Global Forum on Bioethics in Research

Ethics of artificial intelligence in global health research:

Cross-cutting themes

Cape Town, South Africa
29 and 30 November 2022

W: www.gfbr.global **E:** gfbr@who.int

The Global Forum on Bioethics in Research (GFBR) convened in Cape Town on 28 & 29 November 2022, to explore the topic of '**Ethics of AI in global health research'**. The Forum brought together 87 experts from 31 countries, with a focus on the low- and middle-income country (LMIC) context where AI has the potential to address critical skills shortages and improve access to care, but where the ethical challenges are made harder due to existing disparities in infrastructure, knowledge and capacity.

A series of case studies and governance papers were presented and used as the basis for wide-ranging discussion in five themed sessions. This policy overview is structured around the cross-cutting issues that emerged throughout the full meeting. A separate report summarises the meeting discussions, session by session, and the range of views that were expressed in response to the presentations. The full case studies and governance papers can be found on the GFBR website.[1]

The challenge of definitions

- The Organisation for Economic Co-operation and Development defines an AI system as: "**a machine-based system that can, for a given set of human-defined objectives, make predictions, recommendations, or decisions influencing real or virtual environments. AI systems are designed to operate with varying levels of autonomy**"[2].
- 'AI systems' refers to any AI-based component, software and/or hardware, on the basis that AI systems are usually embedded as components of larger systems, rather than being stand-alone systems. **An AI system can process far greater quantities of data with which to assess patterns and correlations on a broader scale than would otherwise be possible**. AI systems can be powered by a number of different techniques, for example, machine learning, deep learning and artificial neural networks. Systems can be either autonomous or semi-autonomous. **For the purpose of this report we refer to 'AI' and 'AI systems' without necessarily specifying the underlying technique**.
- **Challenges regarding definitions, language, concepts and terminology** in the field of AI ethics were recognised at the meeting, as well as the potential impact that framing can have on the ethical questions that are asked and the solution proposed.
- Despite this being an area of ongoing work, a vast amount of guidance has been published on 'AI ethics' broadly.

AI systems and health research

- AI systems have many applications in health research, including:
  - **Basic research** e.g. drug discovery, protein folding predictions, use in genomics, vaccine development.
  - **Clinical research** e.g. to develop AI-based tools for screening and triage, diagnosis, prognosis, decision-support and treatment recommendation, and to manage clinical trial design and conduct, for example pre-screening and identifying suitable patients and analysing trial data in real-time.
  - **Public health research** e.g. AI-based tools to monitor and predict the spread of an epidemic or monitoring and assessing population health, and targeting public health interventions.
  - **Health systems research** e.g. to assess and refine delivery and access to health services.
- For each of these health research purposes, AI could have the role of an:
  - "**algorithm for discovery**" (e.g. being used as a method to generate hypotheses or answer research questions such as discovering associations in population health data that reveal a new disease group or discovering potential drug candidates) or

[1] www.gfbr.global/past-meetings/16th-forum-cape-town-south-africa-29-30-november-2022
[2] OECD/LEGAL/0449 (2019) Recommendation of the Council on Artificial Intelligence

  - "**algorithm for intervention**" (e.g. being used as a component of an intervention, the impact of which is being examined in a particular setting such as a clinical screening tool which generates an output for human consideration).

AI in health research: do we need to re-imagine research ethics?

