|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| ITU Logo | INTERNATIONAL TELECOMMUNICATION UNION  **TELECOMMUNICATION STANDARDIZATION SECTOR**  STUDY PERIOD 2017-2020 | | FG-AI4H-K-028 | |
| **ITU-T Focus Group on AI for Health** | |
| **Original: English** | |
| **WG(s):** | | WG-Ethics | E-meeting, 27-29 January 2021 | |
| **DOCUMENT** | | | | |
| **Source:** | | WHO | | |
| **Title:** | | Draft Guidance document: Ethics and governance of artificial intelligence for health | | |
| **Purpose:** | | Discussion | | |
| **Contact:** | | Andreas Reis Chair, FG-AI4H WG-Ethics WHO | | Email: [reisa@who.int](mailto:reisa@who.int) |
| **Contact:** | | Sameer Pujari Vice chair, FG-AI4H WHO | | Email: [pujaris@who.int](mailto:pujaris@who.int) |

|  |  |
| --- | --- |
| **Abstract:** | This document contains a draft guidance document prepared by WHO on the ethics and governance of artificial intelligence for health. It addresses the key topics in 10 sections and contains three checklists addressing implementors, ministries of health and health care providers.  In line with the discussions held at Meeting J, this draft which is provided for comments by the FG-AI4H members. **In particular**, for comments from AI programmers/designers from the FG-AI4H on the related checklist in Annex I (pages 104-109). Also, TG Drivers are invited to try to apply the principles and checklists in this document (particularly, Annex I) to their TDDs and provide feedback to the authors. An early feedback by **30 November 2020** would be extremely appreciated. |

**Draft Guidance document:  
Ethics and governance of artificial intelligence for health**

**Table of contents**

| Page |
| --- |
| [Section 1: Introduction 5](#_Toc56017556)  [Section 2: Describing Artificial Intelligence in health and medicine 7](#_Toc56017557)  [Section 3: Applications of Artificial Intelligence in Health 9](#_Toc56017558)  [Clinical care 10](#_Toc56017559)  [The evolving role of the patient in clinical care 11](#_Toc56017560)  [The shift from hospital to home-based care 11](#_Toc56017561)  [Application of AI in health research 12](#_Toc56017562)  [Use of AI for health systems management 12](#_Toc56017563)  [Public health informatics (uses of AI for public health) 12](#_Toc56017564)  [The future of AI in health 15](#_Toc56017565)  [Section 4: Ethical challenges for the use of Artificial Intelligence in Health Care 16](#_Toc56017566)  [Assessing when artificial intelligence should be used 16](#_Toc56017567)  [The use of artificial intelligence to achieve universal health coverage 17](#_Toc56017568)  [Artificial intelligence and the digital divide 20](#_Toc56017569)  [Data collection and use 20](#_Toc56017570)  [Accountability and responsibility for decision-making with AI 26](#_Toc56017571)  [Autonomous decision-making 29](#_Toc56017572)  [Implications of replacing human judgment 29](#_Toc56017573)  [Implications of the loss of human control 30](#_Toc56017574)  [AI, bias, and discrimination 33](#_Toc56017575)  [Biases in data 33](#_Toc56017576)  [Biases in deployment and use of AI 34](#_Toc56017577)  [Impacts of AI on labour and employment (in health and medicine) 35](#_Toc56017578)  [Challenges with the commercialisation of health care using AI 37](#_Toc56017579)  [Section 5: Risks associated with the widespread adoption of AI for health 40](#_Toc56017580)  [Safety of AI technologies 40](#_Toc56017581)  [Cybersecurity risks 41](#_Toc56017582)  [Use of AI for drug development 43](#_Toc56017583)  [Section 6: Existing governance frameworks that apply to AI in health 44](#_Toc56017584)  [AI and human rights 44](#_Toc56017585)  [General principles of AI 45](#_Toc56017586)  [AI principles for health 46](#_Toc56017587)  [Bioethics laws and policies 47](#_Toc56017588)  [Governance of health data 47](#_Toc56017589)  [Regulatory standards 47](#_Toc56017590)  [Section 7: Key ethical principles for the use of AI in health 48](#_Toc56017591)  [Human Rights 49](#_Toc56017592)  [Ethical Principles 50](#_Toc56017593)  [Section 8: Liability regimes for AI in health 55](#_Toc56017594)  [Liability for use of AI in clinical care 55](#_Toc56017595)  [Are machine learning algorithms products? 56](#_Toc56017596)  [Compensation for errors 57](#_Toc56017597)  [The role of regulatory agencies and pre-emption 57](#_Toc56017598)  [Considerations for LMICs 58](#_Toc56017599)  [Section 9: Elements of a governance framework for AI in health 60](#_Toc56017600)  [Governance of Data 60](#_Toc56017601)  [Evolving approaches to consent 60](#_Toc56017602)  [Principles and approaches for the protection, collection, and appropriate use of health data 63](#_Toc56017603)  [Data sovereignty 65](#_Toc56017604)  [Data cooperatives 66](#_Toc56017605)  [Data sharing 67](#_Toc56017606)  [Data hubs 67](#_Toc56017607)  [Federated data 68](#_Toc56017608)  [Data sharing and data partnerships with the private sector 68](#_Toc56017609)  [Control and Benefit Sharing 70](#_Toc56017610)  [Control over and benefit sharing of Big Data 70](#_Toc56017611)  [Ownership of AI-based products, services, and methods 72](#_Toc56017612)  [Governance of the private sector 74](#_Toc56017613)  [The role of self-governance 74](#_Toc56017614)  [Public-private partnerships for AI in health care 77](#_Toc56017615)  [Governance and oversight of large technology companies 77](#_Toc56017616)  [Provision of health care by the private sector outside of the healthcare system 79](#_Toc56017617)  [An enabling environment for effective governance of the private sector 79](#_Toc56017618)  [Governance of the public sector 81](#_Toc56017619)  [Assessing whether AI is necessary and appropriate for use by the public sector 81](#_Toc56017620)  [Accountability through transparency and participation 82](#_Toc56017621)  [Appropriate collection, stewardship, and use of data 82](#_Toc56017622)  [Risks and opportunities with the use of AI for provision of public services and social protection 83](#_Toc56017623)  [Regulatory considerations 84](#_Toc56017624)  [Does regulation interfere with innovation? 84](#_Toc56017625)  [Transparency and explainability of AI-based devices 85](#_Toc56017626)  [Addressing bias 86](#_Toc56017627)  [Ethics considerations for low- and middle-income countries and high-income countries with poor health outcomes 87](#_Toc56017628)  [International governance of AI 89](#_Toc56017629)  [Section 10. Building an ethical approach in practice for AI in health 91](#_Toc56017630)  [Ethical and transparent design of technologies 91](#_Toc56017631)  [Bills of rights 93](#_Toc56017632)  [Research agenda for the ethical use of artificial intelligence for healthcare 94](#_Toc56017633)  [Engagement and role of the public and building trust with providers and patients 96](#_Toc56017634)  [Impact Assessments 98](#_Toc56017635)  [Putting prediction to good use 101](#_Toc56017636)  [Policy observatory and model legislation 101](#_Toc56017637)  [References 103](#_Toc56017638)  [Annex I: Checklist for AI Developers 104](#_Toc56017639)  [Designing an AI technology 105](#_Toc56017640)  [Developing an AI technology 107](#_Toc56017641)  [Deploying and improving an AI technology following deployment 109](#_Toc56017642)  [Annex II: Checklist for Ministries of Health 110](#_Toc56017643)  [How should Ministries of Health prepare for the introduction and use of AI technologies? 111](#_Toc56017644)  [How should Ministries of Health address ethical challenges? 113](#_Toc56017645)  [How should Ministries of Health protect the health, safety, and human rights of patients? 114](#_Toc56017646)  [Annex III: Checklist for Health Care Providers (3rd November 2020) 116](#_Toc56017647)  [Is the AI technology appropriate? 117](#_Toc56017648)  [Is the context in which this AI technology will be used appropriate? 119](#_Toc56017649)  [Should a health care provider use the AI technology? 120](#_Toc56017650) |

**List of boxes**

| Page |
| --- |
| [Box X1: The growth of digital identification in the SARS-COV2 pandemic 23](#_Toc56018454)  [Box X2: Examples of AI principles proposed by inter-governmental organizations and countries 45](#_Toc56018455)  [Box X3: Design for Values 91](#_Toc56018456) |

**List of Figures**

| Page |
| --- |
| [Figure X1: Michigan Medicine Human Data and Biospecimen Release Committee Rubric 62](#_Toc56017843)  [Figure X2: Eight questions addressing building trust in data use 64](#_Toc56017844) |

# Section 1: Introduction

Digital technologies, machine learning, and artificial intelligence (AI) will likely transform medicine, medical research, and public health. Technologies that utilise AI are now deployed within health services in OECD countries and its utility is being assessed in low- and middle-income countries. While these technologies hold great promise, there are trans-national ethical, legal, commercial, and social concerns. AI also poses novel ethical challenges in software engineering that extend beyond the purview of traditional regulators and participants within health care systems. These ethical challenges must be negotiated if AI will improve human health, preserve human autonomy, and if equitable access to such technologies is to be ensured.

Optimistic views of AI predict rapid improvements to individual and global public health and assisting countries to achieve universal health coverage. This could include improving accessibility of health services for marginalized and vulnerable populations. AI could also improve public health surveillance, drug development, pharmacovigilance, forecasting, modelling, and monitoring, improve clinical outcomes, and optimize operations and health care provision.

Yet unchecked optimism of AI’s potential benefits could veer into a habit of looking first to technological solutions to address problems that do not lend itself to easy fixes. Such ‘techno-optimism’ could also make matters worse. For example, AI could exacerbate the unequal distribution of health care and widen differences in health care outcomes within and between countries. The benefits of technology are already distributed unevenly and unequally between and within wealthy and low-income countries. The digital divide could worsen inequitable access to health care, whether by geography, gender, age, access to devices, and connectivity.

AI could also perpetuate or exacerbate bias. Use of limited, low-quality, and heterogeneous data in AI could perpetuate and deepen prejudices and disparities in health care. Biased inferences, misleading data analyses and poorly designed health applications and tools could be harmful. It has been shown that commercial prediction algorithms can identify complex health needs, but they can also result in significant racial or ethnic bias. Some AI algorithms are based for example on images of uneven quality and resolution, which could adversely affect the accuracy and reliability of the systems’ output.

AI could present a singular opportunity to augment and improve the capabilities of over-stretched health care workers and providers. Yet AI could have significant negative impacts on the health care workforce – it could reduce the size of the workforce, limit, challenge, and degrade the existing skills of health workers, and oblige health workers to retrain to adapt to the use of AI. Centuries of medical practice are constructed on a carefully constructed relationship of provider and patient. What happens when AI technologies ‘disrupt’ this relationship?

Claims made on behalf of AI technologies will need to be validated, and procedures must be specified to ensure their use in practice replicates validated results. Personnel must be trained to follow such procedures, and adjustments must be made to clinical workflows to ensure that incorporating these systems in clinical practice does not result in unintended adverse consequences.

Though privacy, confidentiality and informed consent are worldwide pillars of patient rights, these rights could be dramatically redefined or undermined as digital technologies take hold and expand. The performance of AI depends (among other factors) on the nature and volume of data and information and the conditions under which such data was gathered. The pursuit of data, whether by government or companies, could undermine privacy and autonomy at the service of unmerited government surveillance or commercial profit. If privacy and autonomy are not assured, there can be a wider impact of the exercise of a full range of human rights including civil and political rights (such as freedom of movement and expression), and social and economic rights (such as access to health care and education).

AI technologies are likely to be designed by companies, although many governments have the capability to develop and deploy these technologies to strengthen health care provision. Some of the world’s largest technology companies are aggressively pursuing the development of new applications and services that these companies either own or invest in. Many of these companies already accumulate large quantities of data, including health data, and exercise significant power in society and the economy. While many of these companies may offer innovative approaches to addressing health challenges, there are significant concerns that such companies may eventually exercise too much power in relation to governments, providers, and patients with respect to health care.

AI technologies is also changing where healthcare is accessed. AI technologies for health are increasingly distributed outside of a regulated health care setting, including through the workplace, social media, and the education system. With the rapid proliferation and evolving uses of AI in health care, including in response to the SARS-COV2 pandemic, government agencies, academic institutions, foundations, non-governmental organisations, and national ethics committees are defining how governments and other entities should effectively use and regulate such technologies. Ethically optimised tools and applications could sustain the widespread use of AI to improve human health and quality of life, while mitigating or eliminating many risks and worst practices.

WHO recognises that ethics guidance which draws upon the shared perspectives of different entities that either develop, use, or oversee such technologies is critical to build trust in these technologies and guard against negative or erosive effects, while avoiding the proliferation of contradictory guidelines. Harmonized ethics guidance should therefore be developed for the design and implementation of AI in global health.

The primary audience for this guidance document is the Ministries of Health that must determine how to integrate and harness these technologies while restricting or prohibiting inappropriate uses. It is also intended for those stakeholders throughout the health care system who will have to adapt to and adopt these technologies, including medical researchers, scientists, health care workers, and patients, whose access to such technologies can both empower people who fall ill but also leave them vulnerable with fewer services and protections. People have always been at the centre of health care at all levels of decision-making, and the inevitable growth of AI in healthcare must continue to put human beings at the centre of healthcare.

The development, adoption, and use of AI requires an integrated approach across multiple government ministries. This includes regulatory agencies, which must validate and define whether, when, and how such technologies are to be used, Ministries of Education that must prepare the current and future workforces (and the general public) on how such technologies function and are to be integrated into everyday practice, and Ministries of Information Technology that should facilitate the appropriate collection and use of health data and narrow the digital divide.

This guidance is also geared to those responsible for the design, deployment, and refinement of these technologies, including technologists and software developers. It is also intended to guide the companies, universities, medical associations, and international organisations that will, alongside governments and Ministries of Health, set policies and practices that determine whether the use of AI for health benefits public and individual health.

WHO recognises that AI is a fast-moving and evolving field, for which many applications, not yet envisaged, will emerge as ever-greater public and private investment is dedicated to the use of AI in health. WHO issued interim guidance on the [use of proximity tracking applications](https://apps.who.int/iris/handle/10665/332200) that are intended to facilitate contact tracing during the SARS-COV2 pandemic, and may consider such guidance for additional tools and applications, and also periodically update this guidance to keep pace with this rapidly changing field.

# Section 2: Describing Artificial Intelligence in health and medicine

There is no universally accepted definition of AI. The term is commonly used to describe computing technology that works similarlyto human intelligence, such as visual recognition, learning, reasoning, problem-solving, decision-making, and adapting to change. The main objectives of AI technologies and applications are to potentially emulate, augment, and enhance human capabilities in solving problems and completing tasks. Although these processes may resemble those performed by humans, their scope in AI systems is strongly limited, and they are not combined in the complex ways that constitute human decision-making. So, for example, an AI system may perform a visual recognition task in a way that is similar to how visual processing takes place in the human brain, but, in the AI system, that processing is not linked to a robust network of other cognitive abilities as it is in a human. In the near term, AI itself is likely to remain confined mainly to “artificial narrow intelligence”, a form of AI in which a technology performs narrowly defined tasks. Hypothetical notions of AI, which assume dramatic advances in the field of artificial intelligence from artificial ‘narrow’ intelligence to artificial ‘general’ intelligence, would mean that AI could autonomously perform a broad range of cognitive tasks integrated into a single source of agency, similar to human intelligence.

One approach to defining AI in health is to define what goals the use of AI is intended to achieve (applications), and how AI is being used to achieve an objective or goal (methods).

There is an ever-expanding set of possible applications of artificial intelligence within health and medicine that defy the use of a precise, ‘textbook’ definition. In the home and community, AI can be used by patients and caregivers for prevention and care support as well as monitoring and communication.[[1]](#footnote-1) In clinic and hospital-based settings, AI is being used for both diagnosis and treatment of numerous diseases, whether through image recognition or machine vision, and to assist macro-level decision-making workflow and process optimisation.[[2]](#footnote-2) AI is also used to assist scientific and medical research, including drug development.[[3]](#footnote-3) Indeed, the scope of AI in health is gradually expanding, especially as the capability of AI evolves.[[4]](#footnote-4) The diverse uses of AI within health care are described in greater detail below (Section 3).

The methods of artificial intelligence are usually equated to machine learning, which is a subset of AI. Machine learning employs a wide range of statistical and mathematical modelling techniques that are built through the analysis of large amounts of data. Such learned patterns are then applied to perform or guide certain tasks and make predictions.

Machine learning can be subcategorised according to how it learns from data, namely unsupervised learning, supervised learning, and reinforced learning. With supervised learning, which is currently the most common in precision medicine, data used to train the model is labelled, i.e. the outcome variable is known. The model learns a pattern to map outputs to different inputs. Unsupervised learning does not involve labelling data; there is no need for human supervision to tell a machine specifically what to look for. Unsupervised learning involves a machine identifying hidden patterns in the data. Reinforcement learning involves machine learning by trial and error to achieve an objective, for which the machine is ‘rewarded’ or ‘penalized’ depending on whether its inferences reach or hinder achieving an objective.[[5]](#footnote-5)

Deep learning is a subset of machine learning that permits software to train itself to perform tasks. Deep learning functions by exposing multi-layered artificial neural networks, or computing systems that are inspired by the neural networks that constitute the brain, to vast amount of data. Deep learning generally improves even when more data is fed into the network if the data has been evaluated to satisfy an appropriate level of quality and measures have been taken to mitigate bias. Although deep learning holds promise, it is also notorious for inaccuracy.

Many machine learning approaches generally require the collection of large amounts of data to produce tangible results. One exception is weakly supervised learning, a branch of machine learning which could be used within medical AI, and which uses noisy, limited, or imprecise sources of data, to provide a supervision signal for labelling large amounts of training data in a supervised setting.

The proliferation of usable data from health care is the result of securing data from numerous sources, including wearable technologies, genetic information, electronic health care records, radiology images, surveillance data, and even from hospital rooms,[[6]](#footnote-6) as well as the evolution and decreasing cost of certain technologies such as sequencing machines. Big Data refers to data sets that are so large, fast, or varied (volume, velocity, variability), that it requires unconventional methods for handling, including data handling methods such as Hadoop or MapReduce. Data scientists apply machine learning techniques to Big Data.

The different sources of data, the ethical and human rights issues with the collection and use of such data, and the appropriate governance of data collection, analysis, and use, are discussed throughout this publication.

# Section 3: Applications of Artificial Intelligence in Health

The use of AI in medicine was first mentioned over 60 years ago, not long after the term ‘artificial intelligence’ was first mentioned in a 1955 academic paper. Even as AI itself went through several cycles of optimism and decline (so-called ‘AI winters’), the use of AI in health care continued. This included assisting with diagnosis, especially where clinical expertise was not easily accessible, and involved early systems such as MYCIN, Iliad, Quick Medical Reference, and Internist-1.[[7]](#footnote-7) AI solutions in health also included identification of risk factors for disease progression and selection of the best course of treatment, given interacting factors such as cost, effectiveness, and risks.[[8]](#footnote-8)

The resurgence of AI occurred during an AI winter that lasted for over two decades (and ended only in the late 2000s), when many of the mathematical and statistical techniques that guide modern AI emerged. AI is now expanding with the accelerating collection of data and investment by both the public and private sector. Building on a relatively long history of AI in health, there are now many uses of AI in health which are each described in this section.

**Uses of AI in medicine**

The use of AI in medicine usually stirs notions of AI replacing clinicians and human decision-making. However, grounded in earlier efforts to define the role of computers in medicine[[9]](#footnote-9) and through regulations that have defined AI as a support tool (as a means to improve judgment), the prevailing sentiment is that AI will augment existing human capabilities and expertise. But if AI is augmenting the abilities of physicians, nurses, and other health care workers, it could also define a new role for patients.

Thus, there are efforts to make AI systems “human-centric”, and models are being developed to work with and for people, or for human–machine teams to work effectively together. AI technologies are changing from a focus on being clinician-centric to patient-centric, with the idea of facilitating shared decision-making with patients, their caregivers, and their communities, although this may not actually be possible in practice. Perhaps the most significant advances, though many are yet not proven and tested, rest in diagnosis and prognosis.

*Diagnosis (and prediction-based diagnosis)*

There are several ways in which AI is being considered as a tool to support diagnosis, including in radiology and medical imaging. Such applications, while widely used compared to other AI applications, are still relatively novel, and AI is not yet used in routine clinical decision support. Currently, AI is being evaluated for diagnosis in oncology in radiological applications (thoracic imaging, abdominal and pelvic imaging, colonoscopy, mammography, brain imaging and dose optimization for radiology treatment), for non-radiological applications (dermatology, pathology), for diagnosis of diabetic retinopathy, for use in ophthalmology, and for RNA and DNA sequencing to guide immunotherapy.[[10]](#footnote-10)

Nevertheless, few such systems have been evaluated in prospective clinical trials. A recent comparison of deep-learning algorithms and health care professionals to detect diseases from medical imaging found that AI is equivalent to human medical judgement in specific domains and applications in specific contexts, while also finding that ‘few studies present externally validated results or compare the performance of deep learning models and health-care professionals using the same sample’.[[11]](#footnote-11) An additional question is whether the performance of AI in these examples will generalise to implementation in practice.

As AI improves, it could also assist medical providers in making faster and more accurate diagnoses. AI could be used for prompt detection of different diseases such as the detection of stroke; pneumonia, breast cancer by imaging[[12]](#footnote-12);[[13]](#footnote-13) in echocardiography for diagnosing coronary heart disease[[14]](#footnote-14) and in screening to predict psychotic episodes by analysis of speech patterns.[[15]](#footnote-15) As AI is used for disease prediction and diagnosis guidance, it could also be adapted to assess the relative risk of disease, which could be used for prevention of lifestyle diseases, such as cardiovascular disease and diabetes. AI might help predict illness or major health events before they occur.

Although current AI-based diagnosis is near-term, and its efficiency and accuracy can be easily evaluated to mitigate potential harm, efficacy and accuracy will be more difficult or impossible to achieve in long-term predictions. The risk of harm therefore increases, as predictions could affect an individual’s health and well-being, restrict access to or unnecessarily expend scarce resources. The ethical implications of prediction-based health care, including the consequences of errors, including the challenge of informed or valid consent, as well as how accuracy may be strongly related to underlying data, are discussed below (See Section 4.X).

### Clinical care

Clinicians might use AI to integrate patient records during consultations, identify patients at risk as well as vulnerable groups, aid difficult treatment decisions, and catch clinical errors. Yet clinical experience and knowledge about patients should remain essential, as AI is not a substitute for clinical due diligence.

Challenges to the progress and deployment of AI in medicine include interactions of social and medical issues and the differing and rapidly evolving skills, preferences, biases, and use patterns of those who use AI. There are also technological challenges for wider use of AI in medicine. Although many prototypes developed in both the public and private sectors performed well in field tests, the technologies often cannot be translated, commercialized, or deployed. An additional obstacle is constant changes in computing and information technology management, whereby systems become obsolete (‘software erosion’) and companies disappear. In resource-poor countries, the lack of digital infrastructure, and the existing digital divide, will impede the use of such technologies.

Health care workers will have to significantly adapt their clinical practice as use of AI increases. Putting diagnostic tools into the hands of lay workers through AI changes the way people think about the role and function of doctors. AI could automate tasks, so that doctors have time to listen and address fears and concerns and ask about unrelated social factors, although clinicians may still worry about responsibility and accountability. Doctors will need new skills to communicate risks, predictions, and trade-offs to patients as well as ethical and legal concerns to understand AI technology. Even if technology makes such gains, they will materialize only if the individuals who manage health systems use them to extend the capability of health systems in other areas.

### The evolving role of the patient in clinical care

AI could eventually change how patients self-manage their own medical conditions, especially chronic diseases such as cardiovascular diseases, diabetes, and their mental health.[[16]](#footnote-16) Patients already undertake significant responsibility for their own care, including taking medicines, improving nutrition and diet, engaging in physical activity, caring for wounds, or delivering injections. AI could assist with self-care, including through conversation agents (e.g. “chat bots”), health monitoring and risk prediction tools, and technologies designed specifically for individuals with disabilities.[[17]](#footnote-17) Many of these AI-guided technologies require the use of either mobile health applications and wearables, and with the growing trend of self-management, so have the use of mobile health applications and wearable technologies increased.[[18]](#footnote-18)

Wearable technologies range from those placed in the body (artificial limbs, smart implants), on the body (insulin pump patches, EEG devices), or near the body (activity trackers, smart watches, and glasses). By 2025, 1.5 billion wearable units may be purchased annually.[[19]](#footnote-19) Wearables will create more opportunities to monitor a person’s health and to capture more data to predict health risks. Such data collection also contributes to the ever-growing practice of ‘biosurveillance’ – a form of surveillance focused on health data and other biometrics, such as facial features, fingerprints, temperature, and pulse.[[20]](#footnote-20) The growth of biosurveillance poses significant ethical and legal concerns, including the uses of such data for medical and non-medical purposes for which explicit consent may not have been obtained, or the repurposing of such data for non-health purposes by a government or company, such as within the criminal justice or immigration system.

AI-technologies may also play a more active role in a patient’s health, such as via ‘just-in-time adaptive interventions’ (JITAIs). These technologies rely on sensors and are intended to provide patients with specific interventions based on the prior and current data the technology collects, as well as notifying a health care provider of specific concerns that may be emerging.[[21]](#footnote-21) The growth and use of sensors and wearables may improve the effectiveness of JITAIs but also heighten concern over the amount of data such technologies are collecting and how such information is used, and the burdens such technologies may shift on patients.

The growing use of digital self-management applications and technologies also raises wider questions over whether such technologies should be regulated as clinical applications (thus requiring greater regulatory scrutiny) or as wellness applications (less regulatory scrutiny). Many digital self-management technologies arguably fall into a ‘grey zone’ between these two categories and may present a risk if such applications are used for a patient’s own disease management or clinical care but remain largely unregulated.

### The shift from hospital to home-based care

The use of telemedicine is one part of a larger shift from hospital-based to home-based care, with the use of AI-guided technologies to facilitate that shift. This includes remote monitoring systems, for example using video-observed therapy for TB, or virtual assistants to support patient care. Even prior to the SARS-COV2 pandemic, over fifty U.S. health care systems were making use of telemedicine services.[[22]](#footnote-22) SARS-COV2 discouraged people in some settings from visiting healthcare facilities and has therefore in certain places accelerated and expanded the use of telemedicine.

### Application of AI in health research

AI can be used to identify patterns in data more quickly and accurately. It can also be used to combine different sources of data to detect patterns and draw inferences. One important source of medical research data is the electronic health record, whether for biomedical research, quality improvement, or the optimisation of health care. From electronic health records, AI can help to identify clinical best practices from data before the more customary diffusion of scientific literature, guideline development, and clinical support tools.

### Use of AI for health systems management

Health systems, even if run through a single-payer government run system, can be overly complex, and involve a range of actors who contribute to, pay for, or benefit from, the provision of health care services. The management and administration of care can be labour-intensive. AI can be deployed to work alongside personnel to assume mundane and repetitive tasks or support complex decision making.

Some of the possible functions of AI within health systems management include: (a) identifying and eliminating fraud or waste; (b) supporting the scheduling of patients; (c) predicting which patients are unlikely to attend a scheduled appointment, and (d) assisting with identification of staffing requirements.[[23]](#footnote-23)

### Public health informatics (uses of AI for public health)

There are several AI tools for population and public health that can be implemented through public health programs. Several concerns with the use of technology for public health surveillance, promotion, and outbreak response must be considered prior to the use of AI for such purposes, including a tension between the public health benefits of surveillance with ethical and legal concerns for an individual’s privacy and autonomy.[[24]](#footnote-24)

Health promotion

AI can be used for health promotion, or to identify specific target demographics or geographies where there are ‘high-risk’ behaviors and target populations that would benefit from health communication and messaging. AI programs can use different forms of data to identify such populations with varying degrees of accuracy that may allow for improved targeting of messaging.

Disease prevention

AI has also been used to address underlying causes of poor health outcomes, such as for addressing risks related to environmental health or occupational health. AI tools can be used to identify bacterial contamination in water treatment plants, simplifying the process of detection and lowering the costs. Sensors can also be used to improve environmental health, such as analysing air pollution patterns or using machine learning to build inferences between the physical environment and healthy behaviours.[[25]](#footnote-25)

Surveillance (including prediction-based surveillance) and emergency preparedness

The use of AI for public health surveillance has been based on collecting evidence and using it to create mathematical models to make decisions. Technology has changed surveillance by the addition of digital “traces”, which are data that are not generated specifically for public health purposes (such as from blogs, videos, official reports and Internet searches) that can be applied to public health. Videos (e.g. YouTube) are a rich source of information that can be transformed into health insights. Public health institutions have not yet made full use of these sources of data. Google Flu Trends, for example, is based on search engine queries about complications, remedies, symptoms, and antiviral medications for influenza, which are used to estimate and predict influenza activity.[[26]](#footnote-26) The characterisation of digital traces as ‘health data’ raises questions as to what types of privacy protections or other safeguards should be attached to such information. It also raises questions of accuracy – while Google Flu Trends first provided relatively accurate predictions in advance of the US Centres for Disease Control, it also overestimated the prevalence of flu between 2011 and 2013 because the system was not re-trained when search behaviour evolved.[[27]](#footnote-27)

Data are useful only when appropriate models are used. Similarly, machine-learning algorithms may be more valuable and effective when augmented by digital traces of human activity. Models have evolved from mechanistic models, such as use of data from air travel networks to predict the possible emergence of pandemics, to “black box” models with unknown components and ingredients. These advances have enabled a more accurate model of human population behaviour, with greater detail, complexity, and capability for forecasting.

Thus, surveillance is already changing, especially real-time surveillance. For example, researchers could detect a surge in cases of severe pulmonary disease associated with the use of electronic cigarettes by mining disparate online sources of information and using Health Map, an online data-mining tool.[[28]](#footnote-28) Similarly, Microsoft researchers have found early evidence of adverse drug reactions from web logs with an AI system. In 2013, the company’s researchers detected side-effects of several prescription drugs before they were found by the U.S. Food and Drug Administration’s warning system.[[29]](#footnote-29) In 2020, the US Food and Drug Administration sponsored a challenge that solicited public submissions to develop computation algorithms for automatic detection of adverse event anomalies using publicly available data.[[30]](#footnote-30)

AI has also been used to assist with both detection and prediction during the SARS-COV2 pandemic, though some believe the techniques and programming developed during SARS-COV2 will only ‘pay dividends’ during a subsequent pandemic.[[31]](#footnote-31) HealthMap first issued a short bulletin about a new type of pneumonia in Wuhan, China at the end of December 2019.[[32]](#footnote-32) Since then, AI has been used to ‘now-cast’ (assess the current state of) the SARS-COV2 pandemic[[33]](#footnote-33), while in some countries, real-time data on the movement and location of people has been used to build AI models that can forecast regional transmission dynamics and guide border checks and surveillance.[[34]](#footnote-34) Prior to the pandemic, WHO had started to develop EPI-BRAIN, a global platform that will allow experts in data and public health to analyze large datasets for emergency preparedness and response (See Section 10.X).

Outbreak response

In addition to the use of AI for surveillance during the SARS-COV2 response, the possible uses of AI for different aspects of outbreak response have expanded during the SARS-COV2 pandemic, and includes studying SARS-COV2 transmission, facilitating detection, developing possible vaccines and treatments, and understanding the socio-economic impacts of the pandemic.[[35]](#footnote-35) While many possible uses of AI have been identified and implemented during the SARS-COV2 pandemic, the actual impact are likely to have been modest, and in some cases, as with early SARS-COV2 screening tools, many AI screening tools ‘were utter junk’ for which companies ‘were trying to capitalise on the panic and anxiety.’[[36]](#footnote-36)

New applications[[37]](#footnote-37) are intended to support the off-line response. This has included proximity tracking applications that are intended to provide users (and possibly health authorities) with a notification if the user has been in proximity (for a duration of time) of an individual who subsequently tests positive for SARS-COV2. WHO and many Ministries of Health have also deployed symptom checkers, which are intended to guide users through a series of questions to assist with determining whether an individual should seek additional medical advice or testing for SARS-COV2. The first symptom checkers were ‘hard coded’ based on accumulated clinical judgment since there was no pre-existing data to apply AI. Such symptom checkers were based only on a simple decision tree, whereas AI systems based on machine learning, require accurate training data, which was scarce for a new disease such as SARS-COV2.[[38]](#footnote-38) New symptom checkers are utilising machine learning to provide advice to patients[[39]](#footnote-39), although the effectiveness of these symptom checkers are not yet known.

AI has also been introduced as a means to map the movement of individuals to approximate the effectiveness of government-mandated orders to remain in confinement, and in some countries, AI technologies have been used to assess individuals who should self-quarantine and be tested. These technologies raise ethical concerns with respect to privacy and discrimination risks as well as possible unnecessary restriction on movement or access to services that heavily impact in the exercise of a range of human rights.[[40]](#footnote-40) As with all AI technologies, the actual effectiveness of such technologies remains questionable without systematic testing and evaluation, and as such the uses described above are not yet established.

*Uses of AI for drug development*

AI could change drug discovery from a labour-intensive to a capital- and data-intensive process with the use of robotics and models of genetic targets, drugs, organs, diseases and their progression, pharmacokinetics, safety, and efficacy. AI could be used in drug discovery and throughout drug development to reduce the development period, as well as to make the process of drug development less expensive and more effective.[[41]](#footnote-41) As an example, AI was used to identify potential treatments for Ebola virus disease, although, as in all drug development, identification of a lead compound may not result in a safe, effective therapy.[[42]](#footnote-42)

At present, there is connection between humans and machines, led either by humans or by AI with human oversight. In the next two decades, as work with machines is optimized, AI could evolve. Computing could allow drug discovery and development by finding novel leads and evaluating whether they meet the criteria for new drugs, structuring unorganized data from medical imaging, searching large volumes of data, including health care records, genetics data, laboratory tests, the Internet of Things, published literature and other types of health big data to identify structures and features, while recreating the body and its organs on chips for AI analysis.[[43]](#footnote-43) By 2040, some trials might be virtual – without animals or humans – based on models of the human body, tumours, safety, efficacy, epigenetics, and other parameters. Prescription drugs could be designed for each person. Such efforts are in part to contribute to precision medicine, or health care that is individually tailored based on a person’s genes, lifestyle, and environment.

For this to occur, challenges in data harmonization, sharing and interoperability will have to be addressed, as well as ethical dilemmas arising for other uses of AI in health care, such as algorithmic bias and a general public demand for “explainability”, which will affect the use of AI in drug development.[[44]](#footnote-44) Ownership of health data, algorithms applied to drug development, and the end-product, may be contested, and issues with liability could also arise. There is also a need for managing expectations of what might be possible from the use of AI, as some of the more extravagant predictions will be hype.[[45]](#footnote-45)

### The future of AI in health

While AI might not replace clinical decision-making, it could increasingly augment the decisions made by clinicians and in places where there are limited resources and take over screening and evaluation if insufficient medical expertise is available. Humans may develop more trust in the accuracy and reliability of AI output if such technologies are subjected to randomised clinical trials of sufficient power to assess safety and efficacy

Yet whether AI can advance beyond narrow tasks depends on numerous factors beyond just the state of AI science and the willingness of providers and patients to trust AI-based technologies. Most studies of AI systems are on paper only and do not involve actual clinical testing. One hurdle to major improvements in the use of AI in health care stems from lingering uncertainties about the actual efforts of these systems in real-world settings.

Securing, standardizing, and analysing large quantities of health data will require addressing numerous ethical challenges, as well as the technical challenges of making disparate and often messy data useful for AI-technologies. There must also be improved methods to test and measure whether AI technologies are effective to engender wider adoption and uptake. The ethical challenges and approaches to resolve these issues are thus an important part of not just ensuring that AI-guided technologies improve human health and provide equitable benefits, but also to address many of the concerns that would and should lead governments, health care providers, and patients to temper enthusiasm and use of these technologies.

# Section 4: Ethical challenges for the use of Artificial Intelligence in Health Care

This section examines several ethical challenges that will emerge with the use of AI in health. The use of AI is not limited to health care, and several ethical concerns discussed below arise with the use of AI in other domains, including criminal justice, social protection, financial services, and humanitarian assistance. Challenges with the routine use of AI that have emerged in other domains, and measures to prevent or mitigate such concerns, should be applied where relevant for the use of AI in healthcare and medicine.

## Assessing when artificial intelligence should be used

There is a risk of hype that overstates what AI can already accomplish, unrealistic estimates of what could be achieved as AI evolves, and the uptake of unproven products and services.[[46]](#footnote-46) This lies in part with the growing appeal of ‘technological solutionism’ – in which technologies such as AI are deployed as a ‘magic bullet’ to solve deeper social, structural, economic, and institutional barriers.[[47]](#footnote-47) The appeal of ‘technological solutionism’ and the promise of technology can lead to overestimating benefits while ignoring challenges and problems that new technologies such as AI may introduce. This can result in unbalanced health care policy and misguided investments in countries that have few resources to spend within a health care system.[[48]](#footnote-48)

First, the AI technology itself must meet existing standards of scientific validity and accuracy that are usually applied to other medical technologies. For example, new digital technologies developed in the early stages of the SARS-COV2 pandemic did not necessarily meet any objective standard of efficacy to justify their use.[[49]](#footnote-49) An emergency does not justify unproven deployment[[50]](#footnote-50), and in fact efforts to ensure that resources were being allocated to where it was most urgently needed should have heightened the vigilance of both companies and governments to ensure such technologies were first deemed accurate and effective.

Second, the benefits of AI can be overestimated when there are erroneous assumptions of the infrastructure and institutional context. This can include the quality and availability of data, as well as the presence of appropriate regulations, stakeholder participation, and oversight – all of which are needed to ensure that ethical and legal concerns can be addressed and human rights are not violated. There must be the legal, technical, and institutional structures in place, as well as protections for individuals.

Third, there may be enough ethical risks and concerns with a use case or a specific AI-technology, which may even provide accurate and useful information and insights, that should discourage a particular use. An AI technology that can predict which individuals are likely to develop Type 2 diabetes or HIV could provide benefits to an ‘at-risk’ individual or community. Yet such predictive technology could foster unnecessary stigma of individuals or communities whose choices and behaviours are questioned and even criminalised, over-medicalise otherwise healthy individuals, create unnecessary stress and anxiety, and open up individuals to aggressive marketing by pharmaceutical companies and other for-profit health care services.[[51]](#footnote-51)

Fourth, even if an AI-technology does not trigger a warning of an ethical risk, the benefits of such a technology may not be justified by the extra expense or cost associated with the procurement, training, and technology investment that is required.[[52]](#footnote-52) Robotic surgery may produce better outcomes, but the opportunity costs associated with such investments must also be considered.

On the other hand, unrealistic expectations of what AI should achieve could also unnecessarily discourage its use. Thus, machines and algorithms (and the data used for algorithms) are expected to be perfect while humans can make mistakes. Medical professionals might overestimate their ability to perform tasks while ignoring or underestimating the value of algorithmic decision tools for which the challenges can be managed, and the evidence indicates the technology confers a measurable benefit. This would result in avoidable morbidity and mortality and make it blameworthy not to make use of a specific AI-technology, especially if the standard of care is already shifting to incorporate its use.[[53]](#footnote-53)

Even after an AI technology is introduced into a healthcare system, there should also be continuous evaluation based on the real-world implementation and impact of the technology, and consideration of an algorithm’s performance as it learns from data that is different from its training data. Regulatory agencies can play an important role to ensure rigorous testing, transparent communication of outcomes, and monitoring of a technology’s performance. Impact assessments can also guide a decision on whether to use AI within an area of health care prior to and after its introduction (See Section 10.X).[[54]](#footnote-54)

The criteria for use of AI may be different when contemplating introduction in low-income countries, wherein financial resources and information and communication technology (ICT) infrastructure lag high-income economies and would discourage use due to the significant investments that would be required. Risk-benefit calculations that discourage a specific use of AI in high-income countries may result in a different outcome in a low-income country which may lack, for example, enough health care workers that could perform certain tasks. The factors that developing countries should consider, including how best to achieve universal health coverage, and the challenges associated with the digital divide, are discussed in the following sub-sections.

## The use of artificial intelligence to achieve universal health coverage

The United Nations Secretary General has noted that the safe deployment of new technologies, including AI, can help the world achieve the United Nations (UN) Sustainable Development Goals (SDGs).[[55]](#footnote-55) This would include the health-related objectives within SDG3. AI could also help to satisfy global commitments to achieve universal health coverage (UHC). Digital health technologies are already used widely in LMICs, including for collection of data, dissemination of health information via mobile phones, and the expanded use of electronic medical records using open-software platforms and cloud computing.[[56]](#footnote-56) There are both opportunities and risks for pursuing the use of AI to achieve universal health coverage.

The use of AI will not proceed at the same pace, present the same opportunities, or carry the same risks, across low- and middle-income countries. Many so-called developing countries have sophisticated economies and digital infrastructure while other countries, such as India, have both world-class digital infrastructure and millions of people lacking electrification. The countries with the largest challenges to adopting AI are least developed countries (LDCs). Yet AI could allow these countries to leapfrog existing models of health care delivery to improve healthcare outcomes quickly.[[57]](#footnote-57)

Schwabe and Wahl (2020) have identified four uses of AI in LMICs: diagnosis, morbidity or mortality risk assessment, disease outbreak and surveillance, and health policy and planning.[[58]](#footnote-58) Assisting health care workers with diagnosis and assessment, or reducing the workload of overburdened health care workers, is needed in many low-income settings that are already facing a chronic health care worker shortage. There are suggestions that AI could fill gaps where health care services or skilled workers are not present.[[59]](#footnote-59) New developments with AI could also, depending on the outcomes of rigorous evaluation, improve identification of disease outbreaks and support surveillance.

