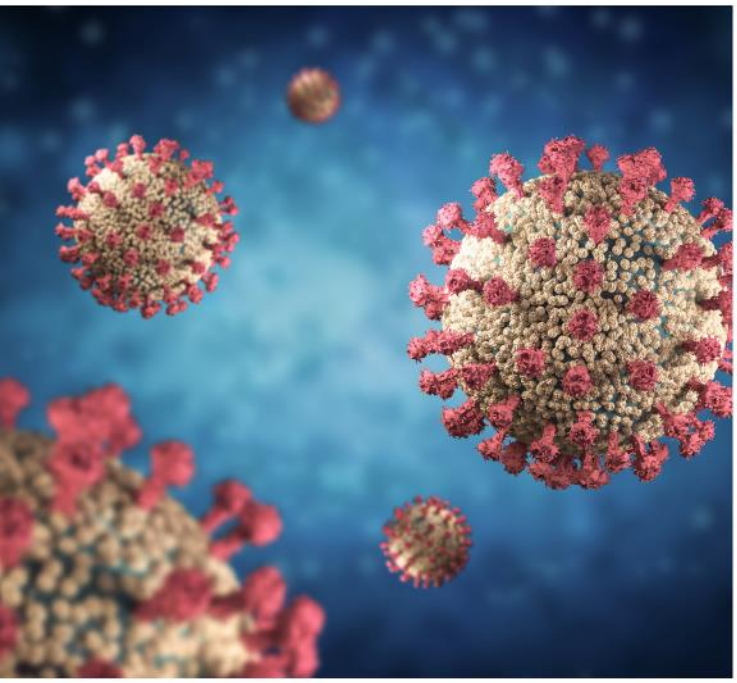
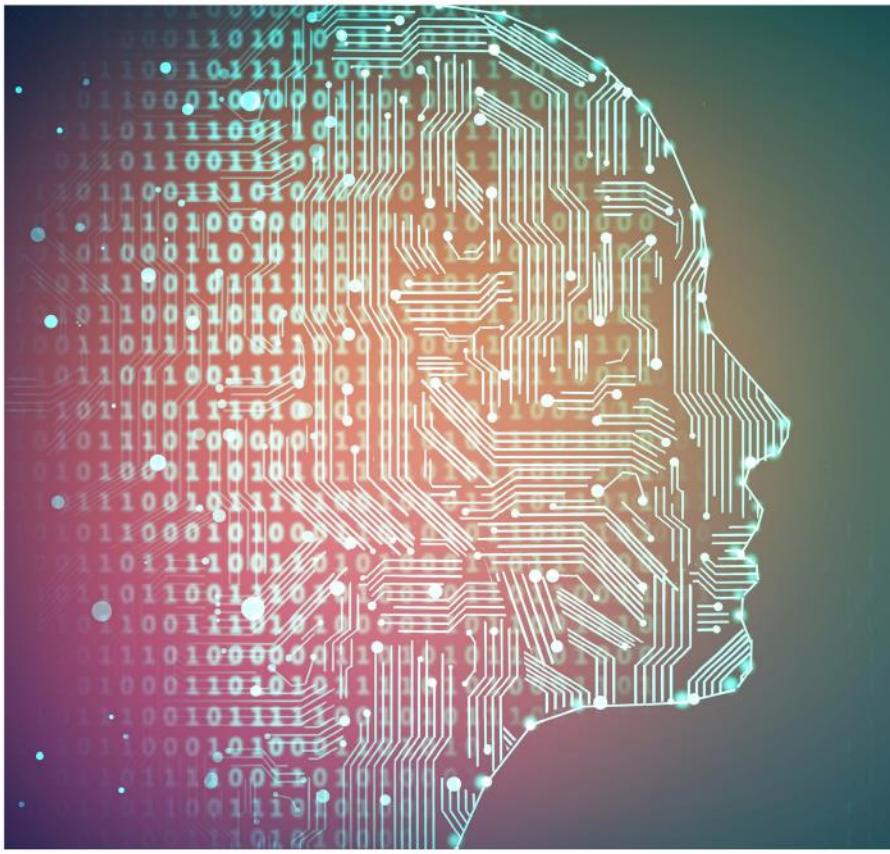
AAAS

Artificial Intelligence and COVID-19:

APPLICATIONS AND IMPACT ASSESSMENT

MAY 2021





AMERICAN ASSOCIATION FOR THE ADVANCEMENT OF SCIENCE

Page left blank intentionally

This report was produced as a part of the Artificial Intelligence: Applications/Implications (AI)² Initiative (https://www.aaas.org/ai2) of the American Association for the Advancement of Science (AAAS). AAAS is the world's largest multidisciplinary scientific membership organization. The goal of the AAAS (AI)² Initiative is to contribute to the responsible development and application of artificial intelligence (AI) such that AI alleviates rather than exacerbates social inequalities.

Report prepared by:
Ilana Harrus, Senior Program Associate (AAAS)
Jessica Wyndham, Program Director (AAAS)

Acknowledgments:
We gratefully acknowledge members of the (AI)² Initiative Advisory Group and of the AAAS Committee on Scientific Freedom and Responsibility who provided input and guidance. In particular, we wish to thank Ronald Arkin, Solon Barocas, Kadija Ferryman, Noreen Herzfeld, Camara Jones, Alexa Koenig, Daniel Kracov, Mark Latonero, Camille Nebeker, Frank Pasquale, David Robinson, Neal Rubin, Dietram Scheufele, Jeramie Scott, and Daniel Weitzner. Special thanks are also due to Theresa Harris and Curtis L. Baxter III for their feedback on the report.

Funding for the AI² Initiative was provided in part by Microsoft.

Disclaimer:
The opinions, findings, and conclusions or recommendations expressed in this publication do not necessarily reflect the views of the AAAS Board of Directors, its council and membership, or Microsoft.

Contact:
AAAS welcomes comments and questions regarding its work. Please send information, suggestions, and any comments to the AI2 project at: https://www.aaas.org/ai2/contact

Cite as:
AAAS (2021). Artificial Intelligence and COVID-19: Applications and Impact Assessment. (Report prepared by Ilana Harrus and Jessica Wyndham under the auspices of the AAAS Scientific Responsibility, Human Rights and Law Program).

American Association for the Advancement of Science
Artificial Intelligence: Applications/Implications
Scientific Responsibility, Human Rights and Law Program
1200 New York Avenue, NW
Washington, DC, 20005 USA

Cover design credit: Paula Fry, Lead Art Associate (AAAS)

Table of Contents

Table of Figures

Acronyms

AAAS: American Association for the Advancement of Science
AI: Artificial Intelligence
BLE: Bluetooth Low Energy
CDC: Center for Disease Control and Prevention
CORD-19: COVID-19 Open Research Dataset
COVID: Coronavirus Disease
CT: Computed Tomography
EDI: EPIC Deterioration Index
EHR: Electronic Health Record
EPIC: Emergency Physicians Integrated Care
ER: Emergency Room
FDA: Food and Drug Administration
FG-AI4H: Focus Group-Artificial Intelligence for Health
FDR: False Discovery Rate
FN: False Negative
FNR: False Negative Rate
FP: False Positive
FPR: False Positive Rate
GPS: Global Positioning System
ICCPR: International Covenant on Civil and Political Rights
ICESCR: International Covenant on Economic, Social and Cultural Rights
ICU: Intensive Care Unit
ITU: International Telecommunication Union
ML: Machine Learning
NIST: National Institute of Standards and Technology
PACT East: Private Automated Contact Tracing
PACT West: Privacy Sensitive Protocols And Mechanisms for Mobile Contact Tracing
QR: Quick Response
SOFA: Sequential Organ Failure Assessment
SSRC: Social Science Research Council
UNESCO: United Nations Educational, Scientific and Cultural Organization
US: United States
WHO: World Health Organization

Abstract

This report provides an overview of the multiple ways that artificial intelligence has been either used or suggested as a tool in the context of the COVID-19 pandemic. The report identifies and analyzes two specific classes of applications of AI that have already been implemented and which present potentially stark implications for marginalized and vulnerable populations (sorting and surveillance.) For each application, their technical attributes are detailed, and the ethical, human rights and differential impacts of these applications explored, drawing upon interviews conducted with experts in the field as well as a wide body of literature. Based on this analysis, a research agenda is proposed to support the fair, just, transparent, and accountable use of those AI applications in the future.

The findings are intended to help raise awareness among populations impacted by the application of AI-based technologies in the current pandemic and in the future, to inform policy makers and health professionals who determine whether and how these tools are deployed during the current pandemic, during future public health crises, and in general, and to broaden the scope of understanding among developers and researchers in the AI community of the implications of the use of these AI-based technologies.

Keywords: Artificial Intelligence, AI, COVID-19, Applications, Surveillance, Resource Allocations, Triage, Ethics, Human Rights.

Executive Summary

As the COVID-19 pandemic struck, Artificial Intelligence (AI) quickly became central to the fight against the virus. From diagnosis to drug development, from forecasting the disease's spread to monitoring and surveillance of the population, the tools of AI were called upon to address the scale and scope of the pandemic. This report catalogues the many and varied uses of AI in the pandemic, outlines the relevant ethical and human rights frameworks, identifies those applications that raise particular ethical and human rights concerns, and explores in detail those concerns as they relate to two general broad categories of application: those for sorting/triage, and those for surveillance. The goal of the report is to inform the future development, use, and policies associated with these applications, whether in the current or future health crisis.

Section One is a survey of the different AI applications used or suggested in the fight against COVID-19 and a classification of those applications according to their field of use. The literature reveals no consensus taxonomy of these applications. The taxonomy adopted for this report classifies AI applications into five large classes: forecasting, diagnosis, containment and monitoring, drug development and treatments, and social and medical management. This classification emphasizes applications implemented at scale in the United States (US).

Section Two surveys the existing ethical frameworks relevant to medical AI-based applications and the overarching human rights principles relevant to the pandemic. Although sometimes expressed differently, four guiding principles form the basis of the ethics frameworks: autonomy (informed consent), beneficence, nonmaleficence, and justice (fairness). Related principles such as privacy and confidentiality are prominent in both the ethics and human rights frameworks. Other relevant human rights of particular relevance are the rights to equality and non-discrimination, to liberty and security of the person, to information, to freedom of movement, to freedom from being subjected to "medical or scientific experimentation" and, finally, the right to enjoy the benefits of scientific progress and its applications. The human rights framework is particularly helpful in conceptualizing those measures that can or should be taken during a public health crisis of the sort posed by the COVID-19 pandemic, the conditions that need to be met for such measures to be implemented, and the framework for determining when such measures should come to an end.

Section Three focuses on those AI applications that raise significant ethical and human rights concerns, especially those with large and disproportionate impacts on marginalized and vulnerable populations: sorting tools, including medical triage applications used in hospitals and all AI-based algorithms used to allocate resources, and surveillance applications that include contact tracing applications and geofencing. These two classes of applications form the basis of a deeper investigation into their technical characteristics and the ethical, legal/regulatory, and societal issues raised by their implementation in practice.

The ethical issues raised by triage applications fall within two broad categories: algorithmic and data issues, the former encompassing concerns about implementation, fairness and bias,

explainability and transparency, and the latter related to data gathering, validation, training, and sharing in the context of both accuracy and fairness of the results. The primary issues raised by surveillance applications are privacy and data governance, with an added concern about their potential discriminatory impact on marginalized populations. Algorithmic issues are also present, especially in the context of contact tracing.

This assessment demonstrates a vital need for technical validation of all applications (both medical triage and contact tracing) to ensure they are medically sound before implementation.

The report also identifies several other gaps in the general understanding of the implications of these applications' widespread use.

The technical and knowledge gaps form the basis for the research agenda laid out at the end of this report. The research agenda identifies those technical knowledge gaps that require further exploration by developers, support by funders, and understanding by users. The research agenda also identifies algorithmic deficiencies that require both a technical fix as well as understanding by policy makers, as well as society more broadly.

Three features of this research agenda are essential to recognize:

- It is not exhaustive. It highlights currently pressing areas for future exploration in a rapidly changing landscape.
- Though categorizing the research gaps as technical, data, and ethical, social and human rights related, it is not always possible to separate the ethical and human rights components of the technical and data gaps.
- While the research agenda focuses on the applications used in the context of COVID-19 that are the focus of this report, the research needs identified would help fill gaps of relevance across a broader range of AI-based applications, generally, and that will persist long after the health crisis driven by COVID-19 is over.

Research Agenda

A) Technical Validation

Contact Tracing Applications:

Before implementation, the following variables need to be measured and made public.

1) Actual False Positive/False Negative rates for identifying people in contact with infected individuals and those who need to quarantine as a consequence of that contact.
2) The calibration of the correlation between the Bluetooth signal strength and the distance between two telephones. In addition, there is a need to quantify:
    - The impact of the physical position of the phones on the subject on the quality of the measurements (Does a phone in a bag provide a measurement as accurate as a phone in a back pocket or inside a coat?).
    - The measure of the maximum distance still classified as a “close contact” by the software and its comparison with the medically relevant distance

associated with a risk of infection. This comparison will determine if the software/application is medically beneficial.

3) Time calibration to estimate the "time in contact" of the phones.

Medical Triage Applications:

Validate medical triage applications by addressing the question:

4) How strong is the evidence that the algorithm used to conduct triage is indeed triaging the patients that are the most affected by the disease?

B) Software Oversight/Standards

"Testing requirements, test cases, and test tools" for both classes of applications (sorting and surveillance) should be informed by the "meaningful use" requirements applied for EHRs and developed by National Institute of Standards and Technology (NIST) in collaboration with the Office of the National Coordinator (ONC).

Sorting Applications: Medical Triage

The creation and implementation of independent software auditors should address the technical validity of the algorithms used for medical triage, and provide an ethical assessment of the tools proposed, to ensure that this technical "certification" is rooted in existing ethical frameworks.

5) What are the anticipated post-pandemic effects on how each unit in the health organization using triage applications interact?
6) What are the post-pandemic effects on the delivery of medical care?
7) Will the system change the dynamic of work within the health care provider?

C) Optimal Time of Operation and Deployment

We recommend running Monte-Carlo-type simulations to model the optimum time of deployment contact tracing applications.

Contact Tracing Applications:

These simulations could include propagation models of the disease coupled with models of people's patterns of travel. Once the simulations show that contact tracing will not be efficient (if too many people are already infected, contact tracing becomes irrelevant to contain the propagation of the disease), consider retiring the applications.

8) When is the implementation of the application optimal to successfully track the majority of cases?
9) How much more efficient in tracking cases are AI-based contact tracing applications compared to traditional contact tracing methods?[1]

[1] The measure of the efficacy of traditional contact tracing is given at: https://www.cdc.gov/coronavirus/2019-ncov/php/contact-tracing/contact-tracing-plan/evaluating-success.html

D) Data Validation

Medical Triage Applications:

10) *Validation*: How relevant are the populations on which any future similar program will be trained to the populations on which it is going to be used? Similarity should not be restricted to race and ethnicity but include age, sex, socio-economic status level, education level, location of residence, and disability status.

E) Bias and Fairness Validation

Medical Triage or Allocation of Resources:

11) Did any differences in the populations to which the algorithms were applied and the training data led to inaccurate results?
12) *Audit:* Could the algorithms' variables or decision points hide or mask issues that trigger unforeseen bias?
13) How are these variables or decision points derived?
14) What would be the consequences on the outcome to choose different variables or decision points in the algorithm?

Surveillance Applications:

15) *Validation*: How robust is the assumption that a cell phone belongs to only one person in the US?
16) How does this assumption vary within different communities in the US and how does this compare to what is measured in the developing world?

Both Sorting and Surveillance Applications:

17) *Forensics:* Examine both applications and determine **all** the underlying assumptions present in the development and implementation.
18) *Completeness:* Is there a systematic or mathematical way to derive or at least identify the maximum number of assumptions for a given application?
19) *Completeness:* Are all the variables used by the algorithm necessary for the application? Inversely, is the algorithm using the correct variables?

F) Data Sharing

Both Types of Applications:

There is a need to identifying the obstacles to broad data sharing—in the context of surveillance, the reluctance to download and use the contact tracing application and in triage applications, the lack of shared medical data to efficiently train the algorithms.

20) Is lack of trust the main issue for lack of data sharing across different communities?
21) Does the lack of data sharing result in an algorithm that disparately impacts different communities?
22) What are the reasons for the existing disparities between the large participation of some groups in health research studies but their reluctance to consent to sharing data?
23) What is effective in increasing the understanding among diverse communities about the value of sharing data in specific contexts and circumstances, and their trust in doing so?

G) Compliance

Contact Tracing Applications:

24) How does compliance with health directives change when instructions are coming from an application rather than a real person?
25) How does the degree of compliance change based on the level of difficulty or inconvenience in following the application's directives (for example, measuring the differences of compliance between an order of social distancing and an order to quarantine)?
26) How and why do compliance attitudes vary across different populations?

H) Post-Pandemic

Contact Tracing Applications:

27) What, if any, health data that have been collected for the purposes of addressing the COVID-19 pandemic can and should be retained for the purposes of addressing future public health crises?

Medical Triage Applications:

28) What value would be served by continuing to use the AI-based medical triage tools developed for the purposes of the COVID-19 pandemic, and does that value out-weigh any potential ethical or human rights concerns?

Both Classes of Applications:

29) Did the implementation of the applications serve the purpose for which they were implemented? What impact, positive or negative, did their use have on the established goal? For example, did their use result in more efficient triage, fairer allocation of resources, larger-scale contact tracing or more accurate geofencing?
30) What measurable parameters should trigger the applications' deployment in the future, and what should trigger their rescindment?

Next Steps

This report, together with a commissioned study on the attitudes of marginalized populations toward AI as applied in the context of health and the COVID-19 pandemic will inform the elaboration of a responsibility framework that will provide a roadmap for developing and implementing just and ethical AI-based medical applications. This roadmap will be conceptualized and articulated by an invited cohort of thought leaders in a wide variety of fields ranging from ethicists to computer specialists and from human rights activists to lawyers and public servants. We anticipate that the result will help both AI practitioners and lawmakers and policy makers to usher a new era for AI-based medical applications.

Introduction

The words “Artificial Intelligence” (AI) evoke super-human learning abilities and machines endowed with the power to analyze astronomically large datasets for purposes benign or beneficial, from making movie recommendations to controlling traffic. At the same time, AI also evokes more nefarious and consequential applications of its power: mass surveillance (Qiang, 2019; Daly, 2019); credit approval (Baesens et al. 2003); bail determination (Angwin et al. 2016; Flores et al., 2016); and selection for enhanced medical follow-up (Obermeyer, 2019). Technologies that both benefit and harm are nothing new. One could argue that they have been with us since the discovery of fire. AI technology, however, feels different to most people because it is different: AI is the first technology that is becoming an integral part of our personal and public decision-making processes. What is more, in some cases, the software itself is the decision-maker.

Nowhere has this potential been so visible than in the applications of AI to the global health crisis that shook the world in 2020 and continues into 2021. From the moment the knowledge of the novel coronavirus became public and, in the months following its emergence from China as a global pandemic, the number of applications claiming or proposing to use AI to combat the virus exploded. The COVID-19 pandemic provided an opportunity for "redemption" for AI practitioners: here was a unique opportunity to prove that AI could be harnessed for the benefit of all humanity and AI developers seized the moment. AI became central to the fight against COVID-19, but the existing issues raised by the use of the technology persist, as the technology, like the pandemic itself, highlights and threatens to exacerbate existing social inequalities. The American Public Media Research Lab (2020) has a website entirely devoted to documenting the impacts the virus has had on minority communities in the United States (US) (see Figure 1).

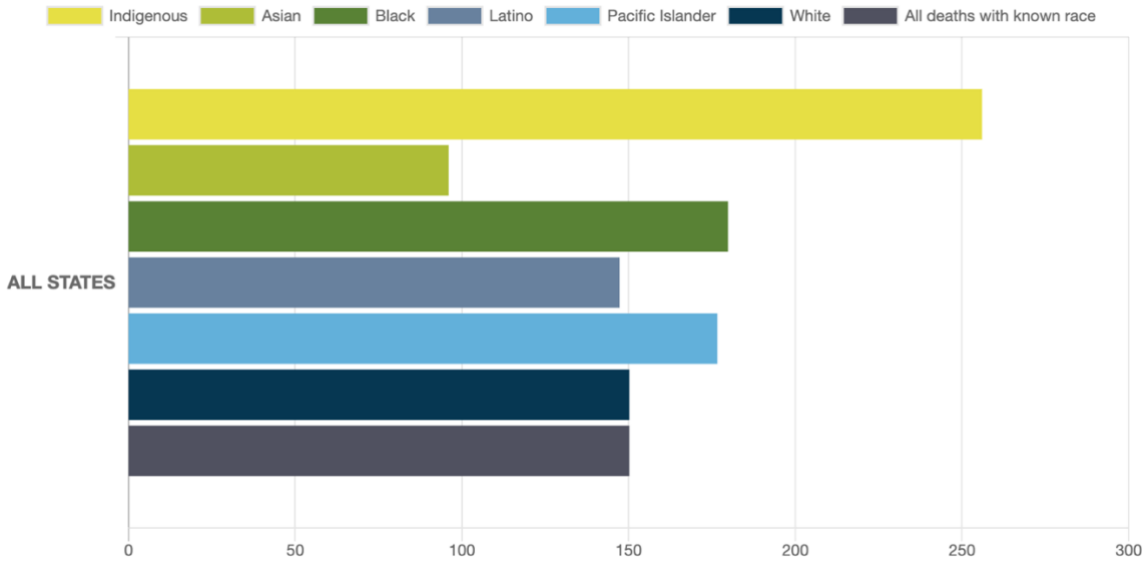


Figure 1: COVID-19 mortality rates per 100,000 cases by race and ethnicity in the US through March 2, 2021.

The pandemic has exacerbated existing issues impacting poor and disenfranchised communities for decades: lack or paucity of access to healthcare, resource inequalities, and disparities in the quality of care received.

One way to consider the disparities in experience across demographic groups is to differentiate between differences arising because of individual characteristics, for example, gender, race, and age, and those caused by broader societal and structural factors, for example, whether someone lives in a multigenerational household, or is employed in an industry at greater health or socioeconomic risk during the pandemic, for example, cashiers, cleaning crews, delivery services, restaurant servers, and trade workers. These workers (and their families) are exposed to the virus in higher proportions. The impacts of these societal factors are difficult to estimate and even harder to correct.

This report was born of the value in taking stock of the ways AI has or could have been deployed in the fight against COVID-19 and to identify the applications that raise the most serious ethical and human rights concerns so that we can apply the lessons learned in the current context in the future. The following section outlines the methodology used to identify and classify the different applications of AI being deployed in the context of the pandemic. Section 2 explores which categories of applications raise particular concerns about differential impacts on underserved populations. Section 3 contains a technical description of each of these applications, the ethical and human rights issues they raise, and their differential impacts on specific populations. Section 4 is the conclusion of this study and Section 5 proposes a path forward tor researchers and developers, policy makers, and funders.

