guidelines, and similar tools
Horvitz et al 2020[101]	Ambiguity in and conflict among conceptions of equity	Engage the broader community
Osoba et al 2019[18]	Ambiguity in and conflict among conceptions of equity	Engage the broader community
Bozkurt et al 2020[267]	Proprietary algorithms or data unavailable for evaluation	Improve governance
Osoba et al 2019[18]	Proprietary algorithms or data unavailable for evaluation	Improve governance
Osoba et al 2019[18]	Proprietary algorithms or data unavailable for evaluation	Evaluate disparities in model performance
Obermeyer et al 2019[10]	Proprietary algorithms or data unavailable for evaluation	Evaluate disparities in model performance
Bowen et al 2022[272]	Proprietary algorithms or data unavailable for evaluation	Increase model reporting and transparency
Mitchell 2019[162]	Proprietary algorithms or data unavailable for evaluation	Increase model reporting and transparency
Harwich and Laycock 2018[73]	Proprietary algorithms or data unavailable for evaluation	Increase model reporting and transparency
O'Neil 2016[72]	Proprietary algorithms or data unavailable for evaluation	Avoid or reduce use of AI
Simonite 2022[266]	Proprietary algorithms or data unavailable for evaluation	Improve diversity, quality, or quantity of data
Kiener 2020[53]	Overreliance on AI apps	Train developers and users
Gianfrancesco et al 2018[75]	Overreliance on AI apps	Evaluate disparities in model performance
Engler 2020[105]	Overreliance on AI apps	Avoid or reduce use of AI
Murphy et al 2021[48]	Underreliance on AI apps	Train developers and users
Murphy et al 2021[48]	Underreliance on AI apps	Engage the broader community
Ferryman and Pitcan 2018[120]	Repurposing existing AI apps outside original scope	Improve governance
Osoba et al 2019[18]	Repurposing existing AI apps outside original scope	Improve governance
Leslie et al 2021[6]	Repurposing existing AI apps outside original scope	Evaluate disparities in model performance
Rodolfa et al 2019[60]	Repurposing existing AI apps outside original scope	Evaluate disparities in model performance

Document	Issue	Proposed strategy
Mitchell 2019[162]	Repurposing existing AI apps outside original scope	Increase model reporting and transparency
Osoba et al 2019[18]	Repurposing existing AI apps outside original scope	Seek or provide restitution for those negatively affected by AI
Röösli and Hernandez-Boussard 2021[4]	Application development or implementation is rushed	Increase model reporting and transparency
Zimmer et al 2021[38]	Unequal access to AI	Train developers and users
Aggarwal et al 2020[219]	Unequal access to AI	Provide resources to those with less access to AI
Mörch et al 2021[113]	Unequal access to AI	Provide resources to those with less access to AI
IEEE 2019[119]	Unequal access to AI	Provide resources to those with less access to AI
O'Brien et al 2022[40]	Unequal access to AI	Provide resources to those with less access to AI
Smythe et al 2021[222]	Unequal access to AI	Provide resources to those with less access to AI
Whitelaw et al 2020[118]	Unequal access to AI	Provide resources to those with less access to AI
Aggarwal et al 2020[219]	Unequal access to AI	Improve diversity, quality, or quantity of data

Appendix B. Stakeholder Interview Guide

This appendix presents the interview guide used in conducting stakeholder interviews. Stakeholder interviews were conducted as part of a broader project examining topics related to both the use of AI in the COVID-19 response, as well as strategies to address the impact of AI on health equity. As a result, this interview guide is also included as a supplemental file in a separate journal article[27] and as an appendix in a separate report on the use of AI in the COVID-19 response. [28]

This interview guide was sent as an email attachment to stakeholder interviewees before the interview. Interviewers also used this document while they were conducting interviews as a guide to receiving informed consent and asking questions. It includes a study description, an informed consent protocol, and a list of potential interview questions.

The remainder of this appendix provides the full text of the interview guide sent to stakeholder representatives before their participation in an interview. References to *you* in this guide refer to the intended audience: a stakeholder representative participating in an upcoming interview.

Study Description

The study description, informed consent protocol, and interview guide will be covered during the interview itself and will also be sent to potential interviewees in advance of any scheduled interview. Interviewers will read the following study description and informed consent protocol before beginning the interview.

This project is funded by the Patient Centered Outcomes Research Institute (PCORI) and is being conducted by the RAND Corporation. The goal of our study is to better understand the use of artificial intelligence (AI) in the clinical and public health response to COVID-19, including its potential impacts on health equity. We are interested in understanding the types of AI applications being used in the COVID-19 response, the functions that these applications perform, and the contexts in which they are used. We are also interested in assessing the evidence surrounding these applications' use, in terms of both their benefits and drawbacks, particularly regarding intergroup health disparities and equity concerns. We will seek to identify strategies to mitigate negative impacts and to enhance positive impacts of AI on health equity.

To do this, we will be producing a literature and evidence review on these subjects. We are also conducting key informant interviews to ensure that we consider a range of perspectives and priorities as we determine the review scope and guiding questions. We expect the final product of our study to be a published report available on PCORI's website.

Informed Consent Protocol

We are conducting interviews with a wide range of people, including patients and patient advocates, clinicians, hospitals and health systems, health care payers and insurers, health policy makers, public health officials, industry representatives, and researchers. We would like to interview you for this study. You have been selected because of your perspective, interest, and experience with issues relevant to this research.

Risks: We do not expect that you would face any risks related to your participation in this interview.

Confidentiality: We will keep your responses during this interview confidential. We will not be recording this interview, although we will be taking written notes. These notes will be accessible only to the study team. We will not include your name in our interview notes, and we will store all interview notes securely and separately from the list of interviewee names and identifying information. We will report themes and variation in responses evident across all our interviews. We may refer to information or opinions you express in the interview in our report, but we will attribute these generically to a "key informant interviewee," and will not attribute comments to anyone by name, position, or affiliation or in any way that could be used to identify you.

Duration: Your participation in this interview will last about 1 hour.

Participation and withdrawal: Participation in this interview is entirely voluntary. Deciding not to participate will have no negative consequences. If you decide to participate, you are free to end the interview at any point or decline to answer any question for any reason.

Questions for interviewer: Do you have any questions about this study or about participation in this interview that you would like answered now?

Informed consent: Do you consent to participate in this interview?

Interview Guide

A list of interview questions follows.

Although we plan to use these prompts and questions to guide our interview discussions, we will not be using these as a verbatim script for interviews. Rather, we will tailor our exact interview approach to each individual interviewee while considering the varying perspectives and backgrounds that different individuals bring to this topic. Some questions may not apply to all interviewees, in which case we will move on to another question or topic.

Throughout the interview, we will encourage interviewees to expand upon their answers or to raise additional topics they feel are important for discussion.

A. Background

Tell us a little about your background on this topic.

A1. In what ways, if any, have you been involved in or affected by the use of AI in the clinical care or public health response to COVID-19?

A2. What experiences or other factors inform your perspective on the use of AI in health care and its potential impact on health equity?

B. Key Questions

Our study is focused on the use of AI in the clinical and public health response to COVID-19, including its potential impacts on health equity. This includes AI-based clinical products and tools that are used to diagnose COVID-19, evaluate patient prognosis, and assess treatment benefits and harms. This also includes AI-based applications that are used in support of public health efforts such as COVID-19 forecasting, contact tracing, resource distribution, vaccine prioritization, and combating health misinformation.

We are in the process of refining the study questions that will be used to guide our literature review. Our current list includes several key questions. We are interested in your answers to these questions and your thoughts on whether these are the right questions to ask in our study.

We will first ask you to answer each of these questions directly. We know these questions cover a broad range of issues, and we will not have time to address them in complete detail. So please just answer with whatever first comes to mind.

B1. Are you familiar with any AI-based applications that have been used in clinical care for COVID-19? Any that may be used in the near future? (If yes to either question, please describe them.)

B2. Are you familiar with any AI-based applications that have been used in the public health response to COVID-19? Any that may be used in the near future? (If yes to either question, please describe them.)

B3. Are you aware of any evidence available on the potential benefits and drawbacks of these applications? (If yes, please elaborate.)

B4. Are you aware of or concerned about potential negative impacts of the use of AI on health equity? Are you aware of any potential positive impacts of AI on health equity?

B5. Are you aware of any strategies that seek to mitigate negative impacts or enhance positive impacts of AI on health equity?

C. Research Priorities

We appreciate your answers to those questions because this is important information for our study. We would like to continue to talk about these topics but to take a step back to think about them a bit differently. We would like to get your opinion on which aspects of these topics you think we should focus our research.

As mentioned earlier, PCORI has asked us to produce a literature review on the use of AI in the COVID-19 response and its impact on health equity.

This literature review has 2 goals. First, it is intended to help PCORI and others involved in health care, such as you, to understand the current state of the field. Second, this report should identify strategies that can mitigate negative impacts and enhance positive impacts of AI on health equity.

We are just starting this project, so we are interested in getting your perspective as we determine our study scope and key research questions to help us address these goals.

C1. What questions about AI in the COVID-19 response, and the evidence surrounding its use, should we seek to answer in our literature review?

C2. Should we make sure to include specific types of AI applications in our review?

C3. What types of potential benefits, harms, and other impacts should we make sure to examine in our review?

C4. What questions about how the use of AI in health care affects health equity should we seek to answer in our literature review?

C5. What types of health equity impacts should we make sure to examine? Should we make sure to look at impacts on particular groups?

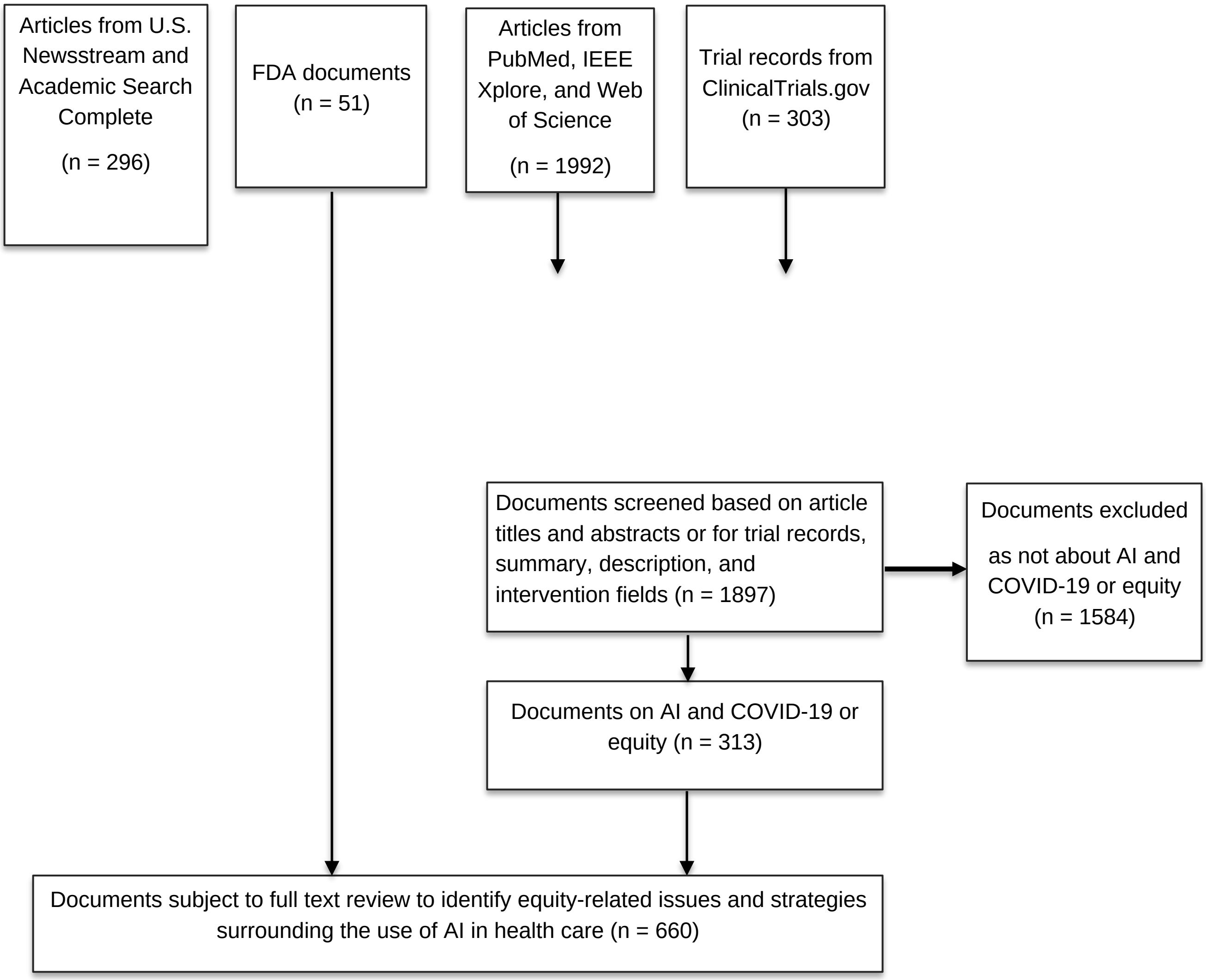
C6. Do you think it would be especially relevant and useful for our study to examine any particular types of documents or data sources?

Appendix C. Literature Searches and Screening

Literature searches and screening were conducted as part of a broader project that examined topics related to both the use of AI in the COVID-19 response and strategies to address the impact of AI on health equity. As a result, the material in this appendix is also included as a supplemental file in a separate journal article[27] and as an appendix in a separate report on the use of AI in the COVID-19 response. [28]

This appendix begins with a literature flow diagram (Figure C.1) following the guidelines in the Preferred Reporting Items for Systematic reviews and Meta-Analyses extension for Scoping Reviews (PRISMA-Scr). [273] We do not provide separate counts of documents on public health or on clinical AI applications because most documents did not differentiate between these categories.

Figure C.1. PRISMA Literature Flow Diagram



Abbreviations: AI, artificial intelligence; FDA, US Food and Drug Administration; IEEE, Institute of Electrical and Electronics Engineers; PRISMA, Preferred Reporting Items for Systematic Reviews and Meta-Analyses.

The remainder of this appendix contains details on the literature searches and document screening methods used in this study.

Literature Search

We conducted systematic searches of PubMed, Web of Science, the IEEE Xplore Digital Library, ProQuest US Newsstream, Academic Search Complete, the ClinicalTrials.gov database, the FDA CDRH document library, and Google web search. The databases that we searched, together with the number of unique documents that we found in each after removal of duplicate results, are summarized in Table C.1.