- The use of AI in health research presents **fundamental new challenges due its scale, complexity, potential impact and the range of actors involved** (including computer scientists, model developers, researchers and significant involvement of the private sector). The **global power inequalities** raise concerns that benefits will mainly be seen by developers in the west while **risk of bias** could actively harm 'out' groups and automation can have unforeseen impacts downstream.
- **AI is stretching paradigms of research ethics** (e.g. informed consent, privacy, the role of research ethics committees (RECs), data governance) leading to questions of whether the paradigms can be adjusted to meet current needs or whether new paradigms are required. **In the digital realm, borders and physical distance are no longer sustained highlighting society's global interdependence and the importance of inter-connectivity**. As society becomes more interdependent, relationality is an increasingly important foundation of ethics, but consent, autonomy etc. are founded on individual rather than collective interests. **Solidarity may be a more appropriate principle to guide the use of AI in health research and to guide re-consideration of research ethics in this field**. Relationality also invites us to create the space and opportunity for harmonisation of principles and approaches to how AI in health research is governed.
- However, **AI emerges out of a reality of existing injustices around data and power imbalances**. Data are neither good nor bad but they gain value and a normative worth from perspective and in their use (e.g. by the data subject, by a researcher, a private company, a government). The use of AI accentuates existing inequities and heightens the challenge of whether the interests and values of different stakeholders can be dovetailed in a fair and ethical way.

Power

- Issues of power feature heavily in the field of AI- based health research, raising principles of equity and fairness:
  - **The power of AI itself and the allure of AI solutionism**, which may divert from other strategies (e.g. addressing underlying public health needs). Also the power of 'black box' AI when the internal workings aren't visible to the user or other interested parties.
  - **The power of data** and the drive to use more and more data in meaningful ways. Once data is translated into uses through AI there's a potential to lose the ability to scrutinise and be transparent which are important characteristics of ethical research.
  - **The power of institutions** – in particular private big data institutions – and how this is being negotiated in research relationships and governance frameworks.
  - **The power of high income countries in setting research prioritise and driving the AI agenda** and also shaping the ethical discourse, and who is excluded as a result.
  - **The powerlessness of individuals and communities** who may not engage because of concerns about AI, resulting in their systemic exclusion from research and developments of AI in health systems. The **reflexivity of researchers should be encouraged to ensure the AI research agenda is helping to address health inequalities viewed through a public health lens** rather than creating small scale technical fixes.
  - **The relative powerlessness of some researchers** working in this field to navigate the ethical dimensions of their work. **Funders have a role in incentivising the**

ethical discussions and empowering researchers to undertake genuine community engagement and upholding a commitment to transparency.

Necessity of AI

- The promise of AI for health is significant but **critical reflection is required on whether it is the best or necessary approach to any given problem**. Ideally, **communities should be involved in defining the problem and solution and work should be done to understand whether there are alternative (non-tech) solutions**. Translation of ethics into policy requires policymakers to be supported by technologists and researchers to understand AI because understanding is a pre-requisite for good governance. **Tools are also needed to help policymakers evaluate the need and impact of AI in their jurisdiction** and to achieve the careful balance between encouraging innovative AI-based health research and appropriate levels of participant protection.

Transferability

- **Inequity can arise due to the poor transferability of AI systems** (e.g. where training data is derived from one context is unrepresentative of the context in which it will ultimately be used). These **issues can be compounded in the context of data poverty** and the difficulty of deriving local data sets in many LMICs.
- AI systems with good technical performance can fail to address a task in a meaningful way, or at worst can **cause harm when the AI system fails to generalise to a new setting**. Ethics requires the **testing of AI systems in a real-life context and on relevant data sets to demonstrate the effectiveness and benefit of the system to local populations**. In the context of AI-based tools intended to support clinical decision making, a ***silent trial*** can provide evidence of effectiveness in local populations by running inference on active cases and making predictions seen only by the research team.
- **Rigor and transparency are required on how an algorithm was trained**. The provenance, creation and use of machine learning data sets should be well documented e.g. regarding who and which groups they are made for and what data were included and not. This is essential to understand the data and the appropriate application of the AI system.

Bias in training data

- **Bias in data sets can arise for a number of reasons, including poor data collection methods, incomplete or inappropriate data or individual or cultural values or biases** (gender, socio-economic status, caste etc.). **Language can also be a source of bias**, given that most AI systems are developed in English and with an English/American bias in their functionality**.** Questions need to be raised as to whether and how an AI system might be biased or unfit for use in any given context and attention should be paid to the possibility that the tool is learning from incomplete or inappropriate data that might perpetuate bias that is already present in research.
- In the context of AI-based health research that relies on data collected from personal digital technology (e.g. mhealth apps on mobile phones), **access to technology and the internet can result in the exclusion of some groups and bias data**.
- An algorithm being developed to help select participants for a clinical trial on depression could digitise the exclusion of people with the most severe symptoms if they are excluded from the training data set. In turn, algorithms that are trained to enrol participants who are less likely to drop out of a study (e.g. those who struggle to adhere to the trial protocol) could result in the intervention having a higher success rate in the clinical trial context than could be achieved in a real-world setting.