1. Compilations and Assessment of AI Applications Linked to COVID-19

1.1 Methodology

Several search databases, including Google Scholar (scholar.google.com), ArXiv (arXiv.org), MedRxiv and BioRxiv (MedRxiv.org), CORD-19 (COVID-19 Open Research Dataset) (Etzioni, 2020), and COVIDScholar (https://www.covidscholar.org/) were used to generate a list as complete as possible of the different ways AI could be used in the fight against COVID-19. The databases were searched for articles combining "Artificial Intelligence", "AI", "COVID-19", "SARS-CoV-2", "Applications", "ML", "Machine Learning", "Deep Learning", and/or "DL."

The results of the “Recommendations” field in ReadCube Papers (papersapp.com) were also included together with information drawn from public conferences organized on AI and COVID-19 as the pandemic progressed (see conferences references in the text below).

1.2 Survey of Surveys

A review of existing articles that are themselves compilations of AI-based applications used in the context of COVID-19 did not reveal any broad agreement on the classification of applications used in the context of the pandemic though it did provide insights into the breadth of applications deployed and enormity of the literature generated about the use of AI in the pandemic.

Several reviews reveal specific uses of AI in the context of COVID-19, including for medical diagnosis (Nguyen 2020; Jamshidi et al. 2020). Chen, Li, et al. (2020) analyzed 1273 publications related to COVID-19, identifying 267 papers linked to AI. Their comparison of the numbers of papers published on a subject with a link or no link to AI provides a first glimpse of the potential of AI (see Figure 3 from Chen, Li, et al. 2020). As shown in Figure 3, one of the most promising areas of use for AI in the fight against COVID-19 in the domain of diagnosis is for image analysis and interpretation labeled “image inspection” in the figure below.

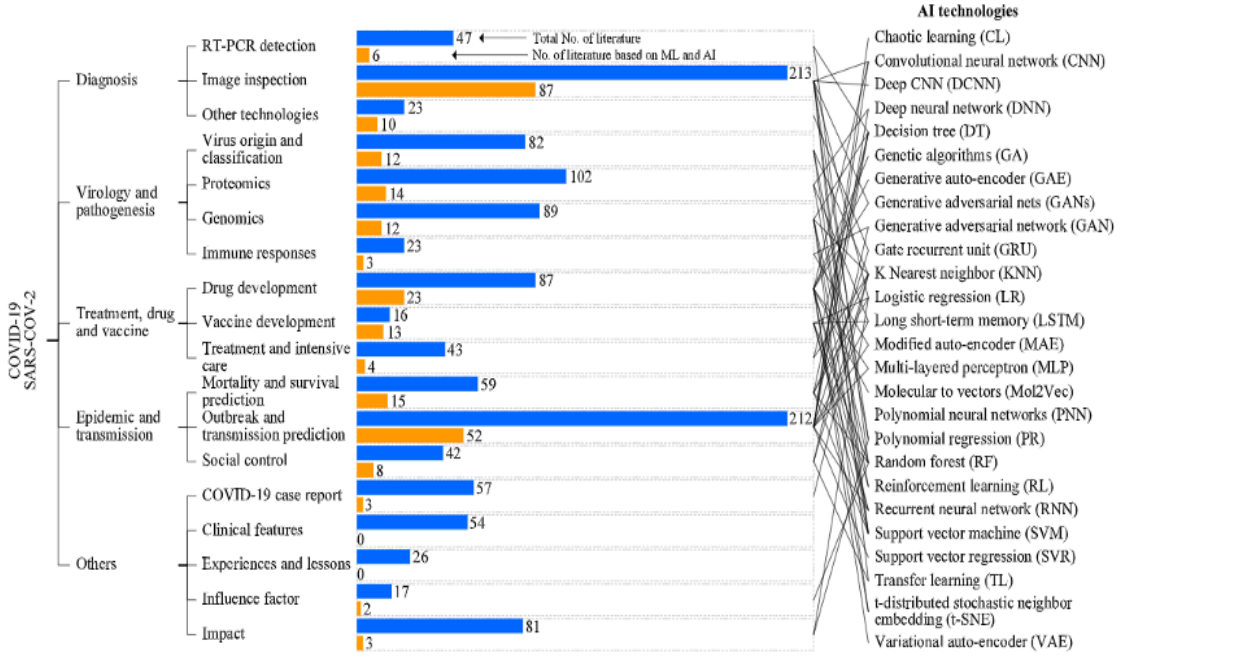


Figure 2: Papers and AI techniques used against both virus and disease. The bars show the number of papers on research topics that do not rely on the use of AI (blue) and those that do (orange)

Drilling one layer deeper, some survey papers examine the use and results of AI applied to one type of image analysis used in diagnosis (see for example, radiology images as described in Table 1 taken from Nguyen 2020).

Papers	Data	AI Methods	Results
[8]	4,356 chest CT exams from 3,322 patients from 6 medical centers: 1,296 exams for COVID-19, 1,735 for CAP and 1,325 for non-pneumonia	A 3D convolutional ResNet-50 [6], namely COVNet	AUC for detecting COVID-19 is of 0.96
[9]	618 CT samples: 219 from 110 COVID-19 patients, 224 CT samples from 224 patients with influenza-A viral pneumonia, and 175 CT samples from healthy people	Location-attention network and ResNet-18 [6]	Accuracy of 86.7%
[12]	5,941 Posterior-anterior chest radiography images across 4 classes (normal: 1,583, bacterial pneumonia: 2,786, non-COVID-19 viral pneumonia: 1,504, and COVID-19: 68)	Drop-weights based Bayesian CNNs	Accuracy of 89.92%
[13]	1,065 CT images (325 COVID-19 and 740 viral pneumonia)	Modified inception transfer-learning model	Accuracy of 79.3% with specificity of 0.83 and sensitivity of 0.67
[14]	Clinical data and a series of chest CT data collected at different times on 133 patients of which 54 patients progressed to severe/critical periods whilst the rest did not	Multilayer perceptron and LSTM [45]	AUC of 0.954
[15]	970 CT volumes of 496 patients with confirmed COVID-19 and 1,385 negative cases	2D deep CNN	Accuracy of 94.98% and AUC of 97.91%
[16]	CT images of 1,136 training cases (723 positives for COVID-19) from 5 hospitals	A combination of 3D UNet++ [17] and ResNet-50 [6]	Sensitivity of 0.974 and specificity of 0.922
[18]	Chest X-ray images of 50 normal and 50 COVID-19 patients	Pre-trained ResNet-50	Accuracy of 98%
[19]	16,756 chest radiography images across 13,645 patient cases from two open access data repositories	A deep CNN, namely COVID-Net	Accuracy of 92.4%
[20]	CT images obtained from 157 international patients (China and U.S.)	ResNet-50	AUC of 0.996
[21]	1,341 normal, 1,345 viral pneumonia, and 190 COVID-19 chest X-ray images	AlexNet [3], ResNet-18 [6], DenseNet-201 [23], SqueezeNet [25]	Accuracy of 98.3%
[22]	170 X-ray images and 361 CT images of COVID-19 from 5 different sources	A new CNN and pre-trained AlexNet [3] with transfer learning	Accuracy of 98% on X-ray images and 94.1% on CT images

Table 1: Summary of deep learning methods for COVID-19 diagnosis using radiology images (Nguyen, 2020).

Another area of specific focus in existing compilations has been containment (Alabool et al. 2020), including, for example, predicting outbreak locations and intensity. A different type of survey concentrates on the technology aspect of each application. For example, Chamola et al. (2020) and Nguyen, Ding, et al. (2020) look at the use of drones for population monitoring and social distancing enforcement. They both also mention use of blockchain, a relatively rare occurrence (in the more than 70,000 papers related to COVID-19 found online, only 0.1% mention blockchain). A detailed survey of the different applications of robots for COVID-19 response can be found in Murphy et al. (2020). The authors reviewed more than 260 reports mentioning the use of robots and published during a four-month period at the early stage of the pandemic. They found more than 200 instances of robots used for COVID-19 and classify them in six categories, including public safety (drone-enforced quarantine enforcement), clinical care (telehealth visits), continuity of work and education (sanitation by ground robots), quality of life (food delivery), laboratory and supply chain automation, and non-hospital care (off-site testing or delivery to quarantined populations). They further refined their analysis by differentiating between ground and aerial robots.

These surveys of applications of AI in the context of the pandemic are instructive in revealing the depth of uses in specific contexts. They are too narrow, however, for the purposes of identifying the breadth of applications and implications for the use of AI in the context of the pandemic. For that purpose, a strong starting point is the GitHub repository called "AI Technology to fight against Coronavirus (COVID-19)" (Jeon, 2020). Jeon developed a powerful visualization of AI applications for COVID-19 (see Figure 2), separating the applications into four categories: 1) forecasting and prevention, 2) emergency operations and response, 3) prevention of infection spread, and 4) treatment and drug research.

Most papers follow a similar categorization as the one found in Jeon (2020). Naudé (2020), for example, one of the earliest survey papers, divided the AI applications for COVID-19 into four categories: 1) tracking and prediction, 2) diagnosis and prognosis, 3) treatments and vaccines and 4) social control. This categorization is different but overlapping with Lalmuanawma et al.

(2020) who considered roughly the same four different aspects of the use of AI in the context of COVID-19 but focused on narrower aspects of the applications: 1) screening and treatment, 2) contact tracing, 3) prediction and forecasting, and 4) drugs and vaccination. These categories, however, are restrictive. Contact tracing, for example, is a good illustration of containing and monitoring the disease but it is not the only one. As explained in the second part of this report, we needed to expand that category to cover other surveillance-type applications.

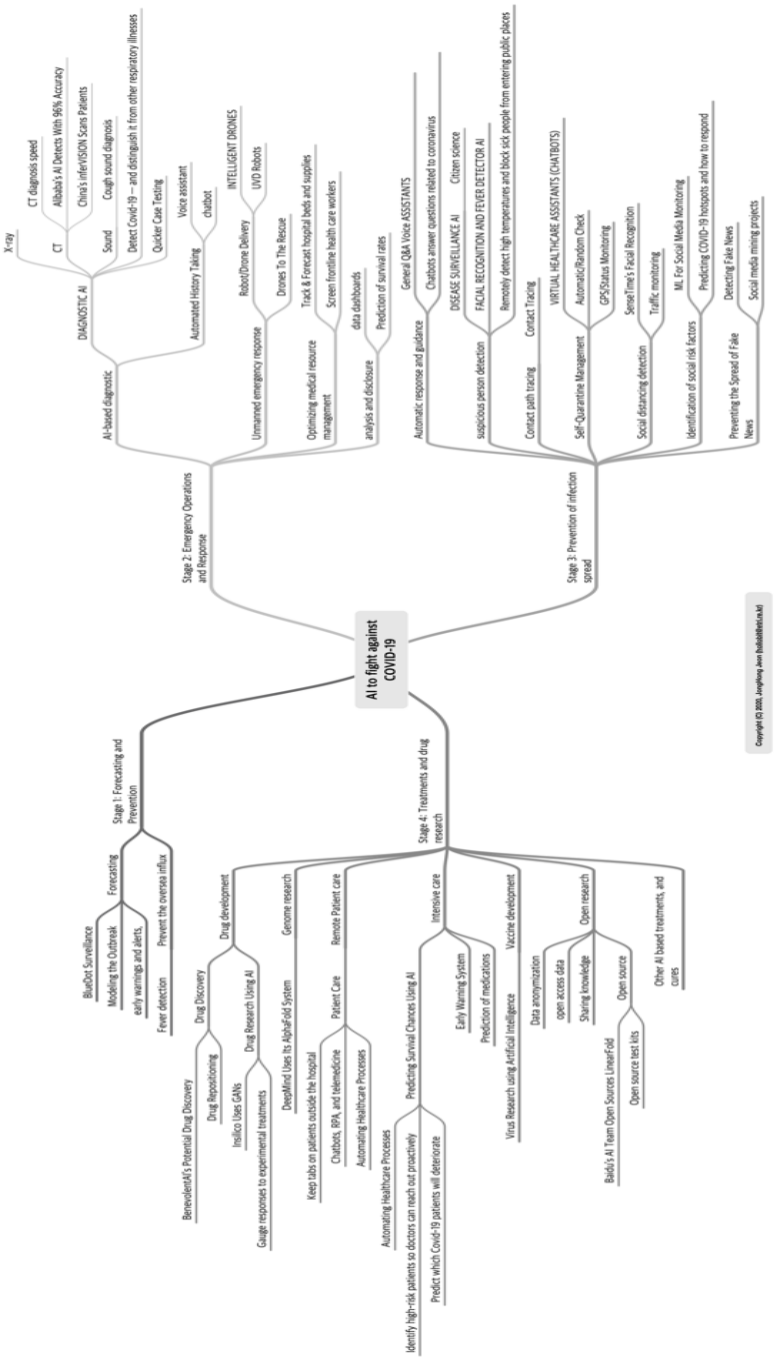


Figure 3: AI techniques against COVID-19 and SARS-CoV-2

Others replicate the categories identified by Jeon with some overlap and different names. For example, Pham et al., (2020) use 1) detection/diagnosis, 2) identifying, tracking and predicting outbreak, 3) "Infodemiology" and "infoveillance" (using social media and marketplace info to assess the disease and fight/circulate misinformation and/or propaganda), and 4) biomedicine and pharmacotherapy when speaking about potential studies.

Not all taxonomies resulted in four categories, although that does appear to have been the dominant approach. Some authors (Vaishya et al., 2020) use seven categories, including, for example, "Reducing the workload of healthcare workers" in addition to the more standard categories identified earlier.

The most recent approach to categorization is that of Latif et al. (2020). The authors of that extensive review (by far the most comprehensive of the papers cited here) increased the number of categories to nine and added four sub-categories for a total of thirteen, allowing for the recognition of subtleties that are often ignored. For example, the authors have a special section dedicated to applications in developing countries. Whitelaw et al. (2020) also look specifically at international collaborations and applications in the international context. The focus of this report, however, is on AI tools deployed in the US context.

Drawing from the inputs and directions derived from the papers mentioned above, this report focuses on applications of AI in five categories: 1) applications applied to forecast the spread of the virus, 2) medical applications to diagnose the disease, 3) applications to contain and monitor the spread of the disease, 4) applications to develop drugs and treatments, and 5) applications for social and medical management including workforce relief and supply chain optimization. These categories reflect the distinct groups responsible for implementing the applications from public officials to the medical community, they also help identify the populations at the “receiving end” of the application. For example, applications used to forecast the spread of the virus are intended for either public officials or public health specialists, not the general population. The fifth category is one that, in other taxonomies, is most commonly split up and integrated in other categories. The social management part is usually coupled with applications linked to “contain and monitor” and the medical management is made part of the “develop drugs and treatments” section. In this report, these applications are categorized together because they are not “directly” connected to COVID-19 and can be more readily applied after the pandemic. Each of these categories and the range/type of applications they encompass are described below.

1.3 Forecasting

Forecasting applications were the first to identify the existence of a new virus. Their success pushed AI to the forefront of the fight against the disease at its outset and they have been deployed to provide a variety of actionable insights throughout the pandemic.

On December 31, 2019, a health-monitoring company based in Canada called "BlueDot" alerted its customers to an outbreak of a new pneumonia-like disease originating from the Wuhan Province in China. This announcement was made seven days earlier than the official alert from

the US Centers for Disease Control and Prevention (CDC) and more than 10 days before the warning coming from the World Health Organization (WHO) (Neiiler, 2020). What made BlueDot's early announcement particularly remarkable was that it was made possible by the automated analysis of large amounts of data—both health-related and non-health-related—using AI techniques to predict both the outbreak of the disease and the way it would spread. BlueDot predicted correctly the few cities where the virus would be detected next.

Another use of AI to predict the spread of, and the level of danger posed by, COVID-19 was presented at a talk at the AAAS Annual Meeting in Seattle, Washington, in February 2020 and later published in *Science* (Li et al., 2020). Li and his collaborators used propagation models developed for the flu and combined them with five years of Chinese travel data around the lunar new year. From their analysis, they inferred that the virus was about 86% unreported. This estimate, based on data from early January 2020 and before the lockdown of Wuhan Province by Chinese authorities, was confirmed by a more recent study done using a much larger dataset (Hao et al., 2020).

These two examples represent a sample of the numerous applications that can be used to predict the spread of the virus. As the virus progressed, some authors used deep learning techniques to predict the spread of the disease state by state in the US (Yang et al., 2020). Yang's paper is notable in that they include in their algorithm an active way to account for the disproportionate number of minorities affected by the virus. Some other examples include models of the spread of the disease in China (Hu, Ge, et al., 2020), and around the entire world (Pal et al., 2020). Some authors also did similar studies at smaller scales to model the spread of the disease in colleges (Bahl et al., 2020) and in similarly small environments (see Liu et al., (2020) for a study on-board the "Diamond Princess" cruise ship).

1.4 Diagnosis

At the outset of any pandemic, it is crucial to have the ability to rapidly diagnose and screen the disease. Indeed, as seen in Nguyen (2020) (see also Figure 3), a large number of AI-based applications were aimed at faster diagnostics. Rapid diagnostics allow for an efficient assignment and triage of people going to hospitals. Multiple different AI-based diagnostic tools have been proposed during the course of the pandemic. Most of them have not been implemented at large scale but only in the context of small trials. This is mostly due to the difficulty in training efficient AI models using data that do not always reflect the population composition on which they will be applied. That issue is examined in more detail in the specific context of triage later in this report.

Because one of the most common manifestations of COVID-19 is pneumonia-like symptoms that affect the lungs, several diagnostic applications focus on the use of image-recognition software to accelerate the reading of lung X-rays and computerized tomography (CT) scans. There are more than 100 publications in MedRxiv and bioRxiv dedicated to this medical application, some as early as March 2020 in China (Zhou et al. 2020) and early April 2020 in Italy (Castiglioni et al. 2020). A majority of these applications use deep learning techniques based on Convolutional Neural Network, either on its own or in combination with Recurrent Neural

Network (Islam et al., 2020), or other neural network methods. AI-based applications can also be used to distinguish chest X-rays from COVID-19 and other diseases like influenza pneumonia (Zhou et al. 2020).

Coughing is another symptom of COVID-19 that can used to diagnose the disease (Imran et al. 2020). A mobile application can do a remote first-degree triage of the person taking a cough test and does not have to be as accurate as in a hospital setting because its goal is not to use that test to cure, but rather to estimate quickly if additional screening and medical attention is required (Patel, 2020). This application is important not only because it can help avoid unnecessary hospital visits and overburdening of limited medical resources, but also because of the possibilities it offers for greater accessibility to timely medical attention. In this case, the emphasis is not on an accurate diagnosis but on a medically viable screening application. Patel and his collaborators took care to develop an application that could be used with any telephone including landline (not necessarily a smartphone). This is particularly significant because the use of medical applications running on smartphones can prove an obstacle to accessibility for those with only a landline phone or a cell phone but not a smartphone. In the US, almost 40% of people 65 years old and above who own a mobile phone of any type own a cell phone but not a smartphone. This is also true for around 25% of the population with a high-school diploma, those making less than $30,000 a year, or living in rural areas (Pew Research Center, 2019). As described later in the report, this disparate access to technology is an issue relevant to other applications proposed in the fight against COVID-19.

1.5 Containment and Monitoring

Several different AI-based measures are being or were used to control the spread of the virus ranging from controlling the movement of people to tools to map the spread and potential reach of the disease. This category of applications has the most direct impact on the general population, not just people directly affected by the disease.

One of the symptoms of any disease is the elevated body temperature of the infected person. One of the first measures implemented in South Asian countries was to check the temperature of people on a massive scale in public places. Drone monitoring of people's temperature was proposed in Connecticut (Shackford, 2020) but most control points were established on the ground in train stations, at the entrance to buildings, and in other public places. Anyone with even a mild fever was refused access to these places. It is now known that temperature is not a reliable indicator of the disease, especially in its early stage: less than 44% of people with the disease have a fever early on (Guan et al., 2020). Similarly, as early as February 2020, China instituted an individual Quick Response (QR) code that allows authorities to control population movements (Hua and Shaw, 2020). The code, generated by a mix of self-reported data and access to the person's health records, produced a red, yellow, or green light that banned, quarantined, or allowed the phone's owner to proceed. The QR code exacerbated the anxiety of the disease for people who had no access to smartphone technology. Older Chinese, for example, experienced a two-fold impact of being the most vulnerable for this disease and the least able to safely leave their homes because of their lack of access to the technology (Wang and Jia, 2020).

Despite the novelty of the virus, it was quickly realized that asymptomatic people could transmit the virus (Bai et al., 2020) which, combined with an incubation period for this disease that varies from 2 to 14 days, accounted for the speed at which the virus spread from one city to the rest of China in less than a month (December 2019/January 2020). That issue, in part, explains the decision by most governments in the world to implement lockdowns and other population-wide restrictions on movement (the notable exception in Europe being Sweden). As time progressed, however, and the disease became more understood, people started to relax their adherence to the measures in place, triggering demands for their stricter enforcement. Proposals in the US included the use of facial recognition software coupled with heat maps to enforce social distancing (Associated Press, 2020; Nguyen, Saputra, et al. 2020) or large-scale measurement of people's body temperature from drones (Shackford, 2020; Mario, 2020) as was done in China (You, 2020-b). Others proposed using Global Positioning System (GPS) monitoring to enforce quarantine (Dobrea & Dobrea, 2020), also called "geo-fencing," which was indeed implemented in multiple countries (Ecuador, Hong-Kong, Israel, and South Korea). Some advocated for the use of machine learning algorithms to predict and prevent the spread of the disease. This was done at the early stage of the crisis, using travel data and similar transmission mechanisms as the flu (Li et al. 2020) or even monitoring wastewater (Venugopal et al., 2020). Some also tried to predict the emergence of COVID-19 hotspots by monitoring social media (Bouffanais & Lim, 2020) with relative success (Lopreite et al., 2021). One of the oldest tools used against the spread of a virus is called "contact tracing," a public health methodology that has been implemented since the influenza pandemic of 1918 (Fairchild et al., 2020).

COVID-19 accelerated the transition of that historic tool from a human-intensive investigation to an AI-based algorithm that could be implemented at large scale. South Korea, for example, implemented a program of contact tracing and strict border and population controls (You, 2020-a) almost immediately. Contact tracing using cell-phone-derived information is attractive because it allows a rapid response and works in cases where people do not necessarily know the persons that they may have infected (e.g., contacts in a supermarket, a restaurant, or in a bus). There are several techniques to implement contact tracing ranging from a simple GPS tracking of the phone to potentially privacy-preserving Bluetooth technologies. There are an increasing number of such applications which, because they require a large number of people, are developed and deployed at a governmental level. A good summary of these applications can be found in Ahmed, Michelin, et al. (2020) and Li & Guo (2020). Section 3 below describes in more detail the way these applications work and their limitations and dangers.