Table C.1. Document Types and Search Results, by Database and Subject

Database	AI in the COVID-19 response	AI and equity in the COVID-19 response	AI and equity	Unique documents
PubMed	Reviews (12/31/19-12/10/21)	All articles (12/31/19-12/10/21)	Reviews (1/1/14-12/10/21)	761
Web of Science	Reviews (12/31/19-12/10/21)	All articles (12/31/19-12/10/21)	Reviews, Highly cited articles (1/1/14-12/10/21)	773
IEEE Xplore	Reviews (1/1/20-12/13/21)	All articles (1/1/20-12/13/21)	Reviews (1/1/14-12/13/21)	60
US Newsstream	All articles (1/1/20-12/13/21)		All articles (1/1/20-12/13/21)	199
Academic Search Complete	All articles (1/1/20-12/13/21)		All articles (1/1/20-12/13/21)	97
ClinicalTrials.gov	All trial records (up to 12/23/21)			303
FDA	All 510(k), Premarket Approval (PMA), De Novo, and EUA documents (up to 12/28/21)			51
Total unique documents: 2244 (1897 of these documents included abstracts, including all results from PubMed, Web of Science, and ClinicalTrials.gov.)				

Abbreviations: AI, artificial intelligence; EUA, emergency use authorization; FDA, US Food and Drug Administration; IEEE, Institute of Electrical and Electronics Engineers.

The ProQuest US Newsstream and Academic Search Complete databases were selected because they include articles and commentaries in publications that give voice to historically marginalized populations (eg, the Baltimore *Afro-American*, the *Philadelphia Tribune*, *Indian Country Today*) in addition to a wide range of other US and English-language periodicals (eg, *Newsweek*, *The Economist*, *New York Times*, *Arizona Daily Star*, *Wall Street Journal*, *Modern Healthcare*).

To find academic literature on AI in the COVID-19 response, we searched for documents that used the word “review” in their title and mentioned at least 1 AI-related term and 1 COVID-19–related term. We focused this search on identifying academic review articles, rather than all articles published on AI and COVID-19, because of the large number of publications in the latter category, many of which focus on AI algorithms that were never deployed for use. We searched for all articles (including nonreview articles) that mentioned AI, COVID-19, and health equity–related terms to ensure we found articles that discussed equity in the context of AI applications used in the COVID-19 response. Finally, we conducted a search for review articles and highly cited articles that discussed AI and health equity topics back to 2014, a year that was selected to include the most recent 8 years of published literature on the subject.

Article type and dates for all searches are noted in Table 2.2. Our searches for documents related to COVID-19 were restricted to articles published between December 31, 2019 (the date on which China reported the possibility of a novel virus to the WHO)[124] and December 13, 2021. Our searches for documents focusing on AI and health equity were restricted to documents published in the past 8 years, beginning January 1, 2014.

Document Screening

To ensure consistency in screening decisions, we used dual-review methods and assessed interviewer reliability. This began with 3 members of the project team each independently examining a random sample of approximately 10% of search results, followed by discussion of discrepancies and refinement of screening criteria. We then used these finalized criteria to conduct single screening of the remaining search results, with random dual-review checks of 25% of the remaining documents to ensure continued consistency in the screening process.

Disagreements in the 2 reviewers’ screening decisions were resolved by a third project team member, with discussion of edge cases conducted on an as-needed basis. Dual-reviewed screening decisions agreed 88% of the time, with a Cohen κ measure of interrater reliability of 0.61 (κ values range from –1.0, indicating perfect disagreement, to 1.0, indicating perfect agreement). Documents subject to dual review were slightly more likely to be screened into our review, with an inclusion rate of 19% compared with 15% for documents that were reviewed by a single team member.

Detailed Information on Search Queries

PubMed Searches

Notes on search fields:

[ti] = Title
[tiab] = Title and Abstract
[mh] = MeSH Terms

PubMed Search #1

AI in COVID-19 Response, review articles

Search conducted on December 10, 2021

Date range: December 31, 2019, to present

English

"machine learning"[ti] OR "artificial intelligence"[ti] OR "deep learning"[ti] OR "neural net*"[ti] OR "support vector machine*"[ti] OR SVM[ti] OR "random forest*"[ti] OR "supervised learning"[ti] OR "unsupervised learning"[ti] OR "reinforcement learning"[ti] OR "unsupervised clustering"[ti] OR "unsupervised classification"[ti] OR "supervised classification"[ti] OR "natural language processing"[ti] OR NLP[ti] OR "gradient boost*"[ti] OR "ensemble model"[ti] OR "expert system*"[ti] OR "rules engine*"[ti] OR "fuzzy logic"[ti] OR algorithm*[ti] OR "Artificial Intelligence"[mh]
AND
Coronavirus[tiab] OR COVID*[tiab] OR "SARS-COV-2"[tiab] OR "2019-nCOV"[tiab] OR "nCOV-19"[tiab] OR "COVID-19"[mh] OR "SARS-CoV-2"[mh] OR "COVID-19 Testing"[mh]
AND
review[ti]

Results: 112

PubMed Search #2

AI and Equity in COVID-19 Response, all articles

Search conducted on December 10, 2021

Date range: December 31, 2019, to present

English

"machine learning"[ti] OR "artificial intelligence"[ti] OR "deep learning"[ti] OR "supervised learning"[ti] OR "unsupervised learning"[ti] OR "reinforcement learning"[ti] OR "unsupervised clustering"[ti] OR "unsupervised classification"[ti] OR "supervised classification"[ti] OR "natural language processing"[ti] OR NLP[ti] OR "expert system*"[ti] OR "rules engine*"[ti] OR "fuzzy logic"[ti] OR algorithm*[ti] OR "Artificial Intelligence"[mh]
AND
Coronavirus[tiab] OR COVID*[tiab] OR "SARS-COV-2"[tiab] OR "2019-nCOV"[tiab] OR "nCOV-19"[tiab] OR "COVID-19"[mh] OR "SARS-CoV-2"[mh] OR "COVID-19 Testing"[mh]
AND
equit*[tiab] OR fair*[tiab] OR unfair[tiab] OR bias*[tiab] OR inequ*[tiab] OR unequ*[tiab] OR equality[tiab] OR inclusiv*[tiab] OR exclude*[tiab] OR race[tiab] OR racial[tiab] OR racism[tiab] OR gender[tiab] OR sex[tiab] OR ethnic*[tiab] OR disab*[tiab] OR dispar*[tiab] OR disproportion*[tiab] OR "social determinant*"[tiab] OR socioeconomic*[tiab] OR income[tiab] OR minorit*[tiab] OR disadvantaged[tiab] OR vulnerab*[tiab] OR marginali*[tiab] OR "Health Equity"[mh] OR "Gender Equity"[mh] OR "Healthcare Disparities"[mh] OR "Prejudice"[mh] OR "Social Determinants of Health"[mh] OR "Minority Health"[mh] OR "Racial Groups"[mh] OR "Socioeconomic Factors"[mh] OR "Race Relations"[mh] OR "Ethnicity"[mh]

Results: 356 – duplicates with PubMed Search #1 = 333

PubMed Search #3

AI and Equity, review articles

Search conducted on December 10, 2021

Date range: January 1, 2014, to present

English

"machine learning"[ti] OR "artificial intelligence"[ti] OR "deep learning"[ti] OR "supervised learning"[ti] OR "unsupervised learning"[ti] OR "reinforcement learning"[ti] OR "unsupervised clustering"[ti] OR "unsupervised classification"[ti] OR "supervised classification"[ti] OR "natural language processing"[ti] OR NLP[ti] OR "expert system*"[ti] OR "rules engine*"[ti] OR "fuzzy logic"[ti] OR algorithm*[ti] OR "Artificial Intelligence"[mh]
AND
health*[tiab] OR clinic*[tiab] OR patient*[tiab] OR hospital*[tiab] OR therap*[tiab] OR medic*[tiab] OR care[tiab]
AND
equit*[tiab] OR fair*[tiab] OR unfair[tiab] OR bias*[tiab] OR inequ*[tiab] OR unequ*[tiab] OR equality[tiab] OR inclusiv*[tiab] OR exclude*[tiab] OR race[tiab] OR racial[tiab] OR racism[tiab] OR gender[tiab] OR sex[tiab] OR ethnic*[tiab] OR disab*[tiab] OR dispar*[tiab] OR disproportion*[tiab] OR "social determinant*"[tiab] OR socioeconomic*[tiab] OR income[tiab] OR minorit*[tiab] OR disadvantaged[tiab] OR vulnerab*[tiab] OR marginali*[tiab] OR "Health Equity"[mh] OR "Gender Equity"[mh] OR "Healthcare Disparities"[mh] OR "Prejudice"[mh] OR "Social Determinants of Health"[mh] OR "Minority Health"[mh] OR "Racial Groups"[mh] OR "Socioeconomic Factors"[mh] OR "Race Relations"[mh] OR "Ethnicity"[mh]
AND
Review[ti]

Results: 337 – duplicates with PubMed Searches #1 and #2 = 316

Total Results from PubMed Searches #1-3 = 761

Web of Science Searches

Science Citation Index Expanded (SCI-EXPANDED), Social Sciences Citation Index (SSCI), Arts & Humanities Citation Index (A&HCI), Conference Proceedings Citation Index – Science (CPCI-S), Conference Proceedings Citation Index – Social Science & Humanities (CPCI-SSH), Emerging Sources Citation Index (ESCI):

Notes on search fields:

TI = Title

KP = Keywords Plus
AK = Author Keywords
TS = Topic
[mh] = MeSH Terms

Web of Science Search #1

AI in COVID-19 Response, review articles

Search conducted on December 10, 2021

Date range: December 31, 2019, to December 10, 2021

English

TI=("machine learning" OR "artificial intelligence" OR "deep learning" OR "neural net*" OR "support vector machine*" OR SVM OR "random forest*" OR "supervised learning" OR "unsupervised learning" OR "reinforcement learning" OR "unsupervised clustering" OR "unsupervised classification" OR "supervised classification" OR "natural language processing" OR NLP OR "gradient boost*" OR "ensemble model" OR "expert system*" OR "rules engine*" OR "fuzzy logic" OR algorithm*) OR KP=("machine learning" OR "artificial intelligence" OR "deep learning" OR "neural net*" OR "support vector machine*" OR SVM OR "random forest*" OR "supervised learning" OR "unsupervised learning" OR "reinforcement learning" OR "unsupervised clustering" OR "unsupervised classification" OR "supervised classification" OR "natural language processing" OR NLP OR "gradient boost*" OR "ensemble model" OR "expert system*" OR "rules engine*" OR "fuzzy logic" OR algorithm*) OR AK=("machine learning" OR "artificial intelligence" OR "deep learning" OR "neural net*" OR "support vector machine*" OR SVM OR "random forest*" OR "supervised learning" OR "unsupervised learning" OR "reinforcement learning" OR "unsupervised clustering" OR "unsupervised classification" OR "supervised classification" OR "natural language processing" OR NLP OR "gradient boost*" OR "ensemble model" OR "expert system*" OR "rules engine*" OR "fuzzy logic" OR algorithm*)
AND
TS=(Coronavirus OR COVID* OR "SARS-COV-2" OR "2019-nCOV" OR "nCOV-19")
AND
TI=(review)

Results: 156 – duplicates with PubMed searches = 79

Web of Science Search #2

AI and Equity in COVID-19 Response, all articles

Search conducted on December 10, 2021
Date range: December 31, 2019, to December 10, 2021
English

TI=("machine learning" OR "artificial intelligence" OR "deep learning" OR "supervised learning" OR "unsupervised learning" OR "reinforcement learning" OR "unsupervised clustering" OR "unsupervised classification" OR "supervised classification" OR "natural language processing" OR NLP OR "expert system*" OR "rules engine*" OR "fuzzy logic" OR algorithm*) OR KP=("machine learning" OR "artificial intelligence" OR "deep learning" OR "supervised learning" OR "unsupervised learning" OR "reinforcement learning" OR "unsupervised clustering" OR "unsupervised classification" OR "supervised classification" OR "natural language processing" OR NLP OR "expert system*" OR "rules engine*" OR "fuzzy logic" OR algorithm*) OR AK=("machine learning" OR "artificial intelligence" OR "deep learning" OR "supervised learning" OR "unsupervised learning" OR "reinforcement learning" OR "unsupervised clustering" OR "unsupervised classification" OR "supervised classification" OR "natural language processing" OR NLP OR "expert system*" OR "rules engine*" OR "fuzzy logic" OR algorithm*)
AND
TS=(Coronavirus OR COVID* OR "SARS-COV-2" OR "2019-nCOV" OR "nCOV-19")
AND
TS=(equit* OR fair* OR unfair OR bias* OR inequ* OR unequ* OR equality OR inclusiv* OR exclude* OR race OR racial OR racism OR gender OR sex OR ethnic* OR disab* OR dispar* OR disproportion* OR "social determinant*" OR socioeconomic* OR income OR minorit* OR disadvantaged OR vulnerab* OR marginali* OR prejudic*)

Results: 466 – duplicates = 250

Web of Science Search #3

AI and Equity, review articles

Search conducted on December 10, 2021
Date range: January 1, 2014, to December 10, 2021
English

TI=("machine learning" OR "artificial intelligence" OR "deep learning" OR "supervised learning" OR "unsupervised learning" OR "reinforcement learning" OR "unsupervised clustering" OR "unsupervised classification" OR "supervised classification" OR "natural language processing" OR NLP OR "expert system*" OR "rules engine*" OR "fuzzy logic" OR algorithm*) OR KP=("machine learning" OR "artificial intelligence" OR "deep learning" OR "supervised learning" OR "unsupervised learning" OR "reinforcement learning" OR "unsupervised clustering" OR "unsupervised classification" OR "supervised classification" OR "natural language processing" OR NLP OR "expert system*" OR "rules engine*" OR "fuzzy logic" OR algorithm*) OR AK=("machine learning" OR "artificial intelligence" OR "deep learning" OR "supervised learning" OR "unsupervised learning" OR "reinforcement learning" OR "unsupervised clustering" OR "unsupervised classification" OR "supervised classification" OR "natural language processing" OR NLP OR "expert system*" OR "rules engine*" OR "fuzzy logic" OR algorithm*)
AND
TI=(health* OR clinic* OR patient* OR hospital* OR therap* OR medic* OR care) OR AB=(health* OR clinic* OR patient* OR hospital* OR therap* OR medic* OR care)
AND
TS=(equit* OR fair* OR unfair OR bias* OR inequ* OR unequ* OR equality OR inclusiv* OR exclude* OR race OR racial OR racism OR gender OR sex OR ethnic* OR disab* OR dispar* OR disproportion* OR "social determinant*" OR socioeconomic* OR income OR minorit* OR disadvantaged OR vulnerab* OR marginali* OR prejudic*)
AND
TI=(review)

