

EXECUTIVE SUMMARY

ETHICS AND GOVERNANCE OF ARTIFICIAL INTELLIGENCE FOR HEALTH

WHO GUIDANCE



ETHICS AND GOVERNANCE OF ARTIFICIAL INTELLIGENCE FOR HEALTH EXECUTIVE SUMMARY

Ethics and governance of artificial intelligence for health: WHO quidance Executive summary

ISBN 978-92-4-003740-3 (electronic version) ISBN 978-92-4-003741-0 (print version)

© World Health Organization 2021

Some rights reserved. This work is available under the Creative Commons Attribution-NonCommercial-ShareAlike 3.0 IGO licence (CC BY-NC-SA 3.0 IGO; https://creativecommons.org/licenses/by-nc-sa/3.0/igo).

Under the terms of this licence, you may copy, redistribute and adapt the work for non-commercial purposes, provided the work is appropriately cited, as indicated below. In any use of this work, there should be no suggestion that WHO endorses any specific organization, products or services. The use of the WHO logo is not permitted. If you adapt the work, then you must license your work under the same or equivalent Creative Commons licence. If you create a translation of this work, you should add the following disclaimer along with the suggested citation: "This translation was not created by the World Health Organization (WHO). WHO is not responsible for the content or accuracy of this translation. The original English edition shall be the binding and authentic edition".

Any mediation relating to disputes arising under the licence shall be conducted in accordance with the mediation rules of the World Intellectual Property Organization (http://www.wipo.int/amc/en/mediation/rules/).

Suggested citation. Ethics and governance of artificial intelligence for health: WHO guidance. Executive summary. Geneva: World Health Organization; 2021. Licence: CC BY-NC-SA 3.0 IGO.

Cataloguing-in-Publication (CIP) data. CIP data are available at http://apps.who.int/iris.

Sales, rights and licensing. To purchase WHO publications, see http://apps.who.int/bookorders. To submit requests for commercial use and queries on rights and licensing, see https://www.who.int/about/policies/publishing/copyright.

Third-party materials. If you wish to reuse material from this work that is attributed to a third party, such as tables, figures or images, it is your responsibility to determine whether permission is needed for that reuse and to obtain permission from the copyright holder. The risk of claims resulting from infringement of any third-party-owned component in the work rests solely with the user.

General disclaimers. The designations employed and the presentation of the material in this publication do not imply the expression of any opinion whatsoever on the part of WHO concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries. Dotted and dashed lines on maps represent approximate border lines for which there may not yet be full agreement.

The mention of specific companies or of certain manufacturers' products does not imply that they are endorsed or recommended by WHO in preference to others of a similar nature that are not mentioned. Errors and omissions excepted, the names of proprietary products are distinguished by initial capital letters.

All reasonable precautions have been taken by WHO to verify the information contained in this publication.

However, the published material is being distributed without warranty of any kind, either expressed or implied. The responsibility for the interpretation and use of the material lies with the reader. In no event shall WHO be liable for damages arising from its use.

Graphic design: The New Division

EXECUTIVE SUMMARY

Artificial Intelligence (AI) refers to the ability of algorithms encoded in technology to learn from data so that they can perform automated tasks without every step in the process having to be programmed explicitly by a human. WHO recognizes that AI holds great promise for the practice of public health and medicine. WHO also recognizes that, to fully reap the benefits of AI, ethical challenges for health care systems, practitioners and beneficiaries of medical and public health services must be addressed. Many of the ethical concerns described in this report predate the advent of AI, although AI itself presents a number of novel concerns.

Whether AI can advance the interests of patients and communities depends on a collective effort to design and implement ethically defensible laws and policies and ethically designed AI technologies. There are also potential serious negative consequences if ethical principles and human rights obligations are not prioritized by those who fund, design, regulate or use AI technologies for health. AI's opportunities and challenges are thus inextricably linked.

AI can augment the ability of health-care providers to improve patient care, provide accurate diagnoses, optimize treatment plans, support pandemic preparedness and response, inform the decisions of health policy-makers or allocate resources within health systems. To unlock this potential, health-care workers and health systems must have detailed information on the contexts in which such systems can function safely and effectively, the conditions necessary to ensure reliable, appropriate use, and the mechanisms for continuous auditing and assessment of system performance. Health-care workers and health systems must have access to education and training in order to use and maintain these systems under the conditions for their safe, effective use.

AI can also empower patients and communities to assume control of their own health care and better understand their evolving needs. To achieve this, patients and communities require assurance that their rights and interests will not be subordinated to the powerful commercial interests of technology companies or the interests of governments in surveillance and social control. It also requires that the potential of AI to detect risks to patient or community health is incorporated into health systems in a way that advances human autonomy and dignity and does not displace humans from the centre of health decision-making.

AI can enable resource-poor countries, where patients often have restricted access to health-care workers or medical professionals, to bridge gaps in access to health services. AI systems must be carefully designed to reflect the diversity of socioeconomic and health-care settings and be accompanied by training in digital skills, community engagement and awareness-raising. Systems based primarily on data of

individuals in high-income countries may not perform well for individuals in low- and middle-income settings. Country investments in AI and the supporting infrastructure should therefore help to build effective health-care systems by avoiding AI that encodes biases that are detrimental to equitable provision of and access to health-care services.

