Artificial intelligence-related health research

Ethics review and oversight

World Health Organization

Artificial intelligence-related health research

Ethics review and oversight

World Health Organization

Artificial intelligence-related health research: ethics review and oversight

ISBN 978-92-4-012407-3 (electronic version)
ISBN 978-92-4-012408-0 (print version)

© World Health Organization 2026

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. Artificial intelligence-related health research: ethics review and oversight. Geneva: World Health Organization; 2026. Licence: CC BY-NC-SA 3.0 IGO.

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

Sales, rights and licensing. To purchase WHO publications, see https://www.who.int/publications/book-orders. To submit requests for commercial use and queries on rights and licensing, see https://www.who.int/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.

Contents

Acknowledgements V
Abbreviations VII
Executive summary VIII
Introduction X
1. Background 1
1.1 What is distinct about ethics oversight of AI-related health research? 1
1.2 What are current international and national standards for ethical and rights-based oversight of AI-based health research? 4
1.2.1 Ethics standards 4
1.2.2 Human rights instruments 4
1.2.3 National laws and policies 5
1.2.4 Guidance for research ethics committees 6
1.3 Why may existing ethical oversight not be fit for purpose? 8
2. Scope, ethical challenges and ethical standards for AI-related health research 9
2.1 Health-related research with data (data science) that uses AI 9
2.1.1 What is health-related data science research that uses AI? 9
2.1.2 Does health-related data science research with AI require ethics review (by RECs)? 9
2.1.3 Which ethical requirements are challenged by health-related data science research with AI? 11
2.2 Research with AI tools and technologies 20
2.2.1 What comprises research with AI tools and technologies? 21
2.2.2 What are the benefits and ethical concerns with AI-based tools to conduct health-related research? 21
2.2.3 What are the benefits and risks with generative AI (large multimodal models)? 22
2.2.4 What are the potential benefits and risks of the use of synthetic data? 24
2.2.5 Testing the performance of AI-based research tools 26
2.3 Health-related research on AI tools and technologies 26
2.3.1 What is health-related research on AI tools and technologies? 26
2.3.2 Does health-related research on AI tools and technologies require ethics oversight or REC review? 27
2.3.3 What novel risks are associated with research on AI tools and technologies? 27
2.3.4 Which ethical requirements are challenged by health-related research on AI tools and technologies? 31
3. Challenges for research ethics committees 36
3.1 Why is the operating model of RECs not fit for purpose? 36
3.1.1 RECs examine research too early 36
3.1.2 RECs are decentralized and may not effectively coordinate and provide oversight of multicountry research 36
3.1.3 Private companies may not submit AI-related health research for ethics review 37
3.2 Why may RECs not yet have appropriate expertise, training, support and capacity for ethics review of AI-related health research? 38
4. How can researchers improve adoption of ethical principles? 40
4.1 Training and certification 40
4.2 Early identification of ethical risks and challenges 41
4.2.1 Societal impact of technology 42
4.2.2 Treatment of data (crowd) workers 42
4.2.3 Transparency 42
4.2.4 Communication and public engagement 43
4.2.5 Existing guidelines for researchers 43

5. What could be the role of third parties for ethical oversight? 45
5.1 Public and private funders of AI research 46
5.2 Health data spaces and data hubs 46
5.3 Data access committees 47
5.4 Community control of data: data sovereignty and data cooperatives 47
5.5 Role of scientific journals and publishers 48
5.6 Scientific medical societies 48
5.7 Role of regulatory agencies and other government oversight bodies 49
5.8 Role of policy-makers (ministries of health and intergovernmental agencies) 50
6. What are the specific concerns with AI-related health research in low- and middle-income countries? 52
6.1 Fairness and inclusivity 52
6.2 Benefit sharing and equitable access 53
6.3 Health data colonialism 53
6.4 Ethics dumping 54
6.5 Power inequities 54
6.6 Capacity-building 54
7. What are the risks and benefits of the use of artificial intelligence for ethics review? 56
8. Conclusion 58
9. Key considerations for diverse actors 60
9.1 Considerations for research ethics committees 60
9.2 Considerations for researchers (and their institutions) 62
9.3 Considerations for third-party stakeholders 63
References 65

Acknowledgements

WHO gratefully acknowledges the many individuals who contributed to this publication.

The development of this document was led by Andreas Reis (Senior Ethics Officer, Research and Ethics Ecosystem Strengthening Unit, Department of Science for Health, World Health Organization [WHO] headquarters), Antonella Lavelanet (Chair of the WHO Headquarters Research Ethics Review Committee, WHO headquarters), Maria M. Guraiib (Secretariat of the WHO Headquarters Research Ethics Review Committee, WHO headquarters), and Sameer Pujari (AI Lead, Department of Data, Digital Health Analytics and Artificial Intelligence, WHO headquarters), under the overall guidance of Tanja Kuchenmüller (Unit Head, Research and Ethics Ecosystem Strengthening Unit, WHO headquarters), Meg Doherty (Director, Department of Science for Health, WHO headquarters), Alain Labrique (Director, Department of Data, Digital Health Analytics and Artificial Intelligence, WHO headquarters), Sylvie Briand (Chief Scientist, WHO headquarters), and Yukiko Nakatani (Assistant Director General, Health Systems, WHO headquarters).

Rohit Malpani (Consultant, Paris, France) was the lead writer. The co-chairs of the Expert Group on Ethics and Governance of AI for Health, Effy Vayena (Swiss Federal Institute of Technology, Zurich, Switzerland) and Partha Majumder (National Science Chair, Government of India, India), provided overall guidance on drafting of the report and leadership of the Expert Group.

WHO Expert Group on Ethics and Governance of AI for Health

Adel Al Shehri, King Abdulaziz City for Science and Technology, Riyadh, Saudi Arabia; Najeeb Al Shorbaji, eHealth Development Association, Amman, Jordan; Maria Paz Canales, Global Partners Digital, Santiago de Chile, Chile; Emmanuel Didier, Centre Maurice Halbwachs, École Normale Supérieure, Paris, France; Arisa Ema, University of Tokyo, Tokyo, Japan; Amel Ghouila, Gates Foundation, Seattle, Washington, United States of America; Jennifer Gibson, WHO Collaborating Centre for Bioethics, University of Toronto, Toronto, Canada; Kenneth Goodman, Institute of Bioethics and Health Policy, University of Miami Miller School of Medicines, Miami, Florida, United States; Sharon Kaur, University of Malaya, Kuala Lumpur, Malaysia; Tze Yun Leong, National University of Singapore, Singapore; Alex John London, Carnegie Mellon University, Pittsburgh, Pennsylvania, United States; Partha Majumder, National Science Chair, Government of India, Kolkata, India; Roli Mathur, Indian Council of Medical Research, Bangalore, India; Timo Minssen, Centre for Advanced Studies in Biomedical Innovation Law, Faculty of Law, University of Copenhagen, Copenhagen, Denmark; Keymanthri Moodley, Stellenbosch University, Cape Town, South Africa; Andrew Morris, Health Data Research UK, London, United Kingdom of Great Britain and Northern Ireland (United Kingdom); Jerome Singh, University of Kwa-Zulu Natal, Durban, South Africa; Jeroen van den Hoven, University of Delft, Delft, Kingdom of the Netherlands; Effy Vayena, Swiss Federal Institute of Technology Zurich, Zurich, Switzerland; Robyn Whittaker, University of Auckland, Auckland, New Zealand; and Yi Zeng, Chinese Academy of Sciences, Beijing, China.

Observers

Sara L.M. Davis, University of Warwick, Coventry, United Kingdom; Agata Ferretti, IBM Research, Zurich, Switzerland; Lee Hibbard, Council of Europe, Strasbourg, France; Jan Piasecki, Jagiellonian University Medical College, Krakow, Poland; Guoyu Wang, Fudan University, Shanghai, China; Ning Wang, University of Zurich, Zurich, Switzerland; Yuzhou Wang, Peking University, Beijing, China; Yu Yang, Fudan University, Shanghai, China; Jie Yin, Fudan University, Shanghai, China; Xiaomei Zhai, Chinese Academy of Medical Sciences, Beijing, China; and Linfan Zhu, Fudan University, Shanghai, China.

External reviewers

Angela Ballantyne, University of Otago, Wellington, New Zealand; Rosie Dobson, University of Auckland, Auckland, New Zealand; Takanori Fujita, Tokyo Foundation, Tokyo, Japan; Masahiro Hashimoto, Keio University, Minato, Japan; Jon Herries, Te Whatu Ora Health New Zealand, Wellington, New Zealand; Calvin Ho, Monash University (School of Law), Melbourne, Australia; Yusuke Inoue, Kyoto University, Kyoto, Japan; Cheng Kai Jin,

Te Whatu Ora Health New Zealand, Auckland, New Zealand; Hiroaki Kakizaki, People Media Inc., Tokyo, Japan; Riki Kyle, Te Whatu Ora Health New Zealand, Waikato, New Zealand; Rodrigo Lins, Instituto de Educação Médica, Rio de Janeiro, Brazil; Eric Meslin, University of Toronto, Toronto, Canada; Eisuke Nakazawa, University of Tokyo, Tokyo, Japan; Marceline Djuidje Ngounoue Epse Ndzie, University of Yaoundé, Yaoundé, Cameroon; Kazushi Nomura, Nomura Clinic, Tokyo, Japan; Takafumi Ochiai, Atsumi and Sakai, Tokyo, Japan; Ryan Radecki, Te Whatu Ora Health New Zealand, Christchurch, New Zealand; Hitomi Sano, Keio University, Minato, Japan; Yuki Shimahara, Medical AI Promotion Institute, Tokyo, Japan; Chaitali Singha, International Development Research Center, Ottawa, Canada; Andrew Sporle, Te Whatu Ora Health New Zealand, Auckland, New Zealand; Lisa Stamp, University of Otago, Wellington, New Zealand; Robert Matthew Strother, University of Otago, Wellington, New Zealand; Rochelle Style, Te Whatu Ora Health New Zealand, Auckland, New Zealand; Hisateru Tachimori, Keio University, Minato, Japan; Yuichiro Yano, Juntendo University, Tokyo, Japan; and Megumu Yokono, Waseda University, Tokyo, Japan.

External contributors

Rosie Dobson, University of Auckland, Auckland, New Zealand; Rodrigo Lins, Instituto de Educação Médica, Rio de Janeiro, Brazil; and Jan Piasecki, Jagiellonian University Medical College, Krakow, Poland.

All external reviewers, experts and contributors declared their interests in line with WHO policies. None of the interests declared were assessed to be significant.

WHO headquarters

Ana Palmero (Department of Science for Health); Tigest Tamrat (Department of Sexual and Reproductive Health and Research); Richelle George; Kanika Kalra; Shada Al-Salamah; Rajeshwari Singh; and Yu Zhao (Department of Data, Digital Health Analytics and Artificial Intelligence).

WHO regional offices

Clayton Hamilton (Data, Artificial Intelligence and Digital Health, Regional Office for Europe); Arshad Altaf (Evidence, Data and Research for Policy and Impact, Regional Office for the Eastern Mediterranean); Mengji Chen (Data Strategy and Innovation, Regional Office for the Western Pacific).

Abbreviations

AI	artificial intelligence
CIOMS	Council for International Organizations of Medical Sciences
CONSORT	Consolidated Standards of Reporting Trials
COVID-19	coronavirus disease 2019
DAC	data access committee
LLM	large language model
LMM	large multimodal model
REC	research ethics committee
SARS-CoV-2	severe acute respiratory syndrome coronavirus 2
SPIRIT	Standard Protocol Items: Recommendations for Interventional Trials
WHO	World Health Organization

Executive summary

For decades, international and national standards have been developed that set out the ethical principles by which research with human participants should be governed. In most countries, a system of research ethics oversight has been put in place to ensure respect for these ethical principles and to protect the dignity, rights and welfare of research participants. Yet presently health-related research is transforming with the use of artificial intelligence (AI), because of the expectation that AI can be applied to generate new knowledge, insights and technologies that improve human health. However, AI-related health research could also undermine established ethical principles and violate human rights protections if novel risks and potential harms are not anticipated and adequately addressed.

Even as the use of AI in health-related research is increasingly common, standards for research ethics may not yet fully account for the proliferation of AI-related health research and the unique risks posed by using AI, or for ethical principles developed to address AI-specific risks, such as the World Health Organization (WHO) guiding principles on the ethics and governance of AI for health.

Only some countries have revised national guidelines to account for the advent of AI-related health research, while institutions may not have updated their standards or may each apply standards that diverge. That these standards either are out of date or may differ from one another means researchers and those parties that oversee research conduct are left without appropriate guidance or may provide guidance that contradicts peers. Institutions that are intended to oversee research conduct may face challenges related to expertise and capacity to provide appropriate oversight of these types of research.

For the purposes of this report, WHO defines AI-related health research through three categories: (a) health-related research with data that use AI; (b) research with AI tools and technologies; and (c) health-related research on AI tools and technologies. AI-related health research has certain characteristics that merit a careful examination of whether the established standards and practices of research ethics oversight are well suited. One characteristic of AI-related health research is a compressed research-to-product life cycle, which means that risks not addressed during research can be manifested in outcomes and products that are widely disseminated or relied upon. Furthermore, there is a low barrier to entry to conduct AI-related health research, including by research teams that may lack credentials or training in research ethics. The development and adoption of appropriate principles, and rules and regulations to update and implement those principles, are therefore necessary. There is also a need to consider how best to support and strengthen researchers and institutions that provide oversight.

This report considers how best to address this growing gap in ethical oversight. The document examines how standards for research ethics may need to evolve to address the risks and benefits of AI-related health research and to integrate emerging ethical principles underpinning the use of AI. The report then considers how ethics review and oversight of different types of AI-related health research by research ethics committees (RECs) and third-party oversight mechanisms could be strengthened within their respective mandates. There are several reasons that RECs may not yet be able to fully carry out their duties in this new field. For example, the current operating model of RECs could be strengthened for how, when and where AI-related health research is carried out. Furthermore, RECs could be supported with additional expertise, training, capacity and support to oversee AI-related health research.

However, responsibility for ethics oversight does not just fall upon RECs. Researchers themselves are often not adjusting their design and conduct of research to anticipate and mitigate risks. Yet it is the responsibility of researchers to apply relevant principles from the initial design of the research through the completion of research and when long-term benefits and consequences materialize. Researchers could strengthen their commitment to ethical standards for responsible research through improved training and certification. They could also improve the conduct of research through several measures, such as (a) early identification of risks and challenges; (b) understanding and communicating the societal impact of a technology; (c) appropriate treatment of data (crowd) workers; (d) a commitment to transparency of known risks with a research project; and (e) strengthening communication and public engagement.

The responsibility to address the many risks of AI-related health research will also be shared with other parties that oversee the end-to-end development of technologies. Even if studies are exempt from ethics review by RECs, the studies may require oversight through other well established mechanisms or new specialized committees not yet created. Third parties can both reinforce the efforts of RECs and address systemic issues that RECs cannot address on a case-by-case basis. The report analyses the role of the following third parties to carry out ethical oversight of AI-related health research: (a) public and private funders of AI research; (b) health data spaces and data hubs; (c) data access committees; (d) data cooperatives; (e) scientific journals and publishers; (f) scientific medical societies; and (g) regulatory agencies and other government oversight bodies. The document also considers how policy-makers can address the broader consequences and impacts of AI-related health research, and the implications for health care systems.

The challenges and opportunities associated with AI-related health research are also analysed in this report as they apply to research in low- and middle-income countries. Even though AI is often framed as a technology that can address intractable health care challenges in these settings, not enough AI research and technology development is currently led by entities in these countries and regions. This imbalance exposes people in low- and middle-income countries to several risks and challenges when research is conducted in their context or with their data, yet without their perspectives and expertise. The report describes these risks and suggests how researchers, RECs, and third-party oversight mechanisms can anticipate and respond to the risks, and take positive steps to build capacity and confer benefits that strengthen the role and participation of researchers in low- and middle-income countries in AI-related health research.

WHO acknowledges that this is a fast-changing area – the technologies and uses of AI for health are evolving rapidly, and the benefits and risks associated with these technologies will change in unexpected ways. AI-related health research may be conducted exclusively in the private sector, which in some contexts may not be subject to oversight, or review that is independent. WHO also recognizes that new demands, whether on RECs, other entities that provide oversight of health research, and the researchers themselves, could be difficult to implement as the potential uses of AI continue to evolve. Furthermore, there are and will be national differences as to how researchers, RECs and oversight mechanisms worldwide could individually and collectively contribute to appropriate conduct and oversight of this research.

Therefore, considerations included in this report (see section 9) are merely a starting point for strengthening ethics oversight for the era of AI-related health research, and it is hoped that they can assist in the development of further standards in the future. Over time the use of AI may become largely unexceptional within health research, thereby obviating the need for a specific report or guidance on the use of AI as much as the wholesale integration of AI-related considerations into general research ethics guidance.

Introduction

Artificial intelligence (AI) refers to the ability of algorithms encoded in technology to learn from data so that they can perform automated tasks without explicit programming of every step by a human.

The World Health Organization (WHO) recognizes that AI holds great promise for human beings to enjoy the highest attainable standard of health and for governments and WHO to achieve the strategic priority of universal health coverage. However, AI also presents many potential risks and harms that must be anticipated and addressed if individuals, communities and societies are to fully benefit. The development and adoption of appropriate standards, and rules and regulations to implement these standards, have become more urgent with the speed of technological advances and the rapid adoption and uptake of AI for diverse and occasionally unforeseeable uses.

The potential of AI has been translated into increased use for research in general and medicine and public health specifically. The growth in AI-related health research has also led most leading English language medical and scientific journals to establish AI-focused, stand-alone journals – including the *New England Journal of Medicine*, the *Journal of the American Medical Association*, the *British Medical Journal*, and *Nature*.

For this report, AI-related health research is defined as:

- research utilizing AI-based technologies to collect or analyse large data sets, including from social media, for any social science, biomedical, clinical, behavioural or epidemiological activity with the intent of generating new knowledge for a health-related purpose;
- the use of AI-based research tools to carry out different parts of a research protocol, for example to formulate hypotheses and design research studies, to identify research participants, to remotely monitor study participants and research activities *(1)*, or to produce quantitative and qualitative scientific and academic output, using for example large multimodal models;
- research upon the utility, safety, efficacy and implementation of AI-based health applications and tools that are directly used on human beings (for example, for detection, diagnosis and treatment) or that are intended to influence human behaviour.

This guidance does not cover the use of AI across drug discovery and development, for which WHO has published a separate report *(2)*.

The growing use of AI in health research could generate new knowledge, insights and technologies that improve health. Yet AI-related health research magnifies some of the ethical issues that have been posed by health research in general and introduces new concerns. These can challenge the existing standards and processes governing health-related research. For example, even if AI-related health research only utilizes anonymized, publicly available data, there are AI-related ethical concerns and risks, such as hallucinations and bias. There is a special concern about research practices in low- and middle-income countries, such as ethics dumping, that require measures to prevent exploitation of people in low-income countries and settings, while not undermining the need for AI-related health research to be led by researchers from those same countries. Existing consent processes for data collection may be inappropriate for how data are subsequently used for AI-related health research.

These factors increasingly create challenges in ethical oversight that should be addressed. Multiple parties share responsibility for ensuring ethical conduct of research, including researchers, research funders, research ethics committees (RECs), regulators, host institutions, scientific journals, and conference organizers. RECs historically have been playing a central role in ensuring the ethical conduct of research that respects the rights, dignity, physical and moral integrity, and interests of participants. However, specific categories of AI-related health research may be exempted from REC review or subject to expedited review, as allowed by national laws and regulations and consistent with international guidelines. For example, a study may be exempt when publicly

available data are analysed or the data are generated by observation of public behaviour, and data that could identify individual persons or groups are anonymized or coded. But many of these exempted studies nonetheless raise ethical concerns, thereby potentially creating a gap in ethical oversight.

While the mandate of RECs could be expanded to address this gap, the volume of research that can be conducted at high speed with AI-based tools and techniques could overwhelm RECs if their mandate is expanded. Furthermore, some RECs may not be well equipped to address the problems posed by AI-related health research or may be asked to review research proposals before risks materialize or are understood by researchers and research participants. Instead, other institutions and actors, such as data access committees, AI funders or health data hubs, could assume a complementary oversight function.

Member States, RECs, and AI-focused researchers have requested WHO to provide a report that examines how ethical oversight of AI-related health research can be adequately ensured in the future. To develop this report, WHO established a specialized working group comprising members of the WHO Expert Group on Ethics and Governance of AI for Health, external experts, and members of the WHO Ethics Review Committee and Secretariat. The primary audience is RECs and researchers, but it is also intended for ministries of health, funders of AI-related health research, oversight mechanisms such as data access committees and regulatory agencies, scientific journals and publishers, and the private sector.

Section 1 provides the context, rationale and background for engagement on this subject matter. Section 2 discusses challenges with ethics oversight for the three categories of AI-related health research: (a) regulation that defines the boundaries and scope of ethics oversight may not be well defined for AI-related health research; (b) existing standards for ethics oversight do not fully address all concerns with AI-related health research; and (c) WHO consensus principles for AI are not yet applied to ethics review of AI-related health research.

Sections 3 through 8 examine how different stakeholders can contribute to appropriate research design, conduct, and oversight of AI-related research. In section 3, we examine why the operating model and capacity of RECs may not be currently adequate for addressing the ethical challenges with AI-related health research. In section 4, we look at how researchers could address ethical issues and challenges when they conduct AI-related health research. In section 5, we examine how different third-party stakeholders, including research funders, data access committees, and scientific journals and publications, could play a role in providing ethical oversight of AI-related health research. Section 6 examines the specific concerns of AI-related health research in low- and middle-income countries and the potential obligations of different stakeholders to strengthen research capacity in those countries while abiding by ethical principles and norms. Section 7 examines the possible benefits and risks of RECs using AI to assist with assessing whether research satisfies ethical requirements. Finally, section 8, the conclusion, considers how different parties, whether funders, RECs, data access committees, or scientific journals, can collectively oversee AI-related health research from design through the publication of results.

Section 9 provides considerations for RECs, third-party oversight mechanisms, research institutions, and governments that suggest how stakeholders could strengthen standards and oversight of AI-related health research. It also provides considerations as to how different stakeholders can improve the design and conduct of AI-related health research.

The considerations included in the report are non-exhaustive and are a starting point. WHO recognizes that this is a rapidly changing field and that there are national differences as to how researchers, RECs and oversight mechanisms worldwide should each individually and collectively contribute to appropriate conduct and oversight of this research.

1. Background

1.1 What is distinct about ethics oversight of AI-related health research?

Major international guidance documents on research ethics, such as the Declaration of Helsinki on Ethical Principles for Medical Research Involving Human Subjects, the Council for International Organizations of Medical Sciences (CIOMS) International Ethical Guidelines for Health-related Research Involving Humans *(3)*, and the Council of Europe Protocol to the Convention on Human Rights and Biomedicine concerning Biomedical Research, provide ethical standards that are in wide use for research involving human beings.. One key requirement of these international guidelines is that research involving human beings should be reviewed by an independent REC (also known as an institutional review board), to ensure that ethical principles are upheld and accounted for within a research proposal and protocol.

WHO works with Member States and partners to promote ethical standards and appropriate systems of review for any course of research involving human subjects and collaborates with other organizations and partners to update such standards and relevant processes, when necessary. In most countries, these ethical standards have been adopted through laws or regulatory provisions. WHO itself has research ethics review committees to ensure that WHO only supports research of the highest ethical standards. They review all research projects involving human participants supported either financially or technically by WHO.

The Ethics Review Committee at WHO headquarters defines research under its purview as the following:

> ... any social science, biomedical, behavioural, or epidemiological activity that entails systematic collection or analysis of data with the intent to generate new knowledge; in which human beings (i) are exposed to manipulation, intervention, observation, or other interaction with investigators either directly or through alteration of their environment, or (ii) become individually identifiable through investigators' collection, preparation, or use of biological material or medical or other records *(4)*.

There are specific characteristics of AI-related health research, and its outputs, that merit special consideration.

First, the research-to-product life cycle can be shorter for AI-based technologies and research *(5)*. Unlike traditional pharmaceutical discovery and development, which can take up to a decade or longer (with notable exceptions, for example the development of treatments and vaccines during the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) pandemic), the timeline from the release of AI-related research to the introduction of a commercial product can occur in as little as 12 months *(5)*. Without adequate ethics oversight during a compressed research process, research can generate risks that are rapidly manifested in products that could be utilized widely in health systems and by individuals. This could result in harm that jeopardizes the human rights protection of patients (including those unable to consent), erodes trust in science, and undermines the long-term viability of AI technologies.

Second, the deployment of AI-related health research developed for one context or population may not be generalizable to other contexts or populations. Rapid deployment of research findings or outputs could have adverse consequences if applied to a context or population for which the findings are irrelevant, counterproductive or harmful.

Third, certain types of AI-related health research (as with certain types of genomics research) do not require direct interaction with research participants. This provides researchers with tremendous flexibility and power to formulate novel hypotheses and to conduct research but may lead researchers to consider ethics oversight irrelevant. However, even if such research only utilizes anonymized, publicly available data, there are other ethical and human rights issues that could arise that merit some sort of ethical oversight.

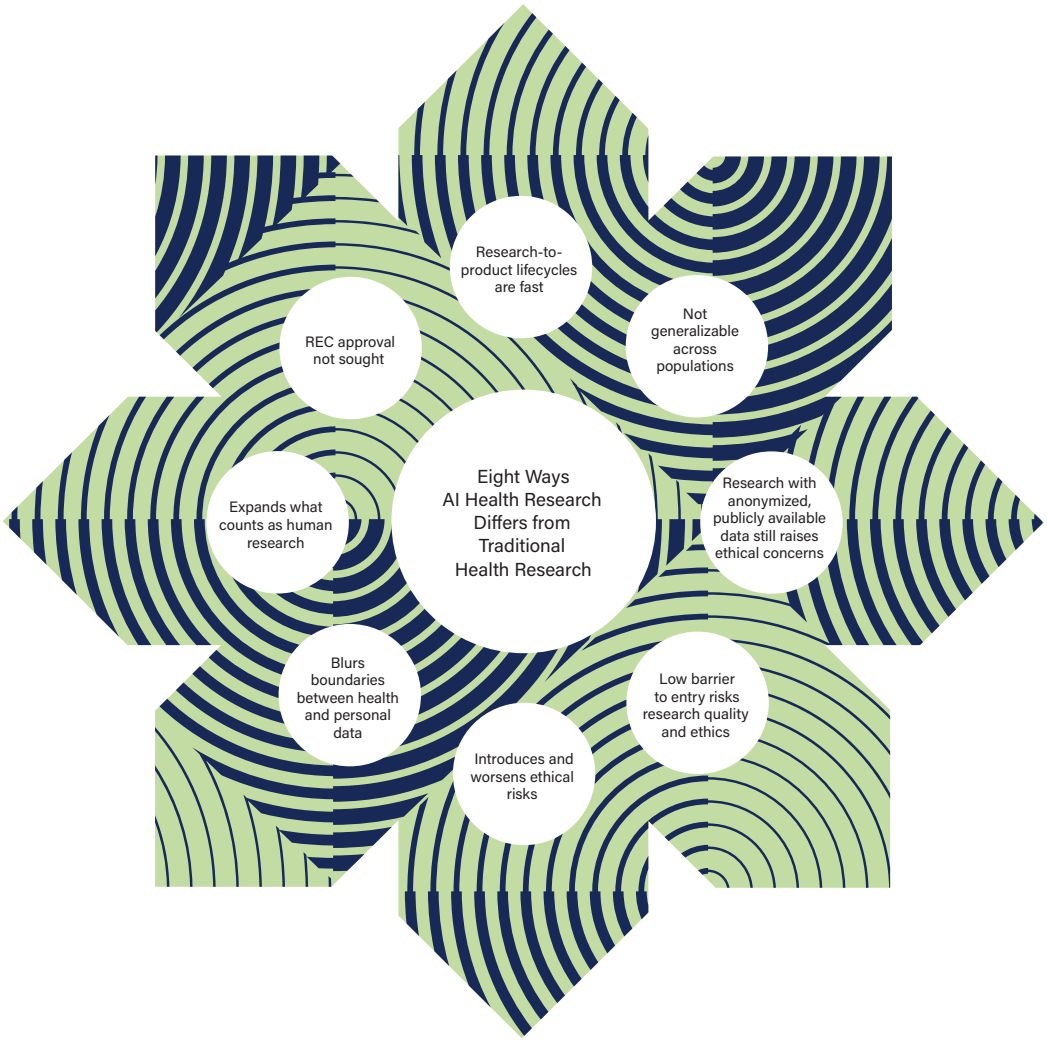
Fourth, there is a low barrier to entry for conducting AI-related health research, as AI and data science tools and techniques are easy to use *(5)*. Researchers who use such tools may not necessarily have the training and credentials, including about research ethics of traditional health researchers *(6)*. This includes companies that are applying AI technologies (and data sets that they may hold or acquire) for health-related purposes. A low barrier to entry can also mean that researchers without appropriate training could conduct AI-related health research that lacks scientific validity or social value.

