Global Forum on Bioethics in Research

Ethics of artificial intelligence in global health research:

Meeting report

Cape Town, South Africa
29 and 30 November 2022

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

CONTENTS

Executive Summary

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 to discuss case studies and governance papers relating to five broad themes. The Forum focused 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. This report summarises the meeting presentations and the range of views that were expressed, while a separate **policy overview** draws together cross-cutting themes from across the five sessions. The full case studies and governance papers can be found on the GFBR website and are linked in this report.[1]

- **Theme 1: Emerging issues in research methodology.** This session focused on new research methodologies involving AI. The first paper described the use of a 'silent trial' as a method to evaluate the effectiveness of an AI tool that is destined for use in a healthcare setting. The second paper presented plans for a digital clinical trial and the potential use of AI to aid the selection of participants. Both papers highlighted that the data used to train the algorithm and the context in which the AI will ultimately be used are critical factors in determining the effectiveness of machine learning and AI tools.

- **Theme 2: Importance of local context and engagement when developing AI tools.** This session focused on the implementation of AI technologies for health. The papers raised a number of issues, including: Who is involved in decisions around what an AI technology looks like when being introduced in a new environment? Who gets to influence the process through which the technologies are introduced and the way the technology becomes used in a local environment? These issues in turn relate to the distribution of power in the implementation environment, who is making investments in technology, disruptions to the work process of health care workers and the need to engage community members and patients to inform the process of the implementation research. The papers also raised methodological issues and fundamental questions about if and when the implementation of AI technologies should be considered 'research'.

- **Theme 3: Collaborative initiatives and data resources to support AI health research.** This session showcased two collaborative initiatives: one that is at the planning stage and one that is more mature with a portfolio of research activities mainly focusing on autonomous systems. Both papers focused on the issue of trustworthiness and were complementary. One initiative is planning how to establish its processes and how to engage communities, while the other is underway and has faced challenges and created best practices. The theme raised the question of how transferable regulatory requirements are from one context to another. For example, is it feasible or indeed right for the EU's General Data Protection Regulations (GDPR) to be enacted and enforced in LMICs? The theme also delved into how responsible, collaborative data sharing practices can be encouraged between public and private organisations in a way that aligns with community values and needs.

- **Theme 4: Regulation of data for health research involving AI.** The governance of AI is inextricably linked to the governance of data, on which algorithms rely. In its report 'Ethics and Governance of AI for Health', the WHO makes recommendations on data governance. The report recognises the

[1] www.gfbr.global/past-meetings/16th-forum-cape-town-south-africa-29-30-november-2022

complicated intersection of national and regional laws and that the amount of data required to train AI, data poverty in many countries and the importance of representative data all add to the challenges in how data governance is approached. This session focused on data governance, specifically in the African context. The first paper analysed the governance of cross-border transfer of data between countries. The second paper evaluated Uganda's data governance regulation against the recommendations in WHO's report.

- **Theme 5: Issues associated with research ethics frameworks and ethics review.** This session focused on the policy context and research ethics frameworks for AI based health research. Two presentations assessed the policy environment in Egypt and Kenya. The third presentation focused on the incorporation of environmental impact assessments into research ethics frameworks. Gaps in REC capacity were also highlighted, along with consideration of how the gaps might be filled and what other committees or oversight mechanisms might be needed.

Introduction

The Global Forum on Bioethics in Research (GFBR) convened on 28 & 29 November 2022, to explore the topic of '**Ethics of artificial intelligence in global health research'**. Artificial intelligence (AI) is increasingly being used in global health research but frameworks, policy and best practice for the ethical development, conduct and oversight of AI heath studies is currently lacking.

The meeting brought together the global bioethics and research community and others to explore ethical challenges including bias, privacy, data provenance and ownership, along with the need for transparency, and engagement during the design and use of AI in global health research. Discussion included the use of AI as part of the research process (e.g. to aid the selection and screening of participants) and research on AI-enabled tools that are destined for use in a healthcare setting. To date, these discussions have predominantly taken place in high-income countries (HICs), and low- and middle-income country (LMIC) perspectives and experiences have been underrepresented.

This report contains extracts from the papers presented at the meeting, and the discussion. The papers were predominantly from LMIC contexts. We thank the presenters for their work. The full case studies are available on the GFBR website.

Case studies and governance papers were invited through an open application process. An international, expert Planning Committee[2] selected the speakers and structured the meeting around the following themes:

- **Theme 1 Emerging issues in research methodology**
- **Theme 2 Importance of local context and engagement when developing AI tools**
- **Theme 3 Collaborative initiatives and data resources to support AI health research**
- **Theme 4 Regulation of data for health research involving AI**
- **Theme 5 Issues associated with research ethics frameworks and ethics review**

[2] Joseph Ali, USA; Caesar Atuire, Ghana; Niresh Bhagwandin, South Africa; Phaik Yeong Cheah, Thailand; Judy Gichoya, USA; Armando Guio, USA; Daudi Jjingo, Uganda; Katherine Littler, Switzerland; Tamra Lysaght, Singapore; Daniela Paolotti, Italy; Jay Shaw, Canada; Effy Vayena, Switzerland

With experts from 31 countries (see map of GFBR participants' countries), the meeting delved into the key ethics and governance issues facing the use of AI in global health research.

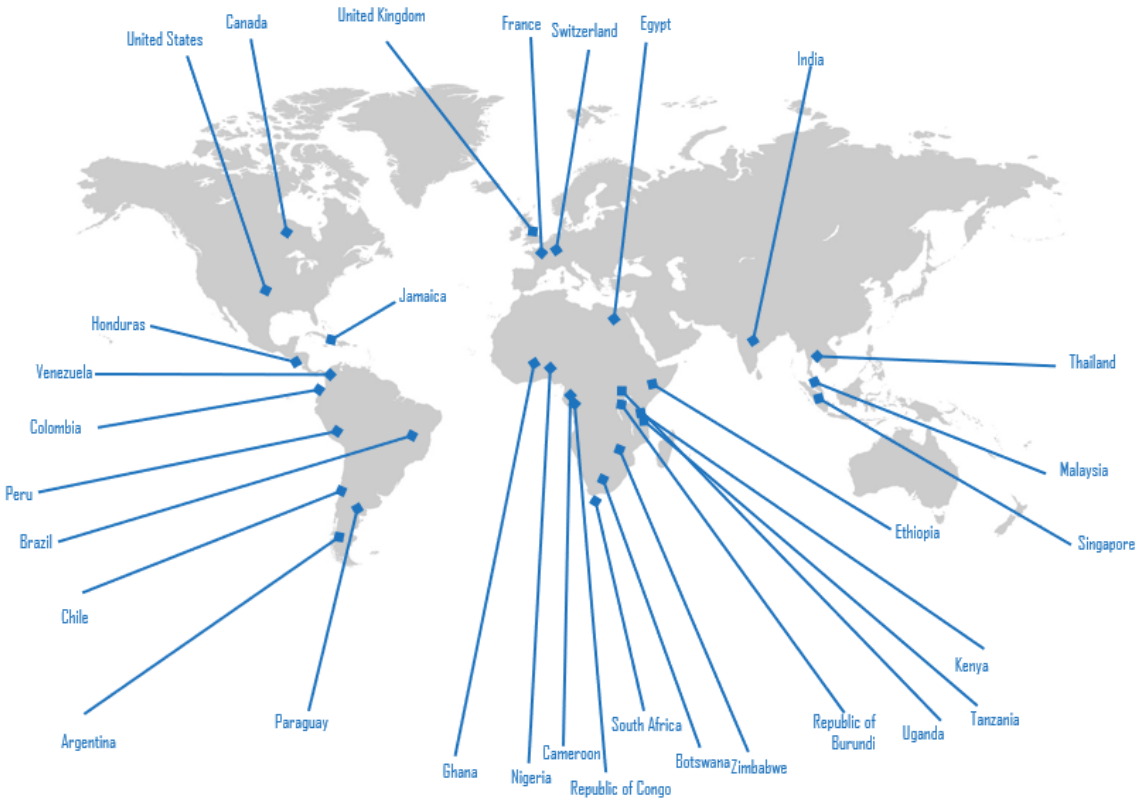

Figure 1 GFBR participants: The 87 participants from 31 countries brought a wide range of expertise to this important topic, including: bioethics, research ethics, artificial intelligence, bioinformatics, policy, regulation, health researchers, and funders, at all career stages. 72% of participants were from LMICs.

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**"[3].
- '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**.

[3] OECD/LEGAL/0449 (2019) Recommendation of the Council on Artificial Intelligence

Keynote

Effy Vayena's (Health Ethics and Policy Lab, Institute of Translational Medicine, ETHZ, Switzerland) keynote talk situated GFBR in the broader context of AI ethics. **Effy** highlighted the **ongoing work around the definitions, language, concepts and terminology in the field of AI ethics and the proliferation of guidance**. A review in 2019 identified 84 high-level guidance documents at international and institutional level on the ethics of AI. Within this intense debate on the ethics of AI, several approaches have been proposed (e.g. human rights based, ethics principles based and the creation of tools). Some have criticised the current approach, highlighting the **gap between principles and technological practice and the need to think more carefully about who is doing AI ethics**. For example, most of the guidance documents were developed in the Global North, many of the groups who developed the guidance included representatives from the private sector, and most guidance did not include civil society, citizens and others who should be involved. With accusation of 'ethics washing' in the field of AI (i.e. talk but no action) **it has been argued that self-regulation through guidance has failed and that hard law is required to enforce the principles** and ethical thinking. This raises questions about AI exceptionalism (is AI is being treated differently from other technologies?) and how to regulate an emerging technology that is advancing so quickly.