This guidance document, produced jointly by WHO's Health Ethics and Governance unit in the department of Research for Health and by the department of Digital Health and Innovation, is based on the collective views of a WHO Expert Group on Ethics and Governance of AI for Health, which comprised 20 experts in public health, medicine, law, human rights, technology and ethics. The group analysed many opportunities and challenges of AI and recommended policies, principles and practices for ethical use of AI for health and means to avoid its misuse to undermine human rights and legal obligations.

AI for health has been affected by the COVID-19 pandemic. Although the pandemic is not a focus of this report, it has illustrated the opportunities and challenges associated with AI for health. Numerous new applications have emerged for responding to the pandemic, while other applications have been found to be ineffective. Several applications have raised ethical concerns in relation to surveillance, infringement on the rights of privacy and autonomy, health and social inequity and the conditions necessary for trust and legitimate uses of data-intensive applications. During their deliberations on this report, members of the expert group prepared interim WHO guidance for the use of proximity tracking applications for COVID-19 contact-tracing.

Key ethical principles for the use of AI for health

This report endorses a set of key ethical principles. WHO hopes that these principles will be used as a basis for governments, technology developers, companies, civil society and inter-governmental organizations to adopt ethical approaches to appropriate use of AI for health. The six principles are summarized below and explained in depth in Section 5.

Protecting human autonomy: Use of AI can lead to situations in which decision-making power could be transferred to machines. The principle of autonomy requires that the use of AI or other computational systems does not undermine human autonomy. In the context of health care, this means that humans should remain in control of health-care systems and medical decisions. Respect for human autonomy also entails related duties to ensure that providers have the information necessary to make safe, effective use of AI systems and that people understand the role that such systems play in their care. It also requires protection of privacy and confidentiality and obtaining valid informed consent through appropriate legal frameworks for data protection.

Promoting human well-being and safety and the public interest. AI technologies should not harm people. The designers of AI technologies should satisfy regulatory requirements for safety, accuracy and efficacy for well-defined use cases or indications. Measures of quality control in practice and quality improvement in the use of AI over time should be available. Preventing harm requires that AI not result in mental or physical harm that could be avoided by use of an alternative practice or approach.

Ensuring transparency, explainability and intelligibility. AI technologies should be intelligible or understandable to developers, medical professionals, patients, users and regulators. Two broad approaches to intelligibility are to improve the transparency of AI technology and to make AI technology explainable. Transparency requires that sufficient information be published or documented before the design or deployment of an AI technology and that such information facilitate meaningful public consultation and debate on how the technology is designed and how it should or should not be used. AI technologies should be explainable according to the capacity of those to whom they are explained.

Fostering responsibility and accountability. Humans require clear, transparent specification of the tasks that systems can perform and the conditions under which they can achieve the desired performance. Although AI technologies perform specific tasks, it is the responsibility of stakeholders to ensure that they can perform those tasks and that AI is used under appropriate conditions and by appropriately trained people. Responsibility can be assured by application of "human warranty", which implies evaluation by patients and clinicians in the development and deployment of AI technologies. Human warranty requires application of regulatory principles upstream and downstream of the algorithm by establishing points of human supervision. If something goes wrong with an AI technology, there should be accountability. Appropriate mechanisms should be available for questioning and for redress for individuals and groups that are adversely affected by decisions based on algorithms.

Ensuring inclusiveness and equity. Inclusiveness requires that AI for health be designed to encourage the widest possible appropriate, equitable use and access, irrespective of age, sex, gender, income, race, ethnicity, sexual orientation, ability or other characteristics protected under human rights codes. AI technology, like any other technology, should be shared as widely as possible. AI technologies should be available for use not only in contexts and for needs in high-income settings but also in the contexts and for the capacity and diversity of LMIC. AI technologies should not encode biases to the disadvantage of identifiable groups, especially groups that are already marginalized. Bias is a threat to inclusiveness and equity, as it can result in a departure, often arbitrary, from equal treatment. AI technologies should minimize inevitable disparities in power that arise between providers and patients, between policy-makers and people and between companies and governments that create

and deploy AI technologies and those that use or rely on them. AI tools and systems should be monitored and evaluated to identify disproportionate effects on specific groups of people. No technology, AI or otherwise, should sustain or worsen existing forms of bias and discrimination.

Promoting AI that is responsive and sustainable. Responsiveness requires that designers, developers and users continuously, systematically and transparently assess AI applications during actual use. They should determine whether AI responds adequately and appropriately and according to communicated, legitimate expectations and requirements. Responsiveness also requires that AI technologies be consistent with wider promotion of the sustainability of health systems, environments and workplaces. AI systems should be designed to minimize their environmental consequences and increase energy efficiency. That is, use of AI should be consistent with global 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 to adapt to the use of AI systems, and potential job losses due to use of automated systems.

Overview of the report

This report is divided into nine sections and an annex. **Section 1** explains the rationale for WHO's engagement in this topic and the intended readership of the report's findings, analyses and recommendations. **Sections 2 and 3** define AI for health through its methods and applications. Section 2 provides a non-technical definition of AI, which includes several forms of machine learning as a subset of AI techniques. It also defines "big data," including sources of data that comprise biomedical or health big data. **Section 3** provides a non-comprehensive classification and examples of AI technologies for health, including applications used in LMIC, such as for medicine, health research, drug development, health systems management and planning, and public health surveillance.

预览已结束, 完整报告链接和二维码如下:

https://www.yunbaogao.cn/report/index/report?reportId=5 23758