Thus, AI could support the management of antiretroviral therapy through use of AI to predict HIV drug resistance and disease progression, which could help physicians optimise therapy. One of the more promising uses of AI may be to support improved detection of TB through a support system that interprets staining images[[60]](#footnote-60), and to assist health systems to identify individuals with TB who may not be reached by health care systems and therefore do not know their status.[[61]](#footnote-61) Predictive analytics could help avert unnecessary morbidity and mortality, such as systems that can predict birth asphyxia. An expert system used in developing countries is 77% sensitive and 95% specific for predicting the need for resuscitation.[[62]](#footnote-62)

And yet the challenges for adopting AI to generate these benefits are considerable and raise concerns that the pursuit of AI could divert attention and resources from proven but underfunded interventions that would reduce morbidity and mortality in low- and middle-income countries.

One concern is the existing digital divide, or the unequal access to telecommunications, internet services, and computing equipment. The digital divide, the lack of supporting and stable infrastructure in some developing countries, and how the digital divide affects the use of AI in health care, are discussed in the next sub-section.

A second set of concerns is with data, which are concerns for all countries but are more challenging in low-and middle-income countries. For AI training, there is the possibility of collecting poor quality data, which can result in bad models that predict artefacts in the data instead of actual clinical outcomes. There can also be an absence of data, which alongside poor quality data, could distort an algorithm’s performance (resulting in inaccurate performance), or which could prevent the use of an AI technology for a specific population due to insufficient usable data.

Additionally, non-uniform data sets that may be collected in low- and middle-income countries may require significant investment to make usable. Compiling data in resource-poor settings is difficult and time consuming and would need to consider the range of responsibilities already shouldered by community health workers in LMICs. Data from the most vulnerable or marginalised populations, including those where health care services are absent or lacking, is not likely to exist or could be inaccurate. Data may also be difficult to collect because of language barriers, and mistrust may lead people to provide incorrect or incomplete information. Often, irrelevant data are gathered, which can undermine the overall quality of a dataset.[[63]](#footnote-63) Even if there are good quality datasets, developers will need to make additional investments to ensure such data is debiased. AI will usually favour the populations that have the most data to contribute, and thus in unequal societies, this may cause AI to be biased towards the majority and disadvantage a minority population. Finally, such additional investments to ensure data is debiased requires data labelling to include individuals from diverse ethnic and social backgrounds.

Generating more data from low- and middle-income countries could raise fears of ‘data colonialism’ – where such data could be used for commercial purposes without due regard for consent, privacy, and autonomy. Such collection of data without informing individuals of intended uses (commercial or otherwise) undermines the agency, dignity, and human rights of these individuals. This is especially a concern because of the possibility that companies from countries with strict regulatory frameworks and data protection laws could expand data collection to low- and middle-income countries without eventually providing products and services to underserved communities and countries.

Third, there may not be sufficient consideration of whether an AI technology is appropriate and adapted to an existing context in LMICs, such as the diverse languages and scripts within countries (or between countries). [[64]](#footnote-64) A lack of investment, for example in translation, can mean that certain applications may not operate correctly, or simply cannot be used by a population. Such lack of foresight points to a wider problem, which is that many AI technologies may be designed by and for high-income populations, and by individuals or companies without enough understanding of the characteristics of target populations in LMICs.

A fourth concern is that these technologies may be introduced in low-and middle-income countries without adequate evidence (including randomised clinical trials) or safeguards prior to roll-out, as has been the case during the SARS CoV-2 pandemic.[[65]](#footnote-65) Regulatory agencies in LMICs may not have the capacity or expertise to assess AI technologies to ensure that systematic errors do not affect diagnosis, surveillance, and treatment. Certain technologies, if not deployed carefully, could exacerbate existing health care disparities, including those related to ethnicity, socioeconomic status, and gender. Such technologies may also be introduced into countries without up-to-date data protection and confidentiality laws (especially for health-related data), or without data protection authorities that can rigorously protect confidentiality and privacy of individuals and communities.

Finally, expanding the use of AI must be done carefully to avoid situations in which large numbers of people may be accurately diagnosed for a health condition but are left without access to appropriate treatment options. There is a duty to provide treatment following testing for and confirmation of disease, and the relative low-cost at which AI diagnostics can be deployed necessitate careful planning to ensure people are not left without treatment options. This also needs to be balanced against not using more accurate diagnostic instruments, which could mean individuals are inaccurately diagnosed and are provided the wrong treatment. Prediction tools that may anticipate a disease outbreak will need to be complemented by robust surveillance systems and other effective measures to respond to an outbreak.

While many of these concerns with the deployment of AI are especially relevant to low- and middle-income countries, the use of AI in high-income countries as a means to expand health care coverage and services in marginalised communities will raise many of the same concerns, including an enduring digital divide in high-income countries, the lack of quality data or collection of data that incorporates pre-existing clinical biases (as well as inappropriate data collection practices), and a lack of treatment options following diagnosis.

## Artificial intelligence and the digital divide

Although the cost of digital technologies is falling, access has not become more equitable. The “digital divide” refers to uneven distribution of access to, use of or effect of information and communication technologies among any number of distinct groups. For example, 1.2 billion women (327 million fewer women than men) in LMICs do not use mobile Internet services because they cannot afford or do not trust the technology, even though the cost of devices should continue to fall.[[66]](#footnote-66) Gender is only one dimension of the digital divide; others include geography, culture, religion, language, and inter-generational differences. The digital divide begets other disparities and challenges, many of which will affect the use of AI and how AI can itself reinforce and exacerbate such disparities. Thus, in 2019, the UNSG High-Level Panel on Digital Cooperation recommended that “by 2030, every adult should have affordable access to digital networks, as well as digitally enabled financial and health services, as a means to make a substantial contribution to achieving the SDGs.”[[67]](#footnote-67)

The human and technical resources required to fully realize the benefits of digital technologies are also unequally distributed, and infrastructure to operate digital technologies may be limited or non-existent. Various technologies require an electricity grid and information and communication technology infrastructure, including electrification, internet connectivity, wireless and mobile networks, and devices. Yet an estimated 860 million people worldwide do not have access to electricity, including 600 million people in sub-Saharan Africa, with growing pressure on the electrical grid in cities due to urbanisation.[[68]](#footnote-68) Even in high-income economies with near-universal electrification and enough resources, the digital divide has persisted. In the United States, for example, millions of Americans in rural areas and in cities still lack access to high-speed broadband services.[[69]](#footnote-69) 60 percent of healthcare facilities outside metropolitan areas also lack broadband.[[70]](#footnote-70)

Even as countries overcome the digital divide, there will be a need to ensure that technology providers ensure that the infrastructure, services, and programs are interoperable to ensure that different platforms and applications can work seamlessly with one another so that an emerging digital health care system is not fragmented.

## Data collection and use

The collection, analysis, and use of health data, including clinical trial data, lab results, and medical records, is a bedrock of medical research and the practice of medicine. Over the last two decades, what qualifies as health data has expanded dramatically. This includes massive quantities of data about individuals from many sources, including genomic data, radiology images, medical records, and different types of non-health data converted into health data.[[71]](#footnote-71) These various types of data, which are collectively ‘biomedical big data’, form a health data ecosystem. This includes data from standard sources (e.g. health services, public health, research) and expanded sources (environmental, lifestyle and socioeconomic, behavioural, and social).[[72]](#footnote-72) [See Figure 1.](https://www.who.int/ehealth/resources/ecosystem.pdf?ua=1)

Thus, there are many more sources of health data, entities that wish to make use of such data, and commercial and non-commercial applications. The successful development of an AI system in health care relies on high-quality data that is used to both train an algorithm and to affirm whether the values of the parameter of an algorithmic model are correct (its validation).

The potential benefits of biomedical big data can be ethically important, since AI technologies, if it uses high-quality data, can improve the speed and accuracy of diagnosis, improve the quality of care, and reduce subjective decision-making. The ubiquity of health data, and the potential sensitivity of health care to data, points to the possible benefits. Health care, compared to other sectors, is still lagging in the adoption of data science and AI (although some would disagree).[[73]](#footnote-73) And informed of the potential benefits of the collection and use of such data, individuals may support the use of such data for their own personal benefit or a wider group of people.[[74]](#footnote-74)

Nevertheless, several concerns may undermine the effective use of health data to provide AI-guided health care or for its use in research and drug development. One concern with health data is the quality of such data, and whether such data is relevant to all populations that an AI-technology intends to serve. Training data will always have one more systematic biases due to under-representation of a gender, age, race, sexual orientation, or other characteristics. These biases will emerge in the modelling and subsequently diffused through the resulting algorithm.[[75]](#footnote-75) Such biases favour the majority unless instructed to do otherwise. Concerns with bias and discrimination using AI in health are discussed separately below (See Section 4.X).

A second major concern is the safeguarding of individual privacy. The collection, use, analysis and sharing of health data has consistently raised broad concerns for individual privacy, whether because a lack of privacy may harm the individual (such as future discrimination on the basis of one’s health status), or may cause a wrong, such as affecting one’s dignity if sensitive health data is shared or broadcast to others.[[76]](#footnote-76) There is a risk that the sharing or transfer of data leaves it vulnerable to cyber-theft or accidental disclosure.[[77]](#footnote-77) Recommendations generated by an algorithm based on an individual’s health data also raises privacy concerns, as a person may have an expectation of privacy for such ‘new’ health data.[[78]](#footnote-78) These privacy-related concerns are heightened for stigmatized and vulnerable populations, wherein data disclosures can lead to discrimination or punitive measures.[[79]](#footnote-79) There are also concerns about the rights of children[[80]](#footnote-80), which could include future discrimination on the basis of the data accumulated about a child, a child’s ability to protect their privacy, and their autonomy to make choices regarding their health care. Measures to collect data or track an individual’s status and to construct digital identities to store such information have accelerated during the SARS-COV2 pandemic. See Box X1.

|  |
| --- |
| Box X1: The growth of digital identification in the SARS-COV2 pandemic  The SARS-COV2 pandemic is itself expanding and accelerating the creation of the infrastructure that could be used to create digital identities that are intended to store health data for several purposes. In China, a QR code system has been established that uses the existing digital payment identities established by Alipay to introduce an ‘Alipay Health Code’, which then on the basis of the data it collects, enables an algorithm to ‘draw automated conclusions as to whether someone is a contagion risk.’[[81]](#footnote-81) To eventually roll out a national effort to immunise millions of people against COVID-19, India is considering applying its existing national digital ID system, Aadhar, to avoid duplication and to track beneficiaries.[[82]](#footnote-82) More widely, many different entities, including travel firms, airports, some governments and political leaders, and the digital ID industry, are calling for the introduction of immunity passports, or a digital ‘credential given to a person who is assumed to be immune from COVID-19 and so protected against re-infection.’[[83]](#footnote-83) For some of these efforts, there are questions as to whether such interventions are effective (scientifically valid), whether it will create forms of discrimination or targeting of certain populations, and whether it may exclude certain segments of the population or not be applicable to many people who do not have access to appropriate technology and infrastructure. It also raises concerns of generating a permanent digital identity for individuals linked to their health and personal data that they may not have consented to and could permanently undermine individual autonomy and privacy.[[84]](#footnote-84) |

Third, health data collected by providers can be repurposed, or shared with government agencies that may take punitive measures against individuals.[[85]](#footnote-85) Such repurposing, or function creep, is a challenge that predates but is also heightened with the use of AI in healthcare. Such data can also be shared with companies that use such data to develop an AI-technology, to inform the marketing of goods and services, or to create prediction-based products used, for example, by an insurance firm[[86]](#footnote-86) or large technology company. Such uses of health data, often unknown to those who have supplied such data, have generated front-page headlines and public concerns.[[87]](#footnote-87) Some companies are already able to draw on large quantities of health data collected through their own products and services and may also acquire supplemental data through a data aggregator or broker[[88]](#footnote-88), while in other cases such companies may rely upon governments to aggregate data that can be used by public, not-for-profit, and private sector entities[[89]](#footnote-89).

Concerns with commercialisation of health data include the individual’s loss of control over his or her data (that there was no explicit consent to such secondary use of the data), with how such data (or outcomes generated by such data) may be used by that company or a third party, with concerns that companies are allowed to profit from the use of such data, and concerns for privacy since companies may not meet a duty of confidentiality, whether purposefully or inadvertently (for example due to a data breach).[[90]](#footnote-90)

Informed consent is increasingly infeasible in an era of biomedical big data.[[91]](#footnote-91) When meaningful consent is possible, it can overcome many concerns, including those related to one’s privacy. Yet the scale and complexity of biomedical big data may make it impossible to keep track of and make meaningful decisions about all the possible uses of personal data.[[92]](#footnote-92) The potential uses of health data may not be known, as data may be eventually linked to and used for a purpose that is far removed from the original source of information. Current and future uses of health data, whether population-level data analytics or predictive-risk modelling, might make patient consent impossible.[[93]](#footnote-93) Even where a use may lend itself to consent, the procedures itself may fall short, or individuals may not be able to consent, including because they may have insufficient access to a health data system.

A separate concern is how to manage the use of health data, including consent for use of such data, after an individual has died. Such data could provide numerous benefits for medical research,[[94]](#footnote-94) such as to improve an understanding of the underlying causes of cancer[[95]](#footnote-95), or to improve the diversity of data that is used for medical AI. However, there is also a need to safeguard the data against unauthorised use, especially as a patient cannot consent to the use of his or her data after dying. Existing laws either have limited circumstances under which such data can be used, but it is restrictive in most jurisdictions.[[96]](#footnote-96) There are proposals to improve the sharing of such data through voluntary and participatory approaches that enable individuals to provide broad consent or selective consent for use of their data after death (much as individuals can provide consent for the use of their organs for medical research).[[97]](#footnote-97)

If patients are unable to safeguard privacy using consent-based mechanisms, other privacy safeguards, including a data holder’s duty of confidentiality, also has its shortcomings. Although confidentiality is a well-recognised pillar of medical practice, duties of confidentiality may not be sufficient to account for the many types of data that are now used to guide AI health technologies, and may also not be sufficient to control the production and transfer of health data.[[98]](#footnote-98)

One proactive approach to preserve privacy is through either the deidentification, the anonymisation, or the pseudo anonymisation of health data. De-identification of data refers to a process of preventing personal identifiers from being connected to information. Anonymisation of personal data is a subcategory of de-identification whereby both direct and indirect personal identifiers are removed, and technical safeguards are implemented so that there is a ‘zero re-identification’ risk, (whereas de-identified data can be re-identified through the use of a key).[[99]](#footnote-99) Pseudo anonymisation, as defined by the European Union’s General Data Protection Regulation, is the ‘processing of personal data in such a way that the data can no longer be attributed to a specific data subject without the use of additional information provided that such additional information is kept separately and is subject to technical and organisational measures to ensure that the personal data are not attributed to an identified or identifiable natural person.’[[100]](#footnote-100)

The use of such techniques could safeguard privacy and encourage data sharing but also raises several concerns and challenges. For example, in the United States, fully de-identified health data can be used for other purposes without consent.[[101]](#footnote-101) De-identification may not always succeed, and the use of ‘data triangulation’ techniques can mean that even if a dataset is de-identified and the dataset is incomplete, it could be reconstructed by a third party and lead to re-identification of an individual.[[102]](#footnote-102) While anonymisation may protect privacy, it can undercut positive benefits provided by health data, including the re-assembly of different fragments of an individual’s health data into a comprehensive profile of a patient that are required for some forms of AI (predictive algorithms of mortality). Furthermore, anonymisation may work at cross-purposes to having control over data. While anonymisation may safeguard privacy, it may undermine a person’s right to control their own data and how it may be used.[[103]](#footnote-103)

A final concern with biomedical big data is that it may foster a divide between those who accumulate, acquire, analyse, and control such data versus those who provide such data but have little control over its use. This can mean having little or no opportunity to know how such data is already being used, whether by a government or company, no opportunity to provide any form of consent to how such data could be used, and often less bargaining power if recommendations obtained from such data have an adverse impact upon an individual or community.[[104]](#footnote-104)

## Accountability and responsibility for decision-making with AI

Much of the momentum with AI rests with the notion that such technologies – whether for diagnosis, care, or systems-level investments, could improve clinical and institutional decision-making for health care. Clinicians and health care workers have numerous cognitive biases and commit diagnostic errors. The National Academy of Sciences found that five percent of U.S. adults seeking health care advice are subject to diagnostic error and that diagnostic errors account for 10% of all patient deaths.[[105]](#footnote-105) At the institutional level, machine learning could reduce inefficiencies and errors and allocate resources in a more appropriate manner, assuming the underlying data is both accurate and representative.[[106]](#footnote-106)

But AI-guided decision-making also introduces several trade-offs and safety risks. One set of trade-offs involves the consequences of displacing human judgment and human control, as well as concerns associated with utilising AI to predict a person’s health status or the evolution of a disease. These considerations are addressed in the subsequent sub-section (See Section X: Autonomous decision-making).

Governments can violate human rights (and companies can fail to respect human rights), undermine human dignity, or cause tangible harms to human health and well-being, when using AI-guided technologies. These violations may be unforeseen when an AI technology is developed and may only emerge after an AI-guided technology evolves with real-world use. If proactive measures such as improved transparency and continuous updating of training data does not avoid harm, then one set of remedies is through civil (and occasionally criminal) liability. The use of liability regimes to address harms caused by AI-guided technologies are addressed separately (See Section X: Liability regimes for AI in health).

This section considers challenges with assigning responsibility and accountability for the use of AI in health care. Responsibility is necessary to ensure that individuals and entities are held to account for adverse effects of their actions, to maintain trust, and to protect human rights. Yet AI technologies have characteristics that affect notions of responsibility (and accountability), including its: opacity, reliance of human input, interaction, and discretion, scalability, capacity to generate hidden insights, and the complexity of the software.

One challenge with assigning responsibility is the emergence of ethics guidance issued by technology companies separately or jointly.[[107]](#footnote-107) Such ethics guidance is intended to set out norms and standards that the companies publicly and voluntarily commit to comply with. On the one hand, it is welcome that technology companies recognize that AI technologies, for health care and in other sectors, are of public concern and must be carefully designed and deployed to avoid different harms, such as violations of human rights or bodily injury, for example. Yet such ethical guidelines can also, depending on how they are implemented, be little more than ‘ethics washing’.[[108]](#footnote-108) First, the public tends to have little to no role in setting out such standards.[[109]](#footnote-109) Second, such guidelines tend to apply to prospective behaviour of technology companies for the technologies they design and deploy (role responsibility), and not historic responsibility for any harms for which responsibility should be duly allocated. This creates a responsibility gap since it does not address causal responsibility, or retrospective harm.[[110]](#footnote-110) Third, oversight of whether companies are complying with its own guidance tends to be done internally with little to no transparency, and without any enforcement institutions or mechanisms that are empowered to act independently to evaluate if commitments are being met.[[111]](#footnote-111),[[112]](#footnote-112) Finally, these commitments are not legally enforceable if violated.[[113]](#footnote-113)

AI affords great power and benefits (including profit) to those who design and deploy such systems. Thus, reciprocity should apply – companies who may reap direct and indirect benefits associated with AI-guided technologies, should also have to shoulder responsibility for its negative consequences (See discussion on liability in Section 8). Companies should also submit to independent audits and oversight for enforcement of its own ethics standards to ensure that such standards are being met and that corrective action is taken when a problem arises.

A second challenge to assigning responsibility is the ‘control problem’ associated with AI, wherein developers and designers of AI may not be held responsible since AI-guided systems function independently of its developer, and may evolve in a way that a developer could claim was not foreseeable.[[114]](#footnote-114) This creates a responsibility gap which could place an undue burden upon a victim who suffers harm, or the clinician or health care worker who uses such a technology but did not have a hand in the technology’s development and design.[[115]](#footnote-115),[[116]](#footnote-116) On the other hand, assigning responsibility to the developer might ensure the developer is incentivised to take all possible steps to reduce harm to the patient.

The ’control problem’ will become even more salient and difficult to address with the emergence of automated AI. Technology companies are making large investments to automate the use of AI due in part to the scarcity of AI developers. Such automation of AI, which can be done through programs including BigML, Google AutoML, and Data Robot, may be attractive to public health institutions that wish to use AI but may lack budgets to hire AI developers.[[117]](#footnote-117) And while automated AI may be more accurate, it may not necessarily be fair, ethical or safe. Since AI programming will be automated, it would also mean that the checks and balances provided by the presence of a human developer to ensure safety and identify errors could be automated. Thus, the control problem is abstracted one step further away from the patient.

A third challenge is the ‘many hands’ problem, or the ‘traceability’ of a harm. Since the development of AI requires contributions from many different actors or agents, assigning responsibility amongst these different actors is difficult, both legally and morally.[[118]](#footnote-118) This creates a diffusion of responsibility amongst all the different contributors to an AI-guided technology. The participation of a machine in making decisions can also discourage assigning responsibility to humans involved in the design, selection, and use of an AI-guided technology.[[119]](#footnote-119)

Such diffusion of responsibility may mean that a victim may not be compensated for harms he or she suffers, the actual harm itself may not be adequately detected (including its cause), the harm itself may not be addressed, and societal trust in such technologies may be diminished if it appears none of the actors that develop and use such technologies can be held responsible.[[120]](#footnote-120)

*Accountability for AI-related errors and harm*

At the individual level, since AI technologies are used only to assist or augment clinical decision-making, and not to replace it, there may be an argument for holding clinicians accountable for any harms that result from the use of AI technologies in health care. Yet this oversimplifies the reasons why a harm may occur, and who should be held accountable for such harm. If there is a mistake in the use of the technology, accountability might rest with the clinician, assuming the clinician has had the opportunity to be trained on the use of the technology that otherwise may not have been a part of his or her medical training.[[121]](#footnote-121) Yet if there is an error in the content of the AI technology, accountability might be better placed upon those who develop or test the AI technology, in lieu of requiring the clinician to judge whether the AI technology is providing useful guidance.[[122]](#footnote-122)

There are other reasons to not hold clinicians solely accountable for decisions made by AI technologies. First, clinicians do not exercise control over an AI-guided technology and its recommendations.[[123]](#footnote-123) Second, since AI technologies tend to be opaque and use black-box algorithms, a physician may not understand how an AI system converts data into decisions.[[124]](#footnote-124) Third, a clinician may not have themselves chosen to use an AI technology, which may have been driven by preferences of a hospital system or other external decision-makers.

Furthermore, making physicians accountable for all harms caused by an AI technology enables technology companies and developers to avoid accountability. This turns human users of such technologies into scapegoats for all faults arising from an AI-technology without having control of what decisions the AI technology makes.[[125]](#footnote-125)

On the other hand, a separate reason not to fully insulate clinicians from accountability for errors in content is to avoid physicians engaging in automation bias, or not considering whether an automated technology meets their needs or those of the patient.[[126]](#footnote-126) When this occurs, a clinician may overlook errors that should have been spotted by human-guided decision-making. While there is a need for physicians to trust an algorithm, it should not lead to the physician ignoring their own expertise and judgment and simply rubber-stamping what is recommended by a machine.[[127]](#footnote-127) As autonomous systems such as driving and warfare, have emerged, there has been growing concern about whether humans can exert “meaningful control” over such technologies or whether the technologies will increasingly make decisions independently of human input and control. (See Section X: Autonomous decision making)

Finally, when decisions are made to use an AI technology across a health care system, assigning accountability is even more complex since the developer, institution, and the physician may all have played a role in the medical harm, and yet none of the participants are fully to blame.[[128]](#footnote-128) In such situations, accountability may not rest with either the provider or the developer of a technology, but with the government agency or institution that selected, validated, and deployed the technology.

## Autonomous decision-making

There has been no ‘full transfer’ of decision-making from humans to machines within health care; at best AI is used only to augment human decision-making. Yet there are signs of full delegation of routine medical functions to AI. Such delegation of clinical judgment introduces concerns with assigning responsibility and accountability if there is medical error. Full delegation also creates a risk of automation bias on the part of the provider, as discussed in the prior section. There are other concerns that could emerge if human judgment is increasingly replaced by machine-guided judgments, and wider ethical concerns with the loss of human control, especially if prediction-based health care becomes the norm. Yet, as with autonomous cars, it is not likely that AI in medicine will ever achieve full autonomy, but may only achieve conditional automation, or automated systems that require human backup.[[129]](#footnote-129)

### Implications of replacing human judgment

There are benefits to replacing human judgment and for humans to cede control over certain aspects of health care. Use of AI systems to make specific, well-defined decisions may be entirely justified, if there is compelling clinical evidence that the system performs that task better than a human. Leaving decisions to humans when machines can perform them more rapidly, accurately and with greater sensitivity can mean that some patients suffer avoidable morbidity and mortality without the prospect of some offsetting benefit.[[130]](#footnote-130)

In some cases, automation of routine, mundane functions, such as recording information, will actually liberate a medical provider to build a relationship with a patient and can enhance clinician–patient relationships while AI-guided machines automate certain aspects of caregiving.[[131]](#footnote-131)

Yet the shift towards applying AI technologies within more complex areas of medicine will present several challenges. One challenge is the likely emergence of ‘peer disagreement’ between two competent experts – an AI machine and a doctor.[[132]](#footnote-132) In such situations, there is no pathway to fuse the decisions together or to reason with an algorithm since it cannot be accessed or engaged to change its mind. There are also no clear rules for determining who is right, and the trust in either a technology or a physician, if left to a patient, may depend upon factors that have no grounding in the ‘expertise’ of the machine or a doctor. Choosing between either option leads to an undesirable outcome – if the doctor ignores the machine, it means the AI has added little value.[[133]](#footnote-133) If the doctor accepts a decision made by a machine, it can undermine the authority of the clinician and weaken his or her accountability. Some may argue that since an algorithm combines the expertise of multiple experts, and many data points, that its recommendation should be preferred.[[134]](#footnote-134)

This challenge with human-computer interactions has been addressed historically through mechanisms that: (a) validate systems, (b) provide appropriate education for users, and (c) continuously validate the system. Yet it may also be ethically challenging for doctors to rely on the judgment of AI since they will have to contend with decisions based on black-box algorithms.[[135]](#footnote-135) The widely held convention is that many algorithms, e.g. those based on artificial neural networks or other complex models, are “black boxes” that make inferences and decisions that even their own developers do not fully understand.[[136]](#footnote-136) Thus, the question arises as to whether doctors can be asked to act on decisions developed through such black-box algorithms.

On the one hand, it has been argued that AI systems must be transparent, namely that clinicians must be able to explain how the system arrived at a diagnosis so that they can audit the decision. And some forms of transparency – e.g. information that can be communicated to a patient – is possible without having to fully explain a black-box algorithm. This includes: (a) transparency on the main features that the AI is using to make a prediction, (b) transparency on clinical feedback of healthcare providers on the AI (and the utility of the AI), (c) transparency on the accuracy rate of the AI in testing, and (d) transparency on the likelihood of a patient getting a low-quality prediction.

Furthermore, some argue that, if a trade-off must be made between even-greater transparency and accuracy, transparency should be preferred. Yet this requirement reaches beyond what may be possible, or even desirable, in the medical context. While it is often possible to explain to patients why a treatment is the best option for a specific condition, it is not always possible to explain how that treatment works or its mechanism of action, because medical interventions are sometimes used for a particular purpose before their mode of action is understood.[[137]](#footnote-137) Rather than being able to explain how an AI model arrives at a particular judgment, it may be more important to explain how a system has been validated and whether a particular use falls within the parameters in which such a system can be expected to produce reliable results. There are also other types of information that a clinician should be provided even if they do not understand exactly how an algorithm functions, including the data it was trained on, how and who build the AI model, and the key variables underpinning the AI model.

### Implications of the loss of human control

The loss of human control, by assigning decision-making to AI-guided technologies, could have impacts for different participants in health care and on the health care system. This includes impacts on the: patient, on the clinician-patient relationship, on the relationship of the health care system to technology providers, and on the choices that societies should make about the standard of care.

Although providing individuals with more opportunities to share data and to obtain autonomous health advice could improve an individual’s agency and support self-care, it could also generate anxiety and fatigue.[[138]](#footnote-138) As more personal data is then collected by such technologies and used by clinicians, it could instead leave patients increasingly excluded from shared decision-making with their providers and leave the patient unable to have agency or autonomy for decisions made about their health.[[139]](#footnote-139) This is because most patients may not have the level of knowledge about how and why AI technologies make certain decisions (and the technologies themselves may not be sufficiently transparent even if a patient is well-informed). In some situations, individuals may feel unable to refuse treatment, in part also because the patient cannot dialogue with or challenge a recommendation made by an AI-guided technology (e.g. a notion that ‘computer knows best’), or that the patient is not provided with enough information or a rationale to provide informed consent.[[140]](#footnote-140) Hospitals and health care providers are unlikely to inform patients that AI has been used as a part of the decision-making process (either to guide, validate, or overrule a provider). On the one hand, there is no precedent for seeking the consent of patients to use technologies for diagnosis or treatment. But the use of AI in medicine, and failing to disclose its use, could fundamentally challenge the core of informed consent. This depends on whether any of the rationales for informed consent – namely protection, autonomy, prevention of abusive conduct, trust, self-ownership, non-domination, and personal integrity, are triggered using AI in clinical care.[[141]](#footnote-141)

The loss of control could equally be felt by a physician who is left out of decision-making that occurs between a patient and an AI health technology. This would mean that clinicians could no longer engage in the back-and-forth that is currently integral to clinical care and shared decision-making between a provider and patient.[[142]](#footnote-142) Some may see the loss of physician control over the patient as beneficial to promoting patient autonomy, but equally there is a risk of surrendering decision-making to an AI-technology, which may be more likely to occur if a technology is presented to the patient as providing better insight than a physician into his or her health status and prognosis.[[143]](#footnote-143) Furthermore, if an AI technology reduces contact between a provider and patient, it could also reduce opportunities for clinicians to offer health promotion interventions to a patient, and in general undermine the supportive care that is usually included with the provision of health care.[[144]](#footnote-144)

Control could be construed as not just being surrendered to a technology, but to companies that will exert power over the development, deployment and use of AI in health care. At present, technology companies are investing resources to accumulate data, computing power and human resources to develop new AI health technologies.[[145]](#footnote-145),[[146]](#footnote-146),[[147]](#footnote-147) This may be done by large companies in partnership with the public sector, as is the case in the United Kingdom[[148]](#footnote-148), but could be done in a way in which different areas of expertise or decision-making are siloed within different companies (and their technologies), with the rules (and standard of care) governed by individual companies that manage such technologies, as opposed to health care systems. In China, several large technology companies, including Tencent[[149]](#footnote-149), Baidu[[150]](#footnote-150), and Alibaba[[151]](#footnote-151), are rapidly expanding the provision of both online and offline health services, backed by the accumulation of data and the use of AI, to develop new points of access to health care. And yet companies, unlike health systems or governments, may ignore the needs of citizens (and the obligations owed to citizens, since there is a distinction between citizens and customers). These concerns heighten the need for regulation and careful consideration of the role of companies to directly provide health care services.

*The use of AI for predictive analytics in healthcare*

Health care has always included and depended in part upon predictions and prognoses. There are many possible benefits to prediction-based health care that relies upon the use of AI. AI could also be used to assess the relative risk of disease, which could be used for prevention of lifestyle diseases, such as heart disease and diabetes. AI could also assist health care providers in predicting illness or major health events before they occur. For example, early studies with limited datasets indicate that AI could be used to diagnose Alzheimer disease years before symptoms appear.[[152]](#footnote-152)

Yet there are a range of risks associated with the use of AI to provide predictions that affect patient care or influence the allocation of resources by a hospital or health care system. While AI-based diagnosis is near-term and its efficiency can be tested, thereby mitigating potential harm, efficacy and accuracy may be more difficult or impossible to achieve in long-term predictions. The risk of harm therefore increases dramatically, as predictions could affect an individual’s health and well-being or restrict or unnecessarily expend scarce resources. Prediction-based health care, even if it is effective for diagnoses or an accurate prediction of disease, may present significant risks of bias and discrimination against individuals because of a predisposition to certain health conditions[[153]](#footnote-153), which could manifest itself in the workplace, health insurance, or access to health care resources.

The use of predictions across health care could also raises an ethical concern with respect to informed consent and individual autonomy if predictions are shared with people who did not consent to surveillance, detection, or the use of predictive models, to draw up inferences about a patient’s future health status (or to provide them with a ‘predictive diagnosis’ that a patient did not proactively request). The Convention for the Protection of Human Rights and Dignity of the Human Being with regard to the Application of Biology and Medicine (Oviedo Convention) states the following: ‘Everyone is entitled to know any information collected about his or her health. However, the wishes of individuals not to be so informed shall be observed.’[[154]](#footnote-154)

Prediction-based technologies could also challenge an individual’s ability to exercise freedom and choice, even outside of the doctor-patient relationship. Such a use of AI, combined for example with ‘nudging’, could transform an application used to promote healthy behaviour into a powerful technology that could exert control over the choices people make in their daily life.[[155]](#footnote-155) This is because nudging, and the multiple channels through which it can be deployed, can be far more effective compared to the sporadic interactions between a health care provider and patient. If AI predicts that a patient is at high risk of developing a certain disease, will individuals still have the right to engage in certain behaviours if that increases the likelihood of certain diseases? Such restrictions on autonomy could be exercised via a doctor, but also via an employer, insurer, or directly by an AI application paired to a wearable device.

Finally, prediction-based technologies could make recommendations based on an inference that optimise markers for health instead of identifying an underlying patient need. Thus, an algorithm that makes a mortality prediction based on training data may learn from the training data that patients who visit the chaplain have an increased risk of death.[[156]](#footnote-156)

## AI, bias, and discrimination

Societal bias and discrimination are often replicated within AI technologies, including the use of AI in the criminal justice system, banking, human resources, and the provision of public services. Thus, an algorithmic risk score intended to help judges determine the future recidivism of defendants in the criminal justice system had a systematic bias against black defendants in the United States, who were twice as likely to be classified as future criminals compared to white defendants.[[157]](#footnote-157) The use of AI within health will also lead to existing biases and discriminatory practices in health care to appear in the data used to train AI models, and thus in the recommendations made by AI-guided technologies.

Biases are often tied to how an algorithm is designed or trained. AI-based technologies are and will continue to be developed by one demographic group and gender, which increases the likelihood of certain biases in the design of the technology.[[158]](#footnote-158) Thus, the Apple Health Kit, which enabled specialised tracking of different health risks, did not include a women’s menstrual cycle tracker in its first releases, which may have been due in part of having no women in the development team.[[159]](#footnote-159) The need for diversity should extend to those who validate an algorithm. They must also be diverse to recognise flaws in the design or functionality of the AI that demonstrate biases within the data.

### Biases in data

There is bias in the data sets used to train[[160]](#footnote-160) the models, whereby ethnic minorities, older people, rural communities, and other disadvantaged groups, are excluded. In general, AI is biased towards the majority data set. Yet biases in data can arise or be exacerbated due to poor clinical practice, such as disparities in health care for racial and ethnic minorities, which can then be reproduced in AI technologies. These systematic biases, if enshrined in the AI, can become normative biases, and can exacerbate and fix (in the algorithm) existing disparities in health care.[[161]](#footnote-161)

Data biases are also caused by other factors. One cause is the digital divide (See Section 4.X). Thus, women are far much less likely than men to have access to a mobile phone or mobile internet in LMICs; 327 million fewer women than men have access to mobile internet[[162]](#footnote-162), meaning that women not only contribute less data towards data sets used to train AI, but are less likely to benefit from such services. Another cause is the unbalanced collection of such data even where the digital divide may not play a factor. This includes the collection of genetic data, which tends to be collected disproportionately from people of European descent.[[163]](#footnote-163),[[164]](#footnote-164)

Such biases can also emerge if certain individuals and communities choose not to participate in the provision of data, resulting in a lack of data that represents those groups. Data on certain population subsets may be difficult to collect if it depends on expensive devices such as wearable monitors.

Representative data may also not be possible to collect if an AI technology is initially trained in local populations which has a distinct health profile compared to where the AI technology is used. Thus, an AI technology that is trained in one country and subsequently used in a quite different country may discriminate against or be ineffective against a population with different characteristics, whether race, ethnicity, or body type. AI is often trained on data for which a company or research organisation has local access, but such an entity may seek to sell an AI algorithm globally without having addressed the insufficiency of the training data.

Overall, these different biases of the training data may especially affect those who face discrimination on the basis of several different characteristics, such as, for example, an individual who may be currently facing prejudice at the same time due to a health condition, race, and gender.

### Biases in deployment and use of AI

Biases can also be introduced when systems are implemented in real world settings. If AI has not considered the diversity of populations that may require its use, whether due to variations in age, disability, co-morbidities, or poverty, AI will discriminate against or work improperly for these disadvantaged populations. Such bias may manifest itself in the workplace, in health insurance or in access to health care resources, benefits and other opportunities. To the extent that AI is designed predominantly in high-income countries, there may be significant misunderstanding of how AI should be deployed in LMICs, including the discriminatory impacts (or worse) that the use of such technologies could have, or that such technologies cannot be used with certain populations.

AI could also, at a population level, encourage use of resources for people who will have the greatest net benefit, e.g. younger, healthier individuals, and divert resources and time from costly procedures intended for the elderly. Thus, if an AI technology is trained to ‘maximise global health’, it may do so by allocating the majority of resources to the healthy to keep such populations healthy, while casting aside a disadvantaged population as an outsider. This would not necessarily be done by accident. Yuval Hariri notes in Homo Deus:

‘First, medicine is undergoing a tremendous conceptual revolution. Twentieth-century medicine aimed to heal the sick. Twenty-first-century medicine is increasingly aimed to upgrade the healthy…Consequently, by 2070 the poor could very well enjoy much better healthcare than today, but the gap separating them from the rich will nevertheless be much greater.’[[165]](#footnote-165)

Some technology groups and universities have attempted to remove bias from AI technologies; however, they have been only partially successful, as algorithmic finesse and computing power cannot squeeze out information that is not present. In other words, a model’s efficacy depends on the datasets used to “train” it and the domain knowledge, if applicable, that was incorporated. If the datasets and knowledge are not sufficiently broad and inclusive, the outcomes will also not be fully inclusive. An alternative approach is to persuade communities that are either reluctant to provide their data or are generally excluded from doing so. A third approach is to decrease the size of the majority dataset so that it is at the same ‘level’ as the minority. This will create a weaker model, but it will be fairer, and thus may require a developer to decide between a more accurate model or a fairer one.

## Impacts of AI on labour and employment (in health and medicine)

AI is viewed equally with optimism and pessimism at it relates to its impact on the health workforce. It is perhaps less contested that nearly all jobs in health care will require a minimum level of digital and technological proficiency. The Topol Review: Preparing the Health Workforce for a digital future, concluded that within two decades, 90% of all jobs in the United Kingdom’s National Health Service would require digital skills, including navigating ‘data-rich’ health care environment, as well as digital and genomics literacy.[[166]](#footnote-166)

Optimistic views of the arrival of AI posit that AI will automate and reduce the burden of routine tasks for clinicians and allow doctors to focus on the challenging work and to focus on engaging with the patient. It could also empower doctors to work across a broader set of areas and providing support in areas for which technology can enable clinical decision-making.

It is expected that digitization of health care and the introduction of AI will create numerous new jobs in health care, such as software development and health care systems analysts, as well as those who will facilitate the use of AI in health care and medicine. One typology identifies three types of jobs – trainers, or those that can evaluate and stress test AI technologies; explainers, or those that can explain how and why an algorithm can be trusted; and ‘sustainers’, or those that need to monitor the behaviour and identify unintended consequences of AI systems.[[167]](#footnote-167)

AI could also expand one of the scarcest resources in health care systems, which is the amount of time that doctors and nurses can attend to patients. If doctors and nurses can hand over repetitive or administrative tasks to AI and therefore spend less time addressing ‘routine care cases’, it could allow providers to have more time to attend to more urgent or complex (rare) cases, as well as improve the overall quality of care offered to patients.[[168]](#footnote-168) Currently, where in some cases AI is already being integrated into health care systems as secondary medical support, and which could best be characterised as a transition period, AI may increase the tasks and add to the workload of doctors and nurses.

Yet AI (as well as telemedicine) could also create inequitable access to health care services (in particular access to health care personnel), where people in rural areas or low-income countries may be asked to make due with expanded access to AI-based services and telemedicine[[169]](#footnote-169) even as individuals in high-income countries and urban areas continue to draw the benefits of in-person care.

There are other concerns with the onset of AI. Health care providers, who already must absorb large amounts of new information to meet the standard of care, may need to learn new skills (use of AI in everyday practice) on a regular basis, and skills may need to rapidly evolve if uptake of AI accelerates in the coming years. Such continuous education may be neither available nor accessible to all health care workers. A separate concern is that AI will automate many of the jobs and tasks required of health care personnel, leading to significant job losses that could affect nearly every part of the health workforce, including specific types of doctors. This is in part since AI has already replaced many jobs in other industries, reduced the total number of people needed to carry out certain roles, or created the expectation that many jobs (e.g. up to 35% of all jobs according to one study in the United Kingdom) would be lost.[[170]](#footnote-170)

There are other scenarios that have been envisaged with the onset of AI. One scenario predicts that AI will lead to short-term instability, causing many job losses in certain areas even as overall employment increases by the creation of new roles, thereby resulting in unemployment for some who may or may not be able to retrain for such new roles. Another scenario posits that such job losses will not materialise either because clinicians or health care workers will have to still fulfil other roles, or that the full integration of such technologies will only occur over a long period of time, during which other roles for health care workers and clinicians will emerge (such as labelling data or designing and evaluating AI technologies).[[171]](#footnote-171)

Even if AI may not displace clinicians, it could lead to less job security and stability for doctors. One trend has been the ‘Uberization’ of health care, with AI facilitating the creation of health care platforms where contractors, including drivers, temporary workers, nurses, physician assistants, and even doctors, work on demand.[[172]](#footnote-172),[[173]](#footnote-173) Over the last decade, health care and education have experienced the fastest growth of ‘gig workers’, or those workers who complete different tasks on a temporary basis with no overall stability of employment.[[174]](#footnote-174) While this could provide more flexible services for people, it could also sever the relationship between recipients and providers of health care and create a sense of insecurity for certain types of health workers.