Transparency and engagement

- There are multiple methods and reasons for undertaking engagement activities e.g. from consultation to co-creation. **Engagement is an important aspect for all health research but is especially warranted for AI-based research** (e.g. because AI is an emerging technology that uses significant amounts of data (often non-health data) and the fact the tools and infrastructure required bring in many different players to the research process, including commercial players). **The method of engagement should be tailored to the context and the purpose for engagement should be transparent.** Local partnerships are important to ensure that **language and methods used to explain AI systems are appropriate to the sociocultural context**, though many local languages do not have the words to adequately describe AI.
- **Understanding the local context is key, requiring multi-disciplinary teams and the involvement of social scientists and anthropologists** from early in the research process to better understand the local priorities and preferences. **Engagement should not merely be about education or sensitization but should aim to identify the interests of all stakeholders involved** (e.g. doctors, community etc.) to ensure development of contextually relevant tools.
- **In the context of implementation research, co-creation processes with end-users can foster a social understanding of AI, address power imbalances** by including a wide range of stakeholders and **embed user feedback** in the adoption of AI systems.

Governance

- The diffuse nature of AI and the speed of adoption has outpaced AI's governance. Challenges for the governance and regulation of AI in health research include:
  - How to **balance the needs for innovation in AI-based health research and potential pitfalls** e.g. quality of data, risk of bias.
  - The 'grey area' of **what counts as research and so what falls under research ethics frameworks** e.g. whether development of a health app counts as research and falls under health research and medical device regulation.
  - How **power and resources are distributed** in this field, which is characterised by a broader range of actors than found in standard health research.
  - Issue of **accountability** in the chain of training and implementing an AI algorithm: who 'owns' the training data and who is responsible for the algorithm's outputs?
  - **Scale and scalability** of AI and unintended consequences of amplification.
  - Problems of **generalisation and transferability of algorithms** between context, potentially introducing issues of bias and the need to collect new data to train the algorithms for different contexts.
  - How to take account of **under-representation as a serious harm** in the data used for increasingly common purposes.
  - **Understanding and listening to people and/or groups who choose to be excluded**.
  - Whether there are **other ways of conceptualising rights – in solidaristic ways – given that** AI is about group rights as well as individual rights.
  - **Avoiding over-regulation** that inhibits the field (e.g. legislators having too much discretion).
- **There are many pieces to the governance jigsaw (e.g. national regulation; international, national, funder or institutional policies and guidance, along with structures and processes for decision making involving RECs, access committees. etc.).** Currently, the European Union's AI Act is the only specific hard law for AI whereas there is a plethora of soft law – e.g. guidelines – which can be nimbler, and quicker to respond to areas of emerging technology. Finding the optimal balance between hard and soft law can be a challenge.

- **Law alone is not enough: ethics and human rights need to inform thinking about where and how concerns around AI can be baked into the broader governance system**. For example, whose role is it to consider the quality, quantity and representativeness of the data and assess issues of bias? How can technology developers and researchers be encouraged to detect and fix bias?
- There is also a **role for self-regulation and encouraging culture change across the governance eco-system. International and regional organisations, governments and funders must take responsibility for creating a research environment that promotes transparency, openness and dialogue on emerging technologies like AI.** AI developers should receive training in research ethics. **Incentives and investment are required for validation of AI systems and to promote the use of data versioning, data transfer agreements, algorithmic impact assessments and data destruction policies**.
- For a fast-moving field like AI, it's important to be creative and **leverage existing systems** to create a culture of responsibility and accountability **– e.g. continuous education, professional adherence**, **codes of conduct** – rather than creating completely new mechanisms.