Fifth, AI-based technologies exacerbate existing ethical concerns and introduce new risks, for example, regarding individual privacy or forms of bias and discrimination. This is due to how AI technologies are designed (and who designs such AI technologies), the data that are used to train AI algorithms (as well as how such data were obtained), and how such technologies are used, and by whom, for health-related purposes. Risks introduced within the design, training and use of AI-based technologies are described by WHO in its guidance on the *Ethics and governance of artificial intelligence for health (7)*. Subsequently, WHO has published new guidance on the risks and benefits of the use of generative AI (large multimodal models) for health *(8)*.

Sixth, AI-related health research is increasingly blurring the divide between "health data" and "personal data". This is because the status of data as either health data or personal data is dynamic, with the possibility that a form of data can depend on how it is used and under what circumstances *(9)*. Furthermore, data that at one point could be characterized as personal can be aggregated and analysed in ways that can be used to infer or predict certain health-related data or characteristics of an individual or population *(10)*. The lack of a clear boundary between different types of data could affect how researchers collect, categorize and use such data, and how RECs and other oversight mechanisms review its collection and use.

Seventh, AI-related health research itself may require an expanded definition of what is "research", especially since many types of studies conducted with AI did not exist before the emergence of new subsets of AI. For example, research could include the development of a tool or product for commercial purposes, studies that involve reinforcement learning (to improve the performance of an algorithm with the use of crowd workers), or the customization of a large multimodal model (LMM) for local use. AI researchers may not realize that their activities are research with human beings and therefore subject to ethical oversight.

Figure 1. Why AI-related health research is unique/different than non-AI (traditional) health research



AI-related health research conducted over the last decade has illustrated the necessity of oversight via RECs. Previous AI-related health research projects that raised ethical concerns and for which ethics review or other forms of oversight were not conducted have led to harm and a potential loss of societal trust. For example, a collaboration between DeepMind and the United Kingdom's National Health Service, which included the transfer of identifiable patient records from the National Health Service without explicit consent (both parties claimed there was implied consent), led to concerns over the use of patient data without consent (and additional uses beyond those in the original research project), the scientific viability of both the research project and its output, the nature of the commercial arrangements between DeepMind and the National Health Service, and the lack of any regulatory oversight of the research project itself *(11)*. As to the latter concern, this included no ethics approval process by an REC *(12)*. That the DeepMind and National Health Service collaboration did not undergo REC review may have demonstrated a willingness of both parties to avoid ethics review for research using identifiable health data (without informed consent) when there were not yet defined rules and guidance for oversight of AI-related health research. Researchers need guidance as to what types of research require ethical oversight, while RECs and other oversight mechanisms require standards that can guide review of such research.

1.2 What are current international and national standards for ethical and rights-based oversight of AI-based health research?

There are diverse laws, policies, standards and benchmarks that apply to the conduct and oversight of health-related research. These include consensus declarations on research ethics, human rights obligations, AI-focused laws and policies, data protection laws and privacy rules, and regulatory standards. This section reviews several of these benchmarks, standards and laws.

1.2.1 Ethics standards

International standards may not have anticipated the advent of AI-related health research and presently have not been updated to account for its growing use. The Declaration of Helsinki, which was updated in October 2024, does not yet fully consider the growth of the digital environment and the conduct of research through digital tools and interfaces. The Declaration of Taipei on Ethical Considerations regarding Health Databases and Biobanks, which was last revised in October 2016, covers the "collection, storage and use of identifiable data and biological materials beyond the individual care of patients" *(13)*, and especially the diverse uses of health databases (and biobanks) for research and other purposes. Relevant standards included in the Declaration of Taipei are enumerated in Section II and in Section 5 with respect to the establishment and management of data hubs and data spaces. The Declaration of Taipei, however, does not account for how health databases should manage research with anonymized data, except to require a notification to individuals who provide identifiable data that "in case the data and material are made non-identifiable the individual may not be able to know what is done with their data/material and that they will not have the option of withdrawing their consent" *(13)*. The World Medical Association has just started a revision process of the Declaration of Taipei.

CIOMS has issued several guidelines that have a bearing upon the use of AI in health research, and especially its consideration of data ethics. While these guidelines do not yet specifically address emerging issues related to the use of AI for health research, many of the standards are relevant. This report highlights both applicable relevant standards in the CIOMS International Ethical Guidelines for Health-related Research Involving Humans and additional considerations that complement what has already been issued by CIOMS.

Guideline 12 of CIOMS examines the obligations upon researchers with respect to the collection, storage and use of data in health-related research, including how researchers should manage informed consent for the collection and use of such data, and the appropriate governance of such data to assure confidentiality, to guide reuse of such data without specific or broad informed consent, and to respect the rights of participants who have provided such data *(3)*.

Guideline 22 of CIOMS requires measures to reduce privacy-related risks for research within the online environment or with digital tools – including to assess the privacy risks, mitigate the risks, and describe the remaining risks in the research protocol. It also delineates how researchers should manage informed consent for such data, as well as the use of publicly available information from a website that does not require direct contact with individuals *(3)*. Guideline 23 states that certain studies may be exempt from review by RECs, for example, when publicly available data are analysed or the data for the study are generated by observation of public behaviour, insofar that data that could identify individual persons or groups are anonymized or coded *(3)*.

1.2.2 Human rights instruments

International and regional human rights conventions, including the Universal Declaration of Human Rights, the International Covenant on Economic, Social and Cultural Rights (including General Comment No. 14 on the right to the highest attainable standard of health), the International Covenant on Civil and Political Rights, the United Nations Declaration on the Rights of Indigenous Peoples, and regional human rights conventions, apply to the conduct of AI-related health research. This includes the right to non-discrimination, the right to privacy, the right of Indigenous peoples to control their own data and intellectual property, the right to the highest attainable standard of health, and the right to enjoy the benefits of scientific progress. The Council of Europe Protocol to the Convention on Human Rights and Biomedicine concerning Biomedical Research was adopted in 2005

and provides legally binding provisions for parties to safeguard human dignity and the fundamental rights and freedoms of the individual regarding biomedical research. Most countries have ratified relevant international covenants or regional conventions, and some have incorporated these standards into constitutions or national laws.

1.2.3 National laws and policies

Most countries have enacted regulations and laws grounded in international ethics standards. However, most governments so far have neither enacted nor revised national guidelines to reflect the emergence of AI-related health research, and many existing laws, because of how such laws are written, do not yet require AI-related health research to conform to research ethics principles or mandate RECs or other oversight mechanisms to review such research. The European Union's Artificial Intelligence Act, enacted in 2024, is one of the world's first laws that seeks to regulate the development and deployment of AI across sectors within the European Union. The Artificial Intelligence Act, however, does not specifically address AI-related health research, though its standards, especially for AI technologies classified as "high risk", could be applied to determine which types of AI-related health research merit REC review or other oversight, or encourage researchers to conduct AI-related health research in a manner that upholds research ethics principles (Box 1).

Box 1. Risk-based approach to AI-related health research

Determining which types of AI-related health research merit review by an REC is complex and requires extended careful deliberation. In this report (see below), WHO does not prescribe how RECs can determine which types of AI-related health research they should review, though the report provides different suggestions.

One approach to determining which forms of AI-based health research should be reviewed would depend on the level of risk associated with the research proposal. This would mirror the risk-based approach adopted by the European Union's Artificial Intelligence Act, which was enacted in 2024 *(14)*. With a risk-based approach, AI research projects could be classified as (a) unacceptable risk, (b) high risk, (c) limited risk, or (d) no risk. Under this type of classification, "high risk" and "unacceptable risk" research would be scrutinized by RECs. However, even though "limited risk" or "no risk" research may not pose the types of ethical challenges posed by higher-risk research, such research could still introduce poor methodological approaches or produce erroneous results that are propagated, thereby undermining a specific field of enquiry, undermining science, and wasting scarce resources and effort. Thus, even if classified as limited or no risk, research may still require ethical oversight, perhaps via an expedited pathway, to ensure it is of high quality and has social value. Even if RECs do not adopt a risk-based approach, they could still require researchers to submit an independently produced risk assessment that provides a risk classification and rationale for the research.

Data protection laws – such as the European Union's General Data Protection Regulation – were written to ensure the adequate protection of an individual's privacy rights while also promoting scientific research and innovation. There are still several unanswered questions as to how data protection laws relate to AI-related health research, and how regulatory oversight through data protection agencies and RECs should complement each other.

Data protection laws generally do not apply to anonymized data. Under the General Data Protection Regulation, data that have been fully anonymized for scientific research are excluded from its regulatory requirements, whereas pseudo-anonymized data, or data for which "identifiers" have been removed so that such data can only be re-identified with these identifiers, are regulated by the General Data Protection Regulation. The General Data Protection Regulation also provides exceptions (derogations) to prohibitions related to the processing of personal data, including "for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes" *(15)*. The regulation also notes that such research should be "proportionate to the aim pursued, respect the essence of the right to data protection and provide for suitable and specific measures to safeguard the fundamental rights and the interests of the data subject". *(15)*.

Such exceptions are also present in other data protection laws. For example, under South Africa's data protection law (Protection of Personal Information Act), the law does not apply to (a) de-identified data that cannot be re-identified; (b) the processing of data by a public body for national security reasons; and (c) personal information that is related to a purely personal or household activity *(16)*.

Data protection laws, including the General Data Protection Regulation, do not yet specifically enumerate what "suitable and specific measures" are required for the processing of personal data for scientific research *(17)*. The General Data Protection Regulation also does not draw any distinction between academic or not-for-profit research and commercial research, even though commercial research may not necessarily be carried out in the public interest. Furthermore, the motivation for certain types of commercial research may be specifically what data protection laws should protect against *(18)*. Even though data protection laws do not specifically refer to the role of RECs and other oversight mechanisms, RECs are one means by which researchers may be able to provide for "suitable and specific measures" *(17)*. The requirements of data protection laws, and their interplay with the role of RECs, is discussed below (see section 2.1).

Indigenous communities have developed frameworks to protect, control and support research using their data. Indigenous data governance principles, such as the CARE Principles for Indigenous Data Governance, describe how Indigenous peoples can protect the rights and interests in their own data while supporting research that could confer group benefits from research using their data. The CARE Principles comprise four elements: (a) collective benefit, (b) authority to control, (c) responsibility (of those working with Indigenous data), and (d) ethics ("Indigenous peoples' rights and well-being should be the primary concern at all stages of the data life cycle and across the data ecosystem") *(19)*.

Finally, the Council of Europe Recommendation CM/Rec(2019)2 *(20)* of the Committee of Ministers to Member States on the protection of health-related data makes specific reference to scientific research, notably to the following:

> 15.9. Where scientific research purposes allow, data should be anonymized; where research purposes do not allow this, pseudonymization of the data – with intervention of a trusted third party at the separation stage of the identification – is among the measures that should be implemented to safeguard the rights and fundamental freedoms of the data subject. These measures must be carried out where the purposes of scientific research can be fulfilled by further processing which does not permit or no longer permits the identification of data subjects.
>
> 15.10. Where a data subject withdraws from a scientific research project, their health-related data processed in the context of that research should be destroyed or anonymized in a manner which does not compromise the scientific validity of the research, and the data subject should be informed accordingly.

1.2.4 Guidance for research ethics committees

National research agencies and some RECs are trying to adapt to the emergence of AI-related health research. The Indian Council of Medical Research has released *Ethical guidelines for application of artificial intelligence in biomedical research and health care (21)*. The framework includes principles to guide the use of AI in health care, ethical review procedures for medical AI research, and guidance on how informed consent should be managed (Box 2).

Box 2. Indian Council of Medical Research *Ethical guidelines for application of artificial intelligence in biomedical research and healthcare*, 2023

The Indian Council of Medical Research under the Department of Health, Ministry of Health and Family Welfare, India, is the apex body in India to promote biomedical and health research, and is involved in setting ethical standards for research involving humans. In the absence of a regulatory framework, the Indian Council of Medical Research decided to create an ethics framework that would assist in the development, deployment and adoption of AI-based solutions in biomedical research and health care delivery. The resulting *Ethical guidelines for application of artificial intelligence in biomedical research and healthcare* were drafted to respond to the needs of innovators, developers, professionals, ethics committees, funding agencies and users. They are intended to encourage innovation while guiding the development and deployment of safe AI-driven approaches to improve health care research and delivery in India.

The national AI guidelines were prepared and published in 2023. They bring out the ethical principles for responsible AI through discussion of these core principles and their applications. The principles included and discussed in detail in the guidance are autonomy, safety and risk minimization, trustworthiness, data privacy, accountability and liability, data quality, accessibility, equity and inclusiveness, collaboration, non-discrimination, fairness principles and validity. Further, the guidelines include guiding principles for stakeholders involved in development, validation and deployment, and also set out the procedure for undertaking an ethics review of research involving AI technologies. This is beneficial to ethics committees, which otherwise are unclear about their role and responsibility with respect to AI-related research studies. The guidelines also discuss special ethics issues, including the selection of training and testing populations for AI-related health research, the implications of accountability, the requirements of quality checks, and requirements for ethical data sharing. In addition, the guidelines document has a comprehensive section on the informed consent process, including the essential information for prospective research participants, the responsibility of researchers, refusal of consent or the right to be forgotten, and the waiver of consent requirements. Finally, the guidance provides a broader framework for the governance of AI technology use for health care and research and an ethics checklist for researchers, developers and ethics committees reviewing projects involving AI-related health research.

Only some RECs have integrated well established laws and standards (including data protection laws) and emerging AI ethics principles into their substantive review procedures, and may be doing so on a piecemeal basis and without the benefit of an overarching framework. Thus, these principles have not been consistently and fully integrated into the work of RECs and are not harmonized across RECs. Therefore, there is no consistency and coherence amongst RECs with respect to AI-related health research.

A lack of alignment across different oversight mechanisms and standard-setting processes, as well as a lack of guidance until now (via the Declaration of Helsinki or CIOMS International Ethical Guidelines for Health-related Research Involving Humans), means that researchers may struggle to ensure that AI-related research protocols satisfy emerging standards and, therefore, adequately protect participants. This challenge is heightened for research consortia and multicentric research projects for which each research site may need to abide by a different set of standards enforced by a local REC or other oversight mechanism *(5)*.

1.3 Why may existing ethical oversight not be fit for purpose?

Ethical oversight plays a critical role in shaping the direction of AI-related health research. Current approaches to health-related ethical oversight in research are not fully applicable to AI-related health research. Specifically, current ethical oversight processes may be inadequate for six reasons:

- Regulation defining the boundaries and scope of research that requires ethics oversight does so in a way that is too vague and inadvertently excludes or ignores several types of AI-related health research. It may also not specify which oversight mechanism is appropriate for diverse types of research.
- Existing guidance for ethics oversight does not fully address all concerns that arise with AI-related health research, and WHO consensus principles on the ethics and governance of AI for health (for example) are not yet applied to ethics review of AI-related health research.
- RECs may not be well suited to how, when and where AI-related health research is carried out.
- RECs have not yet been provided with sufficient expertise, training, capacity and support to adequately oversee AI-related health research.
- Researchers are not yet applying ethical standards to design, implementation and oversight of their studies.
- Third parties that play a role in the oversight of research quality, including research funders, data access committees, scientific journals and publishers, and regulatory agencies, are not adequately accounting for and managing risks and ethical challenges with AI-related health research.

The subsequent sections of this report describe, analyse and suggest ways forward for each of these challenges.

2. Scope, ethical challenges and ethical standards for AI-related health research

2.1 Health-related research with data (data science) that uses AI

This section discusses the ethical challenges and oversight of health-related data science research that uses AI. Subsection 2.1.1 provides a description of this type of research. There is an ongoing debate as to whether health-related data science research with AI should be subject to research ethics review, which is discussed in subsection 2.1.2. In subsection 2.1.3, we examine the different ethical challenges that arise with data science research with AI, irrespective of whether the research qualifies for ethics review by RECs.

2.1.1 What is health-related data science research that uses AI?

Data science extracts meaningful patterns and insights from data, and with the advent of AI, relies on different types of AI to conduct more sophisticated research and analysis. Significant progress has been achieved using smaller AI models, such as classical machine learning models, to answer research questions *(22)*.

The use of AI for health-related data science research opens numerous opportunities for novel research across multiple domains, including applying AI to collect or analyse large data sets for any social science, biomedical, behavioural or epidemiological activity to generate new knowledge. These research studies are built upon developing or using diverse types of data sets, such as public data sets that have been made widely available for use, secondary analysis of existing data sets, social media data, and cell phone and smart phone data (as a source of mobility data). Such data science research can have significant benefits with respect to public health and medical need. Therefore, ethics review can be of paramount importance to protect individual and group rights while not impeding scientifically useful research.

For example, data science research that uses AI could train and evaluate AI models on large data sets of mammogram images to recognize patterns indicative of breast cancer. Evaluation could compare the accuracy of an AI model with those of human physicians. Such research may not require any interaction of the AI model with human participants since the AI model could be trained on previously collected data and then evaluated on a separate data set that had already been collected (and for which a physician provided a diagnosis). That such research may not require the prospective participation of human participants or identifiable data raises questions, for example, as to whether such research should be subject to ethics review or a different type of ethical oversight, how best to safeguard privacy and the right to informed consent, and the treatment of crowd workers required to collect and label data.

2.1.2 Does health-related data science research with AI require ethics review (by RECs)?

RECs traditionally only review research protocols that involve human participants, or research that (a) features a direct interaction with human beings; (b) alters a research participant's environment; or (c) uses identifiable data, even if it does not involve direct interaction with an individual (for example, researchers extract data from medical records). Health-related data science research that uses AI may not involve human participants and may therefore fall outside their remit despite raising ethical concerns. Determining whether a given project on health-related data science research with AI requires ethics review or is exempted necessitates consideration of different factors.

First, health-related data science research using AI is fundamentally different from earlier forms of data science research because it does not involve a direct intervention or interaction between the researcher and a human being, and instead involves certain "novel dynamics" between researchers and participants that are best described as "data intensive", or the analysis of large data sets by researchers to identify valuable insights *(23, 24)*. Researchers may also acquire, use or reuse data in ways that were not previously anticipated when the

data were initially collected. Because these novel forms of data science research are not viewed as a direct intervention or an alteration of a research subject's environment, these protocols may be exempt from research ethics review in some jurisdictions (as is the case in the United States, see below), whereas other jurisdictions require research ethics review *(5, 25)*. For example, the Indian Council of Medical Research requires ethics review for the use of health-related data sets, the secondary use of medical records or health data, or the collection of such data (including digital data) that may be used for health research *(26)*.

Second, such research may be exempt from review because the data are characterized as publicly available, the data set has been anonymized or de-identified, or the research is based on secondary analysis of existing data sets. De-identification prevents connection of personal identifiers to information. Anonymization of personal data is a subcategory of de-identification whereby both direct and indirect personal identifiers are removed, and technical safeguards are used to try to eliminate the risk of re-identification, whereas de-identified data can be re-identified by use of a key *(7)*.

In the United States, research studies using de-identified health records are deemed, under the Federal Policy for the Protection of Human Subjects ("Common Rule"), 1991, to be excluded from regulations governing research with human participants *(27)*. While regulations such as the Common Rule historically did not specify the standard that should be applied to determine whether data are de-identified, the Government of the United States has issued a "de-identification standard" *(28)*. Standards are useful since, as technology evolves, there may be novel ways and methods to re-identify data. Presently, the Common Rule limits review to the collection and processing of "identifiable private information" *(24)*. Similarly, the Human Research Act in Switzerland and the European Union's research ethics legislation do not require REC oversight for research that utilizes anonymized data or the secondary use of data for which broad consent and prior REC approval was obtained *(24)*.

While some jurisdictions have defined what types of research (having been adequately de-identified) are exempt, other jurisdictions have not clearly defined what threshold must be met to qualify as de-identified or anonymized data. RECs may not be equipped to assess whether a data set in question is at risk of de-anonymization. Some forms of data that could be used for health-based research, including voice samples and genomic data, cannot be de-identified. Research also illustrates that de-identification of data may not always be possible or successful. This is because increasingly, "data triangulation" techniques can be used to reconstruct a de-identified, incomplete data set by a third party for re-identification of an individual. It may be impossible to completely de-identify some types of data, such as genome sequences, as relationships with other people whose identity and partial sequence are known can be inferred. Such relationships may allow direct identification of small groups and narrowing down identification to families *(7)*.

Third, AI-related health research that uses certain types of data may not be within the scope of REC review. Data used for AI-related health research are not constrained by information gathered through traditional sources of clinical, medical or scientific sources, and have migrated to diverse online spaces, such as social media sites. These new forms of data are generating novel research studies. However, novel sources of data may be outside the scope of what types of data are covered by RECs because they are publicly available; because they constitute a pre-existing data set that had already been collected by another entity; or because the data may be anonymized and therefore exempt from REC review *(29)*. In the United Kingdom, research projects that do not require ethics approval include those using social media data, geolocation data or anonymized secondary health data with an agreement *(30)*.

Yet excluding AI-based studies that use, for example, social media data, can mean that controversial and ethically questionable research and practices that should require oversight escapes scrutiny. A research study (that did not involve the use of AI but harnessed social media data) on "emotional contagion" co-sponsored by Facebook (now Meta) and Cornell University sought to manipulate a user's news feed to elicit certain emotional reactions (for example, to publish fewer positive posts to see if it would lead to greater expressions of sadness), yet did not undergo an independent REC review before the study was conducted *(31)*. Facebook, which manipulated the posts to collect data, stated it conducted an "internal review", while Cornell University did not conduct an REC review because the experiment had already been run, and Cornell's REC therefore concluded that the researcher "was not directly engaged in human research and that no review by the Cornell Human Research Protection Program was required" *(31)*. Furthermore, Cornell's REC did not review the research for how Facebook collected data because, according to the University, the research involved a pre-existing data set, and therefore researchers at Cornell University were "just analysing data already collected, often by someone else" *(31)*.

The emotional contagion research led to significant public concerns over the company's wanton manipulation of user feeds to elicit emotional reactions, while Facebook may not have even obtained users' informed consent under the terms of its own data use policy at the time of the research, nor did the company notify those users whose feeds were analysed after the research concluded *(31)*. It may have also raised concerns as to whether the study had any impacts on the psychological health and well-being of those who unknowingly participated in the study *(32)*.[1]

Fourth, health-related data science research that uses Indigenous health data may trigger ethics review to ensure that Indigenous data governance principles are applied to the research in question. This may be required irrespective of the purpose for which the data were collected and analysed and may require an REC to conduct ethical review for research that exceeds its traditional purview.

Fifth, AI-related data science health research can introduce novel research methods or roles within research teams that do not yet trigger REC review but could require ethical oversight. For example, AI (and data science) research customarily relies on the cleaning (curation) of large quantities of data, which is often outsourced to third-party data workers (crowd workers) through platforms such as Amazon Mechanical Turk (MTurk) *(5)*. There is ambiguity as to whether crowd workers constitute human participants for ethical and regulatory purposes *(33)*, and also therefore whether such studies, even if they do not otherwise require REC oversight, should require ethical review because of the use of crowd workers.

While the use of AI for health-related data science research can raise concerns that require ethical oversight, the use of AI, for example with certain machine learning tools, can also be applied to conduct a more sophisticated regression analysis and analysis of existing data sets. Researchers should ensure that such AI tools produce accurate results; however, these uses of AI may not merit ethics review by an REC or ethical oversight through a third-party mechanism.

Presently, health-related data science studies that use AI could be exempt from ethics review for any of the reasons noted above. Yet de-identified data are increasingly at risk of re-identification. There are diverse ethical and legal requirements for use of health data. There are also other ethical requirements, such as social value and scientific validity, that should be satisfied when conducting health-related data science studies that use AI. Furthermore, these studies could impact individuals whose data are used, challenge responsible research conduct, and test society's shared ethical values. Therefore, even if these studies are exempt from ethics review, the studies may require oversight through other mechanisms, such as data access committees or new specialized committees not yet established. Other possible oversight mechanisms for health-related data science research that uses AI are discussed in section 5 of this report. The next subsection explores several ethical issues that could merit oversight, even if not by an REC.

2.1.3 Which ethical requirements are challenged by health-related data science research with AI?

There are ethical issues and challenges with health-related data science research using AI that may require specific measures by researchers and may trigger oversight by an appropriate mechanism, whether an REC, data access committee, or potentially a specialized committee mandated to review such research. Six critical issues are as follows:

1. informed consent (and autonomy and privacy) for newly collected data, previously collected data, and social media data;
2. ethical governance of large data sets that are used for data science research;
3. group-level harms;
4. the risk–benefit ratio of research;
5. fair selection of research participants;
6. the treatment of crowd workers.

The following subsections present further information on those issues.

1 The Council of Europe Protocol to the Convention on Human Rights and Biomedicine concerning Biomedical Research (2005) includes, under its explanatory report, the risk to the psychological health of the person concerned.

1. Informed consent, privacy and autonomous decision-making

The ethical foundation of informed consent is the principle of respect for people. To provide informed consent requires a competent individual who has received the necessary information, who has adequately understood the information, and who, after considering the information, arrives at a decision without having been subjected to coercion, undue influence, inducement or intimidation *(34)*.

Informed consent that is based on a researcher and research participant relationship may not be possible to apply directly to research based on a novel relationship of a researcher and data donor for which the researcher uses data from millions of people and for which there is no direct intervention upon the participant *(29)*. It can be difficult for researchers, RECs or other oversight mechanisms to determine if additional safeguards are required to ensure informed consent.

Safeguards that must be followed by researchers are defined under ethical guidelines (such as the CIOMS International Ethical Guidelines for Health-related Research Involving Humans and the Declarations of Helsinki and Taipei) and obligations under data protection laws. Specific requirements of informed consent under these frameworks are different, as are the instances when researchers can seek exceptions to informed consent *(35)*. Data protection laws may differ with respect to informed consent requirements compared to ethical guidelines (or introduce additional measures, especially for the governance of health data).

For any use or reuse of identifiable health data, there are clear norms and requirements that the data subject must give their meaningful informed consent. Exceptions to consent may be exercised for studies that, pursuant to Guideline 10 of the CIOMS International Ethical Guidelines, are conducted under a public health mandate or by public health authorities, such as disease surveillance. Another situation may be, pursuant to Article 16 of the Declaration of Taipei, "in the event of a clearly identified, serious and immediate threat where anonymous data will not suffice" *(13)*.

Health-related data science studies that use AI are largely carried out with anonymized data. Informed consent for use of anonymized data depends on the purpose of the research study, who is conducting the research, and whether the data are (a) newly collected data; (b) previously collected data that are being used for a new research objective; or (c) social media data. We examine these three types of data below.

1a. Informed consent, privacy and autonomous decision-making for newly collected data

The CIOMS International Ethical Guidelines examine how informed consent should be managed as it relates to the collection, storage and use of newly collected data for health-related research. Under Guideline 12, CIOMS requires that for data collected for research-related purposes, a researcher must obtain "either specific informed consent for a particular use or broad informed consent for unspecified future use from whom the data was originally obtained". Similarly, under Article 10 of 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), everyone has the right of respect to a private life in relation to information collected about their health, which includes the right to know (and *not* to know) about information collected.

Data used for AI-related health research are often collected through an "online environment". Under Guideline 22, CIOMS specifically recommends that: "Researchers should inform persons whose data may be used in the context of research in the online environment of: the purpose and context of intended uses of data and information; the privacy and security measures used to protect their data, and any related privacy risks; and the limitations of the measures used and the privacy risks that may remain despite the safeguards put in place".

Furthermore, CIOMS (under Guideline 22), requires that researchers should refrain from using such data if the data donor refuses. To provide a person with the opportunity to opt out, a researcher, pursuant to Guideline 22, must meet the following conditions: (a) persons need to be aware of its existence; (b) sufficient information needs to be provided; (c) persons need to be told they can withdraw their data; and (d) a genuine possibility to object has to be offered.

There are concerns that seeking either specific informed consent or broad informed consent for newly collected data may be difficult when researchers conduct large-scale studies at multiple research sites. WHO guidance on the ethics and governance of AI for health also notes that these forms of consent may not only be unfeasible but also not meaningful for different reasons:

> Patients may be unable to consent to current and future uses of their health data, such as for population-level data analytics or predictive-risk modelling. Even if a use lends itself to consent, the procedures may fall short, individuals might not be able to consent, such as because they have insufficient access to a health data system, or access to health care is perceived or actually denied if consent is not provided *(7)*.

Yet there may be feasible approaches to both ensure informed consent and conduct research. Researchers should keep in mind that for online research, the functions of informed consent can be fulfilled through alternative mechanisms. For instance, researchers could announce that data collection is taking place and offer participants the opportunity to opt out. Data can be collected with varying degrees of intrusiveness – bots interacting with users may be intrusive but can also provide some control, for example by giving users the option to block them *(33)*. Or there may be novel ways to provide informed consent, for example, through a digital platform, as proposed by Brückner et al. et al, *(36)*, to obtain informed consent for which a user can "actively manage data sharing from health apps and wearables, enabling informed, granular decisions for both primary and secondary use through a centralized consent management system".

For the use of newly collected data for AI-related health research, researchers must also assess and mitigate privacy risks. CIOMS Guideline 22 notes several privacy risks with the collection and use of data, and the risks that could result from combining data, including the risk of re-identification of individuals. Therefore, as stated in Guideline 22: "selection and implementation of appropriate measures to mitigate privacy risks by investigators is essential and entails adopting privacy and security controls suited to the intended uses and privacy risks associated with the data."