Perhaps the most certain change will be that with the increased use of AI, the nature of practicing medicine and providing health care will fundamentally change. It could, as noted above, provide health care providers with more time to care for patients, or it could, if patients start to interact with AI more frequently and directly, relegate doctors to spending less time having direct contact with patients, require more time administering technology, analysing data, as well as learning how to use new technologies, and therefore pushing health care workers to exit the practice of medicine.[[175]](#footnote-175)

## Challenges with the commercialisation of health care using AI

The use of AI within health care has been pushed in large part through significant advocacy and investment by companies – from small start-up firms to the largest technology companies, which are mostly based in China and the United States (though their operations and partnerships are global). There is an expectation that these companies can marshal their capital, in-house expertise, computing resources and data to identify and build novel applications that support providers and health care systems.

Some of the services already widely used in health care are for back-office functions and to help to manage health care systems. Some companies involved in the development of technology, such as the pharmaceutical and medical device industry, are integrating AI into their processes and products; while insurance firms are also utilising AI for assessing risk or even automating the provision of insurance, a practice which can raise ethical concerns with respect to algorithmic decision-making.

One prominent use of AI in health care by firms is to support diagnosis, treatment, monitoring and patient adherence. Such applications could generate positive benefits for health care systems. Yet over the last decade, as more technology firms, and especially the largest firms, have sought to enter the health care field, many concerns have emerged. This section examines a range of ethical challenges that rest with the practices of the largest technology firms within the AI for health care field, although some of these concerns apply equally to mid-size firms and start-ups.

One general problem is the lack of transparency. While many of these firms know much about its users, there is little that either its users, civil society, or regulators know about the activities of the firms.[[176]](#footnote-176) These practices can remain hidden in part due to commercial secrecy agreements or the lack of any general obligations for transparent practices, including the role these firms play within health care. Without such transparency (and accountability), the firms themselves have little incentive to act in a manner that does not cross certain ethical boundaries or to disclose deeper problems within its technology or models.[[177]](#footnote-177) In fact, many companies often may want to keep algorithmic models proprietary and secret, as full transparency can lead to criticism of both the technology and the company itself.[[178]](#footnote-178)

A second broad concern is that the overall business model of the largest technology firms involves both the aggressive collection and use of data to make technologies effective and to use surplus data for commercial practices, a practice labelled by Professor Susanna Zuboff as a new form of exploitation known as ‘surveillance capitalism’.[[179]](#footnote-179) Thus, the last decade has been punctuated by several known examples in which large technology firms have sought out large data sets of sensitive health information for the development of AI technologies in health care.[[180]](#footnote-180),[[181]](#footnote-181) While these efforts at acquiring and using such health data may have been done with the express purpose of developing useful AI-based technologies in health, the data itself was not acquired with the express consent of those who provided the data, and the actual benefits conferred by the data upon these firms may be far in excess of what was needed to deliver a specific product, and the firms may not be providing equal benefits to the public that generated the data in the first place.

Such acquisition of sensitive health information can generate legal concerns. First, even if the data is anonymised by an acquiring firm, a firm, due to the amount of information it may already have about users from other sources, would be in a place to be able to combine data and de-anonymise relevant data sets.[[182]](#footnote-182) Second, several large technology firms have been accused and even fined for mishandling data[[183]](#footnote-183) – a concern that should be heightened with firms that acquire health data. And third, as firms continue to accumulate large amounts of data, this can raise anti-trust concerns related to the growing market power of such companies, including barriers that affect smaller companies that may wish to enter such a market.[[184]](#footnote-184)

An additional concern is with the growing power that some companies may exert over the development, deployment and use of AI in health care (including drug development) and the extent to which corporations exert power and influence over individuals and governments. This is especially because data, computing power, human resources and technology can be concentrated within a few companies, including the ownership of such technology through either legal rights (intellectual property protection) or natural monopolies gained through the size of a company’s platform.

While much scrutiny has been trained upon the growing role of large U.S. corporations (such as Google, Facebook, and Amazon) in the development and provision of AI in healthcare, there is a growing ecosystem of large technology companies in China that are playing a growing role in healthcare through similar services and technologies. This includes Tencent, Baidu, and Alibaba, which both are building their own technology platforms as well as collaborating with user platforms such as WeChat to reach millions of people across China.[[185]](#footnote-185) Tencent, for example, has at least three main areas of investment in health, including new AI-based technologies to assist with diagnosis and treatment, a ‘smart hospital’ which is intended to provide a web of online services and data connectivity through a smart health card (which raises its own concerns with data privacy and use – see above), and a Medipedia that is intended to provide health information to users online.[[186]](#footnote-186) Alibaba is working with hospitals to predict patient demand to allocate healthcare personnel or the development of AI-assisted diagnostic tools for radiology.[[187]](#footnote-187)

This power and control of a market may be in part to the ‘first-mover’ advantage that several large firms may have earned through their entry into AI for health care. Even if the data used by a firm can be used by others (for example, data from a public health care system), other firms may be discouraged or unable to replicate the use of such data for a similar purpose, especially if another company has already done so.[[188]](#footnote-188) This power can also mean that the rules set by companies can force even the largest and wealthiest governments to change course. During the SARS-COV2 pandemic, Google and Apple chose to introduce one technical standard (especially related to the where and how data should be stored) for proximity tracking applications, that differed from the preferred approach of several high-income country governments, and which resulted in at least one government changing the technical design of its proximity tracking application to comply with technical standards set out by the two companies. Even though these two companies’ approach may have been consistent with privacy considerations, there is a wider concern that these firms, with control of the infrastructure on which such applications operate, could and can force governments to adopt a technical standard that may have been inconsistent with its own public policy and public health objectives.[[189]](#footnote-189)

When most of the data, health analytics, and algorithms are managed by large technology companies, it will be increasingly likely that those companies will govern decisions that should be in the domain of citizens, societies and governments, because of their control and power over the resources and information that underpin the digital economy.[[190]](#footnote-190) This power imbalance also affects people who should otherwise be treated equitably by their governments, or at least, if treated unfairly, can hold their governments accountable where such inequity arises. Without a strong government role, companies might ignore the needs of citizens, particularly those at the margins of their societies and the global economy.[[191]](#footnote-191)

Oversight by governments might be integrated into public–private partnerships, whereby a collaboration is used for purposes and motives that are not primarily intended to support public health and well-being. Public-private partnerships can also lead to the misappropriation of resources (in most instances, patient data), conflicts of interest in the decision-making of such partnerships, or can forestall or limit the use of regulation, where needed, to protect the public interest.[[192]](#footnote-192),[[193]](#footnote-193)

# Section 5: Risks associated with the widespread adoption of AI for health

In the U.S. and Europe, regulatory agencies are developing regulatory pathways to evaluate and approve AI-based technologies. There are concerns that existing regulatory frameworks may be insufficient to avoid several risks with such technologies, in particular if the technologies are not subjected to randomised clinical trials of sufficient power to assess safety and efficacy, or if regulators rely upon ‘trusted’ developers to assess the performance of their technologies. The ethical norms that should apply to such regulatory frameworks are discussed below (See Section X: Regulation of AI in health care).

For many low- and middle-income countries, which may already not have adequate staffing to assess medicines, vaccines, medical devices, and diagnostic instruments[[194]](#footnote-194), the onset of AI-based technologies could leave such countries with a choice of either not using such technologies or having to approve such technologies without full oversight of their appropriateness, safety, and efficacy for use, and for which the risks of using such technologies (whether due to risks of bias or loss of privacy for example) are potentially greater. Furthermore, governments that are still integrating other technologies into a health care system may struggle to identify enough resources to plan and introduce AI technologies. Thus, the unplanned or unregulated introduction of such technologies in resource-poor countries is a significant risk since no framework, regulations, or regulators may be able to enforce minimum standards. There are several other risks associated with the introduction and use of AI technologies considered in this section.

## Safety of AI technologies

There are numerous risks to patient safety that could arise from the use of AI, and which regulatory agencies may not be able to identify when reviewing a technology for approval. Errors in AI systems, whether incorrect recommendations (e.g. which drug to use, who to treat between two sick patients), or false-negative and false-positive recommendations, can cause injury to a patient.[[195]](#footnote-195) Health care providers also commit judgment errors or other human-induced errors, but the risk with AI is that such an error, if fixed in an algorithm, could cause irreparable harm to thousands of people in a short time if the AI technology is used widely.[[196]](#footnote-196) Thus, an AI-based mobile app developed by DeepMind to predict acute kidney failure, produced two false positives for every correct result, and therefore did not result in improvements in patient outcomes.[[197]](#footnote-197) Even if the system identified some patients requiring treatment, this benefit was cancelled out by overdiagnosis. Such false positives can harm patients if it persuades doctors to take riskier courses of action – such as prescription of a more potent and addictive drug – in response to such predictions.

An AI application could also provide the wrong guidance if there are code errors that result in a health application providing incorrect advice. For example, a code error in the National Health Service (UK) COVID-19 application, which was designed to notify individuals to self-isolate if exposed, had been programmed incorrectly.[[198]](#footnote-198) Thus, a user of the application had to be next to a highly infectious patient for five times longer than the NHS’ own guidance had deemed risky before being instructed to self-isolate. This meant that despite up to 19 million people downloading the application, a ‘shockingly low’ number of people were told to isolate, thereby exposing themselves and others to risks of SARS-COV2 infection.

There is also the possibility that a developer could engage in the unethical design of an AI technology that could optimise an outcome that is tailored to generate profits for the provider or conceal certain practices. Such a design could in fact be accurate (compared to a different modelling techniques) but still generate unmerited sales revenue. Malicious design has affected other sectors, such as the automobile sector, where algorithms used for automobile emissions were programmed to conceal the true emissions profile of a major car manufacturer.[[199]](#footnote-199)

Finally, the use of computers carriers an inherent risk of safety flaws due to insufficient attention to design of machines to minimise risks as well as flaws in the computer code and associated bugs and glitches. Injury or death caused by such flaws and breakdowns are under-reported, and there are also no official figures and few large-scale studies. One study noted that, for example in the United Kingdom, there may be up to 2,000 deaths a year due to computer errors and flaws, and that it is an ‘unnoticed killer’.[[200]](#footnote-200)

*Skills loss of health care workers*

The use of AI to augment, and possibly replace, daily functions of health care workers and physicians, could remove the need for health care workers to maintain certain skills, such as the ability to read an X-ray. At some point, physicians may be unable to carry out such a task without the assistance of a computer, and that AI systems will be asked to take on the repository of medical knowledge that used to rest with human providers.[[201]](#footnote-201) Such dependence on AI systems could erode independent human judgment, and in the worst-case scenario, could leave providers and patients incapable of acting if an AI system fails or is compromised.[[202]](#footnote-202)

## Cybersecurity risks

There is an expectation that as health care systems become increasingly dependent on AI, these technologies will be targeted for malicious attacks and hacking attempts. Such efforts could be used to shut down such systems, to manipulate the underlying training data used for the algorithm, thereby changing its performance and recommendations, or ‘kidnapping’ data for ransom.[[203]](#footnote-203) There could also be attacks on AI developers, who may be targeted through spear-fishing attacks and hacking, which could allow an attacker to modify an algorithm without the knowledge of the AI developer.

Manipulating an algorithm, especially if it runs independent of any human oversight, could be done to generate revenues for specific recipients, with large sums at stake (health care is nearly 20 percent of the U.S. economy). The UK Information Commission Office has noted that cyberattacks on the health sector are the most frequent.[[204]](#footnote-204) Breaches of health data, which represents some of the most sensitive data about individuals, could cause harm to privacy and dignity and the broader exercise of human rights.

## Use of AI for drug development

The potential use of AI to both simplify and accelerate drug development is expected to only be realised over time. There is an expectation that machine learning systems can be used to accurately predict which drugs can be safe and effective and are best suited for human use. There is also an expectation that such predictive models could allow pharmaceutical companies to take ‘regulatory shortcuts’ and conduct clinical trials based on fewer studies and with less patient data.

Yet such approaches carry risks. The predictive models are based itself on algorithms that need to be assessed for their accuracy, which can be difficult due to lack of transparency or explainability of how such algorithms function. Furthermore, reducing the number of trials or patients studied can create concerns that patients may be vulnerable to unforeseen safety risks that are not identified by an algorithm.

Furthermore, such machine learning algorithms may exacerbate bias. For example, if an AI technology uses a dataset that is racially homogenous, biomarkers that an AI technology identifies and that could be responsive to a therapy, may only be appropriate for the race or gender of the applicable dataset, and not for a more diverse population. In such cases, a drug that is approved may not be effective for the excluded population or may even have a harmful impact on a patient’s health and well-being.

# Section 6: Existing governance frameworks that apply to AI in health

Governance frameworks to manage the emergence of AI, and specifically the use of AI in health, have been fragmented and limited. There are numerous sets of principles and guidelines developed for different applications of “ethical” AI in the private and public sectors and in research institutions[[205]](#footnote-205); however, there is no consensus on its definition, best practices or ethical requirements, and different legal regimes and governance models have been associated with each set of principles. There are also several other sets of norms, rules and frameworks, most importantly human rights obligations which are being applied to the use of AI, bioethics laws and policies, data protection laws, and regulatory standards. These are briefly summarised in this section and discussed at length elsewhere in the report.

At the World Health Assembly in 2018, Member States unanimously adopted resolution WHA71.7, which calls on WHO to prepare a global strategy on digital health to support national health systems in achieving universal health coverage.[[206]](#footnote-206) Such a global strategy, and additional governance frameworks and standards established by WHO, will also contribute towards building a governance framework for AI in health.

## AI and human rights

Machine learning systems could advance human rights (including the human right to health) yet could undermine core human rights standards (See Section 7). Human rights organizations have adapted the application of existing human rights laws and standards to AI assessment and are reviewing them in the face of challenges posed by AI. In 2018, the Council of Europe’s Committee of Ministers issued draft recommendations to Member States on the human rights impacts of algorithmic systems.[[207]](#footnote-207) The Council of Europe is also examining the feasibility and potential elements of a legal framework for the development, design, and application of digital technologies, based on its standards on human rights, democracy, and the rule of law. The Toronto Declaration addresses the impact of AI on human rights and situates AI within the universally binding, actionable framework of human rights laws and standards; it provides mechanisms for public and private sector accountability, protects people from discrimination and promotes inclusion, diversity and equity while safeguarding equality and effective redress and remedy.[[208]](#footnote-208)

Legal frameworks for human rights, bioethics and privacy that have been adopted by countries are applicable to several aspects of AI in health. They include Article 8 of the European Convention on Human Rights: the right to respect for private and family life, home and correspondence[[209]](#footnote-209); the Oviedo Convention on Human Rights and Biomedicine, which covers ethical principles of individual human rights and responsibilities[[210]](#footnote-210); Convention 108+ for the Protection of Individuals with Regard to Automatic Processing of Personal Data and Guidelines on the Protection of Individuals with Regard to the Processing of Personal Data in a World of Big Data, prepared by the Consultative Committee of Convention 108+.[[211]](#footnote-211)

Yet, even with robust human rights standards, organizations and institutions recognize that better definition is required of how human rights standards relate and apply to AI, and that new laws and jurisprudence are required to address the intersection of AI and human rights. New legal guidance has been prepared by the Council of Europe. In 2019–2020, the Council set up the Ad-hoc Committee on Artificial Intelligence to conduct broad multi-stakeholder consultations to determine the feasibility and potential elements of a legal framework for the design and application of AI according to the Council of Europe’s standards on human rights, democracy and the rule of law. Further, in 2019, the Council of Europe released Guidelines on Artificial Intelligence and Data Protection, also based on the protection of human dignity, and safeguarding human rights and fundamental freedom.[[212]](#footnote-212) In addition, the Council of Europe’s European Commission for Efficiency of Justice has an ethical charter that includes five principles relevant to use of AI in health.[[213]](#footnote-213)

## General principles of AI

One review of more than 50 proposals for general AI principles was based upon identifying 10 principles: humanity, collaboration, sharing, fairness, transparency, privacy, security, safety, accountability, and long-term AI. None of the proposals included more than two thirds of the principles.[[214]](#footnote-214) Humanity, fairness, privacy, and safety were most cited, while references to long-term AI, collaboration and sharing were less frequent. In another mapping and analysis of current principles and guidelines for ethical use of AI, Jobin et al. noted convergence on transparency, justice, fairness, non-maleficence, and responsibility, while other principles such as privacy, solidarity, human dignity, and sustainability were under-represented.[[215]](#footnote-215)

Several inter-governmental organizations and countries have proposed such principles (Box X2).

|  |
| --- |
| Box X2: Examples of AI principles proposed by inter-governmental organizations and countries   * The Organization for Economic Co-operation and Development (OECD) Recommendation of the Council on Artificial Intelligence, the first intergovernmental standard on AI, was adopted in May 2019 by OECD’s 36 member countries, along with and a range of partner economies which have since adhered to them.[[216]](#footnote-216) The OECD AI Principles[[217]](#footnote-217) also provided the basis for the G20 AI Principles endorsed by Leaders in June 2019. While not legally binding, OECD recommendations do carry a political commitment, and in other policy areas (e.g. privacy and data protection) have proved highly influential in setting international standards and helping governments to design national legislation.  The OECD launched an online platform for public policy on AI, the AI Policy Observatory.[[218]](#footnote-218) (See Section 10.X) The OECD is cooperating closely on this and other initiatives on the ethical implications of AI with the Council of Europe and UNESCO. * The Ibero-American Data Protection Network, which consists of 22 Data Protection Authorities from Spain, Portugal, Mexico, and other countries in Central and South American and the Caribbean, has issued: (a) General Recommendations for the Processing of Personal Data in Artificial Intelligence[[219]](#footnote-219), and (b) Specific guidelines for Compliance with the Principles and Rights that Govern the protection of Personal Data in Artificial Intelligence Projects[[220]](#footnote-220). * In 2019, the Council of Europe Commissioner for Human Rights issued recommendations to ensure that human rights are strengthened rather than undermined by AI: Unboxing Artificial Intelligence: 10 Steps to Protect Human Rights recommendations[[221]](#footnote-221). * The European Commission appointed 52 representatives from academia, civil society and industry to its High-level Expert Group on Artificial Intelligence and issued Ethics Guidelines for Trustworthy Artificial Intelligence.[[222]](#footnote-222) * Japan has issued several guidelines related to the use of artificial intelligence, including AI R&D guidelines as well as AI Utilisation Guidelines.[[223]](#footnote-223) * China has issued the Beijing Declaration on Artificial Intelligence, company principles (such as those of Baidu and Tencent) and principles prepared by academic institutions, such as the Tsinghua University Centre for International Security and Strategy. * In Singapore, a series of AI governance and ethics initiatives was designed to build an ecosystem of trust to support AI adoption. They include Asia’s first Model AI Governance Framework, released in January 2019; an international industry-led Advisory Council on the Ethical Use of AI and Data formed in June 2018; and a research programme on the governance of AI and data use established in partnership with the Singapore Management University in September 2018.[[224]](#footnote-224) |

## AI principles for health

No specific principles for use of AI in health have yet been proposed at the international level. Before WHO’s work on guidance for the ethics and governance of AI in health, the WHO Global Conference on Primary Health Care issued the Astana Declaration, which includes principles for the use of digital technology.[[225]](#footnote-225) The Declaration calls for promotion of rational, safe use and protection of personal data and use of technology to improve access to health care, enrich health service delivery, improve the quality of service and patient safety and increase the efficiency and coordination of care.

UNESCO has recommended development of guidance and principles for the use of AI in general and as related to Big Data in health. UNESCO’s work on the ethical implications of AI is supported by two standing expert committees, the World Commission on the Ethics of Scientific Knowledge and Technology and the International Bioethics Committee. Other work includes the 2017 International Bioethics Committee report on Big Data and Health, which identified important elements of a governance framework[[226]](#footnote-226); the 2017 World Commission on the Ethics of Scientific Knowledge and Technology Report on Robotics Ethics[[227]](#footnote-227); the 2019 Preliminary Study on the Ethics of Artificial Intelligence, which raised ethical concern about education, science and gender[[228]](#footnote-228); the Recommendation on Ethics of AI to be considered by UNESCO’s General Conference in 2021; and the World Commission on the Ethics of Scientific Knowledge and Technology Report on the Internet of Things.

In 2019, the United Kingdom’s National Health Service released a code of conduct with 10 principles for the development and use of safe, ethical, effective data-driven health and care technologies.[[229]](#footnote-229) In October 2019, *The Lancet* and *The Financial Times* launched a joint commission, The Governing Health Futures 2030: Growing up in a Digital World Commission, on the convergence of digital health, AI and universal health coverage, which is working between October 2019 and December 2021.

## Bioethics laws and policies

Bioethics laws and policies play a role in regulating the use of AI, and several bioethics laws have been revised in recent years with due recognition of the growing and irreversible use of AI within science, health care, and medicine. The French government’s most recent revision of its national bioethics law, enacted in 2019, sought to establish standards to address the rapid growth of digital technologies in the health care system. It included standards for human supervision, or human warranty, that requires evaluation by patients and clinicians at critical points in the development and deployment of AI. It also supported free, informed consent for the use of data and the creation of a secure national platform for the collection and processing of health data.

## Governance of health data

There are several types of laws and policies that govern the uptake, process, analysis, transfer, and use of health data. The Council of Europe’s Committee of Ministers issues a recommendation to Member States on the protection of health-related data in 2019.[[230]](#footnote-230) This also includes data protection laws, such as the European Union’s General Data Protection Regulation or the Health Insurance Portability and Accountability Act in the United States. There are also laws that govern the transfer of data between countries, including those defined in trade agreements, intellectual property rules related to the ownership of data, and the role of competition law and policy related to the accumulation and control of data (including health data). These topics are discussed in additional detail in the report.

## Regulatory standards

Regulatory standards for AI technologies should be enforced by health regulatory authorities, and ensure the safety, efficacy, and appropriate use of technologies in the delivery of health care. A WHO Expert Group that is developing regulatory standards for AI in health have identified six areas that regulators should consider when examining new AI technologies, including: documentation and transparency, risk management and the life cycle approach, data quality, analytical and clinical validation, engagement and collaboration, and data protection and information privacy. The governance of AI through regulatory frameworks, and the ethical principles that should be considered within regulatory standards, are discussed below.

# Section 7: Key ethical principles for the use of AI in health

Human dignity and the inherent worth of humans are the central value upon which all other ethical principles rest. The identification and introduction of ethical principles for the application of AI in healthcare and other domains is intended to guide developers, users, and regulators as they improve and oversee the design and use of such technologies.

An ethical principle is a statement that expresses a duty or a responsibility of stakeholders in the context of the development, deployment, and continuing assessment of AI technologies in health. The ethical principles in this document are grounded in set of basic ethical requirements that apply to all persons and that are treated as uncontroversial.

These requirements include:

* Avoid harming others (sometimes called "do no harm" or nonmaleficence).
* Promote the wellbeing of others where possible (sometimes called “beneficence”). Risks of harm should be minimized, while maximizing benefits. Expected risks should be balanced against expected benefits.
* Ensure that all persons are treated fairly which includes the requirement to ensure that no person or group is subject to discrimination, neglect, manipulation, domination, or abuse (sometimes called “justice” or “fairness”).
* Deal with persons in ways that respects their interest in making decisions about their life and their person, including health care decisions, in light of an informed understanding of the nature of the choice to be made, its significance, that person’s interests and likely consequences of the available alternatives (sometimes called “respect for persons” or “autonomy”).

From this list of fundamental moral requirements, additional moral requirements can be derived. For example, besides being recognised as a legal requirement in many countries, safeguarding and protecting individual privacy is important to enable a person’s ability to control sensitive information about themselves and self-determination (respecting their autonomy) and in avoiding harm.

These principles are intended to provide guidance to stakeholders about how these basic moral requirements should direct or constrain their decisions and actions in the specific context of developing, deploying, and assessing the performance of AI technologies for health. These principles are also intended to emphasize issues that arise from the use of a technology that has the potential to alter relationships of moral significance. For example, it has long been recognized that health care providers have special duties to advance these values with respect to patients because of the centrality of health to individual wellbeing, because of the dependence of patients on health professionals for information about their diagnosis, prognosis and the relative merits of the available treatment or prevention options, and the importance of the free and open exchange of information to the success of the provider-patient relationship. If AI systems are delegated clinical tasks that were once reserved for humans, then programmers that design and program such AI technologies should also adhere to these ethical obligations.

Clinicians, systems developers, health system administrators, and policy makers in health authorities and local and national governments can look to these principles for guidance about the responsible development, deployment, and evaluation of AI technologies for health. However, it falls primarily to national governments and health authorities to ensure that regulatory systems are in place to assign responsibility and accountability for satisfying these principles.

Ethical principles enunciated? here should *encourage and assist governments, public and private sector actors* to keep pace with the rapid evolution of AI technologies through legislation and regulation, should empower medical professionals to use AI technologies appropriately, and should assist corporations and programmers to follow such principles in the design of AI technologies.

Ethical principles should also be embedded within the *professional and technological standards* of AI. Software engineers already are guided by standards that address such things as fitness for purpose, documentation and provenance, and version control. Standards are required to guide a program’s interoperability and design, for ongoing education of those who develop and use such technologies, and for governance. Moreover, there are evolving standards for the evaluation and external audit of systems in the contexts of their use. In health computing, there are standards for system integration, electronic health records, system interoperability, implementation, and programming structures.

Though ethical principles do not always identify or clearly address limitations in the uses of such technologies, governments may ban or restrict the use of AI or other technologies when they violate or imperil the exercise of human rights, do not conform to other principles or regulations, or would be introduced in unprepared or other inappropriate contexts. For example, in many countries there may be a lack of data protection laws or inadequate regulatory frameworks to guide the introduction of AI technologies.

## Human Rights

The claim that there are certain basic moral requirements that must constrain and guide the conduct of persons can also be expressed in the language of human rights. Human rights are intended to capture a basic set of moral requirements for conduct to which every person is entitled regardless of race, sex, nationality, ethnicity, language, religion, or any other feature. These rights include human dignity, equality, non-discrimination, privacy, freedom, participation, solidarity, and accountability.

Efforts to enumerate these rights and to fortify their observance through explicit legal mechanisms are reflected in international and regional human rights conventions, including the Universal Declaration on Human Rights, the International Covenant on Economic, Social and Cultural Rights (including General Comment No. 14, which defines the right to health), the International Covenant on Civil and Political Rights, as well as regional human rights conventions, such as the European Convention on Human Rights, the American Convention on Human Rights, and the African Charter on Human and People’s Rights. Not all governments have acceded to key human rights instruments, may have signed but not ratified such charters, or may have expressed reservations to certain provisions, but in general, human rights enumerated in the aforementioned international instruments establish a baseline for the protection and promotion of human dignity worldwide.

Machine learning systems could advance the protection and enforcement of human rights (including the human right to health) yet could undermine core human rights such us non-discrimination and privacy. Human rights and ethical principles are intimately interlinked; because human rights are legally binding, they provide a powerful overarching framework by which governments, international organizations and private actors are obligated to abide. Private sector actors have a responsibility to respect human rights; this responsibility exists independently of state obligations. As part of fulfilling this responsibility, private sector actors need to take ongoing proactive and reactive steps to ensure that they do not cause or contribute to human rights abuses.

Nonetheless, the existence of this framework does not obviate the need for ongoing ethical deliberation. Indeed, much of ethics is intended to expand upon and complement the norms and obligations established through human rights. In many situations, multiple ethical considerations would be relevant and will require an exercise of weighing up and balancing to accommodate multiple principles that are at stake. An ethically acceptable decision depends on articulating the full range of appropriate ethical considerations, ensuring that multiple perspectives are factored into the analysis, and creating a decision-making process that stakeholders will consider fair and legitimate.

## Ethical Principles

This guidance identifies six ethical principles to guide the development and use of AI technology. One mechanism to make these principles operational is to integrate the principles (and human rights and legal obligations) into impact assessments (including human rights impact assessments) of AI technologies that should be carried out by governments, programmers, and health systems. . While ethical principles are universal, their implementation may differ in various contexts, depending on cultural, religious, and other societal factors.

Many ethical issues arising in the use of artificial intelligence and machine learning are not completely new. They have been arising or other applications of information and communication technologies in health, and whenever any computer is used to track a disease, make a diagnosis or prognosis. Computers running different programs have been performing these tasks long before AI became noteworthy. Ethical guidance and related principles have been articulated for fields like telemedicine and data sharing. Likewise, several ethical frameworks have been developed for AI in general (outside of the health sector – See Section 6). The ethical principles emphasized here are the ones identified by the WHO Expert Group as the most appropriate for using AI in health.

1. Protecting autonomy

The adoption of AI can lead to situations where decision making power could be or is in fact transferred to machines. The principle of autonomy requires that any expansion of machine autonomy does not undermine human autonomy.[[231]](#footnote-231)

In the context of health care, this entails that humans must remain in full control of healthcare systems and medical decisions. AI systems should be demonstrably and systematically designed to be in compliance with the principles and human rights with which they cohere, and more specifically they should be designed to assist humans to make informed decisions, whether it is medical providers or patients. Human oversight can vary according to the risks associated with an AI system, but should always be meaningful, and thus should track human values and moral considerations, effectively and transparently. In practice, this could include the ability to decide whether to use a system for a particular healthcare decision, to vary the level of human discretion and decision-making, to develop AI technologies that can provide decisions in a ranked order when appropriate (as opposed to a single decision), to have the capacity to override decisions made by AI systems, and to ensure that any autonomy of machines can be restricted and made ‘intrinsically reversible’.

Respect for autonomy entails related duties to protect privacy and confidentiality and ensure informed and valid consent through the adoption of an appropriate data protection legal frameworks. This must be fully supported and enforced by governments and respected by AI companies and their system designers, programmers, database creators, and others. AI technologies should not be used for experimentation or manipulation of humans in a health care system without informed and valid consent. The use of machine learning algorithms in diagnoses, prognoses, and treatment plans should be incorporated into the informed and valid consent process.

Data protection laws safeguard individual rights and place obligations on data controllers. Such laws are necessary to protect privacy and confidentiality of patient data and to establish patients’ control over their data. Construed broadly, data protection laws should also make it easy for people to access their own health data, and to move or share it as they like. Because machine learning requires large amounts of data – Big Data – these laws are ever-more important.

1. Promoting human well-being and the public interest

AI technologies should not harm people. AI technologies must satisfy regulatory requirements related to safety, accuracy, and efficacy. There must also be measures in place to ensure quality control and quality improvement. Thus, developers and users have a duty on an ongoing basis to measure and monitor performance of AI algorithms, to ensure AI technologies work as designed, and to assess if there is detrimental impact at the individual patient level or group level.

Preventing harm requires that the use of AI technologies does not result in any mental or physical harm. AI technologies that provide a diagnosis or warning that an individual is unable to address due to a lack of appropriate, accessible, or affordable health care, should be carefully managed and balanced against any “duty to warn” that might arise from incidental and other findings, and ensure that appropriate safeguards are in place to protect individuals from stigma or discrimination due to disease status.

1. Ensuring transparency, explainability, and intelligibility

Transparency requires sufficient information (described below) be published or documented prior to deployment of an AI technology, and such information should facilitate meaningful public consultation and debate. Such information should continue to be published and documented in a regular and timely manner after an AI technology is approved for use.

Transparency will improve system quality and protect patient and public health safety. For instance, system evaluators require transparency to analyse errors, and government regulators will rely on transparency to effect proper and effective oversight. An AI technology must be possible to audit if something goes wrong. Transparency should include, among other things, accurate information about the assumptions and limitations of the technology, operational protocols, data properties of the data (including methods of data collection, processing, and labelling), and algorithmic model development.

AI technologies should be explainable to the extent possible and considering the capacity of those to whom an explanation is directed. This is an argument for greater education for those who would request or require an explanation, and such educational information must be tailored to different populations, including for example marginalised populations. Many AI technologies are complex, however, and these might frustrate both the explainer and the one receiving the explanation. There is a possible trade-off between full explainability of an algorithm (at the cost of accuracy) and improved accuracy (at the cost of explainability).

All algorithms must be rigorously tested in real-world settings to ensure it meets standards of safety and efficacy. Such examination and validation should include an AI technology’s assumptions, operational protocols, data properties, and output decisions. Such tests and evaluation must be regular, transparent and of sufficient breadth to consider differences in the performance of the algorithm according to race, ethnicity, gender, age, and other relevant human characteristics. There should be robust independent oversight of such tests and evaluation to ensure they are carried out safely and effectively.

Health care institutions, health systems, and public health agencies should regularly publish information about: the decision-making that has been applied to support the adoption of an AI technology as well as how an AI technology will be periodically evaluated, the uses of AI technologies, known limitations, and role in decision making, which can facilitate external auditing and oversight.

1. Fostering responsibility and accountability

Humans require clear and transparent specification of the tasks the systems can perform, and the conditions under which they can achieve the desired level of performance; this helps ensure that users can deploy an AI technology responsibly. Although AI technologies perform specific tasks, it is the responsibility of various human stakeholders to ensure that they can perform those tasks and that they are used under appropriate conditions.

Responsibility can be enabled by the application of ‘human warranty’. Human warranty implies patient and clinician evaluation at critical points in the development and deployment of AI technologies. Human warranty applies regulatory principles upstream and downstream of the algorithm by establishing points of human supervision. Such supervision is usually focused on critical points identified in a process or dialogue among professionals, patients, and designers. The goal is to ensure that the algorithm stays on a machine learning development path that is both medically effective, can be interrogated, and is ethically responsible; it involves active partnering with patients and the public such as meaningful public consultation and debate[[232]](#footnote-232).

When something does go wrong with an AI technology, someone must be accountable. Appropriate mechanisms of redress should be adopted that enable questioning and redress for individuals and groups that are adversely affected by algorithmically informed decisions. This should include ensuring access to prompt and effective remedies and redress from government and companies that deploy AI technologies in health care. Redress can include compensation, sanctions where needed, and guarantees of non-repetition.

The use of AI technologies in medicine requires the attribution of responsibility within complex systems featuring numerous agents among whom there is distributed responsibility. When medical decisions issued by AI technologies harm individuals, responsibility and accountability processes must be able to clearly identify the relative role of manufacturers and clinical users in the harm. This is an evolving challenge and remains unsettled in most nations’ laws. Aside from legal liability, institutions have a duty to assume responsibility for decisions made by the algorithms they use, even if it not feasible to explain in detail how the algorithms produce its results.

To address a diffusion of responsibility, or in which ‘everybody’s problem becomes nobody’s responsibility’, a faultless responsibility model (‘collective responsibility’), which holds all agents involved in the development and deployment of an AI technology responsible, can be a means to encourage all actors to act with integrity and minimise harm. In such a model, the actual intentions of each agent (or actor) or their ability to control an outcome are not considered.

1. Supporting inclusiveness and equity

Inclusiveness requires that AI in health care is designed to encourage the widest possible appropriate use and access, irrespective of age, gender, income, ability, or other characteristics. Institutions (e.g. companies, regulatory agencies, health systems) should hire employees from diverse backgrounds, cultures, and disciplines to develop, monitor, and deploy AI. AI technologies should be designed by and evaluated with the active participation of those who are required to use the system or will be affected by it, including providers and patients, and such participants should be sufficiently diverse.

AI technology – like any other effective technology – should be shared as widely as possible. AI technologies must not only be available in high-income countries and for use in contexts and for needs that apply only in high-income settings – they must also take into account the types of devices, telecommunications infrastructure, and data transfer capacity in low- and middle-income countries. AI developers and vendors must also consider the diversity of languages, differently abled persons, and forms of communication around the world to avoid barriers to use. Industry and governments should strive to ensure that the pre-existing “digital divide” within and between countries is not widened and ensure equitable access to novel AI technologies.

AI technologies should not be biased. Bias is a threat to inclusiveness and equity because it represents a departure, often arbitrary, from equal treatment. For example, a system designed to diagnose cancerous lesions that is trained on one skin colour may not generate accurate results for patients with different skin colour, putting their health at increased risk.

Unintended biases which may emerge using AI should be avoided or identified and mitigated. AI developers should be aware of the possible biases involved in their design, implementation, and use and the potential harm that biases can cause to individuals and society. These parties also have a duty to take steps to address potential bias and avoid introducing or exacerbating health care disparities including when testing or deploying new AI technologies on vulnerable populations.

AI developers must ensure that AI data, and especially training data, avoids sampling bias and so is accurate, complete, and diverse. If a particular racial or ethnic minority is underrepresented in a dataset, it may be necessary to oversample from that group, relative to their population size, in order to ensure that an AI technology achieves the same quality of results in that population as in better-represented groups.

AI technologies should not exacerbate power disparities between providers and patients, or between companies that create and deploy AI technologies and those that use or rely upon them. Public sector agencies should have access to data gathered by private health-care providers. Everyone should be able to benefit from an AI technology, and not just technology providers. AI technologies should be accompanied by processes to provide patients with knowledge and skills to better understand their own health status and to communicate effectively with health care providers. In the future, health literacy will include an element of information-technology literacy.

Finally, AI technologies must be monitored and evaluated for the effects of their use, including disproportionate effects on specific groups of people wherein such impacts mirror or exacerbate existing forms of bias and discrimination. Special provision must be made for the protection of the rights and welfare of vulnerable persons. There must be mechanisms of redress if such bias and discrimination emerge or are alleged.

1. Promoting AI that is responsive and sustainable

Responsiveness requires designers, developers and users to continuously, systematically and transparently examine an AI technology and determine whether it is responding adequately, appropriately and according to communicated and legitimate expectations and requirements in the context in which it is used. That is, the identification of a health need entails that institutions and governments respond to the need and the context with appropriate technologies and an aim to achieve the public interest in health protection and promotion. When an AI technology is ineffective or engenders dissatisfaction, for instance, the duty to be responsive requires an institutional process to address the problem and resolve it.

Responsiveness also requires that AI technologies be consistent with wider efforts to promote health systems, environmental and workplace sustainability. AI technologies should only be introduced if the technologies can be fully integrated and sustained within a health care system. Too often, especially in under-resourced health systems, new technologies are unused or are not repaired or updated, thereby wasting scare resources that could have been invested in proven interventions. Furthermore, AI systems should be designed to minimise its ecological footprint and to increase energy efficiency, so that use of AI is consistent with society-wide efforts to reduce the impact of human beings on the earth’s environment, ecosystems, and climate. Sustainability also requires governments and companies to address anticipated disruptions in the workplace, including training for health care workers that must adapt to the use of AI, and addressing job losses due to the use of automated systems for routine health care functions and administrative tasks

# Section 8: Liability regimes for AI in health

Even as the performance of machine learning algorithms improves, there will be errors and mistakes, for example whether due to training of an algorithm with incomplete data, programming mistakes, or security flaws. Lawmakers and regulators should update safety rules and frameworks and ensure these regulations are integrated into the design and deployment of AI-guided technologies. Liability rules used within clinical care and medicine should also be modified and updated to consider the unique challenges posed with using AI in health care. This section examines how liability regimes could evolve, approaches to compensation, specific considerations for low- and middle-income, and the role of international institutions and organisations.

## Liability for use of AI in clinical care

The use of AI to support or augment clinical decision-making leaves several questions unresolved. Should a doctor be held at fault if she follows a suggestion of an AI technology which results in a medical error? Or when the clinician ignores the suggestion of an AI technology that would have avoided morbidity or mortality? Answering these questions depends largely on making other choices, including what types of behaviour a legal system wishes to encourage or discourage, and what will be the standard of care as the use of AI becomes more commonplace in clinical practice.

One choice is whether liability rules should encourage clinicians to rely upon AI to confirm their clinical judgments or to encourage doctors to deviate from their own judgments if an algorithm arrives at an unexpected conclusion. If liability rules penalise health care providers for relying on the conclusions of an AI technology that prove to be incorrect, clinicians may only use an AI technology as a means of confirming their own judgment. While this may shield providers from liability, it discourages use of AI to its fullest potential, which is to augment and not just validate human judgment.[[233]](#footnote-233)

On the other hand, if doctors are not penalised for relying upon an AI technology, even if it runs counter to their own clinical judgment, it could encourage doctors to make wider use of AI technologies as a means of improving patient care, or at least consider the use of AI technologies to challenge their own assumptions and conclusions.

A decision of whether a doctor will employ AI also depends upon what is the prevailing standard of care. If AI technologies are viewed as deviating from or are not recognised as within the standard of care, doctors will be discouraged from using such technologies since otherwise meeting the standard of care is a defence (though not absolute) to medical error. If the standard of care expects or requires the use of AI-based technologies, then it would essentially mandate physicians to integrate such technologies into clinical practice.[[234]](#footnote-234)

A separate but related issue is the liability of hospitals and health care systems that select specific technologies. A hospital or health care provider could be held liable for failing to exercise due care in the selection of AI technologies, or in how it maintains, introduces, or uses such technologies in a hospital.[[235]](#footnote-235) More generally, a hospital could be held vicariously liable for the errors made by clinicians within the scope of their employment. This encourages hospitals to both exercise due care in the technologies that are selected and in ensuring that clinicians have clear guidance on how to use such technologies to both improve patient care and avoid errors that result in legal liability for the clinician and hospital.[[236]](#footnote-236) An analogous approach would be to establish hospital liability via ‘negligent credentialing’ – wherein hospitals are liable for not adequately reviewing credentials and practice history of health workers and physicians, hospitals would have a similar duty when introducing AI.[[237]](#footnote-237) Yet this also depends on hospitals and health systems having both the necessary information and tools to identify appropriate AI technologies for clinical use.[[238]](#footnote-238)

## Are machine learning algorithms products?

Even as AI technologies (and the software that powers AI) are integrated into or replace medical devices, there is uncertainty as to whether such technologies can be characterised as products. Product liability, which holds the manufacturer or developer of a technology or a good to account even if not at fault, is a form of strict liability, or the imposition of liability even in the absence of negligence, recklessness, or intent to harm.[[239]](#footnote-239)

Until now, many jurisdictions have been hesitant to apply traditional product liability theories to healthcare software and algorithms. Product liability could apply insofar as an algorithm is integrated within a medical device or diagnostic. Both U.S. and European courts, as well as new regulations, regard software as a medical device based on its intended use.[[240]](#footnote-240) However, developers may escape liability because in many cases the ‘actual uses’ of a product are not the same as ‘intended uses’, even if some of the ‘actual uses’ could be foreseen.[[241]](#footnote-241) Product liability may also not apply if an AI algorithm is construed as a service and not as a product.