Data governance

- AI systems are built on data and the ethical issues around their use reflect debates about data use and sharing in health research. Countries have diverse approaches to data management and protection which tend to be generalised and not specific to AI. **The complicated intersection of national and regional laws, the amount of data required to train AI, data poverty in many countries and the importance of representative data all add to the challenges for how data governance is approached.** WHO's report on the 'Ethics and governance of AI for health' addresses these challenges but capacity building is required in-country to facilitate local implementation of the report's recommendations (including policymakers, RECs and developers of AI).
- Large-scale, collaborative data resources can play a key role in supporting AI-based health research. Data sharing has been recognised for its potential to increase scientific efficiency by maximising the availability and utility of data but **the scale and complexity of how (and how much) data is used to train algorithms can also accentuate issues of privacy, security and consent**. Other challenges for large-scale data resources and data sharing include:
    - **Navigating country-specific rules around data collection, use and storage and moving data between countries**. For example, countries have different approaches to the secondary use of data and whether REC review is required, depending on the context e.g. how specific or broad the original consent was and if data is anonymised.
    - **Data sovereignty and benefit sharing**, especially in Global North and Global South collaborations.
    - **Data storage for LMICs** and the integration of data sets in different locations.
    - The **balance of jurisdiction between supra-national bodies (e.g. African Union, European Union) and the role of national governments**.
    - That **existing legislation (e.g. data protection) may not be directly relevant to AI-based health research** and may not consider data collected, processed or shared in this context.
    - The **interplay of power and decolonisation**. The use of AI in health research can be seen as **recolonising research as many LMICs are data-poor** and do not have sufficient capacity to hold, store and analyse data. In the field of AI, data is primarily held by HICs and by private companies.
    - Understanding if **there is a moral difference between data-sharing to train AI as opposed to other secondary uses of data for health research**. This distinction

might lie in the potential commercialisation of the data, or power of the organisation that owns the AI, or use in terms of the products being developed through this approach.

Fair partnerships and finding responsible methods for sharing data

- Power dissymmetry in international research partnerships can be heightened in the context of AI given that the majority of AI systems are developed in high income countries (HIC). **Power imbalances can result in challenges in science equity regarding the ownership of technology and data**. Learning from initiatives like H3Africa, which focuses on genomic research, data sharing partnerships for AI-based health research should think about how the idea of the 'data commons' underlies the objective of data sovereignty, who controls the data, and about benefit and the role of local intellectual leadership. Issues of **structural injustice should be considered** and what mechanisms already exist locally – in the country or region – for collecting data, ensuring data sovereignty and strengthening governance locally.
- **Data sharing can be constrained by the parameters of data systems**. New systems – such as data trusts – are emerging, especially in HICs. These virtual research environments allow data to be stored in one place, often publicly funded and stewarded by a data access committee. The trusts facilitate the sharing of anonymised data to allow working on shared data statistics. **Approaches that allow remote analysis of data held in third party country databases may help address disparities in computational power for researchers who don't have the resources to download and analyse data locally**.
- **Material transfer agreements are key to defining the terms of data sharing** and are a requirement in many LMICs and HICs.

Benefit sharing and the social value of research

- Fair benefit sharing models are required, especially in public-private partnerships. As recommended by the WHO report, **governments, research institutions and universities involved in the development of AI technologies should maintain an ownership interest in the outcomes so that the benefits are shared and are widely available and accessible, particularly to populations that contributed their data for AI development**. However, it is unclear who should decide what counts as a benefit (the institution, REC or the community) and who should have the role of ensuring that benefit is baked into a research proposal and delivered.
- There could be a **role for governments to negotiate benefit sharing terms with technology partners and set the boundaries for acceptable practice**. Governments could also incentivise – and hold accountable – public-private partnerships to look at health issues that may not otherwise be of interest as the research will generate no profits (similar to efforts to incentivise drugs for neglected diseases).
- **Algorithmic Impact Assessments should be used to evaluate the social value of research** involving AI systems, along with potential **risks and benefits**. However, does a requirement for social value change the nature of research (e.g. exploratory research) and where in the governance system should this be considered? (e.g. RECs could evaluate the assessments, but does this involve changing their role?).