Results: 458 – duplicates with previous searches = 230

Web of Science Search #4

AI and Equity, highly cited articles[1]

---

[1] For information on the Web of Science selection of highly cited papers, see Web of Science Core Collection Help: Citation Products, Published 2020. Accessed December 30, 2021. https://images.webofknowledge.com/WOKRS533JR18/help/WOS/hs_citation_applications.html

Search conducted on December 10, 2021
Date range: January 1, 2014, to December 10, 2021
English

TI=("machine learning" OR "artificial intelligence" OR "deep learning" OR "supervised learning" OR "unsupervised learning" OR "reinforcement learning" OR "unsupervised clustering" OR "unsupervised classification" OR "supervised classification" OR "natural language processing" OR NLP OR "expert system*" OR "rules engine*" OR "fuzzy logic" OR algorithm*) OR AB=("machine learning" OR "artificial intelligence" OR "deep learning" OR "supervised learning" OR "unsupervised learning" OR "reinforcement learning" OR "unsupervised clustering" OR "unsupervised classification" OR "supervised classification" OR "natural language processing" OR NLP OR "expert system*" OR "rules engine*" OR "fuzzy logic" OR algorithm*)
AND
TS=(health* OR clinic* OR patient* OR hospital* OR therap* OR medic* OR care)
AND
TS=(equit* OR fair* OR unfair OR bias* OR inequ* OR unequ* OR equality OR inclusiv* OR exclude* OR race OR racial OR racism OR gender OR sex OR ethnic* OR disab* OR dispar* OR disproportion* OR "social determinant*" OR socioeconomic* OR income OR minorit* OR disadvantaged OR vulnerab* OR marginali* OR prejudic*)

***Results**: 241 – duplicates with previous searches = 214*

Total Results from Web of Science Searches #1-4, excluding duplicates = 773

IEEE Xplore Searches

IEEE Xplore Search #1

AI in COVID-19 Response, review articles

Search conducted on December 13, 2021
Date range: 2020-2021

(“Document Title”:”machine learning” OR “Document Title”:”artificial intelligence” OR “Document Title”:”deep learning” OR “Document Title”:”neural net*” OR “Document Title”:”support vector machine*” OR “Document Title”:SVM OR “Document Title”:”random forest*” OR “Document Title”:”supervised learning” OR “Document Title”:”unsupervised learning” OR “Document Title”:”reinforcement learning” OR “Document Title”:”unsupervised clustering” OR “Document Title”:”unsupervised classification” OR “Document Title”:”supervised classification” OR “Document Title”:”natural language processing” OR “Document Title”:NLP OR “Document Title”:”gradient boost*” OR “Document Title”:”ensemble model” OR “Document Title”:”expert system*” OR “Document Title”:”rules engine*” OR “Document Title”:”fuzzy logic” OR “Document Title”:algorithm*)

AND

(“Document Title”or “Abstract”:Coronavirus OR “Document Title” or “Abstract”:COVID* OR “Document Title” or “Abstract”:”SARS-COV-2” OR “Document Title” or “Abstract”:”2019-nCOV” OR “Document Title” or “Abstract”:”nCOV-19”)

AND

(“Document Title”:review)

Results: 18

IEEE Xplore Search #2

AI and Equity in COVID-19 Response, all articles

Search conducted on December 13, 2021
Date range: 2020-2021

(“Document Title”:”machine learning” OR “Document Title”:”artificial intelligence” OR “Document Title”:”deep learning” OR “Document Title”:”supervised learning” OR “Document Title”:”unsupervised learning” OR “Document Title”:”reinforcement learning” OR “Document Title”:”unsupervised clustering” OR “Document Title”:”unsupervised classification” OR “Document Title”:”supervised classification” OR “Document Title”:”natural language processing” OR “Document Title”:NLP OR “Document Title”:”expert system*” OR “Document Title”:”rules engine*” OR “Document Title”:”fuzzy logic” OR “Document Title”:algorithm*)

AND

("Document Title" or "Abstract":Coronavirus OR "Document Title" or "Abstract":COVID* OR "Document Title" or "Abstract":"SARS-COV-2" OR "Document Title" or "Abstract":"2019-nCOV" OR "Document Title" or "Abstract":"nCOV-19")

AND

("Document Title" or "Abstract":equity OR "Document Title" or "Abstract":equities OR "Document Title" or "Abstract":fair OR "Document Title" or "Abstract":fairness OR "Document Title" or "Abstract":unfair OR "Document Title" or "Abstract":bias OR "Document Title" or "Abstract":biased OR "Document Title" or "Abstract":inequity OR "Document Title" or "Abstract":unequity OR "Document Title" or "Abstract":equality OR "Document Title" or "Abstract":inclusive OR "Document Title" or "Abstract":inclusivity OR "Document Title" or "Abstract":exclude OR "Document Title" or "Abstract":excluded OR "Document Title" or "Abstract":race OR "Document Title" or "Abstract":racial OR "Document Title" or "Abstract":racism OR "Document Title" or "Abstract":gender OR "Document Title" or "Abstract":sex OR "Document Title" or "Abstract":ethnic OR "Document Title" or "Abstract":ethnicity OR "Document Title" or "Abstract":disable OR "Document Title" or "Abstract":disabled OR "Document Title" or "Abstract":disab* OR "Document Title" or "Abstract":disability OR "Document Title" or "Abstract":disparity OR "Document Title" or "Abstract":disparities OR "Document Title" or "Abstract":disproportionate OR "Document Title" or "Abstract":disproportional OR "Document Title" or "Abstract":"social determinant" OR "Document Title" or "Abstract":socioeconomic* OR "Document Title" or "Abstract":income OR "Document Title" or "Abstract":minority OR "Document Title" or "Abstract":minorit* OR "Document Title" or "Abstract":disadvantaged OR "Document Title" or "Abstract":vulnerable OR "Document Title" or "Abstract":vulnerabilities OR "Document Title" or "Abstract":marginalized)

Results: 52

IEEE Xplore Search #3

AI and Equity, review articles

Search conducted on December 13, 2021
Date range: 2014-2021

("Document Title":"machine learning" OR "Document Title":"artificial intelligence" OR "Document Title":"deep learning" OR "Document Title":"supervised learning" OR "Document

Title”:”unsupervised learning” OR “Document Title”:”reinforcement learning” OR “Document Title”:”unsupervised clustering” OR “Document Title”:”unsupervised classification” OR “Document Title”:”supervised classification” OR “Document Title”:”natural language processing” OR “Document Title”:NLP OR “Document Title”:”expert system*” OR “Document Title”:”rules engine*” OR “Document Title”:”fuzzy logic” OR “Document Title”:algorithm*)

AND

(“Document Title” or “Abstract”:health* OR “Document Title” or “Abstract”:clinic* OR “Document Title” or “Abstract”:patient* OR “Document Title” or “Abstract”:hospital* OR “Document Title” or “Abstract”:therap* OR “Document Title” or “Abstract”:medic* OR “Document Title” or “Abstract”:care)

AND

(“Document Title” or “Abstract”:equity OR “Document Title” or “Abstract”:equities OR “Document Title” or “Abstract”:fair OR “Document Title” or “Abstract”:fairness OR “Document Title” or “Abstract”:unfair OR “Document Title” or “Abstract”:bias OR “Document Title” or “Abstract”:biased OR “Document Title” or “Abstract”:inequity OR “Document Title” or “Abstract”:unequity OR “Document Title” or “Abstract”:equality OR “Document Title” or “Abstract”:inclusive OR “Document Title” or “Abstract”:inclusivity OR “Document Title” or “Abstract”:exclude OR “Document Title” or “Abstract”:excluded OR “Document Title” or “Abstract”:race OR “Document Title” or “Abstract”:racial OR “Document Title” or “Abstract”:racism OR “Document Title” or “Abstract”:gender OR “Document Title” or “Abstract”:sex OR “Document Title” or “Abstract”:ethnic OR “Document Title” or “Abstract”:ethnicity OR “Document Title” or “Abstract”:disable OR “Document Title” or “Abstract”:disabled OR “Document Title” or “Abstract”:disab* OR “Document Title” or “Abstract”:disability OR “Document Title” or “Abstract”:disparity OR “Document Title” or “Abstract”:disparities OR “Document Title” or “Abstract”:disproportionate OR “Document Title” or “Abstract”:disproportional OR “Document Title” or “Abstract”:”social determinant” OR “Document Title” or “Abstract”:socioeconomic* OR “Document Title” or “Abstract”:income OR “Document Title” or “Abstract”:minority OR “Document Title” or “Abstract”:minorit* OR “Document Title” or “Abstract”:disadvantaged OR “Document Title” or “Abstract”:vulnerable OR “Document Title” or “Abstract”:vulnerabilities OR “Document Title” or “Abstract”:marginalized)

AND

(“Document Title”:Review)

Results: 24

Total Results from IEEE Xplore Searches #1-3: 94 – duplicates from previous searches = 60

US NewsStream Searches

Magazines, Newspapers, Reports, Blogs, Podcasts, and Websites

Notes on search fields:

ti = Title
ab = Abstract

US NewsStream Search #1

AI in COVID-19 Response, all articles

Search conducted on December 27, 2021
Date range: January 1, 2020, to present

ti("machine learning" OR "artificial intelligence" OR "deep learning" OR "neural net*" OR "support vector machine*" OR SVM OR "random forest*" OR "supervised learning" OR "unsupervised learning" OR "reinforcement learning" OR "unsupervised clustering" OR "unsupervised classification" OR "supervised classification" OR "natural language processing" OR NLP OR "gradient boost*" OR "ensemble model" OR "expert system*" OR "rules engine*" OR "fuzzy logic" OR algorithm*) OR (MAINSUBJECT.EXACT("Artificial intelligence")
AND
(ti(Coronavirus OR COVID* OR "SARS-COV-2" OR "2019-nCOV" OR "nCOV-19") or ab(Coronavirus OR COVID* OR "SARS-COV-2" OR "2019-nCOV" OR "nCOV-19"))

Results: 189 – internal duplicates = 183

US NewsStream Search #2

AI and Equity, all articles

Search conducted on December 27, 2021
Date range: January 1, 2020, to present

ti("machine learning" OR "artificial intelligence" OR "deep learning" OR "supervised learning" OR "unsupervised learning" OR "reinforcement learning" OR "unsupervised clustering" OR "unsupervised classification" OR "supervised classification" OR "natural language processing" OR NLP OR "expert system*" OR "rules engine*" OR "fuzzy logic" OR algorithm*) OR ab("machine learning" OR "artificial intelligence" OR "deep learning" OR "supervised learning" OR "unsupervised learning" OR "reinforcement learning" OR "unsupervised clustering" OR "unsupervised classification" OR "supervised classification" OR "natural language processing" OR NLP OR "expert system*" OR "rules engine*" OR "fuzzy logic" OR algorithm*) OR (MAINSUBJECT.EXACT("Artificial intelligence")
AND
ti(health* OR clinic* OR patient* OR hospital* OR therap* OR medic* OR care) OR ab(health* OR clinic* OR patient* OR hospital* OR therap* OR medic* OR care) MAINSUBJECT.EXACT("Health care")
AND
ti(equit* OR fair* OR unfair OR bias* OR inequ* OR unequ* OR equality OR inclusiv* OR exclude* OR race OR racial OR racism OR gender OR sex OR ethnic* OR disab* OR dispar* OR disproportion* OR "social determinant*" OR socioeconomic* OR income OR minorit* OR disadvantaged OR vulnerab* OR marginali* OR prejudic*) OR ab(equit* OR fair* OR unfair OR bias* OR inequ* OR unequ* OR equality OR inclusiv* OR exclude* OR race OR racial OR racism OR gender OR sex OR ethnic* OR disab* OR dispar* OR disproportion* OR "social determinant*" OR socioeconomic* OR income OR minorit* OR disadvantaged OR vulnerab* OR marginali* OR prejudic*) OR (MAINSUBJECT.EXACT("Distributive justice") OR MAINSUBJECT.EXACT("Discrimination") OR MAINSUBJECT.EXACT("Social exclusion") OR MAINSUBJECT.EXACT("Equality") OR MAINSUBJECT.EXACT("Inequality") OR MAINSUBJECT.EXACT("Racial justice")))

Results: 25 – internal duplicates = 21

Total Results from US NewsStream Searches #1 and 2: 204 – duplicates from previous searches = 199

Academic Search Complete Searches

Magazines, Trade Publications, Newspapers

Notes on search fields:

ti = Title
ab = Abstract
DE = Heading or Keyword
KW = Keyword

Academic Search Complete Search #1

AI in COVID-19 Response, all articles

Search conducted on December 27, 2021
Date range: January 1, 2020, to present

TI("machine learning" OR "artificial intelligence" OR "deep learning" OR "neural net*" OR "support vector machine*" OR SVM OR "random forest*" OR "supervised learning" OR "unsupervised learning" OR "reinforcement learning" OR "unsupervised clustering" OR "unsupervised classification" OR "supervised classification" OR "natural language processing" OR NLP OR "gradient boost*" OR "ensemble model" OR "expert system*" OR "rules engine*" OR "fuzzy logic" OR algorithm*) OR DE "Artificial intelligence"
AND
(ti(Coronavirus OR COVID* OR "SARS-COV-2" OR "2019-nCOV" OR "nCOV-19") or ab(Coronavirus OR COVID* OR "SARS-COV-2" OR "2019-nCOV" OR "nCOV-19")) OR DE "COVID-19"

Results: 91 – duplicates with USNewsStream = 87

Academic Search Complete Search #2

AI and Equity, all articles

Search conducted on December 27, 2021
Date range: January 1, 2020, to present

TI("machine learning" OR "artificial intelligence" OR "deep learning" OR "supervised learning" OR "unsupervised learning" OR "reinforcement learning" OR "unsupervised clustering" OR "unsupervised classification" OR "supervised classification" OR "natural language processing" OR NLP OR "expert system*" OR "rules engine*" OR "fuzzy logic" OR algorithm*) OR ab("machine learning" OR "artificial intelligence" OR "deep learning" OR "supervised learning" OR "unsupervised learning" OR "reinforcement learning" OR "unsupervised clustering" OR "unsupervised classification" OR "supervised classification" OR "natural language processing" OR NLP OR "expert system*" OR "rules engine*" OR "fuzzy logic" OR algorithm*) OR DE "Artificial intelligence"
AND
TI(health* OR clinic* OR patient* OR hospital* OR therap* OR medic* OR care) OR AB(health* OR clinic* OR patient* OR hospital* OR therap* OR medic* OR care) OR KW(health* OR clinic* OR patient* OR hospital* OR therap* OR medic* OR care)
AND
TI(equit* OR fair* OR unfair OR bias* OR inequ* OR unequ* OR equality OR inclusiv* OR exclude* OR race OR racial OR racism OR gender OR sex OR ethnic* OR disab* OR dispar* OR disproportion* OR "social determinant*" OR socioeconomic* OR income OR minorit* OR disadvantaged OR vulnerab* OR marginali* OR prejudic*) OR AB(equit* OR fair* OR unfair OR bias* OR inequ* OR unequ* OR equality OR inclusiv* OR exclude* OR race OR racial OR racism OR gender OR sex OR ethnic* OR disab* OR dispar* OR disproportion* OR "social determinant*" OR socioeconomic* OR income OR minorit* OR disadvantaged OR vulnerab* OR marginali* OR prejudic*) OR (MAINSUBJECT.EXACT("Distributive justice") OR DE "HEALTH equity" OR DE "EQUITY"

Results: 19 – duplicates with previous Academic Search Complete and US Newsstream searches = 16

Total Results from Academic Search Complete Searches #1 and 2: 103 – duplicates from previous searches = 97

ClinicalTrials.gov Searches

Search conducted December 23, 2021

We empirically found that adding the terms “algorithm,” “software,” and “ai” greatly improves the yield of the search without generating an excessive number of results. Therefore, we have included these terms in the search list. The search includes only trials with a start date of 1/1/2020 through the present.