Expanding the remit of product liability may be desirable since otherwise a patient may have difficulty securing compensation (e.g. if a clinician followed the standard of care), and since bringing forward a case that seeks to assign fault to a developer may be too costly and complex. It could also ensure developers take all available steps during the development of an algorithm to reduce the likelihood of error, including using diverse and complete data sets to train an algorithm.

Assessing until what point a developer can be held strictly liable for the performance of the algorithm is complicated by the growing use of neural networks and deep learning within AI technologies, which means that the algorithms may perform differently over time as it is used commercially in a clinical setting.[[242]](#footnote-242) Holding the developer accountable for any error may ensure that a patient can be compensated, but such continuing liability may discourage the use of increasingly sophisticated deep learning techniques, and thereby produce less beneficial observations and recommendations for use in medical care. Liability may depend in part on how much control the developer continues to have over an AI technology, which would determine whether, for example as applied in many EU Member States, whether a ‘development risk defence’ would allow a developer to avoid strict liability.[[243]](#footnote-243)

Finally, even if a developer could be held strictly liable under a product liability framework, a developer could avoid liability under the learned intermediary doctrine, which limits recovery against a manufacturer where a doctor prescribes drugs or devices[[244]](#footnote-244), and wherein the manufacturer has transferred the drug or device to a physician with adequate accompanying information such as warnings about risks.[[245]](#footnote-245) With such adequate warnings, subsequent decisions made by the physician, as the 'learned intermediary’, break the line of causation between a product developer and the patient who has suffered harm.[[246]](#footnote-246)

## Compensation for errors

A liability regime for AI may not be to assign fault, especially when algorithms increasingly evolve in ways that neither developers nor providers can fully control. In other areas of healthcare, compensation is provided without assigning fault or liability, including for example, medical injuries that result from adverse effects caused by vaccines.[[247]](#footnote-247) Such no-fault, no-liability compensation funds could be supplemented by requiring developers (or companies that are developing such technologies) to obtain insurance that would trigger pay-outs upon injury or to pay into an insurance fund, with a separate fund providing compensation in cases where an insurance pay-out is not triggered. In New Zealand, for example, patients seek compensation for medical injuries through a no-fault, no liability scheme, wherein injured patients receive government-funded compensation, thereby giving up the right to seek damages except in rare cases of reckless conduct.[[248]](#footnote-248)

## The role of regulatory agencies and pre-emption

AI technologies, like drugs and devices, will increasingly be subject to regulatory oversight and validation prior to use, especially as their uses expand and as clinicians increasingly rely upon AI technologies. If a commercial algorithm is approved by a regulatory agency, the doctrine of pre-emption may apply, or that a decision taken by a federal agency (or central government agency) to validate a technology supersedes any cause of action guided by civil laws[[249]](#footnote-249). Yet pre-emption may not always be relevant – especially if the regulatory pathway used to approve AI technologies is abbreviated or regulatory approvals are based upon less information of how the algorithm was constructed and trained (and may perform over time)[[250]](#footnote-250). Second, just as a developer may not be held accountable for an algorithm as it evolves, so a doctrine of pre-emption may not be applicable if an algorithm evolves after a regulatory agency has approved the technology.

## Considerations for LMICs

Some of the emerging literature, policy frameworks and court decisions to guide liability regimes arise from the U.S. and European Union, which is also where such technologies have been actively deployed (or there are plans to do so on a wider scale). Yet such guidance may not be as applicable for liability frameworks in low- and middle-income countries.

First, liability may be the first and only line of defence against errors made by machine learning technologies. Many LMICs still lack sufficient regulatory capacity to assess drugs, vaccines, and devices, and the onrush of machine learning technologies could leave these agencies unable to accurately assess and regulate such technologies for the public good. Concerns that such technologies may not operate as intended are heightened by the lack of available, quality data that can be used to train such algorithms.

Second, in many LMICs, access to justice can be unavailable or too expensive (or protracted) for many injured parties, making it not just difficult to provide compensation for those harmed by AI technologies, but also unlikely to serve as a deterrent to those who have responsibility for the development and deployment of such technologies in clinical settings. This lack of protection is often heightened for marginalised populations that are often excluded from any redress within a legal system. It may also be difficult to seek out remedies if the AI technology was developed internationally by a company or developer that has no ‘physical presence’ where the harm occurs.

Third, there may be other legal concerns related to liability that arise in LMICs that are not contemplated or considered in high-income economies. As noted above, this can include a lack of training data that ensures the algorithm performs accurately for patients with a different physical appearance or connectivity challenges that can undermine the use of a technology reliably and safety. There could also be additional rationales that justify the use of AI-technologies in low-income settings that may not be relevant in high-income settings (e.g. except for an AI-guided diagnostic, there is no human or lab capacity to provide a timely and accurate diagnosis for a particular health need).

Therefore, even if legal systems in some low- and middle-income countries may wish to borrow certain approaches as AI technologies are introduced in clinical settings, the means of compensating those who are harmed by such technologies, the means to hold companies accountable for the products that they develop, and the risk-benefit calculations for using or avoiding AI technologies, will require novel approaches and independent thinking.

Recommendations

Legal liability will evolve as it relates to the use of AI in health, including the duties of care owed by providers, hospitals, and developers. Since digital technologies can be developed in one place and deployed worldwide with few barriers, international agencies and institutions should support the efforts of national governments to manage liability.

1. International agencies should ensure that its own clinical guidelines keep pace with the rapid use of AI technologies. Many countries and health care providers rely on the guidance of international institutions, and especially WHO, to provide up-to-date guidance that can inform the local standard of care. An evolving standard of care that considers the use of AI (or its avoidance) can ensure clinicians can act in the best interest of patients while also being protected from errors.
2. WHO and other international agencies can play a role in assessing, validating, and recommending AI technologies that should be used to support clinical care, and to also play an on-going role to monitor and periodically assess if such recommended AI technologies are still functioning according to what was intended. This can encourage the use of pre-emption where it is appropriate, while not expanding its use for technologies that have not been properly validated prior to use in countries that are unable to independently and accurately assess such technologies through their own regulatory structures.
3. WHO and other agencies can seek to establish international norms and legal standards, whether through a separate legal framework or through existing legal frameworks, that can ensure national accountability to protect patients from medical errors. Such efforts could be done through a wider effort by WHO (with partner agencies) to develop model legislation for the use of AI in health (See Section 10.X).

# Section 9: Elements of a governance framework for AI in health

Existing human rights standards, data protection laws, and ethical principles, are all necessary to guide, regulate, and manage the use of AI for health by developers, governments, providers, and patients. There has been a focus by many stakeholders to generate ethical principles for AI in health, and WHO hopes that the principles suggested in this report (See Section 7) can encourage consensus.

Yet the use of AI in health will introduce several challenges that ethical principles cannot resolve, especially as many risks and opportunities with the use of AI are not yet well understood or will change over time as AI technology improves. Furthermore, many principles, as well as existing laws and standards, have often been devised by and for high-income countries. LMICs will face many additional challenges to introduce new AI technologies, which will require not only awareness and adherence to ethical principles, but also the development of appropriate governance for the first time. This section examines several areas for which governance of AI in health will be needed.

## Governance of Data

As the types, quantity, and applications of health data, including for commercial use, have grown, a patchwork of approaches to govern health data have emerged. One overarching challenge is how to safeguard an individual’s privacy and autonomy through the control of his or her data without undermining the purported benefits of the collection and use of health data. Such considerations are likely to apply whether the use of health data is for AI or use of a relational database.

Existing mechanisms that enable individual control of data, such as informed consent, a duty of confidentiality, and deidentification, may not be sufficient to safeguard health data, and may also interfere with positive uses. (See Section 4.X Data Collection and Use). This section examines other principles and mechanisms to enable appropriate governance of health data and some challenges with these options.

### Evolving approaches to consent

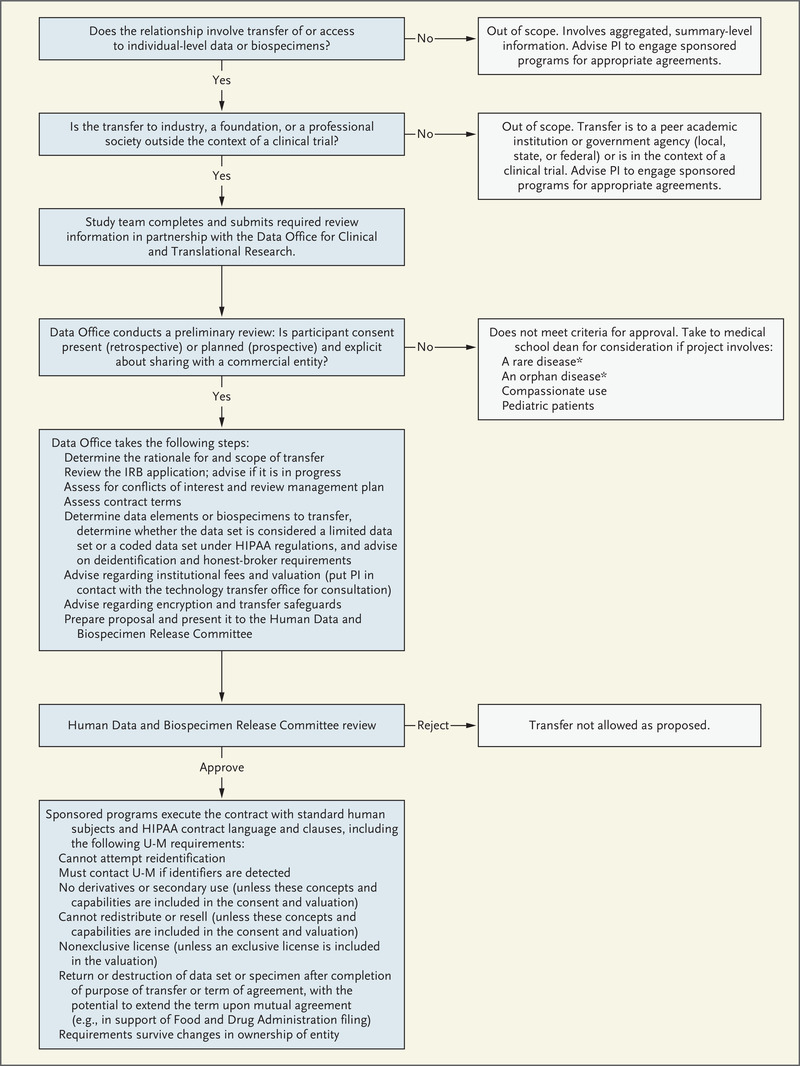
There are other types of consent that could be more workable as the quantity of health data, and the possible uses, expand. Any consent needs to be preceded by a clarification of the consequences of providing consent, including for example which data will be used as well as how it will be used. One form of consent that could improve individual control and choice is electronic informed consent, which uses online forms and communication to facilitate consent for different uses of health data.[[251]](#footnote-251) Electronic informed consent could allow users to have a better understanding of how such data will be used and could improve a user’s control of his or her data, but the content should be presented in a simple format that is easily accessible to the general public, including through the use of illustrations. This can help to ensure that consent is given freely and that risks are understood.[[252]](#footnote-252) Sage Bionetworks, for example, has established its own [toolkit and information guide](https://sagebionetworks.org/tools_resources/elements-of-informed-consent/) for facilitating electronic informed consent.[[253]](#footnote-253) A second approach is dynamic consent, which applies an open-ended approach to consent that allows for users to periodically modify consent over time for uses of data that they may wish to permit and those that they would like to specifically exclude.[[254]](#footnote-254)

Governments may also wish to redefine when consent could be waived because other public interest considerations are implicated. This would imply in certain situations that there is a duty to share health data for the benefit of the public or for other non-monetary benefits, such a better quality of life or health.[[255]](#footnote-255) In other words, consent would be waived because the data is conceived as a public good for which data can be ‘conscripted for publicly minded uses.’[[256]](#footnote-256) This could include situations where there are clear public health benefits for use of data that otherwise would be undermined by too many individuals opting out of sharing such data, although the burden of demonstrating that consent is undermining such a benefit should rest with an entity that seeks to avoid consent. It could also imply that obtaining health data without the specific consent of the individual is justified if there is a benefit, broadly distributed, that outweighs specific violations of one’s privacy where the risks are ‘low’.[[257]](#footnote-257)

Yet such a system of measuring benefits and risks could always lead to the unconsented sharing of data since medical benefits – whether improved surveillance of disease or the development of a new drug, could always be gauged as more important than ‘low-risk’ violations of privacy that may materialise from the use of health data. Additionally, whether patients can share in the benefits may depend, when such data may be provided to commercial actors (which itself may require heightened scrutiny and consent in nearly all circumstances – see below), on commercial actors sharing these benefits widely, wherein often costs of medical products and services are often neither affordable nor available (see below).

Conscripting health data for the public good is especially questionable when the data is shared with a commercial entity, whatever the intended product or service. Recent known instances of patient data shared by not-for-profit entities or academic institutions with private companies without their consent has generated significant concern, especially as patients were neither notified that such data was shared, for what purpose, and with what private entity. Several not-for-profit that have collected health data into centralised biorepositories have now put principles for informed consent for such data sharing into practice, which both ensures a data contributor to have a robust understanding of consent at enrolment and disclosure of any industry partner at the time of consent, as well as prospective, explicit consent process to allow for future secondary use of data.[[258]](#footnote-258) Even with these additional standards in place at one such biorepository operated by the University of Michigan, only 6 of 70 project proposals over a two-year period were denied by a review committee, and this was only due to an absence of adequate initial consent.[[259]](#footnote-259) Figure X1 provides a decision-making process for the approval of such use of health data at Michigan Medicine.

Figure X1: Michigan Medicine Human Data and Biospecimen Release Committee Rubric[[260]](#footnote-260)



A more controversial issue is creating a market or system upon which individuals can buy and sell their health data. Health data is sensitive personal data, which is linked to human agency and dignity. Creating a system that facilitates the sale of personal data can lead to a two-tier society, wherein the wealthy can protect their own rights and afford to limit the use of their data by other parties, whereas people living in poverty may feel compelled to hand over their data to gain access to social or material benefits. Such a system that deploys the sale of data would also be in contravention of several human rights standards.

Furthermore, while the sale of data may contribute to data uses that are commercially valuable but less beneficial to individual or public health; the data market itself may not function properly and undervalue an individual’s data; and the sale of one’s data may lead to the future loss of control of an individual over his or her health data. Such challenges that could emerge with health data have emerged with the commercial sale of blood and related products (plasma).

### Principles and approaches for the protection, collection, and appropriate use of health data

The following are different principles and approaches that have been established for the appropriate protection, collection, and use of health data. These approaches are not always mutually exclusive and can be mutually reinforcing.

One set of principles that guide the collection and use of data are focused upon data quality. The FAIR principles, or findability, accessibility, interoperability, and reusability, are designed to maximise the value and quality of the data[[261]](#footnote-261). These principles do not address concerns with privacy and control of health data, nor do they address concerns with incomplete data that may result in bias and discrimination.

Some governments that are prioritising the collection and use of health data for commercial and public sector interventions have established principles for data collection and use. The United Kingdom’s National Health Service has established five guiding principles for a framework in which data can be used to support health innovation. One notable commitment under these principles is transparency, or that any commercial arrangements should be transparent, clearly communicated, and not undermine public trust and confidence.[[262]](#footnote-262) As discussed below however, many such agreements between the public and private sector are not transparent, raising serious concerns if such agreements are also accompanied by concerns over financial conflicts of interest.

There are also other forms of transparency that could be required. This could include transparency of the sources and methods of obtaining and processing data, how and why certain types of data are excluded, methods that are used to analyse data, and open discussion in publications of data bias.

In New Zealand, an independent ministerial advisory group funded and appointed by the Government engaged in a wide-ranging consultation to build an ‘inclusive, high-trust, and high-control data-sharing ecosystem’.[[263]](#footnote-263) The guidelines include eight questions that summarise what matters most to people with respect to building trust in data use - or whether the use of data provides value, protection, and choice for an individual. The eight questions are in Figure X2.

Figure X2: Eight questions addressing building trust in data use



Although these guidelines are voluntary, each entity that seeks to use such health data have been asked to publish their answers to these questions so that individuals who may provide such data can determine whether the values of the entity aligns with their preferences.[[264]](#footnote-264)

*Data protection*

From a human rights perspective, an individual’s personal data should always be controlled by that individual. An individual’s rights in his or her own data is grounded in concepts that are related to but distinct of ownership, including control, agency, autonomy, and human dignity. Control may include different approaches to individual consent (see above), as well as collective mechanisms to ensure that such data is used appropriately by third parties (see below). Data protection laws are ‘rights-based approaches’ that establish the rights of data that both protect the rights of individuals and establish obligations for data controllers, and which include sanctions and remedies in case of actions that violate these statutory rights. Data protection laws can also provide for exceptions for non-commercial uses by third parties.

Data protection frameworks and regulations are essential for managing the use of health data. The European Union’s General Data Protection Regulation (GDPR), while it applies to citizens of the European Union, also has a global reach because it applies to EU citizens around the world, and therefore can affect the standards and practices of many non-EU entities. GDPR seeks to minimise the amount of data that is collected about an individual to what is necessary, to allow for collection of data only for listed, legitimate purposes or with an individual’s consent, and notify individuals of the receipt of data. Health data are protected under GDPR unless an individual provides specific consent or if the use of data meets certain exceptions, such as for health-related operations or scientific research.

GDPR also introduced ‘data portability’ – or the right of individuals to obtain one’s personal data in a machine-readable format from one controller that can then be sent to another controller.[[265]](#footnote-265) Data portability could, depending on how it is implemented within the European Union, provide individuals with control over their own data, and also enable individuals to share health data with additional entities that could make use of such data. Data portability could help to decentralise the control and distribution of data, and subject to appropriate implementation, could introduce a novel form of data management that both fosters oversight and innovation.

Data protection regulations are enforced through data protection authorities, which develop and administer regulations, provide guidance and technical advice, and conduct investigations. Some governments have introduced additional supervisory authorities to facilitate the use of health data. The UK established a National Data Guardian in 2014 for appropriate management of health data (with respect to confidentiality) while also seeking to improve the use of health data for beneficial purposes.[[266]](#footnote-266) In 2018, the entity was granted the power to issue official guidance on the use of data for health and adult and social care in England.[[267]](#footnote-267)

### Data sovereignty

Several indigenous communities have sought to establish control over their own data through the exercise of data sovereignty. Māori (the indigenous population of New Zealand) have introduced principles for data sovereignty that establishes, for example: control over data, including to protect against future harm, accountability for those who collect, use and disseminate such data to the people who provide such data, an obligation for such data to provide a collective benefit, and free prior and informed consent, which, when not possible, should be accompanied by stronger governance measures.[[268]](#footnote-268) Māori also recognise a need to balance the individual rights of data holders with benefits for the community, and that in some situations the collective rights of the Māori will prevail over those of individuals.[[269]](#footnote-269)

First Nations groups in Canada have also outlined principles that promote their own sovereignty over data. The guiding principles for the First Nations have four elements: ownership of data, control of data, access to data, and possession of data. There is an expectation over time that First Nation tribes will establish protocols that enable wider access to such data for uses that benefit First Nation tribes.[[270]](#footnote-270)

### Data cooperatives

A data cooperative seeks to provide data subjects with control over their own data through a mechanism that stores health data for the members of a cooperative. Data cooperatives allow for secondary uses of such data, while also allowing members of the data cooperative to collectively decide how such data should be used.[[271]](#footnote-271) Data cooperatives allow members to develop common ethical standards and some data cooperatives may develop their own tools and applications so that the data can be used beneficially. [[272]](#footnote-272)

### Data sharing

As health data has proliferated, governments have taken several steps to improve the sharing of data for the benefit of scientific research (but also for the furtherance of commercial development of health AI and other health applications). In 2014, the United States National Institutes of Health introduced a Genomic Data Sharing Policy, which is intended to encourage the ‘broad and responsible sharing of genomic research data’.[[273]](#footnote-273) Subsequently, new legislation enacted in the United States in 2016, the 21st Century Cures Act, expanded the remit and created statutory authority for the Director of the NIH to require those receiving awards from the NIH to share data as well as to provide the means for the NIH to enforce data sharing.[[274]](#footnote-274)

The Act also provides the means to improve the ability of individuals to have access to their own health data, which was finalised in rules issued by the US government in 2020. The rules create requirements for health IT providers to introduce standards-based application programming interface (API) to support an individual’s use and control of electronic health information.[[275]](#footnote-275) Health IT providers must meet three requirements for its API to be certified – it must meet certain technical programming standards that ensure interoperability, must be transparent, and must be ‘pro-competitive’, or promote the efficient exchange, access, and use of health data.[[276]](#footnote-276) The requirements that have been placed on health IT providers – such as anti-blocking or interoperability requirements – point to how governments can mandate and manage commercial activities with respect to the use of AI (and other technologies) in health care.

### Data hubs

There are numerous data hubs that pool different types of health data so that it can be used by third parties (the types of third parties differ depending on the type of data hub). Several data hubs sponsored by governments have emerged. In the United States, two different government-led data hubs are the Precision Medicine Initiative (All of Us)[[277]](#footnote-277), and the U.S. Department of Veteran Affairs health data hub. The European Union is in the process of establishing a European Health Data Space, which is intended to facilitate the exchange and sharing of health data (e.g. health records, genomics, registries) for primary and secondary purposes such as the delivery of primary care and the development of new treatments, medicines, medical devices, and services, while also ensuring people have control of their own health data.[[278]](#footnote-278)

There are also not-for-profit efforts to establish data hubs, such as the efforts of Health Data Research UK, an independent, not-for-profit organization of 22 research institutions across the United Kingdom, whose aim is to collect health data throughout the country and to make it available to public and private entities to understand diseases and find ways to prevent, treat and cure them. Principles of Participation have been developed in consultation with policy makers, the NHS, industry, and the public.[[279]](#footnote-279)

### Federated data

There has been significant growth and support of federated data systems, including collaborations between research institutions, governments, between the public and private sector, and within the private sector. Federal data sharing has been summarised as ‘a promising way to enable access to health data, including genomic data, that must remain inside a country or institution because of their sensitivity.’[[280]](#footnote-280) Data never leaves each participating organisation that hold such data but does allow authorised users to perform queries within a federated network. Thus, the queries allow a user to access such data, for example, to train an algorithm. Proponents have noted that federated data systems allow for each entity to govern the use of its data and that such an approach preserves privacy and security.[[281]](#footnote-281) While federated data sharing may facilitate analysis of larger data sets while maintaining local control, it does not overcome concerns that informed consent is not sought for secondary uses of data.[[282]](#footnote-282)

### Data sharing and data partnerships with the private sector

One of the more difficult questions with the creation of such government, not for profit, or academic data hubs is how such hubs should work with companies, either in accepting data from such companies that could improve the quality of the data, or allowing companies to make use of such data for training or validation of their algorithms. When commercial entities make use of such data, there are concerns, some of which have materialised, that people may have not knowingly given consent to such use of their health data for a commercial purpose. There are additional concerns that such agreements, beyond not informing individuals that data has been provided to third parties for commercial purposes, are not made transparent and disclosed to the public, including the private sector counterparties to such agreements.

Thus, numerous agreements signed between the Mayo Clinic, a major health system based in the United States, with sixteen technology companies, provided the Mayo Clinic with a ‘revenue stream and generated crucial insights for health tech firms eager to commercialise digital products and services.’[[283]](#footnote-283) In some cases, the Mayo Clinic not only shared data with a company but subsequently took an equity stake in these same companies (which would provide the Mayo Clinic with additional revenue). De-identified patient data was shared without requiring consent or even notification of people who supplied such health data for products under development. The names of eight of the firms that signed agreements were not disclosed, and none of the contracts signed between the Mayo Clinic and its technology partners were made public.[[284]](#footnote-284)

In other cases, physicians or scientists within healthcare systems that may have access to raw data that was provided to health technology firms have also founded or invested in such companies. A 2018 investigation found that board members and senior executives at Memorial Sloan Kettering Hospital had either founded or invested in an AI start-up to improve cancer diagnosis, and had also used the hospital’s trove of 25 million patient tissue slides (and six decades of pathology research) for the company’s benefit (Memorial Sloan Kettering had also taken an ownership stake in the company) without open bidding or transparent consideration of whether such data should be shared.[[285]](#footnote-285)

Finally, there are also efforts of companies, on their own, or in collaboration with other companies, to develop health data hubs that may involve combining the data of one or more companies into a common pool of information which can then be used to generate specific products and services. Such partnerships, while they may result in the development of useful products and services, raise concerns about the transparency of such activities, oversight of activities, and whether or not such private carriers of data will seek consent or at least engage those communities and individuals which have provided such data.

Recommendations

1. As the quantity of health data proliferates, new forms of informed consent, such as electronic informed consent and dynamic consent, should be considered as a means of providing individuals with control over their health data.
2. Governments can facilitate informed consent by requiring entities seeking health data to publish their values for the potential use of health data provided by the public. Data protection regulations are essential for managing the use of health data and promote individual autonomy, control, dignity, and agency. Such laws can provide exceptions to informed consent if data is used for non-commercial purposes. Data protection laws should enable data portability for individual users.
3. There may be limited situations in which data can be conscripted for the public good that does not require the informed consent of individuals for such secondary uses. Informed consent should be sought for commercial uses of health data, or there must be robust, prospective informed consent for any such secondary use of data. The burden to justify any exceptional use of health data without informed consent falls to companies to make, and such use of data must be accompanied by transparency of terms and conditions of such agreements, must produce outcomes that are in the public interest and should avoid any commercial conflicts of interest between the entity providing such data and any entity receiving such data..
4. Mechanisms that enable community control over data should be supported. Such forms of community control include efforts to establish data sovereignty or the establishment of data collectives.
5. Data hubs created by governments can play an important role in bringing together high-quality, unbiased data that can contribute to significant scientific and public health advances. Such data hubs must meet the highest standards of informed consent when such data may be used by the private sector, should be transparent in its agreements with companies, and should seek to ensure that the outcomes of such data collaborations provide for the widest possible public benefit.

## Control and Benefit Sharing

The application of big data and artificial intelligence in health care raises several questions of how to define data control and associated intellectual property (IP) rights with medical big data. This includes questions of asserting exclusive rights over health datasets, algorithms and products that incorporate AI, as well as the outcomes of such AI-based technologies, including medicines and diagnostic technologies. There are also several wider questions that need to be resolved, including whether health big data can or should be exclusively controlled by individuals through an appropriate form of governance or by entities that may aggregate such data (control of personal data is discussed separately above).

A separate question is whether novel products that are created solely by a machine can be owned, and if so, whether such ownership rights are conferred to the machine or assigned to an entity that created or controls a machine? There are also questions of how to ensure that the public’s contribution to the development of new AI technologies, such as investments in the development of algorithms, the contribution of data by individuals and health systems, and the creation of health data hubs that are accessed by private actors to develop new AI technologies, can be appropriately valued. Finally, if AI technologies are increasingly protected by exclusive rights, there are wider questions as to whether such technologies will be available, appropriate, and affordable in low- and middle-income countries.

### Control over and benefit sharing of Big Data

The central role of big data for AI, including medical big data for health-care associated uses of AI, has led some to label data as the new ‘oil’, or a valuable commodity that will create increased commercial conflict over its control, use, and access.[[286]](#footnote-286) Such a framing has been critiqued as unhelpful and conceptually inaccurate for several reasons.[[287]](#footnote-287),[[288]](#footnote-288) Unlike oil, the supply of data is virtually infinite and can be re-used in other contexts that can be valuable for commercial or non-commercial applications. There is at least the possibility of control and consent of the use of one’s data. And while oil’s intrinsic value is captured once it is extracted or drilled (subject to processing and refining), data itself is not intrinsically valuable unless data science is applied to generate something of value.

A separate framing considers not so much the commercial value of data but its central importance to the development and deployment of AI-based applications. In such a framing, data is ‘oxygen’, or an indispensable resource that is a part of the public infrastructure needed for AI and data science to serve the public and private sector.[[289]](#footnote-289) Whether big data should be ‘oil’ or ‘oxygen’ (or neither) rests in part upon whether there can and should be exclusive rights associated with Big Data, who should have such exclusive rights, and to what extent should such impede others from access to and use of such data for public or private uses.

There are several types of intellectual property rights that may apply to data, including trade secret protections, copyright, database rights (albeit in only a few jurisdictions), and in rare circumstances, patent rights (data “as such” is not patentable but “functional” data that is used in technical applications may be)[[290]](#footnote-290),[[291]](#footnote-291) It is beyond the scope of this publication to discuss the various IP rights that could apply to large data sets or to Big Data, yet such rights, if they are to be expanded (or minimised) as it relates to large data sets or Big Data depend upon broader policy objectives and ethical considerations.

There is a tension between the sharing of data and the commercial prerogatives that are protected with trade secret frameworks. On the one hand, conferring IP rights for biomedical Big Data could discourage the open sharing of Big Data that is necessary to advance scientific progress and the development of AI within health care and medicine[[292]](#footnote-292).[[293]](#footnote-293) Public or private ‘owners’ of health big data may not grant third parties the right to use such data to develop novel AI technologies, thereby undermining open innovation[[294]](#footnote-294) and providing commercial entities with power to exclude competitors or engage in rent-seeking. On the other hand, a lack of IP rights in biomedical Big Data could discourage some commercial investments.[[295]](#footnote-295) While the 21st Century Cures Act, enacted in the United States in 2016, takes several steps to encourage the sharing of data (see above), it also asserts that proprietary interests trump data sharing interests, and that the ability of the U.S. government to mandate data sharing is limited by policies that prioritise the protection of trade secrets, proprietary interests, confidential commercial information, and intellectual property rights.[[296]](#footnote-296)

An additional concern is whether sharing of health data by communities, health systems, or governments in LMICs will result in the sharing of benefits, especially if such data is used for commercial applications of AI.[[297]](#footnote-297) The lack of benefit sharing may be either because there are no legal conventions or frameworks that mandate benefit sharing for uses of big data, or because entities that negotiate benefit sharing on behalf of LMICs may not have the legal capacity to do so or are negotiating from a weaker position.[[298]](#footnote-298) Benefit sharing may not only include equitable access and availability of technologies that emerge from the sharing of health big data, but also ensuring that there are enough investments in digital infrastructure, research capacity, training and infrastructure that ensures that the benefits of AI and big data are also generated by researchers and companies based in LMICs.[[299]](#footnote-299)

Thus, while intellectual property rights could be adjusted on a case-by-case basis to encourage open innovation, investment, or benefit sharing, it may be that control (and IP rights used to assign control) are inappropriate to encourage the widespread use and application of biomedical big data when taking into account numerous competing considerations, including an individual’s right to privacy and control, society’s interest in scientific progress and the development of AI-guided technologies, commercial interests to exploit such data for profitable activities, and the interest of data contributors (communities, health systems, governments) to share in the benefits that are generated by third parties.[[300]](#footnote-300)

One recommendation has been to focus not on recalibrating or introducing new IP rights, which could impede data sharing or could intensify competing claims to control of data, but instead to establish a legal framework based on custodianship.[[301]](#footnote-301) Custodianship, or responsible oversight infused with ethical values, can ensure access to data, promote fair data sharing practices, and preserve privacy. While those who provide data maintain limited control, certain decisions are delegated to data custodians that have custodial rights, and not control (or IP rights) over big data. Such custodial rights can include protecting data privacy of those who contribute, disseminating research findings, ensuring freedom of scientific enquiry, and providing attribution to those that invest in creating databases and agreeing on terms of use and access.[[302]](#footnote-302)

### Ownership of AI-based products, services, and methods

Products and services that may be created using AI and big data, whether new medicines or new diagnostic methods, could be patented or subject to other IP rights.

One overlying concern with the patenting (or other forms of ownership) of AI-generated inventions and outcomes is that IP rights can exclude affordable access to such products or services; wherein patent holders may engage in rent-seeking behaviours as a means to recuperate investments and earn outsized profits. Since novel medicines, diagnostic methods, or other products and services developed with the use of AI may rely upon publicly generated health data (and other public-sector investments in AI and healthcare infrastructure) to identify, test, and validate a new medicine, method, or product, one issue is whether public investment will be fairly rewarded, including by ensuring affordable access to an end-product. Assessing ownership is especially difficult when a product or research output is the result of a public-private partnership, for which governments may have provided funding and other forms of support but maintain limited or no ownership of the research output. Providing a role for government, both in the development of new AI-based technologies, as well as in the ownership of the outcomes, may provide a fairer outcome for governments and citizens that contribute resources and data to a technology collaboration with the private sector.

A separate concern may be that providing patent monopolies for such inventions, even if it does encourage innovation, may discourage companies who own AI-guided technologies to develop or adapt such products to address the needs of people living in poverty in LMICs. Thus, as AI technologies are used more frequently to develop new technologies to improve health care, including new medicines, the use of incentives outside of the patent system, such as those that separate the cost of research and development from the expectation of high prices, could encourage companies that manage AI technologies to invest in uses of AI, or adapting new products, to meet global public health needs.

There are also several legal issues that remain unresolved and that will affect the patenting of AI-based technologies. One legal issue is whether AI-guided machines which develop new products or services on its own can be an inventor? Some legal experts have argued that recognising machines as inventors will encourage the development of creative and powerful machines that can generate new innovations.[[303]](#footnote-303) However, if such machines are owned primarily by a narrow group of companies, then the benefits of such inventions will also accrue to a small group of companies, who will wield significant power through exclusive rights, and can deploy machines to capture an entire field of technology. In April 2020, the United States Patent and Trademark Office ruled that machines cannot be listed as inventors according to current U.S. patent laws.[[304]](#footnote-304)

A second legal issue is whether diagnostic methods and algorithms can be patented. While in the U.S. securing patent protection for such diagnostic methods and mathematical models is highly restricted, the European Union has provided several grounds for the issuance of patents.[[305]](#footnote-305) On the one hand, while patent monopolies could encourage the development of new technologies that provide improved medical benefits, the availability of patents for such methods and services could limit the diffusion of and access to such diagnostic methods or technologies.

Recommendations

1. Although there are several types of intellectual property rights that can be asserted in large data sets and/or biomedical big data, IP may not be an appropriate means to encourage the widespread use and application of biomedical big data. Custodianship, or responsible oversight infused with ethical values, can ensure access to data, promote fair data sharing practices, and preserve privacy.
2. Governments that play a role in the development of new AI technologies, whether in providing funding, health data, or through partnership in the development and use of AI, should ensure it either maintains an ownership interest in the outcomes or ensure that such outcomes are widely available and affordable.
3. Governments may need to introduce additional incentives, such as those that separate the cost of research and development from the expectation of high prices, to ensure that companies that manage AI technologies invest in uses of AI, or adapt new products and services, to meet global public health needs.

## Governance of the private sector

The private sector plays a central role in the development and delivery of artificial intelligence in healthcare. The ‘private sector’ ranges from small start-ups to the world’s largest technology companies, as well as companies that provide many key materials needed for AI in health, whether it is health care data collected by wearable firms or data aggregators, or software firms that write new algorithms that can be applied for use in health care. Furthermore, many companies that were already engaged in the provision of products and services are transforming their businesses to integrate AI and big data. This includes biopharmaceutical companies, diagnostic and medical device firms, insurance companies, and private hospitals and health care providers. Additionally, companies that are developing AI technologies for use in healthcare are providing these applications and services outside of the healthcare system, raising questions as to how such forms of health care provisions should be regulated.

This section examines several issues related to the governance of such companies: (i) To what extent should oversight and governance of the private sector be enforced by companies collectively or individually; (ii) What are the challenges and opportunities for effective governance associated with public-private partnerships used for AI in health care; (iii) What are the specific challenges associated with the oversight and governance of large technology companies that are involved in the use of AI for health; (iv) how should governments manage the growth of health care services provided by companies independent of the health system, and (v) How can governments ensure that they are providing effective oversight of the private sector.

### The role of self-governance

As companies are often pushing the boundaries of innovation, and acting much more quickly than regulators, governments, and civil society can anticipate, they are often first setting the ‘rules of the road’ through the code that they may write, the services they design, and the corporate practices and terms of services that are offered.[[306]](#footnote-306) Since some innovations have raised significant concerns, companies have aimed to strengthen their own internal processes and measures to assuage concerns and have also pursued collaborations and partnerships.

Thus, some companies have introduced their own ethical principles and internal processes that are intended to integrate ethical considerations into their business operations.[[307]](#footnote-307) This can include integrating ethics into the design of new technologies, including design-related approaches to privacy and safety. Companies have also launched multi-stakeholder initiatives to develop best practices,[[308]](#footnote-308) though no such initiative yet exists for the use of AI in health.

While a company’s active integration of ethics into its operations is welcome, it raises as many concerns as it does hopes, especially that companies may be engaging in ‘ethics-washing’ and that such ethics-related efforts are intended to forestall regulation in lieu of preparing for and adapting to oversight.[[309]](#footnote-309) Even when efforts to formulate and integrate ethics into daily company operations is sincere, there are other challenges that may limit the effectiveness of such measures for the use of AI in health.

First, the incentives and values under which AI firms and developers operate may differ from those of patients, health care providers, and health care systems.[[310]](#footnote-310) For example, large technology companies, which may be situated in only a few countries, may adopt, and seek to mainstream certain values and belief systems which are not appropriate for other countries, health care systems, and communities, which are asked to use such products and services but do not have a role in establishing the culture and norms under which products and services are developed.[[311]](#footnote-311) More generally, while medicine is guided by a bottom-line objective of promoting the health and well-being of patients, an AI developer within a company, even if aiming to develop a product or service that generates specific benefits, is ultimately working towards the bottom line interest of a company to develop and deliver a profitable service or product, and in the case of publicly traded companies, to their shareholders.[[312]](#footnote-312) And while medical professionals have a long-standing fiduciary relationship to patients, AI developers, however well-intentioned, and even taking into account emerging expectations and legal obligations to safeguard individual privacy, do not have any fiduciary duty to patients (or health care providers), thereby complicating any individual or company effort to put the health and well-being of the patient first.[[313]](#footnote-313)

Second, ethical norms that companies adopt may not be easily translated into practice.[[314]](#footnote-314) This can be because AI development, a relatively new technology and area of practice, does not have methods in place that can translate such principles into practice. Or if high-level ethical norms are translated into practical measures, it may still be difficult to reconcile with a culture at companies that is focused upon fast growth, fast failures, and getting first to market. In such situations, ethical principles may be ‘watered down’, modified, or rendered ineffective. It may also be difficult to determine whether ethical norms are being written into source code for an AI technology. This is contrast to the practice of medicine, which over time has built numerous structures, including professional societies and boards, ethics review committees, accreditation and licensing schemes, peer self-governance, codes of conduct, and other mechanisms, that determine and shape what is acceptable on a day to day practice, and identifying both bad practices and bad actors quickly.[[315]](#footnote-315)

Third, there are insufficient legal and professional accountability mechanisms to reinforce the good-faith efforts of firms to turn ethical principles into practice.[[316]](#footnote-316) Unlike the medical profession, AI developers and tech firms lack effective self-governance mechanisms and do not face the legal penalties and accountability that other professions, especially the medical profession, are built upon. Accountability mechanisms within the medical profession reinforce a fiduciary duty between the provider and patient and are backed by sanctions to deter poor practices. At present, AI development does not include such professional or legally endorsed accountability mechanisms.[[317]](#footnote-317)

Fourth, there are questions as to whether companies can govern their own products and services effectively to minimise harmful impacts. Social media companies such as Facebook play an important role in the sharing of health information through its different platforms, including Facebook and WhatsApp. In recent years, there has been significant concern with the spread of misinformation and disinformation on its platform that undermines medical and public health information issued by governments and international agencies, which has been heightened during the SARS-COV2 pandemic. The company has taken several steps to address such forms of misinformation and disinformation, including a partnership with the World Health Organisation to create a Chatbot on Facebook Messenger and WhatsApp to provide accurate information through the WHO Global Alert Platform.[[318]](#footnote-318) Yet at the same time, a study issued by a not-for-profit group, Avaaz, found that the spread of medical disinformation and misinformation on Facebook far exceeded information from trustworthy sources such as WHO. During the early stages of the SARS-COV2 pandemic, this peaked in April 2020, when ‘disinformation sites attracted an estimated 420 million clicks to pages peddling harmful information – such as supposed cures for COVID-19.’[[319]](#footnote-319) Furthermore, only 16 percent of misleading or false articles displayed a warning label by Facebook third-party fact checkers.[[320]](#footnote-320)

The study from the US-based non-profit activism group Avaaz found that the spread of medical disinformation on Facebook far outstripped that of information from trustworthy sources, with the most popular “super spreader” sites receiving four times the number of clicks of bodies such as the US Center For Disease Control and the World Health Organization.[[321]](#footnote-321) The consumption of health misinformation on Facebook peaked in April, when the severity of the pandemic was becoming clear to populations in the US and Europe. In that month alone, disinformation sites attracted an estimated 420m clicks to pages peddling hoaxes that included harmful information — such as supposed cures for Covid-19. Just 16 per cent of the misleading or false articles analyzed by Avaaz displayed a warning label from Facebook’s third-party fact-checkers.[[322]](#footnote-322)

None of these concerns should be a reason for companies not to invest in measures that improve the design, oversight, and self-regulation of their products. This could include, for example, licensing requirements for developers involved in ‘high-risk’ AI, such as the use of AI for health care. Such licensing requirements would bring AI developers more in line with requirements within the medical profession and improve trust of their products and services. In fact, international standards organization have made important contributions to efforts to improve applications of health information technology. From data structure and syntax to privacy and implementation, for instance, the International Organization for Standardization (ISO),[[323]](#footnote-323) Health Level Seven International (HL7),[[324]](#footnote-324) and other organizations have contributed in numerous ways to the governance of information technology, including machine learning. Indeed, such standards have been described as carrying ethical weight.[[325]](#footnote-325)

### Public-private partnerships for AI in health care

Public-private partnerships (PPPs) are commonplace in health care. Unsurprisingly PPPs are emerging within the field of AI for health care. One type of PPP is the provision of raw data by the public sector, such as electronic medical records or other health data collected by health care systems and hospitals, which is then used by one or companies to develop specific products and services, such as diagnostic methods or predictive health care algorithms.