1.6 Drug Developments and Treatments

One of the most promising applications of AI in the fight against COVID-19, is one that uses the power of computation to develop new drugs or identify old ones that can be translated into new vaccines and therapeutics.

Developing drugs and vaccines has always been a lengthy process combining several basic science disciplines like biology, chemistry and pharmacology (UCI, 2020). A summary of the process and its usual timescale can be found in Figure 4 below (Brunning, 2016).

At the most basic level, anywhere between hundreds and thousands of chemical compounds are made and tested to find the one that fulfills the requirements. Only about one in a thousand possible drugs progress from preclinical testing to clinical trials and only one in 10 drugs entering phase 1 of clinical trials ends up being developed for sale (FDA, 2015). All in all, it takes an average of 10 years and more than $2.5 billion to develop one single viable new drug approved by the US Food and Drug Administration (DiMasi et al., 2016). Considering these numbers, it is not surprising that pharmaceutical companies looked for ways to cut costs and development time in the fight against COVID-19.

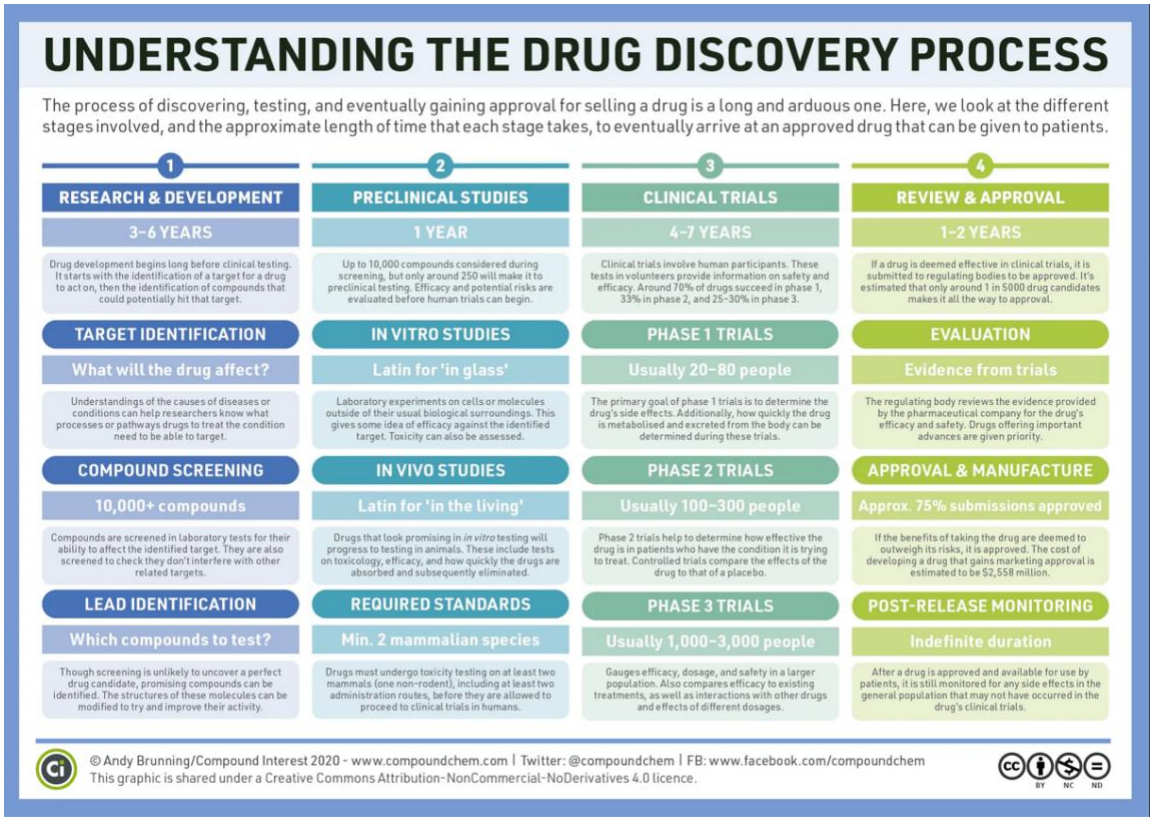


Figure 4: A schematic view of the traditional drug development process.

The power of AI-based algorithms in the context of drug development was quickly realized and starting in mid-2010, a number of pharmaceutical companies merged, acquired or collaborated with AI-centered software companies (Smalley, 2017). AI-based algorithms can be applied at the early stages of drug development, both in reducing the initial number of compounds considered and in eliminating the drugs for which the algorithm can predict a high-probability of adverse reactions. This trend accelerated with COVID-19 and now covers a larger array of the drug development process. A short overview of the possible applications of AI in this context can be found in Ho (2020).

One of the most encouraging paths has been the re-purposing of existing drugs. This has the advantage of cutting the time for approval because the drugs are already in use, with known

and measured side-effects. The approval in this case hinges only on the effectiveness of the drug for use other than the one for which it was initially approved. For example, as early as February 2020 BenevolentAI, a start-up using AI for drug development and identification, suggested the use of a rheumatoid arthritis drug called Baricitinib to alleviate the most severe symptoms of COVID-19 (Richardson et al., 2020). Eli Lilly, the maker of the drug, immediately partnered with the US National Institute of Allergy and Infectious Diseases on a large clinical trial of the drug and indeed, the drug has proven effective (Eli Lilly, 2020). The president of Lilly's biomedicines division acknowledged that his group would not have made the connection between the arthritis drug and COVID-19 without the aid of an AI-based match (Simonite, 2020).

Vaccine development is also affected by AI. A good review of the use of AI in vaccine development can be found in Russo et al. (2020).

Some have argued that there is a need to ensure that the AI-based algorithms used in drug development meet pre-defined standards. This was argued in a presentation at a meeting organized by the FDA[2] (Fisher, 2019). The need for pre-defined standards is particularly important because of the serious concerns raised by the population diversity or lack thereof in the datasets used by the algorithm in the drug-development process. There are deep worries that a drug could prove ineffective or worse, dangerous, when used on populations that have a different response to the biomarker that was used to develop or validate that drug (Ramamoorthy et al., 2015). COVID-19 has demonstrated the immense potential of using AI-based algorithms for the fast development and repurposing of drugs and this field of research will most certainly grow even after the current health crisis.

1.7 Medical and Social Management

Described here are applications of AI used to organize the large body of work that resulted from COVID-19 as well as to manage public perceptions and questions about the pandemic. These applications were conceived as the first interface between people and health authorities and could have a significant impact on large portions of the population, even after the COVID-19 crisis.

On March 16, 2020, the White House launched an initiative to coordinate research efforts among academic, government and industry scientists (Condon, 2020). The initiative resulted in the web-based COVID-19 Open Research Dataset (CORD-19) led by the Allen Institute for AI. This AI-powered website allowed researchers worldwide to quickly share knowledge and information about any aspect of research linked in any way to COVID-19. In addition, Microsoft (a partner in CORD-19) released visualizations of public databases of COVID-19 statistics that allow everyone (not just academics) to study the raw data released by the CDC (2020). The Asian Development Bank (2020) also offered several visualization tools for the spread of the

[2] https://www.fda.gov/news-events/fda-meetings-conferences-and-workshops/promoting-effective-drug-development-programs-opportunities-and-priorities-fdas-office-new-drugs

disease, tools that also allow comparisons between countries, different economies and varying responses during the pandemic.

Other applications of AI-based algorithms in the social management of COVID-19 include the increased use of chatbots for 1) addressing public concerns and questions about the disease (Miner et al., 2020; VolppKevin, 2020), 2) screening patients outside hospital settings (Espinoza et al., 2020; Martin et al., 2020), 3) screening healthcare providers (Judson et al., 2020), and 4) providing telehealth visits unrelated to COVID-19 during the pandemic (Bharti et al., 2020). Chatbot technology and Natural Language Processing are only two aspects of using AI-based algorithms in the pandemic. Healthcare in general has also benefited from the leap-forward in technology adoption that resulted from the pandemic, as reflected in the advance in and acceptance of remote-health care (Bokolo, 2020; Golinelli, 2020; Koonin et al., 2020; Bestsennyy et al., 2020) as well as benefits to the medical supply chain and allocation of resources (Wuest et al., 2020; Sinha et al., 2020; Baryannis et al., 2019).

2. Selection process: Methodology and Results

The following analysis draws from the above understanding of the applications that could or are being used in the fight against COVID-19 to identify the applications that raise particular ethical and human rights concerns, specifically applications that have the potential to cause disproportionate societal impacts on marginalized and vulnerable populations, both during the pandemic and also in the aftermath of the current public health crisis. To achieve this goal, a list was generated of more than sixty categories of applications mentioned in the surveys outlined in Section 1. Each application was then associated with an estimate of the scale of their impact (e.g., are they used on a large scale?), the timescale of their implementation (e.g., are they already in use? If not, when would they be used?), and a note of their benefits and risks in the context of human rights (e.g., privacy, accessibility), potential for ethical concerns, and risks of abuse.

The final part of the analysis was conducted with input from the members of the AAAS AI: Applications/Implications initiative Advisory Group and the AAAS Committee on Scientific Freedom and Responsibility. They were asked to highlight the potential benefits and risks associated with the different categories of applications that we generated, with a focus on 1) potential for a disproportionate impact on marginalized and vulnerable populations, 2) potential for ethical issues arising from these applications, and 3) absence or inadequacy of existing regulations. The focus was applications in use in the US.

The result was the identification of two main categories of AI applications:

1) **Sorting applications:**
    - **Triage applications** used to predict the likelihood of survival, evaluate probability of deterioration, and/or assign a medical score for eventual follow-up.

- **Resource allocation and distribution:** used in chain of supply decisions, like assigning ventilators to Intensive Care Unit (ICU) patients, allocation of scarce resources to hospitals, or, more recently, deciding the order in which individuals should receive a COVID-19 vaccine; and

2) **Surveillance applications**: used in the US for policing quarantine, geo-fencing, and contact tracing.

3 In-depth examination of the selected applications

This section is an examination of each of the selected applications, focused on the ethical and human rights issues that do or may arise from their implementation, and an examination of the differential impacts they can have on populations whether based on race and ethnicity, disability status, geographic location, socio-economic status, age, or another demographic characteristic.

Before beginning that examination, it is vital to point out that hand in hand with the specific ethical and human rights questions associated with AI-based applications to combat COVID-19, are questions of justification and validity. In the enthusiasm generated by the power of these new technologies, it is easy to lose sight of the fact that AI is not a panacea: it may not be the best tool to address a specific need. Furthermore, after the applications have been developed, they must be validated. The question "Do they work?" is one that should be addressed before implementing these applications at scale.

3.1 Ethical and Human Rights Frameworks

Many general frameworks for the ethical use of AI have been proposed. One survey and syntheses of the frameworks is provided by Fjeld et al. (2020) from the Berkman Klein Center for Internet & Society at Harvard University and another by Hagendorff (2019). Other recent reports include the general framework developed in Ricks et al. (2020) and faith-based frameworks including Moore et al. (2019) for the Ethics and Religious Liberty Commission of the Southern Baptist Convention or the Catholic Church in the Rome Call for AI Ethics (2020), signed by IBM and Microsoft[3]. These frameworks address different overlapping principles, some of which are relevant to medical ethics as described below. Fjeld et al. (2020), for example, classify 35 different ethical frameworks in the context of AI, in eight themes: Privacy, Accountability, Safety and Security, Transparency and Explainability, Fairness and Non-discrimination, Human Control of Technology, Professional Responsibility, and Promotion of Human Values.

In the domain of medical ethics, there are four guiding principles: autonomy, which concerns informed consent; beneficence, which expresses the concept that doctors should aim to benefit the patient; nonmaleficence, which conveys the initial Hippocratic Oath to "first do no harm"; and, finally, the principle of justice, which introduces the notion of fairness or equal treatment. These principles are sometimes expanded upon for more clarity. For example, medical research

[3] *Disclosure:* Microsoft is the primary sponsor of this initiative. The company had no involvement in the findings, conclusions, or recommendations of this report.

ethics in US follows the Belmont Report (HHS, 1979) for research ethics and the Menlo Report for information and communication technology research (DHS, 2012). The two frameworks can be combined (see the work done by Nebeker et al. (2019) on their Digital Health Framework) and include principles of privacy and confidentiality, in conjunction with other principles like usability or efficacy that are not included in most guidelines (Blasimme & Vayena, 2019). Additional to those first principles are concerns specific to the use of algorithms in health care (see for example Price, 2018; or Grote & Berens, 2020). In one of the earliest papers on a general ethical framework for digital tools specifically linked to the fight against COVID-19, Gasser et al. (2020) focused on the following tools of most relevance early in the pandemic: contact tracing; symptoms checking; quarantine compliance; and flow modeling. From this taxonomy, they derive an ethical framework with an emphasis on several basic principles shown on the left side of Figure 5 (Gasser et al., 2020).

Data sharing is an ethical issue arising in the context of medical research ethics, not included in the framework shown in Figure 5. Sharing health-related data is much less accepted in African American populations, for example, regardless of income, age, or education level (Sanderson et al. 2017). This is a subject fraught with historical mistrust (Byrd & Clayton, 2002; Skloot, 2010) that plays an important role both in triage and in the context of surveillance applications. The reluctance to share health data is also reported in Native American communities (Harding et al., 2012). There may be differences, however, between the trust in the medical care *received* and the willingness to share biological/health data with doctors while participating in medical research. One study (Wendler et al., 2005) suggests that there are no major differences in the *willingness* of some racial and ethnic groups (African American, Hispanic) to be included in medical research studies compared to their non-minority counterparts.

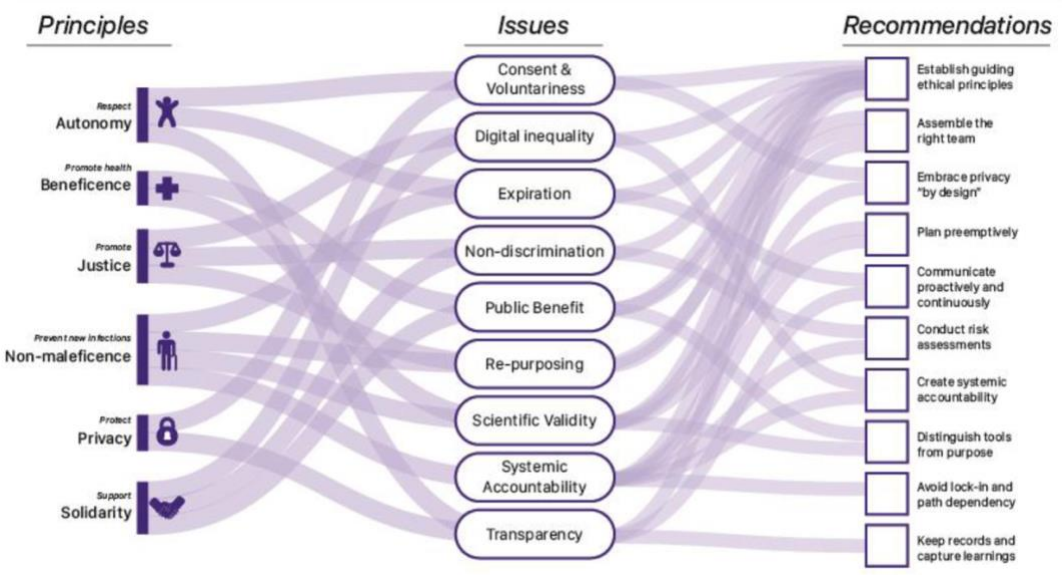


Figure 5: A schematic representation of the relationships between Principles, Issues and Recommendations (Gasser et al., 2020)

While not entirely disputing the statement that they could be more distrustful of the research, the authors hypothesize that two effects could be at work here: first, a shared desire by minority and non-minority people alike to help advance research by participating in these studies, and second, considering them as an alternative way to access health treatments or to receive compensatory payment. It does not necessarily contradict studies on the lack of trust in the overall health system. These attitudes are something that should be noted and taken into account when considering the issues of data sharing. We were unable to find reliable and recent data on the differences in trust and willingness to share health data based on sex, age, socio-economic status, political beliefs, geographic location, sexual orientation, religion, or disability status.

One category that we are not including in this section on AI ethics is the issues linked to ethical rights of machines (Turner, 2018). These issues may or may not become important in the future, but they are still on the fringe of the AI ethics framework discussions.

Complementary but separate from the ethical frameworks described above are human rights. Rooted in the same historical moment following the end of World War II, ethics and human rights share a commitment to the dignity of the person (Adorno, 2009). Overarching ethics and human rights frameworks diverge, however, in form and in substance in ways pertinent to our discussion of the implications of AI. Human rights are said to function as “an internationally accepted ethical discourse” (Baker, 2001). As such, and recognizing the “intellectual, practical and societal challenges that may accompany the task” (Ashcroft, 2010), human rights consist of a global framework that, when applied internationally or nationally, provides legal protection for individuals and groups, accountability for violations, and remedies for such violations.

Bioethicists, in particular, have grappled with the challenge of understanding the relationship between ethics and human rights. Some have suggested that human rights frame *what* should be done while ethics provide a framework for determining *how* those actions should be undertaken (Arras & Fenton, 2009). Another key distinguishing feature is that the human rights framework recognizes not only the dignity of the individual—the emphasis of an ethics framework—but is based on an expanded appreciation for humans in the fullest expression of their beings, including membership of specific demographic groups (e.g., race, class, gender, or religious group) (Plomer, 2009). An appeal to both frameworks is suggested by some scholars to be a means to “maximize the protection available to the vulnerable” (Peel, 2005), both individually and as groups.

The human rights framework incorporates several rights of particular relevance to our consideration of the implications of AI-based tools in the context of COVID-19, in particular the rights to equality and non-discrimination, the rights to privacy, to liberty and security of the person, the right to information, and the right to freedom of movement. These rights of general relevance to the current COVID-19 pandemic, exist alongside other rights of specific relevance to health and the relationship between science, technology and society.

The human rights framework encompasses a right to physical and mental health (Article 12, International Covenant on Economic, Social and Cultural Rights (ICESCR)). That right is

understood as encompassing both freedoms—for example, to be free from non-consensual medical treatment, and to control one's health and body—as well as entitlements—for example, to a health system that "provides equality of opportunity for people to enjoy the highest attainable level of health."

The human rights framework also recognizes a right to enjoy the benefits of scientific progress and its applications (Article 15, ICESCR). The 'precautionary principle' is understood to be an essential element of this right which establishes that "in the absence of scientific consensus, caution and the avoidance of steps are required in case an action or policy might cause severe or irreversible harm to the public or the environment" (UNESCO, 2009). This right, as defined by the United Nations, is also understood to require that national policies be developed consistent with widely accepted scientific evidence. At the same time, the right requires that measures be taken to prevent the use or misuse of science and technology for purposes inconsistent with human rights (Schabas, 2007). The US is not a party to the ICESCR, though the language of the treaty has been used to affect policy change in multiple contexts preceding the pandemic.

The final right of particular relevance to the pandemic context is the right to freedom from being subjected to "torture or to cruel, inhuman or degrading treatment or punishment. In particular, no one shall be subjected without his [sic] free consent to medical or scientific experimentation" (Article 7, International Covenant on Civil and Political Rights (ICCPR)). In the context of COVID-19, this requires full and informed consent by participants in clinical trial research on vaccines, treatments, and experimental technologies. The US is a party to the ICCPR.

While recognizing these universal and interconnected rights, and the obligations of governments to respect, protect and fulfill them, the human rights framework acknowledges that there will be distinct moments in time when, for reasons including public health, that some of these rights may be limited while still adhering to strict conditions for doing so. As explained by the Australian Human Rights Commission (Croucher, 2020): "Measures that limit our rights and freedoms on these grounds must always be necessary and **proportionate to the evaluated risk**, and must respect people's **dignity**, human rights **and fundamental freedoms**. These measures should be in place for the **shortest time possible** consistent with the emergency" (emphasis added).

3.2 Sorting Applications

3.2.1 Triage Applications:

The first focus is on AI-based applications to affect medical triage, both because such applications are already implemented at relatively large scale but also because medical triage is not specific to the COVID-19 crisis and could be used in many other contexts. This analysis does not address screening applications like AI-based medical bots (or robots programmed for medical applications) introduced very early in the pandemic to answer and sort phone calls from the public. The use of medical bots was to decrease the number of people who would visit emergency rooms by sorting people according to symptoms. Rather the focus is on the ethical

issues raised by AI-based triage in hospitals because these have the most devastating potential impact and potential for expansion into other domains beyond the current health crisis.

Medical triage occurs in recognition of situations where the number of incoming patients threatens to overwhelm available resources (Iserson & Moskop, 2007), and there is a need to prioritize which and in what order patients receive medical attention. Triage appears first in the context of the battlefields in the eighteenth century. At that time, the emphasis was on getting care to the soldiers who would die without any medical assistance, without using resources on those whose cases were deemed hopeless. Triage evolved quickly into an efficient way to identify the soldiers who could return to battle the fastest (Iserson & Moskop, 2007). Triage quickly found its way into hospital settings where it is used to prioritize people in emergency rooms. In US hospitals, a number is used for purposes of conducting triage. The most common number used in triage is the Emergency Severity Index which is derived from the acuity of the condition and the number of resources required for care (e.g., X-rays, lab tests). Another common index in the ICU context, is called the Sequential Organ Failure Assessment (SOFA) score which measures the failure of a patient's organs.

Neither of these indices were recommended in the context of COVID-19, when doctors needed to estimate the probability of a patient requiring ICU care or ventilators in the course of their treatments. One number, called the EPIC index, proved capable of predicting the trajectory of care for COVID-19 patients with relatively high accuracy. This number, derived from a proprietary software developed by their namesake EPIC company, is a single number associated with each patient and based on their health and demographic information. The idea is to capture their overall health in one single number to allow for a quick assessment of the urgency of their needs. This number was already used in hospital settings before the COVID-19 pandemic, as a quick way to assess the overall health of a patient.

When the number of cases in US hospitals started climbing, there was an urgent need to assess patients and predict the severe cases with high probability of requiring ICU care and the EPIC index emerged as a possible proxy for that decision point. As early as March 2020, doctors at the Stanford Health Care used the so-called "EPIC Deterioration Index" (EDI) as a way to measure their patients' risks of requiring an ICU stay and/or the use of a ventilator. The index is now used in over one hundred health systems in the US.

ETHICAL CONCERNS AND AI-BASED MEDICAL TRIAGE

Triage, whether based on AI-driven algorithms or performed in a traditional way, raises ethical issues. It is for this reason that medical triage is performed using a pre-defined protocol, outlined within a given framework. The utilitarian framework is the basis for most triage and allocation of resources protocols developed in medical centers in the US (see for example Daugherty-Biddison et al., 2017).