1b. Informed consent, privacy and autonomous decision-making for previously collected data

Health-related data science studies that use AI may rely upon the use of data that were separately and previously collected for research, clinical care or other purposes for which informed consent was not obtained for subsequent uses. CIOMS (under Guideline 12) notes that for the use of "stored data collected for past research, clinical or other purposes without having obtained informed consent for their future use ... the research ethics committee may consider to waive the requirement of individual informed consent if: 1) the research would not be feasible or practicable to carry out without the waiver; and 2) the research has important social value; and 3) the research poses no more than minimal risks to participants or to the group to which the participant belongs".

This standard should apply to the use of previously collected health data for AI-related health research. However, since publication of the CIOMS guidance, researchers can use health data to conduct novel AI-related health research in innumerable ways that were not foreseeable when the data were collected. This does not preclude waiving of informed consent in instances where data are fully anonymized. Given the size of data sets used for AI-related health research, it is likely that obtaining informed consent for novel projects would not be feasible. However, a waiver of informed consent may require researchers to undergo scrutiny through an appropriate oversight mechanism.

As with the collection of new data, reuse of data may also introduce novel privacy risks for individuals who previously provided data, especially if data sets are combined with other information to produce more accurate outcomes. Even if data are anonymized, risks remain of re-identification, as well as risk of data breaches, cybersecurity threats *(37)* or disclosure due to a hacking operation that puts data in the public domain *(38)*. Thus, RECs or another oversight mechanism will need to carefully scrutinize what measures have been put in place to maintain the privacy of data, and the risks of unauthorized disclosure. Box 3 explores one example of how re-identification remains a possibility for data that had been de-identified, and its consequences.

Box 3. X-ray re-identification and its consequences

A clinical researcher was given access to a data set of chest X-rays by a public health service in New Zealand for a research project with the approval of a university-based REC. The researcher de-identified the images and curated a data set that was subsequently made available with a data agreement to researchers worldwide using various platforms. This availability would not be approved under today's data governance laws of the public health service. In addition, one of the research projects later conducted on this data set and other data sets showed that individual chest X-rays could be linked. Therefore, if identifiable data were available on the patients whose X-rays were in the original data set, the de-identified data set images could be re-identified. Due to these issues, a request was made for the data to be removed from the hosting platforms and sharing of the original data set to cease.

The prior collection of data that are then used separately by a different research team also raises questions as to whether oversight should be applied to how the initial data were collected, even if there was not an explicit intent to conduct research with such data. Thus, even though researchers may not be required to obtain informed consent from those people whose data were previously collected, an oversight mechanism may wish to revisit (a) how data were collected and what safeguards were put in place; (b) whether there were any ethical concerns with how those data were initially collected; (c) whether data donors knew their data were being collected; or (d) whether the initial data gathered, even if not immediately used for research, were being collected by a company or research institution with the intent that a third party could separately conduct research. In any of these scenarios, an oversight mechanism should not hesitate to challenge whether the data gathering aligns with ethical norms. Even if data collection practices (and subsequent use of such data) might be legal, they may not satisfy moral standards. To meet moral requirements, researchers could inform data donors, regularly but not too often, of the studies in which their data are being used and provide them with an opportunity to opt out of specific research or all future research studies. Irrespective of how data were originally collected, third parties should only get access to coded or anonymized data.

1c. Informed consent, privacy and autonomous decision-making for social media data

Meta's "emotional contagion" is one example of how researchers can use social media data to conduct data science health-related research in ways that can raise significant concerns for the privacy interests of those whose data are used, and the social value of such studies.

Social media data (and other types of data that may not be initially viewed as having any relationship to health) may not be initially characterized as health data yet can be subsequently used as health data to conduct AI-related health research. In many cases, users who provide or disclose data through a social media site may have (sometimes inadvertently) provided broad informed consent for secondary use of such data, which could encompass health-related research. Yet there is dispute as to whether broad informed consent provided to a social media site is sufficient for health-related research. Guideline 22 of the CIOMS International Ethical Guidelines notes that:

> ... users rarely adequately understand how their data are stored and used. And despite the insights that may result from this high volume of data, legal and ethical standards are unclear due to changing social norms and the blurring of boundaries of public and private information. Although the information may be collected from a public source, researchers should acknowledge that persons may be unwilling to have their data obtained for studies, and should account for the privacy norms in communities sharing information online. Users may not fully understand or appreciate the consequences of their actions, and may feel violated when their information is used in a context they did not anticipate.

There are three ways in which researchers may collect social media data for AI-related health-related research: (a) social media data collected directly from individuals participating in a research study; (b) social media data collected indirectly from data participants who would have a reasonable expectation of privacy; and (c) indirectly collected social media data that are publicly available. Different types of informed consent could be applied to these forms of data collection, whether specific informed consent, an opt-out (or opt-in mechanism), or public notice of the collection of data for research purposes.

For social media data collected through an online environment directly from individuals who are participating in a research study, researchers should obtain specific informed consent. Guideline 22 of the CIOMS International Ethical Guidelines recommends the use of an opt-out procedure for this form of data collection. However, as

research is increasingly conducted only through an online environment, and with the availability of online tools, specific informed consent may be both feasible and necessary to safeguard individual rights.

Researchers might collect social media data indirectly from participants who would have a reasonable expectation of privacy of their data, including information shared in private groups. In such situations, private groups should be given the option to not consent to any use of data for research purposes. Even if private groups do consent to the potential use of their data for research studies, people should be able to opt out of data collection and research, and individuals should be provided with adequate information to decide. This requires research participants to have been informed of the purpose and context of intended uses of data and information; the privacy and security measures used to protect their data, and any related privacy risks; the limitations of the measures used; and the privacy risks that may remain despite the safeguards put in place. In case of refusal by the person approached, researchers should refrain from using the data of this individual. An opt-out procedure could fulfil the following conditions to ensure individuals can make an informed decision: (a) persons need to be aware of the research study's existence; (b) sufficient information needs to be provided to the research participants; (c) persons need to be told that they can withdraw their data to the extent possible; and (d) a genuine possibility to object must be offered.

Finally, researchers may collect social media data indirectly from publicly accessible websites. In this situation, CIOMS Guideline 22 recommends that "researchers collecting data on individuals and groups through publicly accessible websites without direct interaction with people should, at a minimum, obtain permission from website owners, post a notice of research intent, and ensure compliance with published terms of website use". In recent years, research studies (that may not require the use of AI) apply sentiment analysis to social media data to retrospectively assess people's attitudes to a certain phenomenon (for example, vaccination or abortion). The guideline does not require broad consent for reuse of this type of social media data for AI-related health research, nor does it require providing an option to opt out of such reuse of social media data. Nevertheless, given the increased importance of this type of social media data for health research, social media applications or other sites that collect data could, for example, (a) provide clear information to users that their anonymized data could eventually be used for research purposes without future consent; or (b) establish independent data access committees to review requests for social media data to ensure that the research, and the use of the data, adheres to ethical obligations.

All social media data should be de-identified to protect the privacy of those individuals who provided data. Yet since social media data can often be re-identified due to the public availability of social media data, researchers should adhere to the principle of data minimization and introduce additional safeguards. These can include an increased emphasis on the security of the data throughout the research process by controlling or limiting what data are made available, to whom, and in what environment, and deleting data once they are no longer needed *(39)*. Social media data should also be stored separately from other data to mitigate security-related risks.

2. Ethical governance of health data

Appropriate governance of health data has been called a prerequisite to ethical governance of health research, since health data are the "foundation of AI development" – including the training of AI algorithms *(40)*. A core requirement is to anonymize data and to preserve anonymization. Yet governance extends to other requirements. This section explains several other aspects of the ethical governance of health data, including requirements enumerated under data protection laws as well as those that ethical standards for research may require. Under data protection laws, such as the General Data Protection Regulation, these requirements can satisfy the "suitable and specific measures" required to process personal health data.

One requirement is that researchers should be transparent with the use of health data – whether storage, transfer, access requests by third parties, or additional uses of data that are related to but separate from the initial reason for collecting such data. This can include notification for use of data that have been previously collected and are anonymized for health-related data science research with AI. Notification could be achieved through several different avenues, whether "publication on websites and social media, individual notification as well as broad notification through posters, emails, brochures, social media, or web portals" *(17)*.

A second requirement is that researchers should satisfy legal requirements, including those under data protection laws, for data minimization (or limiting the collection of information to what is directly relevant and

necessary to accomplish a specific objective) *(17)*. This requirement may create a quandary for researchers who have to satisfy legal requirements to the detriment of the expected benefit of combining different data sets to improve research outcomes. Combining data sets may not only violate data minimization requirements but also increase the risk of de-anonymization, for example when data sets that are separately anonymized are brought together.

Data protection authorities, RECs or data access committees, where they have oversight authority, must ensure that researchers remain in compliance with data protection laws and do not risk undermining the privacy rights of data donors, especially since individuals may have provided neither specific nor broad informed consent for use of data. Where researchers seek to combine data sets to improve research outcomes, an oversight mechanism should require researchers (a) to justify the scientific value of combining data sets so as to satisfy data minimization obligations; (b) document the risk of re-identification due to combining such data sets; and (c) where there is a risk of re-identification, for researchers to seek specific informed consent for collection and use (of newly collected data), or – for data that had already been collected – to retroactively contact individuals who provided such data, whether health data or social media data, so they can decide whether to opt out. Since retroactively contacting individuals to opt out may be challenging – whether due to a data donor's geographic mobility in some countries, the death of the data donor, or donors who cannot recall the original data collection – researchers should not use data where there is a clear risk of re-identification.

A third requirement for appropriate governance is for researchers to effectively manage data sharing. Data sharing, or the process of making the same data resources available to multiple applications, users or organizations in a form that is broadly usable, generates multiple benefits. These include "resource preservation, knowledge advancement, enhanced data integrity, transparency [and] public accountability" *(37)*. Data sharing helps to advance open science and a culture of openness in science, which also provides numerous benefits. Finally, data sharing may help to overcome known challenges with biased data sets that exclude certain populations and communities.

Yet there are also risks and problems tied to data sharing. Additional uses of an individual's data may undermine a person's privacy interests. The Indian Council of Medical Research states that "[a]dditional consent from patients is required for data sharing if not taken previously. The consent must contain the nature of data, to what extent it is being shared, and possible harm that can occur from sharing data" *(21)*. Broad consent may not be sufficient for data-sharing arrangements that could not have been foreseeable and may involve uses of data that would not be acceptable to a data subject.

There are also undertones of what has been called "health data colonialism", or the acquisition of health data from low- and middle-income countries by AI researchers and developers from high-income countries, whether to build algorithms or to support research *(40)*. Such practices may be driven primarily by commercial gain but can also reflect power imbalances in scientific research between researchers and institutions in high-income countries and researchers, institutions, governments and populations in low- and middle-income countries *(41)*.

Anonymization does not alleviate challenges with data sharing. Individuals may not have wished for their data to be used for additional research purposes that had not been foreseeable, and for which consent may not have been sought. This is especially important for any reuse of data for commercial purposes for which informed consent has not been obtained. Such forms of data sharing must be avoided. Furthermore, even if data are anonymized, risks remain of re-identification, as well as risks to data from breaches, cybersecurity threats *(37)*, or due to a hacking operation that puts data in the public domain *(38)*.

Data sharing may be facilitated through data hubs and data spaces (see section 5), as well as data repositories, which are often used to store potential research data. The simplest repositories are often free to use up to a certain data limit, with deposition and access via a website without any mediation or review by a committee (these are open repositories). Reuse conditions are set by the depositor, for example whether use is open or closed (for example, only metadata are available). For example, Denmark has established a "research machine", which allows users to upload data and link data to an array of registries connected to an individual's unique civil registration number. These sensitive and highly personal data, securely managed by the repository, enable researchers to access an online environment to conduct analysis using different statistical tools, while forbidding the sharing or export of the data *(42)*. Requirements may be set and enforced by a data access committee.

When an REC has authority to review a research project that uses data obtained through a repository, the REC may need to consider whether any repository reuse conditions apply and have been followed, including any requirements set by a data access committee.

The management of data, including international transfers of data, data security and privacy-related risks, is not yet clearly defined within and between countries. Thus, there are gaps in regulatory frameworks, especially as it relates to the sharing and transfer of data between countries (North–South and South–South data sharing) *(41)*, and as it relates to Indigenous data sovereignty. This is mitigated in part by most countries now having data protection laws that provide oversight and protection related to the sharing of data, including international data transfers. For example, Kenya's Data Protection Act requires that internationally funded projects are led by a local principal investigator with accountability for data sharing and use *(40)*. There are also wider efforts to regulate the transfer of data, for example the African Union Data Policy Framework *(37)*. Data transfer agreements can also include safeguards to address these challenges and can be reviewed by an oversight mechanism. However, since some countries do not have laws governing the cross-border transfer of data, research data "may be crossing borders without agreements or export permits in place" *(43)*. These deficiencies require governments to take steps, either individually or jointly, to address this legal vacuum.

Finally, a separate concern is the sale of health data, and the presence of data brokers who buy and sell sensitive health data (such as mental health data). One report, published in February 2023, found that the data broker industry:

> … appears to lack a set of best practices for handling individuals' mental health data, particularly in the areas of privacy and buyer vetting. It finds that there are data brokers which advertise and are willing and able to sell data concerning … highly sensitive mental health information. Some data brokers are marketing highly sensitive data on individuals' mental health conditions on the open market, with seemingly minimal vetting of customers and seemingly few controls on the use of purchased data. Pricing for mental health information varied: one data broker charged $275 for 5,000 aggregated counts of … mental health records, while other firms charged upwards of $75,000 or $100,000 a year for subscription/licensing access to data that included information on individuals' mental health conditions *(44)*.

Concern about the commercialization of health data includes individual loss of autonomy, loss of control over the data (with no explicit consent to such secondary use), how such data (or outcomes generated by such data) may be used by the company or a third party, with concern that companies are allowed to profit from the use of such data, and concern about privacy, as companies may not meet the duty of confidentiality, whether purposefully or inadvertently (for example, due to a data breach) *(7)*. While in some jurisdictions the sale of health data is not regulated, other countries may either forbid the sale of health data or place highly restrictive conditions on any sale. For data sets that are acquired as part of a research study, an oversight mechanism may require proof of provenance for the data sets to ensure they were developed and acquired in line with ethical principles *(45)*.

3. Group-level harms

Even if health data remain anonymized, the use of data-driven AI-related health research allows researchers to identify insights that could relate to a specific group that shares certain traits or characteristics. These insights could generate benefits for groups, for example identifying individuals in a group with genetic markers or other characteristics that have a higher probability of developing diabetes *(46)*. This research may especially benefit those groups that tend to be the most stigmatized and neglected.

Yet this research (even if the conduct of the research itself poses minimal risk) could produce outcomes that stigmatize specific groups, or eventually result in third parties using correlations to discriminate. Thus, certain risk indicators, such as genetic variants, neuroimaging biomarkers of addiction, and molecular biomarkers of chronic illness, could then be used by third parties to justify actions that discriminate against individuals that belong to a group, including higher health insurance rates (or denial of health insurance), or denial of other products or services because individuals of that group are deemed high risk *(38)*.

Furthermore, analysis of data, even if the data remain anonymized, can generate other unique group risks, including privacy-related risks, stigma and impacts on a group's psychological well-being, that can affect vulnerable groups represented in health data and for which researchers can draw inferences through AI-related health research *(5)*. For example, research on facial images has been used to draw inferences, with the use of AI, about an individual's sexual orientation with greater accuracy than on the basis of human judgement *(47)*. Data (for example, social media data) that initially had no apparent connection to one's health or other private information can generate insights that reveal private information, such as one's sexual orientation or a medical condition (such as depression). An algorithm could also discriminate in indirect and non-obvious ways that exceed legal definitions of discrimination and are not evident or comprehensible to humans *(48, 49)*.

Therefore, even as it is important to recognize that AI-related data science research can generate beneficial research outcomes on behalf of groups that share certain traits and characteristics, the conduct of research and the communication of results should avoid stigma or marginalization of groups under study.

4. Risk–benefit assessment

In ethically acceptable research, risks are minimized (both by preventing potential harm and by minimizing their negative impacts should they occur) and are reasonable in relation to the potential benefits of the study. The nature of the risks may differ according to the type of research to be conducted. Risks may occur in different dimensions (for example, physical, social, financial, spiritual or psychological), all of which require serious consideration. Further, harm may occur either at an individual level or at the family, community or population level *(34)*.

Assessing the risk–benefit ratio for health-related data science research with AI is difficult for several reasons. First, for research conducted with large numbers of data subjects, many or most risks or benefits will not be identifiable, and therefore the specific risk–benefit ratio for each data donor may not be possible to quantify. Second, even if data donors are known, the actual risk–benefit ratio may not be possible to identify up front, as researchers may modify or introduce a new research hypothesis, or additional research questions are tested, yet researchers may not resubmit such amendments for oversight. Third, it may be difficult to quantify the probability of a certain event occurring, such as a data hack of de-anonymization of health data.

Fourth, there are challenges to assess AI-related health research with respect to explainability and transparency. Certain research entities, especially companies or private sector actors, may be unwilling to disclose information that could assist with deliberations, including internal assessments of an algorithm's performance, or strategies that a company has taken to protect data from unauthorized disclosures *(38)*. On the other hand, explainability of how an AI algorithm arrives at certain conclusions may not be possible. Yet it may also be neither relevant nor necessary. Explainability may not be a good foundation for decision-making. In fact, ethical judgement can be made without having to rely on an understanding of the algorithm itself *(50)*.

5. Fair selection of research participants

Ethically acceptable research ensures that no group or class of persons bears more than its share of the burdens of participation in research. Similarly, no group should be deprived of its fair share of the benefits of research; these benefits include the direct benefits of participation (if any) as well as the new knowledge that the research is designed to yield *(34)*.

Researchers should understand any biases in the data and identify what steps might be required if a data set is biased or does not include relevant groups that could benefit from the research. Researchers can thereafter mitigate bias by incorporating additional data sets, testing specifically for bias, publishing the results, and describing the results as having limited applicability and with potential risks, or even labelling the final research product as potentially harmful for excluded populations. It can also mean ensuring that the purpose of the research does not exceed the data set's characteristics.

RECs or other oversight mechanisms must be able to determine whether a data set that has been collected and will be used for a study satisfies a fair selection of research participants, understand any biases in the data (for example, through a bias analysis of the data set), and what steps might be required if a data set is biased or does not include relevant groups that could benefit from the research. RECs may also need to determine how to treat novel forms of research that use previously collected or publicly available data sets to answer research questions, as opposed to traditional studies, in which individuals are prospectively enrolled after REC review. For example, RECs may have to decide if the potential benefits and insights of a data science study justify use of a data set that is biased and for which the results will only benefit certain populations that are represented in the data, and whether corrective actions adopted by researchers consider and mitigate bias.

This does not mean that researchers should abandon or forego a study because the data set only has limited applicability to those represented in the data. There are certain inequities in societies that researchers cannot overcome, even though researchers should make reasonable efforts to address existing biases when collecting data for a research study. One such inequity – the representativeness of mobility data collected through mobile phones and smart phones – can lead to biases that replicate historical and modern forms of inequality. Use of mobile phone data for health-related data science research also introduces other ethical risks (Box 4).

Box 4. Ethical benefits and challenges with the use of mobility data for infectious disease and public health research

The coronavirus disease 2019 (COVID-19) pandemic led to the broader use of mobility research from data collected from digital technologies (such as cell phones and smart phones) to track transmission routes, calculate the effects of health policies, or assist with prediction of how a disease outbreak may evolve. Such mobility data can also be used for other public health interventions, including improved city planning, optimization of transportation routes, and optimizing the allocation of health resources. In many settings, the widespread use of cell phones and smart phones provides a data-rich environment that can be used flexibly by researchers.

However, the use of mobility data for data science studies (with or without AI) introduces several ethical challenges. Individuals may not realize that they have consented to the use of their cell phone data, as such consent may be buried within privacy policies that individuals may neither read nor fully comprehend. Re-identification may also remain a problem as algorithms become more sophisticated and if researchers decide to combine data sets. Mobility data may mirror existing inequalities in societies and may be more likely to represent the movements of those who are "male, relatively wealthy and largely urban". There may not be adequate community engagement on the uses of such data (and the benefits of such use), which can reduce public trust. It is possible that researchers may be unable to identify risks that are evident to community members but not to data scientists. Finally, such research may lead to function creep, as "techniques in cell phone data-driven mobility research to promote public health ... will end up being used for questionable surveillance, commercial or political purposes".

Thus, even though such techniques and data sets could engender tremendous public health benefits, RECs and other oversight mechanisms should ensure that researchers proactively address and minimize such risks and engage communities that provide such data to engender trust and to maximize the benefits of such research. *(51)*

One strategy that AI researchers are employing to improve the representativeness of data sets is use of synthetic data (artificially generated data that mimic real-world patterns and characteristics). The growing use of synthetic data is discussed below (section 2.2), including its benefits and risks.

6. Treatment of crowd workers

Crowd workers are temporary workers, often recruited through crowd-sourcing platforms, who can be rapidly enlisted to collectively complete large tasks, such as annotating massive data sets used for the development, validation and use of AI algorithms. Crowd workers (or data labellers) form the backbone of data science and AI-related research – including the collection, creation, labelling and analysis of data *(5, 33)*. Their role and function differ from the use of data collectors who have been enlisted historically to collect data (for example through household surveys). Data science research introduces risks of poor treatment of crowd workers and a lack of labour laws and relevant protections. These risks to the dignity and well-being of crowd workers differ from the conditions of employment faced previously by data collectors or research assistants and have triggered a debate as to whether crowd workers within AI studies should be treated as research participants, or as members of the research team. In both cases, this could trigger ethics review *(5)*. While some experts have called for all

research using crowd workers to be reviewable by an REC, there are nuanced approaches that define specific circumstances under which the role and activities of crowd workers should be characterized as research participants and trigger ethics review *(5)*.

Kaushik et al, *(33)* have identified three categories of research involving crowd workers and whether crowd workers should be treated as research participants for each category. First, there are "clear-cut" cases for which research should be submitted to RECs (or for an REC to declare such research exempt). This is where "researchers interact with crowd workers to produce data *about* those individuals, and then analyse those data to produce generalized knowledge about a population from which those individuals are considered representative samples" *(33)*. This category can include (a) studies in which researchers are determining best practices for using crowd workers; (b) best practices for generating data sets with crowd workers; or (c) research that involves interaction with crowd workers "to produce data about the crowd workers ... to answer research hypotheses, which created generalizable knowledge" *(33)*.

A second category involves research with crowd workers that does not necessarily require ethics review. The crowd workers may assume tasks usually delegated to researchers, such as labelling data. While the crowd worker is participating in the production of data for which knowledge will be generated, the crowd worker is not performing any different function than a researcher and is not producing any information or data about themselves *(33)*. In some countries, such as New Zealand, the ethics committee may consider the safety of the research team, which could include a review of the working conditions and challenges of crowd workers.

A third category includes situations in which research could trigger ethics review, but such research may be assembled or managed in a manner that either intentionally or unintentionally avoids oversight. For example, researchers could collect information both about and from crowd workers but only analyse and generate knowledge about crowd workers in a second, separate study for which the data about the crowd workers have been anonymized (thereby exempting the data from ethics oversight) *(33)*.

Appropriate oversight of research that relies upon crowd workers may face additional challenges. First, many researchers may not realize that research using crowd workers may require REC review. Second, existing laws, such as the Common Rule in the United States, have certain loopholes that prevent RECs from effective oversight *(33)*. Third, there are concerns that research that uses crowd workers may be difficult to reproduce *(33)*.

Even in situations where research using crowd workers (for example to label data) does not trigger ethics review, there still may be serious concerns with the treatment of crowd workers. There can be a mental and psychological toll on people responsible for reviewing content, annotating data used to train LMMs, and removing abusive, violent or mentally disturbing content from a data set. Those responsible for filtering out content are often based in low- and middle-income countries and can suffer psychological distress, without access to counselling or other forms of medical care. Furthermore, crowd workers are often employed by intermediary companies and are paid low wages that are not in accordance with the minimum wage regulations in countries where they are employed *(8)*.

Researchers that rely upon crowd workers should ensure that crowd workers are informed of the purpose of their tasks and compensated fairly. Measures adopted to safeguard crowd workers should be shared with an REC or other oversight mechanism. Beyond the responsibilities of researchers, RECs or other oversight mechanisms, the enforcement of existing labour laws or a revision is necessary to provide protection. Furthermore, there could be obligations placed on platforms that host crowd workers to provide appropriate labour protections, as well as consequences for those platforms that do not satisfy these standards.

2.2 Research with AI tools and technologies

This section addresses the growing use of AI tools and technologies to conduct research. Subsection 2.2.1 describes how AI tools are used by researchers to conduct health research; subsection 2.2.2 examines ethical concerns and benefits of specific tools; subsection 2.2.3 examines the use of generative AI for health research, and how oversight mechanisms can account for risks with these popular tools; subsection 2.2.4 examines the risks and benefits of the use of synthetic data; and subsection 2.2.5 examines some deficiencies with the testing and validation of tools by researchers, and how testing could ensure safe and effective use.

2.2.1 What comprises research with AI tools and technologies?

AI is increasingly used as a tool to support or conduct research. This can include the use of AI as a tool to formulate hypotheses and design research studies, to identify research participants, to obtain informed consent, to analyse results, or to produce scientific and academic output, including with generative AI (LMMs). One challenge in identifying the types of AI tools that are used for research is that there are no agreed definitions, and many developers may be utilizing commercial marketing in lieu of evidence to encourage use of their AI technologies for research. Furthermore, as the use of AI becomes common in health research, treating such tools as a stand-alone, exceptional category may not be useful.

The following are ways researchers are already using AI to assist with the conduct of health research:

- formulate hypotheses and design research studies;
- identify, target, and recruit (or enrol) research participants for a specific research study, including using algorithmic identification, wherein an algorithm is trained on and then analyses various types of data to identify individuals "with a particular pathology or impairment" *(52)*;
- conduct a study, such as using a chatbot to facilitate collection of data from a research participant (even if the chatbot itself is not under research scrutiny);
- obtain informed consent of research participants through use of AI-based chatbots;
- generate synthetic data for use in health research;
- produce examples (such as messaging) that can be used in a study (for example, the use of generative AI to produce public health messaging that is tested for its effectiveness);
- develop a counterfactual to assess whether a proposed research intervention (that may or may not use AI in the study) could produce an improved outcome compared to the status quo;
- use generative AI to select and analyse research results, such as whether to include or exclude specific data, or generate text or figures to disseminate findings.

These use cases could accelerate research, expand the ability of health researchers to reach out to (or emulate with synthetic data) a diverse set of research participants that otherwise are not possible to identify, diversify (with the use of AI) the types of messages or outputs tested on research participants, reduce the burden of conducting research, and improve the decision-making of a research team on whether to proceed with a research study. While these benefits should be welcomed, use of such tools may require scrutiny by an REC or other oversight mechanism before use.

2.2.2 What are the benefits and ethical concerns with AI-based tools to conduct health-related research?

The following are two examples of AI-based tools that are used for health research to illustrate the different benefits and concerns.

1. Use of AI to recruit or enrol patients

The use of AI to recruit or enrol patients, including identification of individuals who do not know their health status but may qualify for a research study, may be an opportunity to provide critical care to individuals. Yet it may undermine informed consent requirements and the autonomy of the relevant person if predictions of health status are shared with people who did not consent to surveillance, detection or use of predictive models to draw inferences about their current or future health status. This could include non-consensual misuse, for example, to identify individuals with tuberculosis who do not know their status or individuals at high risk of HIV infection who are thus candidates for pre-exposure prophylaxis *(7)*.

Identification also involves the generation of new information about an individual, and as it relates to a medical condition, is also sensitive and private. This could have a significant impact on an individual's legal status and psychological well-being, as well as the forms of discrimination that they may face, such as stigma associated with a condition *(52)*. In some cases, irrespective of the potential benefit through such identification, a person may simply not wish to know their status. 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 that: "Everyone is entitled to know any information collected about their health. However, the wishes of individuals not to be so informed shall be observed" *(7)*. To justify use of these technologies, researchers may need to take additional steps to ensure that technologies are used appropriately, including how individuals are informed (ethical communication) *(52)*, their right to refuse participation in a research study (or anything else that flows from identification), and appropriate treatment or medical services for those individuals who were identified without their explicit consent.

A separate concern with these tools is that they could encode certain biases. Thus, if an AI tool is used for identification of research participants, biases encoded in the research tool may undermine the ability of an AI tool to identify those individuals whose characteristics were not represented adequately in the data set *(52)*. This could deepen existing research biases and health inequities and call into question whether a research cohort meets the requirements for fair subject selection. While an REC or other oversight mechanism may not be able to identify the bias in the research tool itself, it could require a researcher to document the due diligence it has performed to ascertain any biases, how biases may affect a research study, and what steps have been taken to address a known bias.

2. Use of AI-based chatbots for informed consent

AI-based chatbots, which have been deployed widely to facilitate human conversation for a product or service, are being tested for possible use to facilitate informed consent for biomedical research. One study used a chatbot to provide potential participants with information about the study, educational material and a quiz. There were several benefits to the use of a chatbot, including increased efficiency of the informed consent process, which took 44 minutes with the chatbot as opposed to 76 minutes through in-person consent, while the total time from referral to consent was five days compared to 16 days with in-person consent (which could be due to scheduling challenges). Most participants reported a positive experience, and nearly all participants successfully completed the quiz *(53)*.