Supporters of PPPs in both government and industry point to the benefit of leveraging the resources and innovative capacity of companies to generate products and services. Such collaborations could presumably enable governments to oversee the activities of private companies engaged in such partnerships and to safeguard the public interest.

However, PPPs raise challenges for ensuring effective governance of the private sector. First, there is a significant information and skills asymmetry between the companies and government agencies that negotiate such partnerships. Companies often have trained professionals that are highly versed in the technology in question and the different parameters of a negotiated partnership. A second challenge is that the ‘social license’ granted to the public sector for use of certain resources, such as patient data, may not extend to private companies that are not trusted, and whose goals and objectives may not align with public expectations.[[326]](#footnote-326) Thirdly, public sector entities face several competing priorities that may undermine a government’s ability to effectively oversee a partnership. A public sector entity may have difficulty reconciling the objective of successful development of a new product or service, an obligation to protect the rights of individuals (and patients), and a wider responsibility to effectively regulate all of the operations of a private sector partner.

Finally, PPPs must determine ownership of products and services developed through a partnership. There are often concerns with other health PPPs that public sector contributions (data, funding, expertise, testing sites) are not taken into account when allocating ownership rights (if any) in the technology, as well as setting out the price of such technologies or the rules under which such a technology is operated.[[327]](#footnote-327)

### Governance and oversight of large technology companies

Large technology companies, especially those located in the United States and China, are expected to play a central role in the development and deployment of AI in health, whether through partnerships, in-house development of AI, or through acquisition of other companies. The role and involvement of these companies raises a related but novel set of considerations for oversight of the private sector.

Large technology companies, which are limited in number, wield significant power within the field of artificial intelligence. This is due to both the resources available to the companies – whether it be their human, economic, and technical resources, the accumulation of data by such companies through the products and services that they offer, the political influence that such companies may be able to exert through their relationships and partnerships with governments and their own staffing (see below), and the ability of such companies to use their platforms to introduce a range of products and services to a large number of users who may use a companies’ products and services for different reasons.

Over time, large technology companies may develop diversified products and services across their platforms. Google is developing a range of diagnostic applications that are still under examination for their safety and efficacy, while its parent holding company, Alphabet, has also launched a new health insurance service that will work in partnership with SwissRe.[[328]](#footnote-328)

Companies may also launch products and services that could compete with, replace, or introduce for the first time, a function or process that usually is managed by a government. Tencent has introduced an application that seeks to determine, based on information voluntarily supplied by individuals, what type of healthcare provider a patient should seek out.[[329]](#footnote-329) This has been done in part to improve upon an on-going challenge in China for patients, on the basis of their own research or intuition, to seek out medical advice from specialists that is not related to a patient’s underlying condition.[[330]](#footnote-330) Furthermore, the growth of telemedicine is providing opportunities for company owned platforms to shift patients to their platforms, which also are enrolling doctors that provide services via the platform. For example, Tencent based WeDoctor, which works with the government, has enrolled at least 240,000 providers on its platform, as well as 2,700 hospitals and 15,000 pharmacies. At least 27 million monthly active users use the ‘healthcare collaboration platform’ for either an AI-guided or a remote consultation, after which users are matched with specialists as appropriate within the healthcare system.[[331]](#footnote-331)

In the long term, this could mean that governments may not so much regulate companies that provide such services as they might have to depend upon such companies to fill in gaps and manage parts of the healthcare system. In other cases, technology companies may supply the underlying infrastructure on which healthcare services are operated, which also creates a dependence of governments upon the services and capacity of technology companies, in lieu of regulating the industry to serve the needs of governments and the public.

As noted above, technology companies have started to issue guiding principles for the use of AI. These principles have generated concerns in part because are sometimes viewed as ethics washing, may create a responsibility gap (with respect to assigning responsibility for retrospective harm), are not developed with the input of the public, and may be administered in a manner that is not transparent to the public or to governments, and which do not involve the participation of the public or the use of an independent authority to manage the oversight and operation of such principles.

### Provision of health care by the private sector outside of the healthcare system

AI applications in health are no longer grounded exclusively in healthcare systems. For example, AI applications for mental health needs are often provided outside of the healthcare system, such as through education, workplace, social media, and even financial services.[[332]](#footnote-332) Yet while there may be support for such expanded uses of health applications as a means of compensating for both increased demand and limited numbers of providers[[333]](#footnote-333), it generates new questions and concerns as to the regulation of such AI applications which are not deployed within the regulatory confines of health care systems.

Some of the concerns include the following: (a) it may be developed without appropriate reference to clinical standards; (b) such applications may not include be user friendly, especially as to how patients seek healthcare; (c) it may raise concerns for safety if individuals are not connected to healthcare services – for example that such applications may not be able to assist individuals with suicidal ideation that make use of AI chatbot; (d) such chatbots may not be tested properly for (and may lack) efficacy and (e) applications may not meet privacy standards required for sensitive health data.[[334]](#footnote-334) Since such applications may not be necessarily deemed to be healthcare services, and may not even be known to governments, it could undermine the overall quality of healthcare, and in particular may relegate those without other options to healthcare to subpar services. Governments will need to identify these applications, set out common standards and regulations (or in some cases prevent such applications from being deployed to the public), and ensure that individuals who use such applications are not severed from access to appropriate healthcare services that cannot be provided through an online platform.

### An enabling environment for effective governance of the private sector

There are several challenges to that may impede appropriate governance of the private sector. One challenge is the power of many companies involved in the delivery of AI for health care. Many companies employ former government officials and regulators, who in part are asked to lobby and influence policymakers and regulators charged with overseeing the use of AI in health care. This can affect the ability of governments to act independently of companies.

A second challenge is that many of the technologies developed by companies are increasingly difficult to evaluate and oversee. This is partly due to the growing complexity of such technologies, including the use of ‘black-box algorithms. This growing complexity has encouraged both governments and companies to consider models of ‘co-regulation’, in which each party relies upon the other to assess and regulate a technology. While such models of oversight may assist governments with understanding a technology under consideration, it may limit a government’s ability to exercise independent judgment and require governments to trust that companies are willing to strictly self-regulate their practices.

Improving governance of the private sector, if not through co-regulation, will require governments to have more independent ‘in-house’ expertise and information to effectively evaluate and regulate company practices. Thus, capacity building of government regulators to effectively regulate company practices, as well as transparency, will both play a key role in improving government oversight of the private sector. Transparency could include improved transparency of the data collected and used by private companies, transparency of how ethical and legal principles are integrated into company operations, and transparency of how products and services are performing in practice (including how algorithms may change over time).

Recommendations

1. Companies should invest in measures that improve the design, oversight, and self-regulation of their products. However, governments should not consider such investments in self-regulation as sufficient. Companies should also consider introducing licensing requirements for developers involved in ‘high-risk’ AI, such as the use of AI for health care.
2. Governments that engage in public-private partnerships should not extend its social license to companies without transparent engagement and acceptance by the public, should always prioritise the protection of individual rights within such partnerships, and should either seek out an ownership rights in such products or ensure that the outcomes of such PPPs are affordable and available to all.
3. Effective oversight of large companies will require different approaches depending on what aspects of healthcare a company seeks to engage upon and the jurisdiction such services are provided.
   1. There must be improved transparency by companies as to how such ethics principles are implemented in practice, including the outcomes of any actions taken to address violations of such principles.
   2. There must be improved participation of the public, including the use of an independent authority to oversee the implementation of such principles.
   3. There must be careful consideration by governments of what aspects of healthcare delivery, financing, services, and access can or should be supplied by companies, and what should remain within the remit of governments.
   4. As healthcare is increasingly delivered or managed through the use of AI and big data, it is incumbent upon governments not just to encourage the development of AI by commercial actors, but to also make use of government resources, knowledge, and know-how to develop publicly-owned products and services that can be used for the public benefit.
4. Governments must ensure that the growing provision of healthcare through online platforms not connected to the healthcare system are identified, regulated (including meeting standards of privacy protection guaranteed within healthcare systems), forbidden for areas of healthcare where the safety and care of patients cannot be guaranteed, and do not leave patients without access to appropriate healthcare services where required.
5. Governments may wish to adopt models of co-regulation to understand an AI technology. However, this may limit a government’s ability to exercise independent judgment and require governments to trust that companies are willing to strictly self-regulate their practices. Governments should also consider:
   1. Building its own internal capacity to effectively regulate companies that deploy AI technologies.
   2. Improving transparency of all aspects of a company’s operations.

## Governance of the public sector

The use of AI within the public sector has expanded over the past few years, though it lags private sector adoption. A 2019 OECD study identified 50 countries that have launched or are planning to launch national AI strategies, of which 36 of those countries plan or have already issued separate strategies for public sector AI.[[335]](#footnote-335) Even though the use of AI has expanded in the public sector, a literature review of nearly 1,700 AI studies found only 59 focused on the use of AI in the public sector.[[336]](#footnote-336)

There is no comprehensive accounting of how government is either advancing the use of AI or integrating AI into its own healthcare associated operations. The OECD has identified six broad roles for the government as it relates to AI:

* Government as a financier or direct investor into AI technologies both in the public and private sector
* Government as a ‘smart buyer’ and co-developer, which includes public-private partnerships and other forms of collaboration with companies.
* Government as a regulator or rule maker
* Government as a convenor and standard setter
* Government as a data steward
* Government as a user and services provider[[337]](#footnote-337)

This section briefly examines how governments should use AI ethically as an investor in AI technologies, as a smart buyer and/or co-developer, and as a user and services provider. It also considers ethical and human rights concerns with the increased use of AI to manage social protection and welfare, programs which often has a direct impact on access to health care services and an indirect impact on human health and well-being.

### Assessing whether AI is necessary and appropriate for use by the public sector

As with any use of AI by medical professionals and health care workers, governments must assess whether AI is appropriate for the intended use. This could include an impact assessment that evaluates the use of AI (see Section 10.X for a full discussion on the use of impact assessments) is a good fit. The government of the United Kingdom has established such an analytical framework for its use of AI, including the following: (a) whether the available data contain the required information; (b) if it is ethical and safe to use the data and consistent with the Government’s own Data Ethics Framework; (c) if there is sufficient quantity of data for the AI to learn from; (d) the task is too large and repetitive for a human to undertake without difficulty; (e) the AI will provide information a team could apply to achieve real world outcomes.[[338]](#footnote-338)

### Accountability through transparency and participation

Governments are increasingly required to disclose the use of algorithms in government services and operations to promote accountability for the use of AI. In France, the government is required to provide a general explanation of how any algorithm it uses functions, personalised explanation of decisions issues by algorithms, justification for decisions, and publication of the source code and associated document about such algorithms. Such use of AI in France should also provide for contestation.[[339]](#footnote-339)

More generally, governments should be transparent with respect to its use of AI, including whether government is: investing in the use of AI, engaged in partnerships with companies, developing AI on its own through state-owned enterprises or government agencies, and transparent with respect to any harms caused by such uses of AI, and measures that have been taken to redress such harm. A Review conducted by the UK Committee on Standards in Public Life found that the British government had not met established principles of openness – noting that ‘under the principle of openness, a current lack of information about government use of AI risks undermining transparency.’[[340]](#footnote-340)

Yet transparency may not be sufficient to ensure that government use of algorithms will not lead to undue harms, especially for marginalised communities and populations. Governments must also improve democratic participation through a wide range of stakeholders to ensure that decisions as to whether or not to introduce an AI system (in healthcare and beyond) are not just merely the remit of civil servants and companies, but are ultimately a decision that is based upon democratic participation of a wider range of stakeholders, including public interest representatives and leaders of vulnerable groups that are often not considered in such decision-making. It is especially critical to include these perspectives in advance, and not only after adverse impacts are identified, which is therefore too late.

### Appropriate collection, stewardship, and use of data

Governments must also develop and implement ethical and legally compliant collection, storage, and use of data. Government use of data could otherwise be prone to abuse, whether through sale or provision of such data to private companies that violate the public trust, or sharing of data obtained or collected for health-care related purposes for other government programs, including immigration enforcement or criminal justice. Such additional uses of data undermine trust in the health care system and the willingness of individuals to provide data and use AI technologies that are intended to improve the administration of health care and medicine.

Governments should also take steps to avoid risks of bias for data that is collected and used for the development and deployment of AI in the public sector. The obligation of the public sector to remain objective is otherwise undermined, since the ‘prevalence of data bias risks embedding and amplifying discrimination in everyday public sector practice.’[[341]](#footnote-341) The aforementioned review of the use of AI by the public sector in the UK also found that ‘data bias is an issue of serious concern, and further work is needed on measuring and mitigating the impact of bias.’[[342]](#footnote-342)

### Risks and opportunities with the use of AI for provision of public services and social protection

One use of AI by governments has been in the provision of public services, including assessments of whether an individual qualifies for certain public services, or what is known more generally as the ‘digital welfare state’. In sum, digital data and technologies are used to automate, predict, and identify or disqualify potential recipients of social welfare.

While some have championed such use of AI as a way of eliminating redundant and repetitive tasks that would both save resources and provide government employees with more time to address difficult issues,[[343]](#footnote-343) there are concerns that the digital welfare state can undermine access to social services and welfare and especially affect poor and marginalised populations. According to a report issued by the UN Special Rapporteur on extreme poverty and human rights, the digital welfare state can become a ‘digital dystopia’, constricting budgets intended for the provision of services, narrowing those who qualify for government services, creating new conditionalities, and introducing new sanctions that discourage the use of services.[[344]](#footnote-344) It also notes that administering the welfare state through a digital ecosystem risks exacerbating inequality since many poor and marginalised individuals are without adequate access to online services.[[345]](#footnote-345) Although the report of the UN Special Rapporteur does not discuss the use of AI to either provide or deny health care services, such uses of AI could affect the provision of health care through the public sector (or, for example, the provision of health insurance through the public or private sector).

Recommendations

1. For AI technologies that government wishes to use in the public sector, governments should conduct transparent and inclusive impact assessments prior to the selection or use of any AI technology. See also Section 10.X for a discussion on impact assessments.
2. Governments should be transparent with respect to the use of AI in the public sector, including but not limited to whether government is investing in the use of AI, engaged in partnerships with companies, developing AI on its own through state-owned enterprises or government agencies, and with respect to any harms caused by such uses of AI.
3. Governments must ensure that decisions of whether to introduce an AI system (in healthcare and beyond) are not just merely the remit of civil servants and companies, but are ultimately a decision that is based upon democratic participation of a wider range of stakeholders, including public interest representatives and leaders of vulnerable groups that are often not considered in such decision-making.
4. Governments must also develop and implement ethical and legally compliant collection, storage, and use of data in the public sector, and should also take steps to avoid risks of bias for data that is collected and used for the development and deployment of AI in the public sector.
5. Governments must ensure that any use of AI for providing health and welfare benefits and social services does not become a digital dystopia that does excludes people from government services, creating new conditionalities, or introduce new sanctions that discourage the use of services, especially for poor and marginalised populations that already suffer from unequal access to health services and related social services.

## Regulatory considerations

The largest regulatory agencies, including the US Food and Drug Administration and European Medicines Agency, have been developing guidance and protocols to attempt to assure safety and efficacy of new AI technologies, while other regulatory agencies may not have the capacity nor expertise to approve such devices. A WHO Working Group has been formed to address regulatory considerations for the use of AI for health care and drug development. This guidance document identifies several ethics related concerns that could be addressed by regulatory agencies, and the challenges that may arise.

### Does regulation interfere with innovation?

One common assertion is that stringent regulations can limit innovation and deprive health care systems, providers and patients of innovations that are beneficial. Therefore, a balance must be struck between protecting the public with promoting growth and innovation.[[346]](#footnote-346) Yet the use of AI in health care through the application of machine learning is still new and often untested, and there are numerous ethical and human-rights concerns that policymakers and regulators must consider. For example, regulators must identify applications and AI-based devices that may be best described as ‘snake oil’, and which may seek to either: misrepresent what an application can accomplish, generate misinformation, or persuade vulnerable individuals to follow health advice that may undermine their well-being.[[347]](#footnote-347) Applications that provide no therapeutic or wellness benefit could also be introduced solely for the purpose of collecting health and biological data that is then used for other commercial or marketing purposes, or to encourage patients to pay for irrelevant or unproven health interventions.[[348]](#footnote-348) For example, data that was obtained by an academic from 300,000 Facebook users who thought they were providing data for a ‘psychological test’. Such data, alongside data of an estimated fifty million other users who were linked to the users (Facebook friends) that was collected, was then sold to Cambridge Analytica, which then used such data to build a software program to both predict and influence choices at the ballot box.[[349]](#footnote-349) Such use of data collected nominally for academic or health purposes could expose health systems, health providers, and companies providing health-related AI services, to significant risk.

Regulation could be varied according to risk, or that for those who are especially vulnerable, including people with mental illnesses, children, and the elderly, should be protected from misinformation and bad advice issued by health applications.[[350]](#footnote-350) Beyond the vulnerability of certain classes of individuals, people living in resource-poor settings, in countries with few resources to regulate and monitor adverse consequences associated with AI applications, and individuals with diseases that result in marginalisation and discrimination, such as HIV and AIDS or TB, may also merit greater protection and oversight by regulatory agencies as compared to applications targeting lifestyle or wellness related users.

### Transparency and explainability of AI-based devices

The ‘black box’ of machine learning creates several challenges for regulators. Authorities may not be able to fully assess new AI technologies because the standard measures that are used to assess safety and efficacy of medical technologies, or scientific understanding and clinical trials, are not appropriate for ‘black-box’ medicine.[[351]](#footnote-351) Complex algorithms are difficult for regulators to understand (including due to a lack of expertise in regulatory agencies) and difficult for developers to explain.

Improving scientific understanding (explainability) of an algorithm has been claimed as necessary so that regulators (and clinicians and patients) can understand how a system has arrived at a decision. Explainability is also a requirement of the European Union’s General Data Protection Regulation and may be introduced into legislation in other countries that are grappling with the proliferation of AI in healthcare and beyond.[[352]](#footnote-352) There are also arguments that, if a trade-off must be made between such transparency and accuracy, transparency should be preferred. Yet this requirement reaches beyond what may be possible, or even desirable, in the medical context. While it is often possible to explain why a treatment is the best option for a specific condition, it is not always possible to explain how that treatment works or its mechanism of action, because medical interventions are sometimes used for a particular purpose before their mode of action is understood.

Trust in decisions and expert recommendations depends on the ability of experts to explain why a system is the best option for achieving a clinical goal. Explanations of this kind should be based on reliable evidence of the superior accuracy and precision of AI systems over alternatives. The evidence should be generated by prospective testing of the system in randomized trials and not the performance of the systems against existing datasets in a laboratory.

Understanding how a system arrives at the judgements it makes may be valuable for a variety of reasons, but it should not take precedence over or replace sound, prospective evidence of that system’s performance in prospective clinical trials. Explanations of how a system arrived at a particular decision could encourage use of machine learning systems for purposes for which they are not well suited, as the models created by such systems are based on associations among a wide range of variables, which are not necessarily causal. If the associations are causal, practitioners might rely on them to make decisions for which the system has not been tested or validated. Requiring every clinical AI decision to be ‘explainable’ could also limit the capacity of AI developers to use the best AI technologies which may outperform older AI technologies, but which are not explainable.[[353]](#footnote-353)

Thus, clinical trials could provide assurance that unanticipated hazards and consequences of AI-based applications can be identified and addressed (and avoided entirely), and additional testing and monitoring of an approved AI-device can measure its performance and any changes that may occur once approved. Clinical trials, especially those carried out on diverse populations, can also determine whether an AI technology may be biased against certain sub-groups, races, and ethnicities (see below). However, clinical trials may not be appropriate because of the cost, because they take significant time to be undertaken properly, because the validity of clinical trial data may be called into question if an algorithm is expected to change over time with the use of new data, and because AI-based technologies and products are increasingly personalised to smaller populations, and therefore more difficult to test on sufficient numbers of individuals.[[354]](#footnote-354)

In lieu of clinical trials, regulators could introduce ‘lighter premarket scrutiny’ that assesses the safeguards that have been put in place by developers, including the quality of data used, development techniques, and validation procedures, alongside ‘robust post-market oversight’. Yet this may be difficult to implement in practice, especially post-market oversight of novel algorithms,[[355]](#footnote-355) and may be too late to forestall harm to people that are especially vulnerable, such as individuals who may not have access to a healthcare provider that can protect patients from misguided diagnosis or advice. There could also be improved transparency of the initial dataset, including the provenance of such data and the data-processing procedures – as well as transparency of the system architecture.[[356]](#footnote-356) Such forms of transparency can enable others to independently validate an AI technology and improve trust with users.

While increased transparency of different components of an AI system, including its source code, data inputs, and analytical approach can improve regulatory oversight, some forms of transparency could also lead to a misplaced focus. Reviewing lines of code would be time consuming and likely illuminating, especially compared to the systems performance, functionality, accuracy (both before and after a system is integrated into a healthcare system).

### Addressing bias

Regulatory agencies could create incentives that encourage developers to identify and avoid biases. This could include, for example, additional measures within a precertification program hosted by the US Food and Drug Administration called the Digital Health Innovation Action Plan. This program already assesses medical software according to different ‘excellence criteria’ including quality.[[357]](#footnote-357) The quality criteria, and other regulatory criteria developed by regulatory agencies, could include the risk of bias in training data.[[358]](#footnote-358) Robust post-marketing surveillance that identifies biases in machine learning algorithms, including through collaboration with providers and communities that are likely to be affected by such biased algorithms, could also improve regulatory oversight.

### Ethics considerations for low- and middle-income countries and high-income countries with poor health outcomes

Developing countries historically and currently face a challenge of having insufficient regulatory capacity, either leaving countries unable to assess the safety and efficacy of new technologies, or having to rely upon other regulatory authorities (in particular regulatory agencies based in high-income countries), including collaborative processes sponsored by WHO, to ensure that new technologies are appropriate for use. It could also be argued that there should be efforts to have global harmonisation of regulatory standards to ensure that all countries can benefit from rigorous testing, transparent communication of outcomes, and monitoring of a technology’s performance.

Yet international harmonisation of regulatory standards (in particular based upon high-income country standards), reliance on other regulatory agencies, or the assurances provided by product developers, assumes that the criteria used to develop or assess a new technology in a high-income country is appropriate for LMIC contexts and populations. This may not be the case. It is likely that AI-health technologies may not be ‘translatable’ between divergent settings, including when comparing LMIC settings to high-income settings.[[359]](#footnote-359) This can not only include the types of data used to train the algorithm, but also the assumptions and definitions included in an AI technology, such as what constitutes ‘healthy’, an assumption that may be driven by a small group of developers located within one company or geography, and which regulators in high-income countries may validate without regard for how such assumptions may translate into LMIC settings.[[360]](#footnote-360)

There are also assumptions by regulators of the ‘context’ in which an AI technology has been introduced. AI technologies may have ‘contextual bias’, wherein algorithms may not recommend safe, appropriate, or cost-effective treatments for low-income or low-resource settings,[[361]](#footnote-361) or countries that have resources but still experience poor health outcomes for segments of the population, as is often the case in some high-income countries. A developer who may have designed a technology for a high-income setting where there most of the population has good health outcomes may neither anticipate nor build an AI technology to anticipate the differences within LMIC settings or in other high-income countries with poor health outcomes, and a regulator, even if it may require prospective clinical trials, may not require data on how such AI technologies operate in LMIC settings or certain high-income settings.

While transparency of: the data used to train algorithms, the context in which an algorithm is trained, and other material assumptions are necessary, such transparency may only forestall use of an AI technology, thus avoiding harm, but not bestowing any benefit. Improving the performance and use of AI technologies in LMICs and certain high-income countries, and ensuring such technologies are adapted to such realities requires different incentives, approaches, and actors to develop technologies that are appropriate for all people.[[362]](#footnote-362)

Recommendations

1. Regulatory standards for new AI technologies should not be circumscribed or ignored in the name of promoting innovation. Such shortcuts can result in harmful, insecure, or dangerous AI technologies being used by providers and patients with the potential for irreversible harms and the loss of public trust.
2. Regulators should require certain forms of transparency of an AI technology, such as its source code, data inputs, and the analytic approach, to improve oversight and assurance of safety and efficacy of AI technologies. For some AI technologies, this may include clinical trials, insofar that the data generated by the AI technology is valid for a sufficiently large population. Clinical trial data may not be sufficient if an AI algorithm changes over time with the use of new data.
3. Regulators should provide incentives to developers to identify and address biases during product design and development and could also integrate such guidelines into precertification programs. Regulators should also mandate or conduct robust marketing surveillance to identify biases.
4. Reliance on regulatory approval of AI technologies by HICs or the use of collaborative registration procedures may facilitate timely approval of AI technologies for LMICs, and especially those countries without adequate resources or expertise. However, reliance on HIC regulatory approval or harmonisation of regulatory standards of LMICs with standards in HICs may be inappropriate due, for example, to the different context in LMICs, and the assumptions of regulators and developers of what constitutes being ‘healthy’.

## International governance of AI

AI continues to play an ever-expanding role worldwide. Defined by its role in the economy, AI already has contributed US$ 2 trillion to global gross domestic product (GDP), and by 2030 could rise to represent more than US$ 15 trillion for the global economy.[[363]](#footnote-363) The import of AI has not just been measured by its economic relevance but also for the role it might play, positively or negatively, to achieve the Sustainable Development Goals (SDGs). According to one study, AI could enable the accomplishment of 134 targets within the SDGs, yet also inhibit 59 targets.[[364]](#footnote-364)

Ethics principles, regulatory frameworks, and national laws continue to proliferate, which do provide some measure of governance over the use of AI. Yet even with the growth of AI standards and laws around the world, there are several reasons why an international framework to regulate AI may be needed.

First, there is still significant variance across countries, organisations, and companies with respect to ethical principles related to AI. Implementation also differs widely across countries, but also differs between governments and companies, with several companies boasting more users than the most populous countries in the world. Global governance can ensure that companies and governments are increasingly adopting one common standard, and to avoid a ‘race to the bottom’ as it relates to the use of AI. Otherwise, the economic gains that could be achieved from AI could encourage some governments and companies to ignore human rights obligations.

Second, global governance could also strengthen the voice and role of LMICs, which are less involved in the development of AI technologies or in setting out international principles. LMICs also lag in implementation of AI, including for health, in part because of the enduring digital divide that undermines the use of AI technologies. Global governance could play a role in improving access to ICT and digital technologies in LMICs, as well as to provide guidance to hold companies accountable for harmful practices in LMICs.

Third, global governance could ensure that all governments can adapt to the changes that will be wrought by AI as these technologies become more sophisticated and powerful. Independent scientific advice and evidence will need to be collected and communicated as AI technologies rapidly evolve, and then translated into policy guidance.

Global governance for AI in health will consist in part of adapting existing governance structures, whether it is the policies and practices of global health agencies, adapting treatment guidelines issued by WHO, or updating global agreements to meet certain health objectives, such as the end of HIV and AIDS, by 2030. It is an open question whether each dimension of the use of AI, such as labour-related impacts of AI, data governance, privacy, ownership, or autonomous decision-making, require specific global standards for AI in health, or only require broad standards that can be applied to different AI uses.

Over the last few years, several efforts have been launched to improve global governance of AI, including a joint initiative of the governments of Canada and France to establish an International Panel on Artificial Intelligence, which is intended to convene global AI experts and develop guidance on different AI topics, including the ‘future of work’, data, and privacy.[[365]](#footnote-365)

Such bilateral or plurilateral initiatives, while welcome, should feed into global processes that are based upon the perspectives of all countries. In particular, the UNSG Roadmap for Digital Cooperation recommended in 2019 for ‘creating a strategic and empowered multi-stakeholder high-level body, building on the experience of the existing multi-stakeholder advisory group, which would address urgent issues, coordinate follow-up action on Forum discussions and relay proposed policy approaches and recommendations from the Forum to the appropriate normative and decision-making forums.’[[366]](#footnote-366) Such a multi-stakeholder body would benefit the wider governance and standard setting required for AI and provide a way forward for many of the specific challenges and questions related to the ethics and governance of AI in health.

Recommendations

1. Global governance of AI in health is needed to ensure that LMICs can participate and shape the use of AI in health, to prevent a ‘race to the bottom’ with respect to protecting human rights, and to ensure all countries can equally adapt to the changes that AI will bring to healthcare systems.
2. While plurilateral and regional initiatives should be welcome, the UN should create, as per the recommendation of the 2019 UNSG Roadmap on Digital Cooperation, a multi-stakeholder high-level body that could address urgent issues, coordinate follow up, and provide policy approaches and recommendations for use by governments and other normative and decision-making forums.

# Section 10. Building an ethical approach in practice for AI in health

This section examines several ways in which AI can be used to improve human health and well-being while minimising the risks, challenges, and ethical concerns that have been discussed throughout this publication. Several ethical approaches of using AI for health, such as public stewardship and oversight of health data (See Section 9.X) are discussed previously in this report.

## Ethical and transparent design of technologies

While technology designers and developers play a critical role in designing AI tools for use in healthcare, there are no credentialing and licensing procedures for developers and programmers to mirror those required of healthcare workers. To remedy the lack of reliable qualifications in ethics in the AI field, it is not enough to simply call for adherence to values of reproducibility, transparency, fairness, and human dignity

New approaches to software engineering have been developed in the last decade that move beyond an appeal to abstract moral values. Improvements in design methodology along these lines are not merely upgraded programming techniques. To support the effective, systematic, and transparent integration of ethical values, methods of designing AI technologies for moral values in health and other sectors have been proposed.

.

One approach to integrating human rights principles is through the process of Design for Values, which is an effort to ground the design process in the values of human dignity, freedom, equality, and solidarity (See Box X3) and to construe them as non-functional requirements.[[367]](#footnote-367) This requires less of a solutions-oriented approach and instead calls for a process-oriented approach which satisfies stakeholder needs consistent with the moral and social values embodied by human rights.

|  |
| --- |
| Box X3: Design for Values[[368]](#footnote-368)  Design for Values is the explicit translation of moral and social values into context-dependent design requirements. It encompasses under one umbrella term several pioneering methodologies, such as Value Sensitive Design, Values in Design, and Participatory Design. Design for Values presents a roadmap to engage stakeholders to translate human rights into context-dependent design requirements via a structured, inclusive, and transparent process. Design for Values seeks to gradually translate abstract values into design requirements. Values are expanded into norms (properties that a technology should exhibit to support certain values), and norms are then specified into a socio-technical design requirement. The process of identifying design requirements permits all stakeholders – such as individuals affected by the technology, users, engineers, field experts, and legal practitioners, to debate design choices and identify the advantages and shortcomings of each design choice.  Thus, a value such as privacy can be interpreted through certain norms, such as informed consent, right to erasure and confidentiality. These norms, through discussion and consultation, can then be converted into specific design requirements, such as positive opt-in (as a means of ensuring informed consent), or for example, homomorphic encryption techniques to assure confidentiality. Other techniques to safeguard privacy, such as k-anonymity, differential privacy, and coarse graining through clustering, could also be selected through a process of consultation. |

Ethical design can only be applied to socio-technical systems within which algorithms are designed. This means the ensemble of software, data, methods, procedure, personnel, protocols, laws, norms, incentive structures, and institutional frameworks should be brought together in a manner so that products and services deliver ethical outcomes to society and its healthcare systems.

More generally, ensuring ethical and transparent design of AI technologies should be done through processes and methods that prioritise inclusivity.[[369]](#footnote-369),[[370]](#footnote-370) Consideration of inclusivity at the outset of designing and developing an AI technology can help overcome barriers to the equitable use of AI in health, such as the digital divide – whether defined by geography, gender, age, culture and religion, or language.

Three such approaches to promote inclusivity are as follows:

* Citizen science: Citizen Science, as defined by the Alan Turing Institute, is ‘when non-professional scientists contribute directly to scientific research. This could be, for instance, by contributing data or performing tasks’.[[371]](#footnote-371) Citizen Science not only gets the public to understand a particular study or technology that may affect them personally but to also ensure that the public is involved in research, discussions of, and tool building. This ensures that there is respectful co-creation of AI technologies that can reduce the distance between the researcher or programmer with the individuals that a technology is intended to serve.
* Open Source Software: Transparency and participation could be enabled by adopting open source software for the underlying design of an AI technology or making the source code of the software publicly available. Open Source software is open for both contributions and feedback, which allows users to understand how the system works, to identify potential issues and to have the opportunity to extend and adapt the software. It is important for Open Source Software design to be implemented in an accessible and welcoming structure and content to allow greater engagement and transparency.
* Increased diversity: Too often, diversity for AI is focused upon increasing diversity in data collected. Although it is necessary, it is not sufficient and might even amplify potential biases inherent within technology design. Minimizing and uncovering potential biases require increasing representation of diverse people who are more familiar with the nature of potential biases, contexts and regulations, throughout the entire process of software development from the design of a technology, to the consultation of stakeholders, labelling of data, testing, and deployment.

Finally, toolkits can play a useful role in providing concrete guidance to technology designers that wish to integrate ethical considerations into their daily work. Software developer kits can provide guidelines that can incorporate a code of ethics, with specific guidelines dedicated to health. Such a toolkit could help to determine, for example, how to manage data, including data collection, de-identification, and aggregation, as well as to how to manage data and safeguard the destination of such data.

Specific toolkits have also been developed to help facilitate certain ethical (and increasingly legal) requirements, such as the Sage Bionetworks Toolkit for the Elements of Informed Consent.[[372]](#footnote-372) The toolkit provides use cases to explain its approach to informed consent, including eConsent, examples of how this should be put in practice, a checklist to ensure programmers have considered all questions that need to be taken into account, and also additional resources.

Annex 1 of this guidance document provides a Checklist for programmers and designers to consider how best to integrate ethical considerations for the development of AI technologies for health.

Recommendations

Ethical and transparent design of technologies is necessary to uphold and promote values such as transparency, fairness, and human dignity. Several approaches that can be implemented to promote ethical design include the following:

1. Design for values can engage stakeholders to translate human rights into context-dependent design requirements via a structured, inclusive, and transparent process, and ultimately abstract values into specific design requirements.
2. Consideration of inclusivity at the outset of designing and developing an AI technology can help overcome barriers to the equitable use of AI in health. Three strategies to ensure inclusivity are: Citizen Science, open source software, and increased diversity through the design and development process of an AI technology.
3. Software developer kits can provide guidelines that can incorporate a code of ethics, with specific guidelines dedicated to health.

## Bills of rights

The Universal Declaration of Human Rights has been instrumental in enshrining the notion of human dignity in international law. Based on the fundamental dignity and equality of all human beings, the notion of patient rights was developed. Patients' rights vary in different countries and in different jurisdictions, often depending upon prevailing cultural and social norms. Different models of the patient-physician relationship have been developed, and these have informed the rights to which patients are entitled. There is growing international consensus that all patients have a fundamental right to privacy, to the confidentiality of their medical information, to consent to or to refuse treatment, and to be informed about relevant risks of medical procedures.[[373]](#footnote-373)

A patient’s bill of rights is a list of such guarantees and may take the form of a law or a non-binding declaration. There are patient’s bills of rights in diverse countries such as India, the United Kingdom, Israel, and Malaysia. There are also several international declarations summarising patient rights, such as the World Medical Association’s Lisbon Declaration.

The emergence of artificial intelligence in health may require patients, medical associations, and governments to revisit such patient bills of rights and to update these charters and laws to account for an era in which algorithms will play a more prominent role in the delivery of health care. WHO has contemplated the development of such a Bill of Rights to accompany the growing use of AI in healthcare. Such an updated bill of rights would preserve the voice of patients within decision making that includes input and recommendations from both medical providers and machines, each of which may have more information than the patient of his or her own health, or wherein a machine may make a prediction about a patient’s future mental or physical health that the patient cannot yet anticipate.

Recommendations

A patient bill of rights plays an important role in defining the basic obligations that providers, healthcare systems, and governments must be required to follow to preserve a patient’s dignity, autonomy, and participation.

1. Existing patient bills of rights should be updated to reflect the emerging changes in health care provision with the use of AI technologies.
2. WHO should work with partner health agencies and appropriate institutions to delineate new rights that must be afforded patients with the wider use of AI technologies in health care.

## Research agenda for the ethical use of artificial intelligence for healthcare

In a fast moving field such as the use of AI in health, there are many unresolved technical questions on how best to use AI, and much more technical research on the best way to use AI in health care systems, including operational and technical research. There are also many ethical quandaries generated by the use of AI, and each new application or use of AI, as was the case with the introduction of proximity tracking applications during the COVID-19 pandemic, introduces additional ethical questions that had not been previously considered.

Recommendations

Some of the ethical concerns will require research to substantiate and explain these challenges, and approaches to address these concerns could be tested and validated with research. The following are recommended research questions that should be considered:

* How is artificial intelligence changing the relationship of healthcare workers and patients? The assumption is that such technologies will facilitate providers to spend more ‘quality time’ with patients, or is the contrary true, making care less humane? Are there specific contextual factors that either improve or undermine quality of care?
* What are that attitudes of healthcare workers and patients with the use of AI? Are these technologies acceptable? Are there differences depending on the type of intervention, the location of the intervention, and the existing acceptability of such technologies (not only within the health care system but in society more broadly)?
* How far has the digital divide exacerbated the introduction and use of AI in health care? Or does AI, combined with telemedicine, reduce the gap of access to care and ensure equitable access to quality care, irrespective of geography and other demographic factors?
* How best can providers and programmers address the biases that will always manifest in applications? What are the barriers to addressing biases?
* What should be the methodology for assessing whether AI is cost-effective and appropriate compared to existing or other ‘low-technology’ solutions in low- and middle-income countries? How should governments and providers assess resource allocation between existing interventions and new technologies that is fair?

## Engagement and role of the public and building trust with providers and patients

Effective use of AI in health will require building trust with the public, providers, and patients. Such social license requires hard fought efforts but can also be surrendered quickly if AI technologies are introduced without due care to the perspectives of those that will be affected by its use. Public engagement and dialogue are one means to ensure AI meets certain core societal expectations for its use in health care and greater trust and acceptance. Public dialogue ensures ascertainment of society’s views, as far as possible, on the ethical dimensions of AI, its design and uses.

One critical issue of concern for the public is the collection and use of patient data in AI and other applications. In the United Kingdom, there have been several examples of employing public debate and dialogue to reflect and address these concerns. Health Data Research, which collects health data to make it available to public and private entities for health-related applications of AI, has used public engagement, including with the Wellcome Trust’s initiative - Understanding Patient Data.[[374]](#footnote-374)

Workshops through the initiative provided a forum for participants to discuss their expectations and concerns with use of patient data in AI and other applications. Before these workshops, 18% of participants considered it acceptable to share anonymized patient data with commercial organizations for reasons other than direct care; after the workshops, the proportion increased to 45%.[[375]](#footnote-375) Patients viewed contributing data as a value exchange, with a societal benefit, and wanted the National Health Service to benefit from their data. They further considered it acceptable for commercial companies to have access to their data, provided that the benefit returned to the public, and that the National Health Service administer the data for the public benefit.

The UK Academy of Medical Sciences found through its own meetings and workshops that

“ongoing engagement with patients, the public and healthcare professionals, including via co-creation, will be critical to ensuring new AI technologies respond to clinical unmet need, are fit for purpose, and are successfully deployed, adopted and used”[[376]](#footnote-376). The Academy conducted public dialogue on the “data driven future” in order to understand awareness, expectations, aspirations and concerns about future technologies that require that patient data be accessed, analysed or linked for clinical diagnosis and management. Respondents considered that any new use of data must have a proven social benefit and that an appropriate organization (such as the Government or the National Health Service) should oversee the data and administer it for the public benefit.[[377]](#footnote-377)

Steps must also be taken to build trust with providers and patients that will increasingly rely upon the use of AI for routine clinical decision-making. In some cases, the willingness of patients to rely upon AI may be much lower than expected. For example, a study conducted by HSBC, the bank, demonstrated that only 8% of respondents surveyed would trust a machine offering mortgage advice compared to 41% trusting a mortgage broker.[[378]](#footnote-378) Not building wider trust could also create significant divisions in a healthcare system, wherein for example, older patients may be unwilling to adapt and use new AI technologies while younger patients may be more amendable.[[379]](#footnote-379)

From such a low level of trust, scandals that emerge from the use of AI in health care that undermine a patient’s economic, personal, or physical security could be fatal. Following the Cambridge Analytica scandal in 2019, an estimated 15% of Facebook users surveyed indicated they would reduce their use of the social networking site. Such trust could be even more quickly and severely eroded within the domain of healthcare if similar scandals or abuses of trust emerge into the public discourse, leading to the evaporation of public trust overnight.[[380]](#footnote-380)

One means to mitigate and manage risk could be through the use of ‘regulatory sandboxes’ – or contained testing grounds for health care providers and developers to test a new AI product or service in a ‘live environment’ with safeguards and oversight that ring-fences the health system from risks and unintended consequences that could materialise.[[381]](#footnote-381) Two examples of regulatory sandboxes are the United Kingdom’s Care Quality Commission and the Singaporean government use of regulatory sandboxes to test new (digital) health models.[[382]](#footnote-382)

A second approach that can build trust and facilitate a ‘graceful transition’ of health care is to redesign training programs and general education.[[383]](#footnote-383) This includes improvements in primary education in science, technology, and mathematics as well as continuing education that can allow health professionals to adapt to the use of AI. It could also enable health care workers to develop skills to avoid automation bias with the use of such technologies in healthcare settings. Otherwise there will be both a skill mismatch and less trust of providers to use AI technologies that would otherwise provide benefits.