According to a utilitarian framework, decisions about triage are made to provide the greatest good to the greatest number of people. Other models used less frequently to triage patients apply egalitarian principles, whether on a first-come, first-served basis or randomized by a lottery or even "prioritarian," where the priority is given to certain people, regardless of needs.

Each framework raises its own ethical issues. For example, the utilitarian framework is rejected by ethicists at the National Catholic Bioethics Center (2020) because such "principles can be used to justify actions that undermine the dignity of the human person." Egalitarian models either ignore the difference of needs (lottery system) or increase the difference between a well-connected/well-informed population and others (that is the case for a first-come, first-served model). A "prioritarian" model raises fundamental issues of fairness and biases in the determination of the priority, (see for example Tolchin et al., (2020); Tolchin, Latham, et al., 2020).

There are some on-going efforts to articulate an ethical framework for triage that could or should be used in the context of AI-based triage. In a paper entitled "Five things every clinician should know about AI ethics in intensive care", Shaw et al. (2020), approach the problem outside any existing ethical framework. Dr. Shaw was interviewed for this report[4] and asked about the lack of an ethical framework. He acknowledged this unconventional approach but explained that he and his collaborators wanted to create "a list of ethical harms or ethical issues that could precipitate harms that need to be understood for applications of AI in intensive care." He argues that before a framework can be produced that could meaningfully navigate any AI-based medical decision, "there needs to be an accurate and empirically grounded understanding of what those harms are in the first place." He pointed out that "considering how early in the development of health-related applications of AI we are and how varied the use of AI in the medical field has been", there is a lack of a single typology of harms for these applications.

The list published in Dr. Shaw's paper reflects what he and his co-authors consider the most important ethical harms (or issues that precipitate ethical harms) that need to be considered by the responsibly practicing clinician in intensive care. The list includes recommendations for emergency-care physicians to 1) develop a general understanding about how AI works; 2) build patient trust not in AI, but in the patient-clinician relationship so the introduction of AI-based applications does not diminish that trust; 3) be able to assess the training data and use only ethically collected data; 4) mitigate bias by learning to identify algorithmic choices and data usage that can result in these biases; and 5) understand what sources of evidence exist to support the use of AI-based applications in clinical care. The burden put on emergency physicians in these recommendations is enormous and it is likely that some of the recommendations will become the responsibility of other members of the health care ecosystem, including administrators. Each of the guiding principles or safeguards recommended by Dr. Shaw and his collaborators are reflected in the ethical issues that we address below.

The ethical questions that exist about the use of AI-based triage during the COVID-19 crisis fall into two categories: **algorithmic issues** such as implementation, fairness and bias, explainability and transparency; and **data issues** including gathering, validation, training and sharing in the context of both accuracy and fairness of the results.

[4] Clips from the interview can be accessed on the website of the initiative at: https://www.aaas.org/ai2/publications/clips-interview-shaw

ALGORITHMIC ISSUES

Implementation

One of the first questions to ask of an application before its implementation should be "Does it work as intended?" In this case, we could ask "How robust is the evidence that the algorithm is indeed triaging the patients that are the most affected?" In a talk describing the use of triage within the Stanford Health Care systems (Etchemendy et al, 2020), Dr. Li described early efforts (early April 2020) to use both EDI and its time dependance to evaluate a deterioration probability of individual patients (2020) and concluded that the method turned out to be fairly accurate. This encouraged others to adopt EDI-based triage for their own patients.
Hu, Jacob, et al. (2020) also studied the challenges of implementing general AI-based tools in the context of the COVID-19 health crisis and formulated recommendations that are specifically relevant to the context of AI-based tools used for the purposes of triage. The recommendations address 1) robust validation of any AI models used in hospitals[5]; and 2) "local adaption", that is, the constant evolution of the algorithm to better represent the local data (populations, demographics) when the algorithms are used in different hospitals. Their second recommendation on "local adaption" is examined in the "data issues" part of this section.

Fairness and bias in triage

All AI-based algorithms are biased. This is not a moral statement, but a description of a fact based on the mathematics behind every classification or prediction. What is called fairness is best described as the independence of results with respect to a pre-defined attribute[6]. This choice of a pre-defined attribute defines several "fair" algorithms which are shown to be incompatible mathematically which each other (Pleiss et al., 2017). In other words, it is impossible for an algorithm to be fair for all its attributes. An excellent summary of the different definitions of fairness and why there is no easy answer to the question "Is this algorithm fair?" is provided in Rodolfa et al. (2019). In particular, they share a helpful example of a decision tree to decide what mathematical concept one should use in trying to build a "fair" application. That flow chart is shown in Figure 6 (from Rodolfa et al., 2019) and illustrates the multiple and different ways an algorithm can be optimized to be "fair."

The different parity "recipes" written in shorthand in the bottom boxes of Figure 6. provide to the programmer a recommendation on the quantity to be equalized in each particular case. For example, "FNR Parity" (or False Negative Rate) means that in the case of an algorithm assigning help to people and assuming that this help is not scarce, the best way to make that distribution "fair", would be to make that "False Negative Rate" equalized for all participants.

In other words, the programmer would make sure everyone has the same rate of false refusal so that the algorithm does not erroneously refuse aid to people who need it (False Negative) even if it means that people who do not really need that aid will also receive it (their False Positive rate can be much higher). This changes if the resources distributed are scarce (small fraction of possible interventions in the decision tree and an emphasis on "recall") or if the

[5] See Section 5 – Research Agenda. Question 4 on page 49.
[6] Issues arising from biased databases are addressed in the Data Issues part of the discussion starting on page 32.

emphasis is not on helping people but on punishing them (first branch out in the decision tree and three different possibilities of variables according to the desired equity).

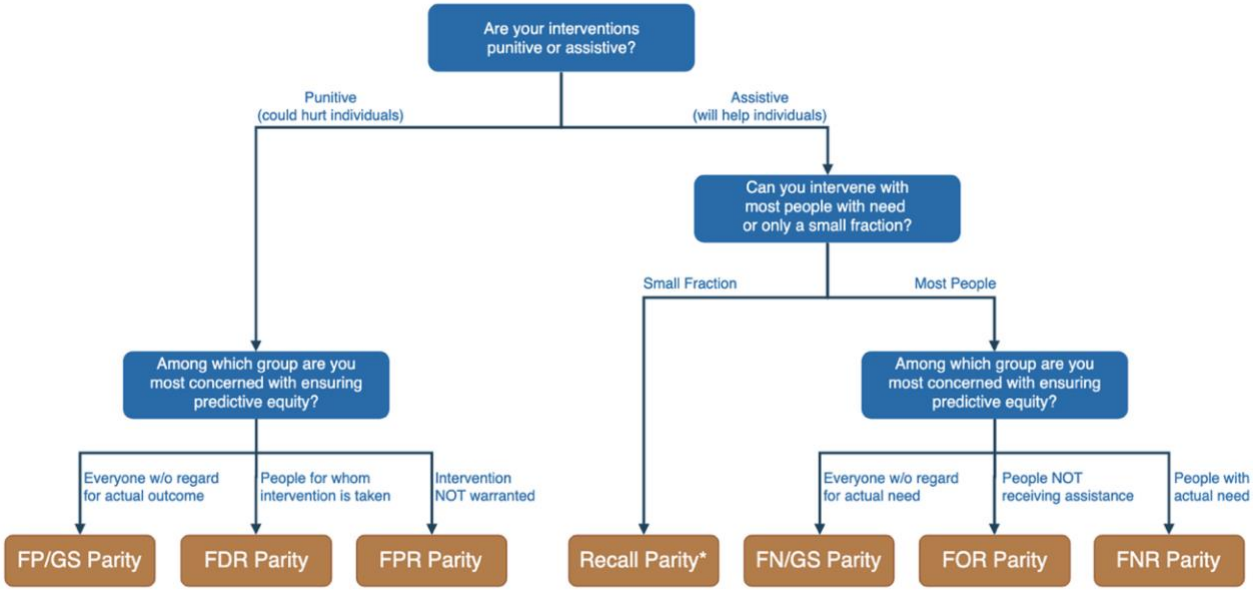


Figure 6: A schematic decision tree to measure/ensure fairness

In the field of medical triage, one of the most cited studies on algorithmic bias was published by Obermeyer et al. (2019). A large national health care system, unnamed in their paper but identified subsequently as UnitedHealth Care, was using algorithms to identify patients that could benefit the most from expensive care management programs. These programs were conceived as a way to provide additional resources (for example, dedicated nurses or additional appointments) to the sicker population of their patients. Obermeyer and his colleagues showed that at the same level of sickness, white patients were referred more frequently to these programs than black patients. In other words, the algorithm showed a racial bias against black patients who had to be sicker than whites to take advantage of these additional resources. This occurred despite the fact that the algorithm specifically excluded race as an input. So where was this disparity coming from?

Obermeyer and his colleagues were able to identify the cause of this difference because they had access to the entire data (inputs, outputs and outcomes). It turned out that, the issue arose from the misguided use of “future health care cost” as the decision point used by the company to triage their patient toward an extensive care management program. As pointed out in the paper, the choice is not unreasonable because as quoted in the paper “it stands to reason that patients with the greatest future costs could have the greatest benefit from the program.” That said, there are large differences in costs associated with white and black patients. The number quoted in the paper is that on average, black patients incurred health care costs about $1,800 lower than whites in that particular health system. The reasons for these differences can be split into socioeconomic barriers to accessing health care (geography, transportation, jobs, child-care issues), education-related obstacles like a lower awareness of reasons to seek care,

and reasons that are particular to the black community such as mistrust of physicians and health-care systems for historical reasons (Byrd & Clayton, 2002; Skloot, 2010).

The solution to the racial bias measured at the beginning of the study was to move away from the "future cost" variable and look for a better variable. To do so, Obermeyer et al. informed UnitedHealth Care of the issue they had uncovered. After checking these claims on a much larger database (more than three million people), the company and Obermeyer et al. started a collaboration to find and implement the use of a variable that would better reflect the patients' health as opposed to their cost to the health system.

The major lesson from Obermeyer's study should be that the choice of variable is what determines what bias an algorithm will have (because, again, bias is unavoidable and intrinsic to algorithms using complex datasets)[7]. There is indeed a growing literature on this so-called "problem formulation" of finding the correct variable to address the problem at hand in a way that minimizes specific biases (see for example, Passi & Barocas, 2019).

Explainability and transparency

The concepts of explainability and transparency are crucial for AI-based triage applications. A patient and a doctor should be both able to understand why and how the algorithm derives its results (Price, 2018). In the case of EDI, this is further complicated by the fact that the algorithm deriving the EPIC number is proprietary adding an additional layer to the "black box" quality of the result. The case of "black box" medicine is not unique to triage but because of the speed at which decisions are made in this context and the gravity of the potential outcomes, the concerns are more acute. Can an algorithm using an index number that is proprietary be trusted?

Dr. Singh, the lead author of a paper describing an attempt to study the different aspects of the use of EDI for medical triage in the COVID-19 context was engaged by AAAS in discussion about this issue (Singh et al., 2020). Dr. Singh is an Assistant Professor of Learning Health Sciences, Internal Medicine, Urology, and Information at the University of Michigan and he directs the Machine Learning for Learning Health Systems lab which focuses on using machine learning and biomedical informatics methods to understand and improve health at scale. The conversation with Dr. Singh spanned a Facebook Live event on the subject of Responsible AI and the issues concerning the use of AI in medical triage as well as emails exchanged in the process of writing this report[8].

Dr. Singh described his lab's effort to understand the potential use of the EDI as a tool to guide medical triage. To accomplish this, they logged all the EDI scores for all their patients, along with all the associated patient data. They then worked with a multidisciplinary team of researchers with the goal of developing an improved (and more transparent) version of the EPIC score. They confirmed the result from Li (2020) that using EPIC (or their own-developed

[7] See Section 5 – Research Agenda. Questions 12, 13, 14, 17, 18, and 19 on page 51-52.

[8] The complete talk can be found at: https://www.aaas.org/events/responsible-ai-medical-triage-during-covid-19-and-beyond

equivalent index), was an accurate way to predict the evolution of the disease in hospitalized patients. Examined in the next section is the next section the issue of bias, also studied by Dr. Singh and his collaborators.

DATA ISSUES

Gathering, validation, sharing and training issues linked to accuracy

Most hospital systems contract out the development of their AI-based algorithms[9]. This means that the algorithms are developed on data gathered from external patients who may or may not reflect the patient population of the hospital using the software. At the beginning of the pandemic, models were trained with data from patients with similar but not-COVID-19 related symptoms. That scarcity of data is the result of a slew of issues (data security, interoperability, data privacy) that render data sharing across health systems a challenge, a fact that impedes the validation of most in-house models. As discussed previously, the triage of patients in the Stanford Health Care System seems to accurately predict the evolution of the disease (Etchemendy et al, 2020). This example needs to be replicated at larger scale.

Validation using datasets that are larger and sampled differently than their own data is crucial for developing algorithms that can be used and accepted widely. The most famous example of this type of accuracy limitation is described by Buolamwini (2017) who details the failures of image recognition software when applied to Black or Asian faces. Addressing data limitations requires deliberate efforts to increase the size and diversity (racial and sex) of the population whose data are being used to test and train the models. In the case of the EDI triage model, this diversity should include race, ethnicity, socioeconomic status, age and education levels as all of these variables impact the overall health of individuals[10].

Sometimes, however, the accuracy of an algorithm is skewed for a reason that is independent of the diversity of the training set. The machines can pick up correlations from clues that are, at first, incomprehensible. In a famous recent example, researchers at Mount Sinai Hospital in New York City evaluated an algorithm used to read and triage chest X-rays to detect pneumonia (Zech et al., 2018). The sicker and bed-ridden patients were given X-rays using a portable machine. The accuracy of the algorithm dropped dramatically when applied to another hospital because unbeknownst to the doctors, the algorithm picked up on the differences between X-rays taken with portable and fixed machines, using them to identify the sicker patients. The algorithm's initial high accuracy was linked to the hospital and machine taking the X-rays rather than the images themselves.

These examples show the necessity of sharing data (the collaboration between these hospitals revealed the issue early on). To increase data sharing, however, among the concerns and challenges that need to be addressed is that of data security. One approach being proposed as a means of achieving greater privacy protection is the use of encrypted data to train models (Xu et al., 2019). Once popularized, this approach could open the door to wider data-sharing.

---

[9] We distinguish here between AI-based and rules-based (standard) algorithms.

[10] See Section 5 – Research Agenda. Questions 10 and 11 on page 51

Gathering, validation, sharing and training issues linked to fairness and bias

The issue of the validity of the training dataset in AI can also be understood in the context of a fair and unbiased process. Applying this optic, the validation process itself is at fault, not the composition of the patient dataset and its mismatch with the training set.

A recent article by Pierson et al. (2021) examines this type of bias in the context of the use of AI in health and triage (but not in relation to COVID-19) for patients suffering from osteoarthritis. For these patients, the severity of the disease and potential further recommendations for additional medical interventions is established using the Kellgren-Lawrence Grade (KLG), a standard radiographic scale developed in England in the 1960s. There is some evidence (Allen et al., 2009; Allen et al., 2010; Eberly et al., 2018) to suggest that at identical KLG levels, there exist pain level disparities among some populations classified along racial, socioeconomical or educational lines compared to the general population also suffering from osteoarthritis. The reasons for these discrepancies are not completely clear (there are no biological reasons that can explain a higher pain for people with a lower education level, for example). One explanation would be a "bootstrap effect": patients from these populations are used to be dismissed and ignored so they unconsciously report higher levels of pain that would make them more "visible" to a doctor.

Pierson et al. (2021) developed a machine learning algorithm (called ALG-P) to predict directly the pain experienced by patients using only radiographs and bypassing the use of the KLG classification to gauge the severity of the disease. They showed that, regardless of the composition of their training data, ALG-P was a better predictor of the pain than KLG, and more importantly, they showed that the more diverse the training data, the better ALG-P performances were. This study is significant because it shows a different approach to diagnostics and triage with a medical measurement as the input, but an output directly geared to the patients (not the potential severity of the disease) and the pain they experienced. These findings have the potential to suggest a way of adapting AI-based algorithms to close the existing gap between the underserved population's access to extended care and that of the general population.

HUMAN RIGHTS CONCERNS LINKED TO AI-BASED MEDICAL TRIAGE

Expiration date

The "expiration date" issue is as much an ethical as a human rights concern. It relates to the idea that measures that push the limits of what is ethical or that would constitute a violation of human rights in 'normal' times but are implemented to address the exigencies of a crisis should end when the crisis ends. There is currently no such end date for the triage applications implemented during the COVID-19 crisis. Indeed, some authors (Rivero, 2020) have foreshadowed that the successful implementation of EPIC-based triage during the COVID-19 pandemic may suggest use of similar applications for routine emergency room triage after the end of the pandemic. This needs to be examined carefully[11].

[11] See Section 5 – Research Agenda. Questions 5, 6, and 7 on page 50; Questions 28, 29 and 30 on page 53

LEGAL AND REGULATORY STATUS OF MEDICAL TRIAGE APPLICATIONS

In September 2019, the FDA issued guidance to "clarify the subset of software functions to which [it] intends to apply its authority" (FDA, 2019). The guidance document does not contain an explicit mention of triage applications, but it does state that applications that are intended for "use in the diagnosis or the cure, mitigation, or prevention of disease" may be subject to regulations. In January 2021, the FDA released another report on the issue (FDA, 2021) summarizing their approach to a pre-market review of AI-driven software. The report emphasizes transparency and performance monitoring by manufacturers at all stages of software development, including using Good Machine Learning Practice (during development), addressing the issue of bias while improving algorithm robustness, and following-up after any software release measuring Real World Performance.

Irrespective of whether a device uses AI or not, the FDA adopts a risk-based approach for assessing models that constitute "software as a medical device." Models that are directly used to drive care decisions or make diagnoses are treated as being high risk for the patients in case of erroneous decisions (class III and class IV) and are subject to premarket approval, a relatively stringent process. Models that are used for informational purposes typically fall into class I and II, for which the regulatory requirements are lesser. The required review in this case typically aims to establish whether the newly proposed device is substantially equivalent to one that is already approved and on the market. Currently the EDI assessment has not received this approval but similar models, such as the Rothman Index, have been approved[12].

Dr. Singh described the ways the EDI was and is still used, his answer illustrating the wide range of uses for this index. The score is used by Dr. Singh's rapid response team, allowing them to identify the sickest patients and proactively contact their nursing staff. The EDI is also used for allocating and assigning beds and anticipating possible demands on ICU beds. The software is only used when the information it provides is useful to support health care decisions. In other words, both the model's ability to predict the outcome and the effectiveness of this information in making care decisions ultimately dictates whether the team does use the model and how. It is never the sole input for that decision[13].

This approach is common among medical practitioners when surveyed about AI in medical care (Sarwar et al., 2019) and reflects the so-called "Kasparov law" which states that a decision taken by a knowledgeable human with inputs from a weak computer program will always be superior to the one taken by a machine alone (Kasparov, 2017). Of note is that Kasparov's Law may run contrary to medical AI liability practice. As shown in Figure 7 (Figure 1 in Price et al., 2021), a doctor may have an incentive to follow AI-based recommendations in all cases. Even in the case of nonstandard care recommended by the algorithm, there is evidence to suggest that a jury will be less likely to hold a doctor liable for a decision taken following the 'advice' of an AI-based tool, even the decision turned out to be incorrect (Price et al., 2021).

[12] The FDA approval of the Rothman index can be found at:
https://www.accessdata.fda.gov/cdrh_docs/pdf17/K172959.pdf

[13] See Section 5 – Research Agenda. Question 3, page 49

Scenario	AI recommendation	AI accuracy	Physician action	Patient outcome	Legal outcome (probable)
1	Standard of care	Correct	Rejects	Injury	Liability
2		Incorrect (standard of care is incorrect)	**Follows**	Injury	No liability
3	Nonstandard care	Correct (standard of care is incorrect)	Rejects	Injury	*No liability*
4		Incorrect	**Follows**	Injury	*Liability*

Figure 7: Comparison of potential legal outcome under law

ALGORITHMIC ISSUES: FAIRNESS AND BIAS

The question of bias in triage based on the EDI score was studied in detail by a group of medical doctors and computer scientists at the University of Michigan (Singh et al., 2020). Dr. Singh and his colleagues studied about 400 adults hospitalized with COVID-19 and focused on the ability of the EDI to predict both the risk of deterioration and the likelihood of having a mild form of the disease and be triaged out to a field hospital built to receive moderately ill patients.

Because the EDI is based on a proprietary score, Dr. Singh and his colleagues had to "un-wrap" the black box that this algorithm represented and address several questions beyond the simple accuracy of the prediction (although, that was measured too). While Dr. Singh and his colleagues studied the accuracy of using EDI for triage, they were explicitly looking for hidden biases related to categories such as race, sex, and age. They were also questioning the compatibility between the data on which the proprietary algorithm was trained and the data on which they were going to apply that code.

Dr. Singh and his colleagues found that EDI value was indeed good at predicting the evolution of the disease in a patient. Their study showed no bias with regard to race, sex or age and confirmed that the most affected patients were older with a host of underlying conditions including (but not limited to) cardiac arrhythmias, chronic kidney disease, congestive heart failure, and diabetes. However, the relatively small sample means that bias, even if present, could have missed by this study. If confirmed, this evidence of a lack of bias is important because it can help set the stage for a fair way (in conjunction with a human expert) to triage patients in settings other than COVID-19.

It is important to note the EPIC-based algorithm's lack of bias related to the patient's age. As has become clearer over time, COVID-19 is mostly fatal for older people. It has a mean fatality rate of around 8% for people older than 80 years old, compared to less than 0.45% for people younger than 65 years old and less than 0.006% for people younger than 25 years old (Berezow,

2020). In this case, therefore, an age-related bias will emerge from a triage algorithm trained solely on outcome data (the elderly fare much worse than younger patients, so an algorithm will learn to triage them out). This is the point made by Dr. Magnus, Director of the Stanford Center for Biomedical Ethics, member of the Ethics Committee for the Stanford Hospital, and a participant in the AAAS Facebook Live event on AI-based medical triage[14]. He pointed out that, in the utilitarian ethical framework, such algorithm would accurately assign a much larger risk to the elderly population, triggering a systematic triage of old people out of the system to benefit the larger population.