The website search engine of ClinicalTrials.gov automatically includes additional search terms related to COVID, so we found that using “COVID” alone is sufficient.

Because the search field has a character limit, we performed an initial pruning and tested each term 1 by 1 to eliminate terms with 0 yield and split the search into fewer subsearches. The results of the single-term search follow.

Results of single-term search

Term	**Hits**
algorithm	118
software	112
ai	54
artificial intelligence	47
machine learning	44
deep learning	15
neural network	14
natural language processing	4
NLP	3
unsupervised classification	1
reinforcement learning	1
support vector	1
rules engine*	0
expert system*	0
ensemble model	0
gradient boost*	0
supervised classification	0
unsupervised clustering	0
unsupervised learning	0
supervised learning	0

Results of single-term search	
Term	**Hits**
random forest	0
SVM	0

The final searches were therefore split into the following 2 searches:

- **Search 1:** (“algorithm” OR “software” OR “ai” OR “artificial intelligence” OR “machine learning” OR “deep learning” OR “neural network”) AND COVID, yielding 299 hits.
- **Search 2:** (“natural language processing” OR “NLP” OR “supervised classification” OR “unsupervised classification” OR “unsupervised clustering” OR “reinforcement learning” OR “random forest” OR “support vector”) AND COVID, yielding 10 hits. Subtracting duplicates leaves 4 hits.

Total Results from ClinicalTrials.gov: 303

FDA Document Searches

Google Searches of FDA CDRH Document Library

AI in COVID-19 (all AI-enabled devices authorized by the FDA via the 510(k), De Novo, or Premarket Approval (PMA) process, in 2020 or 2021)

Search conducted December 28, 2021

Using Firefox, private browser mode (no cookies, accounts, search history, etc)

Six separate Google.com searches, using 3 AI terms, of the 2020 and 2021 FDA CDRH document libraries

“machine learning site:https://www.accessdata.fda.gov/cdrh_docs/pdf21/”

“machine learning site:https://www.accessdata.fda.gov/cdrh_docs/pdf20/”

“artificial intelligence site: site:https://www.accessdata.fda.gov/cdrh_docs/pdf21/”

“artificial intelligence site: site:https://www.accessdata.fda.gov/cdrh_docs/pdf20/”

“neural network site:https://www.accessdata.fda.gov/cdrh_docs/pdf21/”

“neural network site:https://www.accessdata.fda.gov/cdrh_docs/pdf20/”

181 results across all 6 searches - 75 duplicates = 106 results

106 authorization documents concerning 104 device applications were downloaded.

Search of FDA List of AI-Enabled Devices

AI in COVID-19 (all AI-enabled devices authorized by the FDA via the 510(k), De Novo, or PMA process, in 2020 or 2021)

https://www.fda.gov/medical-devices/software-medical-device-samd/artificial-intelligence-and-machine-learning-aiml-enabled-medical-devices

Search conducted December 30, 2021

Webpage states: “Content current as of September 22, 2021”

The list includes 137 AI-enabled devices that were authorized by the FDA during the period from December 31, 2019, through June 17, 2021, the date of the latest authorization.

We downloaded all 143 FDA CDRH authorization documents on these 137 devices.

After removing duplicates from the earlier Google searches of the FDA CDRH document library, we were left with 70 documents concerning 67 device authorizations.

Total combined results for both the FDA list of AI-enabled devices and the Google searches of the FDA CDRH document library were 176 documents concerning 171 device authorizations.

These 171 device authorizations included:

- 1 device authorized under the PMA process
- 163 devices authorized under the 510(k) determination of substantial equivalence process
- 7 devices authorized under De Novo process.

The text of all 176 documents were searched for the presence of any of the following terms: 'COVID', 'coronavirus', 'nCOV', or 'SARS'.

Results: Zero documents that mention any of these COVID-19 terms

--

https://www.fda.gov/medical-devices/emergency-use-authorizations-medical-devices/coronavirus-disease-2019-covid-19-emergency-use-authorizations-medical-devices

The FDA has issued COVID-19 EUAs for many medical devices.

On December 30, 2021, we downloaded 1017 PDF documents from the FDA "Coronavirus Disease 2019 (COVID-19) Emergency Use Authorizations for Medical Devices" webpage (last updated 11/15/2021, https://www.fda.gov/medical-devices/emergency-use-authorizations-medical-devices/coronavirus-disease-2019-covid-19-emergency-use-authorizations-medical-devices) and 8 of its subpages:

- https://www.fda.gov/medical-devices/coronavirus-disease-2019-covid-19-emergency-use-authorizations-medical-devices/blood-purification-devices-euas (content current as of 7/15/2021)
- https://www.fda.gov/medical-devices/coronavirus-disease-2019-covid-19-emergency-use-authorizations-medical-devices/continuous-renal-replacement-therapy-and-hemodialysis-devices-euas (content current as of 7/15/2021)

- https://www.fda.gov/medical-devices/coronavirus-disease-2019-covid-19-emergency-use-authorizations-medical-devices/in-vitro-diagnostics-euas (content current as of 11/15/2021)
- https://www.fda.gov/medical-devices/coronavirus-disease-2019-covid-19-emergency-use-authorizations-medical-devices/remote-or-wearable-patient-monitoring-devices-euas (content current as of 7/15/2021)
- https://www.fda.gov/medical-devices/coronavirus-disease-2019-covid-19-emergency-use-authorizations-medical-devices/infusion-pump-euas (content current as of 7/15/2021)
- https://www.fda.gov/medical-devices/coronavirus-disease-2019-covid-19-emergency-use-authorizations-medical-devices/respiratory-assist-devices-euas (content current as of 7/15/2021)
- https://www.fda.gov/medical-devices/coronavirus-disease-2019-covid-19-emergency-use-authorizations-medical-devices/ventilators-and-ventilator-accessories-euas (content current as of 7/15/2021)
- https://www.fda.gov/medical-devices/coronavirus-disease-2019-covid-19-emergency-use-authorizations-medical-devices/other-medical-device-euas (content current as of 7/15/2021)
- Note: we did not download any documents from the personal protective equipment EUA subpage (https://www.fda.gov/medical-devices/coronavirus-disease-2019-covid-19-emergency-use-authorizations-medical-devices/personal-protective-equipment-euas)

We searched the text of all of these 1,017 EUA pdf documents to determine if any of the authorized devices involved the use of AI, using the following terms.

Search Terms: "machine learning" OR "artificial intelligence" OR "deep learning" OR "neural net*" OR "support vector machine*" OR SVM OR "random forest*" OR "supervised learning" OR "unsupervised learning" OR "reinforcement learning" OR "unsupervised clustering" OR "unsupervised classification" OR "supervised classification" OR "natural language processing" OR NLP OR "gradient boost*" OR "ensemble model" OR "expert system*" OR "rules engine*" OR "fuzzy logic" OR algorithm

Results: 51 COVID-19 EUA documents

--

On March 6, 2022, we also searched the CDC's database of CDC-Authored Genomics and Precision Health Publications using the following search strings:

- *COVID "machine learning"*
- *COVID "artificial intelligence"*
- *SARS-COV-2 "machine learning"*
- *SARS-COV-2 "artificial intelligence"*

CDC publication database URL:
https://phgkb.cdc.gov/PHGKB/cdcPubFinder.action?Mysubmit=init&action=search&query=&dbTypeChoice=All

Results: Zero CDC-authored publications

Abbreviations

AI	artificial intelligence
CDRH	Center for Devices and Radiological Health
EUA	Emergency Use Authorization
FDA	US Food and Drug Administration
IEEE	Institute of Electrical and Electronics Engineers
ML	machine learning
PCORI	Patient-Centered Outcomes Research Institute
PRISMA	Preferred Reporting Items for Systematic Reviews and Meta-Analyses
WHO	World Health Organization

References

1. Panch T, Mattie H, Atun R. Artificial intelligence and algorithmic bias: implications for health systems. *J Glob Health*. 2019;9(2). doi:10.7189/jogh.09.020318
2. Thomasian NM, Eickhoff C, Adashi EY. Advancing health equity with artificial intelligence. *J Public Health Policy*. 2021;42(4):602-611. doi:10.1057/s41271-021-00319-5
3. Rajkomar A, Hardt M, Howell MD, Corrado G, Chin MH. Ensuring fairness in machine learning to advance health equity. *Ann Intern Med*. 2018;169(12):866-872. doi:10.7326/M18-1990
4. Röösli 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(1):190-192. doi:10.1093/jamia/ocaa210
5. Ghassemi M, Mohamed S. Machine learning and health need better values. *NPJ Digit Med*. 2022;5(1):51. doi:10.1038/s41746-022-00595-9
6. Leslie D, Mazumder A, Peppin A, Wolters MK, Hagerty A. Does "AI" stand for augmenting inequality in the era of COVID-19 healthcare? *BMJ*. 2021;372:n304. doi:10.1136/bmj.n304
7. McCoy LG, Banja JD, Ghassemi M, Celi LA. Ensuring machine learning for healthcare works for all. *BMJ Health Care Inform*. 2020;27(3):e100237. doi:10.1136/bmjhci-2020-100237
8. Panch T, Mattie H, Celi LA. The "inconvenient truth" about AI in healthcare. *NPJ Digit Med*. 2019;2(1):77. doi:10.1038/s41746-019-0155-4
9. Panch T, Pearson-Stuttard J, Greaves F, Atun R. Artificial intelligence: opportunities and risks for public health. *Lancet Digit Health*. 2019;1(1):e13-e14. doi:10.1016/S2589-7500(19)30002-0
10. 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):447-453. doi:10.1126/science.aax2342
11. Mude W, Oguoma VM, Nyanhanda T, Mwanri L, Njue C. Racial disparities in COVID-19 pandemic cases, hospitalisations, and deaths: a systematic review and meta-analysis. *J Glob Health*. 2021;11:05015.
12. Okonkwo NE, Aguwa UT, Jang M, et al. COVID-19 and the US response: accelerating health inequities. *BMJ EBM*. 2021;26(4):176-179. doi:10.1136/bmjebm-2020-111426
13. LaVeist TA. *Minority Populations and Health: an Introduction to Health Disparities in the United States*. John Wiley & Sons; 2011.
14. Clark R. World health inequality: convergence, divergence, and development. *Soc Sci Med*. 2011;72(4):617-624. doi:10.1016/j.socscimed.2010.12.008
15. Escarce J. Health inequity in the United States: a primer. Leonard Davis Institute of Health Economics. Accessed December 14, 2022. https://ldi.upenn.edu/wp-content/uploads/2021/06/Penn-LDI-Health-Inequity-in-the-United-States-Report_5.pdf

16. Osoba OA, Welser WIV. *An Intelligence in Our Image: The Risks of Bias and Errors in Artificial Intelligence*. RAND Corporation; 2017.

17. Yeung D, Khan I, Kalra N, Osoba OA. *Identifying Systemic Bias in the Acquisition of Machine Learning Decision Aids for Law Enforcement Applications*. RAND Corporation; 2021.

18. Osoba OA, Boudreaux B, Saunders J, Irwin JL, Mueller PA, Cherney S. *Algorithmic Equity: A Framework for Social Applications*. RAND Corporation; 2019.

19. Mhasawade V, Zhao Y, Chunara R. Machine learning and algorithmic fairness in public and population health. *Nat Mach Intell*. 2021;3(8):659-666. doi:10.1038/s42256-021-00373-4

20. Thomasian NM, Eickhoff C, Adashi EY. Advancing health equity with artificial intelligence. *J Public Health Pol*. 2021;**42**:602–611. doi:10.1057/s41271-021-00319-5

21. Pérez-Stable EJ, Jean-Francois B, Aklin CF. Leveraging advances in technology to promote health equity. *Med Care*. 2019;57(suppl 6):S101-S103. doi:10.1097/MLR.0000000000001112

22. Veinot TC, Ancker JS, Bakken S. Health informatics and health equity: improving our reach and impact. *J Am Med Inform Assoc*. 2019;26(8-9):689-695. doi:10.1093/jamia/ocz132

23. Chen IY, Szolovits P, Ghassemi M. Can AI help reduce disparities in general medical and mental health care? *AMA J Ethics*. 2019;21(2):E167-179. doi:10.1001/amajethics.2019.167

24. Chen IY, Joshi S, Ghassemi M. Treating health disparities with artificial intelligence. *Nat Med*. 2020;26(1):16-17. doi:10.1038/s41591-019-0649-2

25. Berdahl CT, Baker L, Mann S, Osoba O, Girosi F. Strategies to Improve the Impact of Artificial Intelligence on Health Equity: Scoping Review. *JMIR AI*. 2023;2:e42936 doi:10.2196/42936

26. Smedley A, Smedley BD. Race as biology is fiction, racism as a social problem is real: Anthropological and historical perspectives on the social construction of race. *Am Psychol*. Jan 2005;60(1):16-26. doi:10.1037/0003-066x.60.1.16

27. Mann S, Berdahl CT, Baker L, Girosi F. Artificial intelligence applications used in the clinical response to COVID-19: a scoping review. *PLOS Digit Health*. 2022;1(10):e0000132. doi:10.1371/journal.pdig.0000132

28. Mann S, Berdahl CT, Baker L, Girosi F. *Artificial Intelligence in the COVID-19 Response, Volume 1: Applications Used in the Clinical and Public Health Response to COVID-19*. Patient-Centered Outcomes Research Institute. 2022.