A third approach – or the use of human warranty – is discussed previously in this report (See Section 7) – as a means of working directly with providers and patients to promote patient and clinical evaluation at critical points in the development and deployment of AI technologies. Use of human warranty can ensure there is meaningful public consultation and debate[[384]](#footnote-384)

Recommendations

To build trust with providers and patients and to ensure a social license to introduce AI technologies, the following forms of public engagement should be considered:

1. Entities that manage data should conduct workshops and consultations with people that may be asked to share health data to understand what forms of data sharing and use are acceptable and to fully consider and take into account public concerns and expectations.
2. Regulatory sandboxes can build trust with providers and patients before an AI technology is widely introduced, as well as limit risk.
3. Human warranty is an effective means to work directly with providers and patients to promote patient and clinical evaluation of AI technologies during its development and deployment.
4. Training and education programs must be redesigned to both improve the general public’s understanding and acceptance of the use of AI technologies, alongside continuing education programs to assist health care professionals to adapt to the use of AI.

## Impact Assessments

An impact assessment is a process of identifying the future consequences of a current or proposed action, policy, law, regulation, or as will be the case with the use of AI in health, of a new technology or service. Impact assessments can both provide technical information of the possible consequences – both positive and negative, can improve the decision-making process (and be tied to it), improve transparency and participation of the public in decision-making, and also introduce a framework for appropriate follow-up and on-going measurement, which may be especially important for the use of AI since an AI technology may change over time.[[385]](#footnote-385) Finally, impact assessments can measure whether or not a technology will uphold or undermine human rights obligations, including privacy and non-discrimination. There are several types of impact assessments related to the use of AI in health that have either been proposed or have been put into place that should be considered by governments, companies, and providers.

Businesses that design and introduce AI technologies in health have a particular obligation to conduct impact assessments, including human rights impact assessments. The UN Guiding Principles on Business and Human Rights establish a corporate responsibility to respect human rights, including companies implementing due diligence to identify, avoid, mitigate, and remedy human rights impacts for which they are directly responsible for or indirectly involved.[[386]](#footnote-386) Although the UN Guiding Principles do not require businesses conduct human rights impact assessments, such assessments can play a role in companies meeting their obligations.

Human rights impact assessments are a process for identifying, understanding, assessing, and addressing the adverse effects of business projects or activities.[[387]](#footnote-387) Although relatively new, the use of such impact assessments has expanded in recent years. It has also been recognized in national laws as an obligation of companies. For example, the French government enacted a Duty of Vigilance law that requires parent companies to identify and prevent adverse human rights and environmental impacts resulting from their own activities, from activities of companies they control, and from the activities of its subcontractors and suppliers, with whom they have a commercial relationship.[[388]](#footnote-388) Furthermore, an EU Directive may require all companies headquartered in Europe to conduct human rights due diligence, although such discussions will only be completed in 2021.[[389]](#footnote-389)

Other types of impact assessments have also been either proposed or implemented. One approach could be an ‘ethical impact assessment’, which would identify the impacts of AI on human rights, including impacts upon vulnerable groups, labor rights, the environment, and different ethical and social implications. A second approach, as proposed by the AI Now Institute, is an ‘Algorithmic Impact Assessment’, which is intended for public agencies and is a ‘practical framework to assess automated decision systems and to ensure public accountability.’[[390]](#footnote-390) Such an impact assessment is both for affected communities to obtain information to assess how automated decision systems function and to determine if such systems are acceptable, while also providing governments with a tool to assess how such systems are used, whether such systems produce disparate impacts (in particular on the basis of gender, race, or other dimensions), and how to hold such automated systems accountable. This could be useful as governments turn to algorithmic decision making to assist with large- and small-scale decisions that are taken throughout health care and medicine.

Several laws have also been proposed or implemented that also require the use of impact assessments, including for the use of AI in health. In 2019, two United States Senators co-sponsored the Algorithmic Accountability Act, which would have required companies to study and fix flawed algorithms that result in inaccurate, unfair, biased, or discriminatory decisions that would have an impact on people in the United States.[[391]](#footnote-391) It would also require companies, via enforcement by the U.S. Federal Trade Commission, to ‘reasonably address’ the results of such assessments and includes such algorithmic decisions that affect health. Such assessments would be only for ‘high-risk’ decision-making, which would include health information or genetic data, or decision-making or analysis of sensitive aspects of their lives, including an individual’s health and behavior. This law, however, had only been proposed and not enacted.[[392]](#footnote-392)

A separate proposal under the Algorithmic Accountability Act would have required companies to conduct ‘data protection impact assessments’ for high-risk information systems, or those that store or utilize personal information including health information of individuals. Such a data protection impact assessment in fact mirrors an impact assessment that has already been placed into law under the European Union’s General Data Protection Regulation (GDPR), which requires companies to conduct ‘data impact assessments’ of the risks of data processing operations to the ‘rights and freedoms of natural persons’ and their impact on the protection of personal data.[[393]](#footnote-393)

Presently, the European Commission is also contemplating the introduction of a new ‘Inception Impact Assessment’ for artificial intelligence. The proposed legislation would seek to address ‘a number of ethical and legal issues raised by AI’ and would seek to ‘foster the development and uptake of safe and lawful AI that respects fundamental rights across the Single Market by both private and public actors while ensuring inclusive societal outcomes.’ [[394]](#footnote-394)

Recommendations

Impact assessments can improve transparency, increase public participation, anticipate positive and negative consequences of AI technologies, and ensure adherence to human rights obligations.

1. Governments should enact laws and policies that require companies to conduct impact assessments, which could consist of either a human rights impact assessments, ethical impact assessments, data protection (privacy) impact assessments or inception impact assessments for artificial intelligence.
2. Companies and programmers should conduct impact assessments even if governments have not yet mandated it. The outcomes of such impact assessments should be carried out by an independent third party, published, and should be conducted prior to introduction of an AI technology.

## Putting prediction to good use

Even though healthcare has always depended in part upon predictions, the use of AI to make predictions about a patient’s prognoses will create significant opportunities to assess the relative risk of disease, predict illness, and predict major health events, including outbreaks, before they occur. Yet there are also several risks and challenges with the use of predictive analytics, including concerns with the accuracy of such predictions and concerns that such predictions – if it includes a negative prognoses – could both affect an individual’s autonomy and well-being.

Yet there should be ways to ensure predictive analytics can be used to maximise the public interest, which WHO is actively exploring. For example, prior to the SARS-COV2 pandemic, WHO started to develop EPI-BRAIN, a global platform that will allow experts in data and public health to analyze large datasets for emergency preparedness and response.[[395]](#footnote-395) This includes forecasting and early detection of infectious threats and their impact from scenarios, simulation exercises and sharing of insights to improve coordinated decision-making and response.

Recommendations:

1. Governments and international health agencies, such as WHO, should seek to proactively design AI technologies that make use of predictive analytics to assist and augment decision-making of providers and policymakers.
2. Predictive technologies designed by governments, health providers, and health agencies must adhere to ethical standards and human rights obligations, should be open to improvement, and should be available for adaptation and use by other governments and providers on a non-exclusive basis.

## Policy observatory and model legislation

As artificial intelligence plays a more prominent role in healthcare systems, governments have to or are already introducing national policies and laws to govern the use of AI in health. Ensuring that such laws and policies can address the various ethical concerns and opportunities from the use of AI, the OECD launched a policy observatory in 2020 that, according to the OECD: ‘aims to help countries enable, nurture and monitor the responsible development of trustworthy artificial intelligence systems for the benefit of society.’[[396]](#footnote-396)

WHO supports such efforts, and on the basis of the key ethical principles that are included in this report, will explore collaborating with OECD and other inter-governmental agencies to develop a policy observatory that would identify and analyse relevant policies and laws..

Furthermore, WHO may consider developing model legislation that can serve as a reference for governments to develop its own laws that can ensure that appropriate protections, regulations, rules, and safeguards are put in place so that the general public, providers, and patients can have trust in the use of AI within healthcare systems, as well as (for example) the management of data and information in a manner that strengthens the accuracy and utility of AI while not undermining privacy, confidentiality, and informed consent.

Recommendations:

WHO should work with OECD to identify relevant laws and policies that facilitate the ethical introduction and use of AI technologies for health, and should also consider developing model legislation that can serve as a reference for governments seeking to build an appropriate legal framework for the use of AI in health.

# References

# Annex I: Checklist for AI Developers

The following checklist summarises several key principles, ideas, and recommendations included in this report.

This Checklist is for developers and programmers – namely those individuals, research organisations, and companies, which are often the key party involved in the design, deployment, and re-design of AI technologies used for health care. Even though AI developers may not be situated formally within healthcare systems, the products that they design will play an increasingly central role in all aspects of health care. Beyond the specific requirements in this checklist and report, programmers, research organisations, and companies must consider more systematic approaches to ensure that the values, principles, and processes that guide their operations are increasingly aligned with the expectations of other actors within health care systems.

This Checklist is not comprehensive – it is a starting point for the steps that programmers and companies should take to ensure that the technologies they design and deploy can be used to the benefit of patients and providers.

The Checklist considers three areas of enquiry:

* Designing an AI technology
* Developing an AI technology
* Deploying and improving an AI technology following deployment

## Designing an AI technology

1. Clarify objectives

Scientific methods can only be applied to address clearly defined objectives.

*Specific requirements:*

* Consider overall objectives and constraints at every step of the process
* Clearly identify needs and end users of the tool that is being designed, including: who is going to use the tool, how it is going to be used, when and where it is going to be used, whether there is a secondary (indirect) group of users, and if there is a time limitation on tool validity and efficiency.

1. Engage multiple stakeholders and understand contexts

During the initial design of an AI technology, a developer should involve all relevant stakeholders in the full solutioning process. Prioritising inclusivity throughout the design process enables developers to better understand needs and to build adapted solutions for multiple stakeholders. AI technologies used in health care are context-dependent and must work appropriately for multiple stakeholders.

*Specific requirements:*

* Clearly delineate responsibilities – what to do, when, and how
* Have someone who understands the context involved in the design process (to avoid carrying bias from the data and amplifying it through the process)
* Design, discuss, and validate the problem formulation, conceptualization, proposed approach and solution with different stakeholders in the targeted regions, including policy and decision-makers, project owners and leaders, project managers, solution engineers and developers, potential users, domain experts, and ethics and information privacy experts.
* Examine in which context will the AI technology be used (e.g. which region, users background and main languages, main skills, regulatory frameworks)
* Examine how was the study data collected, including looking for any potential bias in the data that could be context dependent
* Determine what are the operation and technical limitations (manpower and expertise, software, and hardware requirements) to design, develop, test, use, and maintain the tool.

1. Define ethical topics and issues and apply ethical principles

Each AI technology will require consideration of ethical issues to focus upon, and such ethical concerns, which often emerge through a period of consultation and risk assessment, should be integrated into a technology’s design and development. Alongside those ethical principles included in this report (See Section 7), and depending on the organisation and project needs, a developer may need to select specific additional ethical principles to base her work upon.

*Specific requirements*

For each topic (such as privacy, safety, fairness and biases, transparency, accountability), clearly define the principles, the norms that explain such principles, and the (technical or operational) solutions and approaches.

1. Assess risks

Developers must engage in risk assessment and management of such risks to support human-centred design and development. Such risk assessment should occur at different stages of a technology’s development and should be reassessed on a regular basis in collaboration with different stakeholders

*Specific requirements:*

* Consider the context, cultural, social, and economic circumstances
* Consider human judgmental heuristics and cognitive biases
* As a part of the risk assessment, examine: (a) what are the expected outcomes, (b) what are the potential unexpected outcomes, (c) what would be the impact and consequences, and (d) what are the risk-mitigating approaches.
* Clearly communicate the above aspects to different stakeholders

## Developing an AI technology

1. Identify regulatory requirements, including support from experts

Depending on the jurisdiction, there might be regulatory frameworks for AI and data management already in place or under development. Compliance is needed for deployment and use of an AI technology. Adherence to such regulatory requirements usually requires working closely with experts or requests experts to review such plans. Even though regulatory frameworks for AI are evolving, most current regulatory frameworks focus on data security and privacy, while others may now also include requirements related to fairness and bias mitigation.

*Specific requirements:*

* Identify different regional or country-based regulations which may apply. For data protection, for example, this could include the European Union’s General Data Protection Regulation or for example Singapore’s Personal Data Protection Act.
* Certain sectoral or domain-based requirements may also apply – including for example the United States Health Insurance Portability and Accountability Act, or if eventually enacted, laws such as the Algorithmic Accountability Act.

1. Establish data management plans

Developers should establish clear data collection, storage, organisation, and access plans for data security. This might be done carefully, particularly when combining data from different sources for a single AI technology needs.

*Specific requirements:*

* Understand what the data sharing requirements and regulations are related to potential users’ countries.
* Understand and determine how confidentiality will be protected
* Determine what type of data is being collected and where and how the data will be stored.
* Determine a time frame: how long the data will be stored, when could the data be shared, etc..
* Determine who is responsible for data governance and ensure there is appropriate follow-up.
* Clearly identify all groups who have access to the data throughout the product’s life cycle.
* Determine what type of secondary use of data could be allowed, if at all.

1. Adopt standards and best practices

Developers should ensure compliance and/or interoperability of the AI technologies they develop with other technologies that will be introduced into health systems.

*Specific requirements:*

Based on the regulations, guidance and application requirements, design and development plans, developers should adopt one or more of the following standards for an AI technology:

* ISO Standards (Security and Privacy)
* NIST Standards (Security and Privacy)
* IEEE 7000 Series (Privacy and Fairness)

## Deploying and improving an AI technology following deployment

1. Engage and educate multiple stakeholders for deployment and maintenance

All relevant stakeholders should be involved in the full solutioning process. Prioritising inclusivity throughout the process will assist with better understanding needs and building adapted solutions that work for multiple stakeholders.

*Specific requirements*

* Clearly delineate responsibilities – what to do, when, and how
* Design, discuss, and validate the proposed approach and solution with different stakeholders in all targeted regions – including policy and decision makers, project owners and leaders, project managers, solution engineers and developers, potential users, domain experts, and ethics and information privacy experts.
* Train business owners, executives, developers, support engineers and users on why, how, and when to use the tool, including the following: (a) what are the main objectives and functions, (b) what are the different features, and (c) what are the differences between different usage scenarios when applicable.
* Engage continuously with multiple stakeholders and provide support to users.

1. Evaluate and improve performance

Developers should formally assess deployment outcomes and improve the design and development of an AI technology, considering the ethical principles which initially guided development. Developers must also continue to assess risks and management will be needed to support deployment, continuous development, and maintenance.

*Specific requirements*

Developers should evaluate accuracy and error risks and assess implications for:

* How long should the results or technology be used for?
* How often does the tool need to be updated?
* Who has responsibility for such updating?

# Annex II: Checklist for Ministries of Health

The following checklist is intended to summarise several key principles, ideas, and recommendations included in this report.

This Checklist is intended for Ministries of Health that will have the primary responsibility for determining how AI technologies should be integrated into health care systems, the conditions under which such technologies should be used, the protections for individuals that must accompany the use of such technologies, and the policies that can address both expected and unexpected ethical challenges which may materialise.

This Checklist is not comprehensive – it is a starting point for Ministries of Health to ensure that the use of AI technologies is consonant with the wider objective of governments to provide affordable, equitable, appropriate, and effective health care.

The Checklist considers three areas of enquiry:

* How should Ministries of Health prepare for the introduction and use of AI technologies?
* How should Ministries of Health address ethical challenges?
* How should Ministries of Health protect the safety, dignity, and human rights of persons?

## How should Ministries of Health prepare for the introduction and use of AI technologies?

1. Institutional Preparedness and Technical Capacity

Ministries of Health must have human and technical resources in place to fully realize the benefits of AI technologies for health while mitigating any possible negative impacts (e.g. to identify unethical outcomes). This means they need a degree of technical competence, a knowledge of human rights, and where necessary, the capability to compensate for missing capacity through external independent advice, civil society support, or international cooperation.

*Specific requirements*

* There should be training and capacity building of government officials, based on established criteria, to evaluate AI technology based on ethical principles.
* There should be the involvement and engagement of healthcare authorities and medical professionals in AI design, and where possible, hands-on guidance of software engineering.
* Civil society, medical staff, and patient groups should be consulted regarding the introduction of AI technology and be considered as part of the external audit and monitoring of its functioning.

1. Infrastructure for AI technologies

Not all countries or regions have the infrastructure to ensure the proper functioning and sustainable benefits of an AI technology. The right infrastructure is a pre-requisite to ensure the proper deployment of AI in a health care system. Without such infrastructure, or if the cost of such infrastructure is too expensive compared to its benefit, the introduction of AI technology should be resisted especially where the AI involves high risk for patients.

*Specific requirements*

* There should be established criteria to identify and measure the infrastructure requirements, including its operation, maintenance, and oversight.
* Where possible, infrastructure should be provided or supported with civil society support and international cooperation.
* Ministries of Health should identify effective alternatives if infrastructure is lacking, AI technology is too expensive and/or the AI involves high-risk to patients.

1. Assessing whether to adopt or avoid an AI technology for healthcare

AI technologies must be deployed safely. AI technologies should only be used if they contribute to achieving universal health coverage. The pursuit of AI could also divert attention and resources away from proven but underfunded interventions that would reduce morbidity and mortality in low- and middle-income countries. Reliable evidence of the superior accuracy and precision of AI systems over alternatives is needed. The evidence should be generated by prospective testing of AI system in randomized trials and not the performance of the systems against existing datasets in a laboratory.

*Specific requirements:*

* There should be criteria to determine the uptake of an AI technology that is specific to the context of the country where it may be introduced.
* AI requires rigorous testing, monitoring, and evaluation. This means prospective testing of AI technologies in randomized trials and not the performance of AI technologies against existing datasets in a laboratory. Regulatory agencies can play an important role in upholding such testing, transparent communication of outcomes, and monitoring of a technology’s performance.
* Impact assessments can guide a decision on whether to use AI within an area of health care.
* Ministries of Health should calculate the cost-benefit ratio for AI adoption and continuously evaluate AI technologies to measure and balance the cost-benefit ratio of investment and uptake of AI.

1. Management of data

Data quality is critical to prevent unintended harmful consequences in the implementation of AI systems. The use of limited, low-quality, and inaccurate data in AI could engender biased inferences, result in misleading data analyses, and promote poorly designed health applications. It has been shown that commercial prediction algorithms can identify complex health needs, but they can also result in significant racial or ethnic bias. Some AI algorithms are based for example on images of uneven quality and resolution, which could adversely affect the accuracy and reliability of the systems’ output. Securing data from numerous sources, including wearable technologies, genetic information, electronic health care records, radiology images, surveillance data, and even from hospital rooms and ensuing its quality, interoperability and confidentiality can be challenging, even questionable, and resource intensive.

*Specific requirements*

* The regulation of data, especially its representativeness, accuracy, harmonization, accessibility, interoperability, and reusability, coupled with the informed consent of the data subject (patients), is necessary for the introduction of AI technologies in health care. Such regulations should also apply to non-medical device data.
* There should be regulations in place regarding the access and use of data coming from consumer-facing digital self-care applications and/or wearable technologies. Data processed by these applications and technologies should be collected, stored, and used in accordance with data minimisation.
* Patients and consumers as data subjects should be allowed to reuse and thereby benefit from their own data and should not be monopolised by an AI technology provider.
* There should be quality control measures implemented to ensure the representativeness of data emanating from different population groups.
* Ministries of Health should ensure that there are mechanisms and procedures in place to collect patient data to enable AI technology to learn from data that relates to the environment, culture, and the specifics of the community the AI technology is intended to be placed

## How should Ministries of Health address ethical challenges?

1. Preserving and enhancing human agency and autonomy

AI should not be pursued as a substitute for clinical due diligence. The primacy of healthcare authorities and the experience of medical professionals are paramount. AI technologies in health should aim to enhance human capacity and capabilities and to empower medical professionals rather than to displace them. AI systems should be built to be ‘human-centric’ and models should work with and for people. Thus, human-machine teams should work together effectively coupled with shared decision-making with patients, their caregivers, and communities.

*Specific requirements:*

* There should be human judgment after the AI technology has made its (disease) prediction and/or proposed treatment recommendation.
* Ministries of Health should designate the types of information a clinician should be provided with to make an independent judgment of the AI-based result/outcome.

1. Transparency of AI technologies for health

AI technologies must be transparent. Transparency is critical to both building trust in AI technologies and to ensure that patient rights are protected. Thus, transparency of an AI technology (even if such a technology is not explainable) may be necessary if doctors are asked to act on decisions developed through black-box algorithms. A failure to disclose information could invalidate informed consent. Whether this is the case depends on whether any of the rationales for informed consent – namely protection, autonomy, prevention of abusive conduct, trust, self-ownership, non-domination, and personal integrity - are triggered by the use of AI in clinical care.

*Specific requirements:*

* Ministry of Health’s experts should be able to assess and evaluate AI technology in accordance with the requirements of transparency and responsibility.
* Ministries of Health must ensure that clinicians can patients and their families how a system has been validated.

1. Ensuring equitable access to healthcare alongside the use of AI

AI might create gaps between diagnoses and treatment. Expanding the use of AI must be done carefully to avoid situations in which large numbers of people may be accurately diagnosed for a health condition but are left without access to appropriate treatment options. There should be careful planning to avoid this.

*Specific requirements:*

* Prediction tools that anticipate a disease outbreak will need to be complemented by robust surveillance systems and other effective measures to respond to an outbreak.
* There is a duty to provide treatment following testing for and confirmation of disease

## How should Ministries of Health protect the health, safety, and human rights of patients?

1. Privacy, confidentiality, and informed consent related to the collection and use of patient data

Health data collection and use should respect the privacy, confidentiality, and informed consent of patients. The autonomy of the data subject is paramount, especially an individual’s ability to exert meaningful control over his/her data. The collection and reuse of such data in any country, without due regard for consent, privacy, and autonomy, as well as data collection without informing individuals of intended uses (commercial or otherwise), undermines the agency, dignity, and human rights of these individuals, and fuels concerns of an individual’s loss of control over his or her data.

Informed consent of the patient/data subject is critical and a pre-requisite for the implementation of any AI system in the healthcare domain, although informed consent may not always be possible with respect to biomedical Big Data. Biomedical big data may foster a divide between those who accumulate, acquire, analyze, and control such data versus those who provide such data but have little control over its use. While anonymization may safeguard privacy, it may undermine a person’s right to control their own data and how it may be used.

*Specific requirements:*

* Up-to-date data protection and confidentiality laws are a pre-requisite.
* Data protection supervisory agencies should have sufficient resources to make privacy protection actionable and effectives.
* There should be independent oversight mechanisms and other forms of redress for the protection of patient privacy and data confidentiality.
* A MOH should have experts evaluate an AI tool for meeting standards of privacy, fairness, and trust.
* A MOH should have a protocol in place for the collection, storage, and sharing of personal data or data that could be personally identifiable, and should ensure that such data is managed in a manner which protects privacy, confidentiality, and informed consent.
* A MOH should ensure that patients have the right to opt-out of data collection and data sharing requirements via an AI technology. There must be explicit consent to secondary use of health data.
* Ministries of Health should minimise the collection of data to what is specifically needed and not collect additional data.
* Ministries of Health should provide training on the privacy implications of the use of AI technologies for patients or health staff. This should be considered as a part of the human capacity building components of AI technology implementation.
* There should be swift and accessible mechanisms of complaint for either patients or health staff to demand the protection of personal data and particularly sensitive health data.

1. Patient agency and autonomy

AI predictive analytics across healthcare raises ethical concerns with respect to informed consent and individual autonomy for decisions made about patient and consumer health. User-facing apps and other services provided by companies might exert influence over those who are vulnerable regarding their health (this is problematic if companies can predict and thereby influence people’s feelings/anxieties and behaviours for commercial advantage). Individuals may feel unable to refuse treatment, in part also because the patient cannot dialogue with or challenge a recommendation made by an AI-guided technology (e.g. a notion that ‘computer knows best’), or that the patient is not provided with enough information or a rationale to provide informed consent. Health-related services and products driven by AI technology might lean towards negatively affecting individual well-being and autonomy.

*Specific requirements*

* Individuals should have the ability to access their own health data.
* Patients should be able to opt out of AI technologies for health.
* There should be a mechanism in place to inform patients of the benefits, risks, value, constraints, novelty, and scope of using an AI tool.
* There should be an assessment of the need for AI technology and a risk assessment of such a technology as it relates to patient autonomy and well-being.

1. Promoting a standard of care and assigning liability

Many LMICs still lack sufficient regulatory capacity to assess drugs, vaccines, and devices, and the onrush of machine learning technologies could leave these agencies unable to accurately assess and regulate such technologies for the public good. Understanding and explaining how a system arrives at the judgements it makes may be valuable for a variety of reasons, but it should not take precedence over or replace sound, prospective evidence of that system’s performance in prospective clinical trials. Clinical trials could provide assurance that unanticipated hazards and consequences of AI-based applications can be identified and addressed (and avoided entirely), and additional testing and monitoring of an approved AI-device can measure its performance and any changes that may occur once approved. Furthermore, relying on AI technologies ultimately entails responsibility, accountability, and liability, as well as compensation for undue damage caused.

*Specific requirements*

* MOH experts should evaluate an AI tool for accountability.
* There should be sufficient regulatory scrutiny for AI technologies relied upon in healthcare.
* For certain low-risk AI technologies, regulators may consider ‘lighter premarket scrutiny’.
* Liability rules used within clinical care and medicine should be modified and updated to consider the use of AI technologies. This should include causal responsibility, objective liability regimes and retrospective harm, as well as mechanisms to assign vicarious liability where appropriate.

1. Ensuring all people are guaranteed redress within a legal system

There must be equitable access to AI health technologies and compensation for undue damage. Ministries of Health should ensure independent oversight of AI technology deployment to ensure that there can be mechanisms of recourse and redress.

*Specific requirements:*

* There should be an independent oversight mechanism to ensure equitable access to healthcare of appropriate quality
* Swift and accessible mechanisms of complaint should be available.

# Annex III: Checklist for Health Care Providers (3rd November 2020)

The following checklist summarises several key principles, ideas, and recommendations included in this report.

This Checklist is intended for health care providers – namely practitioners (e.g. doctors and nurses), hospitals, and health care systems. While programmers may be primarily responsible for the design of AI technologies, and Ministries of Health and regulatory agencies for the approval and selection of such technologies for use in health care systems, health care providers play a critical role in determining which technologies to use, how they should be used, and can also provide direct feedback within their countries, the medical community, and to those that design such technologies to ensure that they meet the best needs of patients who rely on their judgment.

This Checklist is not comprehensive – it is a starting point for similar exercises that health care providers should introduce as the use of AI in health care increases.

The Checklist considers three areas of enquiry:

* Is the AI technology appropriate?
* Is the context in which this AI technology will be used appropriate?
* Should a health care provider use this AI technology?

## Is the AI technology appropriate?

The following requirements to determine if an AI technology is appropriate are not easily implementable. For those technologies designed by companies or programmers outside of the healthcare system, providers that use such technologies will have difficulty obtaining relevant information that could enable a provider to follow the guidance included in this Checklist – for example to assess the appropriateness of the technology. Furthermore, many providers will not have the expertise or resources – even if such information is available – to assess these technologies. Yet providers still must take ensure that such technologies do meet these requirements – whether by determining if such checks have been put in place by a designer, or if an international or national health regulator or agency has validated a technology to the satisfaction of these requirements. Such oversight will also require post-use oversight, especially if the source code and output of the AI technology evolves over time through use and accumulation of new data.

For those AI technologies that are developed locally by a health care system or by a provider, the requirements listed below (as well as those required of programmers and designers – see Annex I), should be more readily satisfied. Such local development of AI technologies may be preferable, especially to ensure the technology itself, and the training data used, is appropriate for the local context. Yet such local development should place certain obligations for a provider to conduct appropriate due diligence before using such technology with a patient.

* + - 1. Transparency of the AI technology

Any AI technology that is deployed must be sufficiently transparent to be open to criticism whether by the public or by internal review teams.

*Specific requirements to promote transparency:*

* Full disclosure of the source code
* Algorithms must be open to be critiqued by an in-house or other appropriate expert
* Sufficient understanding of what data was used to train the algorithm, whether certain groups were systematically excluded from such training data, how training data was labelled and by whom (including expertise and appropriateness of labelling).
* Underlying principles for decision trees should be transparent
* Learned code should be shared
  + - 1. Potential for bias

Bias based on historical or on-going discrimination can be replicated. An AI technology should only be used if such bias can be mitigated and AI should be designed to reduce inequities and bias.

*Specific requirements to address bias:*

* Examine effect of ethnicity or whether to exclude certain ethnic groups from using an AI technology
* Consider majority and minority groups included in the data and whether any under-representation that results in bias can be mitigated
* Where it is not possible to remedy such bias, ensure that this is transparently stated and reflected within decision-making processes (e.g. that a human can take this into consideration)

1. Privacy by design

Health care providers must guard against re-identification, especially where datasets can be linked by third parties to reidentify individuals.

*Specific requirements to safeguard privacy:*

* Understand any issues related to privacy and reverse engineering
* Ensure no leakage of identifiable information

1. Regular challenge and review

Even if an AII technology is deemed appropriate up front, it must be subject to regular challenge and review. This may be necessary due to software erosion, changes in context over time, and changes in the AI technology itself as it continues to learn from new data and evolves.

*Specific requirements for periodic review:*

* Establish a regular technical review process including external review
* Review whether AI is having the intended impact and is improving healthcare

## Is the context in which this AI technology will be used appropriate?

* + - 1. Prerequisites to be able to make the most of AI in health care

For a particular health care service or system, there should be an assessment as to whether it is appropriate to deploy AI.

*Specific requirements for consideration:*

* Ensure that there is sufficient electronic (health) data undergirding the AI technology.
* Ensure that there is the infrastructure to support the use of an AI technology.
* There should be the support of experts, including partnerships with academic organisations and with commercial entities with appropriate agreements with respect to intellectual property, accountability, confidentiality, ethics, access, and commercialisation.
* There should be commonly agreed ethical principles.
  + - 1. Accountability (liability) for decisions supported by AI

There should be clear rules in place that can assess and assign liability – including product liability, personal liability of decision-makers, input liability, and liability to data donors.

*Specific requirements for accountability:*

* Understand accountability of the AI developer/vendor, the clinician, and the health provider organisation that selected and purchases an AI technology.
  + - 1. Understanding local perspectives

There must be an understanding of local perspectives of consumers, particularly indigenous data sovereignty and the collective benefit for the people (*kotahitanga*). This involves determining whether the service has the ‘social license’ to use AI (consumers permission).

*Specific requirements:*

* There should be adequate public and consumer communication and education around AI
* A provider should ascertain and secure a ‘social license’ with consumers
* A provider should address issues of indigenous data sovereignty and governance with its population

## Should a health care provider use the AI technology?

1. AI is interpretable

When implemented, AI needs to be interpreted by a clinician. Human judgement will be critical, and context is key.

*Specific requirements for interpretability:*

* Clinicians need to understand the key data and variables so they can explain the AI to themselves, colleagues, patients, and families.

1. Understand the level of risk

Decisions made by clinicians with AI must be accompanied by a certain level of transparency and understanding that is appropriate or commensurate to any risks.

*Specific requirements with respect to risk;*

* AI should only be used at first for low-risk decisions, leaving high-risk options and decisions to humans initially and until an organisation achieves a certain level of maturity.

1. Responsible use of AI

A health care provider must not only ensure whether an AI technology is technically accurate but also consider whether it is responsible to use AI.

*Specific requirements:*

* Health care providers should be specific about what makes AI the appropriate solution for a particular use.