AI investment in health is growing globally. Yet despite the wealth of guidance on AI ethics there was no specific guidance on AI for health, until the **WHO's 2021 report** '**Ethics and governance of AI for health**'. The work was motivated by the potential of AI to act as an enabling factor (e.g. for people to benefit from Universal Health Coverage and for better protection from public health emergencies etc.) and the hope that technology can bring about good and improve the lives of people who need it most. This hope was balanced by the **recognition that humans must exercise control over AI applications, think about risk mitigation and the distribution of benefit, and recognise the need for norms around use of data and AI technology**. While targeted at health more broadly, the WHO recommendations are generic enough to apply to the use of AI in health research and address issues such as the digital divide, accountability and responsibility, bias, discrimination, commercialisation of products and sustainability.

The WHO guidance describes **six principles** and how they should apply to the health context:

- Protect autonomy
- Promote human well-being, human safety and the public interest
- Ensure transparency, explainability and intelligibility
- Foster responsibility and accountability
- Ensure inclusiveness and equity
- Promote AI that is responsive and sustainable

The guidance recommends **not to separate AI from data**: while they are distinct areas when it comes to ethical questions they are inter-connected. Fundamental questions remain around health data ethics, despite many years of work in this field (e.g. issues of data-sharing and fair collaborations given resource inequities around the globe). Although we are in the era of collecting vast amounts of health data, **health data poverty is a significant issue in many parts of the world**. The WHO report provides **guidance on data governance and makes specific recommendations** about what countries and governments should do and what mechanisms should be in place (e.g. data protection, the role of independent authorities, commitments to transparency and community oversight).

The guidance also addresses the role of **private actors** who have the know-how, **public private partnerships** and questions around **intellectual property rights**. It recognises that tools need to be developed or ways found to ensure that development and deployment of AI in health is done in appropriate ways (e.g. impact assessments). The report has a strong focus on **health equity, and highlights outstanding questions regarding the social value of technology, data use and sharing, how to validate technologies and explainability**. More work is also required on the question of **what kind of evidence should be generated** in order to know that technology is working and is effective.

There is an emerging consensus about AI ethics and acknowledgement that a mix of soft law and hard law will be required, but this is a work in progress. The field has many stakeholders and **the responsibility for AI ethics has to be distributed and needs global co-operation and meaningful engagement with the public**.

1. Emerging issues in research methodology

This session focused on new research methodologies involving AI. The first paper described the use of a 'silent trial' as a method to evaluate the effectiveness of an AI tool that is destined for use in a healthcare setting. The second paper presented plans for a digital clinical trial and the potential use of AI to aid the selection of participants. Both papers highlighted that the data used to train the algorithm and the context in which the AI will ultimately be used are critical factors in determining the effectiveness of machine learning and AI tools.

Presentation 1: A silent trial is critical to accountable and justice-promoting implementation of artificial intelligence in healthcare
Melissa McCradden – The Hospital for Sick Children, Canada

The silent trial refers to the deployment of a model in the anticipated clinical environment, where the model is running inference on active cases and making predictions – however, these predictions are seen only by a research team and do not influence patient care. The predictions are recorded and compared to the true clinical outcomes or human-defined labels that they are predicting (e.g., a radiologist's confirmation of a given diagnosis). See Melissa's presentation for details of a case study trialing a classification model to identify obstructive hydronephrosis in infants.

Melissa McCradden briefly described existing guidance on responsible AI and machine learning practices which focuses on developing a good, reliable algorithm that performs in relation to a specific task (e.g. making a diagnosis). Initiatives include:

- The US Food and Drug Administration and the UK National Institute for Health and Care Excellence guidance on digital health technologies.
- Reporting guidelines for clinical trials that involve AI (e.g. DECIDE-AI, CONSORT-AI and SPIRIT-AI).
- The development of quality assurance frameworks for AI tools to help with post-implementation monitoring.

However, despite the current guidance, **many AI tools don't work when implemented into a healthcare setting or don't address the task they are assigned in a meaningful way**. A recent systematic review of clinical AI tools found that only two fifths of those with good technical performance were also associated with an improvement to patient outcome.

Melissa explained that AI tools are most reliable when the population and environment on which they were trained is highly similar to that in which they are integrated. As such, **there are limitations on model generalisability and some models can fail to generalize to new settings**. She argued that demonstration of benefits to patients is crucial to the ethical integration of AI tools for both patients and healthcare workers.

A silent trial is a specific method that can fill the gap between responsible algorithm development and a prospective clinical trial. The **silent trial evaluates an AI tool by testing a mature algorithm in a real-world context to establish its effectiveness on the local patient population** prior to the model being used to inform patient care. **Melissa** argued that a silent trial is necessary because it provides the kind of information that can establish clinical equipoise, which is the justificatory basis on which interventional trials are considered ethically permissible. This evidence is missing from the responsible machine learning stage.

Recognising the potential for AI systems to reflect biases in society, which can result in systemic disadvantages to racialized, marginalized, and other oppressed groups, **Melissa** argued that the principle of distributive justice suggests two key steps must be taken:

- **Establishing the distributive benefits and burdens of a given system** e.g. conducting an algorithmic audit at the silent trial stage to characterize model performance across the locally relevant population and subgroups of patients.
- **Identifying opportunities for correction, revision, or redress**. This would involve reflection and engagement regarding the suitable options to address potential discrepancies found in that distribution.

In this way, **a silent trial can provide an empirical foundation for ethical decision-making** and leads to informed choices about moving forward – e.g. expanding involvement of different groups in the study and exploring what mitigations to make.

Additional opportunities of a silent trial were highlighted:

- **Patient, family, community engagement** e.g. about the best way to implement an algorithm in a given setting and to understand how patients might want to be communicated with.
- **Human factors analysis** to look at the processes that individuals use during clinical decision making.
- **Clinical utility analysis** to assess and better understand the factors involved in how to improve the clinical utility of an algorithm in a specific setting.
- **Visualisations and interface design.** Sometimes how predictions are shown can bias people to particular choices so it is important to explore the best ways to use an algorithm and visualise predictions to maximise patient benefit.
- **Identifying educational needs** e.g. of patients and families and clinicians who are using the tool.

Silent trials could be used in LMICs and HICs:

- **Prior to purchasing a system to enhance accountability** and protect the interests of patients against AI failure.
- As a **testing period to understand community relevance** and tailor implementation to local values, characteristics and work flows.
- To **understand impact on equity and equality** – e.g. by using an algorithm specifically to advance the interests of particular disadvantaged groups.

PRESENTATION 1: CLARIFICATIONS & DISCUSSION

Consent and engagement

Questions were raised over the governance of silent trials, and it was clarified that in Canada (the context for Melissa's talk), consent for silent trials falls under secondary use of clinically-collected data. This reflects the fact there is no intervention and the trial does not modify treatment decisions for patients. A REC review is required for all proposed uses of patient data and patients can opt out of their data being used for research. A silent trial protocol specifies the threshold required to demonstrate the AI tool's effectiveness, after which the research team would submit a new protocol to the REC for a prospective, post-implementation trial, with clear consent processes.

Some participants perceived a tension in the conceptualisation of a *'silent'* trial at a time when research is trying to de-silence many voices and focus on the importance of community engagement. From an ethics perspective, the practice at present – while legally consistent with Canadian legislation and research ethics guidelines – may not be ideal. However, *'silent'* does not mean *'secret'*, it just means not disrupting or influencing the current clinical process. There could be several models being trialled at any given time making it a challenge in an Intensive Care Unit to talk to everyone about every model, especially when the natural failure rate of the models is high. The relevant piece is that **patients and families need to know how their health data are used, including during the development and evaluation of AI tools**. In moving to a system of proactive engagement with patients and families, **Melissa** and colleagues have engaged the Hospital of Sick Kids Children's Council, along with other specific groups who may be more at risk with AI. The research team have drawn on these insights to understand degrees of concern. **By calibrating towards people who are more concerned about AI, the team hopes to create a system that is more ethically justifiable**.

Research Ethics Committees

There's a **need to build REC capacity to evaluate protocols that involve AI and for protocols to be written in a language that is understandable to the committee**. However, this is a difficult task considering that a lot of knowledge sits with machine learning developers and does not cross over to RECs. **Melissa** and colleagues have published a research ethics framework for the clinical translation of healthcare machine learning and are working to help RECs consider issues of bias and whether it has been sufficiently addressed in any given protocol. GFBR participants agreed on the need to **train RECs and also AI developers to increase their sensitivity to issues of privacy, harm and other risks**.

Accessibility of tools and infrastructure

GFBR participants asked about the necessary infrastructure for silent trials – especially in less high-tech settings. Access to the code or training data is not required to run a silent trial, which is essentially an observational non-implementational trial. If the AI tool in question could be implemented in a local context, then conducting a silent trial would also be feasible.

Research into practice: impact of AI tools on clinical decisions

GFBR participants agreed that a good decision for a patient takes into account more than the output of an AI tool. As such, **AI should be recognized as just one component of a larger picture in which the clinician**

remains responsible for making decisions about care. **Melissa** stressed the importance of providing guidance and support for clinicians on how to use a reliable and beneficial tool, with patients' interests central.