29. PCORI. PCORI's stakeholders. Updated October 9, 2018. Accessed 12/21/2020, https://www.pcori.org/about-us/our-programs/engagement/public-and-patient-engagement/pcoris-stakeholders

30. Arksey H, O'Malley L. Scoping studies: towards a methodological framework. *Int J Soc Res Methodol*. 2005;8(1):19-32. doi:10.1080/1364557032000119616

31. Miake-Lye IM, Hempel S, Shanman R, Shekelle PG. What is an evidence map? A systematic review of published evidence maps and their definitions, methods, and products. *Syst Rev*. 2016;5(1):28. doi:10.1186/s13643-016-0204-x

32. Syrowatka A, Kuznetsova M, Alsubai A, et al. Leveraging artificial intelligence for pandemic preparedness and response: a scoping review to identify key use cases. *NPJ Digit Med*. 2021;4(1):96. doi:10.1038/s41746-021-00459-8

33. World Health Organization. Social determinants of health: health equity. Accessed November 13, 2022. https://www.who.int/health-topics/social-determinants-of-health

34. Braveman P, Arkin E, Orleans T, Proctor D, Plough A. *What is health equity? And what difference does a definition make?* National Collaborating Centre for Determinants of Health. 2017. Accessed February 9, 2023. https://www.rwjf.org/en/insights/our-research/2017/05/what-is-health-equity-.html

35. Vokinger KN, Feuerriegel S, Kesselheim AS. Mitigating bias in machine learning for medicine. *Commun Med*. 2021;1(1):25. doi:10.1038/s43856-021-00028-w

36. Chen IY, Pierson E, Rose S, Joshi S, Ferryman K, Ghassemi M. Ethical machine learning in healthcare. *Annu Rev Biomed Data Sci*. 2021;4(1):123-144. doi:10.1146/annurev-biodatasci-092820-114757

37. Zdaczyk C, Vallee AB. Gender bias in AI: what the United States, Canada, and Western European nations can learn from other countries. OECD.AI Policy Observatory. Accessed December 14, 2022. https://oecd.ai/en/wonk/ai-national-gender-bias

38. Zimmer M, Franco Z, Madiraju P, Echeveste C, Heindel K, Ogle J. Public opinion research on artificial intelligence in public health responses: results of focus groups with four communities. AAAS Center for Public Engagement with Science and Technology. Accessed December 14, 2022. https://www.aaas.org/sites/default/files/2021-09/AI%20in%20Public%20Health%20Focus%20Groups%20-%20Final%20Report%20with%20Appendix.pdf

39. The Alan Turing Institute. Data science and AI in the age of COVID-19. Accessed December 14, 2022. https://www.turing.ac.uk/sites/default/files/2021-06/data-science-and-ai-in-the-age-of-covid_full-report_2.pdf

40. O'Brien N, Van Dael J, Clarke K, et al. Addressing racial and ethnic inequities in data-driven health technologies. Updated February 24, 2022. Accessed December 14, 2022. https://spiral.imperial.ac.uk/bitstream/10044/1/94902/2/Imperial_IGHI_AddressingRacialandEthnicInequities%20_Report.pdf

41. Stathoulopoulos K, Mateos-Garcia JC. Gender diversity in AI research. Accessed December 14, 2022. https://media.nesta.org.uk/documents/Gender_Diversity_in_AI_Research.pdf

42. World Economic Forum. Assessing gender gaps in artificial intelligence. Accessed December 14, 2022. https://reports.weforum.org/global-gender-gap-report-2018/assessing-gender-gaps-in-artificial-intelligence/

43. Google. Google diversity annual report 2022. Accessed December 14, 2022. https://about.google/belonging/diversity-annual-report/2022/static/pdfs/google_2022_diversity_annual_report.pdf

44. Facebook. Annual diversity report. Accessed December 14, 2022. https://diversity.fb.com/wp-content/uploads/2021/07/Facebook-Annual-Diversity-Report-July-2021.pdf

45. Microsoft. Global diversity & inclusion report. Accessed December 14, 2022. https://aka.ms/2021DIReport

46. Ledford H. Millions of black people affected by racial bias in health-care algorithms. News. *Nature*. 2019;574(7780):608-609. doi:10.1038/d41586-019-03228-6

47. Ziarek EP. Triple pandemics: COVID-19, anti-black violence, and digital capitalism. *Philos Today*. 2020;64(4):925-930. doi:10.5840/philtoday20201124377

48. Murphy K, Di Ruggiero E, Upshur R, et al. Artificial intelligence for good health: a scoping review of the ethics literature. *BMC Med Ethics*. 2021;22(1):14. doi:10.1186/s12910-021-00577-8

49. Holstein K, Wortman Vaughan J, Daumé III H, Dudik M, Wallach H. Improving fairness in machine learning systems: what do industry practitioners need? *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. 2019:1-16. doi:10.1145/F3290605.3300830

50. Vourganas I, Stankovic V, Stankovic L. Individualised responsible artificial intelligence for home-based rehabilitation. *Sensors (Basel)*. 2020;21(1). doi:10.3390/s21010002

51. Luengo-Oroz M, Bullock J, Pham KH, Lam CSN, Luccioni A. From artificial intelligence bias to inequality in the time of COVID-19. *IEEE Technol Soc Mag*. 2021;40(1):71-79. doi:10.1109/mts.2021.3056282

52. Mahomed S. COVID-19: the role of artificial intelligence in empowering the healthcare sector and enhancing social distancing measures during a pandemic. *S Afr Med J*. 2020;110(7):610-612. doi:10.7196/SAMJ.2020.v110i7.14841

53. Kiener M. Artificial intelligence in medicine and the disclosure of risks. *AI Soc*. 2020:1-9. doi:10.1007/s00146-020-01085-w

54. Whittaker Z. This startup reworked its privacy-friendly sensors to help battle COVID-19. TechCrunch. Updated July 23, 2020. Accessed December 14, 2022. https://techcrunch.com/2020/07/23/butlr-privacy-sensors-coronavirus/

55. Paul A, Jolley C, Anthony A. Reflecting the past, shaping the future: making AI work for international development. USAID. 2018. Accessed March 7, 2022. https://www.usaid.gov/digital-development/machine-learning/AI-ML-in-development

56. Bynum J, Lewis V. Value-based payments and inaccurate risk adjustment—who is harmed? *JAMA Intern Med*. 2018;178(11):1507-1508. doi:10.1001/jamainternmed.2018.4142

57. Ross C. As the FDA clears a flood of AI tools, missing data raise troubling questions on safety and fairness. February 3, 2021. Accessed March 8, 2022. https://www.statnews.com/2021/02/03/fda-clearances-artificial-intelligence-data/

58. Deonarine A, Lyons G, Lakhani C, De Brouwer W. Identifying communities at risk for COVID-19–related burden across 500 US cities and within New York City: unsupervised learning of the coprevalence of health indicators. *JMIR Public Health Surveill*. 2021;7(8):e26604. doi:10.2196/26604

59. Pham Q, Gamble A, Hearn J, Cafazzo JA. The need for ethnoracial equity in artificial intelligence for diabetes management: review and recommendations. *J Med Internet Res*. 2021;23(2):e22320. doi:10.2196/22320

60. Rodolfa KT, Saleiro P, Ghani R. Bias and fairness (Chapter 11). *Big Data and Social Science*. Chapman & Hall/CRC; 2019.

61. Vyas DA, Eisenstein LG, Jones DS. Hidden in plain sight—reconsidering the use of race correction in clinical algorithms. *N Engl J Med*. 2020;383(9):874-882. doi:10.1056/NEJMms2004740

62. Asiimwe IG, Zhang EJ, Osanlou R, Jorgensen AL, Pirmohamed M. Warfarin dosing algorithms: a systematic review. *Br J Clin Pharmacol*. 2021;87(4):1717-1729. doi:10.1111/bcp.14608

63. Stevens LA, Claybon MA, Schmid CH, et al. Evaluation of the chronic kidney disease epidemiology collaboration equation for estimating the glomerular filtration rate in multiple ethnicities. *Kidney Int*. 2011;79(5):555-562. doi:10.1038/ki.2010.462

64. ITU-T focus group on artificial intelligence for health (FG-AI4H). Data and artificial intelligence assessment methods (DAISAM) reference. 2020. Accessed March 1, 2022. https://www.itu.int/en/ITU-T/focusgroups/ai4h/Documents/all/FGAI4H-I-035.docx

65. Chouldechova A, Roth A. The frontiers of fairness in machine learning. *arXiv preprint*. 2018. Doi:10.48550/ARXIV.1810.08810

66. Peiffer-Smadja N, Rawson TM, Ahmad R, et al. Machine learning for clinical decision support in infectious diseases: a narrative review of current applications. *Clin Microbiol Infect*. 2020;26(5):584-595. doi:10.1016/j.cmi.2019.09.009

67. Google. Case study: responsible AI: the fitbit blood oxygen saturation (Spo2) tracking feature. Accessed December 14, 2022. https://ai.google/static/documents/case-study-blood-oxygen-saturation-tracking.pdf

68. Wynants L, Smits LJM, Van Calster B. Demystifying AI in healthcare. *BMJ*. 2020;370:m3505. doi:10.1136/bmj.m3505

69. Yeung D, Khan I, Kalra N, Osoba O. Identifying systemic bias in the acquisition of machine learning decision aids for law enforcement applications. *RAND Corporation*; 2021.

70. Straw I. The automation of bias in medical artificial intelligence (AI): decoding the past to create a better future. *Artif Intell Med*. 2020;110:101965. doi:10.1016/j.artmed.2020.101965

71. Liu HR, Li L, Wormstone IM, et al. Development and validation of a deep learning system to detect glaucomatous optic neuropathy using fundus photographs. *JAMA Ophthalmol*. 2019;137(12):1353-1360. doi:10.1001/jamaophthalmol.2019.3501

72. O'Neil C. *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*. Crown Publishing Group; 2016.

73. Harwich E, Laycock K. Thinking on its own: AI in the NHS. 2018. Accessed December 14, 2022. https://www.wiltonpark.org.uk/wp-content/uploads/Thinking-on-its-own-AI-in-the-NHS.pdf

74. Caliskan A, Bryson JJ, Narayanan A. Semantics derived automatically from language corpora contain human-like biases. *Science*. 2017;356(6334):183-186. DOI:10.1126/science.aal42

75. Gianfrancesco MA, Tamang S, Yazdany J, Schmajuk G. Potential biases in machine learning algorithms using electronic health record data. *JAMA Intern Med*. 2018;178(11):1544-1547. doi:10.1001/jamainternmed.2018.3763

76. Badreldin N, Grobman WA, Yee LM. Racial disparities in postpartum pain management. *Obstetr Gynecol*. 2019;134(6):1147. doi:10.1097/AOG.0000000000003561

77. Anderson KO, Green CR, Payne R. Racial and ethnic disparities in pain: causes and consequences of unequal care. *J Pain*. 2009;10(12):1187-1204. doi:10.1016/j.jpain.2009.10.002

78. Shahian DM, Jacobs JP, Badhwar V, et al. The Society of Thoracic Surgeons 2018 adult cardiac surgery risk models: Part 1-Background, design considerations, and model development. *Ann Thorac Surg*. 2018;105(5):1411-1418. doi:10.1016/j.athoracsur.2018.03.002

79. Grobman WA, Lai Y, Landon MB, et al. Development of a nomogram for prediction of vaginal birth after cesarean delivery. *Obstet Gynecol*. 2007;109(4):806-12. doi:10.1097/01.AOG.0000259312.36053.02

80. Inker LA, Eneanya ND, Coresh J, et al. New creatinine- and cystatin c–based equations to estimate GFR without race. *N Engl J Med*. 2021;385(19):1737-1749. doi:10.1056/NEJMoa2102953

81. Delgado C, Baweja M, Crews DC, St. Peter WL, Warfield C, Powe NR. A unifying approach for GFR estimation: recommendations of the NKF-ASN task force on reassessing the inclusion of race in diagnosing kidney disease. *J Am Soc Nephrol.* 2021;32(12):2994-3015. doi:10.1053/j.ajkd.2021.08.003

82. Diao JA, Inker LA, Levey AS, Tighiouart H, Powe NR, Manrai AK. In search of a better equation—performance and equity in estimates of kidney function. *N Engl J Med*. 2021;384(5):396-399. doi:10.1056/NEJMp2028243

83. Benito-León J, Del Castillo MD, Estirado A, Ghosh R, Dubey S, Serrano JI. Using unsupervised machine learning to identify age- and sex-independent severity subgroups among patients with COVID-19: observational longitudinal study. *J Med Internet Res*. 2021;23(5):e25988. doi:10.2196/25988

84. Eneanya ND, Yang W, Reese PP. Reconsidering the consequences of using race to estimate kidney function. *JAMA*. 2019;322(2):113-114. doi:10.1001/jama.2019.5774

85. Sjoding MW, Dickson RP, Iwashyna TJ, Gay SE, Valley TS. Racial bias in pulse oximetry measurement. *N Engl J Med*. 2020;383(25):2477-2478. doi:10.1056/NEJMc2029240

86. Fine J, Branan KL, Rodriguez AJ, et al. Sources of inaccuracy in photoplethysmography for continuous cardiovascular monitoring. *Biosensors (Basel)*. 2021;11(4):126. doi:10.3390/bios11040126

87. Public Health Alliance of Southern California, Virginia Commonwealth University. Healthy Places Index (HPI 2.0). 2021. Accessed December 14, 2022. https://healthyplacesindex.org/wp-content/uploads/2021/04/HPI2Documentation2018-02-20-FINALrev2021-04-22.pdf

88. Schmidt H, Ünver U, Williams M, Pathak P, Sönmez T, Gostin L. Blueprint Labs Discussion Paper #2020.11: What prioritizing worse-off minority groups for COVID-19 vaccines means quantitatively: practical, legal, and ethical implications. October 2020. Accessed December 14, 2022. https://blueprintcdn.com/wp-content/uploads/2020/11/Blueprint-Labs-Discussion-Paper-2020.11-SchmidtUnverWilliams-PathakSonmezGostin-1.pdf