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

1. Matheny M, Israni ST, Ahmed M, Whicher D (editors), Artificial intelligence in health care: The hope, the hype, the promise, the peril. National Academy of Medicine. 2019 (http: <https://nam.edu/wp-content/uploads/2019/12/AI-in-Health-Care-PREPUB-FINAL.pdf>, accessed August 2020 [↑](#footnote-ref-1)
2. National Health Service. Artificial intelligence: How to get it right. October 2019. (http:// <https://www.nhsx.nhs.uk/media/documents/NHSX_AI_report.pdf>, accessed August 2020) [↑](#footnote-ref-2)
3. Future Advocacy and Wellcome Trust. Ethical, social and political challenges of Artificial Intelligence in Health. 2018. (http:// <https://wellcome.ac.uk/sites/default/files/ai-in-health-ethical-social-political-challenges.pdf>, accessed August 2020). [↑](#footnote-ref-3)
4. Matheny M, Israni ST, Ahmed M, Whicher D (editors), Artificial intelligence in health care: The hope, the hype, the promise, the peril. National Academy of Medicine. 2019 (http: <https://nam.edu/wp-content/uploads/2019/12/AI-in-Health-Care-PREPUB-FINAL.pdf>, accessed August 2020) [↑](#footnote-ref-4)
5. Hao, K, What is machine learning?, MIT Technology Review, 17 November 2018. (http:// <https://www.technologyreview.com/2018/11/17/103781/what-is-machine-learning-we-drew-you-another-flowchart/>, accessed August 2020) [↑](#footnote-ref-5)
6. Flynn, L, When AI is watching patient care: Ethics to consider. Bill of Health. February 2020. (https:/: <https://blog.petrieflom.law.harvard.edu/2020/02/18/when-ai-is-watching-patient-care-ethics-to-consider/>, accessed August 2020). [↑](#footnote-ref-6)
7. Altman RB, AI in medicine: The spectrum of challenges from modern care to molecular medicine. AI Magazine. Volume 20. No. 3. 1999. [↑](#footnote-ref-7)
8. Reddy S, Fox J, Purohit MP, Artificial intelligence-enabled delivery, Journal of the Royal Society of Medicine; 0(0) 1–7 DOI: 10.1177/0141076818815510. [↑](#footnote-ref-8)
9. Miller RA, Schaffner KF, Meisel A. Ethical and legal issues related to the use of computer programs in clinical medicine. Ann Int Med 1985;102:529-36. [↑](#footnote-ref-9)
10. Bi WL, Hosny A, Schabath MB, et al. Artificial intelligence in cancer imaging: Clinical challenges and applications. *CA Cancer J Clin*. 2019;69(2):127-157. doi:10.3322/caac.21552 [↑](#footnote-ref-10)
11. Liu X, Faes L, Kale AU, Wagner SK, Fu DJ, Bruynseels A, et al. A comparison of deep learning performance against health-care professionals in detecting diseases from medical imaging: a systematic review and meta-analysis. Lancet Digital Health. 2019;1:6. [↑](#footnote-ref-11)
12. Rajpurkar P, Irvin J, Ball RL, Zhu K, Yang B, Mehta H, et al. Deep learning for chest radiograph diagnosis: a retrospective comparison of the CheXNeXt algorithm to practicing radiologists. PLoS Med. 2018;15(11):1002686. [↑](#footnote-ref-12)
13. Bejnordi BE, Veta M, van Diest PJ, van Ginneken P, Karssemeijer N, Litjens J, et al. Diagnostic assessment of deep learning algorithms for detection of lymph node metastases in women with breast cancer. JAMA. 2017;318(22):2199–210.  [↑](#footnote-ref-13)
14. Alsharqi M, Woodward WJ, Mumith JA, Markham DC, Upton R, Leeson P. Artificial intelligence and echocardiography. Echo Res Pract. 2018;5(4):R115–25. [↑](#footnote-ref-14)
15. [Bedi G](https://www.ncbi.nlm.nih.gov/pubmed/?term=Bedi%2520G%255BAuthor%255D&cauthor=true&cauthor_uid=27336038), [Carrillo F](https://www.ncbi.nlm.nih.gov/pubmed/?term=Carrillo%2520F%255BAuthor%255D&cauthor=true&cauthor_uid=27336038), [Cecchi GA](https://www.ncbi.nlm.nih.gov/pubmed/?term=Cecchi%2520GA%255BAuthor%255D&cauthor=true&cauthor_uid=27336038), Slezak DF, Sigman M, Mota NB, et al. Automated analysis of free speech predicts psychosis onset in high-risk youths. NPJ Schizophr. 2015;1:15030. [↑](#footnote-ref-15)
16. The Topol Review: Preparing the healthcare workforce to deliver the digital future. National Health Service. 2019. (<https://topol.hee.nhs.uk/>, accessed August 2020) [↑](#footnote-ref-16)
17. Id. [↑](#footnote-ref-17)
18. Id. [↑](#footnote-ref-18)
19. Presentation by Christian Stammel, Wearable Technologies. 6th Mach 2020 for World Health Organisation working group on ethics and governance of AI in health. [↑](#footnote-ref-19)
20. ‘What is biosurveillance? The COVID-19 measures getting under our skin.’ Digital Freedom Fund. 28th May 2020. (http: <https://medium.com/digital-freedom-fund/what-is-biosurveillance-c8bffe70d16f>, accessed October 2020). [↑](#footnote-ref-20)
21. Chapter 3: Roski, J., W. Chapman, J. Heffner, R. Trivedi, G. Del Fiol, R. Kukafka, P. Bleicher, H. Estiri, J. Klann, and J. Pierce. “How Artificial Intelligence is Changing Health and Health Care.” In Artificial Intelligence in Health Care: The Hope, the Hype, the Promise, the Peril, edited by Michael Matheny, Sonoo Thadaney Israni, Mahnoor Ahmed, and Danielle Whicher. Washington, DC: National Academy of Medicine, 2019. [↑](#footnote-ref-21)
22. Hollander JE, Carr BG, Virtually Perfect? Telemedicine for COVID-19. April 30, 2020. N Engl J Med 2020; 382:1679-1681. DOI: 10.1056/NEJMp2003539 [↑](#footnote-ref-22)
23. National Health Service. Artificial intelligence: How to get it right. October 2019. (http:// <https://www.nhsx.nhs.uk/media/documents/NHSX_AI_report.pdf>, accessed August 2020) [↑](#footnote-ref-23)
24. WHO guidelines on ethical issues in public health surveillance. World Health Organisation. 2017. (<https://apps.who.int/iris/bitstream/handle/10665/255721/9789241512657-eng.pdf;jsessionid=935C008CB079C96FDF2B401C485B49A6?sequence=1>, accessed September 2020). [↑](#footnote-ref-24)
25. Chapter 3: Roski, J., W. Chapman, J. Heffner, R. Trivedi, G. Del Fiol, R. Kukafka, P. Bleicher, H. Estiri, J. Klann, and J. Pierce. “How Artificial Intelligence is Changing Health and Health Care.” In Artificial Intelligence in Health Care: The Hope, the Hype, the Promise, the Peril, edited by Michael Matheny, Sonoo Thadaney Israni, Mahnoor Ahmed, and Danielle Whicher. Washington, DC: National Academy of Medicine, 2019. [↑](#footnote-ref-25)
26. Ginsberg J, Mohebbi M, Patel R, Brammer L, Smolinski MS, Brilliant L. Detecting influenza epidemics using search engine query data. Nature. 2009;457:1012–4. [↑](#footnote-ref-26)
27. Cho, A, Artificial intelligence systems aim to sniff out signs of COVID-19 outbreaks, Science. 12 May 2020. ([https://www.sciencemag.org/news/2020/05/artificial-intelligence-systems-aim-sniff-out-signs-covid-19-outbreaks#](https://www.sciencemag.org/news/2020/05/artificial-intelligence-systems-aim-sniff-out-signs-covid-19-outbreaks), accessed August 2020) [↑](#footnote-ref-27)
28. Hswen Y, Brownstein JS. Real-time digital surveillance of vaping-induced pulmonary disease. NEJM, 2019;381:1778–80 [↑](#footnote-ref-28)
29. White RW, Wang S, Pant A, Harpaz R, Shukla P, Sun W, et al. Early identification of adverse drug reactions from search log data. **J Biomed Informatics.** 2016;59:42–8. [↑](#footnote-ref-29)
30. Precision FDA: Gaining new insights by detecting adverse event anomalies using FDA Open Data. US Food and Drug Administration. (<https://precision.fda.gov/challenges/9>, accessed September 2020). [↑](#footnote-ref-30)
31. Cho, A, Artificial intelligence systems aim to sniff out signs of COVID-19 outbreaks, Science. 12 May 2020. ([https://www.sciencemag.org/news/2020/05/artificial-intelligence-systems-aim-sniff-out-signs-covid-19-outbreaks#](https://www.sciencemag.org/news/2020/05/artificial-intelligence-systems-aim-sniff-out-signs-covid-19-outbreaks), accessed August 2020) [↑](#footnote-ref-31)
32. Id. [↑](#footnote-ref-32)
33. Id. [↑](#footnote-ref-33)
34. Whitelaw S, Mamas MA, Topol E, Van Spall GC, Applications of Digital Technology in COVID-19 planning and response, Lancet Digital Health 2020; 2: e435–40 Published Online June 29, 2020 https://doi.org/10.1016/ S2589-7500(20)30142-4 [↑](#footnote-ref-34)
35. Bullock J, Luccioni A, Pham KH, Nga Lam CS, Luengo-Oroz M, Mapping the landscape of artificial intelligence applications against COVID-19, arXiv:2003.11336v2 [cs.CY] 23 Apr 2020. (<https://arxiv.org/pdf/2003.11336.pdf>, accessed September 2020). [↑](#footnote-ref-35)
36. Hao, K., Doctors are using AI to triage COVID-19 patients. The tools may be here to stay.’ MIT Technology Review. 23 April 2020. (<https://www.technologyreview.com/2020/04/23/1000410/ai-triage-covid-19-patients-health-care/>, accessed October 2020). [↑](#footnote-ref-36)
37. AI and control of COVID-19 coronavirus, Council of Europe. (<https://www.coe.int/en/web/artificial-intelligence/ai-and-control-of-covid-19-coronavirus>, accessed September 2020). [↑](#footnote-ref-37)
38. Olson, P., Coronavirus reveals limits of AI health tools. Wall Street Journal. 29 February 2020. (<https://www.wsj.com/articles/coronavirus-reveals-limits-of-ai-health-tools-11582981201>, accessed October 2020). [↑](#footnote-ref-38)
39. Horowitz BT, ‘Are medical chatbots able to detect coronavirus?’, Health Tech Magazine. 10 September 2020. (<https://healthtechmagazine.net/article/2020/09/are-medical-chatbots-able-detect-coronavirus>, accessed October 2020). [↑](#footnote-ref-39)
40. Whitelaw S, Mamas MA, Topol E, Van Spall GC, Applications of Digital Technology in COVID-19 planning and response, Lancet Digital Health 2020; 2: e435–40 Published Online June 29, 2020 https://doi.org/10.1016/ S2589-7500(20)30142-4 [↑](#footnote-ref-40)
41. Fleming, N., How artificial intelligence is changing drug discovery, Nature. 30 May 2018. (<https://www.nature.com/articles/d41586-018-05267-x>, accessed October 2020). [↑](#footnote-ref-41)
42. New Ebola treatment using artificial intelligence. San Francisco (CA): Atomwise; 2015 (https://www.atomwise.com/2015/03/24/new-ebola-treatment-using-artificial-intelligence/, accessed February 2020). [↑](#footnote-ref-42)
43. Fleming, N., How artificial intelligence is changing drug discovery, Nature. 30 May 2018. (<https://www.nature.com/articles/d41586-018-05267-x>, accessed October 2020). [↑](#footnote-ref-43)
44. Shaywtiz D. Novartis CEO who wanted to bring tech into pharma now explains why it’s so hard. Forbes, 16 January 2019 (https://www.forbes.com/sites/davidshaywitz/2019/01/16/novartis-ceo-who-wanted-to-bring-tech-into-pharma-now-explains-why-its-so-hard/#3ee3e45c7fc4, accessed February 2020). [↑](#footnote-ref-44)
45. Fleming, N., How artificial intelligence is changing drug discovery, Nature. 30 May 2018. (<https://www.nature.com/articles/d41586-018-05267-x>, accessed October 2020). [↑](#footnote-ref-45)
46. Report of the IBC on Big Data and Health. UNESCO. 15 September 2017. See: <http://unescoblob.blob.core.windows.net/pdf/UploadCKEditor/REPORT%20OF%20THE%20IBC%20ON%20BIG%20DATA%20AND%20HEALTH%20%2015.09.17.pdf>, accessed November 2020. [↑](#footnote-ref-46)
47. Morozov, Evgeny, To Save Everything, Click Here. Public Affairs. 2014. [↑](#footnote-ref-47)
48. Matheny, M., S. Thadaney Israni, M. Ahmed, and D. Whicher, Editors. 2019. Artificial Intelligence in Health Care: The Hope, the Hype, the Promise, the Peril. NAM Special Publication. Washington, DC: National Academy of Medicine. See: <https://nam.edu/artificial-intelligence-special-publication/>, accessed November 2020. [↑](#footnote-ref-48)
49. Gasser, Urs, Marcello, Ienca, Scheibner, James, Sleigh, Joanna, and Vayena, Effy, Digital tools against COVID-19: taxonomy, ethical challenges, and navigation aid. The Lancet. Volume 2, Issue 8. See: <https://www.thelancet.com/journals/landig/article/PIIS2589-7500(20)30137-0/fulltext> [↑](#footnote-ref-49)
50. Id. [↑](#footnote-ref-50)
51. Ethical, social, and political challenges of artificial intelligence in health. Future Advocacy. April 2018. See: <https://wellcome.ac.uk/funding/ethical-social-and-political-challenges-using-artificial-intelligence-healthcare>. Accessed November 2020; [↑](#footnote-ref-51)
52. Artificial intelligence: How to get it right. NHSx. October 2019. See: <https://www.nhsx.nhs.uk/media/documents/NHSX_AI_report.pdf>, accessed November 2020. [↑](#footnote-ref-52)
53. London AJ. Groundhog Day for Medical Artificial Intelligence. Hastings Cent Rep. 2018 May;48(3):inside back cover. doi: 10.1002/hast.842. PMID: 29806902. [↑](#footnote-ref-53)
54. London AJ. Groundhog Day for Medical Artificial Intelligence. Hastings Cent Rep. 2018 May;48(3):inside back cover. doi: 10.1002/hast.842. PMID: 29806902. [↑](#footnote-ref-54)
55. Report of the Secretary General on SDG Progress (2019): Special Edition. United Nations. See: <https://sustainabledevelopment.un.org/content/documents/24978Report_of_the_SG_on_SDG_Progress_2019.pdf> [↑](#footnote-ref-55)
56. Wahl B, Cossy-Gantner A, Germann S, et al. Artificial intelligence (AI) and global health: how can AI contribute to health in resourcepoor settings? BMJ Glob Health 2018;3:e000798. doi:10.1136/ bmjgh-2018-000798 [↑](#footnote-ref-56)
57. Singh, Jerome. (2019). Artificial Intelligence and global health: opportunities and challenges. Emerging Topics in Life Sciences. 3. 10.1042/ETLS20190106. [↑](#footnote-ref-57)
58. Schwalbe N., Wahl, B., Artificial intelligence and the future of global health. The Lancet. 395: 1579–86. See: <https://www.thelancet.com/action/showPdf?pii=S0140-6736%2820%2930226-9> [↑](#footnote-ref-58)
59. Schwalbe N., Wahl, B., Artificial intelligence and the future of global health. The Lancet. 395: 1579–86. See: <https://www.thelancet.com/action/showPdf?pii=S0140-6736%2820%2930226-9> [↑](#footnote-ref-59)
60. Xiong Y, Ba X, Hou A, Zhang K, Chen L, Li T. Automatic detection of mycobacterium tuberculosis using artificial intelligence. *J Thorac Dis*. 2018;10(3):1936-1940. doi:10.21037/jtd.2018.01.91 [↑](#footnote-ref-60)
61. Singh, Jerome. (2019). Artificial Intelligence and global health: opportunities and challenges. Emerging Topics in Life Sciences. 3. 10.1042/ETLS20190106. [↑](#footnote-ref-61)
62. Wahl B, Cossy-Gantner A, Germann S, et al. Artificial intelligence (AI) and global health: how can AI contribute to health in resource poor settings? BMJ Glob Health 2018;3:e000798. doi:10.1136/ bmjgh-2018-000798 [↑](#footnote-ref-62)
63. Presentation from Dr Amel Ghoulia. WHO Expert Group on Ethics and Governance of AI in health. 3rd October 2019. [↑](#footnote-ref-63)
64. Schwalbe N., Wahl, B., Artificial intelligence and the future of global health. The Lancet. 395: 1579–86. See: <https://www.thelancet.com/action/showPdf?pii=S0140-6736%2820%2930226-9> [↑](#footnote-ref-64)
65. Schwalbe N., Wahl, B., Artificial intelligence and the future of global health. The Lancet. 395: 1579–86. See: <https://www.thelancet.com/action/showPdf?pii=S0140-6736%2820%2930226-9> [↑](#footnote-ref-65)
66. In tech-driven 21st century, achieving global development goals requires closing digital gender divide. UN News. 15 March 2019. See: <https://news.un.org/en/story/2019/03/1034831>. Accessed November 2020. [↑](#footnote-ref-66)
67. The Age of Digital Interdependence: Report of the UN Secretary General’s High-Level Panel on Digital Cooperation. United Nations. June 2019. See: <https://www.un.org/en/pdfs/HLP%20on%20Digital%20Cooperation%20Report%20Executive%20Summary%20-%20ENG.pdf>, accessed November 2020. [↑](#footnote-ref-67)
68. SDG7: Access to affordable, reliable, sustainable and modern energy for all. International Energy Agency (IEA). October 2020. See: <https://www.iea.org/reports/sdg7-data-and-projections/access-to-electricity>, accessed November 2020. [↑](#footnote-ref-68)
69. Winslow J., America’s Digital Divide. Pew Trust Article. 26 July 2019. (<https://www.pewtrusts.org/en/trust/archive/summer-2019/americas-digital-divide>, accessed September 2020). [↑](#footnote-ref-69)
70. Id. [↑](#footnote-ref-70)
71. Vayena E., Blassime A., Biomedical big data: New models of control over access, use, and governance, Bioethical Inquiry (2017) 14:501–513 DOI 10.1007/s11673-017-9809-6 [↑](#footnote-ref-71)
72. <https://www.who.int/ehealth/resources/ecosystem.pdf?ua=1> [↑](#footnote-ref-72)
73. Presentation by Andrew Morris for WHO Expert Group on AI and Health, October 2019. [↑](#footnote-ref-73)
74. Id. [↑](#footnote-ref-74)
75. Matheny, M., S. Thadaney Israni, D. Whicher, and M. Ahmed. “Artificial Intelligence in Health Care: The Hope, the Hype, the Promise, the Peril.” In Artificial Intelligence in Health Care: The Hope, the Hype, the Promise, the Peril, edited by Michael Matheny, Sonoo Thadaney Israni, Mahnoor Ahmed, and Danielle Whicher. Washington, DC: National Academy of Medicine, 2019. [↑](#footnote-ref-75)
76. McNair, D. and W. N. Price. “Health Care AI: Law, Regulation, and Policy.” In Artificial Intelligence in Health Care: The Hope, the Hype, the Promise, the Peril, edited by Michael Matheny, Sonoo Thadaney Israni, Mahnoor Ahmed, and Danielle Whicher. Washington, DC: National Academy of Medicine, 2019. [↑](#footnote-ref-76)
77. Id. [↑](#footnote-ref-77)
78. Id. [↑](#footnote-ref-78)
79. Xafis V et al, An ethics framework for big data in health and research, Asian Bioethics Review (2019) 11:227–254 https://doi.org/10.1007/s41649-019-00099-x [↑](#footnote-ref-79)
80. White Paper: On artificial intelligence – A European approach to excellence and trust. European Commission. 19February 2020. See: <https://ec.europa.eu/info/sites/info/files/commission-white-paper-artificial-intelligence-feb2020_en.pdf>, accessed November 2020. [↑](#footnote-ref-80)
81. Mozur, Paul, Zhong, Raymond, Krolik, Aaron, In Coronavirus Fight, China gives Citizens a Color Code, with red flags. New York Times. 1 March 2020. See: <https://www.nytimes.com/2020/03/01/business/china-coronavirus-surveillance.html>. [↑](#footnote-ref-81)
82. Angwadi centres, digital tracking in India’s blueprint for COVID-19 vaccine drive. The Wire (Science). 6 November 2020. See: <https://science.thewire.in/health/anganwadi-centres-digital-tracking-in-indias-blueprint-for-covid-19-vaccination-drive/> [↑](#footnote-ref-82)
83. Immunity passports and COVID-19: An explainer. Privacy International. 21 July 2020. See: <https://privacyinternational.org/explainer/4075/immunity-passports-and-covid-19-explainer>. Accessed November 2020. [↑](#footnote-ref-83)
84. The looming disaster of immunity passports and digital identity. Privacy International. 21 July 2020. See: <https://privacyinternational.org/long-read/4074/looming-disaster-immunity-passports-and-digital-identity>. Accessed November 2020. [↑](#footnote-ref-84)
85. Gasser, Urs, Marcello, Ienca, Scheibner, James, Sleigh, Joanna, and Vayena, Effy, Digital tools against COVID-19: taxonomy, ethical challenges, and navigation aid. The Lancet. Volume 2, Issue 8. See: <https://www.thelancet.com/journals/landig/article/PIIS2589-7500(20)30137-0/fulltext> [↑](#footnote-ref-85)
86. Price, W. N., 2nd, & Cohen, I. G. (2019). Privacy in the age of medical big data. *Nature medicine*, *25*(1), 37–43. https://doi.org/10.1038/s41591-018-0272-7 [↑](#footnote-ref-86)
87. Copeland, Rob, Google’s Project Nightingale gathers personal health data on millions of Americans. Wall Street Journal. 11 November 2019. See: <https://www.wsj.com/articles/google-s-secret-project-nightingale-gathers-personal-health-data-on-millions-of-americans-11573496790>, accessed November 2020. [↑](#footnote-ref-87)
88. Fussell, Sidney, Google’s totally creepy, totally legal health-data harvesting. The Atlantic. 14 November 2019. See: <https://www.theatlantic.com/technology/archive/2019/11/google-project-nightingale-all-your-health-data/601999/>, accessed November 2020. [↑](#footnote-ref-88)
89. Lewis, Paul, Conn, David, and Pegg, David, UK government using confidential patient data in coronavirus response. The Guardian. 12 April 2020. See: <https://www.theguardian.com/world/2020/apr/12/uk-government-using-confidential-patient-data-in-coronavirus-response>, accessed November 2020. [↑](#footnote-ref-89)
90. Ross, Carey, At Mayo Clinic, sharing patient data with companies fuels AI innovation — and concerns about consent. STAT News. 3 June 2020. See: <https://www.statnews.com/2020/06/03/mayo-clinic-patient-data-fuels-artificial-intelligence-consent-concerns/>, accessed November 2020. [↑](#footnote-ref-90)
91. Vayena E. and Blasimme A., Biomedical Big Data: New Models of Control over Access, Use, and Governance. Bioethical Inquiry; (2017). 14: 501-513. DOI 10.1007/s11673-017-9809-6 [↑](#footnote-ref-91)
92. Id. [↑](#footnote-ref-92)
93. Id. [↑](#footnote-ref-93)
94. Krutzinna J., Taddeo M., Floridi L. (2019) Enabling Posthumous Medical Data Donation: A Plea for the Ethical Utilisation of Personal Health Data. In: Krutzinna J., Floridi L. (eds) The Ethics of Medical Data Donation. Philosophical Studies Series, vol 137. Springer, Cham. https://doi.org/10.1007/978-3-030-04363-6\_11 [↑](#footnote-ref-94)
95. Shaw DM, Gross JV, Erren TC, Why you should donate your health data (as well as your organs) when you die. STAT News. 14 February 2017. (<https://www.statnews.com/2017/02/14/donate-health-data-death/>, accessed September 2020). [↑](#footnote-ref-95)
96. Id. [↑](#footnote-ref-96)
97. Id. [↑](#footnote-ref-97)
98. Vayena E. and Blasimme A., Biomedical Big Data: New Models of Control over Access, Use, and Governance. Bioethical Inquiry; (2017). 14: 501-513. DOI 10.1007/s11673-017-9809-6 [↑](#footnote-ref-98)
99. At a glance: De-identification, anonymisation, and pseudo-anonymisation under the GDPR. Bryan Cave Leighton Paisner. 24 July 2017. (<https://www.bclplaw.com/en-US/insights/at-a-glance-de-identification-anonymization-and-pseudonymization-1.html>, accessed September 2020) [↑](#footnote-ref-99)
100. General Data Protection Regulation, Article 5. [↑](#footnote-ref-100)
101. Bari, Lisa and O’Neill, Daniel, Rethinking Patient Data Privacy in the Era of Digital Health. Health Affairs. 12 December 2019. See: <https://www.healthaffairs.org/do/10.1377/hblog20191210.216658/full/>, accessed November 2020. [↑](#footnote-ref-101)
102. Rocher, L., Hendrickx, J.M. & de Montjoye, Y. Estimating the success of re-identifications in incomplete datasets using generative models. *Nat Commun* **10,**3069 (2019). https://doi.org/10.1038/s41467-019-10933-3 [↑](#footnote-ref-102)
103. Vayena E. and Blasimme A., Biomedical Big Data: New Models of Control over Access, Use, and Governance. Bioethical Inquiry; (2017). 14: 501-513. DOI 10.1007/s11673-017-9809-6 [↑](#footnote-ref-103)
104. Hariri, Yuval, How to survive the 21st Century. World Economic Forum. 24 January 2020. See: <https://www.weforum.org/agenda/2020/01/yuval-hararis-warning-davos-speech-future-predications/>, accessed November 2020; [↑](#footnote-ref-104)
105. Grote T, Berens P, On the ethics of algorithmic decision-making in healthcare, *Journal of Medical Ethics*2020;**46:**205-211. See: <https://jme.bmj.com/content/46/3/205> [↑](#footnote-ref-105)
106. Id. [↑](#footnote-ref-106)
107. Metcalf, Jacob, Moss, Emmanuel and Boyd, Dana, Owning Ethics: Corporate Logics, Silicon Valley, and the Institutionalisation of Ethics. Social Research: An International Quarterly, Volume 82, Issue 2, Summer 2019, 449 – 476. See: <https://muse.jhu.edu/article/732185> [↑](#footnote-ref-107)
108. Yeung, Karen. A study of the implications of advanced digital technologies (including AI systems) for the concept of responsibility within a human rights framework. Council of Europe. September 2019. See: <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3286027> [↑](#footnote-ref-108)
109. Whitaker, Meredith, Crawford, Kate et al., AI Now Report 2018. AI Now Institute. December 2018. See: <https://ainowinstitute.org/AI_Now_2018_Report.pdf>. [↑](#footnote-ref-109)
110. Yeung, Karen. A study of the implications of advanced digital technologies (including AI systems) for the concept of responsibility within a human rights framework. Council of Europe. September 2019. See: <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3286027> [↑](#footnote-ref-110)
111. Whitaker, Meredith, Crawford, Kate et al., AI Now Report 2018. AI Now Institute. December 2018. See: <https://ainowinstitute.org/AI_Now_2018_Report.pdf>. [↑](#footnote-ref-111)
112. Vincent, James, The Problem with Ethics: Is Big Tech’s embrace of AI Ethics Boards actually helping anyone. The Verge. 3 April 2019. See: <https://www.theverge.com/2019/4/3/18293410/ai-artificial-intelligence-ethics-boards-charters-problem-big-tech> [↑](#footnote-ref-112)
113. Vincent, James, The Problem with Ethics: Is Big Tech’s embrace of AI Ethics Boards actually helping anyone. The Verge. 3 April 2019. See: <https://www.theverge.com/2019/4/3/18293410/ai-artificial-intelligence-ethics-boards-charters-problem-big-tech> [↑](#footnote-ref-113)
114. Yeung, Karen. A study of the implications of advanced digital technologies (including AI systems) for the concept of responsibility within a human rights framework. Council of Europe. September 2019. See: <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3286027> [↑](#footnote-ref-114)
115. Id. [↑](#footnote-ref-115)
116. Habli, Ibrahim, Lawton, Tom, Porter, Zoe, Artificial intelligence in healthcare: Accountability and safety. Bull World Health Organ 2020;98:251–256. doi: <http://dx.doi.org/10.2471/BLT.19.237487> [↑](#footnote-ref-116)
117. Hurtgen Holger et. al, Rethinking AI talent strategy as automated machine learning comes of age, McKinsey and Company. 14 August 2020. See: <https://www.mckinsey.com/business-functions/mckinsey-analytics/our-insights/rethinking-ai-talent-strategy-as-automated-machine-learning-comes-of-age> Accessed November 2020. [↑](#footnote-ref-117)
118. Yeung, Karen. A study of the implications of advanced digital technologies (including AI systems) for the concept of responsibility within a human rights framework. Council of Europe. September 2019. See: <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3286027> [↑](#footnote-ref-118)
119. Id. [↑](#footnote-ref-119)
120. Braun M, Hummel P, Beck S, et al. J Med Ethics Epub ahead of print: [please include Day Month Year]. doi:10.1136/ medethics-2019-105860 [↑](#footnote-ref-120)
121. Artificial intelligence in healthcare. Academy of Medical Royal Colleges. January 2019. See: <https://www.aomrc.org.uk/reports-guidance/artificial-intelligence-in-healthcare/>. [↑](#footnote-ref-121)
122. Id. [↑](#footnote-ref-122)
123. Habli, Ibrahim, Lawton, Tom, Porter, Zoe, Artificial intelligence in healthcare: Accountability and safety. Bull World Health Organ 2020;98:251–256. doi: <http://dx.doi.org/10.2471/BLT.19.237487> [↑](#footnote-ref-123)
124. Id. [↑](#footnote-ref-124)
125. Yeung, Karen. A study of the implications of advanced digital technologies (including AI systems) for the concept of responsibility within a human rights framework. Council of Europe. September 2019. See: <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3286027> [↑](#footnote-ref-125)
126. Artificial intelligence in healthcare. Academy of Medical Royal Colleges. January 2019. See: <https://www.aomrc.org.uk/reports-guidance/artificial-intelligence-in-healthcare/>. [↑](#footnote-ref-126)
127. In brief: Artificial intelligence in health care. Swedish National Council on Medical Ethics. <https://smer.se/en/publications/?date=2020-5> [↑](#footnote-ref-127)
128. Grote T, Berens P, On the ethics of algorithmic decision-making in healthcare, *Journal of Medical Ethics*2020;**46:**205-211. See: <https://jme.bmj.com/content/46/3/205> [↑](#footnote-ref-128)
129. <https://twitter.com/EricTopol/status/1082363519675248640/photo/1> [↑](#footnote-ref-129)
130. London, Alex John, Groundhog Day for Medical Artificial Intelligence. The Hastings Center Report. 26 May 2018. <https://doi.org/10.1002/hast.842> [↑](#footnote-ref-130)
131. The Topol Review: Preparing the healthcare workforce to deliver the digital future. The National Health Service. February 2019. See: <https://topol.hee.nhs.uk/>, accessed November 2020. [↑](#footnote-ref-131)
132. Grote T, Berens P, On the ethics of algorithmic decision-making in healthcare, *Journal of Medical Ethics*2020;**46:**205-211. See: <https://jme.bmj.com/content/46/3/205> [↑](#footnote-ref-132)
133. Id. [↑](#footnote-ref-133)
134. Id. [↑](#footnote-ref-134)
135. Artificial intelligence in healthcare. Academy of Medical Royal Colleges. January 2019. See: <https://www.aomrc.org.uk/reports-guidance/artificial-intelligence-in-healthcare/>. [↑](#footnote-ref-135)
136. Pasquale, Frank, The Black Box Society: The Secret Algorithms that control Money and Information. Harvard University Press. 2015. [↑](#footnote-ref-136)
137. London, Alex John, Artificial intelligence and Black-box medical decisions: Accuracy versus Explainability. The Hastings Center Report. 21 February 2019. See: <https://onlinelibrary.wiley.com/doi/abs/10.1002/hast.973>, accessed November 2020. [↑](#footnote-ref-137)
138. Artificial intelligence in healthcare. Academy of Medical Royal Colleges. January 2019. See: <https://www.aomrc.org.uk/reports-guidance/artificial-intelligence-in-healthcare/>. [↑](#footnote-ref-138)
139. Grote T, Berens P, On the ethics of algorithmic decision-making in healthcare, *Journal of Medical Ethics*2020;**46:**205-211. See: <https://jme.bmj.com/content/46/3/205> [↑](#footnote-ref-139)
140. Grote T, Berens P, On the ethics of algorithmic decision-making in healthcare, *Journal of Medical Ethics*2020;**46:**205-211. See: <https://jme.bmj.com/content/46/3/205> [↑](#footnote-ref-140)
141. Cohen IG, Informed consent and medical artificial intelligence: What to tell the patient?. Georgetown Law Journal. Vol. 108: 1425. (<https://www.law.georgetown.edu/georgetown-law-journal/wp-content/uploads/sites/26/2020/06/Cohen_Informed-Consent-and-Medical-Artificial-Intelligence-What-to-Tell-the-Patient.pdf>, accessed September 2020) [↑](#footnote-ref-141)
142. In brief: Artificial intelligence in health care. Swedish National Council on Medical Ethics. <https://smer.se/en/publications/?date=2020-5> [↑](#footnote-ref-142)
143. In brief: Artificial intelligence in health care. Swedish National Council on Medical Ethics. <https://smer.se/en/publications/?date=2020-5> [↑](#footnote-ref-143)
144. Artificial intelligence in healthcare. Academy of Medical Royal Colleges. January 2019. See: <https://www.aomrc.org.uk/reports-guidance/artificial-intelligence-in-healthcare/>. [↑](#footnote-ref-144)
145. Mullin, Emily, Healthcare is the next battleground for Big Tech. One Zero. 27 January 2020. See: <https://onezero.medium.com/health-care-is-the-next-battleground-for-big-tech-477a7263974> [↑](#footnote-ref-145)
146. Turea, Marina, How the Big 4 Tech companies are leading innovation. Healthcare Weekly. 27 February 2019. See: <https://healthcareweekly.com/how-the-big-4-tech-companies-are-leading-healthcare-innovation/> [↑](#footnote-ref-146)
147. A look back at Alphabet’s moves in 2019. MobiHealthNews. 13 December 2019. See: <https://healthcareweekly.com/how-the-big-4-tech-companies-are-leading-healthcare-innovation/>, accessed November 2020. [↑](#footnote-ref-147)
148. Our work with Google Health UK. NHS Royal Free London. See: <https://www.royalfree.nhs.uk/patients-visitors/how-we-use-patient-information/our-work-with-deepmind/> [↑](#footnote-ref-148)
149. Bridging gaps in healthcare industry with technology. Tencent. 2019. See: <https://www.tencent.com/en-us/articles/2200933.html>. Accessed November 2020. [↑](#footnote-ref-149)
150. How Baidu, Alibaba, and Tencent aim to disrupt Chinese health care. Forkast. 28 January 2020. See: <https://forkast.news/baidu-alibaba-tencent-china-health-care-blo/> [↑](#footnote-ref-150)
151. Jourdan, Adam. AI ambulances and robot doctors: China seeks digital salve to ease hospital strain. Reuters. 28 June 2018. See: <https://de.reuters.com/article/us-china-healthcare-tech-idUKKBN1JO1VB>. Accessed November 2020. [↑](#footnote-ref-151)
152. Ding Y, Sohn JH, Kawczynski MG, Trivedi H, Harnish R, Jenkins NW, Lituiev D, Copeland TP, Aboian MS, Mari Aparici C, Behr SC, Flavell RR, Huang SY, Zalocusky KA, Nardo L, Seo Y, Hawkins RA, Hernandez Pampaloni M, Hadley D, Franc BL. A Deep Learning Model to Predict a Diagnosis of Alzheimer Disease by Using 18F-FDG PET of the Brain. Radiology. 2019 Feb;290(2):456-464. doi: 10.1148/radiol.2018180958. Epub 2018 Nov 6. PMID: 30398430; PMCID: PMC6358051. [↑](#footnote-ref-152)
153. Morley J, Caio C, Machado V, Burr C, Cowls J, Joshi I, Taddeo M, Floridi F, The debate on the ethics of AI in health care: a reconstruction and critical review. <https://digitalethicslab.oii.ox.ac.uk/wp-content/uploads/sites/87/2019/11/The-Debate-on-the-EThics-of-AI-in-Health-Care-pre-print-.pdf> [↑](#footnote-ref-153)
154. Oveido Convention, Article 10(3) [↑](#footnote-ref-154)
155. Ethical, Social, and Political Challenges of Artificial Intelligence in Health. Future Advocacy (A Report with the Wellcome Trust). April 2018. See: <https://wellcome.org/grant-funding/ethical-social-and-political-challenges-using-artificial-intelligence-healthcare> [↑](#footnote-ref-155)
156. Chen, J. H., A. Beam, S. Saria, and E. Mendonça. “Potential Trade-Offs and Unintended Consequences of AI.” In Artificial Intelligence in Health Care: The Hope, the Hype, the Promise, the Peril, edited by Michael Matheny, Sonoo Thadaney Israni, Mahnoor Ahmed, and Danielle Whicher. Washington, DC: National Academy of Medicine, 2019. [↑](#footnote-ref-156)
157. Angwin, Julia, Larson, Jeff, Mattu, Surya, and Kirchner, Lauren, Machine Bias. Pro Publica. 23 May 2016. See: <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>, accessed November 2020. [↑](#footnote-ref-157)
158. West, Sarah Myers, Whitaker, Meredith, Crawford, Kate, Discriminating Systems: Gender, Power, and Race in AI. AI Now Institute. April 2019. See: <https://ainowinstitute.org/discriminatingsystems.pdf>, accessed November 2020. [↑](#footnote-ref-158)
159. Rose, Eveleth, How Self-Tracking Apps Exclude Women. The Atlantic. 15 December 2014. See: <https://www.theatlantic.com/technology/archive/2014/12/how-self-tracking-apps-exclude-women/383673/> [↑](#footnote-ref-159)
160. ‘Training data’ is the data that is used to teach an algorithm how to perform a certain task [↑](#footnote-ref-160)
161. Simonite, Tom, How an algorithm blocked kidney transplants to patients. Wired Magazine. 26 October 2020. See: <https://www.wired.com/story/how-algorithm-blocked-kidney-transplants-black-patients/>, accessed November 2020. [↑](#footnote-ref-161)
162. Bridging the Digital gender divide: Include, Upskill, Innovate. OECD. 2018. See: <http://www.oecd.org/digital/bridging-the-digital-gender-divide.pdf>, accessed November 2020. [↑](#footnote-ref-162)
163. Munshi, Neil, How unlocking the secrets of African DNA could change the world. Financial Times. 5 March 2020. See: <https://www.ft.com/content/eed0555c-5e2b-11ea-b0ab-339c2307bcd4>, accessed November 2020. [↑](#footnote-ref-163)
164. Devlin, Hannah, Genetics research ‘biased towards studying white Europeans’. The Guardian. 8 October 2018. See: <https://www.theguardian.com/science/2018/oct/08/genetics-research-biased-towards-studying-white-europeans>, accessed November 2020. [↑](#footnote-ref-164)
165. Yuval Noah Hariri, Homo Deus: A Brief History of Tomorrow. 2015. Vintage. [↑](#footnote-ref-165)
166. The Topol Review: Preparing the healthcare workforce to deliver the digital future. The National Health Service. February 2019. See: <https://topol.hee.nhs.uk/>, accessed November 2020. [↑](#footnote-ref-166)
167. Chen, J. H., A. Beam, S. Saria, and E. Mendonça. “Potential Trade-Offs and Unintended Consequences of AI.” In Artificial Intelligence in Health Care: The Hope, the Hype, the Promise, the Peril, edited by Michael Matheny, Sonoo Thadaney Israni, Mahnoor Ahmed, and Danielle Whicher. Washington, DC: National Academy of Medicine, 2019. [↑](#footnote-ref-167)
168. The Topol Review: Preparing the healthcare workforce to deliver the digital future. The National Health Service. February 2019. See: <https://topol.hee.nhs.uk/>, accessed November 2020. [↑](#footnote-ref-168)
169. Chen, J. H., A. Beam, S. Saria, and E. Mendonça. “Potential Trade-Offs and Unintended Consequences of AI.” In Artificial Intelligence in Health Care: The Hope, the Hype, the Promise, the Peril, edited by Michael Matheny, Sonoo Thadaney Israni, Mahnoor Ahmed, and Danielle Whicher. Washington, DC: National Academy of Medicine, 2019. [↑](#footnote-ref-169)
170. Davenport T, Kalakota R. The potential for artificial intelligence in healthcare. *Future Healthc J*. 2019;6(2):94-98. doi:10.7861/futurehosp.6-2-94 [↑](#footnote-ref-170)
171. Davenport T, Kalakota R. The potential for artificial intelligence in healthcare. *Future Healthc J*. 2019;6(2):94-98. doi:10.7861/futurehosp.6-2-94 [↑](#footnote-ref-171)
172. Matheny, M., S. Thadaney Israni, D. Whicher, and M. Ahmed. “Artificial Intelligence in Health Care: The Hope, the Hype, the Promise, the Peril.” In Artificial Intelligence in Health Care: The Hope, the Hype, the Promise, the Peril, edited by Michael Matheny, Sonoo Thadaney Israni, Mahnoor Ahmed, and Danielle Whicher. Washington, DC: National Academy of Medicine, 2019. [↑](#footnote-ref-172)
173. Parikh, Rahul K., Should doctors play along with the uberization of health care. Slate. 14 June 2017. See: <https://slate.com/technology/2017/06/should-doctors-play-along-with-the-uberization-of-health-care.html>, accessed November 2020. [↑](#footnote-ref-173)
174. Matheny, M., S. Thadaney Israni, D. Whicher, and M. Ahmed. “Artificial Intelligence in Health Care: The Hope, the Hype, the Promise, the Peril.” In Artificial Intelligence in Health Care: The Hope, the Hype, the Promise, the Peril, edited by Michael Matheny, Sonoo Thadaney Israni, Mahnoor Ahmed, and Danielle Whicher. Washington, DC: National Academy of Medicine, 2019. [↑](#footnote-ref-174)
175. Gawande, Atul, Why doctors hate their computers. The New Yorker. 5 Novembner 2018. See: <https://www.newyorker.com/magazine/2018/11/12/why-doctors-hate-their-computers> [↑](#footnote-ref-175)
176. Powles, Julia and Hodson, Hal, Google Deepmind and healthcare in the age of algorithms. Health Technol. DOI 10.1007/s12553-017-0179-1. [↑](#footnote-ref-176)
177. Powles, Julia and Hodson, Hal, Google Deepmind and healthcare in the age of algorithms. Health Technol. DOI 10.1007/s12553-017-0179-1. [↑](#footnote-ref-177)
178. Ballantyne, Angela and Stewart, Cameron, Big Data and Public-Private partnerships in healthcare and research. Asian Bioethics Review (2019) 11:315–326 <https://doi.org/10.1007/s41649-019-00100-7> [↑](#footnote-ref-178)
179. Shosana Zuboff. The Age of Surveillance Capitalism. 2019. Principle Books. [↑](#footnote-ref-179)
180. Copeland, Rob, Google’s Project Nightingale gathers personal health data on millions of Americans. Wall Street Journal. 11 November 2019. See: <https://www.wsj.com/articles/google-s-secret-project-nightingale-gathers-personal-health-data-on-millions-of-americans-11573496790>, accessed November 2020. [↑](#footnote-ref-180)
181. Hodson, Hal, Revealed: Google AI has access to huge haul of NHS patient data. New Scientist. 29 April 2016. See: <https://www.newscientist.com/article/2086454-revealed-google-ai-has-access-to-huge-haul-of-nhs-patient-data/> [↑](#footnote-ref-181)
182. Rocher, L., Hendrickx, J.M. & de Montjoye, Y. Estimating the success of re-identifications in incomplete datasets using generative models. *Nat Commun* **10,**3069 (2019). https://doi.org/10.1038/s41467-019-10933-3 [↑](#footnote-ref-182)
183. Durkee, Alison, Facebook to pay millions for allegedly mishandling user data (again). Vanity Fair. 30 January 2020. See: <https://www.vanityfair.com/news/2020/01/facebook-settlement-facial-recognition-illinois-privacy>, accessed November 2020. [↑](#footnote-ref-183)
184. Mergers: Commission opens in-depth investigation into the proposed acquisition of Fitbit by Google. European Commission. 4 August 2020. See: <https://ec.europa.eu/commission/presscorner/detail/en/ip_20_1446>, accessed November 2020. [↑](#footnote-ref-184)
185. Jourdan, Adam. AI ambulances and robot doctors: China seeks digital salve to ease hospital strain. Reuters. 28 June 2018. See: <https://de.reuters.com/article/us-china-healthcare-tech-idUKKBN1JO1VB>. Accessed November 2020. [↑](#footnote-ref-185)
186. Bridging gaps in healthcare industry with technology. Tencent. 2019. See: <https://www.tencent.com/en-us/articles/2200933.html> [↑](#footnote-ref-186)
187. Jourdan, Adam. AI ambulances and robot doctors: China seeks digital salve to ease hospital strain. Reuters. 28 June 2018. See: <https://de.reuters.com/article/us-china-healthcare-tech-idUKKBN1JO1VB>. Accessed November 2020. [↑](#footnote-ref-187)
188. Powles, Julia and Hodson, Hal, Google Deepmind and healthcare in the age of algorithms. Health Technol. DOI 10.1007/s12553-017-0179-1. [↑](#footnote-ref-188)
189. Veale, Michael, Privacy is not the problem with the Google-Apple contact-tracing toolkit. The Guardian. 1 July 2020. See: <https://www.theguardian.com/commentisfree/2020/jul/01/apple-google-contact-tracing-app-tech-giant-digital-rights>, accessed November 2020. [↑](#footnote-ref-189)
190. Shosana Zuboff. The Age of Surveillance Capitalism. 2019. Principle Books. [↑](#footnote-ref-190)
191. Yuval Noah Hariri, Homo Deus: A Brief History of Tomorrow. 2015. Vintage. [↑](#footnote-ref-191)
192. Ballantyne, Angela and Stewart, Cameron, Big Data and Public-Private partnerships in healthcare and research. Asian Bioethics Review (2019) 11:315–326 <https://doi.org/10.1007/s41649-019-00100-7> [↑](#footnote-ref-192)
193. Powles, Julia and Hodson, Hal, Google Deepmind and healthcare in the age of algorithms. Health Technol. DOI 10.1007/s12553-017-0179-1. [↑](#footnote-ref-193)
194. Why we need strong regulatory systems to reach universal health coverage. World Health Organization. 1 July 2019. See: <https://www.who.int/medicines/news/2019/strong-reg-systems-to-reach-UHC/en/> [↑](#footnote-ref-194)
195. Artificial intelligence in healthcare. Academy of Medical Royal Colleges. January 2019. See: <https://www.aomrc.org.uk/reports-guidance/artificial-intelligence-in-healthcare/>. [↑](#footnote-ref-195)
196. Id. [↑](#footnote-ref-196)
197. Tomasev, Nenad et al, A clinically applicable approach to continuous prediction of future acute kidney injury. Nature. 31 July 2019. 572: 116-119. See: <https://www.nature.com/articles/s41586-019-1390-1> [↑](#footnote-ref-197)
198. Hern, Alex, Fault in NHS Covid App meant thousands at risk did not quarantine. The Guardian. 2 November 2020. See: <https://www.theguardian.com/world/2020/nov/02/fault-in-nhs-covid-app-meant-thousands-at-risk-did-not-quarantine> [↑](#footnote-ref-198)
199. Contag, Moritz et al, How they did it: An analysis of emission defeat devices in modern automobiles. See: <https://cseweb.ucsd.edu/~klevchen/diesel-sp17.pdf/> Accesses November 2020. [↑](#footnote-ref-199)
200. Baraniuk, Chris, How tech bugs could be killing thousands in our hospitals. New Scientist. 16 May 2018. See; <https://www.newscientist.com/article/mg23831781-700-how-tech-bugs-could-be-killing-thousands-in-our-hospitals/> [↑](#footnote-ref-200)
201. In brief: Artificial intelligence in health care. Swedish National Council on Medical Ethics. <https://smer.se/en/publications/?date=2020-5> [↑](#footnote-ref-201)
202. In brief: Artificial intelligence in health care. Swedish National Council on Medical Ethics. <https://smer.se/en/publications/?date=2020-5> [↑](#footnote-ref-202)
203. Chen, J. H., A. Beam, S. Saria, and E. Mendonça. “Potential Trade-Offs and Unintended Consequences of AI.” In Artificial Intelligence in Health Care: The Hope, the Hype, the Promise, the Peril, edited by Michael Matheny, Sonoo Thadaney Israni, Mahnoor Ahmed, and Danielle Whicher. Washington, DC: National Academy of Medicine, 2019. [↑](#footnote-ref-203)
204. Vayena E, Haeusermann T, Adjekum A, Blasimme A. Digital health: meeting the ethical and policy challenges. Swiss Med Wkly. 2018 Jan 16;148:w14571. doi: 10.4414/smw.2018.14571. PMID: 29376547. [↑](#footnote-ref-204)
205. Jobin, A., Ienca, M. & Vayena, E. The global landscape of AI ethics guidelines. *Nat Mach Intell* **1,**389–399 (2019). https://doi.org/10.1038/s42256-019-0088-2 [↑](#footnote-ref-205)
206. Digital health. A/71/A/CONF./1. World Health Organization. 71st World Health Assembly. Agenda Item 12.4. See: <https://apps.who.int/gb/ebwha/pdf_files/WHA71/A71_ACONF1-en.pdf> [↑](#footnote-ref-206)
207. Addressing the impact of algorithms on human rights. Council of Europe. See: <https://rm.coe.int/draft-recommendation-of-the-committee-of-ministers-to-states-on-the-hu/168095eecf> [↑](#footnote-ref-207)
208. The Toronto Declaration. Amnesty International and Access Now. 16 May 2018. See: <https://www.torontodeclaration.org/declaration-text/english/> [↑](#footnote-ref-208)
209. European Convention on Human Rights. Council of Europe. See: <https://www.echr.coe.int/documents/convention_eng.pdf> [↑](#footnote-ref-209)
210. Convention for the protection of Human Rights and Dignity of the Human Being with regard to the Application of Biology and Medicine: Convention on Human Rights and Biomedicine. Council of Europe. 1997. See: https://rm.coe.int/CoERMPublicCommonSearchServices/DisplayDCTMContent?documentId=090000168007cf98 [↑](#footnote-ref-210)
211. Convention for the Protection of Individuals with regard to Automatic Processing of Personal Data. Council of Europe. 1981. See: <https://rm.coe.int/1680078b37> [↑](#footnote-ref-211)
212. Guidelines on Artificial Intelligence and Data Protection. Council of Europe. 2019. See: <https://rm.coe.int/guidelines-on-artificial-intelligence-and-data-protection/168091f9d8> [↑](#footnote-ref-212)
213. European ethical Charter on the use of Artificial Intelligence in judicial systems and their environment. Council of Europe. December 2018. See: <https://rm.coe.int/ethical-charter-en-for-publication-4-december-2018/16808f699c> [↑](#footnote-ref-213)
214. ### Zeng Y, Lu E, Huangfu C. Linking artificial intelligence principles. Proceedings of the AAAI Workshop on Artificial Intelligence Safety, Honolulu, Hawaii, 2019. Aachen: CEUR-WS; 2019 (https://arxiv.org/ftp/arxiv/papers/1812/1812.04814.pdf, accessed February 2020).