DATA ISSUES

AI training has an insatiable need for data. In the context of health care, gathering, accessing and using data is particularly complicated because of the multiple regulations in place to preserve patients' privacy and promote many other important concerns. In the context of a global health crisis, governments can use their additional emergency powers to mandate health data sharing at the international level—see for example, the "vaccines-for-data" agreement negotiated between Israel and Pfizer (Israeli Health Ministry, 2021). Although the need for data sharing is understandable, the related privacy and ethical concerns call for a careful balancing act (Beauvais & Knoppers, 2020). In addition, the gathering and sharing of health and even biological data without patients' consent has been historically abused. As seen at the beginning for Section 3, the issue of data sharing is important as it exposes deep mistrust in government, particularly in African American communities. Mistrust is also present to some extent in Native American communities (Pacheco et al., 2013). This is important because an algorithm's calculations will reflect the composition of the training set on which it bases those calculations and, depending on the context, the calculations can lead to erroneous conclusions when applied to datasets with different characteristics than the ones used for training[15].

3.2.2 Allocation of Resources:

AI-based allocation of resources in the context of the pandemic is closely linked to the issue of medical triage already discussed. Resource allocation happens in the same environment as medical triage: scarcity of resources that need to be prioritized or the necessity of rapid and efficient distribution of resources. An example is the allocation of Personal Protection Equipment (PPE) at the onset of the pandemic. As a result, the ethical and human rights issues related to triage also mostly apply to AI-assisted approaches to the allocation of resources.

AI as a tool to allocate resources predated the pandemic and was suggested early on to be applied in the pandemic in particular in resource limited countries. Bhansali and Jain (2020) suggested using machine learning algorithms to allocate hospital beds, manage healthcare and essential workforce assignments, and distribute vaccines. In their paper, the authors tackled vaccine distribution from the perspective of countries lacking the infrastructure to sustain the types of public health care campaigns necessary for the COVID-19 vaccine rollout. For example,

[14] The complete talk can be found at: https://www.aaas.org/events/responsible-ai-medical-triage-during-covid-19-and-beyond

[15] See Section 5 – Research Agenda. Question 10, page 51.

they address issues such as identifying people at risk of ignoring the protocol requirement of a second dose.

Multiple frameworks have been proposed for how best to affect widespread vaccination including with priorities based on age (US National Academy of Sciences, 2020), based on race (Dembosky, 2020) and on ZIP codes (Har & Taxin, 2021).

In a paper suggesting an approach for the allocation of scarce medications (but not the vaccine) in the context COVID-19, DeJong et al. (2020) recommend first to give the drugs to patients that will benefit the most from them. Their second recommendation is an egalitarian-inspired provision for no discrimination on the basis of "age, disability, religion, race or ethnicity, national origin, gender, sexual orientation, or perceived quality of life." This is important because patients with disabilities have been denied treatments during the COVID-19 crisis specially because issues of perceived quality of life (Shapiro, 2020).

It is in this context that AI-based algorithms have been discussed as a tool for allocating the vaccine. To our knowledge, however, no AI-based algorithm has actually been used to establish the priority order of the vaccine distribution at large scales. The only example of the claim for an AI-based solution for vaccine distribution was exposed as a scam in Philadelphia: a technology-driven start-up purported to develop an algorithm to deliver the vaccine but did not actually do so, leaving people, particularly in black and poor neighborhoods, to bear the consequences (Farzan, 2021). In addition, at least one rules-based (not AI-driven) algorithm was deployed in the Stanford community which resulted in prioritizing faculty staff before physicians (Guo & Hao, 2020).

While the issues linked to allocation of resources are similar, in most instances, to those seen in triage situations, one difference is that triage is conducted on an individual patient basis whereas the allocation of scarce resources is a statistical challenge. For example, vaccine prioritization is established by category of people, not by their name. The decision is made on variables that characterize the entire group, not an individual. This difference in application has implications for the way the vaccine was prioritized for zip code or race in some communities).

3.3 Surveillance Applications:

In medical vernacular, "surveillance" does not have the negative connotations attached to the word outside the health services industry. In a medical context, "surveillance" mostly refers to the act of identifying and keeping track of an infectious disease and monitoring its spread or progress. Earlier in this report, several of the applications described are used in that context. Surveillance of the sort that relates to identifying and keeping track of people are not currently being applied in the US, at least in part because they raise serious ethical and human rights concerns. There are signs, however, that some of these applications could find their way into the US response to COVID-19. For example, the Q-code implemented in China is being re-branded in the US as a "Green Pass." Below is a brief description of the most common surveillance applications and their ethical and human rights implications.

3.3.1 Contact Tracing

Contact tracing is a commonly used technique to track and monitor the spread of the disease. As explained in Section 1.5, the idea of contact tracing is not new, but the technical ability to carry it out at large scale is now possible mainly because of the ubiquity of smartphones. The basic idea of contact tracing is simple: the ability to control the spread of the disease is linked to the ability to determine who has been infected and with whom they have been in contact.

When one individual is diagnosed with the disease and asked to quarantine, health authorities are empowered to inform everyone who has been in contact with that person in a given period before the diagnosis and ask them to quarantine and report any case of disease themselves. If and when someone in that contact group reports being infected, the process starts.

Although the idea is simple, the implementation is not. In cases of diseases that are spread through extensive contacts (like HIV for example), old-fashioned contact tracing can be achieved by interviewing people, asking them for the names of their acquaintances and people with whom they have socialized recently. This approach was implemented with COVID-19 early on. Given the high rate of infection measured early in the pandemic, however, it soon became clear that time and human resources were insufficient to follow-up each patient's contact history. This is also the case when the disease spreads through the air and by droplets and thus could be transmitted by complete strangers in settings like a store or a bus (WHO-a, 2020).

As a consequence of these challenges early on in the pandemic mobile phone technology was turned to help trace individuals and their contacts using GPS to track their movements (ELLIS, 2020) and determine their proximity to someone who was or has been infected. That method was adopted in several countries (e.g., Hong-Kong, Singapore, China, Taiwan). Alternative approaches were taken in the US where a partnership between Google and Apple gave rise to an application for contract tracing (Google/Apple Exposure Notification or GAEN see Michael & Abbas, 2020)[16] and two identically called applications were developed simultaneously on the West Coast (Privacy sensitive protocols and mechanisms for mobile Contact Tracing, PACT-West by Chan et al., 2020) and the East Coast (Private Automated Contact Tracing or PACT-East by Rivest et al., 2020).

Both PACT applications put privacy of the cellphone phone owner at the center of their design. To summarize briefly, these applications rely on Bluetooth technology bar-strength measurement for proximity estimates (using Bluetooth Low Energy or BLE) and on chirps emitted every minute or so by every smartphone for timing estimates. These chirp values are randomly generated using a seed number that varies every hour. It is very difficult to connect the random seed to the owner of the cell phone. The device keeps two logs to track: 1) the seed values used to generate the chirps emitted (the seed log) and 2) chirps received from any other phone coming within the Bluetooth detection range (contact log). These two logs are stored for about 3 months on the device (long enough to be useful in the pandemic timescale). If a user tests positive for COVID-19, they can, through the application, update their status (in some

[16] Strictly speaking, the distinction should be made between "contact tracing" and "exposure notification" applications. One addresses the location of individuals, the other focuses on their proximity to others.

cases, this update is shared with health authorities) and the chirps' values they emitted for the past several days will be inserted in the "exposure database." The application on someone else's phone then compares the chirps' values in this exposure database with their local file of recorded contact chirps. A match is generated based on the exposure notification formula that takes into account how close and how long was the interaction by looking at the strength of the Bluetooth signal and the number of chirps matched.

Figure 8 illustrates how BLE applications work (Bradshaw, 2020). Of particular significance from an ethics and human rights perspective are certain details of the implementation of the contact tracing applications, in particular, whether the application uses a centralized database, its broadcasting method, and the nature of participation (mandatory or voluntary). These issues are described in more detail in the next two sections.

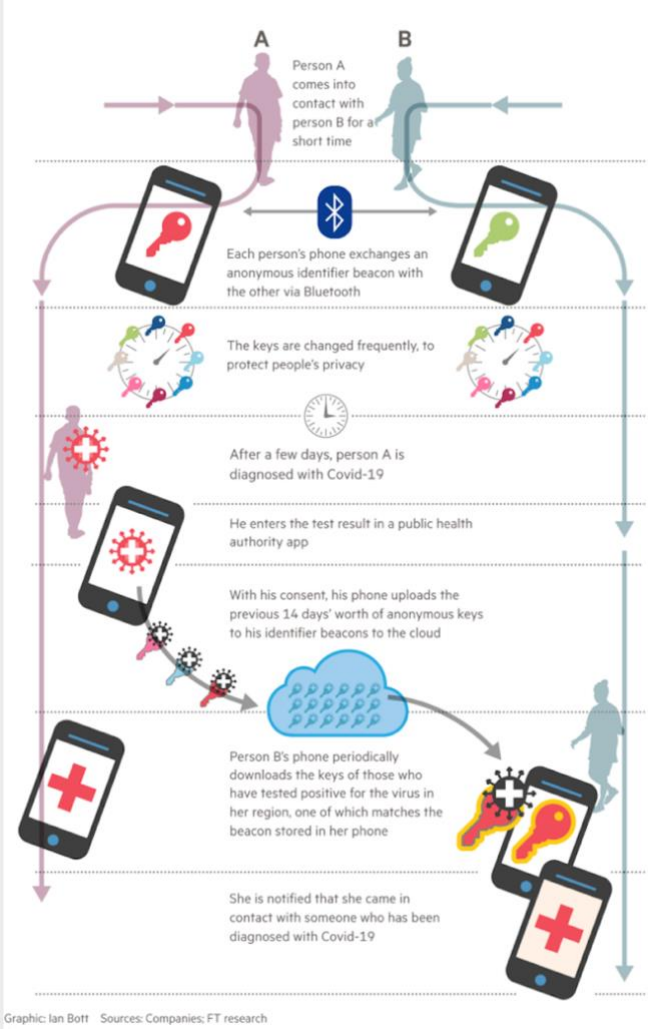


Figure 8: A schematic view of how contact tracing can work

As seen in the context of when to use AI in health applications (see Section 3.1), one of the first questions to ask is about efficacy: "Does it work?" The answer to that question in the context of

contract tracing applications is more complicated than the simple "yes or no" answer found for the triage applications (Sapiezynski et al., 2020). Unlike triage, the success of a contact tracing application relies on both technical and social considerations: the application has to be accurate in determining who was in proximity to an infected individual but also people using the application must be motivated to act when notified.

The technical aspect has not been studied in great detail. One exception is a paper by Hernández-Orallo et al. (2020), which was based on simulations rather than analysis of actual data and includes a series of assumptions. Additional questions range from estimates for False Negative Rate (or FNR which measures the missing contacts with infected people) and more crucially False Positive Rate (or FPR which measures the times a positive contact is flagged, and quarantine is recommended erroneously) to the question of timeframe for the maximum efficiency of the applications ("when as the pandemic is evolving should these applications be deployed?")[17].

There are very few studies on the False Positive (FP) rate of contact tracing applications. This rate measures the number of people receiving an exposure notification but who, in fact, were not exposed long or close enough to be at risk for infection. Measuring this rate requires technical calibrations of the signal strength as a function of the distance between the two phones. This has not been done extensively. One paper, published before the pandemic, studied the correlation between signal strength and distance in Bluetooth technology and found it was difficult to distinguish between phones at a distance of 1m from each other versus those at 3m (Sekara & Lehman, 2014). Another study, done by researchers in Dublin, used volunteers sitting in a train and looked for correlations between the strength of the Bluetooth signal (measured by a GAEN-like application running on their phone) and the distance between them (Leith & Farrell, 2020). The result shows no correlation at distances between 4 and 8 feet. This lack of technical validation is a problem for COVID-19 contact tracing usefulness as one distance is highly relevant for the disease's transmission whereas the other is too large to effect it. Some authors (Hernández-Orallo et al., 2020) assume a false positive rate as large as 0.7, meaning that as many as 70% of cases flagged to have been in contact with someone infected are not correctly assessed. If confirmed, such a large rate would make the applications almost useless[18].

Another question related to contact tracing applications is when, in the timeline of a pandemic, are these applications most useful (assuming that they are accurate)? A study done on traditional contact tracing estimates that they are most useful at the beginning of a pandemic when the number of infected people is low (Eames & Keeling, 2003)[19].

The social aspect of the question about efficacy is linked to the examination of baked-in assumptions like "Everyone has a smartphone", "A mobile phone tracks one person" (both at

[17] See Section 5 – Research Agenda. Questions 9 on page 50; Question 30 on page 53
[18] See Section 5 – Research Agenda. Questions 1, 2, and 3 on pages 58-59
[19] See Section 5 – Research Agenda. Question 8 on page 50

the heart of all tracing applications), or "Everyone will download the application" that are known to be incorrect. These questions are examined in more detail below.

ETHICAL AND HUMAN RIGHTS CONCERNS LINKED TO CONTACT TRACING APPLICATIONS

In their paper on the "Ethics of instantaneous contact tracing using mobile phone apps in the control of the COVID-19 pandemic", Parker et al. (2020) highlights several questions about the use of these applications. One of the ethical issues they raise concerns privacy, and the circumstances in which privacy infringements are justifiable if they can bring societal benefits. The authors argue not only that these privacy infringements would need to be minimized and carefully monitored and protected, but that precautions should be taken for the protection of privacy after the pandemic and the return to full privacy protection. The paper also examines the balance between privacy and freedom, equal access, questions on the compulsory nature of installing such applications and the fate and control of the data generated by the applications when the pandemic ends. Most are issues that will also arise for applications to verify a person's vaccine status, such as "green passports."

Some other issues raised are particularly relevant in the context of differential impact on marginalized and vulnerable populations. As described in the context of application-based diagnostics, most privacy-preserving contact tracing applications cannot work on smartphones that do not have the latest BLE technology enabled (Bradshaw, 2020). Most work only with Android operating systems more recent than Android 6.0+ and Apple iOS 13.5). This accounts for about 85% of Androids[20] and about 81% of iPhones[21]. When combined with the number of users of cellphones that are not smartphones and people who do not own a mobile phone (cell or smartphone), this suggests some limitation in relying on BLE technology to affect public health protection.

Digital contact tracing raises concerns for human rights because of its potential to lead to violations of privacy and enable mass surveillance (see for example, Chapter 6 in Landau, 2021). Buchanan et al. (2020) published a comprehensive review and a detailed analysis of contact tracing applications that underline some of these concerns. One technical consideration is the way the information from application users is collected and stored. Centralized applications (e.g., TraceTogether in Singapore, Aarogya Setu in India, or COVIDSafe in Australia) store all user data on a server that is used by health or governmental authorities to identify people in contact with a contaminated individual. In a decentralized application (e.g., PACT both West and East Coast or GAEN in the US, DP-3T in Switzerland), that matching of contacts with a carrier of the disease is done at the individual level and on a voluntary basis. Any centralized application raises the question of data governance, both during and after the health crisis, confidentiality privacy protections, including measures taken to ensure against data breaches and interference by foreign powers or malicious non-state actors (Buchanan et al., 2020; Landau, 2021). In the US, the PACT and GAEN applications were designed with privacy in mind (Rivest et al., 2020; Chan et al., 2020) while others have been developed with more concern for security (open-source code for transparency and technical validation). These issues still require

[20] See https://www.statista.com/statistics/271774/share-of-android-platforms-on-mobile-devices-with-android-os/
[21] See https://www.statista.com/statistics/565270/apple-devices-ios-version-share-worldwide/

careful consideration given the potential use of this type of application after the pandemic and for other lesser health crises like the flu (Khan, 2020; Parker et al., 2020)[22].

Beyond the underlying technology of the contact tracing application, another consideration is what is done with the data collected and what protocols for communication are in place to disseminate the exposure notification. For example, in Singapore and Hong Kong, health authorities broadcasted a map with the location of all the cases of people infected by the disease (Raskar et al., 2020). Such a broadcast would be illegal in the US for violation of privacy and medical information disclosure (Pandit et al., 2020). In the US, the working model is a "unicast" broadcast with the information sent to people suspected to have been in contact with a person infected. Still, each method requires some kind of centralized health agency and the privacy risks in all cases range from moderate to high. At a more basic level, each method requires adoption of the technology for it to have impact.

A decentralized communication system requires the participation of people and their willingness to share information concerning their health status. The additional issue of this method is the low probability of acting upon a notification received from such an application in particular. Would anyone isolate themselves for ten days just upon receiving a notification from an application? That question is particularly relevant for people for whom quarantine is not easily done (lack of space to isolate, impossibility to work from home, and necessity to work)[23]. The rate of false positives (asking people to quarantine when, in fact, they were not exposed to the virus) for contact tracing applications has not been widely advertised. As mentioned in the previous section, a false positive rate as large as 0.7 would result in a lack of trust in the applications and a further eroding of their usefulness. A study on the issues surrounding contact tracing applications (Sapiezynski et al., 2020) show their low adoption rate as illustrated in Figure 9.

Not shown on this graph, but consistent with its findings, is a report indicating that less than 4% of the French population downloaded the STOP-COVID application released by the government (Rowe et al., 2020) and less than 30% of US population said that they would use a contact tracing application (Gitlin, 2020). Both numbers are well below the 50 to 80% participation rate required for such applications (centralized and decentralized models) to work (see Figure 3 in Hernández-Orallo et al., 2020).

Another important flaw in this type of self-reporting application is security. In a pandemic, it is easy to assume that everyone using such applications is reporting honest data. It is also possible for a foreign power or a malicious non-state actor to report false information to cripple or overwhelm the legitimate response of a country. This consideration of security and potential for attacks is one that is not often examined but it is no less important than privacy and should be carefully weighted as well (Buchanan et al., 2020).

[22] See Section 5 – Research Agenda. Questions 27, 29 and 30 on page 53
[23] See Section 5 – Research Agenda. Questions 24, 25, and 26 on page 53

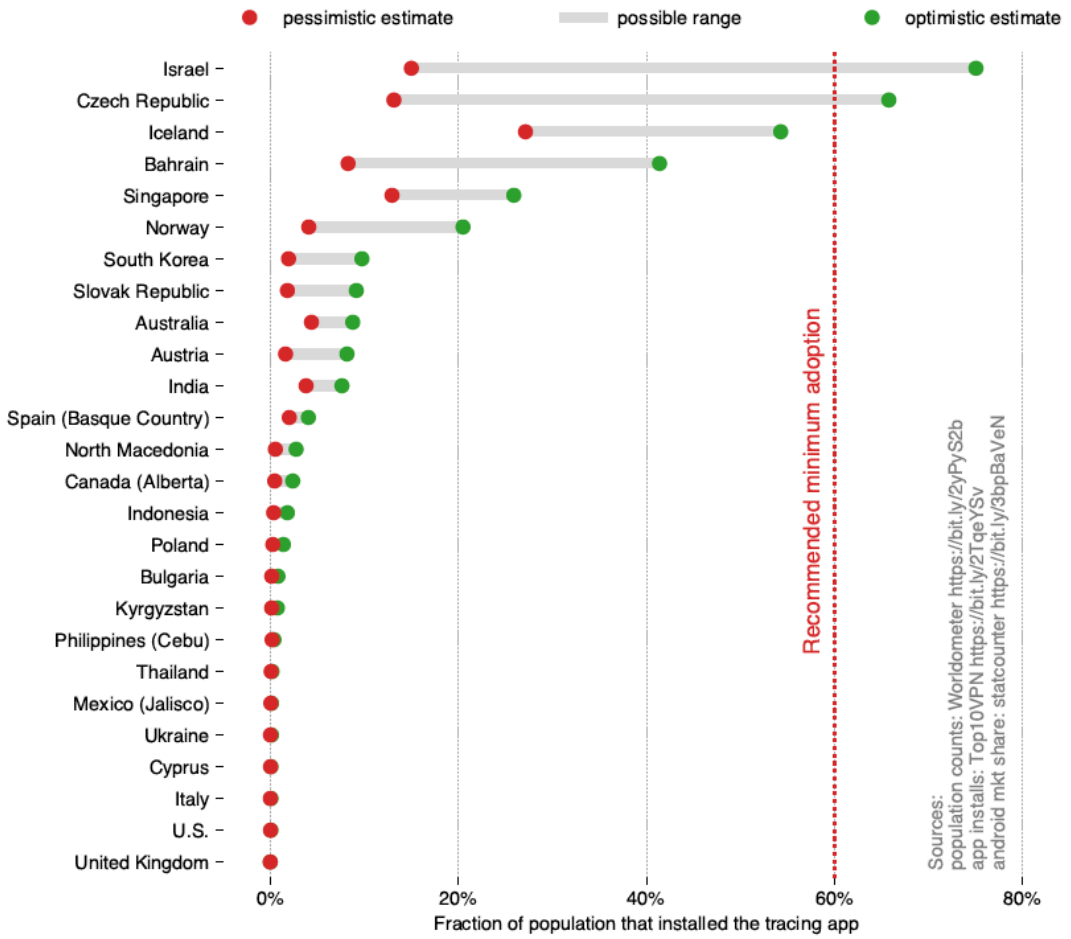


Figure 9: Estimate adoption rate of contact tracing applications

One of the dangers of discussions about these applications is that too often they are evaluated only in term of their efficacy to curb the virus' propagation. A lot of the arguments for curtailing civil liberties have followed an "it works" approach. Examples to show the efficacy of the most extreme forms of contact tracing (for example, a mandatory GPS-based surveillance system) are usually drawn from China's numbers showing the rapid control of the virus following their implementation (Yakabuski, 2020). Further studies are needed to determine the reliability of these statistics over time and in other contexts.

All of the concerns raised by contact tracing have a disproportionate impact on underserved populations (Davis, 2020). As already noted, tools that can only work on smartphones have limited utility when, as seen previously in this report, almost 40% of people 65 years old and above who own a mobile phone of any type, do not own a smartphone. This is also true for about 25% of the population with a high-school diploma, those making less than $30,000 a year, or living in rural areas. (Pew Research Center, 2019). One outstanding and related question is whether for all communities it is possible to assume an equivalence between a mobile phone and one owner, something that is not the case in most of the developing world

(Erikson, 2018). There is no apparent data on this effect for the underserved populations in the US It is, therefore, not possible to say with certainty how big of an effect it is[24].

One final concern relates to surveillance and control of populations that have historically been over-surveilled and controlled (Bedoya, 2016). A recent international survey (Simko et al. 2020) conducted on-line and reaching a mostly young, white, educated, international population shows that less than 40% of people would download a contact tracing application that shared its data with the government. The different conditions of download are shown in Figure 10 (from Simko et al., 2020). No similar study in the US was found, but assuming similar attitudes among marginalized populations in the US and considering that most models assume minimum participation rates around 60% for success (Hernández-Orallo et al., 2020), the 40% rate shown above would render ineffective any implementation of contact tracing applications[25].

3.3.2 Geofencing / Green Passports

Prior to the pandemic, geofencing or the monitoring of the position of a cell phone as a proxy to locate its user, was primarily a marketing tool. For example, advertisers could target people according to their location and send them an advertisement tailored to their physical proximity to a specific store. In the context of the pandemic, geofencing repurposes knowledge of the cellphone user's position. As early as March 2020, geofencing was touted as the panacea for enforcing quarantine by commercial companies eager to capitalize on the fear of the disease (ProtectYu Pres Release, 2020) and was implemented in China and Hong-Kong (Hui, 2020).