Comparator

Melissa clarified that a silent trial in her context is being compared to normal standard of care (SoC). However, existing standards can have their own biases (e.g. racial disparities in referrals for care). GFBR participants discussed whether **silent trials can be used to open-up what good care *should* look like and whether researchers can employ a values-based implementation strategy** rather than a strictly 'is this better or not than SoC' approach.

Presentation 2: The PSYLECT study: opportunities and pitfalls of digitizing a clinical trial in a LMIC
Andre Brunoni – University of São Paulo Medical School, Brazil

Transcranial Electrical Stimulation (tES) is a non-invasive brain stimulation technique with excellent acceptability and moderate effectiveness for major depressive disorder and could be a first-line intervention, especially in patients with low-drug resistance. The study investigated a portable tES that can be operated by patients themselves, reducing costs and enhancing scalability which could have significant advantages especially in contexts where there is a scarcity of skilled personnel. Also, it may ease the logistical burdens and costs associated with a patient's daily visits to the clinic, which impact negatively on the availability and utilization of tES in clinical practice. The randomised, sham-controlled clinical trial investigated the synchronous combination of portable tES and a remotely delivered, self-directed and internet-based behavioral intervention, for the treatment of major depressive disorder in adult patients.

Andre Brunoni's case focused on the *Portable Transcranial Electrical Stimulation (tES) and Internet-based Behavioral Therapy for Major Depression Study* (PSYLECT) in Brazil which involved digital recruitment, data collection and analysis. **The trial used social media and digital marketing strategies to advertise the trial and screened potential participants by telemedicine**, requiring fewer in-person visits. They received an increased level of interest and engagement using the online strategy but only 30% of screened patients met eligibility criteria. A further 30% were later excluded during the initial on-site interview because the online interview didn't capture some important information.

The research team faced a number of ethical issues:

- **Consent**: Screening data are collected to assess eligibility prior to the individual being invited to join the study and prior to them giving informed consent to participate in the trial. This data could be used to train an algorithm to enhance screening accuracy and inform recruitment. **The team recognised the need to consent for the purpose of using the screening data to train an algorithm, in addition to seeking consent from the eligible participants to join the trial.**
- **Equity and exclusion**:
    - People with severe mental health conditions are often excluded from research. If an algorithm is trained using data from less severe patients this could potentially **digitise and perpetuate the exclusion of people with severe mental health conditions when the algorithm is put to use**.
    - Only those who have digital devices can participate, which **risks digital exclusion**.
    - An algorithm that is trained to enroll participants who are less likely to drop out of the study could result in the exclusion of certain groups e.g. those who have a low level of digital literacy.

- **Effectiveness**: Ultimately, a selection process which excludes and eliminates people who are unlikely to be successful in the use of the tES tool means that the success rate of the trial will be higher than it would be in the real-world setting.
- **Standard of care**: The team grappled with the use of chatbots to address participant's questions. They recognised the tension between the benefits of improving scalability and the downside of participants potentially feeling alienated if they do not have prompt access to the research team, which could decrease adherence. In addition, they identified that **ancillary care obligations to participants could require greater – perhaps human – attention to issues not easily capturable by a chatbot**.
- **Collection of sensitive data**: Digitization of a trial brings the opportunity to enhance data granularity by collecting active and passive data using apps and wearables. But this also brings ethical issues regarding the extent of data collected and how the data is being used. Andre recommended that **participants and patients who use apps that collect sensitive personal data for research should be clearly informed regarding the extent and types of data being collected**.
- **Accountability**: The use of home mobiles to capture data can lead to problems (intentional or not) from misuse, overuse, lack of training or device malfunction, which is especially important when it concerns personal health data. This raises the question of **who is responsible**: the research team, the app developer, or the person whose device it was?

PRESENTATION 2: CLARIFICATIONS & DISCUSSION

Inclusion / exclusion

GFBR participants discussed the risks associated with excluding certain groups from the algorithm training data. For example, **the generalizability of the data may be impacted if only people with access to digital services and infrastructure can participate in the trial**. Conversely, **an online approach may have the benefit of addressing barriers of access** – e.g. access to health centres (which may be inaccessible due to distance or stigma etc.) and/or by expanding involvement outside business hours. Andre reported that the study recruited more women than expected, suggesting that digitisation made the trial more accessible to at least some members of that group. In response to the digital literacy issue, the team used traditional media (newspapers and radio) and scheduled on-site screening visits for those uncomfortable with the digital process of screening.

Inclusion or exclusion criteria will be determined by the goal of any given research project. There will be times when making data more diverse doesn't solve the problem and representation is not enough. For example, sometimes over-representation of specific populations will be required, in order to focus on their needs, rather than an algorithm being generalizable. In this context, the **'effectiveness' of the AI tool is being assessed in relation to the specific population on whom the algorithm was trained and will be used**. GFBR participants agreed that this approach requires rigor and transparency on how an algorithm was trained. **Data sets should be well defined e.g. regarding who they are made for (children or other groups) and what data were included and not**. This is essential to understand the data and the appropriate application of the AI tool.

As with all types of research, RECs need to be able to justify the recruitment approach and reach and the inclusion / exclusion criteria used.

THEME 1: DISCUSSION

Interdisciplinarity

GFBR participants agreed on the need for interdisciplinarity in research involving AI. A complementarity of skills is required e.g. software engineers working with clinicians to ensure the devices developed are effective in practice and relevant to the problem being addressed. There is also a role for social scientists to identify unheard voices, to avoid the risk of double burden of exclusion (e.g. of those who can't access healthcare, transport, mobile devices) and to promote reflexivity in the research process.

Language and culture as a barrier and source of bias

Most machine learning and AI tools are developed in English and there is an English/American bias in their functionality. GFBR participants noted that in situations when users are required to answer questions, if the questions are framed in a manner that is biased to certain cultures or languages, the responses provided may not be appropriate. This could impact on the subsequent analysis. **Researchers need to be alert as to whether and how the tool 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.

2. Importance of local context and engagement when developing AI tools

This session focused on the implementation of AI technologies for health. The papers raised a number of issues, including: Who is involved in decisions around what an AI technology looks like when being introduced in a new environment? Who gets to influence the process through which the technologies are introduced and the way the technology becomes used in a local environment? These issues in turn relate to the distribution of power in the implementation environment, who is making investments in technology, disruptions to the work process of health care workers and the need to engage community members and patients to inform the process of the implementation research. The papers also raised methodological issues and fundamental questions about if and when the implementation of AI technologies should be considered 'research'.

Presentation 3: Ethical issues associated with the development of an ear biometric tool for patient identification in Zambia
Alinani Simukanga – University of Zambia, Zambia

In 2005, an electronic management system was deployed in Zambia to coordinate the delivery of HIV care. Since its initiation 'SmartCare' has been expanded to track outpatient care, maternal/child health, tuberculosis treatments, and monitor the status of orphans and vulnerable children. SmartCare relies on a 'CareCard' as the primary identifier. Relying on the CareCard for identification has proven a critical limitation of SmartCare as they are easily lost, damaged, or used inadvertently by another individual and suffer from a high rate of technical error. As a result patients often accumulate multiple cards spreading their medical history across multiple unlinked aliases in SmartCare's database. Ear biometrics are being investigated as an identifier to integrate into SmartCare.

Alinani Simukanga described Project SEARCH (Scanning EARs for Child Health). The project uses ear scans to identify children, as an alternative to facial scans (which cause more privacy concerns and are more expensive) and iris scanners (which can frighten children). In field studies, the project's mHealth identification app (the SEARCH app) has achieved near-perfect subject identification among Zambian infants as young as six months old through biometric analysis of ears. The project aims to show the value of integrating SEARCH's biometric ID system with the SmartCare EMR (the electronic medical records system used in public health institutions in Zambia).

Alinani presented the ethical challenges associated with this work:

- **Data bias**: Publicly accessible datasets of darker-coloured ears captured in controlled conditions were not available in the early stages of the project. The team therefore used datasets collected at the Museum of Science in Boston, USA, to train the biometric tool. Tests conducted with the tool on a dataset in Zambia showed a drop-off in performance. **At risk of building a tool that does not work in local context, the team created a new data set of darker-coloured ears captured from young Zambian infants to train the tool**. Written consent forms were translated into two local languages, describing the use of the data. Where mothers showed concern about the technology, their child was excluded from the study and data collection.
- **Transparency and engagement**: Pre-implementation focus groups and interviews took place with mothers and health officials in Zambia's rural and urban settings to understand how receptive they would be to the proposed app. The engagement was an integral part of the grant applications as the team was keen to understand what issues might be later stumbling blocks. Issues raised included **problems with paternal consent which might be refused because of dislike of western technologies and also because of unwillingness to diverge from the status quo**.

Alinani concluded that further community sensitization is key in tackling the sociocultural issues that cause hesitance toward AI solutions. He proposed that responsibility for community engagement must be shared by researchers and the Zambian Ministry of Health, which has an important role in dispelling any fears that people might have towards AI tools. Researchers also have a very important role to play in clearly laying out the benefits and risks of the tool to the end users.