89. Singh GK. Area deprivation and widening inequalities in US mortality, 1969-1998. Am J Public Health. 2003;93(7):1137-43. doi:10.2105/ajph.93.7.1137

90. Gallifant J, Zhang J, Del Pilar Arias Lopez M, et al. Artificial intelligence for mechanical ventilation: systematic review of design, reporting standards, and bias. *Br J Anaesth*. 2022;128(2):343-351. doi:10.1016/j.bja.2021.09.025

91. Wu E, Wu K, Daneshjou R, Ouyang D, Ho DE, Zou J. How medical AI devices are evaluated: limitations and recommendations from an analysis of FDA approvals. *Nat Med*. 2021;27(4):582-584. doi:10.1038/s41591-021-01312-x

92. Hooker S. Moving beyond “algorithmic bias is a data problem.” *Patterns*. 2021;2(4):100241. doi:10.1016/j.patter.2021.100241

93. Nundy S, Montgomery T, Wachter RM. Promoting trust between patients and physicians in the era of artificial intelligence. *JAMA*. 2019;322(6):497-498. doi:10.1001/jama.2018.20563

94. ACM US Public Policy Council. Statement on algorithmic transparency and accountability. Accessed March 1, 2022. https://www.acm.org/binaries/content/assets/public-policy/2017_usacm_statement_algorithms.pdf

95. Mullainathan S. Biased algorithms are easier to fix than biased people. *New York Times.* Updated December 6, 2019. Accessed December 14, 2022. https://www.nytimes.com/2019/12/06/business/algorithm-bias-fix.html

96. Palmer K. Changing the equation: researchers remove race from a calculator for childbirth. STAT. Updated June 3, 2021. Accessed December 14, 2022, https://www.statnews.com/2021/06/03/vbac-calculator-birth-cesarean/

97. Barocas S, Hardt M, Narayanan A. Fairness and machine learning. Accessed December 14, 2022. https://fairmlbook.org/pdf/fairmlbook.pdf

98. Bateni A, Chan MC, Eitel-Porter R. AI fairness: from principles to practice. *arXiv preprint*. 2022. doi:10.48550/arxiv.2207.09833

99. Corbett-Davies S, Goel S. The measure and mismeasure of fairness: a critical review of fair machine learning. *arXiv preprint*. 2018. doi:10.48550/ARXIV.1808.00023

100. Rodolfa KT, Lamba H, Ghani R. Empirical observation of negligible fairness–accuracy trade-offs in machine learning for public policy. *Nat Mach Intell*. 2021;3(10):896-904. doi:10.1038/s42256-021-00396-x

101. Horvitz E, Clyburn M, Griffiths J-M, Matheny J. Privacy and ethics recommendations for computing applications developed to mitigate COVID-19: white paper series on pandemic response and preparedness, no. 1. May 6, 2020. Accessed March 7, 2022. https://www.nscai.gov/wp-content/uploads/2021/01/NSCAI_White-Paper_Privacy-and-Ethics-Recommendations-for-Computing-Applications-Developed-to-Mitigate-COVID-19.pdf

102. Rodolfa KT, Salomon E, Haynes L, Mendieta IH, Larson J, Ghani R. Case study: predictive fairness to reduce misdemeanor recidivism through social service interventions. *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*. 2020:142-153. doi:10.1145%2F3351095.3372863

103. Bowen S, Khoury MJ, Moonesinghe R. "Precision" health tools and… increased health disparities? *Centers for Disease Control and Prevention, Office of Genomics and Precision Health.* January 8, 2020. Accessed December 14, 2022. https://blogs.cdc.gov/genomics/2020/01/08/precision-health-tools/

104. Geis JR, Brady AP, Wu CC, et al. Ethics of artificial intelligence in radiology: summary of the Joint European And North American Multisociety statement. *Radiology*. 2019;293(2):436-440. doi:10.1148/radiol.2019191586

105. Engler A. A guide to healthy skepticism of artificial intelligence and coronavirus. 2020. Accessed March 1, 2022. https://www.brookings.edu/research/a-guide-to-healthy-skepticism-of-artificial-intelligence-and-coronavirus/

106. Asan O, Bayrak AE, Choudhury A. Artificial intelligence and human trust in healthcare: focus on clinicians. *J Med Internet Res*. 2020;22(6):e15154-e15154. doi:10.2196/15154

107. Singh K, Valley TS, Tang S, et al. Evaluating a widely implemented proprietary deterioration index model among hospitalized COVID-19 patients. *medRxiv*. 2020:2020.04.24.20079012. doi:10.1101/2020.04.24.20079012

108. Singh K, Valley TS, Tang S, et al. Evaluating a widely implemented proprietary deterioration index model among hospitalized patients with COVID-19. *Ann Am Thorac Soc*. 2021;18(7):1129-1137. doi:10.1513/AnnalsATS.202006-698OC

109. Sipior JC. Considerations for development and use of AI in response to COVID-19. Article. *Int J Inf Manage*. 2020;55:6.102170. doi:10.1016/j.ijinfomgt.2020.102170

110. Dankwa-Mullan I, Scheufele EL, Matheny ME, et al. A proposed framework on integrating health equity and racial justice into the artificial intelligence development lifecycle. Editorial Material. *J Health Care Poor Underserved*. 2021;32(2):300-317. doi:10.1353/hpu.2021.0065

111. Mbow M, Lell B, Jochems SP, et al. COVID-19 in Africa: dampening the storm? *Science*. 2020;369(6504):624-626. doi:10.1126/science.abd3902

112. Fernandez-Luque L, Imran M. Humanitarian health computing using artificial intelligence and social media: a narrative literature review. *Int J Med Inform*. 2018;114:136-142. doi:10.1016/j.ijmedinf.2018.01.015

113. Mörch CM, Atsu S, Cai W, et al. Artificial intelligence and ethics in dentistry: a scoping review. *J Dent Res*. 2021;100(13):1452-1460. doi:10.1177/00220345211013808

114. Jobin A, Ienca M, Vayena E. The global landscape of AI ethics guidelines. *Nat Mach Intell*. 2019;1(9):389-399. doi:10.1038/s42256-019-0088-2

115. Naude W, Vinuesa R. Data deprivations, data gaps and digital divides: lessons from the COVID-19 pandemic. Article. *Big Data Soc*. 2021;8(2):12. doi/10.1177/20539517211025545

116. Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. *Nat Med*. 2019;25(1):44-56. doi:10.1038/s41591-018-0300-7

117. Bahia K, Suardi S. The state of mobile internet connectivity. Accessed December 14, 2022. https://www.gsma.com/mobilefordevelopment/wp-content/uploads/2019/07/GSMA-State-of-Mobile-Internet-Connectivity-Report-2019.pdf

118. Whitelaw S, Mamas MA, Topol E, Van Spall HGC. Applications of digital technology in COVID-19 pandemic planning and response. *Lancet Digit Health*. 2020;2(8):e435-e440. doi:10.1016/S2589-7500(20)30142-4

119. IEEE. Ethically aligned design: a vision for prioritizing human well-being with autonomous and intelligence systems, Version 2. Accessed December 14, 2022. https://standards.ieee.org/wp-content/uploads/import/documents/other/ead_v2.pdf

120. Ferryman K, Pitcan M. Fairness in precision medicine. Data and Society. February 2018. Accessed March 29, 2022. https://datasociety.net/library/fairness-in-precision-medicine/

121. Ely RJ, Thomas DA. Getting serious about diversity. *Harvard Business Review*. 2020;98(6):114-122.

122. Chin C, Robison M. How AI bots and voice assistants reinforce gender bias. Brookings Institution. Updated November 23, 2020. Accessed April 12, 2022. https://www.brookings.edu/research/how-ai-bots-and-voice-assistants-reinforce-gender-bias/

123. West SM, Whittaker M, Crawford K. Discriminating systems: gender, race, and power in AI. AINOW. 2019. Accessed April 5, 2022. https://ainowinstitute.org/discriminatingsystems.html

124. Taylor S, Pickering B, Boniface M, et al. Responsible AI--key themes, concerns, & recommendations for European research and innovation. European AI Alliance. Accessed April 5, 2022. https://ec.europa.eu/futurium/en/european-ai-alliance/responsible-ai-key-themes-concerns-recommendations-european-research-and.html

125. Kino S, Hsu YT, Shiba K, et al. A scoping review on the use of machine learning in research on social determinants of health: trends and research prospects. *SSM Popul Health*. 2021;15:100836. doi:10.1016/j.ssmph.2021.100836

126. Valle-Cruz D, Ruvalcaba-Gomez EA, Sandoval-Almazan R, Criado JI. A review of artificial intelligence in government and its potential from a public policy perspective. Proceedings of the 20th Annual International Conference on Digital Government Research; 2019:91-99. doi:10.1145/3325112.3325242

127. McCradden MD, Joshi S, Mazwi M, Anderson JA. Ethical limitations of algorithmic fairness solutions in health care machine learning. *Lancet Digit Health*. 2020;2(5):e221-e223. DOI:10.1016/S2589-7500(20)30065-0

128. Morley J, Machado CCV, Burr C, et al. The ethics of AI in health care: a mapping review. *Soc Sci Med*. 2020;260:113172. doi:10.1016/j.socscimed.2020.113172

129. OECD.AI. Human-centred values and fairness (principle 1.2). April 7, 2022. Accessed December 14, 2022. https://oecd.ai/en/dashboards/ai-principles/P6

130. Shachar C, Gerke S, Adashi EY. AI surveillance during pandemics: ethical implementation imperatives. *Hastings Cent Rep*. 2020;50(3):18-21. doi:10.1002/hast.1125

131. Ortiz K, Nash J, Shea L, et al. Partnerships, processes, and outcomes: a health equity–focused scoping meta-review of community-engaged scholarship. *Annu Rev Public Health*. 2020;41(1):177-199. doi:10.1146/annurev-publhealth-040119-094220

132. Damian AJ, Gonzalez M. Dismantling racism in research. *Lancet*. 2020;396(10246):237. doi:10.1016/S0140-6736(20)31534-8

133. Balcombe L, De Leo D. Digital mental health challenges and the horizon ahead for solutions comment. *JMIR Ment Health*. 2021;8(3):11.e26811. doi:10.2196/26811

134. Bricout J, Greer J, Fields N, et al. The "humane in the loop": inclusive research design and policy approaches to foster capacity building assistive technologies in the COVID-19 era. *Assist Technol*. 2021:9. doi:10.1080/10400435.2021.1930282

135. Zhao IY, Ma YX, Yu MWC, et al. Ethics, integrity, and retributions of digital detection surveillance systems for infectious diseases: systematic literature review. *J Med Internet Res*. 2021;23(10):e32328. doi:10.2196/32328

136. National Center for Advancing Translational Sciences. N3C tribal consultation. Updated December 16, 2021. Accessed March 2, 2022. https://ncats.nih.gov/n3c/about/tribal-consultation

137. Reddy S, Allan S, Coghlan S, Cooper P. A governance model for the application of AI in health care. *J Am Med Inf Assoc*. 2020;27(3):491-497. doi:10.1093/jamia/ocz192

138. The Toronto Declaration. Protecting the right to equality and non-discrimination in machine learning systems. 2018. Accessed April 7, 2022. https://www.accessnow.org/cms/assets/uploads/2018/08/The-Toronto-Declaration_ENG_08-2018.pdf

139. Future of Life Institute. AI principles. Accessed December 14, 2022. https://futureoflife.org/open-letter/ai-principles/

140. Vought RT. Guidance for regulation of artificial intelligence applications. Accessed December 14, 2022. https://www.whitehouse.gov/wp-content/uploads/2020/01/Draft-OMB-Memo-on-Regulation-of-AI-1-7-19.pdf

141. Alami H, Rivard L, Lehoux P, et al. Artificial intelligence in health care: laying the foundation for responsible, sustainable, and inclusive innovation in low- and middle-income countries. *Glob Health*. 2020;16(1):52. doi:10.1186/s12992-020-00584-1

142. OECD.AI Policy Observatory. National AI policies & strategies. Accessed December 14, 2022. https://oecd.ai/en/dashboards

143. OECD.AI Policy Observatory. Chilean participation process on AI. Accessed April 7, 2022. https://oecd.ai/en/dashboards/policy-initiatives/http:%2F%2Faipo.oecd.org%2F2021-data-policyInitiatives-26732

144. Wiens J, Saria S, Sendak M, et al. Do no harm: a roadmap for responsible machine learning for health care. *Nat Med*. 2019;25(9):1337-1340. doi:10.1038/s41591-019-0548-6

145. Google. Artificial intelligence at Google: our principles. Accessed December 14, 2022. https://ai.google/principles/

146. World Health Organization. Ethics and governance of artificial intelligence for health. Accessed September 26, 2022. https://apps.who.int/iris/bitstream/handle/10665/341996/9789240029200-eng.pdf

147. World Health Organization. WHO hub for pandemic and epidemic intelligence: better data. better analytics. better decisions. Accessed December 14, 2022. https://cdn.who.int/media/docs/default-source/2021-dha-docs/who_hub.pdf?sfvrsn=8dc28ab6_5

148. U.S. Food and Drug Administration. Learn if a medical device has been cleared by FDA for marketing. Accessed December 14, 2022. https://www.fda.gov/medical-devices/consumers-medical-devices/learn-if-medical-device-has-been-cleared-fda-marketing

149. U.S. Food and Drug Administration. FDA permits marketing of artificial intelligence algorithm for aiding providers in detecting wrist fractures. Accessed December 14, 2022. https://www.fda.gov/news-events/press-announcements/fda-permits-marketing-artificial-intelligence-algorithm-aiding-providers-detecting-wrist-fractures

150. U.S. Food and Drug Administration, Health Canada, United Kingdom Medicines and Healthcare products Regulatory Agency. Good machine learning practice for medical device development: guiding principles. Updated October. Accessed December 14, 2022. https://www.fda.gov/medical-devices/software-medical-device-samd/good-machine-learning-practice-medical-device-development-guiding-principles

151. Villasenor J. Soft law as a complement to AI regulation. Brookings Institution. Updated July 31, 2020. Accessed December 14, 2022. https://www.brookings.edu/research/soft-law-as-a-complement-to-ai-regulation/

152. Lundqvist B. *Standardization Under EU Competition Rules and US Antitrust Laws: The Rise and Limits of Self-regulation.* Edward Elgar Publishing; 2014.

153. Hytha DA, Aronson JD, Eng A. Technology innovation and the rebirth of self-regulation: how the internet of things, cloud computing, blockchain, and artificial intelligence solve big problems managing environmental regulation and resources. *Int J Commun*. 2019;13:5.