However, there were several limitations to the study (study size, lack of randomization, lack of diversity of the study group). There are also broader concerns with use of an AI-based chatbot: first, it could "dehumanize" informed consent, thereby appearing to prospective research participants as little more than a click-through agreement as opposed to a careful consideration of risks and benefits that each research participant must weigh; second, there is not yet any technical or regulatory guidance for RECs or other oversight mechanisms to determine whether use of a chatbot is acceptable; and third, these uses of a chatbot may be inappropriate for vulnerable individuals (as well as children and adolescents) who require more careful and empathetic engagement. Nevertheless, use of AI-based chatbots may be useful for low-risk research that may already rely upon electronic consent, and additional studies of the benefits and risks of AI-based chatbots, as well as overarching guidance to support their use by researchers and oversight by RECs or other entities, could address these concerns.

2.2.3 What are the benefits and risks with generative AI (large multimodal models)?

Large multimodal models (LMMs) extend the ways in which AI can support scientific and medical research. LMMs can be used for a variety of aspects of scientific research. They can generate text and figures to be used in a scientific article, for writing manuscripts or for a peer review. For example, one survey of over 1600 researchers conducted by the journal *Frontiers* found that over half of researchers used LMMs for peer review of scientific transcripts *(54)*. LMMs can be used to summarize texts, including summaries for academic papers, or can generate abstracts. LMMs can also be used to analyse and summarize data to gain new insights into clinical and scientific research. They can be used to edit text, improving the grammar, readability and conciseness of written documents such as articles and grant proposals. Additionally, LMMs can provide insights from the data with which they were trained. One LMM, trained on millions of academic articles, is claimed to analyse scientific research to answer questions, extract information or generate relevant text *(8)*. Finally, LMMs can be the basis for chatbots for health research intended to reach large numbers of people, even if the chatbot itself is not under examination.

However, the increased use of LMMs poses different risks to individuals and for societies and health systems, and, as generative AI also becomes more common for writing amongst students and academics, may lead to many of the risks and problems associated with generative AI polluting the production of scientific and medical knowledge. These risks, which are enumerated in WHO guidance on LMMs published in January 2024, include the following.

Box 5: Risks of LMMs enumerated in WHO Guidance

- **Lack of accountability.** The authorship of a scientific or medical research paper requires accountability, which cannot be assumed by AI tools. Lack of accountability was the basis for the decision of a major academic publisher and the World Association of Medical Editors not to accept LMMs as a credited author.
- **High-income country bias.** Most of the scientific and medical research used to train LMMs is conducted in high-income countries. Thus, the outputs of any LMM query are likely to be biased towards a high-income country perspective, including the use of an LMM to analyse research results. A bias towards high-income country perspectives can also mean that, for example, an AI-powered chatbot used to generate messages for public study may formulate communications in ways that are biased against certain populations.
- **Limited utility beyond English (and French and Spanish).** Most LMMs – which are often the basis of chatbots or other forms of AI that produce communication – operate only in English (and were trained only on data in English). This can mean that use of chatbots could only work optimally for languages for which there are enough training data on which to train an LMM appropriately *(8)*.
- **Hallucination or misinformation.** An LMM may "hallucinate" by summarizing or citing academic articles or other information that do not exist, or generate an analysis based on a hallucination. Hallucinations are indistinguishable from factually accurate responses generated by an LMM, because even LMMs with reinforcement learning from human feedback are not trained to produce facts but to produce information that looks like facts. One study found that large language models, when provided with a simple set of facts to summarize, would hallucinate at least 3% of the time and as high as 27% *(55)*. Furthermore, current LMMs also depend on human "prompt engineering", in which an input is optimized to communicate effectively with an LMM. Thus, LMMs, even if trained specifically on accurate information, may not necessarily produce correct responses *(8)*.

The benefits and risks of generative AI in research have been recognized as an important issue by governments and regulators. For example, the European Commission, in March 2024, published *Living guidelines on the responsible use of Generative AI in research* "to prevent misuse and ensure that generative AI plays a positive role in improving research practices" *(56)*.

Blau et al. note five broad concerns that generative AI may pose to scientific integrity. They note:

> Generative AI … makes it more difficult for scientists, the larger research community and the public to (1) understand and confirm the veracity of generated content, reviews and analyses; (2) maintain accurate attribution of machine- versus human- authored analyses and information; (3) ensure transparency and disclosure of uses of AI in producing research results or textual analyses; (4) enable the replication of studies and analyses; and (5) identify and mitigate biases and inequities introduced by AI algorithms and training data *(57)*.

Researchers, RECs, and other oversight mechanisms could all play critical roles in preserving and strengthening scientific integrity and addressing other risks associated with generative AI. RECs must be able to account for these risks when examining research that uses LMMs, including anticipated use of an LMM to analyse and summarize research results. They should require transparency from researchers as to how LMMs will be used throughout the research process, including writing both protocols and results, and what safeguards will be put in place, including the identification and mitigation of biases. This will require researchers to clearly delineate between human and AI-generated information and ideas, including appropriate citation of human ideas even if generative AI outputs do not furnish such information *(57)*. Researchers should themselves monitor and test for biases in AI algorithms and outputs and identify and correct biases that could affect research outcomes *(57)*. The use of generative AI may also require RECs or a different oversight mechanism, at the end stages of the research process, to conduct continuing review.

There are also measures that third parties can take to address proliferating use of LMMs within the research process and production of outputs. Leading medical and scientific journals have already responded to the emergence of LMMs, their potential and their impact on scientific research. One academic publishing company has established two rules: (a) an LMM will not be accepted as a credited author of a research paper; and (b) researchers who use LMMs should document their use in the section on methods and in the acknowledgements *(58)*. The World Association of Medical Editors has restricted authorship to humans *(59)*.

Other journals – while not permitting authorship with LMMs – are encouraging the responsible use of generative AI. One leading journal declared: "We have elected instead to allow the use of LLMs for submissions, if authors take complete responsibility for the content and properly acknowledge the use of LLMs. However, this policy does not allow an LLM to be listed as a co-author" *(60)*. To encourage responsible use of LMMs while avoiding undue and harmful reliance, Mann et al. propose three criteria for their use: (a) human vetting and guaranteeing; (b) substantial human contribution; and (c) acknowledgement and transparency (of the use of LMMs) *(61)*.

2.2.4 What are the potential benefits and risks of the use of synthetic data?

There is no one universally accepted definition of synthetic data, and their use is not limited to just health care (and health research) but is and will be applied across many fields. Synthetic data, as defined by Arora et al., are "data generated by a purpose-built AI model, trained on real-world data, such that the synthetic data maintains the aggregate properties of the original data" *(62)*. Synthetic data can either be partially or fully "synthetic". Partially synthetic data combine real-world data with synthetic data, which, for health care-related purposes, can be a proxy for real-world data and protect patient privacy, allowing researchers to still conduct analyses *(63)*.

The interest in and use of synthetic data has been driven by commercial, scientific and ethical rationales. Many companies are employing synthetic data as a means of training generative AI models *(64)*. This is because even as companies require ever more sources of human data to train general-purpose foundation models, there may be fewer data available, either because they have already been utilized or because there are restrictions placed on the sharing and use of data by those who store, manage or oversee the data.

Synthetic data have also been identified as a tool for AI-related health research. First, because the data are either partially or fully synthetic, and contain little or no precise information about individuals, they could help to address concerns with privacy of health data *(65)*. Second, synthetic data could mitigate or overcome bias in data sets or at least balance data sets for those characteristics or qualities that otherwise are not possible to obtain *(65)*. Third, even if synthetic data will not be an exact copy of the underlying data on which they are modelled (fidelity), synthetic data could have utility as a stand-in and can be evaluated for usefulness (though this may not be possible if there are not enough real-world data that can be used as a comparator) *(65)*.

Yet the use of synthetic data for AI-related health research also introduces risks:

First, even though synthetic data are considered a privacy-preserving technique, re-identification of individual data subjects is not impossible, and partially synthetic data can be vulnerable to re-identification. Second, there are concerns that there may be use of synthetic data as a means of circumventing data protection laws, research ethics oversight or other forms of scrutiny *(65)*. Forms of data sharing that would otherwise be deemed to be unethical or in violation of data privacy laws may not apply if a company or other entity created and shared a synthetic version of the data *(62)*. Third, even though synthetic data have been deemed a means to overcome bias, synthetic data generation tools may produce data sets for which utility and fidelity differ significantly depending on the sociodemographic group *(65)*.

Furthermore, synthetic data, if used improperly by researchers, could exacerbate or introduce community- or group-level harm, for several reasons. Synthetic data may oversimplify or generalize complex characteristics or identities of communities that may not be apparent to researchers who rely upon the output. The consequences can mask specific risks or benefits that research could otherwise reveal. As noted by Koul, Duran and Hernandez-Boussard: "When synthetic data fail to preserve relationships between variables, predictive models might systematically underestimate the risks for clinically vulnerable populations, leading to inadequate care strategies" *(66)*.

Second, the production of synthetic data can generate intersectional hallucinations that misrepresent or invent associations or correlations between different demographic or clinical variables that do not exist but are treated as real. Intersectional hallucinations can therefore misrepresent the diversity of a data set and invent associations between different identities – for example between age, race and gender – that do not exist in the underlying data *(67)*. Intersectional hallucinations may make synthetic data unsuitable for certain types of health research *(66)*. And even though techniques can be used to minimize intersectional hallucinations, they must always exist in synthetic data, as otherwise the synthetic data would be a "mirror" of the original data (intersectional fidelity), which would undermine efforts to preserve privacy *(67)*. The necessity of intersectional hallucinations to assure privacy creates a challenge. If synthetic data are required to provide a "high-fidelity representation" of data, including "complex variable relationships", the synthetic data will have to better approximate the underlying real data, which therefore increases privacy risks *(66)*. On the other hand, if the synthetic data are too dissimilar, clinical validity is compromised *(66)*. Researchers will have to carefully consider the trade-off between fidelity and privacy.

Third, synthetic data, instead of overcoming bias, could instead amplify bias in the original data that an algorithm was trained upon *(66)*. There is additional risk of bias if an algorithm to produce synthetic data was trained on source data comprising a small sample size *(68)*.

Therefore, use of synthetic data may require guardrails so that their use respects ethical principles. Researchers who wish to use synthetic data should only do so as a means of strengthening privacy protection, and not as a means of escaping oversight. Additional caution may be required for video- or voice-based synthetic data, since it may be impossible to fully anonymize the data. To guard against both inaccuracy and potential privacy violations, there may be a need to label both where there is intersectional fidelity and where there is intersectional hallucination *(67)*. More broadly, as synthetic data become more commonplace and are included in open-source data sets, they should be clearly labelled.

There are different views as to whether individuals who provide data would have to subsequently provide their informed consent for the use of their data to generate synthetic data. On the one hand, since synthetic data that are produced are not an exact replica and will anonymize the underlying data, there may not be a legal requirement, for example under data protection laws, to seek informed consent. Yet since personal data may have been used to train an algorithm that produces synthetic data, and those data sets may include personally identifiable information, it may not be possible to fully anonymize the data (as may be the case with audio- or video-based data or due to intersectional fidelity). As a result, the use of synthetic data may carry greater risk, including re-identification. Therefore, in some jurisdictions, the use of personal data may require informed consent or a different form of legal authorization before such data can be used.

Synthetic data may also have the unintended consequence of limiting contact and engagement between those who design and carry out research and the people and communities it intends to benefit. It may be more important to find ways to increase engagement and contact between researchers and affected communities, even if there is limited reliance on the collection of data from individuals *(65)*. One reason researchers may use synthetic data is to create a more robust data set on behalf of a population that otherwise is underrepresented. This may not, however, reflect the actual real-world diversity of the population in question. Instead of seeking to build a larger data set only through synthetic data, efforts could be made to supplement the synthetic data set through working with underrepresented communities and collecting a more comprehensive data set *(66)*.

Transparency may be critical to ensure appropriate use. This includes the lineage of the synthetic data, the bias, hallucinations, or privacy risks in the data set, and disclosure by researchers of how synthetic data have been or will be used *(66)*. RECs or other oversight mechanisms may have to take steps, where necessary, to engage in oversight of the prospective use of synthetic data by researchers, including any disclosures that researchers should provide, and safeguards that researchers may need to introduce to improve privacy protection and avoid bias. Governments and standard-setting bodies may require more information and evidence of the utility, performance and risks of synthetic data before issuing recommendations, including the types of information that must be disclosed, to ensure the appropriate use of synthetic data as they become more commonplace in health-related research.

2.2.5 Testing the performance of AI-based research tools

One concern with AI algorithms is that developers often only rely on internal validation (evaluating an AI model on internally held data not used to train or validate the algorithm) or external validation (use of an independent data set). Few developers are publishing the results of their analyses, which are necessary so that the results can be peer-reviewed. For AI systems used in medical settings, for example, only 65 randomized controlled trials of AI interventions were published between 2020 and 2022 *(69)*.

Developers should not just consider internal and external validation, but also evaluate performance under real-world conditions of an AI model that would be used to conduct or support research *(52)*. Such evaluation, and publication of research protocols as well as results, would be essential for transparency, validation and scrutiny by researchers, RECs and other oversight mechanisms.

The type of validation that is required would depend in part on the proposed use of the AI tool, and the type of function it would carry out. A simple AI algorithm that scrapes data from a public website (which may introduce ethical concerns but is not a complex operation) may not require validation or performance testing. A generative AI model used to produce synthetic data for a research study should require additional scrutiny. For example, Koul, Duran and Hernandez-Boussard call for "rigorous validation protocols that combine clinician expertise with computational stress testing to verify the fidelity of synthetic data to real-world medicine" *(66)*.

Based on validation, researchers should demonstrate that algorithms used for AI-related health research are appropriate for the intended use, and thereafter researchers must communicate the limitations of AI tools, both to RECs and eventually in published research results. RECs require information to evaluate (a) the potential use of research tools (how risky or novel is use); (b) whether tools are appropriate for the intended research (including the limitations of a tool); (c) whether risks (such as bias) that do exist have been mitigated, or whether risks are acceptable given the potential benefits of the research tool; and (d) how researchers should communicate the limitations of tools. In cases where documentation is not provided or information pertaining to a tool's validation is not adequate, RECs or other oversight mechanisms may wish to ask for improvements and additional disclosure to inform their deliberations *(45)*.

2.3 Health-related research on AI tools and technologies

This section addresses the third category of AI-related health research – health-related research on AI tools and technologies. Subsection 2.3.1 explains what health-related research on AI tools and technologies is; subsection 2.3.2 examines whether ethics review or other oversight includes such research; subsection 2.3.3 examines the novel risks associated with these AI technologies; and subsection 2.3.4 examines how different ethical requirements for research may be challenged.

2.3.1 What is health-related research on AI tools and technologies?

There are many uses of AI for health, including medical applications utilizing generative AI and the integration of AI into medical devices and software systems. AI is also applied to the administration of health care, for example in the case of health insurance and hospital management, as well as medical education. These applications of AI can be the object of research but may not be subject to oversight or ethical review.

Initially, potential applications of AI, especially using discriminative AI, included screening and triage, diagnosis (and prediction-based diagnosis), clinical care (including decision support and treatment recommendations), and use of AI for public health-related interventions, such as public health messaging (with the delivery of messaging through online channels such as social media and messaging services) *(7)*.

Generative AI is applied to existing use cases, and the technology has also facilitated novel uses of AI for health. In WHO guidance on the ethics and governance of LMMs, WHO identified five broad uses of LMMs: (a) diagnosis and clinical care; (b) patient-centred applications; (c) support for clerical functions and administrative tasks; (d) nursing and medical education; and (e) scientific and medical research (discussed above) and drug development *(8)*. As AI technologies are developed and approved for use in these diverse areas, implementation research, or the systemic approach to understanding and addressing barriers to implementation and scale-up of effective and high-quality health interventions, will become more relevant.

2.3.2 Does health-related research on AI tools and technologies require ethics oversight or REC review?

AI-based health tools or applications that are intended for direct use on human beings or to influence human behaviour should be tested for their utility, safety, efficacy and implementation, and will require research with human participants. Health-related research with human participants to test AI-based health tools or applications should be subject to ethical oversight and REC review.

For other AI-based tools and technologies, there is less clarity. Research on AI-based tools used in health care, for example for health insurance, health financing or the administration of health care, may not involve human participants and therefore is not likely to be subject to REC review. Yet these types of research may introduce ethical concerns, for example related to bias encoded in these technologies that could cause individual or society-wide harm, if they are not properly addressed during research. However, there is not yet a mechanism that could provide ethical oversight of these types of research.

Companies that develop and test AI tools and technologies may conduct health research that is for product development or quality assurance. This research may not be subject to ethics review, and there may be no other type of oversight that could be applied. This is because this research, when carried out by a company, may be characterized as a commercial activity or product improvement, and not as research on human participants. The research may be focused on either (a) the development of an algorithm *(40)*, or (b) improving a product for commercial purposes, such as collecting responses and data to optimize an algorithm (for example, how users respond to and interact with an experimental chatbot). Companies developing an algorithm or conducting product testing may not even submit research to an REC to determine if it is reviewable. Likewise, RECs may determine that research is not within the remit of RECs because it is regarded as quality control or product improvement.

2.3.3 What novel risks are associated with research on AI tools and technologies?

Prior WHO publications on the ethics and governance of AI for health have enumerated ethical issues associated with the use of AI technologies for health. These risks are different from the customary risks that RECs or other oversight mechanisms might address when reviewing a research protocol, and may include the following.

Biases encoded in AI technologies

Since AI technologies are likely to encode certain biases (or introduce bias), researchers may need to demonstrate that entities who designed and developed an AI algorithm have taken steps to avoid, identify and redress biases, including bias in the data sets used to train and validate AI algorithms, bias in how the algorithm was designed, and possible contextual bias in the AI tool or technology – in that the algorithm is designed for optimal use in one context but not in other contexts (such as low- and middle-income settings or rural settings with less infrastructure and internet connectivity). There may also be a need to review the outputs, to the extent that the outputs of an AI technology under study are available to identify biases.

AI-focused privacy-related risks (including those related to facial or voice recognition)

There are other privacy concerns that go beyond concerns with informed consent or the de-anonymization of data. AI technologies, for example, those used for public health surveillance, patient monitoring or proximity tracking applications (such as those used during the SARS-CoV-2 pandemic), can undermine a person's right to privacy, including tracking their movements. Other examples of AI technologies that present novel privacy risks are those AI systems that rely upon facial data or voice data for a medical or health-focused output. For example, research has identified that AI could use a person's voice data (the pitch of a person's voice) as a means of monitoring blood sugar levels *(70)*. A separate research project postulated the use of voice samples to identify individuals with depression or suicidal thoughts (see Box 5 for a case study of such a project).

Box 6. Ethics oversight of a voice recognition diagnostic tool used to identify depression and related disorders

Researchers designed a study targeting male and female adolescents (aged between 13 and 18 years), with a total research population of 25 participants diagnosed with depression and related disorders and undergoing treatment at a paediatric psychiatry clinic. The research team's hypothesis was that detectable differences could be found in voice tone, and that these differences correlated with depression systems. Therefore, depression and suicidal thoughts could be identified through voice samples.

The research team's objective was to explore the potential of voice samples as early markers for depression and other psychiatric mood disorders in adolescents. Early detection of these symptoms could aid in preventing the progression of psychiatric disorders and reducing suicide risk.

A company was assigned to conduct voice analysis using machine learning algorithms to identify associations between voice parameters and psychiatric diagnoses. Participants would be asked to provide three types of voice samples: (a) neutral reading – or reading a preselected neutral text; (b) structured interaction – or reciting approximately 20 sentences and engaging in a two-minute conversation with the interviewer; and (c) personal expression – or describing their day or discussing a topic of interest.

The analysis would focus on short voice segments, particularly low-frequency passages lasting several dozen milliseconds. The parameters to be evaluated included frequency, amplitude and voice quality. Additionally, whole-signal analysis would be performed, incorporating semantic and linguistic elements. With respect to data management, the researchers proposed that the voice samples be anonymized before being shared with the company. The company would not have access to any identifying participant information.

The REC raised and discussed the following ethical issues.

Outcome and ownership

- What will be the specific outcomes of this research?
- Who will hold the intellectual property rights to the diagnostic tool?

Data management and access

- What will happen to the voice samples after the algorithm training is completed?
- Who will have access to the voice samples, and for how long?
- Will the algorithm retain data enabling indirect identification of participants?

Collaboration terms

- What are the parameters of the collaboration between university researchers and the company?
- Is the company restricted in its use of the research results, even if the voice samples are not reused?

Additional issues the REC could have discussed include the following:

- What evidence is there that the proposed study was scientifically valid? How did the researchers formulate the proposed hypothesis?
- What safeguards were put in place to assure informed consent of a particularly vulnerable adolescent population receiving mental health treatment in a clinical setting?
- Did the target population understand how the voice recognition tool was designed and how it could diagnose depression, and any of the risks associated with such an AI-based tool?
- To what extent are the research results generalizable beyond the study population? Is the sample size adequate for power calculations? What are the limitations of the research results, whether positive or negative?

Safety and cybersecurity

AI tools and technologies pose several potential safety risks to patients. While these may be assessed during regulatory review, an REC or oversight mechanism might play an important role in identifying risks up front. Errors in AI systems, including incorrect recommendations (for example, which drug to use, which of two sick patients to treat, misdiagnosis, oversimplification) and recommendations based on false negative or false positive results, can cause injury to a patient or a group of people with the same health condition *(7)*. Model resilience, or how an AI technology performs over time, is a related risk, though measuring model resilience may not be possible for an REC or other oversight mechanism (unless testing has already been conducted by a researcher and disclosed to an REC).

Safety concerns may be heightened because of the ability of companies, in a relatively short period of time, to develop, test and deploy new AI tools without adequate oversight. For example, companies that have constructed databases of electronic health records are able to utilize these platforms to develop, test and rapidly introduce new AI tools over their platforms. One of the largest electronic health record vendors (electronic health record company) announced in early 2024 that over a 12-month period the company had 60 use cases of AI for clinical care in development. The company noted that:

> Our key focus, whether it's in the traditional predictive analytics space or in the generative AI space, has been putting the technology, in this case AI, into the workflow directly so that it's running on the latest information and is able to help drive forward good decision making … The reason why you see 60-plus use cases from the clinical to the bedside to the back office is because we've integrated generative AI into our core development processes. It's just part of the standard approach, where predictive analytics tended to be very point solution [sic] and it just wasn't a part of how a developer thinks *(71)*.

In partnership with another company, the electronic health record company also introduced the other company's "ambient" AI technology to automatically draft clinical notes in seconds after a patient visit, with over 150 000 notes drafted in a 12-month period. While these uses of AI could have significant benefits for both providers and patients, they also raise concerns as to whether these tools should be introduced rapidly, whether there was research to test the safety, efficacy and utility of the tools, and whether there had been adequate regulatory and ethical oversight of research on the tools prior to their introduction. If steps were or were not taken prior to introduction, liability laws may be the only way to manage the use of these technologies and discourage harmful practices.

These concerns arose in part because previously the electronic health record company had rapidly released a proprietary AI-based sepsis prediction model that it had developed internally with use of its own patient records for training and testing. Yet subsequent independent research, contested by the company, found that the predictive tool did not identify two thirds of patients with sepsis, even though it sent out a sepsis alert for nearly one out of five hospitalized patients *(72, 73)*. This led to alert fatigue, thereby undermining the effectiveness and utility of the system *(73)*. Subsequently, with evidence mounting of problems with the initial model, the company released a new version that changed the data variables, the definition (marker) of sepsis onset, and the guidance it provided to health systems to fine-tune the system to local needs *(74)*. The problem with the original algorithm means that the product was used across many health systems in the United States to guide decisions for millions of ill patients that may have faced a life-threatening illness *(74)*.

Additionally, cybersecurity threats are a significant challenge and threat for AI technologies. Cybersecurity threats, such as hacking, could lead to manipulation of data used for training an algorithm, thereby changing its performance and recommendations, and presenting a threat to patient safety *(7)*. Breaches of algorithms could also lead to the unauthorized disclosure of patient data, which could undermine the privacy of both research participants and eventually patients who use the algorithm. Certain types of AI, such as LMMs, are especially prone to hacking and other cybersecurity risks that could lead to the disclosure of private data or to manipulation of responses from an LMM *(7)*.

RECs and other oversight mechanisms will need to develop expertise (or identify outside sources of expertise) to assess safety and cybersecurity risks during the research phase of AI technology and the possibility of problems arising once a technology has been approved for use. This will also require disclosure, by researchers, of known or expected safety and cybersecurity risks of such technologies.

Hallucination and errors associated with the use of an AI technology

One concern with LMMs has been the propensity of chatbots to produce incorrect or wholly false responses from data or information (such as references) "invented" by the LMM and responses that are biased in ways that replicate flaws encoded in training data, or specific biases. Hallucinations and errors, when directed to either health professionals or experts, can lead to negative consequences, especially if either a health worker or a patient suspends their judgement (automation bias) *(7)*.

Automation bias associated with use of an AI-based technology

With automation bias, a clinician may overlook errors that should have been spotted by human-guided decision-making. While physicians must be able to trust an algorithm, they should not ignore their own expertise and judgement and simply rubber-stamp the recommendation of a machine *(7)*. This concern is especially heightened with the advent of LMMs because they generate false, inaccurate or biased responses. LMMs are also likely to encourage automation bias in experts and health care professionals, and patients that use LMMs have neither the expertise nor the judgement to recognize that the LMM is providing misleading or false information *(7)*. To avoid automation bias, research projects themselves may need to examine how best a tool can be deployed to avoid misuse. This was a topic contemplated by an REC with respect to a diagnostic tool for the analysis and interpretation of microbiological diagnoses (Box 6).

Box 7. Ethics oversight of a microorganism recognition tool

A research project proposed to utilize an existing digital database of de-identified images of microscopic preparations of various human microorganisms (bacteria and fungi). The database was created over the years for educational purposes, using samples from patients diagnosed at a public university. The research team's hypothesis was that AI tools could rapidly and accurately diagnose microbiological samples based on microscopic images.

The research team's primary objective was to develop a diagnostic tool for the analysis and interpretation of microbiological diagnoses. While numerous AI-based diagnostic tools exist, none currently focuses on microbiological diagnoses derived from microscopic preparations. In addition to the core diagnostic algorithm, the tool was expected to include a user-friendly interface to enhance its clinical utility.

The database consisted of 25 000 high-quality images of various microorganisms. Each image was accompanied by detailed metadata, including (a) the biological material from which the microorganism was isolated; (b) the specific type of microorganism depicted; and (c) information about how the microorganism's colony was established.

The images were to be classified and aggregated using AI algorithms. These algorithms would undergo validation in clinical settings, though no clinical decisions would be made based on the tool during the research phase.

The database of images was securely stored on the computers of the microbiology department and remains on site. The diagnostic tool would, depending on the outcomes of the research, eventually be utilized and potentially commercialized by the university's patents and innovations department.

The REC raised and discussed the following ethical issues.

Clinical safety and standards

- How can the research team ensure that no clinical assessments are made based on the tool during the research phase?
- What safety standards and verification protocols should be in place for the tool's future clinical applications?

Additional issues the REC could have discussed include the following.

- How will the university commercialize the diagnostic test and what steps will be taken to ensure it is accessible?
- Can the algorithm be used for all patient populations, or are there limits to its generalizability?
- What additional steps would health providers have to take, if at all, to confirm a diagnosis and to avoid automation bias?

Each of these risks, such as the training and validation of an AI tool with biased data sets, should necessitate at a minimum review and discussion when the research is subject to oversight by an REC. However, RECs will face several challenges with assessing these risks (see section 3 below). They may have neither the expertise nor the training (nor an appropriate framework to even account for such risks); the risks and benefits may be difficult for RECs to measure; and the actual risks and benefits of AI technologies may not be evident at the time of REC review. Therefore, even though there is a clear need to account for the risks, RECs may be unable to do so.

2.3.4 Which ethical requirements are challenged by health-related research on AI tools and technologies?

Informed consent

RECs or other oversight mechanisms should be able to examine both what information will be provided to research participants and whether the information will allow research participants to make an informed decision. Informed consent should always require the disclosure of the central role of AI using clear language appropriate for a lay audience. First, researchers may need to disclose possible risks and benefits associated with the use of specific AI technology that would affect a research participant, as well as any unexpected outcomes that may arise with the use of AI. For example, informed consent may require additional information about a specific algorithm (or that type of AI generally), such as the propensity of an LMM to hallucinate (or its known or expected error rate). One difficulty with disclosure of these risks is that they may not materialize until after the research is completed. Thus, both a review of research and the actual informed consent of research participants may happen too early in the research cycle *(5)*.

Second, researchers may need to disclose basic information explaining the scientific validity of the algorithm, including the type of AI model that underpins an AI system, and how the AI model generally functions (even if a researcher cannot explain how a specific algorithm that is being tested works). This may be challenging, since it must be provided to a research participant in understandable language.