     [↑](#footnote-ref-214)
215. Jobin, A., Ienca, M. & Vayena, E. The global landscape of AI ethics guidelines. *Nat Mach Intell* **1,**389–399 (2019). https://doi.org/10.1038/s42256-019-0088-2 [↑](#footnote-ref-215)
216. OECD legal instruments. Recommendations of the Council on Artificial Intelligence. Paris: Organization for

     International Co-operation and Development; 2020 (<https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL>

     0449, accessed February 2020). [↑](#footnote-ref-216)
217. Going digital. Making the transformation work for growth and well-being. Paris: Organization for International Co-operation and Development; 2019 (https://www.oecd.org/going-digital/ai/principles/, accessed February 2020). [↑](#footnote-ref-217)
218. OECD AI Policy Observatory. Paris: Organization for International Co-operation and Development; 2019 (https://www.oecd.org/going-digital/ai/about-the-oecd-ai-policy-observatory.pdf, accessed February 2020). [↑](#footnote-ref-218)
219. General Recommendations for the Processing of Personal Data in Artificial Intelligence. Red IberoAmerica de Proteccion de Datos. 21 June 2019. See : <https://www.redipd.org/sites/default/files/2020-02/guide-general-recommendations-processing-personal-data-ai.pdf> [↑](#footnote-ref-219)
220. Specific guidelines for compliance with the Principles and Rights that govern the protection of personal data in artificial intelligence projects. Red IberoAmerica de Proteccion de Datos. 21 June 2019. See: <https://www.redipd.org/sites/default/files/2020-02/guide-specific-guidelines-ai-projects.pdf> [↑](#footnote-ref-220)
221. Unboxing Artificial Intelligence: 10 Steps to Protect Human Rights. Council of Europe. May 2019. See:

     <https://rm.coe.int/unboxing-artificial-intelligence-10-steps-to-protect-human-rights-reco/1680946e64> [↑](#footnote-ref-221)
222. Ethics guidelines for trustworthy AI. Brussels: European Commission; 2018. [↑](#footnote-ref-222)
223. AI Utilisation guidelines. See : <https://www.oecd.ai/dashboards/policy-initiatives/2019-data-policyInitiatives-24346> [↑](#footnote-ref-223)
224. Singapore wins international award for its artificial intelligence governance and ethics initiatives. Singapore: InfoComm Media Development Authority; 2019 (<https://www.imda.gov.sg/news-and-events/Media-Room/Media-Releases/2019/singapore-wins-international-award-for-its-artificial-intelligence-governance-and-ethics-initiatives>, accessed February 2020). [↑](#footnote-ref-224)
225. Declaration of Astana. Global Conference on Primary Health Care, Astana,25–26 October 2018. Geneva: World Health Organization; 2018 (https://www.who.int/docs/default-source/primary-health/declaration/gcphc-declaration.pdf, accessed February 2020). [↑](#footnote-ref-225)
226. International Bioethics Committee. Report of the IBC on big data and health. Paris: United Nations Educational, Cultural and Scientific Organization; 2017 (https://unesdoc.unesco.org/ark:/48223/pf0000248724, accessed February 2020). [↑](#footnote-ref-226)
227. World Commission on the Ethics of Scientific Knowledge and Technology. Report of COMEST on robotics ethics.

     Paris: United Nations Educational, Cultural and Scientific Organization; 2017

     (https://unesdoc.unesco.org/ark:/48223/pf0000253952, accessed February 2020). [↑](#footnote-ref-227)
228. Preliminary study on the technical and legal aspects relating to the desirability of a standard-setting instrument on the ethics of artificial intelligence. Paris: United Nations Educational, Cultural and Scientific Organization; 2019 (https://unesdoc.unesco.org/ark:/48223/pf0000367422, accessed February 2020). [↑](#footnote-ref-228)
229. Guidance. Code of conduct for data-driven health and care technology. London: Department of Health and Social

     Care; 2019 (<https://www.gov.uk/government/publications/code-of-conduct-for-data-driven-health-and-care->

     technology/initial-code-of-conduct-for-data-driven-health-and-care-technology, accessed February 2020). [↑](#footnote-ref-229)
230. Recommendation CM/Rec(2019)2 of the Committee of Ministers to member States on the protection of health-related data. Council of Europe. 27 March 2019. See: <https://www.apda.ad/sites/default/files/2019-03/CM_Rec%282019%292E_EN.pdf> [↑](#footnote-ref-230)
231. Building on the work of W.D. Ross, Tom L. Beauchamp and James F. Childress formulated a principle-based approach to bioethics in which they added a “principle of respect for autonomy” to Ross’ three other principles. The *Principles of Biomedical Ethics* ([1979] 2001, 5th ed., New York: [Oxford University Press](https://en.wikipedia.org/wiki/Oxford_University_Press)), though highly influential, is not universally accepted as dispositive. [↑](#footnote-ref-231)
232. Public Debate. Council of Europe. See: <https://www.coe.int/en/web/bioethics/public-debate> [↑](#footnote-ref-232)
233. Price WN, Gerke S, Cohen IG. Potential Liability for Physicians Using Artificial Intelligence. JAMA. 2019;322(18):1765–1766. doi:10.1001/jama.2019.15064 [↑](#footnote-ref-233)
234. Id. [↑](#footnote-ref-234)
235. McNair, D. and W. N. Price. “Health Care AI: Law, Regulation, and Policy.” In Artificial Intelligence in Health Care: The Hope, the Hype, the Promise, the Peril, edited by Michael Matheny, Sonoo Thadaney Israni, Mahnoor Ahmed, and Danielle Whicher. Washington, DC: National Academy of Medicine, 2019. [↑](#footnote-ref-235)
236. Price WN, Artificial intelligence in healthcare: Applications and legal implications. University of Michigan Law School. 2017. See: <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2910417> [↑](#footnote-ref-236)
237. Gerke S, Minssen T, Cohen G. Ethical and legal challenges of artificial intelligence-driven healthcare. *Artificial Intelligence in Healthcare*. 2020;295-336. doi:10.1016/B978-0-12-818438-7.00012-5 [↑](#footnote-ref-237)
238. Id. [↑](#footnote-ref-238)
239. Ordish J, Legal liability for machine learning in healthcare. PHG Foundation. August 2018. See: <https://www.phgfoundation.org/documents/briefing-note-legal-liability-for-machine-learning-in-healthcare.pdf>, accessed November 2020. [↑](#footnote-ref-239)
240. Timo Minssen, Marc Mimler, Vivian Mak, When Does Stand-Alone Software Qualify as A Medical Device in the European Union?—The Cjeu’s Decision in *Snitem* and What it Implies for the Next Generation of Medical Devices, *Medical Law Review*, fwaa012, <https://doi.org/10.1093/medlaw/fwaa012> [↑](#footnote-ref-240)
241. Id. [↑](#footnote-ref-241)
242. Report from the Commission to the European Parliament, the Council, and the European Economic and Social Committee: Report on the safety and liability implications of Artificial Intelligence, the Internet of Things and robotics. European Commission. 19 February 2020. See: <https://eur-lex.europa.eu/legal-content/en/TXT/?qid=1593079180383&uri=CELEX%3A52020DC0064> [↑](#footnote-ref-242)
243. Id. [↑](#footnote-ref-243)
244. Price II, William Nicholson, Medical Malpractice and Black-Box Medicine (February 2, 2017). I. Glenn Cohen et al., eds., Big Data, Health Law, and Bioethics (Cambridge University Press, 2018), U of Michigan Public Law Research Paper No. 536, Available at SSRN: <https://ssrn.com/abstract=2910417> [↑](#footnote-ref-244)
245. Husgen J. Product liability suits involving drug or device manufacturers and physicians: the learned intermediary doctrine and the physician's duty to warn. *Mo Med*. 2014;111(6):478-481. [↑](#footnote-ref-245)
246. Id. [↑](#footnote-ref-246)
247. Thomas S, Artificial Intelligence and Medical Liability (Part II), Bill of Health. 10 February 2017. See: <https://blog.petrieflom.law.harvard.edu/2017/02/10/artificial-intelligence-and-medical-liability-part-ii/>, accessed November 2020. [↑](#footnote-ref-247)
248. No Fault compensation in New Zealand: Harmonizing injury compensation, provider accountability, and patient safety. Commonwealth Fund. 2006. [↑](#footnote-ref-248)
249. McNair, D. and W. N. Price. “Health Care AI: Law, Regulation, and Policy.” In Artificial Intelligence in Health Care: The Hope, the Hype, the Promise, the Peril, edited by Michael Matheny, Sonoo Thadaney Israni, Mahnoor Ahmed, and Danielle Whicher. Washington, DC: National Academy of Medicine, 2019. [↑](#footnote-ref-249)
250. Id. [↑](#footnote-ref-250)
251. Vayena E., Blassime A., Biomedical big data: New models of control over access, use, and governance, Bioethical Inquiry (2017) 14:501–513 DOI 10.1007/s11673-017-9809-6 [↑](#footnote-ref-251)
252. Id. [↑](#footnote-ref-252)
253. <https://sagebionetworks.org/tools_resources/elements-of-informed-consent/> [↑](#footnote-ref-253)
254. Vayena E., Blassime A., Biomedical big data: New models of control over access, use, and governance, Bioethical Inquiry (2017) 14:501–513 DOI 10.1007/s11673-017-9809-6 [↑](#footnote-ref-254)
255. I. Glenn Cohen, “Is There a Duty to Share Healthcare Data?,” in *Big Data, Health Law, and Bioethics*, Glenn Cohen et al., 1st ed. (Cambridge University Press, 2018): 209–22. <https://doi.org/10.1017/9781108147972.020>. [↑](#footnote-ref-255)
256. Price, WN, Cohen IG, Privacy in the age of medical big data, Nature Medicine. Vol; 25: January 2019. (<https://www.webdepot.umontreal.ca/Usagers/p1097493/MonDepotPublic/NicholsonPrice-Cohen2019.pdf>, accessed September 2020). [↑](#footnote-ref-256)
257. Id. [↑](#footnote-ref-257)
258. Spector-Bagdady K, Hutchinson R, O'Brien Kaleba E, Kheterpal S. Sharing Health Data and Biospecimens with Industry - A Principle-Driven, Practical Approach. N Engl J Med. 2020 May 28;382(22):2072-2075. doi: 10.1056/NEJMp1915298. PMID: 32459920. [↑](#footnote-ref-258)
259. Id. [↑](#footnote-ref-259)
260. Id. [↑](#footnote-ref-260)
261. Wilkinson, M., Dumontier, M., Aalbersberg, I. *et al.* The FAIR Guiding Principles for scientific data management and stewardship. *Sci Data* **3,**160018 (2016). <https://doi.org/10.1038/sdata.2016.18> [↑](#footnote-ref-261)
262. Creating the right framework to realise the benefits for patients and the NHS where data underpins innovation, United Kingdom Department of Health and Social Care. 15 July 2019. See: <https://www.gov.uk/government/publications/creating-the-right-framework-to-realise-the-benefits-of-health-data/creating-the-right-framework-to-realise-the-benefits-for-patients-and-the-nhs-where-data-underpins-innovation>, accessed November 2020. [↑](#footnote-ref-262)
263. Bhunia, Priyankar. ata Futures Partnership in New Zealand issues guidelines for organisations to develop social license for data use. Open Gov. 27 October 2017. See: [https://opengovasia.com/data-futures-partnership-in-new-zealand-issues-guidelines-for-organisations-to-develop-social-license-for-data-use/#:~:text=The%20Data%20Futures%20Partnership%20is,control%20data%2Dsharing%20ecosystem.%E2%80%9D](https://opengovasia.com/data-futures-partnership-in-new-zealand-issues-guidelines-for-organisations-to-develop-social-license-for-data-use/%23:~:text=The%20Data%20Futures%20Partnership%20is,control%20data%2Dsharing%20ecosystem.%E2%80%9D) [↑](#footnote-ref-263)
264. Id. [↑](#footnote-ref-264)
265. Vayena E., Blassime A., Biomedical big data: New models of control over access, use, and governance, Bioethical Inquiry (2017) 14:501–513 DOI 10.1007/s11673-017-9809-6 [↑](#footnote-ref-265)
266. National Data Guardian: What we do. Government of the United Kingdom. <https://www.gov.uk/government/organisations/national-data-guardian/about>. Accessed November 2020. [↑](#footnote-ref-266)
267. Id. [↑](#footnote-ref-267)
268. Our charter: Tutohinga. Te Mana Raraunga (Maori Data Sovereignty Network). See: <https://www.temanararaunga.maori.nz/tutohinga>. Accessed November 2020. [↑](#footnote-ref-268)
269. Id. [↑](#footnote-ref-269)
270. Schnarch, B., Ownership, Access, Control, and Possession (OCAP) or Self-Determination Applied to Research. International Journal of Indigenous Health. Vol 1, No 1 (2004). See: <https://jps.library.utoronto.ca/index.php/ijih/article/view/28934>, accessed November 2020. [↑](#footnote-ref-270)
271. Vayena E., Blassime A., Biomedical big data: New models of control over access, use, and governance, Bioethical Inquiry (2017) 14:501–513 DOI 10.1007/s11673-017-9809-6 [↑](#footnote-ref-271)
272. Id. [↑](#footnote-ref-272)
273. National Institutes of Health. Genomic Data Sharing Policy (28 August 2014). (<https://www.federalregister.gov/documents/2014/08/28/2014-20385/final-nih-genomic-data-sharing-policy>, accessed September 2020) [↑](#footnote-ref-273)
274. Majumder MA, Guerrini CJ, Bollinger JM, Deegan RC, McGuire AL, Sharing data under the 21st Century Cures Act, Genetics in Medicine, Vol. 19, Number 12, December 2017. (<https://www.nature.com/articles/gim201759.pdf?origin=ppub>, accessed September 2020). [↑](#footnote-ref-274)
275. HHS finalises historic rules to provide patients with more control of their patient data. U.S. Department of Health and Human Services. 9th March 2020. (<https://www.hhs.gov/about/news/2020/03/09/hhs-finalizes-historic-rules-to-provide-patients-more-control-of-their-health-data.html>, access September 2020) [↑](#footnote-ref-275)
276. Id. [↑](#footnote-ref-276)
277. National Institutes of Health: All of Us Research Program. See: <https://allofus.nih.gov/>, accessed November 2020. [↑](#footnote-ref-277)
278. European Health Data Space. European Commission. See: <https://ec.europa.eu/health/ehealth/dataspace_en>, accessed November 2020. [↑](#footnote-ref-278)
279. Digital Innovation Hub Program Prospectus Appendix: Principles for Participation. Health Data Research UK. May 2019. See: <https://www.hdruk.ac.uk/wp-content/uploads/2019/07/Digital-Innovation-Hub-Programme-Prospectus-Appendix-Principles-for-Participation.pdf> accessed November 2020. [↑](#footnote-ref-279)
280. Sharing Sensitive Health Data in a Federated Data Consortium Model An Eight-Step Guide, World Economic Forum. July 2020. See: <http://www3.weforum.org/docs/WEF_Sharing_Sensitive_Health_Data_2020.pdf>, accessed November 2020. [↑](#footnote-ref-280)
281. Id. [↑](#footnote-ref-281)
282. Ross, Carey, At Mayo Clinic, sharing patient data with companies fuels AI innovation — and concerns about consent. STAT News. 3 June 2020. See: <https://www.statnews.com/2020/06/03/mayo-clinic-patient-data-fuels-artificial-intelligence-consent-concerns/>, accessed November 2020. [↑](#footnote-ref-282)
283. Ross, Carey, At Mayo Clinic, sharing patient data with companies fuels AI innovation — and concerns about consent. STAT News. 3 June 2020. See: <https://www.statnews.com/2020/06/03/mayo-clinic-patient-data-fuels-artificial-intelligence-consent-concerns/>, accessed November 2020. [↑](#footnote-ref-283)
284. Id. [↑](#footnote-ref-284)
285. Ornstein, Charles and Thomas, Katie, Sloan Kettering’s Cozy Deal with Start-up ignites uproar. New York Times. 20 September 2018. See: <https://www.nytimes.com/2018/09/20/health/memorial-sloan-kettering-cancer-paige-ai.html>, accessed November 2020. [↑](#footnote-ref-285)
286. The Economist, “The World’s Most Valuable Resource Is No Longer Oil, but Data,” 6 May 2017, (https://www.economist. com/news/leaders/21721656-data-economy-demands-new-approach-antitrust-rules-worlds-most-valuable-resource, accessed August 2020). [↑](#footnote-ref-286)
287. Rajan, A., Data is not the new oil. BBC News Online. 9 October 2017; (<https://www.bbc.com/news/entertainment-arts-41559076>, accessed September 2020) [↑](#footnote-ref-287)
288. Marr B., Here’s why data is not the new oil. Forbes. 5 March 2018. (<https://www.forbes.com/sites/bernardmarr/2018/03/05/heres-why-data-is-not-the-new-oil/#6a65a1453aa9>, accessed September 2020) [↑](#footnote-ref-288)
289. Hilty, R., Big Data: Ownership and use in the Digital Age. Intellectual property and Digital Trade in the Age of Artificial Intelligence and Big Data. Global Perspectives and Challenges for the Intellectual Property System. Issue No. 5: June 2018. Edited by Xavier Seuba, Christopher Geiger and Julien Penin. (<http://www.ceipi.edu/fileadmin/upload/DUN/CEIPI/Documents/Publications_CEIPI___ICTSD/CEIPI-ICTSD_Issue_5_Final.pdf>, accessed August 2020). [↑](#footnote-ref-289)
290. Minssen T. and Pierce J., Big Data and Intellectual Property in the Health and Life Sciences, Big Data, Health Law, and Bioethics, Cambridge University Press, 2018. [↑](#footnote-ref-290)
291. Andanda, P., Towards a Paradigm Shift in Governing Data Access and Related Intellectual Property Rights in Big Data and Health-Related Research. International Review of Intellectual Property and Competition Law. (2019) 50:1052–1081 (<https://link.springer.com/article/10.1007/s40319-019-00873-2>, accessed August 2020). [↑](#footnote-ref-291)
292. Report of the IBC on Big Data and Health. UNESCO. 15 September 2017. (<http://unescoblob.blob.core.windows.net/pdf/UploadCKEditor/REPORT%20OF%20THE%20IBC%20ON%20BIG%20DATA%20AND%20HEALTH%20%2015.09.17.pdf>, accessed August 2020). [↑](#footnote-ref-292)
293. Andanda, P., Towards a Paradigm Shift in Governing Data Access and Related Intellectual Property Rights in Big Data and Health-Related Research. International Review of Intellectual Property and Competition Law. (2019) 50:1052–1081 (<https://link.springer.com/article/10.1007/s40319-019-00873-2>, accessed August 2020). [↑](#footnote-ref-293)
294. Minssen T and Schovsbo J, Big Data in the Health and Life Sciences: What Are the Challenges for European Competition Law and Where Can They Be Found? Intellectual property and Digital Trade in the Age of Artificial Intelligence and Big Data. Global Perspectives and Challenges for the Intellectual Property System. Issue No. 5: June 2018. Edited by Xavier Seuba, Christopher Geiger and Julien Penin. (<http://www.ceipi.edu/fileadmin/upload/DUN/CEIPI/Documents/Publications_CEIPI___ICTSD/CEIPI-ICTSD_Issue_5_Final.pdf>, accessed August 2020). [↑](#footnote-ref-294)
295. Id. [↑](#footnote-ref-295)
296. Majumder MA, Guerrini CJ, Bollinger JM, Deegan RC, McGuire AL, Sharing data under the 21st Century Cures Act, Genetics in Medicine, Vol. 19, Number 12, December 2017. (<https://www.nature.com/articles/gim201759.pdf?origin=ppub>, accessed September 2020). [↑](#footnote-ref-296)
297. Report of the IBC on Big Data and Health. UNESCO. 15 September 2017. (<http://unescoblob.blob.core.windows.net/pdf/UploadCKEditor/REPORT%20OF%20THE%20IBC%20ON%20BIG%20DATA%20AND%20HEALTH%20%2015.09.17.pdf>, accessed August 2020). [↑](#footnote-ref-297)
298. Andanda, P., Towards a Paradigm Shift in Governing Data Access and Related Intellectual Property Rights in Big Data and Health-Related Research. International Review of Intellectual Property and Competition Law. (2019) 50:1052–1081 (<https://link.springer.com/article/10.1007/s40319-019-00873-2>, accessed August 2020). [↑](#footnote-ref-298)
299. Id. [↑](#footnote-ref-299)
300. Id. [↑](#footnote-ref-300)
301. Report of the IBC on Big Data and Health. UNESCO. 15 September 2017. (<http://unescoblob.blob.core.windows.net/pdf/UploadCKEditor/REPORT%20OF%20THE%20IBC%20ON%20BIG%20DATA%20AND%20HEALTH%20%2015.09.17.pdf>, accessed August 2020). [↑](#footnote-ref-301)
302. Andanda, P., Towards a Paradigm Shift in Governing Data Access and Related Intellectual Property Rights in Big Data and Health-Related Research. International Review of Intellectual Property and Competition Law. (2019) 50:1052–1081 (<https://link.springer.com/article/10.1007/s40319-019-00873-2>, accessed August 2020). [↑](#footnote-ref-302)
303. Abbott R, Inventive Machines: Rethinking Invention and Patentability. Global Perspectives and Challenges for the Intellectual Property System. Issue No. 5: June 2018. Edited by Xavier Seuba, Christopher Geiger and Julien Penin. (<http://www.ceipi.edu/fileadmin/upload/DUN/CEIPI/Documents/Publications_CEIPI___ICTSD/CEIPI-ICTSD_Issue_5_Final.pdf>, accessed August 2020). [↑](#footnote-ref-303)
304. Porter J, US Patent Office rules that artificial intelligence cannot be a legal inventor. The Verge. 29 April 2020. (<https://www.theverge.com/2020/4/29/21241251/artificial-intelligence-inventor-united-states-patent-trademark-office-intellectual-property>, accessed August 2020). [↑](#footnote-ref-304)
305. Aboy M et al, How does emerging patent case law in the US and Europe affect precision medicine?. Nature Biotechnology. Vol 37. October 2019. 1118 – 1126. (<https://www.nature.com/articles/s41587-019-0265-1>, accessed August 2020). [↑](#footnote-ref-305)
306. West DM, The role of corporations in addressing AI’s ethical dilemmas, Brookings, 13 September 2018. (<https://www.brookings.edu/research/how-to-address-ai-ethical-dilemmas/>, accessed August 2020) [↑](#footnote-ref-306)
307. Metcalf E, Moss E, Boyd D, Social Research: An International Quarterly, Volume 82, Issue 2, Summer, 2019, pages 449-476. <https://muse.jhu.edu/article/732185> [↑](#footnote-ref-307)
308. Mittelstadt, Brent. (2019). Principles alone cannot guarantee ethical AI. Nature Machine Intelligence. 10.1038/s42256-019-0114-4. [↑](#footnote-ref-308)
309. Metcalf, Jacob, Moss, Emmanuel and Boyd, Dana, Owning Ethics: Corporate Logics, Silicon Valley, and the Institutionalisation of Ethics. Social Research: An International Quarterly, Volume 82, Issue 2, Summer 2019, 449 – 476. See: <https://muse.jhu.edu/article/732185> [↑](#footnote-ref-309)
310. Id. [↑](#footnote-ref-310)
311. Cath C. 2018 Governing artificial intelligence: ethical, legal and technical opportunities and challenges. Phil. Trans. R. Soc. A 376: 20180080. http://dx.doi.org/10.1098/rsta.2018.0080DO - 10.1038/s42256-019-0114-4 [↑](#footnote-ref-311)
312. Mittelstadt, Brent. (2019). Principles alone cannot guarantee ethical AI. Nature Machine Intelligence. 10.1038/s42256-019-0114-4. [↑](#footnote-ref-312)
313. Id. [↑](#footnote-ref-313)
314. Metcalf E, Moss E, Boyd D, Social Research: An International Quarterly, Volume 82, Issue 2, Summer, 2019, pages 449-476. <https://muse.jhu.edu/article/732185> [↑](#footnote-ref-314)
315. Mittelstadt, Brent. (2019). Principles alone cannot guarantee ethical AI. Nature Machine Intelligence. 10.1038/s42256-019-0114-4. [↑](#footnote-ref-315)
316. Id. [↑](#footnote-ref-316)
317. Id. [↑](#footnote-ref-317)
318. WHO launches a chatbot on Facebook messenger to combat misinformation. World Health Organisation. 15th April 2020. (<https://www.who.int/news-room/feature-stories/detail/who-launches-a-chatbot-powered-facebook-messenger-to-combat-covid-19-misinformation>, accessed September 2020). [↑](#footnote-ref-318)
319. Lee D, Murphy H, Facebook accused of failing to tackle medical hoaxes, Financial Times. 20August 2020. (<https://www.ft.com/content/f33f7d61-a8df-40b9-82a8-75f2a41210bc>, accessed August 2020). [↑](#footnote-ref-319)
320. Id. [↑](#footnote-ref-320)
321. Facebook’s algorithm: A major threat to public health. Avaaz. 19 August 2020. See: <https://avaazimages.avaaz.org/facebook_threat_health.pdf>, accessed November 2020. [↑](#footnote-ref-321)
322. Id. [↑](#footnote-ref-322)
323. International Organization for Standardization. See: <https://www.iso.org/home.html> [↑](#footnote-ref-323)
324. HL7 International. See: <http://www.hl7.org/> [↑](#footnote-ref-324)
325. Goodman KW. *Ethics, Medicine and Information Technology: Intelligent Machines and the Transformation of Health Care.* Cambridge: Cambridge University Press, 2016. [↑](#footnote-ref-325)
326. Ballantyne A., Stewart C, Big Data and Public-Private Partnerships in Healthcare and research, Asian Bioethics Review (2019) 11:315–326 <https://doi.org/10.1007/s41649-019-00100-7> [↑](#footnote-ref-326)
327. Id. [↑](#footnote-ref-327)
328. Alphabet’s Verily plans to use big data as health insurance tool. EBN. 25 August 2020. (<https://www.benefitnews.com/articles/alphabets-verily-plans-to-use-big-data-as-health-insurance-tool>, accessed September 2020) [↑](#footnote-ref-328)
329. Presentation by Alexander Ng to WHO Expert Group on AI in health. July 2020. [↑](#footnote-ref-329)
330. Id. [↑](#footnote-ref-330)
331. Ackroyd, Alison Tudor, Tencent-backed WeDoctor makes IPO appointment in Hong Kong and writes prescription for digital healthcare post-pandemic. South China Morning Post. 4 June 2020. See: <https://www.scmp.com/business/banking-finance/article/3087385/tencent-backed-wedoctor-makes-ipo-appointment-hong-kong> [↑](#footnote-ref-331)
332. Marr, Bernard, The incredible ways in which artificial intelligence is now used in mental health. Forbes. 3 May 2019. See: <https://www.forbes.com/sites/bernardmarr/2019/05/03/the-incredible-ways-artificial-intelligence-is-now-used-in-mental-health/?sh=7806594ad02e> [↑](#footnote-ref-332)
333. Gamble, Alyson, Artificial intelligence and mobile apps for mental healthcare: a social informatics perspective, Artificial Intelligence for Mental Health. 13 May 2020. See: <https://www.emerald.com/insight/content/doi/10.1108/AJIM-11-2019-0316/full/pdf?title=artificial-intelligence-and-mobile-apps-for-mental-healthcare-a-social-informatics-perspective> [↑](#footnote-ref-333)
334. Id. [↑](#footnote-ref-334)
335. OECD. Hello World: Artificial intelligence and its use in the public sector. November 2019. (<https://oecd-opsi.org/wp-content/uploads/2019/11/AI-Report-Online.pdf>, accessed August 2020). [↑](#footnote-ref-335)
336. Id. [↑](#footnote-ref-336)
337. Id. [↑](#footnote-ref-337)
338. UK Government. Assessing if artificial intelligence is the right solution. 10 June 2019. (<https://www.gov.uk/guidance/assessing-if-artificial-intelligence-is-the-right-solution>, accessed August 2020). [↑](#footnote-ref-338)
339. OECD. Hello World: Artificial intelligence and its use in the public sector. November 2019. (<https://oecd-opsi.org/wp-content/uploads/2019/11/AI-Report-Online.pdf>, accessed August 2020). [↑](#footnote-ref-339)
340. Artificial Intelligence and Public Standards. The Committee on Standards in Public Life. February 2020. (<https://www.gov.uk/government/publications/artificial-intelligence-and-public-standards-report>, accessed February 2020). [↑](#footnote-ref-340)
341. Id. [↑](#footnote-ref-341)
342. Id. [↑](#footnote-ref-342)
343. Martinho-Truswell E., How AI could help the public sector, Harvard Business Review, 29 January 2019. (<https://hbr.org/2018/01/how-ai-could-help-the-public-sector>, accessed August 2020). [↑](#footnote-ref-343)
344. Digital technology, social protection and human rights: Report of the UN Special Rapporteur for extreme poverty. https://www.ohchr.org/EN/Issues/Poverty/Pages/DigitalTechnology.aspx [↑](#footnote-ref-344)
345. Id. [↑](#footnote-ref-345)
346. Academy of Medical Royal Colleges, Artificial intelligence in healthcare, January 2019. (<https://www.aomrc.org.uk/wp-content/uploads/2019/01/Artificial_intelligence_in_healthcare_0119.pdf> , accessed August 2020) [↑](#footnote-ref-346)
347. Derrington D., Artificial intelligence for health and healthcare, The Mitre Corporation, December 2017. (<https://www.healthit.gov/sites/default/files/jsr-17-task-002_aiforhealthandhealthcare12122017.pdf>, accessed August 2020) [↑](#footnote-ref-347)
348. See : <https://twitter.com/FTC/status/1285578871803437057> [↑](#footnote-ref-348)
349. Revealed: 50 million Facebook profiles harvested for Cambridge Analytica in major data breach. The Guardian. 17 March 2018. See: <https://www.theguardian.com/news/2018/mar/17/cambridge-analytica-facebook-influence-us-election>. Accessed November 2020. [↑](#footnote-ref-349)
350. Academy of Medical Royal Colleges, Artificial intelligence in healthcare, January 2019. (<https://www.aomrc.org.uk/wp-content/uploads/2019/01/Artificial_intelligence_in_healthcare_0119.pdf> , accessed August 2020) [↑](#footnote-ref-350)
351. Price, W. Nicholson, II. "Artificial Intelligence in Health Care: Applications and Legal Implications." The SciTech Lawyer 14, no. 1 (2017). (<https://repository.law.umich.edu/articles/1932/>, accessed August 2020) [↑](#footnote-ref-351)
352. McNair, D. and W. N. Price. “Health Care AI: Law, Regulation, and Policy.” In Artificial Intelligence in Health Care: The Hope, the Hype, the Promise, the Peril, edited by Michael Matheny, Sonoo Thadaney Israni, Mahnoor Ahmed, and Danielle Whicher. Washington, DC: National Academy of Medicine, 2019. [↑](#footnote-ref-352)
353. Id. [↑](#footnote-ref-353)
354. Price, W. Nicholson, II. "Artificial Intelligence in Health Care: Applications and Legal Implications." The SciTech Lawyer 14, no. 1 (2017). (<https://repository.law.umich.edu/articles/1932/>, accessed August 2020) [↑](#footnote-ref-354)
355. Id. [↑](#footnote-ref-355)
356. McNair, D. and W. N. Price. “Health Care AI: Law, Regulation, and Policy.” In Artificial Intelligence in Health Care: The Hope, the Hype, the Promise, the Peril, edited by Michael Matheny, Sonoo Thadaney Israni, Mahnoor Ahmed, and Danielle Whicher. Washington, DC: National Academy of Medicine, 2019. [↑](#footnote-ref-356)
357. Vayena E, Blasimme A, Cohen IG (2018) Machine learning in medicine: Addressing ethical challenges. PLoS Med 15(11): e1002689. https://doi.org/10.1371/journal.pmed.1002689 [↑](#footnote-ref-357)
358. Id. [↑](#footnote-ref-358)
359. McNair, D. and W. N. Price. “Health Care AI: Law, Regulation, and Policy.” In Artificial Intelligence in Health Care: The Hope, the Hype, the Promise, the Peril, edited by Michael Matheny, Sonoo Thadaney Israni, Mahnoor Ahmed, and Danielle Whicher. Washington, DC: National Academy of Medicine, 2019. [↑](#footnote-ref-359)
360. Morley J, Caio C, Machado V, Burr C, Cowls J, Joshi I, Taddeo M, Floridi F, The debate on the ethics of AI in health care: a reconstruction and critical review. <https://digitalethicslab.oii.ox.ac.uk/wp-content/uploads/sites/87/2019/11/The-Debate-on-the-EThics-of-AI-in-Health-Care-pre-print-.pdf> [↑](#footnote-ref-360)
361. Minssen T., Gerke S, Aboy M, Price N, Cohen G, Regulatory responses to medical machine learning, Journal of Law and the Biosciences, 1–18 doi:10.1093/jlb/lsaa002 [↑](#footnote-ref-361)
362. Minssen T., Gerke S, Aboy M, Price N, Cohen G, Regulatory responses to medical machine learning, Journal of Law and the Biosciences, 1–18 doi:10.1093/jlb/lsaa002 [↑](#footnote-ref-362)
363. Pricewaterhouse Coopers, Sizing the Prize: What’s the real value of AI for your business and can you capitalise? 2019. (<https://www.pwc.com/gx/en/issues/analytics/assets/pwc-ai-analysis-sizing-the-prize-report.pdf>, accessed August 2020) [↑](#footnote-ref-363)
364. Vinuesa, R., Azizpour, H., Leite, I. *et al.* The role of artificial intelligence in achieving the Sustainable Development Goals. *Nat Commun* **11,**233 (2020). <https://doi.org/10.1038/s41467-019-14108-y> [↑](#footnote-ref-364)
365. Artificial intelligence: Canada and France work with international community to support the responsible use of AI. Government of France. 2019. ( <https://www.gouvernement.fr/en/artificial-intelligence-canada-and-france-work-with-international-community-to-support-the>, accessed August 2020.) [↑](#footnote-ref-365)
366. Report of the Secretary General : Roadmap for Digital Cooperation. United Nations. June 2020. (<https://www.un.org/en/content/digital-cooperation-roadmap/assets/pdf/Roadmap_for_Digital_Cooperation_EN.pdf>, accessed September 2020). [↑](#footnote-ref-366)
367. Handbook of Ethics, Values, and Technological Design: Sources, Theories, Values, and Application Domains. Editors van den Hoven J., Vermaas PE, van de Poel I., (2015). See: <https://www.springer.com/gp/book/9789400769694#aboutAuthors> [↑](#footnote-ref-367)
368. Aizenberg E, van den Hoven J. Designing for human rights in AI. Big Data & Society. July 2020. doi:10.1177/2053951720949566

     [↑](#footnote-ref-368)
369. IEEE Standards Association. Statement Regarding the Ethical Implementation of Artificial Intelligence Systems (AIS) for Addressing the COVID-19 Pandemic. See: <https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/gieais-covid.pdf?utm_medium=undefined&utm_source=undefined&utm_campaign=undefined&utm_content=undefined&utm_term=undefined>. Accessed November 2020. [↑](#footnote-ref-369)
370. Independent High-Level Expert Group on Artificial Intelligence. Ethics Guidelines for Trustworthy AI. 8 April 2019. See: <https://ec.europa.eu/futurium/en/ai-alliance-consultation/guidelines#Top>. Accessed November 2020; [↑](#footnote-ref-370)
371. Whitaker K., Citizen Science platform with Autistica. The Alan Turing Institute. (<https://www.turing.ac.uk/research/research-projects/citizen-science-platform-autistica>, accessed November 2020). [↑](#footnote-ref-371)
372. Sage Bionetworks. The Elements of Informed Consent: A Toolkit. V.3. January 2020. (<https://sagebionetworks.org/wp-content/uploads/2020/01/SageBio_EIC-Toolkit_V3_21Jan20_final.pdf>, accessed November 2020). [↑](#footnote-ref-372)
373. Patients’ rights. World Health Organisation. See: <https://www.who.int/genomics/public/patientrights/en/>, accessed November 2020. [↑](#footnote-ref-373)
374. Understanding patient data. London: Wellcome Trust; 2019 (https://understandingpatientdata.org.uk/, accessed February 2020). [↑](#footnote-ref-374)
375. Sharing anonymized patient-level data where there is a mixed public and private benefit – a new report. London: Health Research Authority; 2019 (<https://www.hra.nhs.uk/about-us/news-updates/sharing-anonymised-patient-level-data-where-there-mixed-public-and-private-benefit-new-report/>, accessed February 2020). [↑](#footnote-ref-375)
376. Artificial intelligence and health, Summary report of a roundtable held on 16 January 2019. London: Academy of Medical Sciences; 2019 (https://acmedsci.ac.uk/file-download/77652269, accessed February 2020). [↑](#footnote-ref-376)
377. Our data driven future in healthcare. People and partnerships at the heart of health-related technologies

     London: Academy of Medical Sciences; 2018 (https://acmedsci.ac.uk/file-download/74634438, accessed

     February 2020) [↑](#footnote-ref-377)
378. Siau, Keng & Wang, Weiyu. (2018). Trusting Artificial Intelligence in Healthcare. [↑](#footnote-ref-378)
379. Artificial intelligence in healthcare. Academy of Medical Royal Sciences. January 2019. See: <https://www.aomrc.org.uk/reports-guidance/artificial-intelligence-in-healthcare/>. Accessed November 2020. [↑](#footnote-ref-379)
380. Artificial intelligence in healthcare. Academy of Medical Royal Sciences. January 2019. See: <https://www.aomrc.org.uk/reports-guidance/artificial-intelligence-in-healthcare/>. Accessed November 2020. [↑](#footnote-ref-380)
381. Trustworth AI in health: Background paper for the G20 AI Dialogue, Digital Economy, and Trade. OECD. April 2020. See: <https://www.oecd.org/health/trustworthy-artificial-intelligence-in-health.pdf>. Accessed November 2020. [↑](#footnote-ref-381)
382. Id. [↑](#footnote-ref-382)
383. Fihn, S. D., S. Saria, E. Mendonça, S. Hain, M. Matheny, N. Shah, H. Liu, and A. Auerbach. “Deploying AI in Clinical Settings.” In Artificial Intelligence in Health Care: The Hope, the Hype, the Promise, the Peril, edited by Michael Matheny, Sonoo Thadaney Israni, Mahnoor Ahmed, and Danielle Whicher. Washington, DC: National Academy of Medicine, 2019. [↑](#footnote-ref-383)
384. https://www.coe.int/en/web/bioethics/public-debate [↑](#footnote-ref-384)
385. What is Impact Assessment. Fact Sheet. International Association for Impact Assessment. See: <https://www.iaia.org/uploads/pdf/What_is_IA_web.pdf>. Accessed November 2020. [↑](#footnote-ref-385)
386. Guiding Principles on Business and Human Rights: Implementing the United Nations Protect, Respect, and Remedy Framework. Office of the High Commissioner of Human Rights. 2011. See: <https://www.ohchr.org/documents/publications/guidingprinciplesbusinesshr_en.pdf>. Accessed November 2020. [↑](#footnote-ref-386)
387. Human Rights Impact Assessments. National Action Plans on Business and Human Rights. See: <https://globalnaps.org/issue/human-rights-impact-assessments/>. Accessed November 2020. [↑](#footnote-ref-387)
388. French Corporate Duty of Vigilance Law. Corporate European Observatory. See: <https://corporatejustice.org/documents/publications/french-corporate-duty-of-vigilance-law-faq.pdf>. Accessed November 2020. [↑](#footnote-ref-388)
389. New EU Law requiring human rights due diligence on the cards for 2021. Linklaters. 28 July 2020. See: <https://www.linklaters.com/en/insights/blogs/linkingesg/2020/july/new-eu-law-requiring-human-rights-due-diligence-on-the-cards-for-2021>. Accessed November 2020. [↑](#footnote-ref-389)
390. Algorithmic Impact Assessments: A Practical Framework for Public Agency Accountability.’ AI Now Institute. April 2018. See: <https://www.ftc.gov/system/files/documents/public_comments/2018/08/ftc-2018-0048-d-0044-155168.pdf>. Accessed November 2020. [↑](#footnote-ref-390)
391. MacCarthy M. An Examination of the Algorithmic Accountability Act of 2019. Transatlantic Working Group. 24 October 2019. See: <https://www.ivir.nl/publicaties/download/Algorithmic_Accountability_Oct_2019.pdf>. Accessed November 2020. [↑](#footnote-ref-391)
392. Ibid.  [↑](#footnote-ref-392)
393. Data Protection Impact Assessment. GPPR.EU. See: <https://gdpr.eu/data-protection-impact-assessment-template/>. Accessed November 2020. [↑](#footnote-ref-393)
394. Proposal for a legal act of the European Parliament and the Council laying down requirements for Artificial Intelligence. European Commission. See: <https://ec.europa.eu/info/law/better-regulation/have-your-say/initiatives/12527-Artificial-intelligence-ethical-and-legal-requirements>. Accessed November 2020. [↑](#footnote-ref-394)
395. World Health Organization. EPI-Brain webpage. See: <https://www.epi-brain.com/>. Accessed November 2020. [↑](#footnote-ref-395)
396. OECD. OECD.AI Policy Observatory Fact Sheet. 2020. See: <https://www.oecd.org/going-digital/ai/about-the-oecd-ai-policy-observatory.pdf>, accessed November 2020. [↑](#footnote-ref-396)