Geofencing can also be used to mark infected areas and alert health or police departments (Culham, 2020). Geofencing has not been introduced in the US due to its potential use to curtail freedom of movement and the potential for abuse it could enable (Wesner, 2019). For similar reasons "immunity passports", that were floated early in the pandemic, were rejected (Kofler & Baylis, 2020; Privacy International, 2020; and Ada Lovelace Institute, 2021).

Since the deployment of the COVID-19 vaccine, however, there have been discussions in some US states about introducing a "green passport" that would allow its bearer to circulate freely while at the same time prevent anyone without it access to some public places and/or the ability to travel. Australia has already implemented a "vaccine passport" (Macmillan & Norman, 2021). Anyone receiving the vaccine is recorded in a centralized database, the Australian Immunization Register, containing that person's entire immunization history. Other countries and regions that have done so include Denmark, the European Union, Israel and the Netherlands.

[24] See Section 5 – Research Agenda. Questions 15 and 16 on page 51
[25] See Section 5 – Research Agenda. Questions 20 to 23 on page 52

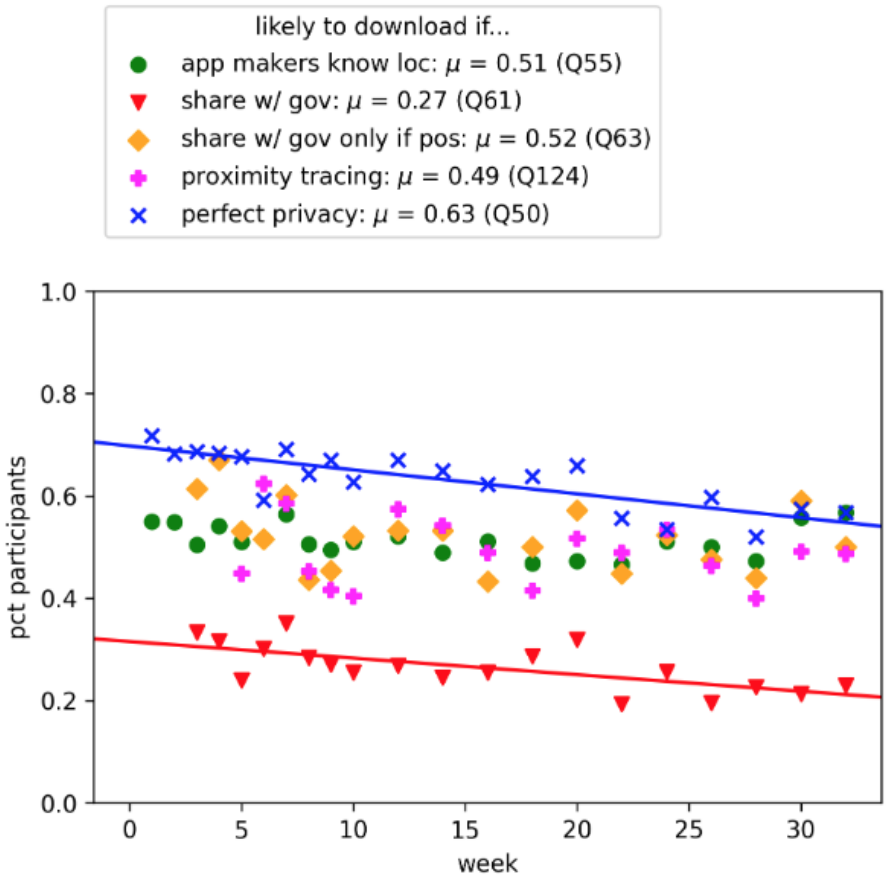


Figure 10: Time evolution of people's attitude toward sharing data from contact tracing applications

ETHICAL AND HUMAN RIGHTS CONCERNS LINKED TO GEOFENCING/GREEN PASSPORTS APPLICATIONS

"Green passports" that are digital in nature provide the bearer with a unique digital fingerprint, most often a QR-code, that is provided through a centralized government vaccine verification process. Concerns about such a tool relate to both their public health utility as well as their human rights impacts.

The WHO has advised against the issuing of "green passports" or immunity certificates mostly due to concerns about the period of validity of immunity conferred to people who had the disease (WHO-b, 2020). Most other commentators in this field have focused on the ethical and human rights concerns raised by the tool, in particular the risks of entrenching a stratification between different populations (Voo et al., 2021) and the same data concerns that arise in the context of contract tracing.

Not all individuals will be vaccinated. Some religious groups, for example, have expressed opposition to the vaccines available in the US because they were either developed (Johnson & Johnson) or tested (Pfizer and Moderna) using fetal cell lines (Wadman, 2020). The Vatican issued a statement allowing vaccination for practicing Catholics in countries (like the US) where

no other choices are available[26]. The text also recognizes the moral dilemma such decisions pose and allows practicing Catholics to choose alternative solutions. A recent report from the Royal Society (2021) identifies other groups that may be reluctant to get the vaccine, including people at risk for serious allergic reactions, elderly with one or more pre-existing conditions, and pregnant women, and raise the question of whether a "green passport" required to work might force someone to get a vaccine when it is not medically advised.

In a report on the human rights implications of a digital health passport, Beduschi (2020) warns that such a passport "may interfere with the respect and protection of data and human rights, in particular the rights to privacy, equality and non-discrimination, and the freedom of movement, assembly, and to manifest one's religion or beliefs." The author recognizes that such measures may be justified but argues for safeguards and an adequate balance between individual rights and public interest. How to affect such a balance is considered by Brown et al. (2020) who argue that the tradeoff between restricting the movement of few for the benefit of the many is one that our societies could and should make. From a human rights perspective, the question becomes one of what is necessary and proportionate in the context of the public health crisis, and when are the measures no longer necessary[27].

As discussed in the context of contract tracing, because of the scope and nature of the data collected, and the potential for abuse. In the US, groups such as the Social Science Research Council's (SSRC) Public Health, Surveillance, and Human Rights Network have formed to map the current state of surveillance and the "new normal" of COVID-19 (SSRC, 2020). Although not centered on digital health passports, the report does mention the issues linked to the fallacy of a "voluntary" application installation necessary for everyday life. The SSRC calls for "critical thinking that anticipates how to maximize positive health effects, while preventing breaches of privacy and non-health related uses of such data is necessary" (SSRC, 2020).

4. Conclusions

More than a year since the COVID-19 pandemic began, AI's power in the fight against the disease has been both visible and largely unmonitored. AI-based applications have been adapted from pre-pandemic uses for diagnosis and drug development, for forecasting the disease's spread and monitoring population movement. New applications have also been developed to address novel and acute needs in the pandemic, including in the provision of healthcare.

At this moment of rapid development and deployment of AI, this report was prepared to capture the breadth of AI applications relevant to COVID, and to analyze these for their social impacts, particularly those that gave rise to the most serious ethical and human rights

[26] See entire text at: https://www.vatican.va/roman_curia/congregations/cfaith/documents/rc_con_cfaith_doc_20201221_nota-vaccini-anticovid_en.html

[27] See Section 5 – Research Agenda. Questions 27, 29 and 30 on page 53

concerns. The goal being to ensure that these concerns were adequately known, an information base provided to address the concerns in the context of the pandemic, but also to ensure that when the global health crisis abates careful consideration is given to whether and, if so, with what oversight and consideration are these applications integrated into non-pandemic life. The review of AI-based technologies being deployed in the course of the pandemic revealed the breadth of applications being used and proposed, and the need in each case to assess the extent to which each application meets both technical validation requirements but also the public health need for which it was deployed. In the case of both contract tracing applications and medical triage tools, technical validation questions remain and standards by which to validate the tools are lacking. Data validation is an additional important aspect that determines the effectiveness of the tool which, if not explicitly screened for potential biases in the data can lead to biased, unfair and unjust applications.

Data fairness is just one of the ethical and human rights issues that arises in the context of the application of AI-based tools in the context of the COVID-19 pandemic. As the report acknowledges, a pandemic presents a particularly unique circumstance, and tackling the crisis presents unique challenges. The human rights framework recognizes that such a crisis may necessitate actions that are contrary to human rights but demands that any such actions be necessary in the context, that they not be contrary to that basis tenets of human dignity, and that they be limited in time and scope.

The human impacts of the AI-based technologies used in the context of the pandemic are potentially immense, be they at the individual scale in the context of medical triage, or the societal scale in the context of contact tracing. The specific experiences of marginalized populations impacted by these technologies is inadequately documented, a gap that needs to be filled as lessons are drawn from the current crisis and the real potential exists for the continuation or redeployment of these tools in the future.

To that end, the last part of this report, detailed in Section 5, contains the lessons learned from this detailed inspection of the selected applications. It proposes a path forward for researchers and developers, policymakers, and funders. That path implies addressing fundamental issues in the conditions in which the applications are validated and trained while ensuring viable data governance and privacy protections. The report also identifies several other gaps in the general understanding of the implications of these applications' widespread use. These gaps are described below.

5. Moving Forward: Proposed Research Agenda

The preceding assessment of ethical and human rights concerns arising in the context of the implementation of AI-based applications in the context of the fight against COVID-19 revealed several unanswered questions worthy of research. The research agenda is divided into three sections: the first pertains to the technical aspects of the applications addressed, specifically with regard to the algorithms and the data used to develop them; the second concerns the

societal context in which the applications are developed and used; and the final section focuses on the specific human rights and ethical issues raised by the applications.

Three features of this research agenda are essential to recognize:

- It is not exhaustive. It highlights currently pressing areas for future exploration in a rapidly changing landscape.
- Though categorizing the research gaps as technical, data, and ethical, social and human rights related, it is not always possible to separate the ethical and human rights components of the technical and data gaps.
- While the research agenda focuses on the applications used in the context of COVID-19 that are the focus of this report, the research needs identified would help fill gaps of relevance across a broader range of AI-based applications, generally, and that will persist long after the health crisis driven by COVID-19 is over.

Technical

In medical applications, the first and most important question to be answered is that of scientific validity: Do the applications work? Do they achieve their intended goals? There are two different aspects to these questions, one concerning the strict technical validation of the application and the other linked to the medical value of the application results (the application works but is the result medically useful?).

Another important question linked to the algorithms is that of the period over which they are deployed and how that timescale should be determined.

A) Technical validation

When AI-based contact tracing applications were proposed (and implemented in some countries), they were hailed as "game-changers." The technical idea behind each of these is sound and the applications "work" in a way that they can indeed connect in time and space one person to another due to the proximity of their phone (in the example of a Bluetooth-based application). The relevant questions though should not be Can these applications identify a close connection between two telephones? but rather, Are the results from these applications, medically useful? In the case of contact tracing applications, there is a significant lack of research that serves to adequately answer that question.

Contact Tracing Applications:

Before implementation, the following variables need to be measured and made public.

1) Actual False Positive/False Negative rates for identifying people in contact with infected individuals and those who need to quarantine as a consequence of that contact
2) The calibration of the correlation between the Bluetooth signal strength and the distance between two telephones. In addition, there is a need to quantify:

- The impact of the physical position of the phones on the subject on the quality of the measurements (Does a phone in a bag provide a measurement as accurate as a phone in a back pocket or inside a coat?)
- The measure of the maximum distance still classified as a "close contact" by the software and its comparison with the medically relevant distance associated with a risk of infection. This comparison will determine if the software/application is medically beneficial.

3) Time calibration to estimate the "time in contact" of the phones.

Medical Triage Applications:
Validate medical triage applications by addressing the question:

4) How strong is the evidence that the algorithm used to conduct triage is indeed triaging the patients that are the most affected by the disease?

B) Software Oversight/Standards

Oversight and standards are necessary in the context of AI and health because, without either one it is not possible to measure the success of the tools (do they accomplish their goals?) and their impacts.

As early as 1997, Miller and Gardner published a list of recommendations for the responsible monitoring and regulation of clinical software systems (1997a & 1997b). Two of their recommendations (endorsed by multiple organizations including the American College of Physicians Board of Regents) were a local oversight of clinical software systems through the creation of independent "software oversight committees" and a system of four monitoring classes based on the level of clinical risks associated with the software. There is also currently a drive to establish standards in AI-assisted medical diagnoses (not included in our detailed analyses) by the World Health Organization in collaboration with the International Telecommunication Union (WHO/ITU, 2020). Triage applications are classified by the FDA as "high-risk" patient-specific systems, and as such should require the highest degree of oversight.

Anderson & Aydin (1994) enumerate ten criteria that would optimize the ethical oversight of computer-based tools. The list includes questions on results, cost, and training, with the following particularly relevant to AI in the context of the current pandemic:

- What are the anticipated long-term effects on how organizational units interact?
- What are the long-term effects on the delivery of medical care?
- Will the system have an impact on control in the organization?

In addition, standards, similarly to the guidance and criteria on the management of electronic health records (EHRs) developed by NIST in collaboration with the Office of National Coordinator are needed in the context of AI (Goodman et al., 2014). Such standards would address "testing requirements, test cases, and test tools" for both classes of applications (sorting and surveillance) and be informed by the "meaningful use" requirements applied for EHRs (https://www.nist.gov/itl/ssd/meaningful-use).

Sorting Applications: Medical Triage

The creation and implementation of independent software auditors should address the technical validity of the algorithms used for medical triage and provide an ethical assessment of the tools proposed, to ensure that this technical "certification" is rooted in existing ethical frameworks.

5) What are the anticipated post-pandemic effects on how each unit in the health organization using triage applications interact?
6) What are the post-pandemic effects on the delivery of medical care?
7) Will the system change the dynamic of work within the health care provider organization?

C) Optimal Time of Operation and Deployment

As there are talks about extending the reach of surveillance applications for future health crises and perhaps for routine flu seasons, there is a need to understand at what time their implementation has the maximum impact and if there should be a sunset clause to retire them once the health crisis has passed. During the COVID-19 crisis, the implementation time varied from country to country as some (South Korea or Singapore) had the experience of the Middle East Respiratory Syndrome health crisis and processes already in place. This was not the case for the US, for example, and it was not until the April/May 2020 time frame that contact tracing applications started to be released. It is not clear that the applications (assumed here to be medically and technically validated) could help at that stage, when the virus had already spread throughout the country. Studies combining disease propagation with applications adoption models should be able to define the optimum time for intervention using contact tracing. Monte-Carlo simulations, for example, are used for risk assessments or long-term predictions. First proposed by Stanislaw Ulam and John von Neumann (Eckhardt, 1987) to predict a set of outcomes based on a range of inputs and a probability distribution for these outputs, Monte-Carlo models are run thousands of times to produce their likely outcomes.

For Contact Tracing Applications:

Running Monte-Carlo or Monte-Carlo-type simulations would be valuable to answer the following questions:

8) When is the implementation of the application optimal to successfully track the majority of cases?
9) How much more efficient in tracking cases are AI-based contact tracing applications compared to traditional contact tracing methods?[28]

D) Data Validation

Issues of data validation are not new to the applications reviewed in this report (Chen, Pierson, et al., 2020), but they deserve to be mentioned nevertheless because of their importance in a medical setting. Models must be trained on data that has been validated. This is particularly important in applications, like the triage one examined here, that are

[28] The measure of the efficacy of traditional contact tracing is given at: https://www.cdc.gov/coronavirus/2019-ncov/php/contact-tracing/contact-tracing-plan/evaluating-success.html

deriving health status from measurements of patients' vitals like blood pressure or heart rate. In this case, validation includes ensuring a similar population composition to the one on which the application is going to be used.

Medical Triage Applications:

10) *Validation:* How relevant are the populations on which any future similar program will be trained to the populations on which it is going to be used? Similarity should not be restricted to race and ethnicity but include age, sex, socio-economic status level, education level, location of residence, disability status.

E) Bias and Fairness Validation

Validation of algorithms to check for bias or unfairness includes looking for hidden biases derived from uninformed choices in variables, the so-called formulation problem, such as was documented by Obermeyer (2019), or examining baked-in assumptions that turn out to be erroneous, like the scale used in measuring knee pain described in Pearson et al. (2021). Understanding the technical roots of such biases is vital and reveal themselves at the technical level (i.e., by changing the variables used in the decision-making process).

Another baked-in assumption which is the basic premise of any contact tracing application, is the equivalence made between a cell phone signal and only one person. A shared or communal cell phone renders irrelevant any information derived from contact tracing. To our knowledge, there is no study showing the validity of this hypothesis in the US across different subsets of the population.

Medical Triage or Allocation of Resources:

11) Did any differences in the populations to which the algorithms were applied and the training data lead to inaccurate results?
12) *Audit:* Could the algorithms' variables or decision points embedded within algorithms hide or mask issues that trigger unforeseen bias?
13) How are these variables or decision points derived?
14) What would be the consequences on the outcome to choose different variables or decision points in the algorithm?

Surveillance Applications:

15) How robust is the assumption that a cell phone belongs to only one person in the US?
16) How does this assumption vary within different communities in the US and how does this compared to what is measured in the developing world?

Both Sorting and Surveillance Applications:

17) *Forensics:* Examine both applications and determine **all** the underlying assumptions present in the development and implementation
18) *Completeness:* Is there a systematic way to derive or at least identify the maximum number of assumptions for a given application?

19) *Completeness:* Are all the variables used by the algorithm necessary for the application? Inversely, is the algorithm using the correct variables?

Ethical, Social and Human Right issues

Described previously are some of the ethical and human right concerns linked to the implementation of the applications addressed. Additional concerns arising from a normalization of the measures taken during the health crisis are described below.

F) Data sharing

An accurate AI-based algorithm can only be developed using large amounts of data. Data sharing, however, relies upon careful consideration of multiple data governance concerns: data collection (informed consent on data gathering), data access (once collected, what is done with it and who has access to it), and data benefits (what is the benefit to participate for the populations involved). There is a need to understand attitudes toward these issues in the different subsets of the populations that are the center of this study. As explained in Section 3, there is also a need to separate the issues of the willingness to participate in health research pilots, with that of the lack of trust toward different entities involved. Some populations may be willing to enroll in studies despite their lack of trust, because of other considerations that can take precedence (desire to help advance science, desire to help society to fight diseases, inability to access health care practitioners otherwise, monetary considerations). Although there are some studies and surveys (some contradictory) on these attitudes within African American communities, there are very few data specific to other groups based on sex, age, socio-economic status, political beliefs, geographic location, sexual orientation, religion, or disability status.

Both Types of Applications

Studying the obstacles for widespread data sharing – in the context of surveillance, the reluctance to download and use the contact tracing application and in triage applications, the lack of shared medical data to efficiently train the algorithms.

20) Is lack of trust the main issue for lack of data sharing across different communities?
21) Does the lack of data sharing result in an algorithm that disparately impacts different communities?
22) What are the reasons for the existing disparities between the large participation of some groups in health research studies but their reluctance to consent to sharing data?
23) What is effective in increasing the understanding among diverse communities about the value of sharing data in specific contexts and circumstances, and their trust in doing so?

G) Compliance

Most contact tracing systems rely on people to comply with instructions (e.g., to quarantine, or self-isolate) upon receiving them from a given application. Studies on the compliance rate upon receiving such instructions, if they exist, have not been widely

disseminated. Yet, compliance rates have broad implications for the effectiveness of these tools and should be taken into account in any plans to expand the implementation of these applications beyond the current crisis. There is no current plan to make compliance mandatory.

Contact tracing applications:

24) How does compliance with health directives change when instructions are coming from an application rather than a real person?
25) How does the degree of compliance change based on the level of difficulty or inconvenience in following the application's directives (measuring the differences of compliance between an order of social distancing and an order to quarantine)?
26) How and why do compliance attitudes vary across different populations?

H) Post-Pandemic Concerns/Review

There are clear indications of a thrust to keep some of the applications tested or implemented during the COVID-19 past the current crisis. For example, the triage applications implemented with relative success to identify the people most at risk for future ICU admission could be used in "normal" triage in ER. If confirmed, this large-scale implementation would require a better understanding of how the application's output is used by the physicians and how this value inform or modify their health care decisions.

We have already commented on the necessity to implement a sunset clause on applications and measure (with simulations) their shelf life. In addition, each application needs a "post-mortem" analysis to assess their potential usefulness in any future health crisis.

Contact tracing applications:

27) What, if any, health data that has been collected for the purposes of addressing the COVID-19 pandemic can and should be retained for the purposes of addressing other public health crises in the future?

For medical triage applications:

28) What value would be served by continuing to use the AI-based medical triage tools developed for the purposes of the COVID-19 pandemic, and does that value out-weigh any potential ethical or human rights concerns?

For both classes of applications

29) Did the implementation of the applications serve the purpose for which they were implemented? What impact, positive or negative, did their use have on the established goal? For example, did their use result in more efficient triage, fairer allocation of resources, larger-scale contact tracing or more accurate geofencing?
30) What measurable parameters should trigger the applications' deployment in the future, and what should trigger their rescindment?

Next Steps

This report, together with a commissioned study on the attitudes of marginalized populations toward AI as applied in the context of health and the COVID-19 pandemic will inform the elaboration of a responsibility framework that will provide a roadmap for developing and implementing just and ethical AI-based medical applications. This roadmap will be conceptualized and articulated by a cohort of thought leaders in a wide variety of fields ranging from ethicists to computer specialists and from human rights activists to lawyers and public servants. We anticipate that the result will help both AI practitioners and lawmakers and policy makers to usher a new era for AI-based medical applications.

6. Bibliography

Ada Lovelace Institute (2021). 'What place should COVID-19 vaccine passports have in society? Findings from a rapid expert deliberation chaired by Professor Sir Jonathan Montgomery.' [online] Ada Lovelace Institute. Available at: https://www.adalovelaceinstitute.org/summary/covid-19-vaccine-passports/

Adorno, R. (2009). Human Dignity and Human Rights as a Common Ground for a Global Bioethics. *Journal of Medicine and Philosophy* 34, (2009): 223-240.

Ahmed, N., Michelin, R. A., Xue, W., Ruj, S., Malaney, R., Kanhere, S. S., … & Jha, S. K. (2020). A survey of COVID-19 contact tracing apps. *IEEE Access*, 8, 134577-134601.

Alabool, H., Alarabiat, D., Abualigah, L., Habib, M., Khasawneh, A. M., Alshinwan, M., & Shehab, M. (2020). Artificial intelligence techniques for Containment COVID-19 Pandemic: A Systematic Review. https://assets.researchsquare.com/files/rs-30432/v1/5d76ab69-8b3a-45db-8d5d-11a495138cdc.pdf Accessed 08/31/2020

Allen, K. D., Helmick, C. G., Schwartz, T. A., DeVellis, R. F., Renner, J. B., & Jordan, J. M. (2009). Racial differences in self-reported pain and function among individuals with radiographic hip and knee osteoarthritis: the Johnston County Osteoarthritis Project. *Osteoarthritis and cartilage*, *17*(9), 1132-1136.