Presentation 4: Adherence vs agency: AI for behaviour change in health

Niyoshi Shah – Quicksand Design Studio, India

Human-centred design (HCD) was used to explore how an AI system called *Vajrahands* may be deployed for the quality improvement of hand hygiene in India's public health system. *Vajrahands* was installed in the labour ward of seven district and sub-district hospitals across three states to cover a broad range of work environments. It functioned in two parts: (1) A camera was set up at a selected basin to capture people's hand movements in real time — with no other identifiable information — for the algorithm to check if they had performed the nine steps of handwashing as recommended by WHO. This was supported by a display monitor where the staff got live feedback for each episode. (2) The data collected is aggregated day-wise and made available to the management on a dashboard for better monitoring and evaluation. The HCD process was used to understand the hospital's experience of both these components.

Niyoshi Shah described HCD as a creative process of problem-solving using co-creation. It begins with ethnography where designers learn about the needs, preferences, social context and constraints of their end users. This research is translated into provisional solutions that are tested on the ground early on and often to come up with a final product, service, system or strategy that is people-centred.

Ethical issues arising in the HCD process included:

- **Adherence and agency**: Staff had to un-learn their current hand-washing protocol and correctly perform the technique recommended by WHO in order to be marked right by the algorithm. The team worked on initial feedback to produce a new protocol in which staff simply had to follow a series of GIFs on the handwashing steps by the timer to get a perfect score. Making the interface more directive reduced the scope to have legitimate variation that people may practise when carrying out the WHO technique, raising questions about the balance between adherence and agency.
- **Soft coercion:** The hospital management received reports with data visualisation to give them a high-level summary of the ward's performance. Some used the reports to motivate their staff with fear, telling them that the ward's performance was being watched by the government.
- **Gaming the system**: A short competition cycle was introduced with rewards to incentivise the handwashing protocol. Some sites focused on getting their scores right instead of using the intervention as a learning exercise. They asked senior staff to use *Vajrahands* more often to balance out the day's compliance rate if it dropped or basins were closed to stop poor adherence being recorded.

Niyoshi explained that the Indian health system is highly hierarchical where development projects – like *Vajrahands* – may be sanctioned by the state without meaningful dialogue and buy-in of the implementing hospitals. This creates a fertile ground for soft coercion and number play as the decisionmakers who sit between the state and staff feel the pressure to meet programme outcomes. **Niyoshi** argued that research regulatory frameworks need to account for the political context and realities of implementation and that HCD can be used to foster a social understanding of AI and address power imbalances by including a wide range of stakeholders in its participatory approach.

Presentation 5: Feasibility, acceptance and ethical considerations of a mobile clinical decision support system in Botswana
Kagiso Ndlovu – University of Botswana, Botswana

The shortage of dermatology specialists in Botswana necessitates efficient use of the limited resources and continuous empowerment of those commonly engaged in the management of prevalent skin conditions. VisualDx is an AI driven mobile clinical decision support system. The system combines machine learning algorithms and vision science with a structured clinical knowledge base to allow non-specialist healthcare providers to capture patient-specific findings, build custom differentials, and view images and treatment recommendations. The DermExpert™ feature in VisualDx uses a Convolutional Neural Network to estimate diagnosis and lesion categories from an input image. In this study, DermExpert™ was tested by healthcare workers supporting dermatology clinics and medical students participating in dermatology coursework or rotations at health facilities and universities across Botswana.

Kagiso Ndlovu presented a collaboration between the University of Botswana and VisualDx (an international private organization) to assess feasibility and acceptance of a mobile clinical decision support system in Botswana called DermExpert™. The app involves taking picture of lesions which are analysed by AI drawing on a library of over 80 million images to make diagnostic suggestions. Feedback was gathered on the current app and improvements were identified that would better support health care providers in the region.

Kagiso identified a number of benefits for adopting AI-driven mobile clinical decision support systems in Botswana:

- Automation of tasks.

- Addressing a critical shortage in skilled medical personnel and the very low doctor/patient ratio.
- Improved access to services in remote areas.
- Empowerment of health professionals.
- Getting more value out of available data.
- AI as a 'technology balancer' to counteract cognitive and cultural bias.

Ethical considerations during the research implementation included:

- **Safety and transparency**: The system was peer reviewed with expert validated content and there was continuous sensitisation to upskill users for safe use. The AI model was non-prescriptive and there was federated learning across sites and second opinions.
- **Data privacy**: The system complies with data protection standards, and patient images were discarded immediately after analysis. Only de-identified and generalized demographic information about the patient was collected.
- **Algorithmic fairness and biases**: The model is trained with over 80 million images including different ethnicities and skin tones.

Kagiso concluded that the tool was well received by those who tested it despite the health system in Botswana not being ready for the technology. In-country there is a lack of quality data and the government is only just beginning to grapple with its regulatory role to establish 'good machine learning practices' and how to oversee how algorithms behave in real-world contexts (assessing bias etc.). This is particularly challenging given how quickly technology is evolving. Finally, he stressed that **health systems are unique in terms of their technology and care processes and that AI may not always be the best solution**.

THEME 2: DISCUSSION

Power relations and ownership

The case studies raised the issue of how power and resources are distributed in implementation research involving AI-based tools. GFBR participants noted that all three tools in the case studies were developed by the private sector. Generally, there is a lack of capacity in the public sphere for developing and maintaining these tools and where hardware / software is gifted to an organisation it is often with minimal support for maintenance in the longer term.

Stakeholder engagement

- **What do we mean by engagement?** There are multiple methods and reasons for undertaking engagement activities: from informing stakeholders to actively involving them in decision-making. **The method of engagement should be tailored to the context and the purpose for engagement should be transparent.** Engagement within the case studies focus on identifying problems and 'sensitizing' stakeholders to the AI tool. Some GFBR participants perceived a lack of scope for the engagement to change what was happening. For example, mothers of children who feared the biometric ear scan were excluded from the study. In the context of *Vajrahands*, the healthcare workers inputted into the design of the AI tool but they were not able to opt out of using it.
- **Who are the stakeholders?** This will be context specific and could include a range of people (e.g. people who provide data to train an algorithm through to clinicians who will eventually use the AI tool).
- GFBR participants identified a particular need for **community engagement** (e.g. involving the people who will be subject to the AI tool). Researchers should be imaginative and rethink **engagement in low**

resource contexts in sensitive ways. For example, the use of town halls, women's groups and other local networks and the use of radio. Community advisory boards could also advise on engagement (as they have for data science research). **AI tools need to be developed in local languages** instead of being in English, though many languages lack the necessary words to describe AI, making it hard to communicate.

- The call for community engagement is not specific to AI health research and applies to all research. GFBR participants warned against **AI exceptionalism but recognised that the nature of AI can bring added complications** given that it 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.
- **Role of researchers**: GFBR participants recognised the pressures on academic researchers (e.g. to publish) which may conflict with their ability to engage with stakeholders during the research process. Also, researchers may not be best equipped to carry out the engagement themselves e.g. on the level of dealing with fear of technology which people may find hard to articulate and could be linked with the local context. Realistically researchers need to be incentivised but also not expected to carry engagement responsibilities alone. It was noted that some research funders require public engagement as part of a grant and provide funding to support engagement specialists and activities.
- **Regulators and policy makers**: Many regulators and policy makers will be unfamiliar with AI, and so may not know the right question to ask about the risks and benefits of AI-based research. **Researchers and machine learning developers could play a key role engaging with and helping policy makers and regulatory bodies understand AI**. However, such engagement may be a challenge in the context of regulators in many LMICs being overwhelmed and not having the time or resources to engage.

Relevance of research to the local context

Understanding the local context is key, requiring multi-disciplinary teams and leadership – or genuine involvement – of local researchers. Involvement of social scientists and anthropologists from early in the research process is also important to better understand the local priorities and preferences.

Trust is required to bridge communication, especially when talking about such a complex, emerging technology. Some GFBR participants recommended engaging local leaders **who are trusted by the community** or advocacy groups who can be more effective at interactions locally because they understand the local dynamics. **Engagement should not merely be about education or sensitization but should aim to identify the interests of all stakeholders involved** (e.g. doctors, diverse parts of the community etc.). to ensure development of contextually relevant tools.

Some GFBR participants highlighted the issue of funder dependency and how local priorities may not align with the funder's strategy and purpose. RECs could check that research protocols engage with the importance of the local contexts and demonstrate how this will be taken into account throughout the research.

In AI for health research, what counts as research?

Niyoshi's project was granted a non-research determination and exempt from REC review because they were: (1) collecting non-identifiable data, (2) investigating a trend of public importance and (3) were working for the state (even though state may not always represent people's interests). Everyone who used the hand washing station had to use the device and there was no option to opt out.

GFBR participants discussed the question: '**when does the use of data to develop an AI tool becomes research use of data?**'. Some suggested this could relate to the level of risk or when there is a competition of interests (e.g. in human health data the competition between personal privacy and others' interests). Or it could relate to the data provenance (e.g. use of data from health records being viewed as research vs social media and public data that might be exempted by RECs). Some argued that if an AI tool is developed using data taken from humans, or humans are involved in implementation projects, the proposal should be considered as research and ethical clearance should be obtained. This should especially be the case if there is potential to create harm e.g. if anonymized data has the potential to be deanonymized.