154. Wynants L, Van Calster B, Bonten MMJ, et al. Prediction models for diagnosis and prognosis of COVID-19 infection: systematic review and critical appraisal. *BMJ*. 2020;369:11.m1328. doi:10.1136/bmj.m1328

155. Arora A. Conceptualising artificial intelligence as a digital healthcare innovation: an introductory review. *Med Devices (Auckl)*. 2020;13:223-230. doi:10.2147/mder.S262590

156. Obermeyer Z, Nissan R, Stern M, Eaneff S, Bembeneck EJ, Mullainathan S. Algorithmic bias playbook. Chicago Booth. June 2021. Accessed March 8, 2022. https://www.chicagobooth.edu/research/center-for-applied-artificial-intelligence/research/algorithmic-bias/playbook

157. ClinicalTrials.gov. Testing the accuracy of a digital test to diagnose Covid-19. Accessed March 1, 2022. https://clinicaltrials.gov/ct2/show/NCT04407585

158. Khan SM, Liu XX, Nath S, et al. A global review of publicly available datasets for ophthalmological imaging: barriers to access, usability, and generalisability. *Lancet Digit Health*. 2021;3(1):E51-E66. doi:10.1016/S2589-7500(20)30240-5

159. Arges K, Assimes T, Bajaj V, et al. The Project Baseline Health Study: a step towards a broader mission to map human health. *NPJ Digit Med*. 2020;3(1):84. doi:10.1038/s41746-020-0290-y

160. The "All of Us" Research Program. *N Engl J Med*. 2019;381(7):668-676. doi:10.1056/NEJMsr1809937

161. Teo J. Commentary: strategies to address COVID-19 vaccine hesitancy and mitigate health disparities in minority populations. *Front Public Health*. 2021;9:2.723583. doi:10.3389/fpubh.2021.723583

162. Mitchell M, Wu S, Zaldivar A, et al. Model cards for model reporting. Presented at: Proceedings of the Conference on Fairness, Accountability, and Transparency. Atlanta, GA, USA; 2019. https://doi:10.1145/3287560.3287596

163. Kundu A, Chaiton M, Billington R, et al. Machine learning applications in mental health and substance use research among the LGBTQ2S+ population: scoping review. *JMIR Med Inform*. 2021;9(11):e28962. DOI:10.2196/28962

164. Zhao Y, Wood EP, Mirin N, Cook SH, Chunara R. Social determinants in machine learning cardiovascular disease prediction models: a systematic review. *Am J Prev Med*. 2021;61(4):596-605. doi:10.1016/j.amepre.2021.04.016

165. UnitedHealthcare and the AMA collaborate to understand and address social barriers preventing people's access to better health. April 2, 2019. Accessed November 11, 2022. https://www.unitedhealthgroup.com/newsroom/2019/2019-04-02-uhc-ama-social-barriers.html

166. Gajane P, Pechenizkiy M. On formalizing fairness in prediction with machine learning. *arXiv preprint*. 2017. doi:10.48550/ARXIV.1710.03184

167. Verma S, Rubin J. Fairness definitions explained. 2018 IEEE/ACM International Workshop on Software Fairness (FairWare); 2018:1-7. doi:10.23919/FAIRWARE.2018.8452913

168. Dwork C, Hardt M, Pitassi T, Reingold O, Zemel R. Fairness through awareness. Proceedings of the 3rd Innovations in Theoretical Computer Science Conference; 2012:214-226. doi/10.1145/2090236.2090255

169. Grobman WA, Sandoval G, Rice MM, et al. Prediction of vaginal birth after cesarean delivery in term gestations: a calculator without race and ethnicity. *Am J Obstet Gynecol*. 2021;225(6):664.e1-664.e7. doi:10.1016/j.ajog.2021.05.021

170. Morris JE, Grubbs V, Hahn M, Richmond S. Abolish race-based medicine in kidney disease and beyond. *San Francisco Examiner*. November 27, 2019. Accessed December 14, 2022. https://www.sfexaminer.com/opinion/abolish-race-based-medicine-in-kidney-disease-and-beyond/

171. Gaffney T. A yearslong push to remove racist bias from kidney testing gains new ground. STAT. July 17, 2020. Accessed December 14, 2022. https://www.statnews.com/2020/07/17/egfr-race-kidney-test/

172. Delgado C, Baweja M, Crews DC, St. Peter WL, Warfield C, Powe NR. A unifying approach for GFR estimation: recommendations of the NKF-ASN task force on reassessing the inclusion of race in diagnosing kidney disease. *Am J Kidney Dis*. 2022;79(2):268-288.e1. doi:10.1053/j.ajkd.2021.08.003

173. Maternal-Fetal Medicine Units Network. Vaginal birth after cesarean calculator, Version 2.0. Updated May 2021. Accessed April 6, 2022, 2022. https://mfmunetwork.bsc.gwu.edu/web/mfmunetwork/vaginal-birth-after-cesarean-calculator

174. Office of the Attorney General for the District of Columbia. AG Racine introduces legislation to stop discrimination in automated decision-making tools that impact individuals' daily lives. Accessed March 7, 2022. https://oag.dc.gov/release/ag-racine-introduces-legislation-stop

175. Office of the Attorney General for the District of Columbia. Stop Discrimination by Algorithms Act of 2021. Accessed March 30, 2022. https://oag.dc.gov/sites/default/files/2021-12/DC-Bill-SDAA-FINAL-to-file-.pdf

176. Gichoya JW, Banerjee I, Bhimireddy AR, et al. AI recognition of patient race in medical imaging: a modelling study. *The Lancet Digital Health*. 2022;4(6):e406-e414. doi:10.1016/S2589-7500(22)00063-2

177. Williams IK. Can A.I.-driven voice analysis help identify mental disorders? *New York Times*. April 5, 2022. Accessed February 8, 2023. https://www.nytimes.com/2022/04/05/technology/ai-voice-analysis-mental-health.html

178. Simonite T. How coders are fighting bias in facial recognition software. *Wired*. March 29, 2018. Accessed April 6, 2022. https://www.wired.com/story/how-coders-are-fighting-bias-in-facial-recognition-software/

179. Feehan AK, Denstel KD, Katzmarzyk PT, et al. Community versus individual risk of SARS-CoV-2 infection in two municipalities of Louisiana, USA: an assessment of area deprivation index (ADI) paired with seroprevalence data over time. *PLoS One*. 2021;16(11):e0260164. doi:10.1371/journal.pone.0260164

180. Ingraham NE, Purcell LN, Karam BS, et al. Racial and ethnic disparities in hospital admissions from COVID-19: determining the impact of neighborhood deprivation and primary language. *J Gen Intern Med*. 2021;36(11):3462-3470. doi:10.1007/s11606-021-06790-w

181. Magesh S, John D, Li WT, et al. Disparities in COVID-19 outcomes by race, ethnicity, and socioeconomic status: a systematic review and meta-analysis. *JAMA Network Open*. 2021;4(11):e2134147-e2134147. doi:10.1001/jamanetworkopen.2021.34147

182. Bellamy RK, Dey K, Hind M, et al. AI fairness 360: an extensible toolkit for detecting and mitigating algorithmic bias. *IBM J Res Dev*. 2019;63(4/5):4:1-4:15. doi:10.1147/JRD.2019.2942287

183. Friedler S, Scheidegger C, Venkatasubramanian S, Choudhary S, Hamilton E, Roth D. A comparative study of fairness-enhancing interventions in machine learning. *arXiv preprint*. 2018. doi:10.48550/ARXIV.1802.04422

184. Kleinberg J, Lakkaraju H, Leskovec J, Ludwig J, Mullainathan S. Human decisions and machine predictions. *Q J Econ*. 2018;133(1):237-293. doi:10.1093/qje/qjx032

185. Kamishima T, Akaho S, Asoh H, Sakuma J. Fairness-Aware Classifier With Prejudice Remover Regularizer. In: Flach, P.A., De Bie, T., Cristianini, N. (eds) Machine Learning and Knowledge Discovery in Databases. ECML PKDD 2012. Lecture Notes in Computer Science(), vol 7524. Springer, Berlin, Heidelberg. 2012:35-50. doi:10.1007/978-3-642-33486-3_3

186. Zhang BH, Lemoine B, Mitchell M. Mitigating unwanted biases with adversarial learning. *arXiv preprint*. 2018:335-340. DOI:doi:10.48550/arXiv.1801.07593

187. Zafar MB, Valera I, Gomez Rodriguez M, Gummadi KP. Fairness beyond disparate treatment & disparate impact: learning classification without disparate mistreatment. *arXiv preprint*. 2017:1171-1180. doi:10.48550/arXiv.1610.08452

188. Kamiran F, Karim A, Zhang X. Decision theory for discrimination-aware classification. Proceedings of the 2012 IEEE 12th International Conference on Data Mining; 2012:924-929. doi:10.1109/ICDM.2012.45

189. Pleiss G, Raghavan M, Wu F, Kleinberg J, Weinberger KQ. On fairness and calibration. Advances in Neural Information Processing Systems 30; 2017.

190. Linardatos P, Papastefanopoulos V, Kotsiantis S. Explainable AI: a review of machine learning interpretability methods. *Entropy (Basel)*. 2020;23(1):18. doi:10.3390/e23010018

191. Antoniadi AM, Du YH, Guendouz Y, et al. Current challenges and future opportunities for XAI in machine learning-based clinical decision support systems: a systematic review. *Appl Sci Basel*. 2021;11(11):23.5088. doi:10.3390/app11115088

192. Caruana R, Lou Y, Gehrke J, Koch P, Sturm M, Elhadad N. Intelligible models for healthcare: predicting pneumonia risk and hospital 30-day readmission. Presented at: Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. Sydney, NSW, Australia; 2015. https://doi:10.1145/2783258.2788613

193. Lundberg M, Lee S. A unified approach to interpreting model predictions. *arXiv preprint*. 2017. Doi:10.48550/ARXIV.1705.07874

194. Ribeiro M, Singh S, Guestrin C. "Why should I trust you?": explaining the predictions of any classifier. *arXiv preprint*. 2016. doi:10.48550/ARXIV.1602.04938

195. Oala L, Fehr J, Gilli L, et al. ML4H auditing: from paper to practice. *PMLR*. 2020;280-317.

196. Rajpurkar P, Chen E, Banerjee O, Topol EJ. AI in health and medicine. *Nat Med*. 2022;28(1):31-38. doi:10.1038/s41591-021-01614-0

197. Hughes JA, Hannah W, Kikkert P, et al. We are not Pontius Pilate: acknowledging ethics and policy. *IEEE* 2020;2975-2984. DOI:10.1109/SSCI47803.2020.9308312

198. Pierson E, Cutler DM, Leskovec J, Mullainathan S, Obermeyer Z. An algorithmic approach to reducing unexplained pain disparities in underserved populations. *Nat Med*. 2021;27(1):136-140. doi:10.1038/s41591-020-01192-7

199. Saleiro P, Kuester B, Hinkson L, et al. Aequitas: a bias and fairness audit toolkit. *arXiv preprint*. 2018. doi:10.48550/arXiv.1811.05577

200. Panigutti C, Perotti A, Panisson A, Bajardi P, Pedreschi D. FairLens: auditing black-box clinical decision support systems. *Inf Process Manage*. 2021;58(5):102657. doi:10.1016/j.ipm.2021.102657

201. Kasturi SN, Park J, Wild D, Khan B, Haggstrom DA, Grannis S. Predicting COVID-19-related health care resource utilization across a statewide patient population: model development study. *J Med Internet Res*. 2021;23(11):10.e31337. doi:10.2196/31337

202. Kelly CJ, Karthikesalingam A, Suleyman M, Corrado G, King D. Key challenges for delivering clinical impact with artificial intelligence. *BMC Med*. 2019;17(1):9.195. doi:10.1186/s12916-019-1426-2

203. Parikh RB, Teeple S, Navathe AS. Addressing bias in artificial intelligence in health care. *JAMA*. 2019;322(24):2377-2378. doi:10.1001/jama.2019.18058

204. Wahl B, Cossy-Gantner A, Germann S, Schwalbe NR. Artificial intelligence (AI) and global health: how can AI contribute to health in resource-poor settings? *BMJ Glob Health*. 2018;3(4):7.e000798. doi:10.1136/bmjgh-2018-000798

205. Adebayo JA. FairML: ToolBox for diagnosing bias in predictive modeling. Massachusetts Institute of Technology; 2016. Accessed February 9, 2023. https://dspace.mit.edu/handle/1721.1/108212

206. Veale M, Binns R. Fairer machine learning in the real world: mitigating discrimination without collecting sensitive data. *Big Data Soc*. 2017;4(2):2053951717743530. doi:10.1177/2053951717743530

207. Harrus I, Wyndham J. Artificial intelligence and COVID-19: applications and impact assessment. American Association for the Advancement of Science. May 2021. Accessed November 18, 2022. https://www.aaas.org/sites/default/files/2021-05/AIandCOVID19_2021_FINAL.pdf

208. Vollmer S, Mateen BA, Bohner G, et al. Machine learning and artificial intelligence research for patient benefit: 20 critical questions on transparency, replicability, ethics, and effectiveness. *BMJ*. 2020;368:l6927. doi:10.1136/bmj.l6927

209. Cruz BGS, Bossa MN, Solter J, Husch AD. Public Covid-19 X-ray datasets and their impact on model bias--a systematic review of a significant problem. *Med Image Anal*. 2021;74:16.102225. doi:10.1016/j.media.2021.102225

210. Park SH, Kim YH, Lee JY, Yoo S, Kim CJ. Ethical challenges regarding artificial intelligence in medicine from the perspective of scientific editing and peer review. *Sci Ed*. 2019;6(2):91-98. doi:10.6087/kcse.164

211. Hernandez-Boussard T, Bozkurt S, Ioannidis JPA, Shah NH. MINIMAR (MINimum Information for Medical Al Reporting): developing reporting standards for artificial intelligence in health care. *J Am Med Inf Assoc*. 2020;27(12):2011-2015. doi:10.1093/jamia/ocaa088

212. O'Neill J, Tabish H, Welch V, et al. Applying an equity lens to interventions: using PROGRESS ensures consideration of socially stratifying factors to illuminate inequities in health. *J Clin Epidemiol*. 2014;67(1):56-64. doi:10.1016/j.jclinepi.2013.08.005

213. Villasenor J. Products liability law as a way to address AI harms. Brookings Institution. October 31, 2019. Accessed March 11, 2022. https://www.brookings.edu/research/products-liability-law-as-a-way-to-address-ai-harms/

214. Chung J, Zink A. Hey Watson-can I sue you for malpractice--examining the liability of artificial intelligence in medicine. *Asia Pacific J Health L Ethics*. 2017;11:51.