Allen, K. D., Oddone, E. Z., Coffman, C. J., Keefe, F. J., Lindquist, J. H., & Bosworth, H. B. (2010). Racial differences in osteoarthritis pain and function: potential explanatory factors. *Osteoarthritis and cartilage*, *18*(2), 160-167.

Anderson, J. G. & Aydin, C. E. (2005). Overview: Theoretical perspectives and methodologies for the evaluation of healthcare information systems. *Evaluating the organizational impact of healthcare information systems*, 5-29.

Angwin, J., Larson, J., Mattu, S., & Kirchner, L. (2016). Machine Bias. There is software that is used across the county to predict future criminals. And it is biased against blacks. https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing Accessed 01/20/2020

American Public Media Research Lab, (2020). Color of Coronavirus. https://www.apmresearchlab.org/covid/ Accessed 09/08/2020

Arras, J. D. and Fenton, E. M. (2009). Bioethics and Human Rights: Access to Health-Related Goods. *Hastings Center Report*, 39 no. 5 (2009): 27-38.

Ashcroft, R. E. (2010). Could Human Rights Supersede Bioethics? *Human Rights Law Review* 10, no. 4 (2010): 639-660.

Asian Development Bank (2020). ADB COVID-19 Policy Database, https://covid19policy.adb.org/ Accessed 09/01/2020

Associated Press (2020). "A northwestern Indiana city wants to install a facial recognition video network camera system downtown in hopes of slowing the spread of the coronavirus." https://www.wlfi.com/content/news/Indiana-city-considering-cameras-to-help-in-contact-tracing-572626341.html Accessed 10/20/2020

Baesens, B., Setiono, R., Mues, C., & Vanthienen, J. (2003). Using Neural Network Rule Extraction and Decision Tables for Credit-Risk Evaluation. *Management Science* 2003 49:3, 312-329. https://doi.org/10.1287/mnsc.49.3.312.12739

Bahl, R., Eikmeier, N., Fraser, A., Junge, M., Keesing, F., Nakahata, K., & Wang, L. Z. (2020). Modeling COVID-19 Spread in Small Colleges. *arXiv* preprint arXiv:2008.09597.

Bai, Y., Yao, L., Wei, T., Tian, F., Jin, D. Y., Chen, L., & Wang, M. (2020). Presumed asymptomatic carrier transmission of COVID-19. *Jama*, *323*(14), 1406-1407.

Baker, R. (2001). Bioethics and Human Rights: A Historical Perspective. *Cambridge Quarterly of Healthcare Ethics* 10 (2001): 241-252

Baryannis, G., Validi, S., Dani, S., & Antoniou, G. (2019). Supply chain risk management and artificial intelligence: state of the art and future research directions. International Journal of Production Research, 57(7), 2179-2202.

Beauvais, M. and Knoppers, B. M. (2020). COVID-19: Who's Afraid of Data Sharing? Canadian Science Policy Centre https://sciencepolicy.ca/posts/covid-19-whos-afraid-of-data-sharing/ Accessed 02/21/2021

Bedoya, A. (2016). The Color of Surveillance: Government Monitoring of the African American Community https://www.law.georgetown.edu/privacy-technology-center/events/2016-conference Accessed 02/07/2021

Beduschi, A. (2020). Covid-19: Human Rights Implications of Digital Certificates for Health Status Verification https://socialsciences.exeter.ac.uk/media/universityofexeter/collegeofsocialsciencesandinternationalstudies/lawimages/research/Policy_brief_-_Digital_Health_Passports_COVID-19_-_Beduschi.pdf Accessed 02/02/2021

Berezow, A. (2020). COVID-19 Infection fatality rates by sex and age. American Council on Science and Health https://www.acsh.org/news/2020/11/18/covid-infection-fatality-rates-sex-and-age-15163 Accessed 12/10/2020

Bestsennyy, O., Gilbert, G., Harris, A., & Rost, J. (2020). Telehealth: A quarter-trillion-dollar post-COVID-19 reality? https://www.mckinsey.com/industries/healthcare-systems-and-services/our-insights/telehealth-a-quarter-trillion-dollar-post-covid-19-reality Accessed 01/20/2021

Bhansali, A. & Jain, V. (2020). Using Machine Learning for Healthcare Resource Allocation in COVID-19: Opportunities and Challenges for LMICs. Center for Global Development 09/15/2020 https://www.cgdev.org/blog/using-machine-learning-healthcare-resource-allocation-covid-19-opportunities-and-challenges Accessed 12/27/2020

Bharti, U., Bajaj, D., Batra, H., Lalit, S., Lalit, S., & Gangwani, A. (2020, June). Medbot: Conversational artificial intelligence powered chatbot for delivering tele-health after covid-19. In *2020 5th International Conference on Communication and Electronics Systems (ICCES)* (pp. 870-875). IEEE.

Blasimme, A., & Vayena, E. (2019). The ethics of AI in biomedical research, patient care and public health. Patient Care and Public Health (April 9, 2019). *Oxford Handbook of Ethics of Artificial Intelligence*, Forthcoming.

Bokolo, A. J. (2020). Exploring the adoption of telemedicine and virtual software for care of outpatients during and after COVID-19 pandemic. *Irish Journal of Medical Science* (1971-), 1-10.

Bouffanais, R., & Lim, S. S. (2020). Cities—Try to predict superspreading hotspots for COVID-19. https://www.nature.com/articles/d41586-020-02072-3 Accessed 11/30/2020

Bradshaw, T. (2020). Two Billion phones cannot use Google and Apple contact-tracing tech: System developed by Silicon Valley relies on technology missing from older handsets *ArsTechnica* April 2020 https://arstechnica.com/tech-policy/2020/04/2-billion-phones-cannot-use-google-and-apple-contract-tracing-tech/?comments=1 Accessed 08/20/2020

Brown, R. C., Kelly, D., Wilkinson, D., & Savulescu, J. (2020). The scientific and ethical feasibility of immunity passports. *The Lancet Infectious Diseases*.

Brunning, A. (2016) Understanding the drug discovery process. *Compound Interest* https://www.compoundchem.com/2016/01/16/drug-discovery/ Accessed 04/21/2021

Buchanan, W. J., Imran, M. A., Ur-Rehman, M., Zhang, L., Abbasi, Q. H., Chrysoulas, C., … & Papadopoulos, P. (2020). Review and critical analysis of privacy-preserving infection tracking and contact tracing. *Frontiers in Communications and Networks*, https://doi.org/10.3389/frcmn.2020.583376

Buolamwini, J. A. (2017). Gender shades: intersectional phenotypic and demographic evaluation of face datasets and gender classifiers. Thesis: S.M., Massachusetts Institute of Technology, School of Architecture and Planning, Program in Media Arts and Sciences, https://dspace.mit.edu/handle/1721.1/114068 Accessed 04/20/2020

Byrd, W. M. and Clayton, L. A. (2001). An American Health Dilemma: Race, Medicine, and Health Care in the United States, 1900-2000. ISBN-13: 978-0415927376, NY, Eds: Routledge

Castiglioni, I., Ippolito, D., Interlenghi, M., Monti, C. B., Salvatore, C., Schiaffino, S., Polidori, A., Gandola, D., Messa, C., & Sardanelli, F. (2020). Artificial intelligence applied on chest X-ray can

aid in the diagnosis of COVID-19 infection: a first experience from Lombardy, Italy. https://doi.org/10.1101/2020.04.08.20040907

CDC (2020). Health Equity Considerations and Racial and Ethnic Minority Groups, https://www.cdc.gov/coronavirus/2019-ncov/community/health-equity/race-ethnicity.html?CDC_AA_refVal=https%3A%2F%2Fwww.cdc.gov%2Fcoronavirus%2F2019-ncov%2Fneed-extra-precautions%2Fracial-ethnic-minorities.html Accessed on 08/14/2020

Chamola, V. Hassija, V. Gupta V., and Guizani, M. (2020). A Comprehensive Review of the COVID-19 Pandemic and the Role of IoT, Drones, AI, Blockchain, and 5G in Managing its Impact, in *IEEE Access*, vol. 8, pp. 90225-90265, 2020, doi: 10.1109/ACCESS.2020.2992341. https://ieeexplore.ieee.org/document/9086010 Accessed 08/23/2020

Chan, J., Gollakota, S., Horvitz, E., Jaeger, J., Kakade, S., Kohno, T., … & Tessaro, S. (2020). PACT: Privacy sensitive protocols and mechanisms for mobile contact tracing. *arXiv preprint* arXiv:2004.03544.

Chen, I. Y., Pierson, E., Rose, S., Joshi, S., Ferryman, K., & Ghassemi, M. (2020). Ethical machine learning in health care. *arXiv e-prints*, arXiv-2009. https://ui.adsabs.harvard.edu/abs/2020arXiv200910576C/abstract Accessed 02/20/2021

Chen, J., Li, K., Zhang, Z., Li, K., & Yu, P. S. (2020). A Survey on Applications of Artificial Intelligence in Fighting Against COVID-19. *arXiv preprint* arXiv:2007.02202. https://arxiv.org/abs/2007.02202 Accessed 08/01/2020

Condon, S. (2020). White House leads effort to publish COVID-19 open research data set https://www.zdnet.com/article/white-house-leads-effort-to-publish-covid-19-open-research-data-set/ Accessed 03/30/2020

Croucher, R. (2020). https://humanrights.gov.au/about/news/e-bulletin/community-update-covid-19-april-2020 Accessed 02/12/2021

Culham, D. (2020). Geofencing informs COVID-19 Response. *Tyler Technologies Blog*. https://www.tylertech.com/resources/blog-articles/geofencing-informs-covid-19-response Accessed 12/25/2020

Daly, A. (2019). Algorithmic oppression with Chinese characteristics: AI against Xinjiang's Uyghurs. In: *Artificial intelligence: Human rights, social justice and development.* APC, pp. 108-112. ISBN 9789295113138

Daugherty-Biddison, L., Gwon, H., Regenberg, A., Schoch-Spana, M. M., & Toner, E. (2017). Maryland Framework for the Allocation of Scarce Life-Sustaining Medical Resources in a Catastrophic Public Health Emergency.

Davis, S.L.M. (2020). Contact Tracing Apps: Extra Risks for Women and Marginalized Groups. *Health and Human Rights Journal* https://www.hhrjournal.org/2020/04/contact-tracing-apps-extra-risks-for-women-and-marginalized-groups/ Accessed 02/11/2021

DeJong, C., Chen, A. H., & Lo, B. (2020). An ethical framework for allocating scarce inpatient medications for COVID-19 in the U.S. *Jama*, 323(23), 2367-2368.

Dembosky, A. (2020). California may consider "Historical Injustice" when allocating COVID-19 Vaccine, Shots: *Health News From NPR* https://www.npr.org/sections/health-shots/2020/12/16/942452184/california-may-consider-historical-injustice-when-allocating-coronavirus-vaccine Accessed 12/24/2020

DHS (2012) The Menlo Report Ethical Principles Guiding Information and Communication Technology Research. Department of Homeland Security. Science and Technology https://www.dhs.gov/sites/default/files/publications/CSD-MenloPrinciplesCORE-20120803_1.pdf Accessed 04/20/2021

DiMasi, J. A., Grabowski, H. G., & Hansen, R. W. (2016). Innovation in the pharmaceutical industry: New estimates of R&D costs. *Journal of health economics*, 47, 20-33.

Dobrea, D. M., & Dobrea, M. C. (2020). An autonomous UAV system for video monitoring of the quarantine zones. *Roman J Inf Sci Technol*, 23, S53.

Eames, K. T., & Keeling, M. J. (2003). Contact tracing and disease control. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, *270*(1533), 2565-2571.

Eberly, L., Richter, D., Comerci, G., Ocksrider, J., Mercer, D., Mlady, G., … & Schenck, R. (2018). Psychosocial and demographic factors influencing pain scores of patients with knee osteoarthritis. *PloS one*, *13*(4), e0195075.

Eckhardt, R. (1987). Stan Ulman, John Von Neumann, and the Monte Carlo Method. *Los Alamos Science* Special Issue 1987 p131 http://www-star.st-and.ac.uk/~kw25/teaching/mcrt/MC_history_3.pdf Accessed 03/27/2021

Eli Lilly (2020). *Baricitinib has Significant Effect on Recovery Time, Most Impactful in COVID-19 Patients Requiring Oxygen. Further results from NIAID-sponsored ACTT-2 Trial. Results presented at the special International Society for Influenza and other Respiratory Virus Diseases Antiviral Group Virtual Conference on 'Therapeutics for COVID-19.* [Press Release] https://investor.lilly.com/news-releases/news-release-details/baricitinib-has-significant-effect-recovery-time-most-impactful Accessed 10/30/2020

ELLIS (2020). European Laboratory for Learning and Intelligent Systems Society (ELLIS) against COVID-19 https://ellis.eu/covid-19/events/ellis-against-covid-19-01-04-2020 Accessed 04/01/2020

Erikson, S. L. (2018). Cell phones ≠ self and other problems with big data detection and containment during epidemics. *Med. anthropology quarterly* 32, 315–339.

Espinoza, J., Crown, K., & Kulkarni, O. (2020). A Guide to Chatbots for COVID-19 Screening at Pediatric Health Care Facilities. *JMIR Public Health and Surveillance*, 6(2), e18808.

Etchemendy, J, et al. (2020). COVID-19 and AI: A Virtual Conference. https://hai.stanford.edu/agenda-covid-19-and-ai-virtual-conference Accessed 04/01/2020

Etzioni, O. (2020). Fighting COVID-19 with Open Access and AI, https://towardsdatascience.com/fighting-covid-19-with-open-access-and-ai-9a4df3cbe8c0 Accessed 05/10/2020

Fairchild, A.M., Gostin, L. O., & Bayer, R. (2020). Contact tracing's long, turbulent history holds lessons for COVID-19. *The Conversation*, https://theconversation.com/contact-tracings-long-turbulent-history-holds-lessons-for-covid-19-142511 Accessed 10/20/2020

Farzan, A.N. (2021). Philadelphia let 'college kids' distribute vaccines. The result was a 'disaster,' volunteers say. *Washington Post*, January 27, 2021. https://www.washingtonpost.com/nation/2021/01/27/philly-fighting-covid-vaccine/ Accessed 01/27/2021

FDA (2015). The Beginnings: Laboratory and Animal Studies, https://www.fda.gov/drugs/drug-information-consumers/beginnings-laboratory-and-animal-studies Accessed 10/31/2020

FDA (2019). Policy for Device Software Functions and Mobile Medical Applications Guidance for Industry and Food and Drug Administration Staff, https://www.fda.gov/media/80958/download Accessed 02/20/2021

FDA (2021). Artificial Intelligence/Machine Learning (AI/ML)-Based.: Software as a Medical Device (SaMD) Action Plan https://www.fda.gov/media/145022/download Accessed 02/21/2021

Fisher, C. (2019). Use of artificial intelligence within drug development programs. Workshop on Promoting Effective Drug Development Programs: Opportunities and Priorities for FDA's Office of New Drugs, November 7, 2019 https://www.fda.gov/media/132612/download Accessed 10/30/2020

Fjeld, J., Achten, N., Hilligoss, H., Nagy, A., & Skrikumar, M. (2020). Principled Artificial Intelligence. Mapping consensus in ethical and rights-based approaches to principles for AI https://wilkins.law.harvard.edu/misc/PrincipledAI_FinalGraphic.jpg Accessed 01/30/2020

Flores, A. W., Bechtel, K., & Lowenkamp, C. T. (2016). False positives, false negatives, and false analyses: A rejoinder to machine bias: There's software used across the country to predict future criminals. and it's biased against blacks. *Fed. Probation*, 80, 38.

Gasser, U., Ienca, M., Scheibner, J., Sleigh, J., & Vayena, E. (2020). Digital tools against COVID-19: Framing the ethical challenges and how to address them. *arXiv preprint* arXiv:2004.10236.

Gitlin, J. M. (2020). More than 7 in 10 American won't use contact tracing apps, data shows. *Ars Technica*, 06/15/2020. https://arstechnica.com/science/2020/06/more-than-7-in-10-americans-dont-want-contact-tracing-data-shows/ Accessed 12/20/2020

Golinelli, D., Boetto, E., Carullo, G., Nuzzolese, A. G., Landini, M. P., & Fantini, M. P. (2020). Adoption of Digital Technologies in Health Care During the COVID-19 Pandemic: Systematic Review of Early Scientific Literature. *Journal of medical Internet research*, 22(11), e22280.

Goodman, K. W., Cushman, R., & Miller, R. A. (2014). Ethics in biomedical and health informatics: Users, standards, and outcomes. In *Biomedical informatics* (pp. 329-353). Springer, London.

Grote, T. & Berens, P. (2020). On the ethics of algorithmic decision-making in healthcare *Journal of Medical Ethics*; 46**:**205-211

Guan, W. J., Ni, Z. Y., Hu, Y., Liang, W. H., Ou, C. Q., He, J. X., … & Du, B. (2020). Clinical characteristics of coronavirus disease 2019 in China. *New England journal of medicine*, 382(18), 1708-1720.

Guo, E., & Hao, K. (2020). This is Stanford vaccine algorithm that left out frontline doctors. *Technology Review* https://www.technologyreview.com/2020/12/21/1015303/stanford-vaccine-algorithm/ Accessed 02/10/2021

Hagendorff, T. (2020). The ethics of AI ethics: An evaluation of guidelines. *Minds and Machines*, 30(1), 99-120.

Hao, X., Cheng, S., Wu, D., Wu, T., Lin, X., & Wang, C. (2020). Reconstruction of the full transmission dynamics of COVID-19 in Wuhan. *Nature*, *584*(7821), 420-424.

Har, J. & Taxin, A. (2021). California uses ZIP codes, outreach to boost vaccine equity. *APNew*s. 02/08/2021 https://apnews.com/article/california-zip-codes-vaccine-equity-89caad537f1cf9507102afcf4ddec0d7 Accessed 02/09/2021

Harding, A., Harper, B., Stone, D., O'Neill, C., Berger, P., Harris, S., & Donatuto, J. (2012). Conducting research with tribal communities: Sovereignty, ethics, and data-sharing issues. *Environmental health perspectives*, *120*(1), 6-10.

Hernández-Orallo, E., Calafate, C. T., Cano, J. C., & Manzoni, P. (2020). Evaluating the effectiveness of COVID-19 Bluetooth-Based smartphone contact tracing applications. *Applied Sciences*, *10*(20), 7113.

HHS (1979) The Belmont Report. Ethical Principles and Guidelines for the Protection of Human Subjects of Research. Department of Health, Education, and Welfare.

https://www.hhs.gov/ohrp/regulations-and-policy/belmont-report/read-the-belmont-report/index.html Accessed 04/21/2021

Ho, D. (2020). Addressing COVID-19 Drug Development with Artificial Intelligence. *Advanced Intelligent Systems*, 2(5), 2000070.

Hu, Y., Jacob, J., Parker, G. J., Hawkes, D. J., Hurst, J. R., & Stoyanov, D. (2020). The challenges of deploying artificial intelligence models in a rapidly evolving pandemic. *Nature Machine Intelligence*, 2(6), 298-300.

Hu, Z., Ge, Q., Jin, L., & Xiong, M. (2020). Artificial intelligence forecasting of covid-19 in China. *arXiv preprint* arXiv:2002.07112.

Hua, J., & Shaw, R. (2020). Corona virus (Covid-19) "infodemic" and emerging issues through a data lens: The case of china. *International journal of environmental research and public health*, 17(7), 2309.

Hui, M. (2020). Hong-Kong is using tracker wristbands to geofence people under coronavirus quarantine https://finance.yahoo.com/news/hong-kong-using-tracker-wristbands-111714176.html Accessed 10/20/2020

International Covenant on Economic, Social and Cultural Rights (ICRSCR) (1966). 993 U.N.T.S. 3, G.A. Res. 2200 (XXI), 21 U.N. GAOR Supp. (No. 16 at 49, U.N. Doc. A/6316), entered into force January 3, 1976.

Imran, A., Posokhova, I., Qureshi, H. N., Masood, U., Riaz, M. S., Ali, K., … & Nabeel, M. (2020). AI4COVID-19: AI enabled preliminary diagnosis for COVID-19 from cough samples via an app. *Informatics in Medicine Unlocked*, *20*, 100378.

Iserson, K. V., & Moskop, J. C. (2007). Triage in medicine, part I: concept, history, and types. *Annals of emergency medicine*, 49(3), 275-281.

Islam, M. M., Islam, M. Z., Asraf, A., & Ding, W. (2020). Diagnosis of COVID-19 from X-rays Using Combined CNN-RNN Architecture with Transfer Learning. 2020.08.24.20181339. https://doi.org/10.1101/2020.08.24.20181339

Israeli Health Ministry (2021). Real-world epidemiological evidence collaboration agreement, https://govextra.gov.il/media/30806/11221-moh-pfizer-collaboration-agreement-redacted.pdf Accessed 02/21/2021

Jamshidi, M., Lalbakhsh, A., Talla, J., Peroutka, Z., Hadjilooei, F., Lalbakhsh, P., & Sabet, A. (2020). Artificial Intelligence and COVID-19: Deep Learning Approaches for Diagnosis and Treatment. *IEEE Access*, 8, 109581-109595.

Jeon, J. (2020). AI Technology to fight against Coronavirus (COVID-19), https://github.com/hollobit/COVID-19-AI Accessed 06/10/2020

Judson, T. J., Odisho, A. Y., Young, J. J., Bigazzi, O., Steuer, D., Gonzales, R., & Neinstein, A. B. (2020). Case Report: Implementation of a Digital Chatbot to Screen Health System Employees during the COVID-19 Pandemic. *Journal of the American Medical Informatics Association*.

Khan, J. P. (Ed.). (2020). Digital contact tracing for pandemic response: Ethics and governance guidance*. Johns Hopkins University Press.*

Kasparov, G. (2017). Deep thinking: where machine intelligence ends, and human creativity begins. Hachette UK.

Kofler, N. & Baylis, F. (2020). Ten reasons why immunity passports are a bad idea. *Nature* 581, 379-381 (2020)
https://www.nature.com/articles/d41586-020-01451-0 Accessed 01/29/2021

Koonin, L. M., Hoots, B., Tsang, C. A., Leroy, Z., Farris, K., Jolly, B., … & Harris, A. M. (2020). Trends in the use of telehealth during the emergence of the COVID-19 pandemic—United States, January–March 2020. *Morbidity and Mortality Weekly Report*, *69*(43), 1595.