However, others cautioned that **restrictive procedures may hamper innovation in AI for health**. A careful **balance is needed between encouraging innovation and appropriate levels of participant protection** but reconciling the two is not always easy. In this context, GFBR participants discussed whether secondary use of existing data warrants REC approval and participant consent. **Countries have different approaches to whether secondary use of data requires REC review and much depends on the context e.g. how specific or broad the original consent was**. It may also depend on whether the data is anonymised, though some GFBR participants expressed concern that de-identification may not eliminate classification of different groups and potential group harms. Many participants commented that regulations in their countries for secondary use of data are not strong so the onus falls upon RECs and regulators to be vigilant where government regulation is lacking.

Necessity of AI

GFBR participants reflected on the hype around AI and the **need to avoid tech solutionalism**. While the promise of AI for health is significant, **critical reflection is required on whether AI is the best or necessary approach to any given problem**. A GFBR participant asked if a cost-benefit analysis was performed for *Vajrahands* and whether AI was really required or whether a human-level intervention might have been more cost-effective. There was no evaluation with respect to effectiveness. The team relied on the literature and WHO recommendation on best practice for handwashing as an infection control measure.

Often it is the people providing solutions (often commercial entities) who decide there is a need and this judgement can be very subjective. Ideally, communities should be involved in defining the problem and solution in the first place. And where a need has been identified work should be done to understand whether there are alternative (non-tech) ways of solving the problem. **Niyoshi** proposed the creation of an evaluation tool for policymakers to measure the need for AI in their jurisdiction from a rights perspective, as one way of creating checks and balances against tech solutionism.

Reliance on AI tools

The hype and potential around AI may make the outputs of AI-based tools appear more reliable than they are. GFBR participants agreed that although AI-based tools have the potential to make better use of existing capacity in resource-limited countries, an expert is still required to make the final decision. It is important to conduct user feedback during implementation research to assess:

- How much the users are internalizing that the tool is for decision-support and not replacing the healthcare professional as decision-maker.
- Feasibility.
- Whether the tool does actually help make decisions quicker and better.

3. Collaborative initiatives and data resources to support AI health research

This session showcased two collaborative initiatives: one that is at the planning stage and one that is more mature with a portfolio of research activities mainly focusing on autonomous systems. Both papers focused on the issue of trustworthiness and were complementary. One initiative is planning how to establish its processes and how to engage communities, while the other is underway and has faced challenges and created best practices. The theme raised the question of how transferable regulatory requirements are from one context to another. For example, is it feasible or indeed right for the EU's General Data Protection Regulations (GDPR) to be enacted and enforced in LMICs? The theme also delved into how responsible, collaborative data sharing practices can be encouraged between public and private organisations in a way that aligns with community values and needs.

Presentation 6: Ethical considerations in implementing the Data Advancing Wellness in Africa (DAWA) Project
Gakii Masunga – Harvard Medical School, USA

The DAWA Project (**DAWA,** Swahili for "Medicine") aims to build innovative technologies to collect and analyse big data, employed to better understand the determinants of cardiometabolic disease and mental health outcomes across sub-Saharan Africa (SSA), while advancing the health and wellness of Africans everywhere. It aims are to: (1) Create an African non-communicable disease big data commons across SSA to be used to generate timely health information and wellness products at varied levels of complexity including at the patient, practice, policy and economic level (2) To study the multifactorial determinants of, and outcomes from, cardiometabolic disease and mental illness in SSA using big data analytics.

Gakki Masunga presented plans for the **D**ata for **A**dvancing **W**ellness in **A**frica Project, a collaboration between four African institutions located in South Africa, Nigeria, Tanzania, Uganda and three US institutions located in Boston, Massachusetts**.** The project responds to the surge in non-communicable disease (NCD) burden in sub-Saharan Africa (SSA), which is on track to overtake infectious, maternal and neonatal causes combined. DAWA will involve data scientists, engineers, software technicians, global health experts, epidemiologists, biostatisticians, and community representatives with a shared aim to advance data science applications for curbing the NCD burden in SSA.

The DAWA Project is considering the following ethical issues:

- **Trust and trustworthiness**: Transparency is required to build and maintain trust with the participants who will contribute personal data to the project. This is especially important in the context of the project being a collaboration with US institutes and given the historical exploitation of the African continent. Having local researchers lead the study could help promote trust and allay concerns people have about participation. There need to be clear benefits for participants and for these to be communicated.
- **Bias**: **Gakki** noted that adaption of algorithms in LMICs is not widespread and the potential risks associated with bias are largely unexplored. The project will collect identifier data for sensitive/protected groups (e.g. tribe, clan or religion) to assess the extent of their inclusion/exclusion. Checks will be performed to ensure that collected data covers all these groups to enhance fairness.
- **Cultural and moral values**: **Gakki** warned that it is unwise to assume that Western individualistic moral viewpoints inform the behaviours and thought processes about justice and fairness in Africa and equally, it is misguided to assume that just one theory could be generalized to the entire continent. The

project seeks to understand the underlying values and moral viewpoints specific to the participants' cultural context as these may inform their health-seeking behaviours.

- **Fair collaborations:** The project will critically examine the power relationship and differential between the collaborative institutions. Mutual trust and respect need to be established between researchers, and participants need to see this has been established.
- **Access to technology**: The project relies in part on the widespread use of cellular phones across Africa. Gakki stressed the need for concerted efforts to ensure women are well represented in the datasets, understanding that women in Africa generally have less access to cellular phones and the internet in comparison to men.
- **Public engagement:** The project will incorporate the perspectives of relevant stakeholders in the design and development of its AI systems, in order to capture issues that are important to them. Focus groups will be held with proposed participants to gauge their level of understanding of the use of AI in health research and provide education to fill knowledge gaps.
- **Datasheets and metrics**: Gakki recommended that the provenance, creation, and use of machine learning datasets should be well documented. Best practice can be embedded to inform the technical design and development of AI tools for health research and to mitigate potential risks (e.g. bias and discriminatory outcomes). The project will identify the most applicable metrics to assess fairness on a case-by-case basis (e.g. demographic parity, equality of opportunity and equality of odds).

PRESENTATION 6: CLARIFICATIONS & DISCUSSION

The collection of significant amounts of data for research – whether using AI or not – raises concerns about privacy and the use of sensitive data. Gakki clarified that people who find the questions too invasive can opt out of participation. While the project wants to recruit a wide range of people, the challenge is to find enough participants who don't have these concerns. Potential participants' privacy and use concerns will be addressed, including through use of local coordinators. The project has an embedded ethics study that is considering how to build the questionnaires, working with policymakers as possible end-users of the data and engaging other stakeholders.

It was noted that in the DAWA Project, as with the ear biometric case study, **the response to people *not* participating due to privacy or burden concerns is that they will be excluded**. This raises a question about who is being excluded as a result, and **what does this systematic exclusion mean to equity and for principles of participant / stakeholder engagement**?

GFBR participants discussed the **challenges of power and science equity regarding the ownership of technology and data in the context of international collaboration**. Initiatives similar to the DAWA Project, such as H3Africa, have come across challenges of data sovereignty. The DAWA Project was encouraged to think about how the idea of the 'commons' underlie the objective of data sovereignty, who controls the data, and about benefit and the role of African intellectual leadership. Ethics requires consideration of **structural injustice** including thinking about what mechanisms already exist on the continent for collecting data and ensuring data sovereignty – and working with those. Otherwise there is a risk of poor governance continent-wide. GFBR participants agreed on the importance of these conversations as part of the discussion and planning from the start, aiming for better data governance across the continent, and actions in response to existing structural injustices.

Presentation 7: Responsible research and development in AI for healthcare: what we are learning from establishing a national collaborative platform in the UK
Kate Devlin – King's College London, UK

The UK's Trusted Autonomous Systems Hub is the world's largest research programme in Trustworthy AI and Autonomous Systems, which was established by three UK universities – Southampton, Nottingham, and King's College London. It is the focal point of the £33m UKRI TAS Programme, involving six TAS Nodes dedicated to the topics of functionality, resilience, security, governance and regulation, verifiability, and trust. TAS serves as a platform for the best practice for regulation and development of autonomous systems and has over 100 international industry partners. Health and AI is a large part of its remit. See the governance paper for more details and examples of health research supported by the Hub.

Kate Devlin presented the development and running of The Trusted Autonomous Systems Hub, a national UK network on trust in AI that foregrounds ethical, responsible and inclusive development. The Hub funds projects, sets up networks, advises on policy, and invites researchers, industry, non-governmental organizations and the public to engage and contribute use-cases/datasets or collaborate on research projects, tech transfer, and training activities. The Hub is built around the core principles of responsible research and innovation with equal attention to equality, diversity and inclusion. It has a strong emphasis on participatory design and is explicitly values-based, with trustworthiness and responsibility as key values.

Kate described the successes, failure and course correction that has taken place in setting up the Hub and explained what being **trustworthy in principle means in practice.** For example, the Hub:

- Has grant criteria that **require projects to be collaborative and interdisciplinary and to centre on stakeholder engagement and responsibility**.
- Promotes **early career leadership**.
- Ensures tangible **ethical approaches**.
- Has developed an **operational framework**.
- **Works with a range of stakeholders**, including industry, academia, end-users and the public.
- Works on **participatory and collaborative project design**, as well as education and outreach, with the goal of being transparent about a future with autonomous systems.
- Undertakes **advocacy and engagement** – providing policy briefings, doing public engagement, responding to government consultations.