Third, researchers may need to disclose any third parties involved in the development of the algorithm, and whether third parties will have access to any of the data or if the models under study will be trained with data produced during a research study. It is common for AI tools that are being tested to have been developed by commercial companies or utilize a platform developed by a technology company.

Fourth, informed consent requires that researchers disclose and seek consent that data generated through a research study could be used beyond the research study itself, for example to be provided to a third party or to train an AI model *(75)*. This includes seeking consent for subsequent use of data either inputted into or produced by a generative AI model prior to initial conduct of a research study with a generative AI model, since those data cannot be erased and could be disclosed and used in ways that a researcher may not be able to control. Finally, researchers may need to inform participants that for some AI algorithms (such as LMMs or generative AI) data supplied to or produced from a research study cannot be erased. This risk could be disclosed to participants before they consent to a research study, including that LMMs can disclose sensitive information to the public or could be used to train an algorithm or future unspecified algorithm. For some other forms of AI where erasure is possible (such as rule-based AI), researchers could emphasize that participants have the freedom to withdraw from a research study at any time and that a participant can request that their data are erased.

Informed consent may not necessarily require that the research participant is informed about exactly how the algorithm arrives at decisions (or explainability of an algorithm). Many medicines and vaccines that are tested through clinical trials have a mechanism of action that may not be fully explainable to either the researcher or participant, yet this does not make the research unethical. At the same time, even if an algorithm is explainable, it may not necessarily have any material impact on the decision-making of a research participant.

Value of research

RECs and other oversight mechanisms should carefully assess the purported value of research on AI tools and technologies, for several reasons. First, a great deal of hype has been attached to AI, with overstatement of what AI can accomplish and unrealistic estimates of what can be achieved. This has often led to charges of "junk science" (for example during the SARS-CoV-2 pandemic), and some screening tools developed for rapid diagnosis have been described as "utter junk", with companies "trying to capitalize on panic and anxiety" *(7)*.

Second, oversight mechanisms may need to determine whether research is valuable when known error rates for specific types of AI – such as LMMs – are high. Hallucinations are not an exception in many LMMs but a common feature *(8)*. Thus, an REC or other oversight mechanism might have to ask whether known tendencies of LMMs to hallucinate and make serious errors are acceptable, and, for all forms of AI, to determine if the AI algorithm has technical merit.

Third, while an REC or other oversight mechanism may be able to measure clinical benefits, it may struggle to define and quantify the social value of an AI intervention *(75)*. This may especially be the case in countries with limited resources and multiple bottlenecks to the provision of effective clinical care. Just because technology can produce a correct outcome does not mean it will translate into actual and sustainable clinical benefits in a health system that is not well equipped to make use of the tool (for example, due to a lack of information and communication technology infrastructure). Thus, researchers could be expected to not just generate information about a technology's clinical effectiveness but also indicate under what conditions a tool will deliver a benefit that would justify a ministry of health diverting scarce resources to scale up its use. Such actions could include determining whether introducing a new technology would be valuable if the technology itself evolves rapidly (as is often the case with AI-based technologies), in which case the introduction of a technology, which can require significant investment, could be wasteful if that technology were quickly replaced with a new version.

Scientific validity and integrity and clinical equipoise

The CIOMS guidance notes the importance of the scientific validity and integrity of research, stating that: "Scientifically unsound research involving humans is unethical in that it may expose them to risk or inconvenience for no purpose" *(3)*. This highlights the importance of scientific review of AI-related health research even prior to ethics review. There are several challenges with health research on AI tools and technologies that may call validity into question.

First, research on an AI tool may often be conducted virtually. While researchers may find that research conducted virtually is low cost and easy to implement, they may also find that it is difficult to measure the efficacy of an algorithm and the extent to which an algorithm was responsible for the measured outcome (and the durability of the result). To be able to both establish a baseline measure and conduct effective follow-up would create problems with respect to anonymization, privacy and data minimization.

Second, metrics used by programmers to measure an algorithm's performance in a research setting may not be a sufficient predictor of outcomes in a real-world (clinical) environment, although early-stage studies, much as a phase 1 study for an investigational compound, could at least be an indicator that a research study should proceed to additional research studies that then measure an algorithm's performance in a real-world environment. Given the difficulty with demonstrating the scientific validity of an AI tool or technology, researchers and RECs should keep informed of appropriate methods to evaluate AI tools.

Third, an AI model that is tested on limited data (or data predominantly on one gender or ethnicity) may not be relevant or generalizable to subpopulations for which the algorithm is not tested, which means that outcomes may only be possible to "selectively deploy" to those populations that are represented in the data used to train an algorithm *(76)*. As far as possible, data used to train an algorithm should be diverse. For specific diseases that are geographically unique or are only known to affect a specific population, training a relevant AI tool with limited data can be acceptable even if it means that such research outcomes may only be possible to selectively deploy to those populations that are represented in the training data. For those AI technologies that have the potential to be widely useful, there could be progressive efforts to train AI tools on diverse data; until this is possible, researchers should be transparent as to the limitations of such research studies and populations for which such an AI tool may be neither effective nor useful. In fact, for randomized controlled trials on AI interventions, measures of success based on individual outcomes may be less appropriate than outcomes for groups or populations tested with an AI intervention *(75)*.

Fourth, it may be difficult to compare an AI intervention with the standard of care, which may differ across clinical settings *(75)*. These requirements mean that, to the extent possible, ethical oversight should not just examine the inputs (especially data) to train an algorithm but also the AI-generated outputs to ensure compliance with ethical obligations. There are different problems with research conducted with generative AI (LMMs), including whether research conducted with LMMs can be reproduced. LMMs exhibit prompt sensitivity, and results of a query to an LMM can be highly dependent on how a researcher or research subject prompts an LMM *(8)*. This raises the question as to whether the result provided by an LMM is "intrinsic to the model" or is an artefact of a prompt. This also means that certain qualities that may be inferred from research conducted on an LMM, or its biases, may not be because of the model but because of how research participants pose questions or queries to the model *(77)*. There are also concerns with "contamination" – namely, is the performance of an LMM (for example, to answer a medical query) genuine or because the LMM has been trained on certain

questions and answers (recalling information previously fed into the algorithm, and not generating a novel response) *(77)*?

A separate concern is whether there is clinical equipoise with respect to a proposed AI intervention, or sufficient dispute or uncertainty as to whether the proposed intervention is better than a standard intervention or placebo. Since many AI models are built on retrospective data, this may not necessarily mean that the algorithm will perform better than existing AI models used for the same purpose or non-AI-based interventions.

As noted above, AI models are built on retrospective data. In retrospective studies with machine learning to predict outcomes in a single analysis, the algorithm can perform better than but might not be very different from a logistic regression model. This may mean that a novel AI model may not necessarily perform better than the standard of care. Such studies also do not address real-world experience in a treatment setting (where data are continually being added to a patient's case and AI analysis would have to be made multiple times) and impact on the outcome (how long before the outcome can the prediction be made correctly and whether, knowing the risk of the undesirable outcome, it can be avoided). For example, a randomized clinical trial achieved success in predicting sepsis three hours before the placebo group, but mortality and length of hospital stay remained the same *(78)*. Additional evidence or information, beyond retrospective testing with existing data sets, may be necessary before approving research on an AI model. RECs may wish to require researchers to produce additional evidence or information, for example through a proof-of-concept study, before approving research into new AI models.

Fair participant selection

As with data science studies that employ AI, there are also concerns with regard to fair participant selection for studies conducted on AI technologies for use in health. Fair participant selection requires that those selected for the study are based on a study's scientific objectives and not based on unrelated factors such as vulnerability or privilege. There is a risk that researchers select trial participants based on efficiency – that is, research subjects will be those that have easier access to AI-based services, such as those who both have personal resources and are not negatively impacted by the digital divide. Researchers may also select trial participants – intentionally or not – that conform to the data that were used to train the algorithm. This is thus more likely to produce successful outcomes. To ensure fair subject selection, research participants could reflect the varied demographics of a target population on account of (for example) sex, ethnicity and age. Research participants might not solely be selected on the basis of their conformity with the data that were used to train the algorithm. Constructing a study with diverse research participants may ensure that researchers identify biases encoded in the AI model.

Researchers who wish to ensure that the testing (and training) of an algorithm is inclusive may have to balance these efforts against the preferences of specific communities to avoid inclusion in research or testing, for example Indigenous communities or people living with a disease *(40)*. Finally, if an algorithm is used to select participants, this may lead to concerns that the selection it makes, based on how the algorithm is trained, could also display bias or is guided by considerations that are beyond a study's scientific objectives. To determine whether researchers have employed appropriate criteria, RECs will need appropriate methods to measure whether the selection of a research population is based on fairness only.

Privacy and confidentiality

Research on AI-based technology could involve the collection and analysis of data for the algorithm to produce a recommendation or output. RECs may be increasingly required to contemplate unique sources of data and novel applications of AI, and whether rights to privacy and confidentiality have been respected for the data that are collected or the outputs of an AI technology. One example – described in the case study presented in Box 7 – involves inputting the data of deceased individuals to build an AI model for facial reconstruction.

Box 8. Ethics oversight of an AI model for facial reconstruction based on DNA samples

The study involved 400 deceased individuals for whom autopsies were required either by legal mandate (prosecutor's orders) or for medical purposes. The primary objective of research was to develop a predictive model for reconstructing human facial features based on DNA, that could be used (for example) in forensic medicine. Since many facial characteristics are heritable, it was hypothesized that DNA could provide a reliable basis for reconstruction. However, the researchers noted that environmental, biogeographical and gender-specific factors must also be accounted for. The models, if proven accurate, were intended for applications in anthropological research and forensic practice. The study employed two methods of facial phenotyping: (a) 3D scanning; and (b) computed tomography scanning. Additional factors, including body mass index, age and gender, were to be integrated into the analysis. Whole genome sequencing would be conducted using the MethylationEPIC platform.

The predictive modelling process included (a) generating 3D facial representations using MeshMonk software; (b) performing advanced 3D image analysis using R software; and (c) combining these facial data points with genomic data to create a comprehensive predictive model. The results would be validated using an independent data set collected from living donors, which included facial and blood data. All data would be securely stored on university servers with restricted access. The REC raised and discussed the following ethical issues.

Consent and legal permissions

- Are there any ethical or legal barriers to using biological material (faces and blood) from deceased individuals?

Future applications and commercialization

- What are the potential future uses of the predictive algorithm?
- Are there any plans for the commercialization of this technology?

Many algorithms that could be the focus of such research will be commercial algorithms, including LMMs. Use of LMMs by patients and laypersons may not be private and may not respect the confidentiality of personal and health information that they share. Users of LMMs for other purposes have tended to share sensitive information, such as company proprietary information. Data that are shared through an LMM do not necessarily disappear, as companies may use them to improve their AI models, even though there may be no legal basis for doing so (though the data may eventually be removed from company servers). A related problem is sharing of information on an LMM with other users of the LMM, whether because the other user specifically requests the LMM to disclose the information or, in the case of one LMM, due to mistaken disclosure of other people's chat histories (even if not the substance of their conversations). Thus, if a person's identifiable medical information is fed into an LMM, it could be disclosed to third parties. This not only undermines a patient's privacy and confidentiality, it could also run afoul of data protection laws *(8)*.

Benefit sharing and future equitable access

One question for research ethics review to consider is whether the population that will bear the risks of participating in the research is likely to benefit. First, knowledge that the research is designed to yield should be communicated by researchers in simple and straightforward language, including the objectives of the research and the outcomes of the research. Second, no group that participates in the research should be deprived of access to research and outcomes, including any new products and technologies, at an affordable price, as well as future discoveries with clinical salience. Researchers and their institutions could also satisfy ethical requirements for benefit sharing by making appropriate investments in training, capacity-building and resources for the development and use of AI tools and technologies appropriate for research (see section 4).

Yet even if benefit sharing may be achievable for study populations, AI technologies may not necessarily be widely useful, including for the reasons described above – that is, the forms of bias that may be encoded in the data, design, or the contexts for which an AI system was designed for use (or not). There are other reasons that an AI system may not be beneficial, including for those populations who participated in such research.

First, AI systems that require human input and data may not be well suited to all populations. This may be because an algorithm is not well suited to certain types of biometric data (such as voice or skin pigments); additionally, such forms of biometric data may not be consistent with a country's laws or policies or may unnecessarily expose populations to privacy risks for use of such AI.

Second, the affordability of an LMM, the languages for which it can be used optimally, and the unavailability of follow-up treatment based on a recommendation of an AI technology could each undermine equitable sharing of the benefits of AI technologies. Examples include the following.

- AI technologies, and especially LMMs, are available only by paying a fee or subscription, as both developing and operating an LMM can be expensive. It has been estimated that ChatGPT (a well known and commercially popular LMM) costs US$ 700 000 per day to operate. Some companies are introducing subscription fees for new versions of LMMs, which could make certain LMMs unaffordable, not just in low- and middle-income countries, but also for individuals, health systems or local governments in resource-poor settings in high-income countries *(8)*. For example, the newest version of ChatGPT was offered at US$ 200 per month in late 2025. This means that an estimated one half of the world's population earns a monthly income that is less than the per month cost of the newest version of the algorithm, while many other people would not be able to pay such a monthly fee without having to sacrifice most or all necessities.
- Most LMMs, at present, operate well only in English. Thus, while they can receive input and provide outputs in other languages, they are more likely to generate false information or misinformation *(8)*.
- Even if AI technology may be useful for a specific population, the benefits an AI technology may generate, such as improved screening and diagnosis, may not be complemented by improvements in access to treatment *(75)*.

RECs and other oversight mechanisms could weigh barriers to equitable access when evaluating risks and benefits of AI technologies. These barriers include (a) discriminatory practices during the development of the algorithm, such as biases encoded in the data; (b) other limitations of such data that prevent the wide use of the algorithm; (c) failure of certain types of algorithms to operate well for certain populations, due, for example, to the types of biometric data that may need to be inputted, the digital infrastructure required for its use, or a developer not maintaining the performance of the algorithm for its long-term use in a specific setting; and (d) lack of affordability or appropriateness of the algorithm for a certain population.

Finally, even if an AI tool or technology that was the focus of AI-related health research is commercialized or acquired by a private sector actor, benefit sharing and post-trial access requirements should be satisfied by private sector actors even if they did not originally carry out or participate in the research.

3. Challenges for research ethics committees

RECs may not review all types of AI-related health research due to limitations of their mandate. Nevertheless, RECs will continue to play a critical role in the effective oversight and ethical review of AI-related health research. Therefore, it is critical to examine how RECs can be strengthened to review AI-related health research that falls within their remit.

This section examines two challenges that RECs face with respect to oversight of AI-related health research: (a) the current operating model of RECs is not yet well suited to how, when and where AI-related health research is carried out; and (b) RECs have not been provided with sufficient expertise, training, capacity and support to oversee AI-related health research. Based on the challenges identified in this section, several considerations are

included in the "Key considerations" section of this report (see section 9).

3.1 Why is the operating model of RECs not fit for purpose?

The REC operating model has evolved to oversee traditional health-related research that used neither AI nor data science methods. This approach is not well suited to how, when and where AI-based research is carried out for three reasons.

1. *RECs examine research too early during the research process.* Risks associated with the conduct of AI-related health research often materialize during the latter stages of research or emerge as broader social impacts once applied outside a research setting.
2. *RECs are decentralized and find it difficult to coordinate for coherent oversight of multicountry research.* AI-related health research often involves multiple partners or institutions across many settings and countries. This is possible in part because certain types of AI-related health research can be carried out through online channels, and, for some research projects, there may not even be trial sites as research is conducted, for example, through social media platforms. Yet multiple RECs may issue conflicting recommendations for multicountry or multi-site research.
3. *Private companies may not submit AI-related health research for ethics review.* The private sector is conducting AI-related health research and may either not submit their research to RECs or may utilize corporate RECs (third party or in-house), which can be a poor substitute. Furthermore, RECs may not be prepared or are unable to address concerns related to the potential commercialization of corporate research, and whether benefits will be distributed fairly.

The following subsections provide further information on those factors.

3.1.1 RECs examine research too early

As per international guidelines, RECs must examine research ex ante – or before research takes place. Continuing review, which should be conducted by RECs at regular intervals while the study is ongoing, is often neglected or not carried out due to the lack of resources. While continuing review is important for most studies, it is of particular importance for AI-related health research, as often ethical concerns only surface during the conduct of the research project. For example, AI and data science research often relies upon identifying secondary uses of large data sets, via novel hypotheses, during the latter stages of the research process. Even if the secondary uses of the data sets are lawful (for example, it is a publicly available data set), this could still expose data donors to certain ethical risks that were not known to the REC when it reviewed the research, and therefore may not be properly scrutinized *(23)*. Certain risks associated with AI-related health research also may not surface until the later stages of research (such as bias). These late-stage risks may also be difficult for RECs to identify due to a lack of expertise of REC members *(5, 25)*.

3.1.2 RECs are decentralized and may not effectively coordinate and provide oversight of multicountry research

RECs are ill suited to review research carried out across multiple sites or institutions, as is commonplace for AI-related health research *(5)*. RECs, within or across countries, are often not governed by one centralized body. While multicentre pharmaceutical trials have increased over recent decades, creating their own governance challenges, AI-related research is taking this to a new level, as often there may not even be any centres or trial sites at which ethics review could be required. It may be straightforward for researchers to conduct AI-related health research across multiple countries, including through a collaboration of multiple institutions, the use of research companies (as contractors) and crowd workers to collect data, and the use of data sets collected from multiple countries. An emerging area of AI-related health research is multicountry research using social media platforms. This research can be conducted by companies – often in partnership with academic institutions or other research partners – in many countries simultaneously to test a diverse, varied group of research participants rapidly and efficiently.

It may be desirable for researchers to use data sets from multiple countries (or conduct research with an AI algorithm in multiple countries) to avoid bias and improve accuracy. Researchers could, pursuant to Article 23 of the Declaration of Helsinki, have the research protocol approved by RECs in both the sponsoring and host countries. Yet multicountry AI research requiring review by distinct RECs could produce inconsistent outcomes or may require reliance on RECs that have varying levels of expertise and resources *(5)*. This inconsistency can undermine the authority of RECs and create challenges for researchers, who must potentially conform to conflicting recommendations based on legal requirements and ethical norms that are different across diverse jurisdictions *(5)*.

A different challenge with social media research is that, even if a company or research partnership seeks approval for the research in the country or countries they are based in, they may not have any host country research institutions that participate in the study. This is in part because a research institution may not be required to conduct the research, and in fact could only complicate the administration of a study. Yet it also raises the question: by whom and how should such a protocol be reviewed in each country where social media research is carried out? Furthermore, research review in every country where the research is conducted may not be feasible, either for the research team or for the countries of study. There may, however, be situations in which comprehensive research review is necessary.

Even though multicountry research poses challenges to REC oversight, there are examples of RECs successfully coordinating ethics review. Countries can, through appropriate bilateral or multilateral cooperation, implement joint review processes. For example, the United States Food and Drug Administration issued guidance for researchers conducting "multicentre clinical research" to rely upon a centralized REC review *(79)*. However, if multicentre research involves sites in different countries, the standards for each country may also need to be aligned and increasingly harmonized.

3.1.3 Private companies may not submit AI-related health research for ethics review

AI-related health research is increasingly carried out by private companies. Corporate research projects are often not subject to the scrutiny of RECs, for various reasons: because companies are not accustomed to submitting research to RECs; because companies, concerned about disclosure of their intellectual property, do not wish to disclose such data publicly *(5)*; or because the research does not qualify as human research (including the use of data scraped from social media sites). Furthermore, while RECs are usually associated with specific research institutions (such as hospitals, universities or health care systems), companies and corporate researchers do not have formal links to an REC, and corporate researchers, unlike physicians, do not have formal obligations to patients. Instead, in some cases, companies are either delegating such oversight to internal review – nearly every large technology company has an ethics committee *(80)* or is outsourcing review to third-party RECs (sometimes for-profit RECs). This may be preferable for the private sector because the speed at which corporate research is carried out may be inconsistent with the time and timelines of RECs.

A separate but related issue involves the growing trend of public–private partnerships carrying out AI-related health research. Public–private research can lead to concerns around who benefits from research – especially if it has both public interest and profit-driven objectives. There is also a concern as to how data and outcomes that may be generated from a public–private collaboration are shared, and whether data can be used in furtherance of commercial objectives (when research participants may have thought that data were intended for a public interest objective) *(5)*. A final concern is that a public–private partnership might avoid ethics review if the research is conducted solely by the private partner, which chooses not to submit the research for ethics review.

RECs may need to consider the growing commercialization of health research, and the role of the private sector in both conducting health research and deciding how it may use the results. This can extend to public–private partnerships, public research for which the results could be handed over to the private sector, and private sector research that may be reviewed by an REC. Commercial research, as discussed above, may respond more to the specific requirements of shareholders and corporate owners to identify new ways to maximize profits rather than to the public interest aspect of the research. Company scientists may not have adequate grounding in research ethics principles, and there may be no mechanism within a company to hold research teams accountable. This is especially since many large technology companies have disbanded or deprioritized their AI ethics teams over the last few years *(81)*.

Much corporate research, even if carried out to improve a product or service, may not need to meet commonly accepted standards of transparency and accountability, since funding and oversight of the research can be carried out entirely within a company. Furthermore, data protection laws do not necessarily distinguish between public interest research and private research, with both forms of research receiving exemptions from specific legal obligations *(18)*.

Thus, while commercialization may offer certain benefits, there are critical questions that RECs must consider with respect to ownership of both the research inputs and the research outputs, and whether research outcomes, specific insights or information, or validation of an AI tool or system will be shared widely (for example, a fair distribution of benefits) *(38)*, especially for those individuals whose data contributed to the research. RECs may also wish to examine how and where a company may deploy a technology. Even as RECs may need to assume greater oversight of and responsibility for commercial research, they may not have the capacity to carry out oversight effectively and could be quickly overwhelmed by many studies.

3.2 Why may RECs not yet have appropriate expertise, training, support and capacity for ethics review of AI-related health research?

RECs may not always have in-house expertise to review AI-related health research. This can be due to a lack of technical expertise on the methods used within AI and data science, as well as a lack of subject matter expertise as to the novel risks and challenges that the research creates *(38)*. RECs that are situated within a research institution or university that is investing heavily in AI-related research may be better placed to review AI-related health research, as opposed to government or state-based RECs in which little or no AI research takes place *(38)*. This lack of expertise may be exacerbated by how AI researchers prepare research protocols for review. There are actors involved in AI-related research that are not usually involved in health research, including large technology companies as well as data and computer scientists. They may present research in a way not attuned to RECs.

Therefore, RECs may not have the expertise to (a) assess risks posed by AI-related health research; and (b) identify what steps can be recommended to mitigate risks – including risks that can be introduced during the research process and those that may emerge with wider use of research, such as downstream social consequences *(5)*. At present, training may not be available to address gaps in the expertise of REC members *(5)*, nor are there resources to consult external experts that can provide ad hoc support to an REC *(38)*. Gaps in expertise and a lack of training are common to RECs globally *(37)*.

Even if RECs had adequate expertise and training, the capacity of RECs to review AI-related research may be limited, as there may be too many studies for RECs to review *(23)*. Many RECs are staffed by individual experts working on a volunteer basis, meaning that REC members, even with the right expertise, may not have the requisite time to appropriately review an application *(5)*. This can lead to RECs engaging in a box-ticking exercise as opposed to a substantive review. RECs may not necessarily communicate well with one another, even within a single institution, as to whether to share expertise and training, or to coordinate review of a single, multicountry research study that requires approvals from multiple RECs *(5)*.

Even as RECs may not have adequate capacity and expertise, they may also lack information from external sources that could improve their ability to conduct an appropriate review. One crucial source of information could be beneficiaries or communities that may either participate in or be affected by AI-related health research, including individuals who support RECs on a permanent basis through their own expertise or carry out outreach to relevant beneficiaries and communities.

Finally, researchers play a crucial role in ensuring RECs can carry out an appropriate review of AI-related health research with information that committees otherwise lack. Unless researchers are transparent and disclose material information to RECs – for example, with respect to risks identified through internal validation (such as the error rate of an LMM), the types of data used to train the algorithm (and the biases of such an algorithm), and any concerns with safety and cybersecurity – RECs may be unable to conduct effective oversight.

4. How can researchers improve adoption of ethical principles?

The responsibility of researchers should not just be a box-ticking exercise to meet ethical obligations but instead should be to apply ethical principles from the initial design of the research through the completion of research, and with respect to long-term benefits and consequences. Researchers must not only consider whether an AI algorithm meets technical specifications, but also understand and determine in which context an algorithm may be appropriate. This section examines the end-to-end roles and responsibilities of researchers to abide by ethical principles, standards and practices. Based on this section, the Expert Group identified several considerations for researchers (see section 9).

4.1 Training and certification

In contrast to the pharmaceutical sector with its long tradition of health research, many AI researchers, whether those in technology companies or computer scientists by training, may not have any relevant background in, experience of, or exposure to the health sector. Researchers may have incentives and values that differ from those of patients, health care providers and health care systems that will be the subjects and beneficiaries of research, but the latter usually have no role in establishing the culture or norms in which products and services are developed with AI *(7)*.

While medicine is guided by the objective of promoting the health and well-being of patients, an AI researcher, especially when working for a company, is ultimately working in the interests of the company to develop a profitable service or product and, in the case of publicly traded companies, for their shareholders. While medical professionals have a long-standing fiduciary relationship with patients, AI researchers, however well intentioned and cognizant of emerging expectations and legal obligations to protect individual privacy, have no fiduciary duty to patients or health care providers. This complicates any attempt by an individual or a company to put the health and well-being of patients first *(6)*. Even if an AI researcher may be well intentioned and aware that there are risks, this does not necessarily translate into identifying, understanding and communicating the risks of a specific study or AI technology.

There are insufficient legal and professional accountability mechanisms to reinforce good-faith efforts of AI researchers to act in accordance with ethical norms for health research, or to even make researchers aware of norms. AI researchers and technology firms may have no effective self-governance mechanisms. AI researchers and computer scientists may also not know of the role that RECs assume within research to both protect research participants and improve the design and outputs of AI-related health research. By contrast, in the medical profession, there are accountability mechanisms to reinforce the fiduciary duty of medical researchers to patients, and these duties are reinforced by sanctions to deter poor practices *(6)*.

Yet failure to address ethical considerations up front, or to engage with RECs and other oversight mechanisms, may expose researchers, their institutions and funders to legal risks and liability. It is reasonable to expect researchers to anticipate and consider the immediate and downstream harm from AI research, and to weigh near-term benefits against future harm (for example, when an algorithm is used in clinical practice). Alongside the immediate liability that researchers could face with respect to harms caused to research participants (RECs could also face liability for harm caused to research participants), there may eventually be claims for liability of researchers for downstream harms that they did not adequately consider and study or did not warn about when disclosing research results.

Thus, to protect research participants, improve the quality of research, and address legal risks that researchers could eventually face, consideration should be given to strengthening the awareness, capacity and engagement of researchers on these core ethical principles and associated laws and policies (such as data protection laws), and identifying and actively addressing ethical issues that RECs and other oversight mechanisms will have to grapple with. This would require more capacity and training to understand and identify the risks and challenges of AI-related health research, as well as the role and function of RECs and other oversight mechanisms. Such

training and capacity-building could be done through several avenues, including professional training at universities or technical schools or through employers. Academic institutions and research organizations should be encouraged to integrate AI ethics into their policies and training programmes comprehensively.

There could also be ethics certification of AI researchers, much as other health professionals or health researchers require certification, to carry out development of AI technologies for health. This could include training in responsible conduct of research, as required by the United States National Institutes of Health *(82)*, with additional training on AI-related health research. WHO has also launched three different on-line training courses: (a) Research ethics online training (V2);[2] (b) Ethics and governance of artificial intelligence for health: WHO guidance;[3] and (c) Integrating ethics and governance into the design of artificial intelligence tools for health (Case study: Cervical cancer screening).[4] Guidelines and regulations could introduce requirements for those who can participate in AI-related health research, which can be introduced through relevant training and accreditation of researchers. Thereafter, institutions and funding agencies could require researchers to provide proof of having completed such training and certification as a prerequisite for funding.

Furthermore, AI researchers (and developers), even if they have improved their understanding and expertise, could be required to include clinicians or the health services in the design of their research, or to even ask health systems that may benefit from AI technologies to connect researchers with clinicians or service providers who can provide critical input. This would help to ensure the validity of AI applications. In situations where AI researchers have not worked with and consulted health systems, a health system or other procurement agency may determine that as a reason to not utilize the outcomes of a research project, or to not approve an AI technology that was the subject of a research study for use.

Finally, in addition to working more closely with clinicians and health services, there are other forms of expertise that could be added to multidisciplinary research teams – including machine learning and AI experts, health communication experts, ethicists, end users, individuals or groups with lived experience, underserved groups and implementation scientists *(83)*. Resources to construct multidisciplinary teams could be provided by funders of AI research if they wish to ensure that their investments and grants are to have a public health benefit (see section 5 below), and so that researchers who are confronted by ethical challenges have the resources in place to address problems with the full backing of the institutions that support their work.

4.2 Early identification of ethical risks and challenges

RECs, as noted above, are often unable to judge risks associated with AI technologies because risks do not materialize until the research has been completed. This problem can be addressed in part by researchers considering and trying to address key ethical risks with AI tools and AI-related health research early in the development of an algorithm.