Lalmuanawma, S., Hussain, J., & Chhakchhuak, L., (2020). Applications of machine learning and artificial intelligence for Covid-19 (SARS-CoV-2) pandemic: A review. *Chaos, Solitons & Fractals, Volume 139, Elsevier.*
https://www.sciencedirect.com/science/article/pii/S096007792030456 2?via%3Dihub. Accessed 07/15/2020

Landau, S. (2021) People Count: Contact-tracing apps and public health. The MIT Press. Cambridge Massachusetts. ISBN 9780262045711

Latif, S., Usman, M., Manzoor, S., Iqbal, W., Qadir, J., Tyson, G., Castro, I., Razi, A., Boulos, M. N. K., Weller, A., & Crowcroft, J. (2020). Leveraging Data Science To Combat COVID-19: A Comprehensive Review. *IEEE Transactions on Artificial Intelligence*, PP(99), 1–1.
https://doi.org/DOI 10.1109/TAI.2020.3020521

Leith, D & Farrell, S. J. (2020). Measurement-based Evaluation of Google/Apple Exposure Notification API for Proximity Detection in a Light-rail Tram. *PLoS One 5, no. 9*

Li, J., & Guo, X. (2020). COVID-19 Contact-tracing Apps: A Survey on the Global Deployment and Challenges. *arXiv preprint* arXiv:2005.03599.

Li, R. (2020). Machine Learning Enabled Systems for Delivering Care to Critically Ill Patients. Stanford HAI “Covid-19 and AI : A Virtual Conference“ Session IV.
https://hai.stanford.edu/video-archive-covid-19-and-ai-virtual-conference
Talk starts at 17:03 (session 4); Accessed 12/01/2020

Li, R., Pei, S., Chen, B., Song, Y., Zhang, T., Yang, W., & Shaman., J. (2020). Substantial undocumented infection facilitates the rapid dissemination of novel coronavirus (SARS-CoV2). *Science* ; eabb3221 doi: 10.1126/science.abb3221

Liu, F., Li, X., & Zhu, G. (2020). Using the contact network model and Metropolis-Hastings sampling to reconstruct the COVID-19 spread on the "Diamond Princess." *Science Bulletin.*

Lopreite, M., Panzarasa, P., Puliga, M., & Riccaboni, M. (2021). Early warnings of COVID-19 outbreaks across Europe from social media. *Scientific reports*, *11*(1), 1-7.

Macmillan, J. and Norman, J., (2021). Australians to have proof of their coronavirus vaccination on their smartphone, ABC News February 07, 2021, https://www.abc.net.au/news/2021-02-07/covid-19-vaccine-certificates-to-be-on-australians-smartphones/13130350

Mario, X. (2020). Can Drones take your temperature? *Drone Tech Planet* https://www.dronetechplanet.com/can-drones-take-your-temperature/ Accessed 10/02/2020

Martin, A., Nateqi, J., Gruarin, S., Munsch, N., Abdarahmane, I., Zobel, M., & Knapp, B. (2020). An artificial intelligence-based first-line defense against COVID-19: digitally screening citizens for risks via a chatbot. *Scientific reports*, 10(1), 1-7.

Michael, K., & Abbas, R. (2020). Behind COVID-19 contact trace apps: the Google–Apple partnership. *IEEE Consumer Electronics Magazine*, 9(5), 71-76.

Microsoft (2020). Understanding our progress against COVID-19 https://www.microsoft.com/en-us/ai/ai-for-health-covid-data Accessed 10/06/2020

Miller, R. A., & Gardner, R. M. (1997a). Summary recommendations for responsible monitoring and regulation of clinical software systems. *Annals of Internal Medicine*, *127*(9), 842-845.

Miller, R. A., & Gardner, R. M. (1997b). Recommendations for responsible monitoring and regulation of clinical software systems. *Journal of the American Medical Informatics Association*, *4*(6), 442-457.

Miner, A. S., Laranjo, L., & Kocaballi, A. B. (2020). Chatbots in the fight against the COVID-19 pandemic. *Digital Medicine*, 3(1), 1-4.

Moore et al. (2019). Artificial Intelligence: An Evangelical Statement of Principles, Ethics and Religious Liberty Commission of the Southern Baptist Convention. https://erlc.com/resource-library/statements/artificial-intelligence-an-evangelical-statement-of-principles/ Accessed 02/04/2021

Murphy, R. R., Gandudi, V. B. M., & Adams, J. (2020). Applications of robots for COVID-19 response. *arXiv preprint arXiv:2008.06976*. https://arxiv.org/pdf/2008.06976.pdf Accessed 04/05/2021

National Academies of Sciences (2020). Framework for Equitable Allocation of COVID-19 Vaccine. Consensus Study Report; https://www.nap.edu/catalog/25917/framework-for-equitable-allocation-of-covid-19-vaccine Accessed 12/25/2020

National Catholic Bioethics Center (2020). Ethical concerns with COVID-19 triage protocols. *Ethics & Medics*, 45(5), 5-6.

Naudé, W. (2020). Artificial intelligence vs COVID-19: limitations, constraints and pitfalls. *AI & Society*, 1.

Nebeker, C., Bartlett Ellis, R. J., & Torous, J. (2020). Development of a decision-making checklist tool to support technology selection in digital health research. *Translational behavioral medicine*, *10*(4), 1004-1015. See framework at: https://recode.health/tools/ Accessed 04/20/2021

Neiiler, E., (2020). An AI Epidemiologist Sent the First Warnings of the Wuhan Virus, *Wired Magazine*, https://www.wired.com/story/ai-epidemiologist-wuhan-public-health-warnings/. Accessed 07/20/2020

Nguyen, C. T., Saputra, Y. M., Van Huynh, N., Nguyen, N. T., Khoa, T. V., Tuan, B. M., … & Ottersten, B. (2020). Enabling and Emerging Technologies for Social Distancing: A Comprehensive Survey. *arXiv preprint* arXiv:2005.02816. https://arxiv.org/abs/2005.02816

Nguyen, D., Ding, M., Pathirana, P. N., & Seneviratne, A. (2020). Blockchain and AI-based Solutions to Combat Coronavirus (COVID-19)-like Epidemics: A Survey. https://www.techrxiv.org/articles/Blockchain_and_AI-based_Solutions_to_Combat_Coronavirus_COVID-19_-like_Epidemics_A_Survey/12121962/1 Accessed 08/18/2020

Nguyen, T. T. (2020). Artificial Intelligence in the Battle against Coronavirus (COVID-19): A Survey and Future Research Directions. *TechRxiv. Preprint*. https://doi.org/10.36227/techrxiv.12743933.v1 https://www.techrxiv.org/articles/Artificial_Intelligence_in_the_Battle_against_Coronavirus_COVID-19_A_Survey_and_Future_Research_Directions/12743933/1 Accessed 08/20/2020

Obermeyer, Z., Powers, B., Vogeli, C., Mullainathan, S. (2019). Dissecting racial bias in an algorithm used to manage the health of populations, *Science* 25 Oct 2019: Vol. 366, Issue 6464, pp. 447-453 DOI: 10.1126/science.aax 2342

Pacheco, C. M., Daley, S. M., Brown, T., Filippi, M., Greiner, K. A., & Daley, C. M. (2013). Moving forward: breaking the cycle of mistrust between American Indians and researchers. *American journal of public health*, *103*(12), 2152-2159.

Pal, R., Sekh, A. A., Kar, S., & Prasad, D. K. (2020). Neural network-based country wise risk prediction of COVID-19. *arXiv preprint* arXiv:2004.00959. Accessed 09/01/2020

Pandit, C., Kothari, H., & Neuman, C. (2020). Privacy in time of a pandemic. In *2020 13th CMI Conference on Cybersecurity and Privacy (CMI)-Digital Transformation-Potentials and Challenges (51275)* (pp. 1-6). IEEE.

Parker, M. J., Fraser, C., Abeler-Dörner, L., & Bonsall, D. (2020). Ethics of instantaneous contact tracing using mobile phone apps in the control of the COVID-19 pandemic. Journal of Medical Ethics, 46(7), 427-431.

Passi, S., & Barocas, S. (2019). Problem formulation and fairness. In *Proceedings of the Conference on Fairness, Accountability, and Transparency* (pp. 39-48).

Patel, S. (2020). ACM Keynote: New Ways of Thinking of the Mobile Phone for Healthcare and the current Pandemic. https://aiforgood.itu.int/events/acm-keynote-new-ways-of-thinking-of-the-mobile-phone-for-healthcare-and-the-current-pandemic/ (cough is mentioned at 0:36:20)

Peel, M. (2005). Human Rights and Medical Ethics. *Journal of the Royal Society of Medicine* 98 (2005): 171-173.

Pew Research Center, (2019). Demographics of Mobile Device Ownership and Adoption in the United States. https://www.pewresearch.org/internet/fact-sheet/mobile/ Accessed 01/20/2021

Pham, Q., Nguyen, D. C., Huynh-The, T., Hwang, W., & Pathirana, P. (2020) Artificial Intelligence (AI) and Big Data for Coronavirus (COVID-19) Pandemic: A Survey on the State-of-the-Arts. in *IEEE Access*, vol. 8, pp. 130820-130839, 2020, doi: 10.1109/ACCESS.2020.3009328. https://ieeexplore.ieee.org/document/9141265 Accessed 08/01/2020

Pierson, E., Cutler, D. M., Leskovec, J., Mullainathan, S., & Obermeyer, Z. (2021). An algorithmic approach to reducing unexplained pain disparities in underserved populations. *Nature* Medicine, 27(1), 136-140.

Pleiss, G., Raghavan, M., Wu, F., Kleinberg, J., & Weinberger, K. Q. (2017). On fairness and calibration. In *Advances in Neural Information Processing Systems* (pp. 5680-5689).

Plomer, A. (2009). Human Dignity, Human Rights, and Article 6 (1) of the Biotech Directive in Embryonic Stem Cell Patents: European Law and Ethics eds, Aurora Plomer et al. (New York: Oxford University Press), 202-226.

Price, W. N. (2018). Big data and black-box medical algorithms. *Science translational medicine*, *10* (471).

Price, W. N., Gerke, S., & Cohen, I. G. (2021) How much can potential jurors tell us about liability for medical artificial intelligence? *The Journal of Nuclear Medicine* 62. 15.

Privacy International, (2020). The looming disaster of immunity passports and digital identity, Newsletter https://privacyinternational.org/long-read/4074/looming-disaster-immunity-passports-and-digital-identity Accessed 02/10/2021

ProtectYu Press Release (2020). Protect-Yu is a NEW innovative Geofencing technology for the control, tracking and monitoring of COVID-19 quarantined patients, populations or persons

under mandatory confinement measures. https://markets.businessinsider.com/news/stocks/protect-yu-is-a-new-innovative-geofencing-technology-for-the-control-tracking-and-monitoring-of-covid-19-quarantined-patients-populations-or-persons-under-mandatory-confinement-measures-1029041038?op=1 Accessed 12/25/2020

Qiang, X. (2019). The Road to Digital Unfreedom: President Xi's Surveillance State. *Journal of Democracy* 30(1), 53-67. doi:10.1353/jod.2019.0004

Ramamoorthy, A., Pacanowski, M. A., Bull, J., & Zhang, L. (2015). Racial/ethnic differences in drug disposition and response: review of recently approved drugs. *Clinical Pharmacology & Therapeutics*, 97(3), 263-273.

Raskar, R., Schunemann, I., Barbar, R., Vilcans, K., Gray, J., Vepakomma, P., … & Werner, J. (2020). Apps gone rogue: Maintaining personal privacy in an epidemic. *arXiv preprint arXiv:2003.08567*.

Richardson, P., Griffin, I., Tucker, C., Smith, D., Oechsle, O., Phelan, A., & Stebbing, J. (2020). Baricitinib as potential treatment for 2019-nCoV acute respiratory disease. *Lancet* (London, England), 395(10223), e30.

Ricks, B. et al. (2020). Creating trustworthy AI. A Mozilla white paper on challenges and opportunities in the AI era. https://foundation.mozilla.org/en/insights/trustworthy-ai-whitepaper/ Accessed 01/20/2021

Rivero, N. (2020). The pandemic is automating emergency room triage. *Quartz*, August 21, 2020. https://qz.com/1894714/covid-19-is-boosting-the-use-of-ai-triage-in-emergency-rooms/ Accessed 09/10/2020

Rivest, R. L., Callas, J., Canetti, R., Esvelt, K., Gillmor, D. K., Kalai, Y. T., … & Shen, E. (2020). The PACT protocol specification. Private Automated Contact Tracing Team, MIT, Cambridge, MA, USA, *Tech. Rep*. 0.1.

Rodolfa, K. T., Saleiro, P., and Ghani, R. (2019). Bias and Fairness (Chapter 11) in *Big Data and Social Science* https://textbook.coleridgeinitiative.org/chap-bias.html

Rome Call for AI Ethics (2020). https://romecall.org/ Accessed 12/10/2020

Rowe, F., Ngwenyama, O., & Richet, J. L. (2020). Contact-tracing apps and alienation in the age of COVID-19. *European Journal of Information Systems*, 29(5), 545-562.

Russo, G., Reche, P., Pennisi, M., & Pappalardo, F. (2020). The combination of artificial intelligence and systems biology for intelligent vaccine design. Expert Opinion on Drug Discovery, 1-15.

Sanderson, S. C., Brothers, K. B., Mercaldo, N. D., Clayton, E. W., Antommaria, A. H. M., Aufox, S. A., … & Holm, I. A. (2017). Public attitudes toward consent and data sharing in biobank research: a large multi-site experimental survey in the U.S. *The American Journal of Human Genetics*, 100(3), 414-427.

Sapiezynski, P., Pruessing, J., & Sekara, V. (2020). The Fallibility of Contact-Tracing Apps. *arXiv preprint arXiv:2005.11297*.
Sarwar, S., Dent, A., Faust, K., Richer, M., Djuric, U., Van Ommeren, R., & Diamandis, P. (2019). Physician perspectives on integration of artificial intelligence into diagnostic pathology. *NPJ digital medicine*, *2*(1), 1-7.

Schabas, W. A. (2007). Study of the Right to Enjoy the Benefits of Scientific and Technological Progress and Its Applications. *In Human Rights in Education, Science and Culture: Legal Developments and Challenge*s, eds. Yvonne Donders et al (New York: UNESCO, 2007), 285.

Sekara, V., & Lehmann, S. (2014). The strength of friendship ties in proximity sensor data. *PloS one*, *9*(7), e100915.

Shackford, S. (2020). Community Anger Shuts Down Connecticut City's Plan To Use Drones for Coronavirus Monitoring, https://reason.com/2020/04/24/community-anger-shuts-down-connecticut-citys-plan-to-use-drones-for-coronavirus-monitoring/ Accessed 10/20/2020

Shapiro, J. (2020). Oregon Hospitals Didn't Have Shortages. So Why Were Disabled People Denied Care? *NPR* December 2020. https://www.npr.org/2020/12/21/946292119/oregon-hospitals-didnt-have-shortages-so-why-were-disabled-people-denied-care Accessed 01/04/2021

Shaw, J. A., Sethi, N., & Block, B. L. (2020). Five things every clinician should know about AI ethics in intensive care. *Intensive Care Med* 47**,** 157–159. https://doi.org/10.1007/s00134-020-06277-y

Simonite, T. (2020). AI uncovers a Potential Treatment for Covid-19 Patients. *Wired Magazine*, 04.17.2020 issue https://www.wired.com/story/ai-uncovers-potential-treatment-covid-19-patients/ Accessed 10/30/2020

Simko, L., Chang, J. L., Jiang, M., Calo, R., Roesner, F., & Kohno, T. (2020). COVID-19 Contact Tracing and Privacy: A Longitudinal Study of Public Opinion. *arXiv preprint* arXiv:2012.01553. https://arxiv.org/abs/2012.01553

Sinha, A., Bernardes, E., Calderon, R., & Wuest, T. (2020). Digital Supply Networks: Transform Your Supply Chain and Gain Competitive Advantage with Disruptive Technology and Reimagined Processes. McGraw Hill Professional.

Singh, K., Valley, T. S., Tang, S., Li, B. Y., Kamran, F., Sjoding, M. W., … & Nallamothu, B. K. (2020). Evaluating a Widely Implemented Proprietary Deterioration Index Model Among Hospitalized COVID-19 Patients. *Annals of the American Thoracic Society*.

Skloot, R. (2010). The immortal life of Henrietta Lacks. *Broadway Books*.

Smalley, E. (2017). AI-powered drug discovery captures pharma interest. *Nat Biotechnol* 35, 604–605. https://doi.org/10.1038/nbt0717-604

Social Science Research Council (2020). Public Health, Surveillance, and Human Right Networks. https://covid19research.ssrc.org/public-health-surveillance-and-human-rights-network/ Accessed 02/10/2021

The Royal Society (2021). Twelve criteria for the development and use of COVID-19 vaccine passports. https://royalsociety.org/-/media/policy/projects/set-c/set-c-vaccine-passports.pdf Accessed 03/05/2021

Tolchin, B., Hull, S. C., & Kraschel, K. (2020). Triage and justice in an unjust pandemic: ethical allocation of scarce medical resources in the setting of racial and socioeconomic disparities. *Journal of medical ethics*.

Tolchin, B., Latham, S. R., Bruce, L., Ferrante, L. E., Kraschel, K., Jubanyik, K., … & Mercurio, M. R. (2020). Developing a Triage Protocol for the COVID-19 Pandemic: Allocating Scarce Medical Resources in a Public Health Emergency. *The Journal of clinical ethics*, *31*(4), 303-317.

Turner, J. (2018). *Robot rules: Regulating artificial intelligence*. Springer.

UNESCO (2009). Venice Statement on the Right to Enjoy the Benefits of Scientific Progress and its Applications, July 16- 17, 2009, s. 12(f); https://unesdoc.unesco.org/ark:/48223/pf0000185558 Accessed 03/20/2021

University of California Irving (UCI), School of Pharmacy & Pharmaceutical Sciences, A Short History of Drug Discovery, https://pharmsci.uci.edu/about/a-short-history-of-drug-discovery/ Accessed 10/15/2020

Vaishya, R., Javaid, M., Khan, I., Haleem, A. (2020). Artificial Intelligence (AI) applications for COVID-19 pandemic, Diabetes & Metabolic Syndrome: *Clinical Research & Reviews*, Volume 14, Issue 4, July–August 2020, Pages 337-339 doi: https://doi.org/10.1016/J.DSX.2020.04.012

Venugopal, A., Ganesan, H., Raja, S. S. S., Govindasamy, V., Arunachalam, M., Narayanasamy, A., … & Vellingiri, B. (2020). Novel wastewater surveillance strategy for early detection of COVID–19 hotspots. *Current opinion in environmental science & health*.

VolppKevin, G. (2020). Asked and answered: Building a chatbot to address covid-19-related concerns. NEJM Catalyst Innovations in Care Delivery.

Voo, T. C., Reis, A. A., Thomé, B., Ho, C. W., Tam, C. C., Kelly-Cirino, C., … & Munsaka, S. (2021). Immunity certification for COVID-19: ethical considerations. *Bulletin of the World Health Organization*, *99*(2), 155.

Wadman, M. (2020) Abortion opponents protest COVID-19 vaccines' use of fetal cells, *Science*, June 5, 2020 https://www.sciencemag.org/news/2020/06/abortion-opponents-protest-covid-19-vaccines-use-fetal-cells Accessed 03/05/2021

Wang, T., & Jia, F. (2020). The impact of health QR code system on older people in China during the COVID-19 outbreak. *Age and Ageing*.

Wendler, D., Kington, R., Madans, J., Van Wye, G., Christ-Schmidt, H., Pratt, L. A., … & Emanuel, E. (2005). Are racial and ethnic minorities less willing to participate in health research? *PLoS Med*, *3*(2), e19.

Wesner, K. L. (2019). Is the Grass Greener on the Other Side of the Geofence: The First Amendment and Privacy Implications of Unauthorized Smartphone Messages. *Case W. Res. JL Tech. & Internet*, *10*, iii.

Whitelaw, S., Mamas, M., Topol, E., Van Spall, H. (2020). Applications of digital technology in COVID-19 pandemic planning and response. *The Lancet Digital Health*. Volume 2, Issue 8, E435-440
https://www.thelancet.com/journals/landig/article/PIIS2589-7500(20)30142-4/fulltext Accessed 07/14/2020

WHO-a, (2020). Modes of transmission of virus causing COVID-19. Scientific Brief https://apps.who.int/iris/bitstream/handle/10665/331601/WHO-2019-nCoV-Sci_Brief-Transmission_modes-2020.1-eng.pdf Accessed 02/11/2021

WHO-b, (2020). "Immunity passports" in the context of COVID-19, Scientific Brief https://www.who.int/news-room/commentaries/detail/immunity-passports-in-the-context-of-covid-19 Accessed 02/12/2021

WHO/ITU, (2020). FG-AI4H Working Group & Ah-hoc Group ToRs https://www.itu.int/en/ITU-T/focusgroups/ai4h/Pages/wg.aspx Accessed 09/01/2020

Wuest, T., Kusiak, A., Dai, T., & Tayur, S. R. (2020). Impact of COVID-19 on Manufacturing and Supply Networks—The Case for AI-Inspired Digital Transformation. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3593540 Accessed 11/30/2020

Xu, R., Joshi, J. B., & Li, C. (2019). Cryptonn: Training neural networks over encrypted data. In *2019 IEEE 39th International Conference on Distributed Computing Systems (ICDCS)* (pp. 1199-1209). IEEE.

Yakabuski, K. (2020). Why would we trust China's official COVID-19 numbers? https://www.theglobeandmail.com/opinion/article-why-would-we-trust-chinas-official-covid-19-numbers/ Accessed 02/10/2021

Yang, T., Sha, L., Li, J., & Hong, P. (2020). A Deep Learning Approach for COVID-19 Trend Prediction. *arXiv preprint* arXiv:2008.05644.

You, J. (2020-a). Lessons from South Korea's Covid-19 policy response. *The American Review of Public Administration*, 50(6-7), 801-808.

You, T. (2020-b). China uses drones with thermal cameras to check quarantined residents' temperatures and drop face masks in bid to control coronavirus outbreak, https://www.dailymail.co.uk/news/article-7974935/China-uses-drones-THERMAL-CAMERAS-check-quarantined-residents-temperatures.html Accessed 10/20/2020

Zech, J. R., Badgeley, M. A., Liu, M., Costa, A. B., Titano, J. J., & Oermann, E. K. (2018). Variable generalization performance of a deep learning model to detect pneumonia in chest radiographs: a cross-sectional study. *PLoS medicine*, *15*(11), e1002683.

Zhou, M., Chen, Y., Wang, D., Xu, Y., Yao, W., Huang, J., … & Qu, J. (2020). Improved deep learning model for differentiating novel coronavirus pneumonia and influenza pneumonia. *medRxiv*. 2020.03.24.20043117; https://doi.org/10.1101/2020.03.24.20043117