The Hub always tries to embed engagement from the start to the end of a project (including acknowledgement of contribution). Engagement is also expected in the projects funded by the Hub. For example, a project assessing mental health apps went out to 21 different universities and students were asked first-hand for their views on the apps. There have been some difficulties with engagement due to delays in getting approval from the UK's National Health Service to undertake engagement work.

Kate reflected that the Hub's clear aims and governance structure have worked well, promoting transparency. Challenges in running the Hub relate to the large scale of the project, lack of integration (especially with government work in this area), academically competing on projects that focus on results (while the Hub also focuses strongly on process), lack of interdisciplinarity from the arts and humanities and the short term nature of core research funding.

THEME 3: DISCUSSION

Data governance

GFBR participants identified a number of challenges for international data resources:

- **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**. It was noted that the African Union is working on governance of this area, focussing on harmonisation of policy frameworks across the region.
- 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.

Challenges of data sharing and data protection legislation

GFBR participants shared their experiences of research with other countries being hampering by data sharing problems. In one collaboration between African countries, no countries would release their data across country borders despite there being a single funder and a collaborative agreement. This reflects a new tendency to prevent data-stealing and misuse due to previous poor experience. The research team had to conduct the analyses at country level then group findings, rather than share the underlying data.

The EU's General Data Protection Regulation (GDPR) is being adopted by several countries outside the EU, including by a number of LMICs. **There were mixed views on whether the GDPR should be the global standard**, with some considering that it should not. It was noted that while several African countries are adopting GDPR, no African country has an adequacy determination by the European Commission, and nor does the US. For researchers willing to share, it can be difficult to negotiate the legislative requirements between the countries.

GFBR participants noted there can be a **conflict between collecting data to help assess the representativeness of data sets (e.g. race and ethnicity) and GDPR's stricter requirements for consent when collecting these 'sensitive categories'**. Some GFBR participants considers that if AI-based research requires consent from every individual who provides 'sensitive' data, it could become unfeasible.

There's a need to recognise public perceptions of owning one's own data, even if this is not legally correct in a particular jurisdiction. Data ownership can be an unhelpful concept: it can be **more helpful to think about control and access in the light of different interests in the data, rather than ownership and property rights** especially given how variable jurisdictions are.

Does the nature of AI research raise any unique issues for data sharing and governance?

Challenges around data-sharing are not unique to AI-based health research. However, GFBR participants agreed that some aspects of AI do accentuate issues of privacy, security and consent:

- The **scale and complexity** of how (and how much) data is used to train algorithms.
- **Ambiguity over whether the development phase of an AI system is characterised as research and as falling within research regulation and requiring consent**.
- 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 **generalisation and transferability of algorithms** between contexts, potentially introducing issues of bias and the need to collect new data to train the algorithms.
- The **interplay of power and decolonisation**. Some participants saw **the use of AI as recolonising research as many LMICs are data-poor** and do not have the capacity to hold, store and analyse data. In the field of AI, data is primarily held by HICs and by private companies.

Fair partnerships and finding responsible methods for sharing data

GFBR participants discussed what it means to have a fair partnership especially when partner countries have different ethics and governance standards and disparities in available resources. For example, there can be competitive concerns which are heightened when some researchers in the partnership have more resources to analyse data. GFBR participants agreed that the many social and political (and commercial) pressures on researchers to publish can be counter-productive to fair data-sharing and partnerships. **Researchers need to be incentivised to embed ethics into their processes and partnerships as well as focusing on outputs**. Material transfer agreements were considered an important element of fair data sharing within partnerships.

Novel data sharing solutions

Data sharing can be constrained by the parameters of data systems. New solutions – 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. GFBR participants considered this an interesting approach, **taking AI to be trained on a series of different data in different places** and using **federated networks** to avoid the need for sharing. 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**.

Public-private partnerships

Fair benefit sharing models are important for all research but especially for public-private partnerships where there is potential for commercial gain. 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 proposal and delivered**. Some GFBR participants thought this role would ideally fall to the REC but recognised that they are too under-resourced to follow-up on benefit sharing commitments.

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 incentivising drugs for neglected diseases).

The competitive nature of the private sector can make public-private partnerships challenging. However, GFBR participants shared examples of public-private projects that have worked well. In Tanzania, data was collected

from farmers by a private insurance company and this data was used to form a new community health care platform, Community Health Fund. It is important to share success stories of public-private partnerships and to see if they can work in other contexts, such as health research.

Figure 2. During a re-cap after Themes 1 ,2 & 3 GFBR participants were asked to identify the main ethical consideration that arose in the talks and discussion.

4. Regulation of data for health research involving AI

The governance of AI is inextricably linked to the governance of data, on which algorithms rely. In its report 'Ethics and Governance of AI for Health' the WHO makes recommendations on data governance. The report recognises the complicated intersection of national and regional laws and that the amount of data required to train AI, data poverty in many countries and the importance of representative data all add to the challenges in how data governance is approached. This session focused on data governance, specifically in the African context. The first paper analysed the governance of cross-border transfer of data between countries. The second paper evaluated Uganda's data governance regulation against the recommendations in WHO's report.

Presentation 8: Governance of cross-border transfer of data in sub-Saharan Africa
Nezerith Cengiz – Stellenbosch University, South Africa

Data access, sharing and transfer between countries are crucial to effectively managing current and future health pandemics. This requires high-quality, comprehensive datasets that can inform policymaking and enhance healthcare decision-making. Data access and sharing, however, raises questions about personal privacy, the adequacy of governance mechanisms to regulate cross-border data flows, and ethical issues relating to the collection and use of personal data in the interests of public health. This paper explored governance considerations that ought to apply to the collection, transfer, and use of data; and provided an overview of the prevailing data sharing governance landscape in SSA's most research-intensive countries.

Nezerith Cengiz presented an assessment of data sharing regulations across the five most research-intensive countries in SSA: South Africa, Nigeria, Kenya, Ethiopia and Uganda. She explained that sharing large datasets is required for the development of AI tools and could help strengthen research capacity in the region. However, **countries in SSA have diverse approaches to data management and protection which impacts on data sharing**. Further, data protection legislation tends to be generalised and is not specific to AI. The presentation highlighted the legal protections afforded to data subjects in the respective countries regarding the cross-border transfer of their personal data. Protections include data subject consent, notification to the data protection authority and a requirement (or not) for data protection equivalency in the recipient country.

Nezerith explained that key limitations in data sharing in SSA include:

- Inability to gain timely access to relevant data.
- Lack of trust over data use.
- Academic competition.
- Insufficient data transfer regimes.
- Different data needs of different stakeholders which makes a common approach difficult.
- Scope of using data in public health emergencies being limited by incomplete and poor quality data.

The Africa Union Commission is developing a data policy framework for Africa with the aim of addressing some of these limitations and harnessing digital technologies and innovation in an attempt to bridge the digital divide. While this work is ongoing, there is currently no uniform law, meaning cross-border transfers have to be evaluated on case by case basis.

Nezerith proposed that:

- Standard contractual provisions and templates for cross-border data transfers should be developed by data protection authorities in Africa.
- SSA countries should strengthen their digital infrastructure for capturing and storing data to help build analytical capacity.
- The principles of transparency, fairness, and accountability would help with the establishment of a reliable and accessible digital ecosystem in SSA to underpin AI-based health research.

Presentation 9: Regulation of health data for AI in Uganda
Harriet Nankya – Makerere University, Uganda

In 2021 the WHO issued a report recommending policies, principles and practices for ethical use of AI for health. One of the key principles endorsed was protecting human autonomy; which requires, among other things, the protection of privacy and confidentiality of data and obtaining valid informed consent through appropriate legal frameworks for data protection. This paper compared the provisions of Uganda's data protection laws to the WHO recommendations.

Harriet Nankya presented an assessment of Uganda's data protection laws with reference to the WHO's recommendations on the regulation of health data for AI. She explained that AI systems in Uganda have been subject to sector-specific laws or subject-specific guidelines such as data protection acts, cyber-security laws and anti-discrimination regulations. This piecemeal approach has created regulatory gaps that have ethical implications.

Uganda's Data Protection and Privacy Act, 2019 and the Data Protection and Privacy Regulations, 2021 are intended to support the protection of privacy during the collection, processing and storage of personal data.

The laws mirror a number of international instruments including the EU GDPR, and UN Declaration of Human Rights. The Ugandan Act also guarantees the protection of privacy in the digital world. **Harriet** explained how the provisions of the Act and Regulations, and the remit of the independent Personal Data Protection Office, align with WHO's recommendations. She also highlighted gaps and points of tension e.g.:

- The laws focus on individual rights and protection and do not provide a mechanism for community oversight of data.
- The laws do not specify that 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.
- There is no specific reference to impact assessments of AI technologies – but there are requirements for data protection impact assessments.

Harriet concluded that Uganda's current regulation and principles contain important rights and could help shape how AI for health research is developed. But there is a need for more specific safeguards for AI and processing big data. In this context, she recommended that The National Information Technology Authority of Uganda should elevate its regulatory function to protect the integrity of personal data used in AI-based health research.

THEME 4 DISCUSSION

Contextualising global recommendations

GFBR participants discussed the **advantages of having global recommendations from WHO that can be customised by country but also the challenges of achieving local ownership** and knowing who within the country should act as advocate and disseminate the recommendations. **Capacity building is required in-country to assist local implement of the report's recommendations** (including policy makers, RECs and developers of AI). Language was identified as a potential barrier, as some countries do not have words to explain AI.