First, researchers should, according to updated requirements published by standard-setting institutions, intergovernmental agencies, institutions or national RECs, be aware of and try to address possible risks during the design and development process, including those relating to safety and cybersecurity. This requires applying "ethics by design" at the outset to mitigate key risks, such as expected biases, or barriers to affordability and accessibility *(7)*, instead of considering ethics as an afterthought. This also ensures that researchers remain up to date with requirements of standard-setting institutions, intergovernmental agencies, institutions or national RECs.

Second, researchers, as one step, and prior to conducting research or commercial testing, should test out and "red team" errors that may be features of AI tools. "Red teaming" is an evaluation of a model or system that identifies vulnerability in real-world simulations that might result in undesirable behaviour, such as an algorithm providing a biased opinion, or having a specific security flaw, so that a developer can correct the model or system to ensure its reliability and safety. In WHO guidance on the ethics and governance of LMMs, the guidance recommends that: "Developers should ensure that LMMs are designed to perform well-defined tasks with the necessary accuracy and reliability to improve the capacity of health systems and advance patient interests. Developers should also be able to predict and understand potential secondary outcomes." Techniques

2 https://whoacademy.org/coursewares/course-v1:WHOAcademy-Hosted+H0122EN+2025_Q1?source=edX
3 https://whoacademy.org/coursewares/course-v1:WHOAcademy-Hosted+H0061EN+H0061EN_Q4_2024?source=edX
4 https://whoacademy.org/coursewares/course-v1:WHOAcademy-Hosted+H0062EN+H0062EN_Q4_2024?source=edX

to meet such requirements include "premortems" and "red teaming" *(8)*. However, while red teaming may be one means to identify specific risks and improve the accuracy and reliability of algorithms, it is also not a panacea and is not a solution to every risk that AI technologies may introduce *(84)*.

Beyond the proactive measures that can be taken by developers and researchers, RECs and other oversight mechanisms can also pose extra questions that researchers will need to answer, which will then prompt researchers to consider ethical issues that will need to be addressed with downstream uses.

4.2.1 Societal impact of technology

Beyond the specific benefits for and risks to beneficiaries of AI technologies, researchers should also examine and consider the wider social impacts of AI technology when setting out research, including for example its impact on health equity and on the ability of humans to maintain epistemic authority over health care and medicine, and the environmental impacts (carbon and water footprint) of widespread use of an AI technology. These considerations can play a role in how technology is designed, and how research is carried out (including whether use of AI tools to support research is appropriate), and should also be communicated transparently to oversight mechanisms, though oversight mechanisms may not be well placed to contemplate and address such societal impacts compared to policy-makers (see section 5 below). To provide relevant information, researchers could submit a social impact statement with their research proposal when seeking a grant *(5)*. Researchers could also include calculations of estimates of AI energy use and concomitant carbon emissions.

4.2.2 Treatment of data (crowd) workers

Researchers should not assume that a project using crowd workers does not require ethical oversight. Researchers, even if they are unsure of whether their use of crowd workers merits ethical review, should declare that crowd workers are or were used in the conduct of a research project. Where use of crowd workers could require ethical review, research should be submitted to an appropriate REC. This includes implementation research that uses crowd workers, for example, to collect information on the use of an AI tool or technology in a specific context. In these cases, ethical standards for the conduct of implementation research, such as those supported by WHO *(85)* through the planning, implementation and post-research phase, should be applied.

Irrespective of whether research conducted with crowd workers falls within the scope of REC oversight, researchers (as well as their institutions) should uphold a responsibility to ensure appropriate treatment of crowd workers as it relates to labour practices, and other health-related consequences associated with the use of crowd workers. Researchers and their institutions should also ensure that crowd workers, through careful vetting and due diligence of intermediary companies that supply crowd workers as a service, and irrespective of where such crowd workers reside, are provided with labour protections. Governments should update their labour standards to extend benefits to all data workers, to promote a "level playing field" among researchers and companies and to ensure that labour standards are maintained and improved over time *(33, 43)*.

4.2.3 Transparency

Once risks have been identified through practices to improve safety and accuracy, such as red teaming and internal evaluations, as well as consideration of the societal impacts, researchers should be transparent with findings so that RECs can take risks into account in their deliberations, as well as research partners and participants. For research participants, information about risks is important to ensure that they can provide meaningful informed consent.

Transparency should also extend to reporting on clinical trials on AI tools and technologies, including commercial research, that have been registered but for which no results have been published. WHO defines the timely reporting of trial results as publication in an open-access repository within 24 months of trial completion, while some countries may have stricter requirements and shorter time periods to report results. Reporting has improved – by October 2024, 83.4% of all trials in the European Union had published their results, whereas in the United States 77.4% had published their results *(86)*. One concern with AI-related health research is that reporting will not be widespread, especially if researchers and institutions had not previously operated in a health care or medical research setting, or in the case of companies that wish to only present positive results for regulatory and commercial purposes. Furthermore, since AI-related health research often does not apply

traditional trial designs, research may not be registered in traditional trial registries. Nevertheless, there needs to be mechanisms to improve transparency with respect to the existence of research and its findings. Finally, researchers should disclose the use of AI tools and technologies for conducting research, including which AI tools and technologies, their performance, and how AI was applied (for example, the prompts used with generative AI models).

4.2.4 Communication and public engagement

Researchers can also strengthen public and community engagement. This can be to inform and empower communities and the public on what AI research is and its risks and benefits and thus the implications of their involvement, and to improve community engagement in the planning of studies and the design of technologies. It can also help to improve trust and facilitate oversight. Ultimately, effective use of AI for health will require the trust of the public, providers and patients. Social licence requires hard-fought efforts that can be surrendered quickly if AI technologies are introduced without due care for the perspectives of those affected by their use. Public engagement and dialogue are means to ensure that use of AI for health meets certain core societal expectations, leading to greater trust and acceptance. Public dialogue also allows ascertainment of society's views, as far as possible, on the ethical dimensions of AI, its design and uses *(7)*.

As noted in section 6, there are concerns over inequities that persist between researchers in the global North and research partners, communities, health systems and RECs in the global South. High-quality research is only possible if it is adaptable to local contexts and addresses the specific health challenges of the research setting. For research conducted by those in high-income countries in a low- or middle-income country, there should be robust engagement with local stakeholders in the design and development of research.

Researchers must be transparent with respect to the purpose, risks and expected use of AI technologies and consider heightened protections with respect to the collection and use of health data, especially when there may be neither the appropriate standards nor the regulatory and enforcement apparatus to implement protections *(41)*.

Finally, even after research is completed, researchers should also be engaged on the ethical risks and concerns with outputs of their research when they are used beyond a research setting. This is in part because RECs may not have any engagement following their initial review.

4.2.5 Existing guidelines for researchers

Over the last decade, there have been several efforts to standardize and improve the conduct of AI-related health research – both in how research protocols are written and in the reporting of results of randomized clinical trials. The following are brief descriptions of two important consensus guidelines – the SPIRIT-AI extension and the CONSORT-AI extension. Both guidelines play an important role in ensuring that AI-based research both discloses adequate information and meets certain ethical requirements.

- **SPIRIT-AI extension.** The SPIRIT 2013 statement strengthens clinical trial protocols through minimum reporting requirements. The SPIRIT-AI extension "is a new reporting guideline for clinical trial protocols evaluating interventions with an AI component" *(87)*. It includes 15 new items that should be included in clinical trial protocols for the testing of AI technologies. While the SPIRIT 2013 statement includes several requirements related to research ethics, the extension includes only one specific item related to research ethics, or to delineate "whether and how the AI intervention and/or its code can be accessed, including any restrictions to access or re-use" *(87)*.
- **CONSORT-AI extension.** The CONSORT 2010 statement establishes minimum guidelines for reporting results of randomized trials with completeness and transparency. The CONSORT-AI extension provides additional guidance on reporting clinical trials that evaluate interventions with an AI component. The 2010 guidelines include several items that could relate to AI trials, including limitations of the trial (including potential biases), generalizability of the trial findings, and interpretation of the results accounting for the benefits and harms, although the CONSORT-AI extension does not provide additional guidance to researchers. The CONSORT-AI extension includes several specific recommendations that touch on ethics – for example, with respect to harms, the CONSORT-AI extension states that researchers should "describe results of any analysis of performance errors and how errors were identified, where applicable" *(88)*.

Other guidelines have also emerged; a summary of these guidelines, developed by Collins et al., is listed in Table 1.

Table 1. Reporting guidelines for AI-related health research

Reporting guideline	Full name	Scope
STARD-AI	Standards for Reporting of Diagnostic Accuracy	Studies evaluating the diagnostic accuracy of an AI-based test (in preparation)
TRIPOD+AI	Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis	Studies developing or evaluating the performance of a prediction model using AI, including machine learning methods
CLAIM	Checklist for Artificial Intelligence in Medical Imaging	Medical imaging studies using AI
DECIDE-AI	Decisions in health Care to Introduce or Diffuse innovations using Evidence	Early-stage clinical evaluation (including safety, human factors evaluation) of decision support systems driven by AI
CHEERS-AI	Consolidated Health Economic Evaluation Reporting Standards	Studies describing health economic evaluations to estimate the value for money (cost–effectiveness) of AI interventions
SPIRIT-AI	Standard Protocol Items: Recommendations for Interventional Trials	Protocols for clinical trials evaluating an intervention with an AI component
CONSORT-AI	Consolidated Standards of Reporting Trials	Clinical trial reports evaluating an intervention with an AI component
PRISMA-AI	Preferred Reporting Items for Systematic Reviews and Meta-Analyses	Systematic reviews and meta-analyses of AI interventions (in preparation)

Source: Adapted from Collins et al. *(89).*

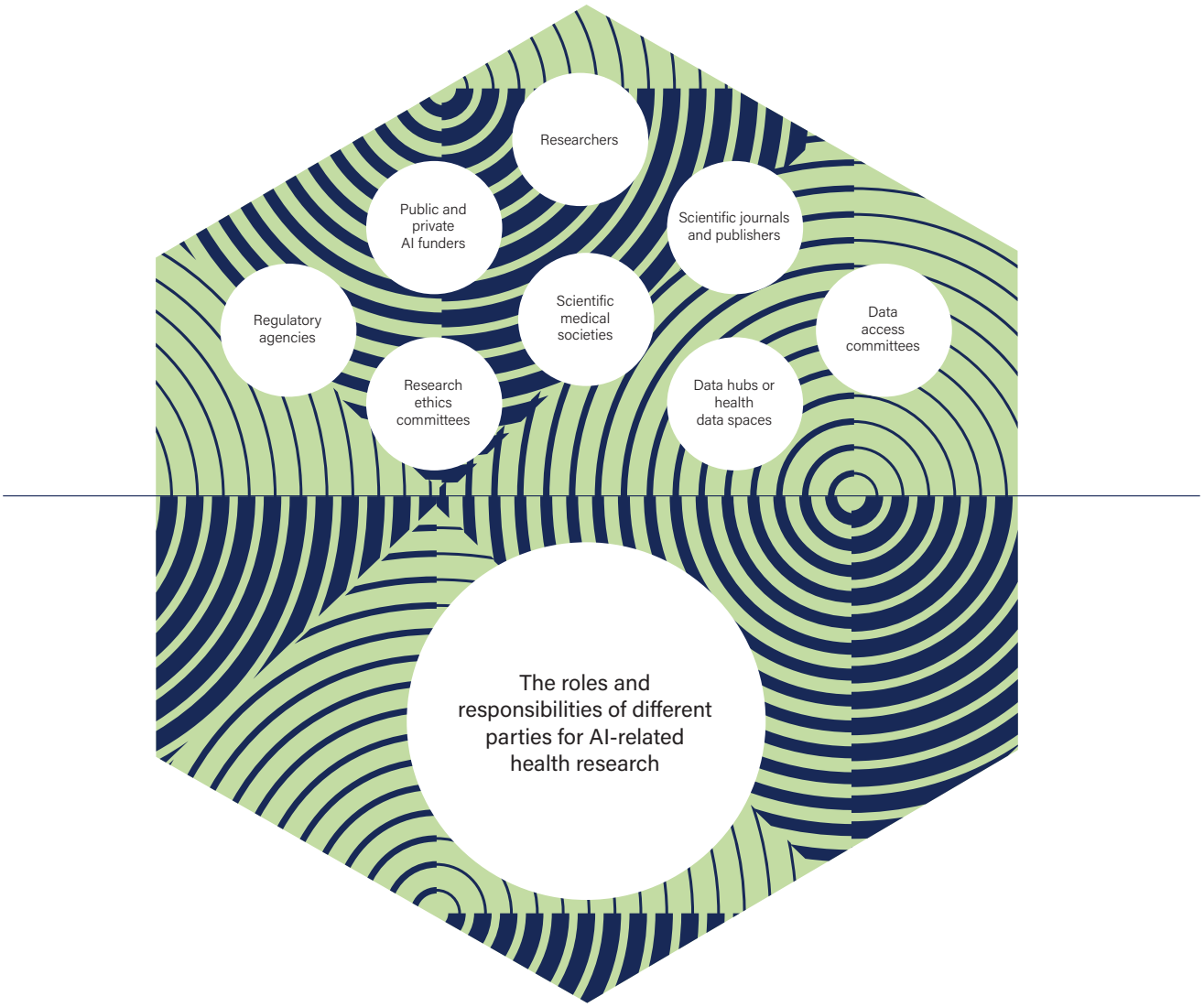
These technical guidelines, and other research practices outlined above, can ensure that researchers play a more central role in reducing ethical risks of AI-related health research and improving the participation and decision-making of other key actors, including RECs, beneficiaries and other parts of the research community. Beyond technical guidelines, researchers should also follow international human rights standards related to the conduct of research, and respect and abide by cultural norms, values and practices where research is situated. Researchers should also disclose additional information, such as other parties involved in the development and training of an AI algorithm or the development of a research protocol, so that RECs and other oversight mechanisms can identify potential conflicts at an early stage and provide additional supervision and feedback.

Ensuring that researchers understand and follow ethical principles for AI-related health research should be a shared responsibility of governments and the institutions where researchers are based. In addition, to ensure researchers are incentivized to conduct research meeting these ethical principles and standards, research funders can encourage or require desired practices when assessing whether to fund research proposals *(5)*. The responsibilities of research funders are discussed below (see section 5).

5. What could be the role of third parties for ethical oversight?

The responsibility to address the many risks of AI-related health research will need to be shared with other parties that participate in oversight. Policy-makers must also consider the broader consequences and impacts of AI-related health research and the implications for health care systems. Since it is likely that RECs will have neither the mandate nor the capacity to review all AI-related health research, these third parties, as oversight mechanisms, can both reinforce the efforts of RECs and also serve as a substitute for those efforts by addressing systemic issues that RECs cannot address on a case-by-case basis. This section considers the roles and responsibilities of various stakeholders to strengthen ethical oversight. Based on this section, the Expert Group identified several considerations for third parties, as described below (see section 9).

Figure 2. Roles and responsibilities for AI-related health research



5.1 Public and private funders of AI research

Funders of AI research can play several roles in strengthening ethical conduct of AI-related research (and of researchers), although not all funders may be prioritizing ethical research but other goals, such as maximizing return on investment.

For all research teams, whether in the global North or global South, funders should consider providing support that facilitates the construction of multidisciplinary research teams that have diverse forms of expertise and perspectives. Furthermore, AI funders should invest in the capacity of those institutions, researchers and countries that may currently lack the necessary infrastructure to support AI research, including data storage, processing capabilities and internet access. Overall, AI research needs to prioritize the sustainable funding of AI research by and within low- and middle-income countries so that research institutions can play a more equitable role in the conduct of AI-related health research. Such funding can also address power inequities between research institutions in the global North and global South.

Public and private funders could establish clear ethical requirements that all supported research should meet and that are in alignment with these guidelines. These should be enforceable, including through funding agreements. This could mean that funders assess applications in part on whether research applicants identify risks and propose means to mitigate or avoid any concerns. Funding agreements could require high-income country researchers to include budget lines for partner-led research in low- and middle-income countries. This may also require funders to maintain a more active role in research after funding is provided, including via reporting requirements of recipients and evaluation of the outcomes of research, and whether it satisfies ethical obligations. Funders should also ensure that sufficient resources are allocated to enable compliance, including for example ethical infrastructure, such as secure data storage as well as REC capacity-building.

5.2 Health data spaces and data hubs

Data hubs pool various types of health data for use by third parties. Several government-sponsored data hubs have emerged. In the United States, two hubs are the Precision Medicine Initiative (All of Us) and the Department of Veteran Affairs health data hub *(7)*. Data spaces can build an ecosystem of health data for which there are common rules, standards and protections. Both data hubs and data spaces can play an important role in facilitating the collection and use of health data to advance scientific research while protecting against the risks to those who contribute their data and the proposed users of data. Data hubs and data spaces customarily rely on the anonymization of data. For those data banks that collect identifiable health data, the Declaration of Taipei sets out four principles for governance: (a) protection of individuals; (b) transparency; (c) participation and inclusion; and (d) accountability *(13)*.

One closely watched effort is the establishment, by the European Union, of a European Health Data Space, which is a "health-specific data sharing framework establishing clear rules, common standards and practices, infrastructures and a governance framework of electronic health data by patients and for research, innovation, policy-making, patient safety, statistics or regulatory purposes" *(90)*. The European Health Data Space has been praised for facilitating and supporting data sharing as a means of encouraging scientific research, including AI-related health research, for the public interest. It has also taken steps to safeguard the rights of individuals under data protection rules (General Data Protection Regulation) by introducing purpose limitations, anonymization (or pseudonymization) requirements, and security-related protections. It also includes an opt-out provision so that individuals whose health data are accessible can refuse uses to which they had not consented *(91)*.

However, the European Health Data Space has also been criticized. There are concerns that information provided to individuals (when determining whether to opt out) is not sufficient for those individuals to exercise informed consent *(92)*. Furthermore, the European Health Data Space could do more to support autonomy by providing individuals with the right to opt in (instead of opt out) *(91)*. There are also concerns that the European Health Data Space does not consider broader social concerns – including the fair distribution of benefits that may result from scientific research with commercial uses, such as the development of new medicines, vaccines and treatments *(91)*. Entities established at the national level – national health data access bodies – that are intended to oversee and sanction data transfers for secondary use of data may have little agency to limit or restrict such transfers of data, since secondary use of patient data is allowed "almost by default" *(92)*. Finally, there will be challenges for health data access bodies to work effectively with data protection authorities and

RECs, and to ensure that research entities do not "forum shop" amongst these entities to validate a practice that requires scrutiny and oversight *(92)*.

And while the European Health Data Space and national data hubs may have adequate resources to operate, many data hubs and data spaces, which could play an important role in contributing to adherence to ethical principles, currently lack adequate resources.

5.3 Data access committees

A data access committee (DAC), according to Cheah and Piasecki, is "a formal or informal group of individuals who have the responsibility of reviewing and assessing data access requests .... Many groups, consortia, institutional and independent DACs have been set up but there is currently no widely accepted framework under which DACs operate" *(93)*. DACs, as separate from or in conjunction with RECs, could also play a role in the governance of AI-related health research, but may require additional resources to play such a role. While RECs "protect research subjects by applying ethical principles and rules of law, DACs should promote data sharing while mitigating any potential risks and should be a mechanism to implement institutional data sharing policies" *(93)*.

DACs can review health data requests to assure that relevant AI-related health research conforms to ethical principles, or, where RECs have already reviewed such research and provided guidance, that data access requests conform to the feedback provided by an REC, as well as principles set out by a DAC. In some cases, where a research project may be exempted from ethics review, thereby creating an ethics oversight gap, a DAC can act as the primary oversight mechanism to scrutinize potentially unethical research practices *(25)*, while also facilitating data sharing for impactful health research.

There are concerns that DACs, like RECs, have certain functional weaknesses that could undermine their effectiveness and capacity to identify ethical harm and consequences. This can include failing to anticipate group harm that may be caused by research conducted on large data sets *(93)*. Such functional weaknesses of DACs can be overcome in part by also consulting outside experts on AI-related health research *(25)*.

Overall, DACs can play an important role in facilitating data sharing and could increasingly assume a leading role in providing ethical oversight to ensure the reviewed research avoids serious ethical risks and challenges.

5.4 Community control of data: data sovereignty and data cooperatives

Measures have been taken to provide discrete communities with control over their data, including health data, through the exercise of data sovereignty or creation of data cooperatives. Several Indigenous communities have sought to establish control over their data through data sovereignty. For example, Māori (the Indigenous population of New Zealand) have introduced principles for data sovereignty that establish, for example, control over data, including to protect against future harm, accountability to the people who provide such data by those who collect, use and disseminate them, an obligation for data to provide a collective benefit, and free prior and informed consent, which, when not obtainable, should be accompanied by stronger governance. Māori also recognizes that the individual rights of data holders should be balanced by benefits for the community and that in some situations the collective rights of the Māori will prevail over those of individuals. Community control can specify the terms on which data can be used for scientific research, and how data outputs benefit those who contributed *(7)*.

First Nations, Métis and Inuit groups in Canada have also each outlined principles for sovereignty over their data. First Nations groups, for example, have introduced principles with four elements: ownership of data, control of data, access to data and possession of data. It is expected that, over time, First Nations peoples will establish protocols to allow wider access to data for uses that benefit them.

A data cooperative gives people who provide data control over their data by storing the data for the members of a cooperative. Data cooperatives allow secondary use of data while allowing members of the cooperative to decide collectively how the data should be used. Data cooperatives allow members to set common ethical standards, and some have developed their own tools and applications to ensure that the data are used beneficially *(7)*.

These different approaches may not be scalable globally and highlight the need for local consideration of the ethics of AI research. Yet they are an opportunity for specific groups of individuals who could either not benefit from scientific research or face possible violations of their privacy and autonomy to assume greater control of their data and to ensure they can be used by researchers for the common good.

Such approaches to managing data for appropriate use, and capturing benefits, may be preferred to compensating individuals for providing their data for research, as has been suggested. While compensation is usually provided for scientific research in exchange for a participant's time and inconvenience, compensation for data that have already been collected or entail little inconvenience may only weaken protections of individual privacy and autonomy in exchange for relatively little financial benefit.

5.5 Role of scientific journals and publishers

Scientific journals, publishers and conference organizers are often the target destination for much scientific research and can have a strong influence on the course of research. This is because researchers are incentivized and rewarded by producing novel and innovative work, evidenced by publications in relevant scientific journals or conferences *(5)*.

Yet there are concerns regarding the effectiveness of these gatekeepers in encouraging ethical health research, including with AI. First, there may be a lack of harmonization amongst academic journals (with each other and with funders and other bodies) on AI ethics *(23)*. This can both encourage researchers to not abide by ethical principles and confuse researchers that must contend with conflicting standards. Second, publishers and organizers may not yet require statements or other information, for example on the broader impact of AI-related health research *(5)*. Third, publishers may not be examining research to ensure that it does not involve "ethics dumping", especially in the case of research that consists of collaborations between institutions and researchers in high- and low-income countries *(41)*.

Currently practices diverge, depending on the research publication. One concern is with pre-print publications, which do not require peer review and therefore may legitimize study findings that may not otherwise be published.

5.6 Scientific medical societies

Scientific medical societies play a key role in promoting high standards in medical education, practice and ethics *(94)*. They also promote the interests of members and the public. During the SARS-CoV-2 pandemic, scientific medical societies were instrumental in presenting trustworthy information about the virus and available vaccines.

Scientific medical societies carry out information campaigns in their area of activity, such as disease awareness, immunization and rational use of resources in health care. Scientific medical societies are also involved in communicating important scientific discoveries to the public in an understandable and simple way. They can discuss and present to the public the ethical issues in health research involving AI, issue criticism of ethically dubious studies, and communicate to the public a careful and correct interpretation of health care studies involving AI and their limitations *(95)*.

Scientific medical societies are a major source of information, education and continuing study for the benefit of health care professionals through events, courses and conferences. They are often a reference for current developments in a field of study and produce guidelines and recommendations for patient management (which may or may not include future AI interventions). Some also manage important scientific journals. Scientific medical societies can play a vital role in bringing ethical research with AI to its educational programmes, and discussing and encouraging high-quality research with AI among their members and in medical conferences. Furthermore, ethics committees can be created by a scientific medical society itself (many already have them) and its members can manage event programmes and officially endorse ethical guidelines for AI-related health research that can be applied to its journals and conferences *(96)*.

5.7 Role of regulatory agencies and other government oversight bodies

Regulatory agencies play an increasingly important role in approving AI-supported health technologies. At present, many regulatory agencies, including those in high-income countries, are developing or updating laws and regulations to guide the regulatory approval of AI-based health technologies. In 2023, WHO published new guidance on *Regulatory considerations on artificial intelligence for health*, which includes six pillars for ensuring appropriate oversight to evaluate AI health technologies for safety and efficacy *(97)*. Regulatory agencies can integrate known ethical risks with research on AI-supported health technology into their own review standards, and specifically examine research studies to ensure that ethical risks identified by an REC have been addressed throughout the research process. A new approach, championed by the Organisation for Economic Co-operation and Development and other institutions, is "anticipatory governance", which enables regulatory agencies to conduct upstream oversight of AI-supported technologies as they are developed. However, many countries do not have regulatory agencies that can regulate AI-related health research and its outputs, or regulatory agencies may be understaffed.

Even in the absence of regulatory agency capacity, countries are establishing other mechanisms to provide appropriate oversight of requests to access health data to train or validate AI algorithms, and to endorse (or reject) proposals to test or implement AI tools within the health system *(98)*. Box 8 provides a description of the approach used in New Zealand, and the government's oversight of a specific AI-supported health technology.

Box 9. New Zealand's national AI governance group and its oversight of a predictive mortality algorithm

In recent years in New Zealand, 20 district health boards were integrated into one national public health system – Health New Zealand | Te Whatu Ora. At this same time, there has been no specific regulation or legislation pertaining to AI or software as a medical device for New Zealand. For these reasons, it was decided to establish a national AI governance group within Health New Zealand. This group reviews proposals to access public health system data for the development or validation of AI tools, and endorses proposals to implement AI tools within the health system. The group also provides general advice to staff, for example on the use of publicly available generative AI models in the workplace. The group is supported by an internal AI laboratory that will work with proposers (generally clinicians or health services) to complete a questionnaire and potentially to test the accuracy and bias of existing tools on Health New Zealand data. The questionnaire is based on a comprehensive checklist developed internally *(99)* building on existing international frameworks with the addition of consideration of consumer and Indigenous Māori perspectives. The group itself reflects the breadth of the eight key perspectives in the checklist: technical, data, legal, ethical, equity, Māori, consumer, and clinician.

New Zealand introduced an End-of-Life Choice Act in 2021 that legalized assisted dying. To be eligible for assisted dying in New Zealand, a person must meet certain criteria, including having a terminal illness that is likely to end their life within six months. As a result, people became interested in whether an AI tool could help identify patients who fitted the time frame criteria. While a predictive algorithm was found to be feasible and potentially accurate, the national AI governance group identified several implementation issues. These included difficult discussions around where and when the information would be made available, who would be able to see it, and the potential unintended consequences of that visibility to anyone involved in the patient's care. Automation bias is always a consideration, and in this case the potential for clinicians to subconsciously become reliant upon this information was also considered a risk.

5.8 Role of policy-makers (ministries of health and intergovernmental agencies)

While RECs and other oversight mechanisms can address many of the specific issues that arise within AI-related health research, they may not be best placed to deal with the systemic trends and issues that often materialize during individual research projects. There are several concerns that could first be addressed by policy-makers – whether by ministries of health, by legislators or through intergovernmental discussions.

Data protection laws. Data protection laws can play a critical role in safeguarding the right to privacy of human participants in scientific research, while also providing clear exceptions to ensure that scientific research can be conducted to promote innovation and advance public interest. In many countries and settings, there may be a need to revisit data protection laws to ensure that they are keeping pace with the rapid evolution of the types and uses of AI, how AI is used for scientific and AI-related health research, and who is conducting such research. This can include revisions to:

- improve the coherence and coordination of data protection laws with international ethics standards, including requirements that scientific research that utilizes health data that are not anonymized should be subject to review by an REC;
- make a clear distinction between commercial and non-commercial research, including that commercial research that utilizes health data that are not anonymized may not receive a research exemption under data protection laws, and may be subject to review by an REC;
- improve the coordination and information sharing between data protection authorities and RECs with respect to scientific research conducted under their authority.

Cross-border transfer of data. While countries may have national data protection laws, they may not have laws governing the cross-border transfer of data, which can lead to research data crossing borders without agreements in place. Governments, either individually or collectively, could take steps to put rules in place to ensure that any such transfer of data for AI-related health research has appropriate safeguards.

Synthetic data. The growing use of synthetic data in medical research, as discussed above (see section 2.2), has both benefits and risks for their use in medical research. Policy-makers could define how synthetic data should be treated under existing laws and policies, such as data protection laws and relevant privacy standards, and define new legal standards and rules for the generation of synthetic data, transparency related to the production and use of synthetic data, and evaluation of such data, including by RECs and other oversight mechanisms.

Labour protections for crowd workers. Crowd workers can face unsafe working conditions due to the nature of the data that they are asked to label, and may also not be compensated fairly for their work. Governments of countries that host crowd workers, which often tend to be low- and middle-income countries, should enforce existing laws to protect the labour rights of crowd workers, or update labour laws to account for the risks and concerns associated with this type of employment.