Consent

- **The approach to consent will depend on context** (e.g. the data type, how data was collected, from whom, by whom, for what purpose(s)) and jurisdictional differences regarding what data falls under data privacy and research regulatory requirements. However, **there are ambiguities.** For example**: Should the development phase of an AI system be characterised as research and as falling within research regulation and requiring consent?** There is also the potential to **conflate privacy protection consent and**

research ethics consent (e.g. where data are collected from personal mhealth apps via Terms and Conditions and subsequently used for research).

- **The complexity of AI systems and lack of familiarity with core concepts and terminology can pose a number of challenges for consent** including how understanding and comprehensibility will be managed, and determining the appropriate level of disclosure regarding how the technology works, who will have access to data, how it will be used and identifying risks (e.g. of re-identification). Even where consent is acquired, it may be insufficient to compensate for the **power dissymmetry between the collectors of data and the individuals who are the sources**, especially in the context of commericalisation of AI systems.
- **Some people may choose not to consent due to privacy or other concerns**, resulting in their exclusion. Who is being excluded as a result of this, and **what does this systematic exclusion mean to equity and for principles of participant engagement**?
- There may be circumstances in which **consent should be disaggregated for specific purposes within a research project**: for example, consent to collect data to train an algorithm to help with recruitment into a clinical trial being distinct from the consent required to participate in the actual trial.

Research ethics committees

- RECs can play an important role in the governance of AI research by asking the right questions. They could have a role in:
  - Reviewing the **study design** and **asking if AI is necessary and the best solution** in the given context.
  - **Considering social value, benefit sharing, whose interests are being served and what are the costs**, including opportunity costs for taking an AI approach (e.g. by requiring and reviewing an algorithmic impact assessment).
  - Ensuring the research has taken account of – and is appropriate to – the **local context and needs**.
  - **Assessing bias and the selection of the study population** to promote inclusivity, diversity and relevance of the data or at least ensuring that issues of bias have been addressed.
  - Ensuring that **appropriate data sharing agreements and policies** are in place.
  - **Overseeing how algorithms behave in real-world context and considers the longer-term implications of health research**. This would be a significant new role for RECs which generally do not consider long-term impact.
- However, REC's role have not kept pace with the pace of AI adoption. To be effective, **RECs' capacity to evaluate protocols involving AI must be strengthened through training (e.g. on how to assess bias) and through access to experts in AI**. Or there should be a **mechanism for independent verification of the technical aspects of AI-based research** which can be submitted as part of the REC review. Even so, the feasibility of RECs taking on all these roles can be compromised due to limits on time and resources.
- Some of the issues – like **assessing algorithmic bias** – often only come to light during implementation. Therefore, it would be difficult for a REC to assess this and it may be better assessed by another body e.g. akin to a data safety monitoring board.

Environmental impact assessment

- **It is the duty of researchers – and everyone involved in the research ecosystem – to consider and reduce environmental impact of health research**. **Research ethics frameworks for AI health research should integrate environmental and social sustainability considerations** and these should be used by researchers, RECs and policy-makers to engage with questions about where and how data are stored and how algorithms can be optimized for environmental and social considerations. Such a framework should include:

- An **assessment of risk and benefits to the environment**, weighing these against the potential health benefit to individuals/communities.
- **Being attentive to the adverse environmental impacts that can emerge from using digital technologies during research and taking steps to reduce them.**
- **Researchers being aware of the composition of datasets** they use, and any possible biases, and focus on benefit sharing of research outcomes.
- Consideration of the **environmental justice issues associated with those involved in the manufacture, use and disposal of digital tools used during the research process.**
- **Risk/benefit considerations for AI research to include those affected by the manufacture, use and disposal of digital products**.