Patchwork of regulatory and governance approaches

GFBR participants shared their experience of setting up data protection laws in their countries. Long processes often resulted in **laws that were outdated by the time they came into force**, or the law addressed the issue too narrowly. Some participants proposed that **guidance and standard operating procedures can be quicker to implement and have more influence**, in particular in the context of their country's Ministry of Health having limited resources to develop regulation. This calls into questions whether legal mechanisms which require enforcement are the best solution or if it would be more effective to **focus on guidance and culture change to promote an ethical culture across the AI health research ecosystem**.

Although there appears to be a global trend towards hard law for data protection, currently, the only specific hard law for AI is the European Union's AI Act. The law assigns applications of AI to three risk categories: banned AI (e.g. that can be weaponised), high-risk which are subject to specific legal requirements (e.g. technology can help in some limited circumstances, but has the potential to be abused so should be highly regulated), and unregulated AI. GFBR participants questioned who can input on how specific uses are classified, the implications of these classifications and how they are operationalised for inter-country collaborations.

GFBR participants expressed differing views on how AI used in health research should be governed:

- Some preferred the use of **soft law – e.g. guidelines – with a strong regulatory body to monitor compliance**. This was seen as being nimbler, and quicker to respond to areas of emerging technology.
- Other participants proposed **hard law to define the systems for regulation rather than the actual activities – enabling those regulatory systems then to be soft and more responsive to changes in technology**, balancing different needs, establishing guidelines etc.
- The US AI Incident Database, a repository to record the times AI has caused harm, was given as an interesting example of non-regulatory monitoring.
- Some participants suggested that **LMICs should have laws to protect against data colonisation in the context of AI**.

Some unique challenges for the regulation of AI in health research were identified:

- How to **balance the needs of innovation in AI-based health research and potential pit falls** e.g. quality of data, risk of bias.
- **Scale and scalability** of AI and unintended consequences of amplification.
- Responding to the way that AI can **make choices** from patterns in data and **learn**.
- How to take account of **under-representation as a serious harm** in the data used for increasingly common purposes.
- 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.
- The 'grey area' of **what counts as research** e.g. whether development of a health app counts as research and falls under health research and medical device regulation.
- **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.

Going beyond the requirements of law

There are many pieces to the governance jigsaw (e.g. RECs, access committees, funders etc.). 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 the technology developers and researchers be encouraged to detect and fix bias? **GFBR participants saw a role for RECs to assess bias, or at least to check that these issues have been addressed**.

GFBR participants also saw a role for self-regulation. **Incentives are required for important research like validating an AI model**, which may not be a funder's priority. Investing in **data versioning, data transfer agreements and data destruction policies** was also considered important – so researchers have a record of when they worked with whom on what data, and the nature of the data set is well characterised. For a fast-moving field like AI, it's important to be creative and **leverage existing systems** – e.g. continuous education, professional adherence, codes of conduct – rather than creating completely new mechanisms.

AI exceptionalism

The cases presented in this session addressed laws on data protection that are not specific to AI health research. GFBR participants discussed whether AI shifts the parameters in terms of anonymization and whether

it poses a greater risk of re-identification via machine learning, in comparison to other kinds of health research. In addition, **is there a moral difference between data-sharing for AI as opposed to other secondary use of data for health research**? How have researchers dealt with representativeness of data in other areas of health research and are we requiring a different standard for AI (and if so, why?). While participants did not have immediate answers, they cautioned against AI exceptionalism.

5. Issues associated with research ethics frameworks and ethics review

This session focused on the policy context and research ethics frameworks for AI based health research. Two presentations assessed the policy environment in Egypt and Kenya. The third presentation focused on the incorporation of environmental impact assessments into research ethics frameworks. Gaps in REC capacity were also highlighted, along with consideration of how the gaps might be filled and what other committees or oversight mechanisms might be needed.

Presentation 10: Recommendations for the development of ethical guidelines for AI-related health research in Egypt *Ahmed Samir Abdelhafiz – National Cancer Institute, Cairo University, Egypt*

Egypt is taking steps to improve its capacity building capabilities in the field of AI. A National Council for AI was founded in November 2019 by the Egyptian government as a partnership between the government, academic institutions, and leading companies in the field of AI. Healthcare is among the initiative's priority sectors. In 2020 the Egyptian parliament issued a data protection law and a separate clinical research law. However, neither are specifically focused on AI research and the clinical research law doesn't cover many aspects of pre-clinical research. This paper focused on consent, commercialization of data, benefit-sharing, and the role of RECs as key guardians of ethical guidelines for AI-related health research.

Ahmed Samir Abdelhafiz described the governance context for AI health research in Egypt and identified the following ethical challenges:

- **Informed consent**, including:
    - The difficulty of predicting who will have access or how data will be used.
    - Explainability, given the complexity of AI research, its core concepts and terminology.
    - The potentially greater risks of re-identification than in other kinds of research.

  Ahmed recommended **qualitative research to explore participants' attitudes and preferences regarding consent model(s) for AI research**. He articulated the pros and cons of some models (e.g., specific consent, dynamic consent) but noted it is premature to recommend a specific model since little is known about which model best aligns with public preference and is scientifically practicable. For example, the high rate of phone number changes in some LMICs was cited as a potential barrier for implementing dynamic consent.
- AI research often involves public-private collaborations raising issues about the **commercialization of data and the implications for benefit sharing**. Ahmed recommended that the Egyptian **government should negotiate benefit sharing terms with technology partners or set the boundaries of acceptable practice**.
- Ahmed recommended **an evaluation of the knowledge, perceptions and attitude of REC members to AI research**. This assessment should inform training which could be co-ordinated by The Supreme Council for the Review of Clinical Research Ethics and The Egyptian Network of Research Ethics Committees.

Presentation 11: The proverbial black box that is ethics of AI in global health research: Are our Kenyan RECs well equipped to review ethics?
Brenda Odero – Strathmore University, Kenya

In Kenya, the Data Protection Act provides a regulatory framework for data collectors, processors, and data participants. This is supported by the National Guidelines for Ethical Conduct of Biomedical Research Involving Human Participants in Kenya (2020), which offers some ethical guidelines on how to handle data while protecting participants. The existing guidance on the application of digital innovation by the National Commission for Science, Technology, and Innovation (2021) regulates the use of digital technology, including AI in science, technology, and innovation in Kenya. However, all these regulations and guidelines give minimal attention to the use of AI in health research. This paper examined the gaps in the Act and National Guidelines as a resource for reviewing AI health research in Kenya.

Brenda Odero assessed the challenges RECs face when reviewing AI-based health research protocols, with reference to Kenya's current laws and guidelines and the Emanuel et al.'s international research ethics framework[4]. Kenya ranks third in sub-Saharan Africa with a score of 45.5% according to the Government Artificial Intelligence Readiness Index 2021 by Oxford Insights. This high level of readiness is due to government and researchers who are pushing the AI agenda often with mentorship from the US and UK, and through the influence of the private sector.

Brenda recommended ways that traditional research ethics guidelines and procedures may be adapted to respond to AI-based health research:

- **Community engagement** should be part of the research proposal and be assessed during REC review. In particular, **community consultations can be used to gather feedback during Algorithmic Impact Assessments, as a way of promoting accountability and trust**. It was acknowledged, however, that community engagement has many challenges e.g. who to engage, how, for what purpose, etc. (see Theme 2).
- **Collaborative partnerships** are required e.g. ad hoc committees to conduct stakeholder engagement, and experts in AI research to advise RECs.
- **Data-sharing agreements** should be provided for REC review to ensure that data rights are understood and respected and to promote transparency and fairness.
- The **social value** of research should be considered, along with potential **risk and benefits**, for example, by using an Algorithmic Impact Assessment. RECs should review these assessments and ensure that the inclusion of any vulnerable groups is justified. **This would be a new role for RECs to think about the future impacts and social value of research on society**.
- RECs should assess the **study design and whether AI is required and appropriate for the context**. This was recommended on the basis that research AI models are predominantly designed using western epistemologies and worldviews, which can limit their applicability to African contexts. The use of western languages, graphics, and aesthetics inherent in AI designs may contradict the Kenyan reality especially in clinical research. The REC should also review the **selection of the study population** to promote inclusivity, diversity and relevance of the data to be collected to the study aims, to mitigate bias.

[4] Emanuel, E.J., Wendler, D., Killen, J., & Grady, C. (2004). *What makes clinical research in developing countries ethical? The benchmarks of ethical research*. The Journal of infectious diseases, 189(5), 930-937

- **Ongoing respect for participants and communities** should be shown through the dissemination and utilization of results (not just publishing).

PRESENTATION 11: CLARIFICATIONS & DISCUSSION

GFBR participants reflected on the feasibility of RECs taking on the proposed roles, especially when **many RECs in LMICs are under-resourced and have inadequate training on the science and ethics of AI**. The onus should also fall on researchers and AI developers to be thinking through the ethical issues prior to arriving at RECs. This **requires training of AI developers in research ethics**. Some suggested the adoption of short online training with a certificate which is required for research proposals to be considered by RECs to help ensure a common baseline of ethics knowledge.