Societal implications and long-term implications and value of AI-related health research. While RECs do contemplate the social value of research, deliberation may not yet consider the long-term societal implications of AI-related research. RECs may not be well suited to address these broader, systemic questions, such as the broader implications of a type of AI system for the practice of medicine (including consequences such as automation bias or de-skilling of medical professionals), the environmental impacts of ever-growing use of AI and data science (including its carbon and water footprint), and concerns related to the commercialization of health technologies developed through health research. Broader societal impacts associated with AI-related health research may need to be considered more carefully at a higher level. Stanford University's introduction of an ethics and society review as part of the grant-making process for AI-related health research is one means for ethical reflection. Such reflection and oversight could also be conducted by other appropriate entities, including ministries of health or intergovernmental organizations.

These entities may also wish to consider specific mechanisms – such as a research observatory or commission – to monitor the long-term impacts of AI research on health outcomes, especially in low- and middle-income countries, including the benefits and harms of AI technologies that were applied to health research. There should

be feedback mechanisms to allow communities and stakeholders to voice their concerns and experiences with AI technologies to improve future research, oversight and implementation. These mechanisms, and community input, would help to adapt guidelines for AI-related health research over time as AI technologies, their uses, and risks continue to evolve.

A wider effort, going beyond individual research institutions, may also be warranted. Blau et al, *(57)* propose the establishment, in the United States, of a strategic council on the responsible use of AI in science. Such a council would (a) provide guidance on the appropriate uses of AI; (b) study, monitor and address the evolving uses of AI in science; (c) address new ethical and social concerns, including equity; and (d) also identify and address emerging threats to scientific norms. Such a strategic council could also be beneficial in other countries, on a regional basis, or through an appropriate intergovernmental body.

6. What are the specific concerns with AI-related health research in low- and middle-income countries?

Presently, most AI research and technology development is funded, designed and led by researchers in North America, Europe and East Asia *(100)*. AI technologies and research, which may eventually be introduced in low- and middle-income countries, may be marginally relevant in those settings since developers and researchers neither consult nor develop AI technologies with experts, organizations and technologists from low- and middle-income countries. The imbalance between where AI technologies are developed and where they may be tested and deployed, especially in low- and middle-income country contexts, may expose people in those countries to research studies that may be poorly designed for those contexts, risks of exploitative practices, and outcomes that may be neither appropriate nor accessible in those settings.

While all AI-related health research carries risks, these are further exacerbated in low- and middle-income country contexts. Researchers and RECs globally, but especially those situated in high-income countries, should be responsible and held accountable for addressing risks and challenges that are likely to emerge in low- and middle-income countries, and should be encouraged to proactively partner and engage with scientists in low- and middle-income countries in the development of technologies and the design of research. The Declaration of Helsinki also requires, under Article 23, that "when collaborative research is performed internationally, the research protocol must be approved by research ethics committees in both the sponsoring and host countries" *(101)*.

This section examines six issues: (a) fairness and inclusivity of AI-related health research; (b) benefit sharing and access; (c) data colonialism; (d) ethics dumping; (e) redressing power inequities between researchers in high-income countries and those situated in low- and middle-income countries; and (f) encouraging capacity-building. The Expert Group identified several considerations that could be taken into account by researchers, RECs and third-party oversight mechanisms (see section 9).

6.1 Fairness and inclusivity

The development of AI technologies can exclude, avoid or neglect the needs of people in low- and middle-income countries during training, testing, deployment, evaluation and maintenance of models. This raises questions as to whose values are prioritized when an AI system is developed, studied and implemented, and whether such an AI system is appropriate and economically viable for communities in certain settings *(75)*.

Data science studies based on existing data sets may be largely biased against or exclude people in low- and middle-income countries (for example, genomic data sets largely comprise white Europeans) *(7)*. Testing of AI-supported health technologies may be difficult to conduct in settings that have little access to digital health technologies or health care more generally and thereby discourage researchers from investing in research in these settings. If technologies are not to be tested in low- and middle-income countries, the primary concern for RECs or other oversight mechanisms should be that the technologies are not introduced beyond the countries and populations for which they were tested. A focus on high-income countries may be especially likely now since nearly all AI expertise, resources, infrastructure and funding is based within high-income countries.

While there may be legitimate reasons for exclusion, including concerns with collecting data from settings that may amount to exploitation or data colonialism (see below), RECs should be provided with guidance to determine how to minimize exclusion and to ensure that if benefits do not accrue widely, exclusion of low- and middle-income countries should be considered when assessing risks and benefits (especially since AI-related health research is ethically justifiable in part because of the broad public benefit that health research should deliver).

Over time, exclusion of low- and middle-income countries from the benefits of AI-related health research should be discouraged, in part through appropriate oversight. To do so, AI-related health research should be based on diverse data sets. This requires collaboration with low- and middle-income country institutions and researchers for the co-creation of data sets, research and findings. Mere inclusion of low- and middle-income country data sets in research is inadequate. Research teams must also work cooperatively to assess the risks and benefits of such research and whether the research design and outputs are sufficient for benefits of research to accrue widely.

6.2 Benefit sharing and equitable access

Researchers are required to meet obligations related to benefit sharing, while RECs and other oversight mechanisms should examine broader efforts to ensure equitable access to the benefits of such research. Alongside those obstacles detailed above (such as the affordability of such AI technologies), there are additional challenges that may hinder equitable access in low- and middle-income countries.

One reason is that, even prior to conducting research, AI technologies will not have been designed for use in low- and middle-income countries, including the eventual cost of implementation and ongoing maintenance. Second, after research is completed on an AI system, a developer may not introduce the technology in a low-or middle-income country setting, may not optimize it to be used appropriately (even if it is introduced), or may not update the algorithm to continue working appropriately for populations in low- and middle-income countries over time after it is commercialized. At present, RECs and other oversight mechanisms may not have the requisite information to assess whether an AI technology is designed to facilitate equitable access in low- and middle-income countries and may also not have the information to assess whether a research team will ensure that the benefits of research are shared equitably, especially since oversight of the research may occur too early to predict how a technology will be introduced and maintained. Therefore, researchers should disclose adequate information to RECs so that an assessment can be made as to whether AI technology will facilitate equitable access in low- and middle-income countries and whether research outcomes will be shared equitably.

6.3 Health data colonialism

One concern with AI-related research is that researchers predominantly based in high-income countries (but who could be from any third country) engage in "health data colonialism", or the acquisition of data from low- and middle-income countries to advance either commercial objectives or a study's objective (if by an academic) without due regard and respect for consent, privacy, and autonomy, and the interests of the community from whom data are obtained *(40)*. The collection of data from populations in low- and middle-income countries could be desirable to avoid excluding populations from the benefits of AI research, and yet collecting data without due regard to safeguards undermines the agency, dignity and human rights of individuals, undermines trust in those who conduct research, and undermines long-term trust in technologies that could have a beneficial impact in low- and middle-income countries. The desire of RECs and other oversight mechanisms globally as well as governments and institutions to improve inclusivity of AI-related health research may lead to researchers pointing to inclusion of populations in low- and middle-income countries as a justification for certain data practices that otherwise would be forbidden. These practices may not just be ethically problematic but could also violate laws and policies that enshrine data sovereignty or go against the growing momentum for governments to subject digital data to restrictions and rules over how the data of its citizens and residents are accessed, transferred or managed. RECs will play an important role in avoiding exploitative practices because, if RECs are not able to identify and curb these practices, it may encourage research entities in high-income countries to engage in ethics dumping, whereby "weaknesses and gaps in ethics policies and systems of lower income countries are intentionally exploited for intellectual or financial gains through research and publishing by higher income countries with a more stringent or complex ethical infrastructure in which such research and publishing practices would not be permitted" *(41)*.

All participants in AI-related health research, including sponsoring institutions, funders, researchers and different oversight mechanisms, should ensure that the objective of encouraging inclusivity of AI-related health research does not lead to exploitative practices with respect to acquiring and using health data.

6.4 Ethics dumping

"Ethics dumping" occurs because of a perception amongst researchers, wherever situated, that ethics oversight in a specific country will permit or will be unable to restrict practices that otherwise are forbidden elsewhere. It can also occur because of "power differentials, patronizing conduct, such as a false belief of superiority by the high-income country, inequitable and unfair distribution of burdens and benefits, cultural insensitivity, double ethical standards, or the lack of due diligence and transparency" *(41)*. All these causes are ones that could or do exist for AI-related health research, including dominance of research by white males in high-income countries that could exhibit both cultural insensitivity and a false belief of superiority, and also a lack of transparency within the sector. Researchers, RECs and other third parties, such as scientific journals, have a shared responsibility to identify instances of ethics dumping and to take concrete steps to restrict or discourage the practice.

6.5 Power inequities

Health research is often plagued by power inequities between researchers in high-income countries and partners in low- and middle-income countries. Therefore, research partners in low- and middle-income countries may also participate in research that they would conduct differently. There are multiple reasons. One reason is that high-income country researchers simply have more opportunities, such as privileged access to funding sources and relationships with scientific publishers that put them in charge of setting out research on their own terms. Therefore, low- and middle-income country partners have little choice but to participate only as junior partners *(41)*. In some cases, researchers in high-income countries that put forward unethical approaches may do so unknowingly due to a lack of knowledge and experience within a low- or middle-income country (and may not take steps to defer to the expertise of a research partner in the low- or middle-income country). Furthermore, with power residing with high-income country researchers, pressure is placed on research partners in low- and middle-income countries who may require funding to accept approaches to research that they may otherwise deem unethical. The participation of researchers in low- and middle-income countries could lend an imprimatur of credibility to research that therefore overcomes possible objections from RECs.

Overcoming power inequities requires greater capacity-building of researchers, research institutions and RECs in low- and middle-income countries (see below). One means that the WHO Research Ethics Review Committee has employed to reduce power inequities is limitation of the use of locally collected data to in-country research teams before externally based researchers may make use of such data. While this could address one problem, it does limit the exploration and use of such data for a time-limited period.

6.6 Capacity-building

Scientific investigation should involve collaboration with research entities in low- and middle-income countries that is based on mutual benefit and respect. This means equal partnership in the conceptualization, design and deployment of research, and in the equal sharing of credit and other benefits flowing from such research. Equitable partnerships will both build the capacity of researchers in low- and middle-income countries to carry out research on their own terms and according to their own priorities and improve the understanding of researchers in high-income countries of a country's health system, priorities and approaches. Such partnership can also encourage the development of research that reflects the actual needs and realities of low- and middle-income countries by researchers who are better placed to identify and eventually address unmet needs. Researchers can also provide an important check on potential practices that may be inappropriate in their low- or middle-income country setting (whether high-income country researchers are engaged in ethics dumping or introducing a practice that may be acceptable in their own country but not in another context).

Funders have a responsibility to reduce power inequities between researchers in high-income countries and low-income countries. One way to meet this responsibility would be for funders to provide funding and other resources to low- and middle-income country researchers for the development and use of AI tools and technologies. Capacity-building of researchers in low- and middle-income countries could also include short-term measures to fill gaps, such as providing low- and middle-income country researchers with remote access to AI-based research tools, reduced subscription rates for AI tools, and resources to build and use AI tools and technologies.

Alongside equal partnership with and capacity-building of researchers in low- and middle-income countries, it is critical that RECs and other oversight mechanisms in low- and middle-income countries also have the resources, expertise and training to provide effective oversight of AI-based health research and enforce ethical standards related to the conduct of AI-based research in their related settings. Because many of the challenges associated with AI-related health research are relatively novel and are likely to keep evolving, oversight may not keep pace with key concerns. Nevertheless, it is likely that RECs and oversight mechanisms in low- and middle-income countries may face even greater challenges to keep pace with novel ethical problems, especially if most new technologies and practices (and expertise to understand the different implications of such research approaches) are developed within companies or academic institutions in the United States, Europe and East Asia.

7. What are the risks and benefits of the use of artificial intelligence for ethics review?

Both generative AI (LMMs) and automated text analysis have been identified as tools to assist RECs with review of health research, and there is an assumption that commercially available LMMs such as ChatGPT are already used by both researchers and RECs during the ethics review process *(102)*. It is envisioned that data scientists and experts from ethics committees could work together to identify which parts of the ethics review process could benefit from automation or AI-based support *(102)*. This could be followed by data collection to facilitate the analysis of texts, including the annotations or responses of ethics reviewers to specific textual prompts drawn from previous protocols and decisions by RECs *(102)*. LMMs could potentially be used as an administrative tool to screen research protocols to ensure that submission requirements have been met. During an ethics review process, an appropriately trained and tested LMM could, for example, be utilized to review a research protocol quickly, identify ethics issues and provide recommendations or solutions for research teams to follow *(103)*.

Box 10: Potential benefits of using AI for research ethics review

Purported benefits of utilizing AI include:

(a) reducing inconsistency between RECs (for example with respect to multi-site studies) and within the same RECs over time and committee composition;

(b) improving the speed of ethics review, especially as research protocols are often lengthy with dense and complex information;

(c) reducing the burden on RECs and thereby enabling REC members to spend more time on learning and best practices (as opposed to rote review); and

(d) amplifying what reviewers usually identify as the most salient points for review, with LMMs and automated text analysis supplying reviewers with precedents or principles to support their judgement *(61, 102, 103)*.

RECs already utilize search and translation software that rely upon AI *(102)*, although the use of generative AI and other AI tools would probably have a more dramatic impact upon ethics review. AI would thus be envisioned as an "adjunct" or "first pass" screening tool that can help improve REC review, especially as RECs continue to face challenges with respect to institutional support and staffing in conducting timely reviews *(61)*.

However, the use of generative AI for ethics oversight carries certain risks. First, there is the well known problem of hallucinations and errors made by generative AI models that could materialize with the use of LMMs for ethics review. Second, confidential information inputted into an LMM may be eventually disclosed to third parties or in the public domain either because a user specifically requests the LMM to disclose such information or, in the case of one LMM, due to mistaken disclosure of other people's chat histories (even if not the substance of their conversations) *(8)*. Confidentiality breaches are particularly important in industry-sponsored clinical trials. Third, reviewers may not only rely upon LMMs to identify issues, principles and precedents, but may also over-rely upon and defer to the recommendations and judgements of LMMs in lieu of their own judgement and expertise (a form of automation bias) *(61, 102)*. This is a pertinent risk in low- and middle-income countries where skills development in research ethics review processes is suboptimal and building capacity amongst REC members is necessary as a first stage. Fourth, REC members usually have specialized roles and expertise for which no one LMM could feasibly support the work of each member *(102)*. This may limit the utility and effectiveness of an LMM. Fifth, RECs operate in different contexts, and the languages, principles, precedents and approaches utilized by RECs in one context may bear little or no relevance to the deliberations of an REC in a different region or context. LMMs may also not be able to account for the "local institutional and cultural contexts" in which they may be used *(61)*. Finally, RECs work through deliberation and dialogue, with one another and with researchers, to improve a study. Such human-centred deliberation over a research study and ethics considerations should not be replaced by AI *(102)*.

It is likely that some RECs or REC members are already using commercially available LMMs that are not specifically suited to ethics oversight. Sridharan and Sivaramakrishnan have studied the effectiveness of four large language models (not specifically designed for ethics review), and concluded that, with appropriate prompts, the models did provide appropriate information related to informed consent but performed "sub-optimally in identifying the suitability of the placebo arm, risk mitigation strategies and potential risks to study participants" *(104)*. Thus, it may be preferable that REC members and other interested parties proactively design an LMM that can be used responsibly by RECs *(102)*. This type of curated LMM or "application-specific LMM", which can reduce hallucinations and align the output with domain-specific knowledge, is possible given various techniques such as fine-tuning, retrieval augmented generation, and chain of thought *(78)*. However, resource constraints and the digital divide may make curated LMMs less accessible in low- and middle-income countries. For any use of an LMM, RECs and its members should determine if they are exposing themselves to any legal risks if they rely upon an LMM that misses a serious concern, which could lead to avoidable harm to a research participant.

Finally, the development of LMMs that could be used by RECs could also be used by researchers to improve research protocols and ethics review applications. While this could help researchers strengthen a research protocol to address key ethics issues, it could also be used to "game" ethics review by avoiding phrasing that would result in red flags by an LMM *(102)*. Ultimately, improving adherence to ethics principles by researchers will require investments in training, even as researchers may rely on AI to prepare research protocols and ethics-related materials *(102)*.

8. Conclusion

The risks and benefits of AI-related health research may not be evident when RECs customarily conduct their reviews. Furthermore, assigning sole responsibility for ethics oversight to RECs is unlikely to address all ethical risks associated with AI-related health research, and RECs themselves do not have the mandate under some country's laws and policies to review certain types of AI-related health research. In addition, RECs already lack the necessary resources to fulfil their current obligations.

Thus, while RECs may need to have both additional resources and capacity to provide effective oversight, it may be equally important that other mechanisms and obligations support or even replace the work of RECs, while ensuring RECs themselves are reviewing research when the review can have maximum impact and benefit. For AI-related health research of tools or technologies, WHO supports a process in which research ethics is a shared responsibility across the entire research life cycle. But it may also be one in which RECs may not themselves be responsible for reviewing research, but such a responsibility could be handed to other oversight mechanisms, such as a data access committee, or a specialized review committee that does not yet exist.

For AI-related health research, responsibility initially rests with both researchers and the funders or sponsors of AI-related health research. Researchers may struggle with satisfying ethical requirements during the earliest phases of technology development, and engagement of RECs may be too early to obtain information to assess research or may not be possible if the research does not qualify for ethics review. Irrespectively, researchers could incorporate ethical principles within the design of the research and mitigate key risks that they identify during the early stages of the research process. This may require researchers to have the training and understanding of the key ethical risks that may arise with research, or the willingness to consult with third parties who can provide relevant expertise.

AI research funders can play a complementary role by assessing adherence to ethics principles in determining which research applications to fund, setting out requirements, including through contracts, for AI-related health research, including the training and certification of researchers on ethics-related considerations, as well as providing resources so that research teams have the expertise to take diverse views and considerations into account.

RECs should be approached once researchers are likely to conduct research with human participants. This will require researchers to be well trained as to when research is likely to "cross over" and qualify as research with human participants. It will also require researchers to recognize when research, even if conducted with publicly available and anonymized data or synthetic data, and therefore not subject to REC review, may nevertheless require ethical oversight because of other risks and ethical concerns.

The Expert Group recognized that RECs cannot be a catch-all solution to the challenges and risks associated with AI-related health research and determined that expanding the mandate of RECs may overwhelm a well established and carefully calibrated oversight process. Data access committees, scientific and medical societies, health systems and other oversight mechanisms may need to assume a more prominent role in the ethical oversight of AI-related health research. There may also be a need to develop entirely new review mechanisms.

Even when RECs are required to review research, they may not be able to fully review research again at a later stage but should remain engaged with the research itself and provide continuous monitoring. They may also wish to closely align their monitoring and review with other review mechanisms that may be introduced before research is completed. This includes any review by a data access committee, as well as any review of research by a data protection agency (insofar that such research does not qualify for exemption under a national data protection law).

As noted above, additional oversight can be provided by scientific journals and regulatory agencies that may be required to review and approve a technology that had been the object of a research study. Thus, even as REC

review remains indispensable to the oversight of AI-related health research, the burden should be shared with other parties, and each party should have a distinct role within the development, design and conduct of such research that complements the efforts of other oversight bodies.

The rapid adoption of the outcomes of AI-related health research in a clinical context soon after research is completed requires not just improved research ethics norms, as discussed in this report, but also a robust ethical framework of learning health systems that will need to be developed.

Finally, to provide coherent guidance and therefore to ensure consistency across a diverse range of entities involved in oversight, standard-setting bodies should assume responsibility. Standard-setting bodies, such as CIOMS and the World Medical Association, should play a critical role in updating existing standards to account for the growth of AI-related health research. The Expert Group noted that these bodies should convene and update their guidelines to account for the different types of AI-related health research discussed in this report.

This could include determining which types of AI-related health research are subject to review by RECs, as well as suggesting other mechanisms or third parties that may provide appropriate ethical oversight of AI-related health research. Standard-setting bodies should also consider how best to update their guidelines on a more frequent basis to account for the rapid emergence and adoption of novel research practices that use AI tools and technologies, the types of AI-supported technologies under study, and the wider implications of AI for health care and society. Finally, to assist all parties with the growing use of AI in health-related research globally, standard-setting bodies should consider development of toolkits that compile checklists, templates, case examples and training resources to facilitate adherence to new or updated standards.

9. Key considerations for diverse actors

The following are considerations that could inform the current activities and future deliberations of RECs (and their institutions), researchers, third-party oversight mechanisms, standard-setting bodies, and other parties concerned with AI-related health research.

9.1 Considerations for research ethics committees

AI-related health research may only satisfy key ethical requirements if RECs can be provided with an appropriate mandate, resources, training and support. The following considerations suggest ways in which RECs can adapt to and successfully review AI-related health research that may fall within their remit.

A. Resources, training and expertise for RECs

To improve the capacity of RECs to oversee AI-related health research, the following measures could be considered.

4. **Resource provision.** Institutions responsible for the management of RECs should provide them with additional resources to deal with AI-related research. Adequate support could include monetary and human resources, administrative support, and support to adapt existing standard operating procedures.
5. **Educational instruction and training.** Institutions should provide educational instruction and specific training so that REC members can understand and evaluate AI-related health research. Staying up to date on advancements in the field is part of capacity development of REC members.
6. **Diverse and knowledgeable membership.** REC membership should include diverse expertise to provide effective oversight of AI-related health research, including members familiar with AI-specific ethics questions.
7. **Cooperation amongst RECs.** Networks of RECs should aim to collaborate at the local, national and regional levels to facilitate experience sharing and consultation on AI-related research ethics questions.

B. Additional institutional capacity to support REC oversight

In addition to increased training, resources, expertise and time for individual RECs to conduct ethics oversight, ethics review for AI-related health research could be strengthened by competent authorities at the national or regional level through the following option.

1. **National or regional specialized REC.** In countries with relatively little AI-related health research, national authorities could consider establishing a national specialized REC that reviews AI-related health research conducted in its territory, or works with neighbouring countries to introduce a regional REC that reviews AI-related health research.

C. Strengthening the oversight by RECs of research that involves the private sector

To strengthen oversight of research that includes the sponsorship or participation of the private sector, the following considerations are suggested.

1. **Need for transparency.** Commercial developers and researchers should be fully transparent with information that can assist RECs and other oversight mechanisms to assess AI-related health research. There could be independent mechanisms in place nationally or regionally to facilitate reporting, and external monitoring to determine if such obligations are met.
2. **Government review of company oversight mechanisms.** Governments and regulators should ensure that a company's internal oversight mechanisms or privately funded and operated RECs are exercising independent judgement and upholding international ethics standards.
3. **REC oversight of public–private research.** AI-related health research often includes both the public and private sector, whether (a) research is conducted through public–private partnerships; (b) research is initiated in the public sector and completed in the private sector; or (c) research is completed in the public sector and acquired by the private sector. For public–private research, research ethics oversight should be put in place to safeguard the ethical principles, goals and objectives of the public sector within a joint research partnership.

D. Strengthening the oversight of RECs for multicountry or multi-site research

To improve oversight of multicountry or multi-site research, including research using social media platforms, the following should be considered.

1. **Promotion of cooperation.** Increased efforts are needed to promote bilateral or multilateral cooperation, joint review processes and common ethical standards.
2. **Risk-based oversight of research using social media platforms.** The REC that originally reviews research should consider whether additional approval from other locations is needed on the basis of certain criteria, for example where the study (a) poses more than a minimal risk; (b) involves the development of a product or service that may eventually be deployed in a country in question; or (c) has implications beyond the use of data from that country.

E. Information required of researchers by RECs

RECs could introduce criteria with respect to the types of information, including impact assessments, that researchers should submit. To ensure that RECs can be provided with sufficient appropriate information to undertake a full analysis and issue useful decisions, RECs could require that researchers disclose the following types of information.

1. **Risk assessment.** The risk assessment would include the risk classification of an AI research project and the rationale for the risk assessment.
2. **Impact assessments.** Impact assessments would improve an REC's ability to contemplate the broader societal impacts of the research. They could include an equity impact assessment (such as concerns with bias or inadequate benefit sharing), an environmental impact assessment that examines the carbon and water footprint of AI interventions if scaled up for use, or any internal evaluations carried out by researchers.
3. **Data set assessment.** This could include assessment of the data sets that had been used to train, develop and validate AI as well as the results of any tests used to assess bias.

F. Beneficiary, community and vulnerable subgroup engagement

RECs can play a critical role in building public understanding of, trust in and engagement with the use of AI for health research (and the wider consequences of the use of AI across health systems) and promoting or recommending adoption by researchers of participatory and community-engaged approaches to AI-related health research. RECs can also improve their own oversight process by ensuring an inclusive membership for the review of AI-related health research.

RECs could consider the following actions.

1. **Inclusive representation.** At least one representative of a beneficiary group (such as a patient group) or a representative of crowd workers should be included, reflecting the added value of their representation on an REC and their ability to conduct outreach to assist beneficiaries in considering the risks and benefits of relevant AI-related health research on an ad hoc basis, where such specific community or group interests are affected.
2. **Representation with lived experience.** RECs could consider inclusion of at least one representative of any population with lived experience for whom AI-related health research could have differing impacts or implications.

G. Unethical research practices in low- and middle-income countries

1. **Health data colonialism.** RECs should ensure that the objective of encouraging inclusivity of AI-related health research does not lead to exploitative practices with respect to acquiring and using health data in a manner that amounts to health data colonialism.
2. **Ethics dumping.** RECs have a responsibility to identify instances of ethics dumping and to take concrete steps to restrict or discourage such practices.

9.2 Considerations for researchers (and their institutions)

The responsibility of researchers should not just be a box-ticking exercise to meet ethical obligations but instead should be to apply ethical principles from the initial design of research through the completion of research and with respect to long-term benefits and consequences. Researchers can consider the following actions to abide by ethical principles, standards and practices.

1. **Implement training and certification.** Complete training and obtain an appropriate certification, where feasible and available, to carry out AI-related health research.
2. **Build inclusive partnerships with low- and middle-income country institutions and researchers.** Collaborate with low- and middle-income country institutions and researchers for the co-creation of data sets, research and findings. Work cooperatively to assess the risks and benefits of research and whether the research design and outputs are sufficient for benefits of the research to accrue widely. Ensure that joint efforts can identify instances of ethics dumping, with concrete steps to restrict or discourage these practices.
3. **Involve health experts early.** Include health experts or health services in the responsible design of their research, or request health systems that may benefit from technologies to connect researchers with an appropriate health expert who can provide critical input.
4. **Apply "ethics by design".** Be aware of and try to address possible risks during design and development, including those relating to safety and cybersecurity, bias or privacy-related risks. This could require applying "ethics by design" at the outset. It could also require remaining up to date with requirements published by standard-setting institutions, intergovernmental agencies, institutions or national RECs.
5. **Test for and identify errors.** Prior to conducting research involving human beings or commercial testing, perform necessary state-of-the-art tests and identify errors that may be features of AI tools, as resources and capacity allow.
6. **Respect human rights standards and cultural values.** Follow international human rights standards related to the conduct of research. Respect cultural norms, values and practices where research is situated.
7. **Consider broader impacts.** Examine and consider, to the extent feasible and for which information is available, the wider social impacts of AI technology when setting out research, including for example its impact on health equity and the environmental impacts of widespread use of an AI technology (carbon and water footprint).
8. **Disclose equitable access commitments.** Disclose adequate information to RECs or other oversight mechanisms so that an assessment can be made as to whether and under which conditions an AI technology will facilitate equitable access in low- and middle-income countries and research outcomes will be shared equitably.
9. **Protect crowd workers.** Uphold a responsibility to ensure appropriate treatment of crowd workers as it relates to labour practices and other health-related consequences associated with their employment. Researchers and their institutions could also ensure through careful vetting and due diligence of intermediary companies that supply crowd workers that, irrespective of place of residence, all crowd workers are provided with labour protections. Researchers (as well as their institutions) could also consider upholding this obligation irrespective of whether research conducted with crowd workers falls within the scope of REC oversight.
10. **Be transparent with internal assessments and evaluations.** Be transparent with findings from internal evaluations and testing, including negative or inconclusive results, so that RECs or other oversight mechanisms can more accurately consider possible risks during their deliberations, and thus inform the future decision-making of research partners, subjects and participants.
11. **Be transparent with the usage of AI tools and technologies for research.** Disclose the use of AI tools and technologies, including which AI tools and technologies are used, their performance, and how AI was applied (for example, prompts used with generative AI models).
12. **Avoid conflicts of interest.** Disclose additional information, such as other parties involved in the development and training of an AI algorithm or the development of a research protocol, so that RECs and other entities that conduct oversight can identify potential conflicts of interest at an early stage and provide additional supervision and feedback.

9.3 Considerations for third-party stakeholders

The responsibility to address the many risks of AI-related health research will need to be shared with other parties that participate in oversight. The following considerations can strengthen the role that different third-party stakeholders could play to improve oversight and conduct of AI-related health research:

A. Public and private funders of AI-related health research

Public and private funders could request or require researchers through funding agreements to undertake the following.