215. Sullivan HR, Schweikart SJ. Are current tort liability doctrines adequate for addressing injury caused by AI? *AMA J Ethics*. 2022;21(2):160-166. doi:10.1001/amajethics.2019.160.

216. Miller K. When algorithmic fairness fixes fail: the case for keeping humans in the loop. Stanford University Human-Centered Artificial Intelligence. March 8, 2022. Accessed December 28, 2022. https://hai.stanford.edu/news/when-algorithmic-fairness-fixes-fail-case-keeping-humans-loop

217. Newman BK. Societal upheaval during the COVID-19 pandemic underscores need for new AI data regulations. *TechCrunch*. July 10, 2020. Accessed March 18, 2022. https://techcrunch.com/2020/07/10/societal-upheaval-during-the-covid-19-pandemic-underscores-need-for-new-ai-data-regulations

218. Humans will add to AI's limitations. *The Economist.* 2022. Accessed March 17, 2022. https://www.economist.com/technology-quarterly/2020/06/11/humans-will-add-to-ais-limitations

219. Nakul Aggarwal MA, Sanjay Basu, John J. Curtin, et al. Advancing artificial intelligence in health settings outside the hospital and clinic. *NAM Perspect*. November 30, 2020. doi:10.31478/202011f

220. Neter E, Brainin E. eHealth literacy: extending the digital divide to the realm of health information. *J Med Internet Res*. 2012;14(1):e1619. DOI:10.2196/jmir.1619

221. Azzopardi-Muscat N, Sørensen K. Towards an equitable digital public health era: promoting equity through a health literacy perspective. *Eur J Public Health*. 2019;29(suppl 3):13-17. doi:10.1093/eurpub/ckz166

222. Smythe S, Grotluschen A, Buddeberg K. The automated literacies of e-recruitment and online services. *Stud Educ Adults-NIACE*. 2021;53(1):4-22. doi:10.1080/02660830.2020.1855870

223. Program GSAC. Global South AI4COVID Program. Accessed March 11, 2022. https://covidsouth.ai/

224. Price WNI. Risks and remedies for artificial intelligence in health care. November 14, 2019. https://www.brookings.edu/research/risks-and-remedies-for-artificial-intelligence-in-health-care/

225. Bigman YE, Yam KC, Marciano D, Reynolds SJ, Gray K. Threat of racial and economic inequality increases preference for algorithm decision-making. *Comput Hum Behav*. 2021;122:11.106859. doi:10.1016/j.chb.2021.106859

226. Young AT, Amara D, Bhattacharya A, Wei ML. Patient and general public attitudes towards clinical artificial intelligence: a mixed methods systematic review. *Lancet Digit Health*. 2021;3(9):e599-e611. doi:10.1016/s2589-7500(21)00132-1

227. Avoundjian T, Dombrowski JC, Golden MR, et al. Comparing methods for record linkage for public health action: matching algorithm validation study. *JMIR Public Health Surveill*. 2020;6(2):e15917. doi:10.2196/15917

228. McCoy D, Mgbara W, Horvitz N, Getz WM, Hubbard A. Ensemble machine learning of factors influencing COVID-19 across US counties. *Sci Rep*. 2021;11(1):11777. doi:10.1038/s41598-021-90827-x

229. Kawlra G, Sakamoto K. Spatialising urban health vulnerability: an analysis of NYC's critical infrastructure during COVID-19. *Urban Stud*. 2021;21.00420980211044304. doi:10.1177/00420980211044304

230. Chakraborti S, Maiti A, Pramanik S, et al. Evaluating the plausible application of advanced machine learnings in exploring determinant factors of present pandemic: a case for continent specific COVID-19 analysis. *Sci Total Environ*. 2021;765:142723. doi:10.1016/j.scitotenv.2020.142723

231. Cao ZC, Tang F, Chen C, et al. Impact of systematic factors on the outbreak outcomes of the novel COVID-19 disease in China: factor analysis study. *J Med Internet Res*. 2020;22(11):10. e23853. doi:10.2196/23853

232. Marvel SW, House JS, Wheeler M, et al. The COVID-19 pandemic vulnerability index (PVI) dashboard: monitoring county-level vulnerability using visualization, statistical modeling, and machine learning. *Environ Health Perspect*. 2021;129(1):17701. doi:10.1289/ehp8690

233. Li M, Zhang Z, Cao W, et al. Identifying novel factors associated with COVID-19 transmission and fatality using the machine learning approach. *Sci Total Environ*. 2021;764:142810. doi:10.1016/j.scitotenv.2020.142810

234. Chen XX, Wang Z, Bromfield SG, DeVries A, Pryor D, Willey V. Identification and comparison of patient characteristics for those hospitalized with COVID-19 versus influenza using machine learning in a commercially insured US population. *Pragmat Obs Res*. 2021;12:9-13. doi:10.2147/por.S304220

235. Shao Y, Ahmed A, Liappis AP, Faselis C, Nelson SJ, Zeng-Treitler Q. Understanding demographic risk factors for adverse outcomes in COVID-19 patients: explanation of a deep learning model. *J Health Inform Res*. 2021;1-20. doi:10.1007/s41666-021-00093-9

236. Baqui P, Marra V, Alaa AM, Bica I, Ercole A, van der Schaar M. Comparing COVID-19 risk factors in Brazil using machine learning: the importance of socioeconomic, demographic and structural factors. *Sci Rep*. 2021;11(1):15591. doi:10.1038/s41598-021-95004-8

237. Jha IP, Awasthi R, Kumar A, Kumar V, Sethi T. Learning the mental health impact of COVID-19 in the United States with explainable artificial intelligence: observational study. *JMIR Ment Health*. 2021;8(4):e25097. doi:10.2196/25097

238. Zencity. How the City of Philadelphia Better Aligned Vaccine Hesitancy Outreach. Accessed March 12, 2022. https://zencity.io/case_studies/how-the-city-of-philadelphia-aligned-vaccine-hesitancy/

239. Nguyen QC, Huang YR, Kumar A, et al. Using 164 million Google Street View images to derive built environment predictors of COVID-19 cases. *Int J Environ Res Public Health*. 2020;17(17):13.6359. doi:10.3390/ijerph17176359

240. Ancochea J, Izquierdo JL, Soriano JB. Evidence of gender differences in the diagnosis and management of Coronavirus disease 2019 patients: an analysis of electronic health records using natural language processing and machine learning. *J Womens Health (Larchmt)*. 2021;30(3):393-404. doi:10.1089/jwh.2020.8721

241. Magua W, Zhu X, Bhattacharya A, et al. Are female applicants disadvantaged in National Institutes of Health peer review? Combining algorithmic text mining and qualitative methods to detect evaluative differences in R01 reviewers' critiques. *J Womens Health (Larchmt)*. 2017;26(5):560-570. doi:10.1089/jwh.2016.6021

242. Radhakrishna R. California's equity tools for COVID-19: the healthy places index and beyond. AcademyHealth. August 3, 2021. Accessed March 28, 2022. https://academyhealth.org/sites/default/files/disadvantage_indices_webinar_slides_combined_003.pdf

243. Erickson J, Abbott K, Susienka L. Automatic address validation and health record review to identify homeless Social Security disability applicants. *J Biomed Inform*. 2018;82:41-46. doi:10.1016/j.jbi.2018.04.012

244. Brown HK, Carty A, Havercamp SM, Parish S, Lunsky Y. Identifying reproductive-aged women with physical and sensory disabilities in administrative health data: a systematic review. *Disabil Health J*. 2020;13(3):9.100909. doi:10.1016/j.dhjo.2020.100909

245. Potash E, Ghani R, Walsh J, et al. Validation of a machine learning model to predict childhood lead poisoning. *JAMA Netw Open*. 2020;3(9):e2012734. doi:10.1001/jamanetworkopen.2020.12734

246. Pan I, Nolan LB, Brown RR, et al. Machine learning for social services: a study of prenatal case management in Illinois. *Am J Public Health*. 2017;107(6):938-944. doi:10.2105/ajph.2017.303711

247. UnitedHealthcare Group. UnitedHealthcare introduces the use of predictive analytics to expand its capabilities to address social determinants of health. July 8, 2021. Accessed March 14, 2022. https://www.unitedhealthgroup.com/newsroom/2021/2021-7-8-uhc-predictive-analytics-social-determinants-health.html

248. Chonde DB, Pourvaziri A, Williams J, et al. RadTranslate: an artificial intelligence-powered intervention for urgent imaging to enhance care equity for patients with limited English proficiency during the COVID-19 pandemic. *J Am Coll Radiol*. 2021;18(7):1000-1008. doi:10.1016/j.jacr.2021.01.013

249. Matheny, M, Thadaney Israni S, Ahmed M, and Whicher D, Editors. 2022. *Artificial Intelligence in Health Care: The Hope, the Hype, the Promise, the Peril*. Washington, DC: National Academy of Medicine.

250. Naseem M, Akhund R, Arshad H, Ibrahim MT. Exploring the potential of artificial intelligence and machine learning to combat COVID-19 and existing opportunities for LMIC: a scoping review. *J Prim Care Commun Health*. 2020;11:2150132720963634. doi:10.1177/2150132720963634

251. Balbin JR, Yap AIT, Calicdan BD, Bernabe LAM. Arrhythmia detection using electrocardiogram and phonocardiogram pattern using integrated signal processing algorithms with the aid of convolutional neural networks. *2021 IEEE International Conference on Automatic Control & Intelligent Systems (I2CACIS)*, Shah Alam, Malaysia, 2021, pp. 146-151, doi:10.1109/I2CACIS52118.2021.9495913

252. Conceicao T, Braga C, Rosado L, Vasconcelos MJM. A review of computational methods for cervical cells segmentation and abnormality classification. *Int J Mol Sci*. 2019;20(20):38.5114. doi:10.3390/ijms20205114

253. Sajal MSR, Ehsan MT, Vaidyanathan R, Wang S, Aziz T, Mamun KAA. Telemonitoring Parkinson's disease using machine learning by combining tremor and voice analysis. *Brain Inform*. 2020;7(1):12. doi:10.1186/s40708-020-00113-1

254. Desiere S, Struyven L. Using artificial intelligence to classify jobseekers: the accuracy-equity trade-off. *J Soc Policy*. 2021;50(2):367-385. doi:10.1017/S0047279420000203

255. Dutta S, Wei D, Yueksel H, Chen P-Y, Liu S, Varshney K. Is there a trade-off between fairness and accuracy? A perspective using mismatched hypothesis testing. *PMLR.* 2020:2803-2813.

256. Cooper AF, Abrams E, Na N. Emergent unfairness in algorithmic fairness-accuracy trade-off research. *Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society*. 2021. Doi:10.1145/3461702.3462519

257. Liu S, Vicente LN. Accuracy and fairness trade-offs in machine learning: a stochastic multi-objective approach. *Comput Manag Sci*. 2022;1-25. doi:10.1007/s10287-022-00425-z

258. World Health Organization. Ethics and governance of artificial intelligence for health: WHO guidance. Accessed September 26, 2022. https://www.who.int/publications/i/item/9789240029200

259. Rose S, Bergquist SL, Layton TJ. Computational health economics for identification of unprofitable health care enrollees. *Biostatistics*. 2017;18(4):682-694. doi:10.1093/biostatistics/kxx012

260. BBC News. IBM abandons ‘biased’ facial recognition tech. June 9, 2020. Accessed November 11, 2022. https://www.bbc.com/news/technology-52978191

261. Johnson S. A.I. is mastering language. Should we trust what it says? *New York Times Magazine.* 2022. Accessed December 27, 2022. https://www.nytimes.com/2022/04/15/magazine/ai-language.html

262. Morey BN, Chang RC, Thomas KB, et al. No equity without data equity: data reporting gaps for Native Hawaiians and Pacific Islanders as structural racism. *J Health Polit Policy Law*. 2022;47(2):159-200. doi:10.1215/03616878-9517177

263. Grote T, Berens P. On the ethics of algorithmic decision-making in healthcare. *J Med Ethics*. 2020;46(3):205-211. doi:10.1136/medethics-2019-105586

264. Seastedt KP, Moukheiber D, Mahindre SA, et al. A scoping review of artificial intelligence applications in thoracic surgery. *Eur J Cardiothorac Surg*. 2021;61(2):239-248. doi:10.1093/ejcts/ezab422

265. Wilkinson K, Sheets L, Fitch D, Popejoy L. Systematic review of approaches to use of neighborhood-level risk factors with clinical data to predict clinical risk and recommend interventions. *J Biomed Inform*. 2021;116:103713. doi:10.1016/j.jbi.2021.103713

266. Simonite T. When it comes to health care, AI has a long way to go. *Wired*. January 16, 2022. https://www.wired.com/story/health-care-ai-long-way-to-go/

267. Bozkurt S, Cahan EM, Seneviratne MG, et al. Reporting of demographic data and representativeness in machine learning models using electronic health records. *J Am Med Inf Assoc*. 2020;27(12):1878-1884. doi:10.1093/jamia/ocaa164

268. Goreke V, Sari V, Kockanat S. A novel classifier architecture based on deep neural network for COVID-19 detection using laboratory findings. *Appl Soft Comput*. 2021;106:8.107329. doi:10.1016/j.asoc.2021.107329

269. NYC Health+Hospitals. Health department launches coalition to confront racism in medical algorithms. November 24, 2021. Accessed November 11, 2022. https://www.nychealthandhospitals.org/pressrelease/health-department-launches-coalition-to-confront-racism-in-medical-algorithms/

270. Zarsky T. The trouble with algorithmic decisions: an analytic road map to examine efficiency and fairness in automated and opaque decision making. *Sci Techn Human Values*. 2016;41(1):118-132. doi:10.1177/01622439156055

271. Arrieta AB, Díaz-Rodríguez N, Del Ser J, et al. Explainable artificial intelligence (XAI): concepts, taxonomies, opportunities and challenges toward responsible AI. *Inf Fusion*. 2020;58:82-115. doi:10.1016/j.inffus.2019.12.012

272. Bowen S, Khoury MJ, Moonesinghe R. "Precision" health tools and…increased health disparities? Centers for Disease Control and Prevention, Office of Genomics and Precision Health. Accessed March 8, 2022. https://blogs.cdc.gov/genomics/2020/01/08/precision-health-tools/

273. Tricco AC, Lillie E, Zarin W, et al. PRISMA extension for scoping reviews (PRISMA-ScR): Checklist and Explanation. *Ann Intern Med*. 2018;169(7):467-473. doi:10.7326/m18-0850