Some GFBR participants suggested that RECs should have some understanding of AI science as 'good science is ethical science'. Others saw a role for **ad hoc AI experts to input during the REC review or having a mechanism for the technical side of AI-based research to be signed off prior to arriving at the REC** – similar to a data management plan that is prepared in advance but is part of the REC's review.

Some of the roles – like assessing algorithmic bias – often only come to light during implementation so would be difficult for a REC to assess. Some aspects may be better assessed by another body akin to a data safety monitoring board.

Presentation 12: Reframing research ethics frameworks to include environmental sustainability

Gabrielle Samuel – King's College London, UK

Dominant research ethics paradigms have historically revolved around ethics principles that are concerned with the protection, rights, safety, and welfare of individual research participants. Emanuel and Weijer (2005) emphasised an ethical principle of 'respect for community' to sit alongside other more individually focused ethical principles, drawing attention to understanding the socio-political impact of research on communities[5]. While considerations of community harm have expanded the moral status considerations of research ethics frameworks, this doesn't capture the adverse environmental (and consequential human health) harms generated from the manufacturing, use, and waste disposal of equipment, tools, and technologies associated with research. This paper made a case for why environmental considerations are important to include in an AI research ethics framework.

Gabrielle Samuel explained that research ethics frameworks that are currently used for AI research lack a normative consideration for adverse environmental impacts of the research. While the use of AI is considered a potential enabler for many sectors, including healthcare and access to care, it is not a no-cost solution. Adverse environmental impacts associated with AI include:

- Very high **electricity** demands to power and cool equipment in data centres and to power the training of algorithms being developed during health-related research.

[5] Emanuel E.J., Weijer C. Protecting communities in research: from a new principle to rational protections. In: Childress JF, Meslin EM, Shapiro HT, eds. *Belmont Revisited: Ethical Principles for Research with Human Subjects*. Washington DC: Georgetown University Press; 2005

- Relying on **mining**, including rare minerals necessary to manufacture digital technologies to store and process data. This raises environmental justice concerns when mining adversely impact biodiversity in local mining areas and, in unregulated environments, individuals who live and/or work in or near mines can also be exposed to environmental harms that promote poor health outcomes.
- **Large amounts of e-waste** due to the obsoleteness of software used to run AI and digital research. Only around 20% are recycled and the rest typically ends up in LMICs in e-waste dump sites, with impacts on the health of people who recycle from the sites.
- **Water usage** required to cool data centres.

There has been no real discussion about environmental impacts in research ethics frameworks and the moral gaze has focused only on humans. **Gabrielle** explained that we could keep the moral gaze on humans, and take account of environmental impacts on human health or we could **consider the environment as having intrinsic value, promoting ecocentric ethics**. She proposed modification to the substantive principles in Emanuel et al.'s (2008)[6] research ethics framework to include considerations associated with the adverse environmental impacts of AI research:

- **Social value** should explicitly include risk and benefits to the environment and weigh these against the potential health benefit to individuals/communities.
- **Respect for persons, communities, and environment**. Stated in the previous framework as 'respect for participants' and 'community partnership', the addition of 'respect for environment' means being attentive to the adverse environmental impacts that can emerge from using digital technologies during research and taking steps to reduce them.
- **Fair collection, storage, and use of data** (previously 'fair participant selection'). Researchers must be cognisant of the composition of datasets they use, and any possible biases (what categories are present/missing in the data? How is data categorised and by whom? What implicit assumptions come from these categories? How diverse is the data and what are the limitations of the datasets being used?). Attention should also focus on benefit sharing of research outcomes.
- **Fair consideration of those affected by the research process** (previously 'fair participant selection' - additional recommendation). Consideration should be given to the environmental justice issues associated with those involved in the manufacture, use and disposal of digital tools used during the research process, who are also the least likely to be beneficiaries of the research.
- **Favourable risk/benefit ratio**. Risk benefit considerations for AI research need to include not only research participants and/or those affected by the research outcomes, but also those affected by the manufacture, use and disposal of digital products.

This revised framework can be used by researchers, RECs and also, importantly, policy-makers. Policy-makers need to understand that increasing digital efficiency will not necessary decrease adverse environmental effects, because it will likely increase use. The same can be said for open science practices. Researchers sharing data, rather than generating and reproducing data from scratch, could reduce environmental impact but more likely the **greater efficiency drivers will promote greater data use** and the opposite outcome. Rather than relying on increasing digital efficiency, **policy-makers should engage with questions about where and how data are stored and how algorithms can be optimized for environmental considerations**. Resources should also be shared more evenly with more low-tech research focusing on the social determinants of health instead of

[6] Emanuel EJ, Wendler D, Grady C. An ethical framework for biomedical research. In: E. J. Emanuel, C. Grady, R. A. Crouch, R. Lie, F. G. Miller, D. Wendler, eds. *The Oxford textbook of clinical research ethics*. Oxford, UK: Oxford University Press; 2008:123-35.

focusing on big tech solutions that may have a higher environmental impact. **All health research – not only AI-based – should be critically examined for its environmental effects and not be seen as intrinsically good**.

PRESENTATION 12 CLARIFICATIONS & DISCUSSION

Carbon calculators were discussed as a tool for assessing environmental impact. While useful, they are only one factor and they can hide the nature of the data behind the calculator (e.g. data sets may be out of date/incomplete/irrelevant with uncertainty such as whether the particular energy consumption is based on fossil fuel or renewable energy). In addition, they are carbon-centric and don't cover other environmental aspects. At the moment there's an unclear relationship between carbon measurement and how this leads to change in practice.

GFBR participants agreed on the **need to integrate environmental and social sustainability in impact assessments into research ethics frameworks,** highlighting the importance of multi-disciplinarity and learning from environmental ethics. Social justice issues are prioritised less than environmental impact and metrics are lacking. While metrics are needed to help articulate the issues, we also need to be alert to how they hide messiness of data, and how they are used. Metrics need to be accompanied by quantitative and qualitative evidence for contextualisation.

Annex 1: Background information on GFBR and meeting content

The Global Forum on Bioethics in Research (GFBR) is the principal global platform for debate on ethical issues pertaining to international health research. Its core aims are to give voice to low- and middle- income country (LMIC) perspectives in debates about global health research ethics and to promote collaboration.

The Forum meets annually to address a specific topic in research ethics. The meeting is case study based to enable participants to understand the practical issues 'on the ground' in addition to broader ethical and governance questions. 16 case studies and 33 governance papers were submitted for this meeting. 5 cases studies and 7 governance papers were selected for oral presentation at the meeting (see links throughout the report). Ten other case studies and governance papers were selected for short Pecha Kucha presentations.

Pecha Kuchas

1	**Analysing a local imbalance of power ethics: University of Ghana vs. Data Commission** Athanasius Egyarkoh Afful, University of Ghana, Ghana
2	**Ethical concerns in the use of AI in patient safety research: an examination of the adequacy of Nigerian laws** Dorcas Akinpelu, University of Ibadan, Nigeria
3	**Who minds the machines? Developing a governance framework for pre-market authorisation of responsible AI applications in healthcare in South Africa** Irvine Sihlahla, University of Kwazulu-Natal, South Africa

4	**Future nanomedicines: building a regulatory framework for the first in-human nanoswarm cancer clinical trial** Matimba Swana, University of Bristol, UK
5	**International AI research: the issue of moral pluralism** Serene Ong, National University of Singapore, Singapore
6	**A shift to openness: open consent and open science in AI health research in South Africa** Meshandren Naidoo, University of KwaZulu-Natal, South Africa
7	**A regulatory framework for AI-health research in the Caribbean** Derrick Aarons, The Caribbean Public Health Agency, Jamaica
8	**How to translate universal principles to local realities: the Chilean experience in AI** Sofia Salas, Clínica Alemana Universidad del Desarrollo, Chile
9	**Developing a governance framework for data science health research in Nigeria** Oluchi Maduka, Center for Bioethics and Research, Nigeria
10	**Adaptability of India's Health Data Regulations** Rupanjali Karthink, Duke University, USA

Participants are selected through a competitive process and come from a diverse range of disciplines, countries and career stages. Awards are available to LMIC colleagues to cover travel and accommodation.

All participants were encouraged to report the meeting recommendations in their home countries and to continue the discussion in their local context. Fellowships were available for LMIC participants to explore issues that arose during the GFBR meeting in greater detail, establish new collaborations, and develop new ideas for resolving issues that could not be resolved at the meeting itself. 18 fellowship applications were received after the meeting and 6 were selected for funding. For details see: gfbr.global/fellowships.

Annex 2: List of abbreviations

GFBR: Global Forum on Bioethics in Research
LMIC: Low- and middle-income country
HIC: High income country
H3Africa: Human Heredity and Health in Africa
REC: Research ethics committee
WHO: World Health Organization
OECD: Organisation for Economic Co-operation and Development

SoC: Standard of care

AI: Artificial intelligence

ML: Machine learning

Acknowledgements: We are grateful to the presenters whose work forms the basis of this report and want to thank all the GFBR participants for their engagement with the theme and each other. Thanks in particular to the session and breakout group chairs, and to the notetakers.

GFBR funders*:* We would like to extend our thanks to our local host and funder, the South African Medical Research Council for their support in the preparation of the meeting. Many thanks to GFBR's other funders: Wellcome; the National Institutes of Health; the UK Medical Research Council.

Author: Adrienne Hunt

Case study and governance paper write-ups and presentations from this meeting are available on the GFBR website.