1. **Identify AI-specific risks.** Request researchers, in funding applications, to identify potential AI-related risks to research participants as well as AI-related societal risks, and how researchers may wish to mitigate, address or examine risks during the research process.
2. **Be transparent.** Request researchers to both produce and be transparent with the different forms of data (including the source of data) and information that could inform RECs and other entities (such as data access committees and scientific journals) on the risks and benefits of the proposed AI-related health research, including impact assessments, possible error rates of AI, and known concerns with safety and security of an algorithm.
3. **Build capacity.** Reduce power inequities between researchers in high-income and low-income countries. One way to meet this consideration would be to provide funding and other resources to low- and middle-income country researchers for the development and use of AI tools and technologies.
4. **Ensure fair distribution of risks and benefits.** Require international research partnerships to set out and implement a fair distribution of risks and benefits associated with research and its outcomes.
5. **Complete training.** Require researchers to have already completed AI-specific research ethics training to qualify for funding, with restrictions on funding if researchers do not demonstrate completion (and comprehension) of such training.
6. **Report back lessons learned.** Require researchers, at the completion of the research process, to report back to funders their successes and challenges with ethical risks so that funders can apply lessons learned.

B. Health data spaces and data hubs

Health data spaces and data hubs could develop and introduce common standards and principles that could be adopted and enforced by institutions that join a health data space or participate in a data hub. In addition, they could consider the following actions.

1. **Apply ethical principles.** Identify, set out and enforce ethical principles systematically, including appropriate standards for informed consent, as a condition for researchers to have access to de-identified or anonymized health data for which the data hub has control.
2. **Engage in cooperative processes.** Work cooperatively with data protection agencies and RECs to enforce ethical principles systematically.

C. Data access committees

Data access committees could consider the following actions.

1. **Review health data requests.** Such a review could ensure that relevant AI-related health research conforms to its standards. This includes requiring researchers to conform with data standards that facilitate reuse, including for AI-related health research, so that the full potential of the data can be realized.
2. **Consider REC feedback.** If RECs have already reviewed research and provided guidance, data access requests could also take into consideration the feedback provided by an REC.

D. Scientific journals and publishers

Scientific journals and publishers should consider the following actions.

1. **Ensure publication integrity.** Only publish high-quality and ethical AI-related health research.
2. **Build partnerships.** Establish partnerships with regular reviewers trained in reviewing AI-focused research articles, including assessment of ethical aspects. Journals could also, where possible, add an ethics specialist to review AI-focused research articles.
3. **Ensure data transparency.** Require submitting authors to provide accessibility to raw data or software code, to the extent possible, so that peer reviewers can assess research methods.
4. **Counteract ethics dumping.** Identify instances of ethics dumping and take concrete steps to restrict or discourage such practices.

E. Scientific medical societies and academies

Scientific medical societies and academies, where they exist, can contribute to high ethical standards of AI-related health research in the following ways.

1. **Promote discussion and information sharing.** Discuss and communicate ethical issues with AI-related health research (within their respective area of activity) with the public, and provide accurate information about studies and their limitations.
2. **Encourage education.** Incorporate ethical considerations of AI-related health research within its educational programmes, events and conferences. Continuing medical education on ethics of AI-based health research could be organized or supported by every medical society.
3. **Help develop ethical guidelines.** Formulate or adopt ethical guidelines for AI-related health research.
4. **Support institutional processes.** Create and support ethics committees or consultation services within scientific medical societies and academies.

F. Regulatory agencies and other government oversight bodies

Analyse and identify risks. Regulatory agencies and other oversight bodies should analyse known risks with AI-related health research within their own review standards and specifically examine research studies on AI-based health technologies to ensure that ethical risks already identified by an REC were addressed throughout the research process.

References

1. Abiodun TN, Okunbor D, Osamor VC. Remote Health Monitoring in Clinical Trial using Machine Learning Techniques: A Conceptual Framework. Health and Technology. 2022;12(2):359-64 (https://doi.org/10.1007/s12553-022-00652-z).
2. World Health Organization. Benefits and risks of using artificial intelligence for pharmaceutical development and delivery. World Health Organization; 2024. Licence: CC BY-NC-SA 3.0 IGO (https://iris.who.int/handle/10665/375871).
3. Council for International Organizations of Medical Sciences. International ethical guidelines for health-related research involving humans: Prepared by the Council for International Organizations of Medical Sciences (CIOMS) in collaboration with the World Health Organization (WHO). Geneva: Council for International Organizations of Medical Sciences; 2016. Licence: CC BY-NC-SA 3.0 IGO (https://cioms.ch/wp-content/uploads/2017/01/WEB-CIOMS-EthicalGuidelines.pdf).
4. World Health Organization. Research Ethics Committee (ERC): Guidelines on submitting research proposals for ethics review. World Health Organization; 2018 (https://www.who.int/groups/research-ethics-review-committee/guidelines-on-submitting-research-proposals-for-ethics-review).
5. Petermann M, Tempini N, Kherroubi Garcia I, Whitaker K, Strait A. Looking before we leap: Expanding ethical review processes for AI and data science research. University of Exeter; 2022 (https://www.adalovelaceinstitute.org/report/looking-before-we-leap/).
6. Mittelstadt B. Principles alone cannot guarantee ethical AI. Nature Machine Intelligence. 2019;1(11):501-7 (https://doi.org/10.2139/ssrn.3391293).
7. World Health Organization. Ethics and governance of artificial intelligence for health. Geneva: World Health Organization; 2021. Licence: CC BY-NC-SA 3.0 IGO (https://iris.who.int/handle/10665/341996).
8. Ethics and governance of artificial intelligence for health: large multi-modal models. WHO guidance. Geneva: World Health Organization; 2024. Licence: CC BY-NC-SA 3.0 IGO (https://iris.who.int/handle/10665/375579).
9. Purtova N. Health Data for Common Good: Defining the Boundaries and Social Dilemmas of Data Commons. In: Adams S, Purtova N, Leenes R, editors. Under Observation: The Interplay Between eHealth and Surveillance. Cham: Springer International Publishing; 2017:177-210 (Law, Governance and Technology Series, https://doi.org/10.1007/978-3-319-48342-9_10).
10. Purtova N. The illusion of personal data as no one's property. Law, Innovation and Technology. 2015;7(1):83-111 (https://ssrn.com/abstract=2346693).
11. Powles J, Hodson H. Google DeepMind and healthcare in an age of algorithms. Health and Technology. 2017;7(4):351-67 (https://doi.org/10.1007/s12553-017-0179-1).
12. Hodson H. Did Google's NHS patient data deal need ethical approval? [website]. New Scientist; 2016 (https://www.newscientist.com/article/2088056-did-googles-nhs-patient-data-deal-need-ethical-approval/).
13. WMA Declaration of Taipei on ethical considerations regarding health databases and biobanks. World Medical Association; 2016 (https://www.wma.net/policies-post/wma-declaration-of-taipei-on-ethical-considerations-regarding-health-databases-and-biobanks/).
14. Resseguier A, Ufert F. AI research ethics is in its infancy: the EU's AI Act can make it a grown-up. Research Ethics. 2023;20(2):143-55 (https://doi.org/10.1177/17470161231220946).
15. General Data Protection Regulation. Regulation (EU) 2016/679 of the European Parliament and of the Council. Official Journal of the European Union. 2016;679:1-88 (https://eur-lex.europa.eu/eli/reg/2016/679/oj).
16. Protection of Personal Information Act (POPI Act). Section 6. South Africa: Parliament of South Africa; 2021 (https://popia.co.za/).
17. Smit J-AR, Mostert M, van der Graaf R, Grobbee DE, van Delden JJM. Specific measures for data-intensive health research without consent: a systematic review of soft law instruments and academic literature. European Journal of Human Genetics. 2024;32(1):21-30 (https://doi.org/10.1038/s41431-023-01471-0).
18. Meszaros J, Ho C-h. AI research and data protection: Can the same rules apply for commercial and academic research under the GDPR? Computer Law & Security Review. 2021;41:105532 (https://doi.org/10.1016/j.clsr.2021.105532).
19. Australian Research Data Commons. CARE Principles for Indigenous Data Governance [website]. Australian Research Data Commons; 2025 (https://ardc.edu.au/resource/the-care-principles/).

20. Council of Europe. Recommendation CM/Rec(2019)2 of the Committee of Ministers to Member States on the protection of health-related data. Strasbourg: Council of Europe; 2019 (https://edoc.coe.int/en/international-law/7969-protection-of-health-related-date-recommendation-cmrec20192.html).
21. DHR-ICMR Artificial Intelligence Cell. Ethical guidelines for application of artificial intelligence in biomedical research and healthcare. India: Indian Council of Medical Research; 2023.
22. Schmidt E. This is how AI will transform the way science gets done. 1 MAIN ST, 13 FLR, CAMBRIDGE, MA, 02142, USA: MIT Technology Review; 2023; 126: 9-17 (https://www.technologyreview.com/2023/07/05/1075865/eric-schmidt-ai-will-transform-science/).
23. Ferretti A, Ienca M, Sheehan M, Blasimme A, Dove ES, Farsides B et al. Ethics review of big data research: What should stay and what should be reformed? BMC Medical Ethics. 2021;22(1):51 (https://doi.org/10.1186/s12910-021-00616-4p).
24. Ferretti A, Ienca M, Hurst S, Vayena E. Big Data, Biomedical Research, and Ethics Review: New Challenges for IRBs. Ethics & Human Research. 2020;42(5):17-28 (https://doi.org/10.1002/eahr.500065).
25. McKay F, Williams BJ, Prestwich G, Bansal D, Treanor D, Hallowell N. Artificial intelligence and medical research databases: ethical review by data access committees. BMC Medical Ethics. 2023;24(1):49 (https://doi.org/10.1186/s12910-023-00927-8).
26. Mathur R, Swaminathan S. National ethical guidelines for biomedical & health research involving human participants, 2017: A commentary. Indian Journal of Medical Research. 2018;148(3):279-83 (https://doi.org/10.4103/0971-5916.245303).
27. Federal Policy for the Protection of Human Subjects ('Common Rule'). U.S. Department of Health and Human Services; 2025 (https://www.hhs.gov/ohrp/regulations-and-policy/regulations/common-rule/index.html).
28. Office for Civil Rights. Guidance Regarding Methods for De-identification of Protected Health Information in Accordance with the Health Insurance Portability and Accountability Act (HIPAA) Privacy Rule [website]. U.S. Department of Health and Human Services; 2025 (https://www.hhs.gov/hipaa/for-professionals/special-topics/de-identification/index.html).
29. Ienca M, Ferretti A, Hurst S, Puhan M, Lovis C, Vayena E. Considerations for ethics review of big data health research: A scoping review. PLOS ONE. 2018;13(10):e0204937 (https://doi.org/10.1371/journal.pone.0204937).
30. Bouhouita-Guermech S, Gogognon P, Bélisle-Pipon J-C. Specific challenges posed by artificial intelligence in research ethics. Frontiers in Artificial Intelligence. 2023;Volume 6 - 2023 (https://doi.org/10.3389/frai.2023.1149082).
31. Myer R. Everything we know about Facebook's secret Mood-Manipulation Experiment. The Atlantic. (https://www.theatlantic.com/technology/archive/2014/06/everything-we-know-about-facebooks-secret-mood-manipulation-experiment/373648/).
32. Council of Europe. The Council of Europe Protocol to the Convention on Human Rights and Biomedicine concerning Biomedical Research (2005) includes, under its explanatory report, the risk to the psychological health of the person concerned.: Council of Europe; 2005 (https://www.coe.int/en/web/impact-convention-human-rights/convention-on-human-rights-and-biomedicine#/).
33. Kaushik D, Lipton ZC, London AJ. Resolving the Human-subjects Status of Machine Learning's Crowdworkers. Queue. 2022;21:101 - 27 (https://doi.org/10.1145/3639452).
34. World Health Organization. Standards and operational guidance for ethics review of health-related research with human participants. World Health Organization; 2011. Licence: CC BY-NC-SA 3.0 IGO (https://iris.who.int/handle/10665/44783).
35. European Data Protection Supervisor. A preliminary opinion on data protection and scientific research. European Data Protection Supervisor; 2020 (https://share.google/rv4QIVny3oI6tmesQ).
36. Brückner S, Dridi A, Deshmukh A, Kirsten T, Lauber-Rönsberg A, Riedel R et al. A user-driven consent platform for health data sharing in digital health applications. npj Digital Medicine. 2025;8(1):699 (https://doi.org/10.1038/s41746-025-02147-3).
37. Cengiz N, Kabanda SM, Moodley K. Cross-border data sharing through the lens of research ethics committee members in sub-Saharan Africa. PLOS ONE. 2024;19(5):e0303828 (https://doi.org/10.1371/journal.pone.0303828).
38. Ferretti A, Ienca M, Velarde MR, Hurst S, Vayena E. The Challenges of Big Data for Research Ethics Committees: A Qualitative Swiss Study. Journal of Empirical Research on Human Research Ethics. 2022;17(1-2):129-43 (https://doi.org/10.1177/15562646211053538).
39. Sloan L, Jessop C, Al Baghal T, Williams M. Linking Survey and Twitter Data: Informed Consent, Disclosure, Security, and Archiving. Journal of Empirical Research on Human Research Ethics. 2019;15(1-2):63-76 (https://doi.org/10.1177/1556264619853447).

40. Shaw J, Ali J, Atuire CA, Cheah PY, Español AG, Gichoya JW et al. Research ethics and artificial intelligence for global health: perspectives from the global forum on bioethics in research. BMC Medical Ethics. 2024;25(1):46 (https://doi.org/10.1186/s12910-024-01044-w).
41. Teixeira da Silva JA. Handling Ethics Dumping and Neo-Colonial Research: From the Laboratory to the Academic Literature. Journal of Bioethical Inquiry. 2022;19(3):433-43 (https://doi.org/10.1007/s11673-022-10191-x).
42. Legido-Quigley C, Wewer Albrechtsen NJ, Bæk Blond M, Corrales Compagnucci M, Ernst M, Herrgård MJ et al. Data sharing restrictions are hampering precision health in the European Union. Nature Medicine. 2025;31(2):360-1 (https://doi.org/10.1038/s41591-024-03437-1).
43. Cengiz N, Kabanda SM, Esterhuizen TM, Moodley K. Exploring perspectives of research ethics committee members on the governance of big data in sub-Saharan Africa. South African Journal of Science. 2023;119(5/6) (https://doi.org/10.17159/sajs.2023/14905).
44. Kim J. Data brokers and the sale of Americans' mental health data. Duke University: Sanford School of Public Policy; 2023 (https://techpolicy.sanford.duke.edu/data-brokers-and-the-sale-of-americans-mental-health-data/).
45. Research ethics for AI research projects: Guidelines to Support the Work of Ethics Committees at Universities. Zevidi: Centre Responsible Digitality; 2023 (https://zevedi.de/wp-content/uploads/2023/02/ZEVED_AI-Research-Ethics_web_2023.pdf).
46. Huang Y-J, Chen C-h, Yang H-C. AI-enhanced integration of genetic and medical imaging data for risk assessment of Type 2 diabetes. Nature Communications. 2024;15(1):4230 (https://doi.org/10.1038/s41467-024-48618-1).
47. Wang Y, Kosinski M. Deep neural networks are more accurate than humans at detecting sexual orientation from facial images. J Pers Soc Psychol. 2018;114(2):246-57 (https://doi.org/10.1037/pspa0000098).
48. Mittelstadt BD, Allo P, Taddeo M, Wachter S, Floridi L. The ethics of algorithms: Mapping the debate. Big Data & Society. 2016;3(2):2053951716679679 (https://doi.org/10.1177/2053951716679679).
49. Sandvig C, Hamilton K, Karahalios K, Langbort C. Auditing Algorithms : Research Methods for Detecting Discrimination on Internet Platforms. 2014 (https://www.semanticscholar.org/paper/Auditing-Algorithms-%3A-Research-Methods-for-on-Sandvig-Hamilton/b7227cbd34766655dea10d0437ab10df3a127396?utm_source=direct_link).
50. McCradden MD, Stedman I. Explaining decisions without explainability? Artificial intelligence and medicolegal accountability. Future Healthcare Journal. 2024;11(3):100171 (https://doi.org/10.1016/j.fhj.2024.100171).
51. Rennie S, Atuire C, Mtande T, Jaoko W, Litewka S, Juengst E et al. Public health research using cell phone derived mobility data in sub-Saharan Africa: Ethical issues. South African journal of science. 2023;119(5-6):1-7 (https://doi.org/10.17159/sajs.2023/14777).
52. London AJ, Karlawish J, Largent EA, Hey SP, McCarthy EP. Algorithmic identification of persons with dementia for research recruitment: ethical considerations. Informatics for Health and Social Care. 2024;49(1):28-41 (https://doi.org/10.1080/17538157.2023.2299881).
53. Rothstein MA. Should Chatbots Be Used to Obtain Informed Consent for Research? Ethics & Human Research. 2023;45(6):46-50 (https://doi.org/10.1002/eahr.500190).
54. Naddaf M. More than half of researchers now use AI for peer review - often against guidance. Nature. 2026;649(8096):273-4 (https://doi.org/10.1038/d41586-025-04066-5).
55. Metz C. Chatbots may 'hallucinate' more often than many realize. The New York Times. (https://www.nytimes.com/2023/11/06/technology/chatbotshallucination-rates.html).
56. European Commission. Living Guidelines on the Responsible Use of Generative AI in Research. European Commission: ERA Forum Stakeholders; 2025 (https://share.google/HU87ZoEtJbOXbSaul).
57. Blau W, Cerf VG, Enriquez J, Francisco JS, Gasser U, Gray ML et al. Protecting scientific integrity in an age of generative AI. Proceedings of the National Academy of Sciences. 2024;121(22):e2407886121 (https://doi.org/10.1073/pnas.2407886121).
58. Tools such as ChatGPT threaten transparent science; here are our ground rules for their use. Nature. 2023;613(7945):612 (https://doi.org/10.1038/d41586-023-00191-1).
59. Zielinski C, Winker MA, Aggarwal R, Ferris LE, Heinemann M, Lapeña JF, Jr. et al. Chatbots, generative AI, and scholarly manuscripts: WAME recommendations on chatbots and generative artificial intelligence in relation to scholarly publications. Colomb Med (Cali). 2023;54(3):e1015868 (https://doi.org/10.25100/cm.v54i3.5868).
60. Koller D, Beam A, Manrai A, Ashley E, Liu X, Gichoya J et al. Why We Support and Encourage the Use of Large Language Models in NEJM AI Submissions. NEJM AI. 2024;1(1):AIe2300128 (https://doi.org/10.1056/AIe2300128).

61. Porsdam Mann S, Vazirani AA, Aboy M, Earp BD, Minssen T, Cohen IG et al. Guidelines for ethical use and acknowledgement of large language models in academic writing. Nature Machine Intelligence. 2024;6(11):1272-4 (https://doi.org/10.1038/s42256-024-00922-7).
62. Arora A, Wagner SK, Carpenter R, Jena R, Keane PA. The urgent need to accelerate synthetic data privacy frameworks for medical research. Lancet Digit Health. 2025;7(2):e157-e60 (https://doi.org/10.1016/s2589-7500(24)00196-1).
63. Giuffrè M, Shung DL. Harnessing the power of synthetic data in healthcare: innovation, application, and privacy. npj Digital Medicine. 2023;6(1):186 (https://doi.org/10.1038/s41746-023-00927-3).
64. Kundaliya D. Open AI and other firms are using synthetic data to train AI models [website]. Computing; 2023 (https://www.computing.co.uk/news/4120522/openai-firms-synthetic-train-ai-models).
65. Susser D, Schiff DS, Gerke S, Cabrera LY, Cohen IG, Doerr M et al. Synthetic Health Data: Real Ethical Promise and Peril. Hastings Center Report. 2024;54(5):8-13 (https://doi.org/10.1002/hast.4911).
66. Koul A, Duran D, Hernandez-Boussard T. Synthetic data, synthetic trust: navigating data challenges in the digital revolution. Lancet Digit Health. 2025;7(11):100924 (https://doi.org/10.1016/j.landig.2025.100924). Licence: NIHMS2130947.
67. Johnson E, Hajisharif S. The intersectional hallucinations of synthetic data. AI & SOCIETY. 2025;40(3):1575-7 (https://doi.org/10.1007/s00146-024-02017-8).
68. Foraker R, Morrow JD, Johnson JA, Wilcox AB, Forster AJ, Payne PRO. Understanding synthetic data: artificial datasets for real-world evidence. BMJ Evidence-Based Medicine. 2025:bmjebm-2024-113617 (https://doi.org/10.1136/bmjebm-2024-113617).
69. Lenharo M. The testing of AI in medicine is a mess. Here's how it should be done. Nature. 2024;632(8026):722-4 (https://doi.org/10.1038/d41586-024-02675-0).
70. Kaufman J, Jeon J, Oreskovic J, Fossat Y. Linear effects of glucose levels on voice fundamental frequency in type 2 diabetes and individuals with normoglycemia. Scientific Reports. 2024;14(1):19012 (https://doi.org/10.1038/s41598-024-69620-z).
71. Landi H. HIMSS24: How Epic is building out AI, ambient technology for clinicians [website]. Fierce Health; 2024 (https://www.fiercehealthcare.com/ai-and-machine-learning/himss24-how-epic-building-out-ai-ambient-technology-clinicians).
72. Wong A, Otles E, Donnelly JP, Krumm A, McCullough J, DeTroyer-Cooley O et al. External Validation of a Widely Implemented Proprietary Sepsis Prediction Model in Hospitalized Patients. JAMA Intern Med. 2021;181(8):1065-70 (https://doi.org/10.1001/jamainternmed.2021.2626).
73. Jercich K. Research suggests Epic Sepsis Model is lacking in predictive power [website]. Healthcare IT News; 2021 (https://www.healthcareitnews.com/news/research-suggests-epic-sepsis-model-lacking-predictive-power).
74. Ross C. Epic's overhaul of a flawed algorithm shows why AI oversight is a life-or-death issue. StatNews. (https://www.statnews.com/2022/10/24/epic-overhaul-of-a-flawed-algorithm/).
75. Youssef A, Nichol AA, Martinez-Martin N, Larson DB, Abramoff M, Wolf RM et al. Ethical Considerations in the Design and Conduct of Clinical Trials of Artificial Intelligence. JAMA Netw Open. 2024;7(9):e2432482 (https://doi.org/10.1001/jamanetworkopen.2024.32482).
76. Goetz L, Seedat N, Vandersluis R, van der Schaar M. Generalization-a key challenge for responsible AI in patient-facing clinical applications. NPJ Digit Med. 2024;7(1):126 (https://doi.org/10.1038/s41746-024-01127-3).
77. Kapoor S, Narayanan A. Leakage and the reproducibility crisis in machine-learning-based science. Patterns. 2023;4(9):100804 (https://doi.org/10.1016/j.patter.2023.100804).
78. Persson I, Macura A, Becedas D, Sjövall F. Early prediction of sepsis in intensive care patients using the machine learning algorithm NAVOY® Sepsis, a prospective randomized clinical validation study. J Crit Care. 2024;80:154400 (https://doi.org/10.1016/j.jcrc.2023.154400).
79. Using a Centralized IRB Review Process in Multicenter Clinical Trials. U.S. Food and Drug Administration; 2006 (https://www.fda.gov/regulatory-information/search-fda-guidance-documents/using-centralized-irb-review-process-multicenter-clinical-trials).
80. Friesen P, Douglas-Jones R, Marks M, Pierce R, Fletcher K, Mishra A et al. Governing AI-Driven Health Research: Are IRBs Up to the Task? Ethics Hum Res. 2021;43(2):35-42 (https://doi.org/10.1002/eahr.500085).
81. Vynck GD, Oremus W. As AI booms, tech firms are laying off their ethicists [website]. The Washington Post; 2023 (https://www.washingtonpost.com/technology/2023/03/30/tech-companies-cut-ai-ethics/).
82. NIH Office of Intramural Research. Responsible Conduct of Research Training [website]. National Institute of Health, USA; 2024 (https://oir.nih.gov/sourcebook/ethical-conduct/responsible-conduct-research-training).

83. National Health and Medical Research Council. NHMRC Artificial Intelligence Workshop Report. Australia; 2024 (https://www.nhmrc.gov.au/about-us/publications/nhmrc-artificial-intelligence-workshop-report#block-views-block-file-attachments-content-block-1).
84. Feffer M, Sinha A, Deng WH, Lipton ZC, Heidari H. Red-Teaming for Generative AI: Silver Bullet or Security Theater? Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society. 2024;7(1):421-37 (https://doi.org/10.1609/aies.v7i1.31647).
85. Luyckx VA, Reis A, Maher D, Vahedi M. Highlighting the ethics of implementation research. Lancet Glob Health. 2019;7(9):e1170-e1 (https://doi.org/10.1016/s2214-109x(19)30310-9).
86. Vesper I. Clinical-trial reporting is on the rise. Nature. 2024;634(8034):S21-s3 (https://doi.org/10.1038/d41586-024-03317-1).
87. Cruz Rivera S, Liu X, Chan A-W, Denniston AK, Calvert MJ, Darzi A et al. Guidelines for clinical trial protocols for interventions involving artificial intelligence: the SPIRIT-AI extension. Nature Medicine. 2020;26(9):1351-63 (https://doi.org/10.1038/s41591-020-1037-7).
88. Liu X, Cruz Rivera S, Moher D, Calvert MJ, Denniston AK, Chan A-W et al. Reporting guidelines for clinical trial reports for interventions involving artificial intelligence: the CONSORT-AI extension. Nature Medicine. 2020;26(9):1364-74 (https://doi.org/10.1038/s41591-020-1034-x).
89. Collins GS, Moons KGM, Dhiman P, Riley RD, Beam AL, Van Calster B et al. TRIPOD+AI statement: updated guidance for reporting clinical prediction models that use regression or machine learning methods. Bmj. 2024;385:e078378 (https://doi.org/10.1136/bmj-2023-078378).
90. European Commission. Questions and answers - EU Health: European Health Data Space (EHDS). European Commission; 2022 (https://ec.europa.eu/commission/presscorner/api/files/document/print/en/qanda_22_2712/QANDA_22_2712_EN.pdf).
91. Staunton C, Shabani M, Mascalzoni D, Mežinska S, Slokenberga S. Ethical and social reflections on the proposed European Health Data Space. Eur J Hum Genet. 2024;32(5):498-505 (https://doi.org/10.1038/s41431-024-01543-9).
92. Marelli L, Stevens M, Sharon T, Van Hoyweghen I, Boeckhout M, Colussi I et al. The European health data space: Too big to succeed? Health Policy. 2023;135:104861 (https://doi.org/10.1016/j.healthpol.2023.104861).
93. Cheah PY, Piasecki J. Data Access Committees. BMC Med Ethics. 2020;21(1):12 (https://doi.org/10.1186/s12910-020-0453-z).
94. da Silva CFA, Virgüez E, Eker S, Zdenek CN, Bergh C, Gerarduzzi C et al. The future of scientific societies. Science. 2023;380(6640):30-2 (https://doi.org/10.1126/science.adh8182).
95. García-Alegría J, Garrido-López P. The role of scientific societies in a post-COVID world. Rev Clin Esp (Barc). 2021;221(8):468-9 (https://doi.org/10.1016/j.rceng.2021.04.004).
96. Montori A, Onorato M. Why there is a need of an ethics committee in scientific medical societies. Dig Dis. 2008;26(1):32-5 (https://doi.org/10.1159/000109383).
97. Regulatory considerations on artificial intelligence for health. Geneva: World Health Organization; 2023. Licence: CC BY-NC-SA 3.0 IGO (https://iris.who.int/handle/10665/373421).
98. Fact Sheet of AI Safety in Japan 2024. Japan: J-AISI -Japan AI Safety Institute; 2025 (https://aisi.go.jp/assets/pdf/j-aisi_factsheet_2024_en.pdf).
99. Whittaker R, Dobson R, Jin CK, Style R, Jayathissa P, Hiini K et al. An example of governance for AI in health services from Aotearoa New Zealand. NPJ Digit Med. 2023;6(1):164 (https://doi.org/10.1038/s41746-023-00882-z).
100. Maslej N, Fattorini L, Perrault R, Gil Y, Parli V, Kariuki N et al. Artificial Intelligence Index Report 2025. Standford University: Human-Centered Artificial Intelligence; 2025 (https://doi.org/10.48550/arXiv.2504.07139).
101. WMA Declaration of Helsinki – Ethical Principles for Medical Research Involving Human Participants. World Medical Association; 2024 (https://www.wma.net/policies-post/wma-declaration-of-helsinki/).
102. Nickel PJ. The Prospect of Artificial Intelligence-Supported Ethics Review. Ethics Hum Res. 2024;46(6):25-8 (https://doi.org/10.1002/eahr.500230).
103. Kolstoe S. AI could transform ethics committees. The Conversation. 2024 (https://theconversation.com/ai-could-transform-ethics-committees-224424).
104. Sridharan K, Sivaramakrishnan G. Leveraging artificial intelligence to detect ethical concerns in medical research: a case study. Journal of Medical Ethics. 2025;51(2):126 (https://doi.org/10.1136/jme-2023-109767).

World Health Organization
20 Avenue Appia
CH-1211 Geneva 27
Switzerland
https://www.who.int